Chapter 3

A Flexible Estimator for Counts with a Dummy Endogenous Regressor[1]

3.1 Introduction

Count data models try to explain the behavior of discrete and non negative dependent random variables (Winkelmann and Zimmermann 1994, Cameron and Trivedi 1998 provide excellent surveys). Applications of these models include health care utilization, recreational demand, number of patents or bankruptcy among others. One of the most popular models for count data assumes that the discrete variable follows a Poisson probability function. However, despite its popularity, such a requirement often fails to hold. Among other features, the Poisson model imposes a restriction of equidispersion (i.e., the conditional mean should be equal to the conditional variance) which most data sets fail to accommodate. A popular solution in the literature has been to include a term which accounts for unobserved heterogeneity. When this random variable follows a Gamma distribution, such an extension leads to the widely known Negative Binomial (NB) model (Hausman, Hall and Griliches 1984, Cameron and Trivedi 1986).

Another customary characteristic of count variables is the high relative frequency of zeros. Unfortunately, the NB distribution does not show enough flexibility to accommodate this feature. Therefore the literature has moved to more flexible specifications

[1] This is a joint work with Andrés Romeu.

that could solve these problems. A non exhaustive list includes hurdle models (Mullahy 1986, Pohlmeier and Ulrich 1995), semiparametric (Gurmu *et al.* 1996,1998,1999), finite class models (Deb and Trivedi 1997), Univariate Poisson Polynomial models (Cameron and Johansson 1997) and Negative Binomial Polynomial model (Creel 1999). In general, these estimators have been shown to work better than the standard NB model in terms of fit and information criteria.

All these approaches do not consider the case when a dummy variable is endogenously determined. Our model tries to combine both the flexibility required to adequately fit count variables and the problems appearing in the presence of a binary endogenous regressor. Such a circumstance typically may hold when the unobserved heterogeneity is correlated with some of the regressors. If it was ignored we may get biased estimates of the parameters of interest since we cannot isolate the effect of the regressor on the distribution alone.

The previous literature that has dealt with a dummy endogenous regressor in a count data setting can be divided in two branches. The first one (Mullahy 1997, Windmeijer and Santos Silva 1997) do not need to assume a distribution for the unobservables and use GMM techniques based on the conditional mean. In order to obtain consistent estimates, the analyst needs to find a function of the data with conditional expectation equals to zero.[2] The only proposed functional form for the conditional mean that seems to preserve positiveness and allow for such a transformation is the popular linear exponential. On the contrary, if one specifies the distribution of unobservables, such a transformation is not needed since the unobservables can be integrated out, and this allows for a variety of functional forms. Consequently, although these approaches are distribution free, they need strong assumptions on the functional form of the conditional mean. In light of the findings of the literature on exogenous regressors, this assumption is likely to be inadequate for some datasets, for instance for those with excess of zeros. The second branch (Terza 1998) also sticks to the linear exponential specification and in addition bivariate normality is assumed. The framework proposed by Terza allows for a variety of functional forms specification since the unobservables will be integrated out. However one (TSM) of the two estimators proposed by Terza (1998) require the linear exponential specification and the second (WNLS) both the linear exponential mean and the Poisson distribution of the counts. Winkelmann (1998) also make extensive use of the assumption of bivariate

[2] Under assumptions regarding the expectation of the unobservable conditional on instruments.

normality to develop a count data model with endogenous reporting.

The lack of flexibility in the functional form of the conditional mean contrasts with the importance that has been given to the conditional mean in linear models of selection. According to Vella (1995) in a survey devoted to selection models: "there is a growing feeling that the parametric procedures perform well if the conditional mean of the model is correctly specified". In fact, Newey *et al.* (1990) concludes that "the specification of the regression function and the set of instrumental variables appears to be more important than the specification of the error distribution for these data". Newey *et* al (1990) and Blundell and Windmeijer (2000) do not find departures from the bivariate normality assumption and Vella (1995) just finds relatively weak departures with causes negligible differences in the point estimates.

On the basis of this background, and given the flexibility needed to adequately fit count data models, there is room for a model that gives enough flexibility to the conditional mean and other moments when dealing with an endogenous dummy variable, even if the distribution of unobservables need to be specified. Notice from above that by specifying the distribution of unobservables, we have access to a wider variety of functional forms for the model.

However estimation of the model using only some of the first moments, is not likely to be successful. Heckman *et al.* (1990) developed a method of moment estimator for a flexible model, a mixture of exponentials. According to them "alternative methods based on maximum likelihood appears to be much more promising". Deb *et al.* (2000) replicated the exercise for a finite mixture count data model with identical conclusion. It is not strange that the first moments give very little information about the data generating process when one is dealing with a flexible specification. That is why we have chosen maximum likelihood instead of method of moments.

The starting point is the Terza (1998) model which is introduced for expositional purposes in section 2. We will concentrate in the specification of the conditional distribution for the count. Under a Poisson specification, the parameters may be estimated using full information maximum likelihood (FIML). Since one could be interested in knowing whether this parametric choice is correct, we also show here how to compute goodness of fit measures.

In section 3, we introduce flexibility assuming that the count follows a polynomial expansion over a baseline Poisson density, instead of using a simple Poisson or Negative

Binomial distribution. This approach extends the flexible model of Cameron and Johansson (1997), who in turn adapted the original Gallant and Nychka (1987) model, to deal with endogenous binary variables. This extension is based on the fact that the baseline density already accounts for some of the unobserved heterogeneity. Hence, we expect that a low degree of the polynomial would be enough to provide a good fit. With a linear exponential conditional mean it is relatively straightforward to recover consistent estimates of the impact of regressors. This is not the case for polynomial expansions since the first moment is not log-linear. This is why we also discuss how to recover equivalent estimates of elasticity measures.

In section 4, we test our model using two data sets already analyzed in the literature: the first one is a data set on the demand of trips by households, previously analyzed in Terza (1990, 1998). Here, the Poisson model fails to accommodate the shape of the empirical distribution mainly for the first counts of the support. Instead, our flexible model is able to adapt to the observed data and significantly improves the fit. We also report consistent estimates of the mean effect of the dummy on the counts. The second example confronts our estimator with data showing an even higher degree of non-Poisson behavior, evidenced by an important overdispersion and relative excess of zeros. The data appear in Deb and Trivedi (1997) who analyze the determinants of the number of physician visits by the elderly using a mixture of Poisson densities. These authors acknowledge that possibly some of the regressors could be correlated with unobservables but minimize its impact and do not correct their model accordingly. Our main finding in this case is that a good fit can also be achieved using a polynomial expansion in a model that explicitly deals with the endogeneity problem.

3.2 Count data models with endogenous dummy regressors

The baseline model is the one proposed by Terza (1998). The count dependent variable for the $i-th$ individual, y_i, takes only non negative integer values. Its probability function $f(y_i|x_i, d_i, \varepsilon)$ depends on a binary variable ($d_i = 0, 1$), a vector of covariates (x_i) and a latent random variable ε. The model for the binary variable is assumed to be generated by the process $d_i = 1$ if $z_i'\alpha + v_i > 0$ and $d_i = 0$ otherwise where z_i is another vector of covariates for individual i, α is a conformable vector of parameters and v is an error term. It is assumed that conditional on the exogenous variables $w = (x, z)$, the vector $(\varepsilon,\ v)$ follows a bivariate normal distribution with zero mean and covariance matrix:

$$\Sigma = \begin{pmatrix} \sigma^2 & \sigma\rho \\ \sigma\rho & 1 \end{pmatrix}. \tag{3.1}$$

The joint density for the observations of the pairs $(y_i,\ d_i)$ conditional on covariates w_i is given by

$$\begin{aligned} f(y_i, d_i|w_i) &= \int_{-\infty}^{+\infty} f(y_i|x_i, d_i, \varepsilon) f(d_i|z_i, \varepsilon) dF(\varepsilon) = \\ & \int_{-\infty}^{+\infty} f(y_i|x_i, d_i, \varepsilon) \left[d_i \int_{-\infty}^{z_i\alpha} f(v|\varepsilon) dv + (1 - d_i) \int_{z_i\alpha}^{\infty} f(v|\varepsilon) dv \right] dF(\varepsilon). \end{aligned} \tag{3.2}$$

Using that $v = (\rho/\sigma)\varepsilon + u$ where $u \sim N\left(0,\ \sqrt{1-\rho^2}\right)$ is independently distributed with respect to ε we have that

$$\begin{aligned} f(y_i, d_i|w_i) &= \int_{-\infty}^{+\infty} f(y_i|x_i, d_i, \varepsilon) \left[d_i \Phi^*(\varepsilon) + (1 - d_i)(1 - \Phi^*(\varepsilon)) \right] dF(\varepsilon), \quad (3.3) \\ \text{where } \Phi^*(\varepsilon) &= \Phi \left[\frac{z_i\alpha + (\rho/\sigma)\varepsilon}{\sqrt{1-\rho^2}} \right] \end{aligned}$$

and $\Phi(.)$ denotes the cumulative distribution function of a standard normal. In his application, Terza (1998) performs a two-stage method of moments (TSM) based on deriving $E[y_i|w_i,\ d_i]$. He assumes the standard linear exponential specification for the mean of the count variable, that is,

$$E[y_i|x_i, d_i, \varepsilon] = \exp\{x_i\beta + d_i\gamma + \varepsilon\}. \tag{3.4}$$

This moment equation can not be used because of the unobservability of the ε. However, after some algebra an appropriate first order moment conditional on observables can be derived and a Heckman (1978) type estimator may be computed. Moreover, since the estimation errors are not homoskedastic Terza (1998) proposes then to use a Weighted Nonlinear Least Squares (WNLS). This WNLS requires a specific assumption about the probability function of the count variable loosing some of the robustness in the initial TSM approach. He presents estimates for the Poisson case, while the negative binomial is also suggested.

We will use Terza's model and FIML as a benchmark (say PFIML model), keeping the assumption of a Poison density, i.e., $y_i|x_i,\ d_i,\ \varepsilon \sim P(\lambda_i)$. Although it is computationally harder, this method presents some advantages with respect to TSM and WNLS.

First, the efficiency gains issue is well known if the restriction on $f(y_i|x_i,\ d_i,\ \varepsilon)$ is true, since the FIML will asymptotically reach the Cramer-Rao lower bound. Notice that in the particular case of NWLS, robustness is not a comparative advantage of the previous since we need to assume a Poisson density for $f(y_i|xi,\ d_i,\ \varepsilon)$ either.

Second, all the parameters are separately identified, more specifically ρ and σ. Given that the variance covariance matrix in (3.1) needs to be positive definite, we reparametrize the model in such a way that we restrict the estimate of ρ to be between -1 and 1, and standard errors for this parameters can be obtained using the delta method. This feature was not directly available in the TSM or WNLS approach where parameter could take values outside the bounds. Doing inference about ρ is important since a simple t-test for the exogeneity of the binary variable is readily available and because ρ may have an appealing structural interpretation. For instance, if the count variable represents visits to doctors and the dummy indicates insurance status, then a positive ρ is an indication of adverse selection in the insurance market. On the contrary, negative ρ could indicate cream skimming by insurance companies (Coulson *et al.* 1995). Moreover, as we will see later, the identification of ρ and σ will play a role in obtaining predicted frequencies of counts.

Third, a formal test for the Poisson assumption, conditional on the other assumptions of the TSM model (i.e., the joint normality and the linear exponential specification of the conditional mean of the count) can be performed. The Poisson FIML provides under the null hypothesis, the asymptotically efficient estimate required to perform a Hausman specification test of the null of Poisson distribution against exponential mean models where the consistent estimate is given by the TSM. It is also possible to go further and test jointly all of the distributional assumptions, i.e., the Poisson and the bivariate normal distributions- through a Hausman test. This could be done using the PFIML and a consistent estimator of the conditional mean of the count obtained by the Generalized Method of Moments (GMM), as suggested by Windmeijer and Santos Silva (1997), Mullahy (1997) and Grogger (1990). This test requires the availability of convenient instruments.

Finally, FIML allows one to obtain the expected frequency for different values of the count variable and compare it with the observed frequencies. This is needed when building goodness of fit measures that have been used by Gurmu and Trivedi (1996) and Cameron and Johansson (1997) in models which ignore the problem of endogeneity. This cannot always be done using the WNLS, since nothing ensures that the estimates of ρ are between -1 and 1. Also this technique is particularly useful to detect the excess of zeros problem. Moreover, Andrews' goodness of fit test (Andrews, 1988a, 1988b) can also be computed on the basis of the differences against fitted and expected frequencies. These statistics have been used in a count data context without endogenous regressors by Deb and Trivedi (1997). We will discuss the basic issues here.

Let us partition the range of the count variable in J intervals where $c_1 \geq c_2 \geq \ldots \geq c_{J-1}$ are the endpoints. The observed frequency p_j of the interval $j = 1, 2, \ldots, J$ is given by

$$p_j = \frac{1}{n} \sum_{i=1}^{n} 1[c_j \leq y_i \leq c_{j+1}],$$

where 1[.] is the indicator function. The expected frequency for the interval $j, \widehat{p}_j$, requires some more computation. If the regressor d_i was uncorrelated with the errors ε then we could use that $f(y|w,d) = \int f(y|w,d,\varepsilon)dF(\varepsilon)$, to compute the frequency of count j conditional on regressors and then average. This is not possible any more since under correlation we need to integrate with respect to the density of conditional also on d_i. Instead of deriving this conditional density a much simpler method is to get the marginal probability of the count variable as

$$f(y|w) = f(y,1|w) + f(y,0|w). \tag{3.5}$$

Consequently one would use $f(y|w)$ estimated to add over the range points of y in every interval j and then average over the whole sample and get $\widehat{p}_j$. With this in mind, a very simple goodness of fit measure is given by the sum over j of the absolute differences $\Delta_j = |p_j - \widehat{p}_j|$. The goodness of fit test is basically a moment conditions test where we use the fact that $\Delta_j \to 0$ almost surely under the law of large numbers (see Andrews 1988a, 1988b for more details). Numerical integration is needed at some steps of the implementation. The reader interested reader may consult the appendix on computational methods.

3.3 Polynomial Poisson Full Information Maximum Likelihood (PPFIML)

As we will see later in the examples, FIML estimation using the Poisson is likely to be inadequate. The consistency of the estimates relies on every one of the four basic assumptions: a Poisson density for the distribution conditional on the unobserved heterogeneity, a linear exponential specification for the conditional mean, the bivariate normality of the error terms and the linear structure of the model for the binary variable d. We will concentrate on relaxing the first two assumptions: the Poisson density and the specification for the conditional mean.

One of the options to relax assumptions about functional forms has been to perform a series expansion from a baseline density. The use of polynomial expansions of a normal density was proposed in the continuous dependent variable case by Gallant and Nychka (1987). Application to a binary choice model has been performed by Gabler, Laisney and Lechner (1993). In count data settings we must cite the work by Gurmu *et al.* (1996, 1998, 1999), Cameron and Johansson (1997) and Creel (1999). We are not aware of any other application of series expansion in a model with endogenous binary regressor on a count data context.

Following the approach of Cameron and Johansson (1997) we will use a squared polynomial expansion over a Poisson baseline probability function. The resulting probability function is obtained by multiplying the baseline by a squared polynomial in the dependent variable $h^2(y,a)$ of degree K, where a is the vector of coefficients. The polynomial has to be raised to the square in order to preserve the non negativity of the density function. To ensure that the resulting probability function sums to unity it is necessary to divide the expression by a normalizing constant $\Psi_K(\lambda, a)$. Following Cameron and Johansson (1997) we have that

$$\begin{aligned} f_{SNP}(y|x,d,\varepsilon) &= \frac{\left(\sum_{i=1}^{K} a_i y^i\right)^2 P(y|\lambda)}{\Psi_k(\lambda,a)}, \text{where} \\ P(y|\lambda) &= \frac{\exp(-\lambda) * \lambda^y}{y!} \text{ and} \\ \lambda &= \exp\left(x_i\beta + d_i\gamma + \varepsilon\right). \end{aligned} \tag{3.6}$$

Estimation is done by maximum likelihood using (3.3) and (3.6). The estimates of all parameters are consistent and asymptotically normal distributed with variance computed by the standard sandwich form. As in Cameron and Johansson (1997) or Creel (1999), we do not consider technical issues on the ability of the expansion to approximate arbitrarily well any model as long as we let $K \to \infty$. The proof for the continuous case appears in Gallant and Nychka (1987).

Our model differs from Cameron and Johansson (1997) in at least two things:

first, we take into account endogeneity. Second, we allow that the latent variable enter in the specification through the baseline density, relieving the adaptative task of the polynomial expansion. We expect that this latter effect helps to get parsimonious results for the degree of the polynomial. Creel (1999) used a negative binomial as the baseline density and he found that small degrees of the polynomial expansion sufficed to obtain a good fit. In fact, the negative binomial can be obtained by integrating out a Poisson density with a gamma distributed latent variable. Such a latent variable is normally distributed in our context.

An important issue will be then to determine the order of the polynomial. In this sense, we must be cautious in order to avoid overfitting. To fix the polynomial degree we will use the goodness of fit test proposed in the previous section and other statistical tools: likelihood ratio tests, score tests and information criteria. The information criteria are defined by

$$\begin{aligned} BIC &= -2ln(L) + Pln(N) \text{ and} \\ CAIC &= -2ln(L) + P(ln(N) + 1), \end{aligned}$$

where P represents the number of parameters to be estimate, BIC stands for Bayesian Information Criteria and $CAIC$ for Consistent Akaike Information Criteria. Gallant and Tauchen (1995) advise to use the BIC as a parsimonious criteria on the size of the polynomial. The BIC imposes a bigger penalty on the number of parameters than the standard Akaike, but not as big as the CAIC does. Considering a penalty on the number of parameters is interesting, since one would like to avoid overparameterized models.

Contrary to the Poisson-Negative Binomial case, the mean of the count variable conditional on both observable and unobservable variables is no longer given by the parameterized λ. Instead, following Cameron and Johansson (1997), it is given by

$$E\left(y|w,d,\varepsilon\right) = \sum_{j=0}^{K}\sum_{i=0}^{K} a_i a_j m_{i+j}(\lambda), \tag{3.7}$$

where $m_j(.)$ denotes the j^{th} non central moment of the Poisson density and we stress the dependence on the baseline density mean λ. It is clear from (3.7) that the departure form the standard linear exponential specification of the conditional mean implies that we must modify the interpretation of the coefficients on the variables. In fact, for the case where there are no series expansion the expression in (3.7) reduces to λ, but in general for the $K \geq 1$ case the coefficients no longer admit an interpretation as elasticities.

In order to recover an estimate of the impact of covariates in the counts we should compute $E[y|x,d]$ which is a non-linear function of the parameters of interest. The derivation of such an expression is a bit more complex than for the $K=0$ model. Since y is a discrete non negative random variable, its mean is given by

$$\begin{aligned} E[y|x,d] &= \sum_{y=1}^{\infty} y * f(y|w,d), \qquad (3.8) \\ \text{where } f(y|w,d) &= \frac{f(y|w,d)}{f(d|w)}. \end{aligned}$$

The numerator can be replaced by the estimate of the joint density, and an estimate of the marginal density of d may also be obtained using

$$f(d|w) = \int_{-\infty}^{+\infty} [d\Phi^*(\varepsilon) + (1-d)(1-\Phi^*(\varepsilon))] dF(\varepsilon). \qquad (3.9)$$

Thus, the percentage mean effect of the change given by dummy regressor d can be computed using (3.8), (3.9) and

$$\frac{1}{N}\sum_{i=1}^{N} \frac{E(y|x_i, d=1) - E(y|x_i, d=0)}{E(y|x_i, d=0)} * 100. \qquad (3.10)$$

Notice that this conditional expectation is a function of the covariates observations. To summarize this information we will report two different measures. The first one computes the quantity in (3.10) evaluated at different points of the covariates. We chose three of these: the mean point, the upper point and the lower point. Here, upper (lower) means that we choose covariates' values in the range of the sample space yielding the largest (smallest) λ. If the fit of the PFIML was poor with respect to the Polynomial Poisson at the left tail of the distribution, then we would also expect larger differences in the mean effect estimates at these points. The second measure provides the frequency plot of the computed means.

3.4 Some applications

3.4.1 Data on frequency of recreational trips

Terza (1998) uses data on the number of trips by households (TOTTRIPS) to specify a model where vehicle ownership (OWNVEH) is included as a potentially endogenous dummy regressor. Table 3.1. describes the variables in the dataset. The variables have been divided in two groups attending to its determination status: endogenous (number of total trips and vehicle ownership) and exogenous (regressors).

We take as given the specification proposed by Terza (1998). In particular, we exclude the ADULTS variable from the count equation. This constitutes an overidentifying restriction. Consequently, the identification of the model do not crucially depend on the non-linear restriction imposed by the bivariate normality assumption.

We will first be concerned with the endogeneity of the OWNVEH variable. It is likely that unobserved variables as the personal predisposition (or aversion) to travel may be positively (or negatively) correlated with the decision of purchasing a vehicle. For instance, an individual may like to travel while detesting traffic jams, and such an aversion will be negatively correlated with the ownership of a vehicle. If this is the case, we should be aware of isolating the effect of vehicle ownership on the number of trips induced by this correlation.

Tables 3.2 and 3.3. show the results of Nonlinear Least Squares (NLS), TSM and WNLS estimation methods. The last two estimators, TSM and WNLS, have been originally proposed by Terza (1998). TSM and WNLS correct for endogeneity using a estimator similar to the one proposed by Heckman (1978) but adapted to the count data setting. The value of the OWNVEH coefficient estimated with TSM and WNLS increases between a 75% and a 30% with respect to NLS. This indicates that the sign of the correlation between ε and the endogenous dummy is negative. The WNLS pursues a more efficient estimation at the price of restricting the parametric family of the conditional counts to be a Poisson. For instance, a test of the significance of some variables like FULLTIME may lead to different conclusions under TSM or WNLS. We must take into account that the Poisson assumption may not verify.

Some descriptive statistics of TOTTRIPS are shown in Table 3.4, where we include some evidence on the non-Poisson behavior of this variable. The variance exceeds five times the mean and the number of zeros is up to 17 times greater than expected from a Poisson with mean parameter equal to the sample mean.

Also the conditional analysis shows that Poisson distribution is not suitable. In Table 3.5, the Andrews' test rejects the null of a correct specification at 5% for the K=0 model. Using an informal test, Terza (1998) also found evidence of misspecification for the Poisson assumption. This motivates the estimation under a more flexible specification which in principle would allow to test the Poisson against a wider set of alternatives. We started with the K=1 specification and sequentially increased the size of the polynomial. In term of goodness of fit, a considerable gain is obtained by the model with K=2, with respect to K=0 and K=1. As Table 3.5 and Figure 3.1 show the models with K=0 and K=1 underpredict the frequency of zeros and overpredict the frequency of counts one and two, as usually happens when the empirical distribution puts an excess of mass in the zero counts. In particular, the measure of distance between observed and predicted frequency decreases considerably and the test does not reject the null for K=2 and higher.

This leads to the problem of taking a decision on where to stop adding new terms to the polynomial expansion. We used several measures for this: information criteria, likelihood ratio tests jointly with the goodness of fit measure. The results on Table 3.6 give a strong evidence in favor of the model with two terms on the series expansion. The log likelihood ratio test strongly rejects the null hypothesis of K=0 and K=1 against the alternative of K=2. On the other hand, the null hypothesis of K=2 is not rejected against the alternative of K=3 or K=4, at even 15% of significance level. In terms of information criteria as shown in Table 3.6, the model with K=2 is the preferred one for any of the information criteria considered. Given that the first coefficient of the polynomial of the model (a1) with K=2 shows a small significance, it is expected that these results would improve if we restricted this coefficient to be zero.

Tables 3.7 and 3.8 show that the OWNVEH coefficient moves around 2.2. up to 2.3 for K=2,3,4 to be compared with the 2.05 in the K=0. Although the change is not important in size, the two coefficient do not have the same structural meaning. In principle the researcher should not be interested in coefficient by themselves but only on the way they can affect (cause) the characteristics of the count variable (for instance, its mean). In order to make comparisons of these mean effects, one should compute the expressions in (3.10). Table 3.9 shows the change in mean in number of trips due to vehicle ownership at three different points: the mean of the covariates, the upper point and the lower point (the exact values of covariates at this point are given in the table). In any case, the increase in the expected mean induced by OWNVEH is overpredicted by the K=0 model. Particularly interesting is the difference for the counts at the lower. Here the K=0 model does not reject the null of a zero impact while the effect is significant for the K=2 model. This is not surprising if we recall that the Poisson model had a worse fit for lower counts.

Figure 3.2 shows the distribution of the change in the expectation of the number of trips across individuals. That is, it shows the distribution of (3.10) without averaging but for every individual in the sample. The covariates are evaluated at the sample mean. Notice that the Poisson distribution overpredicts the impact of vehicle ownership by putting more mass on higher percentages.

3.4.2 Data on demand for medical care by the elderly

Deb and Trivedi (1997) consider data from the National Medical Expenditure Survey (NMES) conducted in 1987 and 1988. We will use a subsample of individuals aged 66 or more in the West part of USA. Table 3.10 describes the variables in the dataset.

Most of the individuals aged 65 or more are covered by Medicare, a public insurance that protects again health care costs. In addition, the individuals have the choice to contract a supplemental private insurance coverage (PRIVINS). The influence of insurance status on the utilization and costs of health care services is a very important topic in health economics (a non exhaustive list include Cameron *et al.* 1988, Manning *et al.* 1987, Coulson *et al.* 1995, Chiappori *et al.* 1998, Holly *et al.*1998, Street *et al.* 1999 and Vera-Hernández 1999). If this utilization were very sensitive to the generosity of insurance, the potential problems caused by moral hazard could be severe. In fact, Besley (1988) relates the optimal copayment rate to the compensated elasticity of the demand for health care with respect to out-of-pocket expenditures.

For studies using non-experimental data, the endogeneity of the insurance status in the equation for utilization is an important issue (see for example Cameron *et al.* 1988). This endogeneity is motivated by the relation between unobservable health characteristics and insurance choice. If adverse selection is a prevalent feature of the market, the ones that enjoy a more generous insurance are the ones with poor unobservable health conditions. This would cause a positive correlation between wide coverage insurance status and unobserved heterogeneity. On the contrary, if private insurance companies are able to select the most healthy individuals (cream skimming), we would expect the correlation to be negative. If endogeneity was neglected, the positive correlation will overestimate the insurance effect, while the negative one will underestimate it. Other studies that do take into account the endogeneity of insurance status in a count data context are Coulson *et al.* (1995) and Vera-Hernández (1999). In their paper Deb and Trivedi implement no correction of the endogeneity bias although they acknowledge that it could be present.

As a measure of health service utilization we use the number of physician office visits (OFP) in a quarter. Other measures like number of hospitalizations or number of physician non-office visits were also available. We chose OFP because this measure showed an accentuated non-Poisson behavior. This is particularly evident in view of Table 3.11. The variable shows a relative frequency of zeros (75.22) and variance to mean ratio of overdispersion (7.55) which a Poisson distribution fails to accommodate by far.

For the sake of parsimony, some restrictions were imposed in the specification of the probit equation. The constant term, the number of chronic diseases, the age, the sex, the marital status and MEDICAID were excluded after fitting preliminary standard probit models for the PRIVINS variable. The inclusion of the first five variables might induce multicollinearity in the probit part while adding low explanatory power (none of these variable was found to be individually significant at 5% and the Likelihood Ratio test of joint significance showed a p-value of 0.40), and they were excluded accordingly. On the other hand, the exclusion of the MEDICAID variable was due to the fact that this variable was a nearly perfect classifier (84% of individuals had either private insurance or MEDICAID coverage). For this model, we do not have an overidentifying restriction, since we did not find a convincing continuous instrument. Therefore, the model is identified thanks to the distributional assumptions. Finally, seven observations with zero or negative family income were deleted.

With this specification we calculated the NLS, TSM and WNLS estimators (see Tables 3.12 and 3.13). None of the TSM coefficients in the count equation except the one affecting the PRIVINS shows a change of sign. Moreover, this coefficient shows a small significance in the TSM and WNLS in opposition to the NLS case. However, the fact that there exist additional changes of sign and significance in the WNLS estimates with respect to the NLS and TSM may suggest that misspecification bias could be playing an important role here. The WNLS and the PFIML should approach asymptotically under a Poisson conditional count. Indeed, the results for the K=0 and the WNLS are similar for most of the coefficients with no change of sign. However, this is not the case for the PRIVINS (which is now bigger and significant) and correlation coefficients (which shows a negative sign).

The comparison of the empirical and predicted probabilities in Table 3.14 and Figure 3.3 leads us to conclude that the above results could be distorted due to misspecification problems. The fit for the Poisson K=0 model is poor, mainly for the zero, one and two counts and accordingly, the goodness of fit test rejects the null of a Poisson and order one polynomial expansion. The fit improves for order two and three polynomials.

On one hand, the models with K=2 and K=3 show better information criteria than K=0 and K=1 (see Table 3.15). On the other hand, in this case the information criteria do not discriminate between K=2 and K=3, since the first is favored by the Consistent Akaike and the second by the Bayesian criteria. We definitely chose the K=3 model since the order two polynomial is rejected against the order three alternative by the likelihood ratio test as shown in Table 3.15. We stopped here, but in order to determine if a polynomial of order four would significantly improve the fit we performed a Lagrange multiplier test. The advantage of using Lagrange tests in this context is that we do not require to estimate the unrestricted larger model which in our case requires an important computational effort (see computational appendix). The test did not reject the null hypothesis of K=3 with a p- value of 17%.

Once we feel confident on the fit of our model we computed the sensitivity analysis of the counts to changes in the endogenous dummy. This effect plays an important role in health economics: it measures the sensitivity of health care utilization due to the insurance status. Table 3.16 shows the estimation of this effect at three different points. It is particularly interesting to notice that the impact of insurance is close to zero in size and significance at the upper point, but not at lower extreme point or mean covariates. The upper point contain covariates values that indicate poor health conditions while the lower indicate good ones. Therefore it seems plausible in this case to conclude that office physician visits by people with poor health conditions is little affected by their insurance status. Finally, notice that the insurance effect predicted by the NLS is around 40%, very close to the mean effect at the mean point of covariates in the K=3 model (45%) and not so much to the mean effect at the lower extreme point (68%). However the NLS estimate is far away from the upper extreme (2.5%) casting doubts on the NLS when imposing the restriction of identical estimated percentage change to all of the individuals.

Figure 3.4 shows the distribution of change in the treatment effect due to insurance status across individuals. More mass is put at the 35%-50% interval of the percentage change for the K=2 model, while the K=0 tends to accumulate on higher values. In general, the K=0 tends to overpredict the percentage change.

The coefficient estimates for the K=3 model are shown in table 3.17 and 3.18. Although our emphasis in the paper is more on the methodology that in the results, we would like to comment on the results obtained for some variables. The coefficients of the discrete choice equation for PRIVINS are in Table 3.18. They show that the socioeconomic determinants are more important than observable health related variables for the private insurance choice. Previous literature as Cameron *et al.* (1988) also find this result for Australia, as well as we do in the second thesis chapter. Table 3.17 shows the coefficients for the number of visits equation. In parallel with the literature, the health related variables are important in determining the number of visits. In particular, the number of chronic conditions and self-assessment of health are statistically significant at conventional levels. We do not find a significant effect of age. It might be either because of our sample size or because we are only considering people over 65. We do not find either a sex related effect. Among socioeconomic determinants, years of schooling seems to play an important role, while race, marital status and income do not. The non significant of the later might be because of Medicaid which covers Medicare copayments of those with less income. Both PRIVINS and MEDICAID are statistically significant different from zero at conventional levels.

We would like to highlight the result found above that private insurance status did not significantly influenced number of visits for people with poor health conditions. We would like briefly to comment on the implications of this result for the optimal copayment. If health care consumption of unhealthy people is little responsive to copayments, high copayments put the individual at considerable risk[3] while not reducing expenditures. Consequently, if this result was confirmed with a higher sample size, the policy recommendation would be that people with poor health conditions should face smaller copayments. This obviously limits the role of cost sharing policies as a cost-containment measure.

3.5 Extensions

This paper has focused on a flexible way to simultaneously deal with count data and a dummy endogenous regressor. The flexibility was given in the count part of the model, but not in the bivariate normality assumption. While previous literature have found that this is a plausible assumption, it would be desirable to design a test for it. This constitutes the mainly methodological extension for the future. A possible way to deal with it is to expand the bivariate normal and perform a lagrange multiplier test on the coefficients of the expansion. Another methodological extension would be to compare the performance of this estimator with the two part model so popular in health economics.

[3]This assumes that an important part of health care consumption is unpredictable. Notice that, for instance, routine treatment of chronic conditions do not have this property.

Allowing the influence of the dummy variable to vary according to covariates have revealed as fruitful from a policy perspective. More work on this should be done in order to model this dependence in a more transparent way, possibly using interaction terms.

3.6 Conclusions

In this paper we propose a way to simultaneously deal with flexible modelling of a count variable and a dummy endogenous regressor. The framework require distributional assumptions on the unobservables in order to consider functional forms that departure from the linear exponential specification. It is natural to consider the departure from the linear exponential specification given the results on the literature on exogenous regressors.

We propose and estimate a flexible Polynomial Poisson FIML which tries to deal with those cases where the count variable shows a persistent non-Poissoness even when we account for unobserved heterogeneity. In addition, we compute measures of fit and procedures à la Andrews to test the assumptions of the model based on the observed differences between fitted and empirical frequencies. We test the model using two data sets on number of trips by households and number of physician office visits, already analyzed in the literature. The results show that flexible estimation of the conditional probability function of the count helps to improve significantly the fit of the model. We also find the largest differences in the estimate of the mean effect can be found when the conditional density has a relatively low predicted mean. Therefore, the flexibility in the count part is also important to estimate the treatment effect.

Given the non-linearity of the conditional mean of the count, the dummy endogenous regressor might have different effects depending on the covariates values. We have shown that this might be important for policy purposes. In particular, we have obtained that private insurance status do not significantly influence number of visits to doctors for those with observable poor health conditions.

3.7 Tables

Table 3.1. Description of variables. Number of trips by households

Variable	**Mean**	**Std.**	**Description**
			Endogenous
Tottrips	4.5511	4.9351	Number of trips by members of the household in 24 hrs.
OwnVeh	0.8492	0.3581	1 if household owns at least one motorized vehicle.
			Exogenous
WorkSchl	0.2622	0.3278	% of total trips for work vs. personal.
Hhmem	2.9289	1.6127	number of individuals in the household.
DistoCbd	0.2887	0.4932	distance to the central business district in kilometers.
AreaSize	0.3761	0.4848	1 if area bigger than 2,5 million population.
FullTime	0.9792	0.8475	number of full time workers in the household.
DistoNod	2.0272	3.1378	distance from home to the nearest transit node in blocks.
RealInc	0.8042	0.9197	household income divided by median income of census tract.
Weekend	0.2236	0.4170	1 if 24 hours survey period is Saturday or Sunday.
Adults	2.0797	0.8978	number of adults in the household 16 years or older.

Scaling: DistoCbd has been divided by 30, DistoNod by 5 and RealInc by 3

Table 3.2. Count equation estimates.

No polynomial models. Number of trips by households

	NLS	**TSM**	**NWLS**	**K=0**
Constant	-0.600 0.225	-1.445 0.528	-1.005 0.181	-1.386 0.217
Workschl	-0.527 0.143	-0.554 0.147	-0.363 0.128	-0.340 0.127
Hhmem	0.166 0.027	0.148 0.031	0.134 0.028	0.150 0.019
DistoCbd	-0.149 0.136	-0.268 0.172	-0.057 0.024	-0.044 0.043
Areasize	-0.034 0.097	-0.008 0.100	0.038 0.085	0.049 0.073
Fulltime	0.189 0.048	0.105 0.101	0.220 0.073	0.246 0.042
DistoNod	0.002 0.001	0.021 0.012	0.019 0.013	0.021 0.009
RealInc	0.041 0.048	0.020 0.052	0.007 0.051	0.082 0.030
Weekend	-0.155 0.112	-0.165 0.115	-0.029 0.080	-0.097 0.080
Ownveh	1.607 0.185	2.796 0.613	2.079 0.312	2.060 0.256
σ				0.728 0.139
ρ				-0.697 0.032

Asymptotic standard errors at the bottom row of each cell.

Table 3.3. Discrete choice equation estimates. No polynomial models.

Number of trips by households

	Probit	**K=0**
Constant	-0.633 0.237	-0.547 0.362
Workschl	0.152 0.265	0.316 0.352
Hhmem	0.003 0.068	0.051 0.086
DistoCbd	0.629 0.399	0.665 0.367
Areasize	-0.206 1.242	-0.249 0.152
Fulltime	0.871 0.155	1.008 0.187
Adults	0.381 0.145	0.267 0.176
DistoNod	0.048 0.033	0.052 0.031
RealInc	0.472 0.177	0.337 0.227

Asymptotic standard errors at the bottom row of each cell.

Table 3.4. Descriptive analysis. Number of trips by households

Number of observations	577
Mean	4.55
Variance	24.35
Variance to mean	5.35
Empirical to expected kurtosis	3.91
Proportion of zeros to sample size	0.18
Poisson predicted frequency of zeros	0.01
Ratio real/predicted	17.65

Note: Poisson predictions were computed using the sample mean.

Table 3.5. Fitted vs. empirical frequencies. Number of trips by households

Count	**Empirical**	**Fitted**				
		K=0	K=1	K=2	K=3	K=4
0	0.185	0.155	0.155	0.184	0.184	0.187
1	0.119	0.157	0.157	0.124	0.125	0.1166
2	0.109	0.133	0.133	0.113	0.112	0.124
3	0.124	0.108	0.108	0.105	0.104	0.114
4	0.091	0.086	0.086	0.096	0.098	0.083
5	0.053	0.068	0.068	0.07	0.076	0.076
6	0.062	0.054	0.054	0.066	0.067	0.055
7	0.053	0.042	0.043	0.041	0.045	0.048
8	0.045	0.033	0.034	0.032	0.037	0.033
9	0.024	0.026	0.027	0.027	0.022	0.033
10	0.024	0.021	0.021	0.023	0.028	0.024
$\geqslant 11$	0.105	0.109	0.109	0.104	0.105	0.108
Sum of differences ($x10^{-4}$)		2.9	2.8	1.2	1.31	1.4
Goodness of Fit (Andrews)		23.59	22.36	9.34	9.07	9.55
P-Value		0.01	0.02	0.59	0.61	0.57

Table 3.6 Information criteria and log-likelihood ratio test. Number of trips by households

	K=1	**K=2**	**K=3**	**K=4**
Number of parameters	22	23	24	25
Sample size	577	577	577	577
$\frac{1}{N} * \ln(L)$	-2.665	-2.655	-2.654	-2.6522
$CAIC$	3237.925	3233.815	3239.374	3245.186
BIC	3215.925	3210.815	3215.374	3220.186
Null\Alternative	**P-Values Log-likelihood ratio tests**			
K=0	0.0639	0.0005	0.0008	0.00113
K=1		0.0007	0.0013	0.00198
K=2			0.1796	0.18769
K=3				0.21367

Table 3.7. Count equation estimates. Polynomial models. Number of trips by households

	K=1	**K=2**	**K=3**	**K=4**
Constant	-1.341 0.223	-2.499 0.387	-2.411 0.404	-2.638 0.522
Workschl	-0.351 0.131	-0.410 0.140	-0.416 0.172	-0.435 0.159
Hhmem	0.157 0.029	0.187 0.028	0.194 0.029	0.189 0.029
DistoCbd	-0.041 0.036	-0.040 0.082	-0.039 0.671	-0.070 0.068
Areasize	0.061 0.083	0.084 0.103	0.099 0.159	0.063 0.174
Fulltime	0.261 0.046	0.305 0.056	0.313 0.081	0.317 0.095
DistoNod	0.021 0.010	0.029 0.011	0.028 0.021	0.033 0.018
RealInc	0.067 0.025	0.114 0.028	0.093 0.032	0.025 0.041
Weekend	-0.095 0.078	-0.113 0.089	-0.112 0.131	-0.126 0.099
Ownveh	2.051 0.257	2.3074 0.299	2.282 0.325	2.249 0.353
σ	0.733 0.141	0.934 0.169	0.926 0.165	0.893 0.194
ρ	-0.676 0.032	-0.671 0.073	-0.659 0.063	-0.681 0.094
$a1$	-0.027 0.001	0.053 0.145	0.034 0.149	-0.188 0.422
$a2$		0.231 0.154	0.210 0.020	0.643 0.664
$a3$			-0.005 0.005	-0.102 0.163
$a4$				0.014 0.023

Asymptotic standard errors at the bottom row of each cell

Table 3.8. Discrete choice equation estimates. Polynomial Models.

Number of trips by households

	K=1	**K=2**	**K=3**	**K=4**
Constant	-0.550 0.311	-0.505 0.347	-0.504 0.778	-0.518 0.386
Workschl	0.311 0.317	0.317 0.313	0.315 0.477	0.324 0.349
Hhmem	0.048 0.069	0.052 0.073	0.048 0.079	0.051 0.114
DistoCbd	0.669 0.373	0.649 0.373	0.652 0.865	0.665 0.435
Areasize	-0.253 0.155	-0.257 0.156	-0.262 0.165	-0.248 0.183
Fulltime	1.004 0.182	0.994 0.176	0.993 0.193	0.987 0.191
Adults	0.270 0.183	0.251 0.179	0.253 0.313	0.247 0.183
DistoNod	0.052 0.032	0.049 0.031	0.050 0.034	0.047 0.031
Realinc	0.351 0.234	0.331 0.232	0.343 0.316	0.373 0.231

Asymptotic standard errors at the bottom row of each cell.

Table 3.9. Conditional mean and change in the expectation of number of household trips with respect to OwnVeh variable

	K=0	K=2
Upper Extreme Covariates		
OwnVeh	31.9 8.2	27.9 34.5
Not OwnVeh	18.0 7.9	22.5 11.2
% Change	77.5 46.0	24.5 210.7
Lower Extreme Covariates		
OwnVeh	1.2 0.4	1.2 0.8
Not OwnVeh	1.3 0.6	2.6 1.9
% Change	-11.1 22.9	-52.2 13.8
Mean Covariates		
OwnVeh	4.6 0.1	4.7 0.1
Not OwnVeh	1.6 0.2	2.0 0.3
% Change	179.3 47.5	136.8 40.7

Upper extreme covariates values: Workschl=0, Hhmem=13, DistoCbd=0, Areasize=1, Fulltime=4, DistoNod=10, RealInc=10, Weekend=0.

Lower extreme covariates values: Workschl=1, Hhmem=1, DistoCbd=10, Areasize=0, Fulltime=0, DistoNod=0.2, RealInc=0.02, Weekend=1.

Mean covariates: Workschl=2.02, Hhmem=2.92, DistoCbd=0.28, Areasize=0.37, Fulltime=0.97, DistoNod=2.02, RealInc=0.8, Weekend=0.22.

Table 3.10. Description of variables. Number of physician office visits

	Mean	**Std.**	**Description**
			Endogenous
Ofp	6.359	6.929	Number of physician office visits in a quarter
Privins	0.778	0.415	=1 if person is covered by private insurance
			Exogenous
Exclhlth	0.115	0.319	=1 if self-perceived health is excellent
Poorhlth	0.108	0.311	=1 if self-perceived health is excellent
Numchron	0.150	0.131	Number of chronic conditions
Adldiff	0.216	0.411	=1 if person has a limiting condition for daily activities
Age	0.741	0.065	age in years
Black	0.054	0.226	=1 if person is African Amercian
Male	0.408	0.491	=1 if person is male
Married	0.574	0.494	=1 if person is married
School	0.575	0.190	Number of years of education
Faminc	0.062	0.066	Family income in $500.000
Employed	0.118	0.323	=1 if person is employed
Medicaid	0.120	0.325	=1 if person is covered by Medicaid

Scaling: Numchron has been divided by 10, Age by 100 and School by 20

Table 3.11. Descriptive analysis. Number of physician office visits.

Number of observations	791
Mean	6.3
Variance	48
Variance to mean	7.5
Empirical to expected kurtosis	4.3
Proportion of zeros to sample size	0.13
Poisson predicted frequency of zeros	0.001
Ratio real/predicted	75.2

Note: Poisson predictions were computed using the sample mean.

Table 3.12. Count equation estimates. No Polynomial Models. Number of physician office visits

	NLS	**TSM**	**WNLS**	**K=0**
Constant	1.669 0.507	1.896 0.652	1.472 0.706	0.650 0.468
Exclhlth	-0.510 0.129	-0.528 0.134	-0.527 0.144	-0.552 0.123
Poorhlth	0.064 0.119	0.035 0.131	0.119 0.151	0.138 0.120
Numchron	1.385 0.274	1.407 0.280	1.562 0.293	1.826 0.264
Adldiff	0.119 0.124	0.082 0.139	0.174 0.142	0.173 0.111
Age	-0.747 0.646	-0.734 0.644	-0.974 0.637	-0.824 0.604
Black	-0.082 0.177	-0.230 0.286	-0.051 0.411	-0.089 0.187
Male	-0.011 0.092	-0.012 0.091	0.003 0.082	-0.010 0.080
Married	-0.106 0.094	-0.106 0.094	-0.138 0.090	-0.099 0.085
School	0.460 0.218	0.627 0.359	0.592 0.504	0.651 0.259
Faminc	-0.188 0.475	-0.029 0.781	0.016 0.866	-0.479 0.605
Employed	0.040 0.173	0.016 0.174	-0.100 0.155	-0.144 0.103
Medicaid	0.418 0.128	0.408 0.131	0.402 0.153	0.475 0.129
Privins	0.341 0.111	-0.054 0.698	0.664 0.970	1.006 0.203
σ				1.083 0.104
ρ				-0.447 0.122

Asymptotic standard errors at the bottom row of each cell.

Table 3.13. Discrete choice equation estimates. No Polynomial Models

	Probit	**K=0**
Exclhlth	-0.237 0.168	-0.219 0.168
Poorhlth	-0.239 0.167	-0.204 0.161
Adldiff	-0.354 0.132	-0.345 0.125
Black	-1.098 0.208	-0.345 -1.078
School	1.584 0.154	1.514 0.153
Faminc	3.464 1.115	3.387 1.169
Employed	-0.301 0.173	-0.275 0.173

Asymptotic standard errors at the bottom row of each cell

Table 3.14. Fitted vs empirical frequencies.Number of physician office visits

Count	**Empirical**	**Fitted**			
		K=0	**K=1**	**K=2**	**K=3**
0	0.1302	0.0985	0.0983	0.1263	0.1257
1	0.0961	0.1290	0.1291	0.1024	0.1045
2	0.1062	0.1207	0.1209	0.0968	0.0970
3	0.0910	0.1022	0.1026	0.0938	0.0929
4	0.0860	0.0841	0.0845	0.0864	0.0856
5	0.0746	0.0688	0.0693	0.0758	0.0754
6	0.0657	0.0565	0.0571	0.0642	0.0643
$\geqslant 7$	0.3502	0.3401	0.3383	0.3543	0.3546
Sum of differences$(x10^{-4})$			1.49	0.37	0.39
Goodness of fit (Andrews)			21.784	2.39	2.52
P-Value			0.0027	0.9347	0.9254

Table 3.15. Information criteria and Log-likelihood ratio test. Number of physician office visits.

	K=0	**K=1**	**K=2**	**K=3**
Number of parameters	23	24	25	26
Sample Size	791	791	791	791
$\frac{1}{N}\ln(L)$	-3.3097	-3.3039	-3.2966	-3.2922
$CAIC$	5412.46	5410.95	5407.11	5407.76
BIC	5389.46	5386.95	5382.11	5381.76
Null\Alternative	**P-Values Log-likelihood ratio tests**			
K=0		0.00243	0.0000	0.0000
K=1			0.0006	0.0000
K=2				0.0080

Table 3.16. Conditional mean and change in the expectation of number of physician office visits with respect to Privins variable

	K=0	K=3
Upper Extreme Covariates		
Privins	30.6 1.4	28.8 0.9
Not Privins	30.9 2.6	28.1 2.1
% Change	-0.9 6.6	2.5 5.9
Lower Extreme Covariates		
Privins	0.8 0.3	0.8 0.3
Not Privins	0.5 0.2	0.4 0.2
% Change	53.2 18.4	68.4 24.7
Mean Covariates		
Privins	6.8 0.2	6.6 0.2
Not Privins	4.4 0.4	4.5 0.4
% Change	52.0 16.9	45.7 15.7

Upper extreme covariates values: Exclhlth=0, Poorhlth=1, Numchron=0.7, Adldiff=1, Age=0.66, Black=0, Male=0, Married=0, School=0.9, Faminc=0.0001, Employed=0, Medicaid=1.

Lower extreme covariates values: Exclhlth=1, Poorhlth=0, Numchron=0, Adldiff=0, Age=0.96, Black=1, Male=1, Married=1, School=0, Faminc=0.48, Employed=1, Medicaid=0.

Mean covariates: Exclhlth=0.11, Poorhlth=0.1, Numchron=0.15, Adldiff=0.21, Age=0.74, Black=0.05, Male=0.4, Married=0.57, School=0.57, Faminc=0.06, Employed=0.11, Medicaid=0.12.

Table 3.17. Count equation estimates. Polynomial Models.

Number of physician office visits

	K=1	K=2	K=3
Constant	0.752 0.475	-0.067 0.561	0.062 0.547
Exclhtlh	-0.549 0.117	-0.650 0.142	-0.634 0.132
Poorhlth	0.146 0.107	0.168 0.142	0.176 0.123
Numchron	1.902 0.263	2.144 0.321	2.178 0.291
Adldiff	0.118 0.109	0.203 0.124	0.149 0.128
Age	-0.700 0.605	-0.985 0.707	-0.842 0.705
Black	-0.178 0.187	-0.104 0.215	-0.178 0.218
Male	-0.033 0.085	-0.013 0.093	-0.036 0.091
Married	-0.090 0.082	-0.116 0.099	-0.102 0.095
School	0.760 0.232	0.801 0.321	0.881 0.268
Faminc	-0.095 0.603	-0.586 0.674	-0.191 0.679
Employed	-0.174 0.104	-0.184 0.121	-0.206 0.120
Medicaid	0.508 0.137	0.561 0.153	0.584 0.152
Privins	0.738 0.176	1.167 0.231	0.901 0.211
σ	1.051 0.068	1.323 0.116	1.279 0.069
ρ	-0.261 0.130	-0.423 0.135	-0.281 0.148
$a1$	-0.0292 0.0007	0.019 0.087	-0.005 0.079
$a2$		1.191 0.089	0.167 0.009
$a3$			-0.0040 0.0001

Asymptotic standard errors at the bottom row of each cell.

Table 3.18. Discrete choice equation estimates. Polynomial Models

	K=1	**K=2**	**K=3**
Exclhlth	-0.233 0.164	-0.219 0.161	-0.231 0.163
Poorhlth	-0.225 0.165	-0.203 0.161	-0.222 0.165
Adldiff	-0.349 0.128	-0.345 0.125	-0.349 0.127
Black	-1.097 0.207	-1.082 0.202	-1.097 0.206
School	1.564 0.155	1.512 0.154	1.554 0.155
Faminc	3.438 1.225	3.404 1.172	3.431 1.217
Employed	-0.288 0.177	-0.275 0.174	-0.285 0.176

Asymptotic standard errors at the bottom row of each cell.

3.8 Figures.

Figure 3.1: Empirical versus …tted frequency of counts. Number of trips by households. Empirical (__), K=0 (- -), K=2 (. . .).

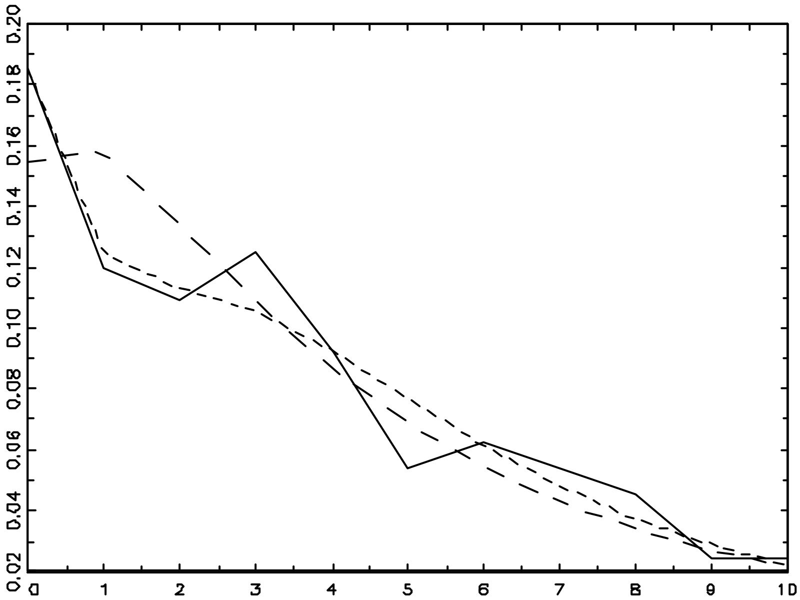


Figure 3.2: Distribution of change in expectation of number of trips due to OwnVeh. Evaluated at mean covariates across the sample. In basis points, K=0 (_ _), K=2 (- -).

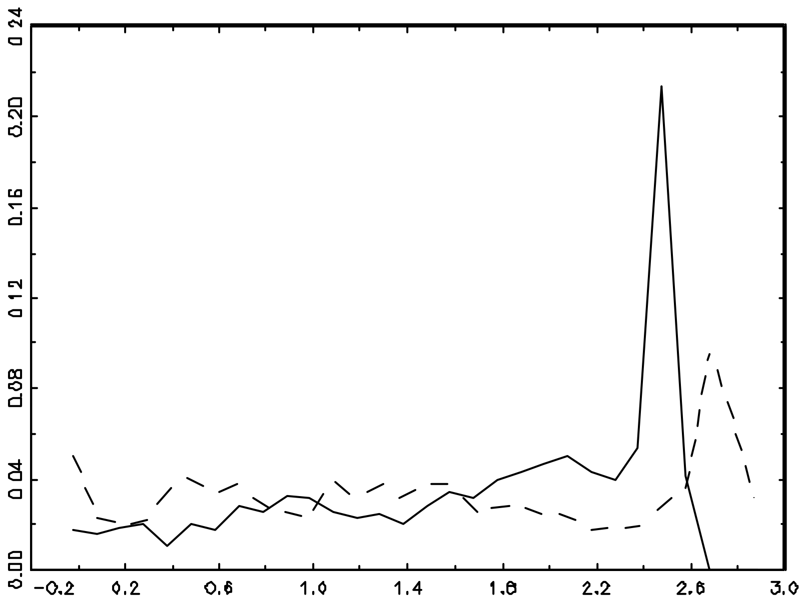


Figure 3.3: Emprical vs. …tted frequency of counts. Number of physician o¢ce visits. Empirical (__), K=0 (- -), K=3 (. .).

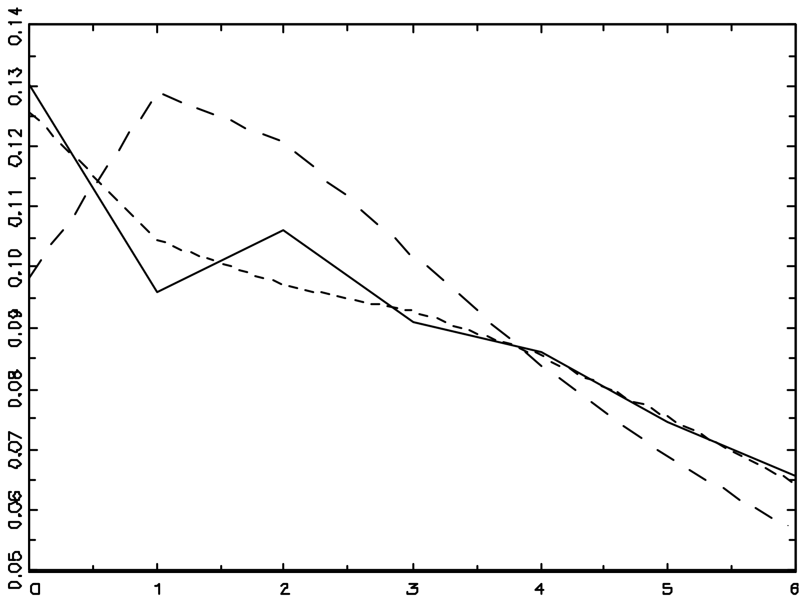
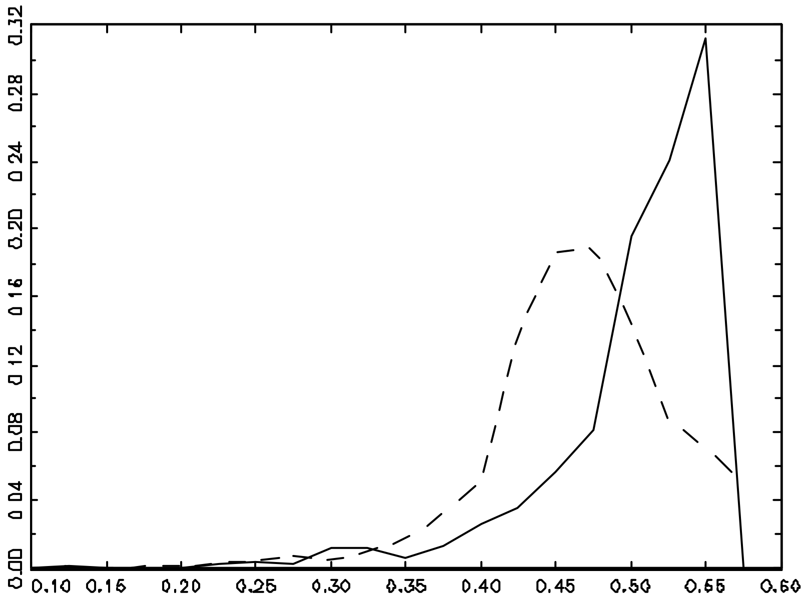


Figure 3.4: Distribution of change in expectation in number of visits due to Privins. Evaluated at mean covariates across the sample. In basis points, K=0 (__), K=3 (- -).



3.9 Computational appendix.

All computations were done using GAUSS 3.2 for MS-DOS in a Pentium III 450Mhz. microprocessor. The numerical routine for integration of unobserved heterogeneity in (3.3) is based on the Gauss-Legendre quadrature (INTQUAD package). The procedure requires to define finite upper and lower bounds of integration and the number of points for quadrature evaluation. This problem was initially solved by setting this bounds as four times the current standard deviation of ε.

The objective function was optimized using the Broyden-Fletchell-Golden-Shannon (BFGS) algorithm. Since local optima is a problem often encountered when using series expansion we tried several initial conditions. Each run with the BFGS took between some minutes and two hours, depending on the polynomial degree and the initial condition. In order to check for global optima, we implemented the Simulated Annealing (SA)[4], which is a robust algorithm designed to find global optima. The SA algorithm is a random search method which tries to escape from local optima by randomly accepting downhill moves. The decision to accept downhill moves is made by the Metropolis criteria depending on two parameters: temperature and strength. Following the advice in Goffe et al. (1994) we tried initial runs to determine the optimal starting temperature and strength. To avoid overflow errors we restricted the area of search using wide enough upper and lower bounds centered around the best BFGS optimum. To ensure that the global optima was found we put a big number of function evaluations per iteration. Moreover, given the difference in time for convergence between the BFGS (around a one or two hours for K>0) and the SA (at least one week) we were also concerned to know whether a global solution could also be found using derivative-based methods. For the dataset on the number of trips, the SA algorithm matched the best result obtainned using BFGS. For the dataset on number of visits, the SA algorthim improved the best result obtainned using BFGS for K=3, while matched the result for K=1 and K=2. In fact, the abnormally high condition number of the covariates inner product matrix (1497) can explain why we did not find the global optima using the BFGS for K=3.

Moreover, we found that the number of visits dataset was more problematic than the data from Terza. Here we found that the Hessian was ill-conditioned even for the K=0 case, which is not strange given the condition number mentioned above. This

[4] using the code written by E.G. Tsionas.

caused numerical problems: negative eigenvalues appeared and the value of the objective function at the optimum changed significantly when we moved the bound from plus/minus four to five standard deviations. Hence, we decided to explore two possible explanations: a bad performance of the integral and the computation of the Hessian procedure. First, we decided to replace the normal specification of ε by a truncated normal distribution. Doing the appropriate changes of variable $\left(\xi = \frac{\epsilon}{\sqrt{2}\sigma}\right)$ equation yields (3.3),

$$\begin{aligned} f(y,d|w) &= \frac{1}{\sqrt{\pi}(2\Phi(a)-1)}\int_{-a}^{+a} f(y|d,w,\xi) * \\ &\quad [d\Phi(\frac{z\alpha+\sqrt{2}\rho\xi}{\sqrt{1-2\rho}}) + (1-d)(1-\Phi(\frac{z\alpha+\sqrt{2}\rho\xi}{\sqrt{1-2\rho}}))]\exp(-\xi^2). \end{aligned}$$

Since is $N(0,1/2)$we chose $a = \frac{2}{\sqrt{2}}$. This approach allows to fix the bounds of integration independently of the parameter. We also increased the number of quadrature evaluation points at the cost of extra computing effort. We also explored the possible instability induced by the computation of the Hessian. We found that the GAUSS package used a two steps hessian procedure. We replaced this with a four steps hessian which was found to be very useful not only for this application but in many other contexts (all codes are available from the authors on request).

Finally, although we required more computing-time[5] (using the SA the k=3 case needed more than two weeks to converge) we found all these patches sufficed to get a stable numerical procedure).

[5]The increase in computing time is mostly because of the increase in the integral accuracy.

Bibliography

[1] Andrews, D.W.K. (1988a), Chi-Square Diagnostic Tests for Econometric Models: Theory, *Econometrica*, 56, 1419-1453.

[2] Andrews, D.W.K. (1988b), Chi-Square Diagnostic Tests for Econometric Models. Introduction and Applications, Journal of Econometrics, 37, 135-156.

[3] Besley, T. J. (1988), Optimal Reimbursement Health Insurance and the Theory of Ramsey Taxation, *Journal of Health Economics*, 7, 321-336.

[4] Blundell, R. and Windmeijer, F. (2000), Identifying Demand for Health Resources Using Waiting Times Information. *Health Economics, 9, 6, 465-474.*

[5] Cameron, A.C. and Trivedi, P.K. (1986), Econometric Models Based on Count Data: Comparisons and Applications of some Estimators and Tests, *Journal of Applied Econometrics*, 1, 29-53.

[6] Cameron, A.C., Trivedi, P.K., Milne, F, and Pigott, J. (1988), A Microeconometric Model of the Demand for Health Care and Health Insurance in Australia. *Review of Economic Studies,* 55, 85-106.

[7] Cameron, C. and Johansson, P. (1997), Count Data Regression using Series Expansions: with Applications, *Journal of Applied Econometrics*, 12, 203-233.

[8] Cameron, A.C., Trivedi, P.K. (1998),*Regression Analysis of Count Data*, Econometric Society Monograph 30. New York: Cambridge University Press.

[9] Chiappori, P.A., Durand, F. and Geoffard, P.Y, (1998), Moral Hazard and the Demand for Physician Services: First Lessons from a French Natural Experiment, *European Economic Review*, 42, 499-511.

[10] Coulson, N.E., Terza, J., Neslulan, C.A. and Stuart, C.B. (1995), Estimating the Moral Hazard Effect of Supplemental Medical Insurance in the Demand for Prescription Drugs by the Elderly, *American Economic Review Papers and Proceedings*, 85, 122-126.

[11] Creel, M. (1999), A Flexible and Parsimonious Density for Count Data, unpublished manuscript, Universitat Autònoma de Barcelona.

[12] Deb, P. and Trivedi, P.K. (1997), Demand for Medical Care by the Elderly: A Finite Mixture Approach, *Journal of Applied Econometrics*, 12, 313-336.

[13] Deb, P., Ming, X., and Trivedi, P.K. (2000), A Comparison of Maximum Likelihood and Moment Based Estimators for Finite Mixture Models. Indiana University Purdue University Indianapolis.

[14] Gabler, S., Laisney, F. and Lechner, M. (1993), Seminonparametric Estimation of Binary-Choice Models With an Application to Labor-Force Participation, *Journal of Business and Economic Statistics*, 11, 61-80.

[15] Gallant, A.R. and Nychka, D. (1987), Semi-Nonparametric Maximum Likelihood Estimation, *Econometrica*, 55, 363-390.

[16] Goffe, W.L., Ferrier, G.D. and Rogers, J. (1994), Global Optimization of Statistical Functions with Simulated Annealing, *Journal of Econometrics*, 60, 65-99.

[17] Grogger, J. (1990), A Simple Test for Exogeneity in Probit, Logit and Poisson Regression Models, *Economic Letters*, 33, 329-332.

[18] Gurmu, S. and Trivedi, P.K. (1996), Excess Zeros in Count Models for Recreational Trips,J*ournal of Business and Economic Statistics*, 14, 469-477.

[19] Gurmu, S., Rilstone, P. and Stern, S. (1999), Semiparametric Estimation of Count Regression Models, *Journal of Econometrics*, 88,123-150.

[20] Gurmu, S. (1997), Semi-Parametric Estimation of Hurdle Regression Models with an Application to Medicaid Utilization, *Journal of Applied Econometrics*, 12, 225-242.

[21] Hausman, J.A., Hall B. H. and Griliches, Z. (1984), Econometric Models for Count Data with Applications to the Patents-R&D relationship, *Econometrica*,52, 909-38.

[22] Heckman, J., (1978), Dummy Endogenous Variables in a Simultaneous Equation System.*Econometrica*, 46, 931-959.

[23] Heckman, J., Robb, Walker (1990), Testing the Mixtures of Exponentials Hypothesis and Estimating the Mixing Distribution by the Method of Moments, *Journal of the American Statistical Association*, 85, 582-589.

[24] Holly, A., Gardiol, L., Domenighetti, G. and Bisig, B. (1998), An Econometric Model of Health Care Utilization and Health Insurance in Switzerland, *European Economic Review*, 42, 513-522.

[25] Kenkel, D.S., Terza, J.V. (1999), The Effect of Physician Advice on Alcohol Consumption: Count Regression with an Endogenous Treatment Effect,unpublished manuscript, Penn State University.

[26] Manning, W.G., Newhouse, J.P., Duan, N., Keeler, E.B., Leibowitz, A. and Marquis, M.S., (1987), Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment,*The American Economic Review*, 77, 251-277.

[27] Mullahy, J. (1997), Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior,*Review of Economic and Statistics*, 79, 586-593.

[28] Mullahy, J. (1998), Much Ado about Two: Reconsidering Retransformation and the Two-Part Model in Health Econometrics, *Journal of Health Econonomics*, 17, 247-281.

[29] Newhouse, J.P. *et al.* (1993), *Free for all?*, Harvard University Press.

[30] Newey, Powell, Walker (1990), Semiparametric Estimation of Selection Models: Some Empirical Results, *American Economic Review Papers and Proceedings*, May, 324-328.

[31] Pohlmeier, W. and Ulrich, V. (1995), An Econometric Model of the Two-Part Decisionmaking Process in the Demand for Health Care,*The Journal of Human Resources*, 30, 339-61.

[32] Street, A. Jones, A. and Furuta, A (1999), Cost-sharing and Pharmaceutical Utilisation and Expenditure in Russia, *Journal of Health Economics*, 18, 459-472.

[33] Terza, J.V., Wilson, P.W. (1990), Analyzing Frequencies of Several Types of Events: a Mixed Multinomial-Poisson Approach, *Review of Economic and Statistics* 72,108-115.

[34] Terza, J. V. (1998), Estimating Count Data Models with Endogenous switching: Sample selection and Endogenous Treatment Effects, *Journal of Econometrics*, 84, 129-154.

[35] Vella, F., (1995), Estimating Models with Sample Selection Bias: a Survey, *The Journal of Human Resources* 33, 1, 127-169.

[36] Vera Hernández, A.M. (1999), Duplicate Coverage and Demand for Health Care. The Case of Catalonia, *Health Economics*, 8, 579-598.

[37] Windmeijer, F.A.G. and Santos-Silva, J.M.C. (1997), Endogeneity in Count Data Models: An Application to Demand for Health Care, *Journal of Applied Econometrics*, 12, 281-294.

[38] Zimmermann, K.F., (1995), Recent Developments in Count Data Modelling: Theory and Application, *Journal of Economic Surveys*, 9, 1-24.

[39] Winkelmann, R., (1998), Count Data Models with Selectivity, *Econometric Reviews,* 17, 4, 339-359.