Working Paper No. 452.00

A Semi-Nonparametric Estimator for Counts With an Endogenous Dummy Variable

Andrés Romeu-Santana
Department of Economics, Universitat Autònoma de Barcelona, aromeu@idea.uab.es
Ángel M. Vera-Hernández
Department of Economics, Universitat Autònoma de Barcelona, avera@idea.uab.es

Abstract

Treating endogeneity and flexibility in such a way that efficiency is not sacrificed has become a rising point of interest in count data models. We use a polynomial expansion of a Poisson baseline density to compute the full information maximum likelihood (FIML) estimator. In order to test the model we propose measures of goodness of fit, information criteria, likelihood ratio and scores tests for evaluation. We also show how to compute statistics for sensitivity analysis. Then, we test our model using data on number of trips by households and number of physician office visits, finding that low order polynomials may be enough to improve fit significantly.

Resumen

El tratamiento de la endogeneidad y la flexibilidad sin sacrificio de la eficiencia se ha convertido en un tema de creciente interés en modelos de número de eventos. En este trabajo utilizamos una expansión polinomial de una densidad base Poisson para calcular el estimador de maximo versosimilitud de información completa (MVIC). Con el fin de contrastar el modelo proponemos medidas de bondad de ajuste, ratio de verosimilitudes y contrates del gradiente para la evaluación. A continuación contrastamos el modelo utilizando datos sobre el número de desplazamientos de los hogares y el número de consultas al médico de cabecera, encontrando que una polinomio de grado pequeño puede bastar para mejorar elajuste significativamente.

KEY WORDS: Polynomial Poisson expansion; Flexible functional form; Treatment effect; Sensitivity analysis.

1. INTRODUCTION

Count data models try to explain the behavior of discrete and non negative dependent random variables (Winkelmann and Zimmermann 1994 and 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 adaptative specifications 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 literature addressing this problem has concentrated on correcting the specification of the moment conditions to account for possible endogeneity of the regressors (Terza 1998, Windmeijer and Santos Silva 1997, Mullahy 1997 and Grogger 1990). These models use as a benchmark the standard assumption of a linear exponential specification for the mean of the count variable conditional on both observable and unobservable variables (Kenkel and Terza, 1999 relax this assumption by assuming an inverse Box Cox functional form for the conditional expectation of counts). The identification of the parameters of interest is made on the basis of this moment restriction. However, even if the restriction is correctly specified and we feel confident about its robustness, one could think of using higher order moment conditions and improve the efficiency of the estimates.

Efficiency is important when decisions are to be made on the basis of the inference process. For instance, if we think of a model where we try to explain health care utilization, the parameter affecting insurance status plays a key role if one suspects of moral hazard problems and a precise estimate would be needed. Therefore, our work is a flexible approach which allows to improve optimality of the estimation in models where endogeneity of a dummy may be present.

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 semi-nonparametric (SNP) 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 condition 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 semi-nonparametric 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 pretends to confront 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.

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 $f(y_i \mid x_i, d_i, \varepsilon)$ 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 (ε,v) follows a bivariate normal distribution with zero mean and covariance matrix

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

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

$$\begin{aligned} f(y_i,d_i|w_i) &= \int_{-\infty}^{-\infty} f(y_i|w_i,d_i,\varepsilon) f(d_i|w_i,\varepsilon_i) dF(\varepsilon) = \\ &\int_{-\infty}^{-\infty} f(y_i|w_i,d_i,\varepsilon_i)[d_i \int_{-\infty}^{z_i\gamma} f(v|\varepsilon)dv + (1-d_i)\int_{z_i\gamma}^{\infty} f(v|\varepsilon)dv] dF(\varepsilon). \end{aligned} \tag{2}$$

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

$$\begin{aligned} f(y_i|w_i,d_i) &= \int_{-\infty}^{-\infty} f(y_i|w_i,d_i,\varepsilon_i)[d_i\Phi^*(\varepsilon)+(1-d_i)*(1-\Phi^*(\varepsilon)]dF(\varepsilon) \\ &\text{where } \Phi^*(\varepsilon)=\Phi\left[\frac{z_i\alpha+(\rho/\sigma)\varepsilon}{\sqrt{1-\rho^2}}\right] \end{aligned} \tag{3}$$

and $\Phi(\cdot)$ 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 $E[y_i \mid w_i, d_i]$. He assumes the standard linear exponential specification for the mean of the count variable:

$$E[y_i|w_i,d_i,\varepsilon]=\exp\{x_i\beta+d_i\gamma+\varepsilon\} \tag{4}$$

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 an 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 $f(y_i \mid x_i, d_i, \varepsilon)$ 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 $f(y_i \mid x_i, d_i, \varepsilon)$ either.

Second, all the parameters are separately identified, more specifically ρ and σ. Given that the variance covariance matrix in (1) needs to be positive definite, we may reparameterize 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 ρ 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 (see Coulson, 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. This statistics has 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 (count) $j=1,2,\ldots,J$ is given by:

$$p_j = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{c_j \leq y_i < c_{j+1}} \tag{5}$$

The expected frequency for the interval $\hat{p}_j$ 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 ε 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:

$$f(y|w) = f(y,1|w) + f(y,0|w) \tag{6}$$

Then 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 $\hat{p}_j$. With this in mind, a very simple goodness of fit measure is given by the sum over j of the absolute differences $\Delta_j = |p_j - \hat{p}_j|$. 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 may consult the appendix on computational methods.

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.

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 where 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 $\psi_p(\lambda,a)$. Thus, following Cameron and Johanson (1997):

$$
\begin{aligned}
&f_{SNP}(y|w,d,\varepsilon)=\frac{(\sum_{i=1}^{K} a_i y^{\,i})^2 P(y|\lambda)}{\Psi(\lambda,a)},\\
&P(y|\lambda)=\frac{\exp(-\lambda)\lambda^y}{y!}\\
&\lambda=\exp(\beta w+\gamma d+\varepsilon).
\end{aligned}
\tag{7}
$$

Estimation is done by maximum likelihood using (3) and (7). 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:

Schwarz: $BIC=-2\ ln\ (L)+P\ ln(N)$
Consistent Akaike: $CAIC=-2\ ln\ (L)+P\ (ln\ (N)+1)$

where P represents the number of parameters to be estimated. 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)

$$E(y|w,d,\varepsilon)=\sum_{i=0}^{K}\sum_{i=0}^{K}a_i a_j m_{i-j}(\lambda), \qquad (8)$$

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 (8) 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 (8) 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 $E[y \mid x,d]$ 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:

$$E(y|x,d)=\sum_{y=1}^{\infty} y f(y|x,d), \quad where\ f(y|w,d)=\frac{f(y,d|w)}{f(d|w)} \tag{9}$$

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 the estimated coefficients as

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

Thus, the percentage mean effect of the change given by dummy regressor x_j is given by

$$\frac{E(y|x_{i(j)},x_{ij}=1)-E(y|x_{i(j)},x_{ij}=0)}{E(y|x_{i(j)},x_{ij}=0)}. \tag{11}$$

Strictly speaking, this in only valid for dummy regressors. For the continuous case we would use the standard notion of partial derivative. Notice that this conditional expectation is a function of the covariates observations. Therefore we get a vector of such quantities from $i=1,\dots,N$. To summarize this information we will report two different measures.

The first one computes the quantity in (11) 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) conditional expected count. That is, at the upper (lower) point we have the conditional distribution of the count with the largest (smallest) conditional mean. 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.

4. SOME APPLICATIONS

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 B.1 in appendix B 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 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.

The first columns of Table B.2 in appendix B shows the results of Nonlinear Least Squares (NLS), TSM and WNLS estimation methods. 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 1, 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 2, 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. The results for the K=1 specification did not imply a significant change with respect to the K=0 and in fact, the single polynomial coefficient shows a small significance. Changes start to appear when we move sequentially to the K=2,3,4 models. 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 2 and Figure 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 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, the model with K=2 is the preferred one for any of the Schwarz and Consistent Akaike info criteria reported. Given that the first coefficient of the polynomial of the model (a_1) with K=2 shows a small significance, it is expected that these results would improve if we restricted this coefficient to be zero.

Table B.2 in the appendix B shows 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 PFIML. Although the change is not important in size, the two coefficient do not have the same structural meaning. In principle the researcher shouldn't 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 (11). Figure 3 shows the distribution of the percentage change across individuals. Notice that the Poisson distribution overpredicts the impact of vehicle ownership by putting more mass on higher percentages.

On the other hand, Table 4 shows the change in mean due to vehicle ownership as well as percentage change 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.

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.

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, 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 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). 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 5. 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 therefore 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). Finally, seven observations with zero or negative family income were deleted.

With this specification we calculated the NLS, TSM and WNLS estimators (see results at Table B.4 in appendix B. 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. The low significance of the presumably endogenous variable (PRIVINS) and of the RHO coefficient is a symptom of no-endogenous determination of this variable. 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.

We estimated the PFIML and PPFIML models up to a polynomial of third order. Recall that 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 RHO coefficients (which shows a negative sign).

The comparison of the empirical and predicted probabilities in Table 7 and Figure 2 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 6). 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 Schwarz. 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 6. 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 don't 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 8 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 covariate 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 estimation is far away from the upper extreme (2.5%) casting doubts on the suitability of NLS when imposing the restriction of identical estimated percentage change to all of the individuals.

Figure 4 shows the distribution of percentage change in mean 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.

5. CONCLUSIONS

In this paper we contemplate the scenario proposed by Terza (1998) where unobserved heterogeneity in a count model is correlated with a dummy regressor. Full information maximum likelihood will allow us to obtain precise estimates which is crucial for positive and normative purposes made on the basis of the inference. However this method imposes several restrictions on the conditional distribution of the counts. We propose and fit an alternative and 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. Consequently 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.

Acknowledgements: We wish to thank Michael Creel for useful help and advice, and professor Joseph Terza for kindly providing the data and the TSM and WNLS programs. We also wish to thank professors Markus Glaser and Frank Windmeijer and participants at the ENTER meeting at University College of London. All remaining errors are our responsibility. We benefitted from financial support of the Comissionat per a Universitats i Recerca de la Generalitat de Catalunya grant no. 1997FI-436, Universitat Autònoma de Barcelona AP92-34967274 and from Spanish Ministry of Education DGICYT PB96-1160.

APPENDIX A. COMPUTATIONAL ISSUES

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) is based on the Gauss-Legendre quadrature (INTQUAD package). The procedure requires to define fixed 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 variable (the σ parameter in our specification). Then, the objective function was optimized using the Broyden-Fletchell-Golden-Shannon algorithm. Several runs were performed using different starting values. This derivative-based algorithm was found to be enough to find the global optimum in the case of k=0. However, for higher degrees of the polynomial size we found that the algorithm often converged to local optima. Local optima is a problem often encountered when using series expansion.

Then we decided to implement a more robust search method like Simulated

Annealing (SA) using the code written by E.G. Tsionas. 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 pair of 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 or hybrid methods. In general we found that the best result of several trials with a BFGS coincided with the SA optima for K=1 and K=2, although finding such a point with higher polynomial degrees was found to be harder. We also used the algorithm implemented in Cameron and Johansson (1997) which basically combines random search with a derivative-based method. We found that this algorithm was not robust enough for high order polynomials in our particular context.

Moreover, we found that the Health data set was more problematic than the data from Terza. Here we found that the Hessian was ill-conditioned even for the K=0 case. A first exploration of the correlations among dependent variables did not find a significantly high degree of pairwise linear dependence. However, the condition number of the covariates inner product matrix was abnormally high (1497). This 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 ($\xi = \frac{\varepsilon}{\sqrt{2}\sigma}$), equation (3) yields

$$f(y,d|w) = \frac{1}{\sqrt{\pi}(2\Phi(a)-1)}\int_{-a}^{a} f(y|d,w,\xi)[d\Phi(\frac{z'\alpha+\sqrt{2}\rho\xi}{\sqrt{1-\rho^2}})+ (1-d)(1-\Phi(\frac{z'\alpha+\sqrt{2}\rho\xi}{\sqrt{1-\rho^2}}))\exp(-\xi^2)$$

Since ξ 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 σ. 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 code which was found to be very useful not only for this application but in many other contexts (all code is available from the authors on request).

Finally, although we required more computing-time (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.

APPENDIX B: VARIABLE DESCRIPTION AND MODEL ESTIMATES

Table B.1. *Number of trips by household. Description of variables*

Variable	*Mean*	*Std.*	*Description*
			Endogenous
Tottrips	4.5511	4.9351	Number of trips by members of the household in the 24 hrs. period prior to the interview
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 business or pleasure
Hhmem	2.9289	1.6127	number of individuals in the household
DistoCbd	0.2887	0.4932	distance to the central business district in kilometers. Divided by 30.
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. Divided by 5.
RealInc	0.8042	0.9197	household income divided by median income of census tract in which household resides. Divided by 3.
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

Table B.2. *Number of trips by households. Model estimates.*

Variables	NLS	TSM	NWLS	K=0	K=1	K=2	K=3	K=4
				Count variable equation (X,d)				
Constant	-0.6006	-1.4458	-1.0055	-1.3867	-1.3418	-2.4997	-2.4112	-2.6385
	0.2251	0.5289	0.1818	0.2170	0.2239	0.3878	0.4046	0.5224
Workschl	-0.5273	-0.5543	-0.3633	-0.3407	-0.3515	-0.4101	-0.4168	-0.4353
	0.1438	0.1474	0.1288	0.1279	0.1310	0.1401	0.1725	0.1596
Hhmem	0.1663	0.1487	0.1349	0.1507	0.1574	0.1874	0.1949	0.1893
	0.0276	0.0313	0.0287	0.0190	0.0181	0.0285	0.0299	0.0297
Distocbd	-0.0049	-0.2688	-0.0573	-0.0442	-0.0419	-0.0408	-0.0396	-0.0709
	0.0045	0.1722	0.0242	0.0435	0.0363	0.0821	0.6716	0.0688
Areasize	-0.037	-0.0088	0.0384	0.0492	0.0618	0.0840	0.0991	0.0630
	0.0976	0.1006	0.0853	0.0738	0.0837	0.1031	0.1592	0.1748
Fulltime	0.1890	0.1059	0.2207	0.2464	0.2612	0.3052	0.3138	0.3176
	0.0487	0.1019	0.0736	0.0427	0.0466	0.0568	0.0814	0.0950
Distonod	0.0047	0.0216	0.0197	0.0215	0.0215	0.0291	0.0287	0.0330
	0.0024	0.0128	0.0136	0.0093	0.0109	0.0112	0.0217	0.0182
Realinc	0.0139	0.0200	0.0071	0.0822	0.0676	0.1145	0.0939	0.0257
	0.0162	0.0522	0.0514	0.0309	0.0258	0.0281	0.0323	0.0419
Weekend	-0.1557	-0.1650	-0.0296	-0.0974	-0.0956	-0.1134	-0.1126	-0.1260
	0.1122	0.1150	0.0835	0.0805	0.0788	0.0891	0.1310	0.0999
Ownveh	1.6070	2.7960	2.0792	2.0601	2.0510	2.3074	2.2826	2.2490
	0.1859	0.6134	0.3121	0.2563	0.2575	0.2990	0.3252	0.3539
a1					-0.0272	0.0537	0.0349	-0.1885
					0.0015	0.1458	0.1492	0.4225
a2						0.2316	0.2102	0.6432
						0.1542	0.0200	0.6644
a3							-0.0056	-0.1022
							0.0050	0.1630
a4								0.0142
								0.0232
				Binary dependent variable equation (Z)				
Constant		-0.6335		-0.5476	-0.5500	-0.5057	-0.5042	-0.5186
		0.2378		0.3626	0.3613	0.3471	0.7786	0.3869
Workschl		0.1525		0.3160	0.3118	0.3177	0.3154	0.3241
		0.2652		0.3522	0.3175	0.3137	0.4776	0.3495
Hhmem		0.0036		0.0511	0.0487	0.0521	0.0484	0.0513
		0.0687		0.0860	0.0691	0.0739	0.0798	0.1144
Distocbd		0.6929		0.6652	0.6691	0.6497	0.6520	0.6653
		0.3998		0.3676	0.3979	0.3734	0.8659	0.4355
Areasize		-0.2065		-0.2492	-0.2533	-0.2576	-0.2620	-0.2481
		-1.2423		0.1528	0.1559	0.1568	0.1650	0.1836

Fulltime	0.8718		1.0080	1.0045	0.9949	0.9931	0.9875
	0.1559		0.1879	0.1821	0.1764	0.1933	0.1918
Adults	0.3815		0.2678	0.2707	0.2516	0.2530	0.2470
	0.1456		0.1762	0.1835	0.1796	0.3137	0.1831
Distonod	0.0484		0.0520	0.0524	0.0496	0.0501	0.0472
	0.0332		0.0315	0.0327	0.0317	0.0347	0.0318
Realinc	0.4724		0.3378	0.3510	0.3312	0.3434	0.3730
	0.1774		0.2276	0.2349	0.2321	0.3160	0.2310
Variance-Covariance Parameters							
ρ		-1.1193	-0.6974	-0.6769	-0.6715	-0.6596	-0.6814
			0.0325	0.0306	0.0732	0.0634	0.0942
σ			0.7287	0.7330	0.9342	0.9261	0.8936
			0.1393	0.1416	0.1693	0.1654	0.1944

NOTE: Asymptotic standard error in the bottom row of each cell

Table B.3. *Number of physician office visits. Description of variables*

Variable	*Mean*	*Std.*	*Description*
			Endogenous
OFP	6.3590	6.9293	*Number of physician office visits*
PRIVINS	0.7780	0.4153	*=1 if the person is covered by private health insurance*
			Exogenous
EXCLHLTH	0.1150	0.3193	*=1 of self-perceived health is excellent*
POORHLTH	0.1087	0.3115	*=1 if self-perceived health is poor*
NUMCHRON	0.1503	0.1318	*Number of chronic conditions (cancer, heart attack, gall bladder problems, emphysema, arthritis, diabetes, other heart disease) Divided by 10.*
ADLDIFF	0.2162	0.4119	*=1 if the person has a condition that limits activities of daily living*
AGE	0.7411	0.0651	*age in years. Divided by 100*
BLACK	0.0544	0.2269	*=1 if the person is African American*
MALE	0.4083	0.4918	*=1 if the person is male*
MARRIED	0.5740	0.4948	*=1 if the person is married*
SCHOOL	0.5755	0.1902	*Number of years of education. Divided by 20.*
FAMINC	0.0629	0.0663	*Family income in $10.000.Divided by 50.*
EMPLOYED	0.1188	0.3238	*=1 if the person is employed*
MEDICAID	0.1201	0.3253	*=1 if the person is covered by Medicaid*

Table B.4. *Number of physician office visits. Model estimates*

Variable	NLS	TSM	WNLS	K=0	K=1	K=2	K=3
			Count Variable Equation				
Constant	1.6694	1.8968	1.4729	0.6506	0.7521	-0.0679	0.0629
	0.5070	0.6525	0.7069	0.4680	0.4756	0.5612	0.5473
Exclhlth	-0.5103	-0.5289	-0.5275	-0.5522	-0.5491	-0.6500	-0.6349
	0.1292	0.1342	0.1443	0.1231	0.1175	0.1426	0.1323
Poorhlth	0.0640	0.0352	0.1193	0.1384	0.1463	0.1686	0.1761
	0.1198	0.1310	0.1515	0.1201	0.1079	0.1421	0.1233
Numchron	1.3850	1.4079	1.5620	1.8261	1.9029	2.1441	2.1783
	0.2741	0.2802	0.2938	0.2648	0.2635	0.3217	0.2918
Adldiff	0.1190	0.0829	0.1745	0.1732	0.1188	0.2031	0.1493
	0.1243	0.1399	0.1422	0.1116	0.1091	0.1246	0.1285
Age	-0.7473	-0.7344	-0.9740	-0.8244	-0.7007	-0.9850	-0.8429
	0.6462	0.6444	0.6370	0.6049	0.6055	0.7070	0.7058
Black	-0.0828	-0.2309	-0.0513	-0.0898	-0.1782	-0.1045	-0.1786
	0.1771	0.2861	0.4112	0.1813	0.1877	0.2152	0.2188
Male	-0.0117	-0.0120	0.0039	-0.0103	-0.0334	-0.0137	-0.0360
	0.0922	0.0916	0.0825	0.0807	0.0805	0.0931	0.0917
Married	-0.1063	-0.1064	-0.1387	-0.0999	-0.0901	-0.1167	-0.1028
	0.0941	0.0944	0.0903	0.0850	0.0824	0.0993	0.0953
School	0.4602	0.6276	0.5925	0.6515	0.7601	0.8017	0.8815
	0.2189	0.3591	0.5048	0.2598	0.2326	0.3213	0.2687
Faminc	-0.1880	-0.0296	0.0165	-0.4799	-0.0959	-0.5861	-0.1917
	0.4753	0.7815	0.8663	0.6057	0.6037	0.6741	0.6793
Employed	0.0409	0.0165	-0.1004	-0.1447	-0.1743	-0.1848	-0.2067
	0.1730	0.1743	0.1551	0.1030	0.1047	0.1213	0.1205
Medicaid	0.4189	0.4088	0.4020	0.4753	0.5082	0.5614	0.5846
	0.1288	0.1318	0.1536	0.1293	0.1372	0.1539	0.1528
Privins	0.3416	-0.0546	0.6641	1.0061	0.7387	1.1671	0.9015
	0.1117	0.6980	0.9701	0.2031	0.1769	0.2314	0.2114
A1					-0.0292	0.0195	-0.0055
					0.0007	0.0873	0.0793
A2						0.1911	0.1672
						0.0892	0.0093
A3							-0.0049
							0.0001
			Binary Dependent Variable Equation				
Exclhlth		-0.2374	-0.2374	-0.2191	-0.2339	-0.2191	-0.2315
		0.1688	0.1688	0.1609	0.1644	0.1613	0.1639
Poorhlth		-0.2395	-0.2395	-0.2040	-0.2255	-0.2039	-0.2220
		0.1679	0.1679	0.1612	0.1658	0.1618	0.1653

Adldiff	-0.3545	-0.3545	-0.3455	-0.3498	-0.3459	-0.3491
	0.1320	0.1320	0.1252	0.1282	0.1255	0.1279
Black	-1.0987	-1.0987	-1.0780	-1.0971	-1.0821	-1.0972
	0.2082	0.2082	0.2018	0.2070	0.2021	0.2065
School	1.5843	1.5843	1.5140	1.5640	1.5123	1.5547
	0.1543	0.1543	0.1532	0.1559	0.1544	0.1556
Faminc	3.4645	3.4645	3.3878	3.4382	3.4049	3.4312
	1.1158	1.1158	1.1696	1.2255	1.1725	1.2171
Employed	-0.3017	-0.3017	-0.2750	-0.2880	-0.2755	-0.2858
	0.1739	0.1739	0.1739	0.1773	0.1742	0.1769
Variance-Covariance Parameters						
Rho		0.3379	-0.4478	-0.2616	-0.4235	-0.2813
			0.1221	0.1303	0.1352	0.1488
Sigma			1.0835	1.0518	1.3234	1.2792
			0.1049	0.0684	0.1163	0.0696

NOTE: Asymptotic standard error in the bottom row of each cell

Bibliography

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

------- (1988b), "Chi-Square Diagnostic Tests for Econometric Models. Introduction and Applications", *Journal of Econometrics*, 37, 135-156.

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

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.

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

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

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.

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.

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

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.

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.

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

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

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

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

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

Gurmu, S. (1997), "Semi-Parametric Estimation of Hurdle Regression Models with an Application to

Medicaid Utilization", *Journal of Applied Econometrics*, 12, 225-242.

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.

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

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.

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.

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.

Mullahy, J. (1998), "Much Ado about Two: Reconsidering Retransformation and the Two-Part Model in Health Econometrics", 17, 247-281.

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.

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.

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

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.

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

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

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 Econometric*s, 12, 281-294.

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

TABLES

Table 1. *Number of trips by households. Descriptive statistics*

Number of observations	577.00
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 sample mean

Table 2. *Number of trips by household. Fitted vs. Empirical*

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
≥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. *Number of trips by households. Goodness of fit measures and likelihood ratio tests*

	Specifications				
	K=0	**K=1**	**K=2**	**K=3**	**K=4**
Num. Param	21	22	23	24	25
Sample Size	577	577	577	577	577
*(1/N)*ln(L)*	-2.668	-2.665	-2.655	-2.654	-2.6522
Cons. Akaike	3233.991	3237.925	3233.815	3239.374	3245.186
Schwarz BIC	3212.991	3215.925	3210.815	3215.374	3220.186
	P-Values				
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

NOTE: The second part of the table shows the P-values of likelihood ratio tests. Null hypothesis models are in rows and alternative in columns.

Table 4. *Number of trips by Households. Sensitivity Analysis to OWNVEH*

(X,Z) Point		**K=0**	**K=2**
Upper Extreme Covariates[A]	Ownveh	31.9	27.9
		8.2	34.5
	Not Ownveh	18.0	22.5
		7.9	11.2
	% Change	77.5	24.5
		46.0	210.7
Lower Extreme Covariates[B]	Ownveh	1.2	1.2
		0.4	0.8
	Not Ownveh	1.3	2.6
		0.6	1.9
	% Change	-11.1	-52.2
		22.9	13.8
Mean Covariates[C]	Ownveh	4.6	4.7
		0.1	0.1
	Not Ownveh	1.6	2.0
		0.2	0.3
	% Change	179.3	136.8
		47.5	40.7

NOTE: *Each rows represents the estimate of the expectation of dep. variable conditional on OWNVEH=1, OWNVEH=0 and the percentage change from first to second at three different covariates points. Standard estimates appear below entries.*
[A]*WORKSCHL=0, HHMEM=13, DISTOCBD=0, AREASIZE=1, FULLTIME=4, DISTONOD=10, REALINC=10, WEEKEND=0*
[B]*WORKSCHL=1, HHMEM=1, DISTOCBD=10, AREASIZE=0, FULLTIME=0, DISTONOD=0.2, REALINC=0.02, WEEKEND=1*
[C]*WORKSCHL=0.26, HHMEM=2.92, DISTOCBD=0.28, AREASIZE=0.37, FULLTIME=0.97, DISTONOD=2.02, REALINC=0.8, WEEKEND=0.22*

Table 5. *Number of physician office visits. Descriptive statistics*

Number of observations	791.0
Mean	6.3
Variance	48.0
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 6. *Number of physician office visits. Goodness of fit measures and likelihood ratio tests*

	Specifications			
	K=0	**K=1**	**K=2**	**K=3**
Num. Param	23	24	25	26
Sample Size	791	791	791	791
*(1/N)*ln(L)*	-3.3097	-3.3039	-3.2966	-3.2922
Cons. Akaike	5412.46	5410.95	5407.11	5407.76
Schwarz BIC	5389.46	5386.95	5382.11	5381.76
	P-Values			
K=0		0.00243	0.0000	0.0000
K=1			0.0006	0.0000
K=2				0.0080

NOTE: The second part of the table shows the P-values of likelihood ratio tests. Null hypothesis models are in rows and alternative in columns.

Table 7. *Number of physician office visits. Predicted vs. Empirical*

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
≥7	0.3502	0.3401	0.3383	0.3543	0.3546
Sum of differences($x10^{-4}$)		1.48	1.49	0.37	0.39
Goodness of fit (Andrews)		21.21	21.784	2.39	2.52
P-Value		0.0034	0.0027	0.9347	0.9254

Table 8. *Number of physician office visits. Sensitivity analysis to PRIVINS*

(X,Z) Point		K=0	K=2
Upper Extreme Covariates	Privins	30.6	28.8
		1.4	0.9
	Not Privins	30.9	28.1
		2.6	2.1
	% Change	-0.9	2.5
		6.6	5.9
Lower Extreme Covariates	Privins	0.8	0.8
		0.3	0.3
	Not Privins	0.5	0.4
		0.2	0.2
	% Change	53.2	68.4
		18.4	24.7
Mean Covariates	Privins	6.8	6.6
		0.2	0.2
	Not Privins	4.4	4.5
		0.4	0.4
	% Change	52.0	45.7
		16.9	15.7

NOTE: *Each rows represents the estimate of the expectation of dep. variable conditional on PRIVINS=1, on PRIVINS=0 and the percentage change from first to second at three different covariates points. Standard estimates appear below entries.*
[A]*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*
[B]*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*
[C]*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*

FIGURES

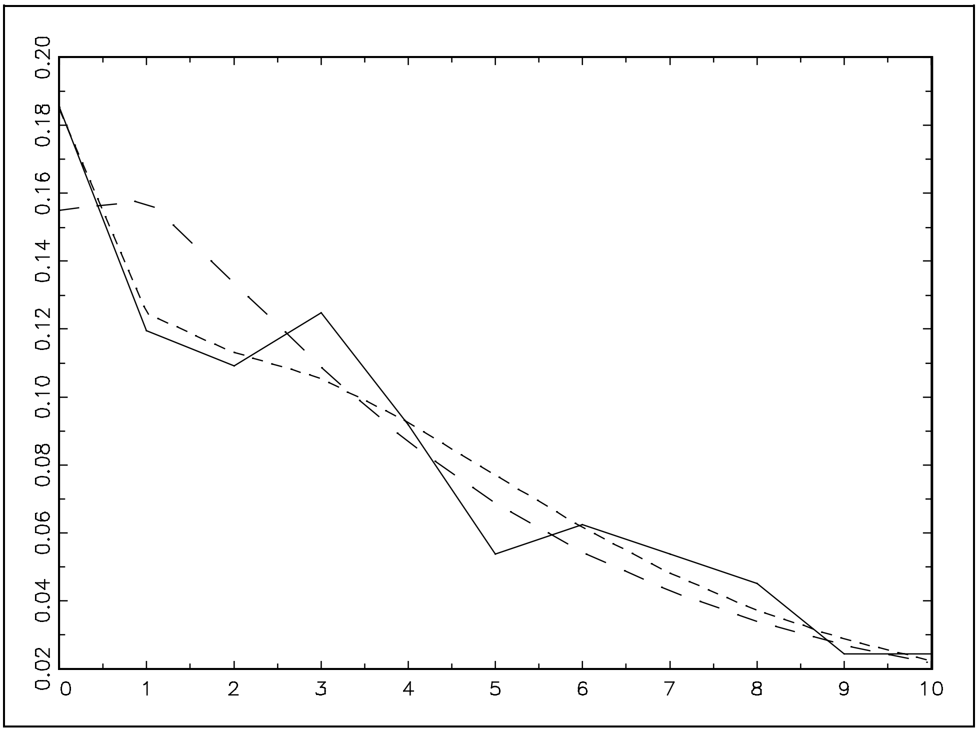

Figure 1. Number of trips. The figure plots the empirical frequency (—) of the counts in the horizontal axis and the predicted probabilities for the K=0 model (---) and the K=2 model (···).

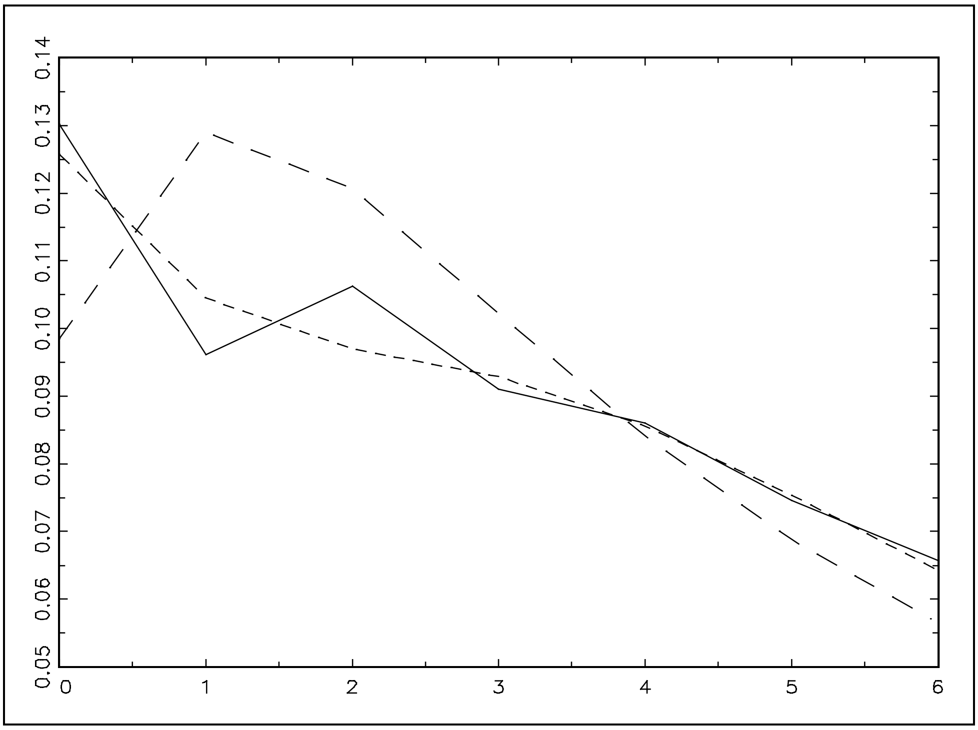

Figure 2. Number of physician office visits. The figure plots the empirical frequency (——) of the counts in the horizontal axis and the predicted probabilities for the K=0 model (---) and the K=2 model (···).

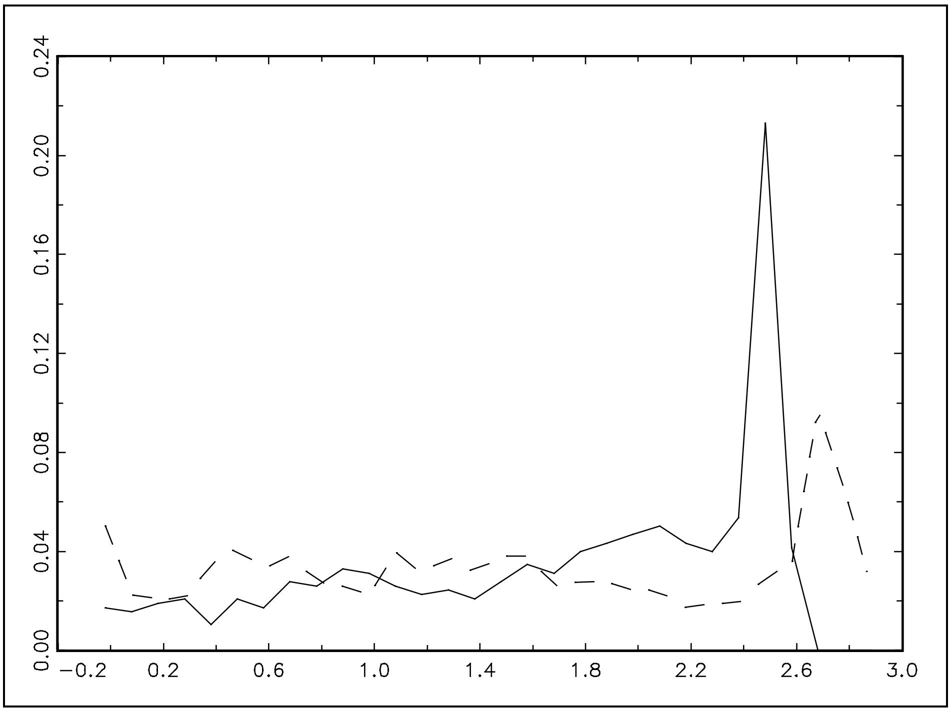

Figure 3. Number of trips by household. The figure shows the distribution of the percentage change in the mean effect due to vehicle ownership (see section 3) for K=0 (——) and K=2 (---).

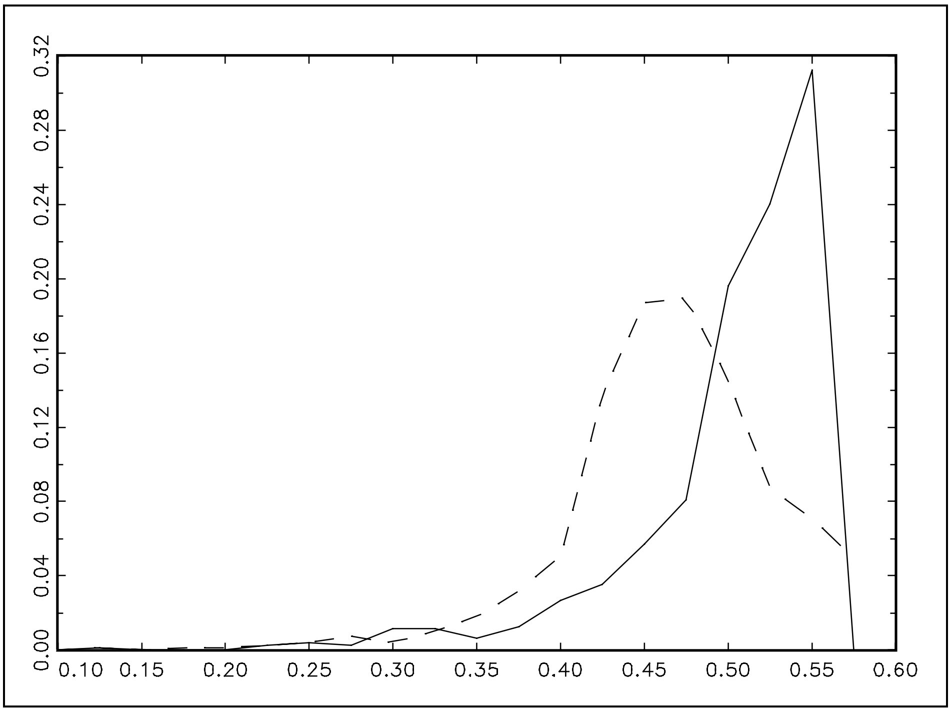

Figure 4. Number of physician office visits. The figure shows the distribution of the percentage change in the mean effect due to insurance status (see section 3) for K=0 (——) and K=3 (---).