11

Nonlinear Models for Health and Medical Expenditure Data

Heath sector variables are seldom continuous and fully observed. For example, they can be discrete (e.g., death), censored (e.g., health care expenditure), integer counts (e.g., visits to doctor), or durational (e.g., time to death). Multivariate analysis of such dependent variables requires nonlinear estimation. In this chapter, we consider the main (parametric) nonlinear estimators that are of relevance to the analysis of health sector inequalities. The literature is extensive, and our coverage is necessarily rudimentary, with a focus on practicalities rather than theory.

Binary dependent variables

There are numerous examples of health sector variables that take only two values—dead/alive, ill/not ill, stunted/not stunted, goes to doctor/doesn't go to doctor, and so on. In some cases, there are only two possible values of the underlying characteristic, for example, dead/alive. In other cases, the underlying characteristic is continuous, for example, degrees of illness, but only two categories are observable in the data—ill/not ill.

Let y_i be the characteristic of interest. Conventionally, $y_i = 1$ indicates that observation i possesses the characteristic, for example, illness, and $y_i = 0$ indicates that it does not. In general, a model of binary response can be defined by the following:

$$E[y_i | \mathbf{X}_i] = \Pr(y_i = 1 | \mathbf{X}_i) = F(\mathbf{X}_i\beta) \tag{11.1}$$

where $E[\ ]$ and $\Pr(\ )$ indicate expected value and probability, respectively. Different functional forms for $F(\ )$ define different specific models. For example, in the linear case, $F(\mathbf{X}_i\beta) = \mathbf{X}_i\beta$, we have the linear probability model (LPM). It is often claimed that the LPM can be consistently estimated by ordinary least squares (OLS). Horrace and Oaxaca (2006) prove that this is true only in the restrictive case that $\mathbf{X}_i\beta$ has a zero probability of lying outside the (0,1) range. A related problem is that the predicted probability given in equation 11.1 is not constrained to the (0,1) range, making results difficult to interpret in such circumstances. A further problem is that the errors are nonnormal and heteroscedastic, and so the estimator is not efficient and conventional standard errors are invalid. That can be (partially) fixed by weighted least squares.

An obvious, and common, response to these problems with OLS is to choose some functional form for $F(\ )$ that constrains estimated probabilities to lie in the (0,1) range. The two most popular choices are the cumulative standard normal

distribution, which gives the probit model, and the cumulative standard logistic distribution, which gives the logit model. Thinking about binary responses being driven by some underlying but unobservable (latent) characteristic helps to motivate such models. For example, let y_i^* indicate propensity to contract illness. When this crosses some threshold, say $y_i^* > 0$, the individual is ill. Specifying the latent variable to be a linear function of observable and unobservable factors, $y_i^* = \mathbf{X}_i\beta + \varepsilon_i$, and choosing a distribution for the error term as either standard normal or logistic gives the probit and logit models. Estimation is carried out by maximum likelihood.

Because the normal and logistic distributions are similar, the choice of a probit or logit specification is not important in most cases. Care must be taken not to compare probit and logit coefficients directly, however. In both cases, parameters are estimable only up to a scaling factor, equal to the unknown standard deviation of the error, which is not estimable given the binary nature of the dependent variable. Only the relative, not the absolute, effect of explanatory variables are estimable. Because variances differ between the normal and logistic distributions, logit coefficients must be multiplied by 0.625 to be comparable with probit coefficients (Amemiya 1981). Dividing probit estimates by 2.5 and logit estimates by 4 will make them comparable with those from the LPM (Wooldridge 2002).

Further care must be taken in the interpretation of estimates from latent variable models. The (scaled) parameters β give the (relative) partial effects on the latent index y_i^*, but these effects are usually not of primary interest. The partial effects on the probability of possessing the characteristic are more informative. For example, an estimate of how the probability of being sick changes with income is more easily interpreted than an estimate of how a latent index of sickness propensity varies with income. From equation 11.1, the estimated partial effect of a continuous regressor (X_k) on the (conditional) probability is given by the following:

$$\frac{\partial \Pr\left(\mathbf{X}_i\hat{\beta}\right)}{\partial X_{ki}} = \frac{\partial F\left(\mathbf{X}_i\hat{\beta}\right)}{\partial X_{ki}} = f\left(\mathbf{X}_i\hat{\beta}\right)\hat{\beta}_k, \tag{11.2}$$

where $f(\ )$ denotes the probability density function and is standard normal and logistic in the probit and logit cases, respectively. For a dummy regressor (X_K), the estimated partial effects can be calculated as follows:

$$F\left(\hat{\beta}_1 X_{i1} + \ldots + \hat{\beta}_{K-1} X_{iK-1} + \hat{\beta}_K\right) - F\left(\hat{\beta}_1 X_{i1} + \ldots + \hat{\beta}_{K-1} X_{iK-1}\right). \tag{11.3}$$

It is clear from equations 11.2 and 11.3 that these partial effects are not constants but are observation specific. There are two options, either calculate equations 11.2 and 11.3 at interesting values of all regressors, such as means or medians, or calculate the partial effect for each observation and take the average of these. The latter is preferable, but the former is somewhat more convenient. In large samples, the partial effect at the means should approximate the mean of the partial effects (Greene 2000). Calculating at medians, rather than means, ensures that values of dummy regressors are either 0 or 1 and, for regressors that are nonlinear transformations of variables, for example, quadratics and logs, it avoids the problem that the mean of the transformation is not the transformation of the mean. However, using medians can create problems of interpretation. For example, it may lead to infeasible combinations of the X's, setting all values to zero for a set of mutually exclusive indicators with less than 50 percent of the sample in each category. Such problems are avoided by computing the partial effect for each observation and then taking the mean or median of these.

Box 11.1 *Example of Binary Response Models—Child Malnutrition in Vietnam, 1998*

We compare the linear probability, logit, and probit models in estimating correlates of a discrete state of child malnutrition, defined as height-for-age more than two standard deviations below the average in a well-nourished (U.S.) population (see chapter 4). The data are for children younger than 10 years of age and are taken from the 1998 Vietnam Living Standards Survey (VLSS). This analysis complements that of a continuous measure of nutritional deprivation presented in the previous chapter.

In the following table we present estimates of the parameters of the respective models. Standard errors are adjusted for the clustered nature of the sample and are robust to general heteroscedasticity (see chapter 10). No adjustment is made for stratification, and sample weights are not applied, it being assumed that stratification is on exogenous factors (see chapter 10). There is a great deal of consistency across the estimators in the levels of significance of the coefficients. As suggested above, dividing logit and probit coefficients by 4 and 2.5, respectively, makes them approximately comparable to the LPM coefficients. For the coefficient on the male dummy, that gives 0.0669 (= 0.2675/4) for logit and 0.0646 (= 0.1614/2.5) for probit, which are both larger than the LPM coefficient of 0.0563. More

(continued)

Estimates from Binary Response Models of Stunting, Vietnam 1998 (children <10 years)

Dependent variable = 1 if height-for-age z-score less than –2

	LPM (OLS)	*Logit (MLE)*		*Probit (MLE)*	
	Coeff.	*Coeff.*	*Partial effect*	*Coeff.*	*Partial effect*
Child's age (months)	0.0079***	0.0403***	0.0100***	0.0245***	0.0097***
	(0.00075)	(0.00394)	(0.00100)	(0.00238)	(0.00100)
Child's age squared (/100)	–0.0053***	–0.0271***	–0.0068***	–0.0165***	–0.0066***
	(0.00058)	(0.00293)	(0.00074)	(0.00177)	(0.00071)
Child is male	0.0563***	0.2675***	0.0661***	0.1614***	0.0639***
	(0.01281)	(0.06072)	(0.01489)	(0.03688)	(0.01451)
(log) hhold. consumption per capita	–0.1849***	–0.9403***	–0.2347***	–0.5639***	–0.2248***
	(0.01726)	(0.09026)	(0.02255)	(0.05301)	(0.02116)
Safe drinking water	–0.0447*	–0.2017*	–0.0504*	–0.1208*	–0.0482*
	(0.02685)	(0.11669)	(0.02906)	(0.07146)	(0.02844)
Satisfactory sanitation	–0.057**	–0.3344***	–0.0822***	–0.1982***	–0.0782***
	(0.02306)	(0.11838)	(0.02860)	(0.06990)	(0.02728)
Years of schooling of head of household	0.0013	0.0047	0.0012	0.0028	0.0011
	(0.00219)	(0.01070)	(0.00267)	(0.00642)	(0.00256)
Mother has primary school diploma	–0.0041	–0.0106	–0.0027	–0.0079	–0.0031
	(0.02008)	(0.09218)	(0.02301)	(0.05571)	(0.02221)
Intercept	1.5681***	5.4812***		3.2734***	
	(0.13511)	(0.69589)		(0.41134)	
Sample size	5,218				

Note: Robust standard errors in parentheses. Adjusted for clustering and heteroskedasticity. Partial effects calculated at medians of regressors.

LPM = linear probability model, OLS = ordinary least squares, MLE = maximum likelihood estimator.

***, **, and * indicate significance at 1%, 5%, and 10%, respectively.

Box 11.1 *continued*

directly, we can compare the partial effects of the regressors on the probability that a child is stunted. For the LPM, these marginal effects are given by the coefficients themselves and so are constants. For the logit and probit models, we have calculated the partial effects at the median values of the regressors. In general, the estimated partial effects from logit and probit are very close and are larger in magnitude than those from the LPM.

In the next table, we summarize the distributions of the partial effects estimated from the probit model. This form of presentation makes it clear that partial effects vary across individuals. For example, the mean effect of satisfactory sanitation is to reduce the probability of stunting by 0.0689, from an estimated population average probability of 0.3737. In absolute terms, the strongest partial effect of satisfactory sanitation is a reduction in the probability by 0.0790, but this is from a predicted baseline probability for that individual of 0.5281. The weakest absolute effect is a reduction in the probability of only 0.0076, but this is large in relation to the respective baseline probability of 0.0118.

Partial effects can be calculated with respect to variables of inherent interest, rather than transformations of these. For example, in the table, we present the partial effect of a currency unit increase in household consumption, as well as the effect of a marginal increase in the log of consumption. Partial effects of variables entered in quadratic form, such as age, can be calculated but are of limited interest. The partial effect of age itself is a function of the partial effects of the first and second powers of age (given in the table). This function can be calculated but, given the quadratic nature of the function, the partial effect changes sign. It is of more interest to examine a picture of the quadratic function and locate its turning point (six years and two months, in this example).

Partial Effects on Probability That Child Is Stunted, Vietnam 1998 (children <10 years)

(derived from probit estimates in table above)

	Mean	*Std. dev.*	*Min*	*Max*
Child's age (months)	0.0086	0.00160	0.0008	0.0098
Child's age squared (/100)	–0.0058	0.00108	–0.0066	–0.0005
Child is male	0.0568	0.01045	0.0046	0.0643
(Log) Household consumption p.c.	–0.1982	0.03675	–0.2250	–0.0174
Household Consumption p.c. (D)	–0.0001	0.00007	–0.0006	–0.0000
Safe drinking water	–0.0430	0.00743	–0.0482	–0.0043
Satisfactory sanitation	–0.0689	0.01240	–0.0790	–0.0076
Years of schooling of head of hhold.	0.0010	0.00018	0.0001	0.0011
Mother has primary school diploma	–0.0028	0.00051	–0.0031	–0.0002

Note: Data are weigted. D = Vietnamese dong.
Source: Authors.

Computation

Stata, like many packages, has preprogrammed routines for probit and logit:

```
probit depvar varlist [pw=weight], robust
logit depvar varlist [pw=weight], robust
```

where `depvar` and *`varlist`* represent dependent and independent variables, respectively; `[pw=weight]` is optional to give weighted (on `weight`) estimates; and `robust` is optional for heteroscedasticity robust standard errors. If the survey data are from a cluster sample, standard errors can be corrected for within-cluster

correlation using the option `cluster(psu)`, where `psu` is a variable identifying the primary sampling units (see chapter 10).[1]

A special routine is available to give probit partial effects at specific regressor values:

```
dprobit depvar varlist [pw=weight], robust
```

By default, this calculates partial effects at the means. To obtain the effects at other values, such as medians, the following can be used:

```
local vars "varlist"
foreach x of local vars {
      qui sum `x' [aw=weight], d
      sca `x'_md=r(p50)
}
matrix define medians=(var1_md, var2_md,….)
dprobit depvar varlist [pw=wt], robust at(medians)
```

where *var1*, *var2* …. are the names of the regressors in *varlist*. There is no such preprogrammed routine for logit partial effects, but Stata's general routine for partial effects, `mfx`, can be used. Simply run a logit and afterward

```
mfx compute, at(median)
```

where `at()` specifies the values at which effects are to be calculated; `mean`, `median`, `zero`, defined values, or a combination of these can be selected. This is much slower than `dprobit`. It can be speeded up by requesting that standard errors not be calculated through the option `nose`.

To calculate partial effects for each observation, run `probit` or `logit`, then obtain predictions of the latent index (`xb`) and probability of a nonzero dependent variable (`p`) for each observation by

```
predict xb if e(sample), xb
predict p if e(sample), p
```

where `if e(sample)` is optional and restricts the prediction to observations used in the estimation. Define two locals containing the names of the continuous variables (e.g., `cont1, cont2`, etc.) and those of the dummy variables (e.g., `dummy1`, `dummy2`, etc.),

```
local cont "cont1 cont2 ..."
local dummies "dummy1, dummy2, ..."
```

For continuous regressors, define a variable that will be used to transform the coefficients

```
gen t_var=normden(xb)          | for probit
gen t_var=p*(1-p)              | for logit
```

and, using equation 11.2, obtain the partial effects from

```
foreach c of local cont {
      gen pe_`c'=t_var*_b[`c']
}
```

[1]If the survey is stratified and the analyst also wishes to take that into account in computation of the standard errors, Stata's survey estimators for probit/logit can be used.

For dummy regressors, use equation 11.3, and obtain the partial effects for probit from the following:

```
foreach d of local dummies {
      gen pe_`d'=p-norm(xb-_b[`d'])
      replace pe_`d'=norm(xb+_b[`d'])-p if `d'==0
}
```

For logit, use

```
foreach d of local dummies {
      gen pe_`d'=p-(exp(xb-_b[`d'])/(1+exp(xb-_b[`d'])))
      replace pe_`d'=(exp(xb+_b[`d'])/(1+exp(xb+_b[`d'])))-p
        if `d'==0
}
```

Finally, obtain summary statistics of the distribution of the following partial effects:

```
summ `cont' `dummies' [fw=weight], detail
```

where `[fw=weight]` applies weights and should be included where these exist.

This procedure will generate, for example, estimates of the population means of the partial effects. For inference, standard errors of these estimates would have to be generated by the delta method (Wooldridge 2002).

Limited dependent variables

A limited dependent variable is continuous over most of its distribution but has a mass of observations at one or more specific values, such as zero. The most important example in the health sector is medical expenditure, which is zero for many individuals over a survey recall period, such as 12 months. For example, in 1998 the average Vietnamese spent 153,000 Vietnamese dong (D) ($1 = 13,987D) out-of-pocket on medical care during a 12-month period, but 17 percent spent nothing at all.

There are a multitude of statistical approaches to modeling of a limited dependent variable—for example, the two-part model, the Tobit model, the sample selection model, hurdle models, and finite mixture models. For a comprehensive survey, see Wooldridge (2002). Here, we restrict attention to the most popular approaches to modeling medical expenditures. For an excellent survey of this literature, see Jones (2000). Equity analysis of medical expenditures may focus on their income elasticity, on variation in the price elasticity of health care with household income, on the responsiveness of medical expenditure to health shocks, or the extent to which this responsiveness is reduced by unequally distributed insurance coverage.

Two-part model

The most straightforward approach is the two-part model (2PM). In its most popular form, this comprises a probit (or logit) model for the probability that an individual makes any expenditure on health care and OLS, applied only to the subsample with nonzero expenditures, to estimate correlates of the positive level of expendi-

ture. Given that typically the distribution of medical expenditures is right skewed, invariably the log of expenditure is modeled in the second part OLS.

Application of OLS to only part of the sample raises the possibility of sample selection bias. The issue has been the subject of a great deal of discussion (Jones 2000). In summary, consistency of the 2PM for the model parameters rests on strong assumptions. Nonetheless, if the aim is simply to predict conditional means and not to make inferences about individual parameters, then the 2PM may perform reasonably well (Duan et al. 1983). On that basis, the model will often be adequate for analysis of health sector inequalities, where we simply want to *predict*, for example, medical expenditure conditional on income, age, gender, and so on.

Following Jones (2000), let the probability that medical expenditure (y_i) is positive be determined by observable $(\mathbf{X}_{1i})$ and unobservable (ε_{1i}) factors. Let $\ln(y_i)$ be the log of positive medical expenditure, with covariates $\mathbf{X}_{2i}$, and unobservable determinants ε_{2i}. Consistency of the 2PM is predicated on an assumption of conditional mean independence (Jones 2000).

$$E\left[\ln(y_i) \mid y_i > 0, \mathbf{X}_{2i}\beta_2\right] = E\left[\ln(y_i) \mid \mathbf{X}_{1i}\beta_1 + \varepsilon_{1i} > 0, \mathbf{X}_{2i}\beta_2\right] = \mathbf{X}_{2i}\beta_2 \tag{11.4}$$

In other words, conditional on expenditure being positive, the unobservable determinants of its log have zero mean. To justify the assumption, either unobservable factors that influence the positive level of expenditures (ε_{2i}) must be independent of those governing the probability of a positive expenditure (ε_{1i}), or the two error terms must have some peculiar joint distribution that gives a conditional distribution centered around zero. The latter would be an extreme and nontestable assumption (Jones 2000). The former assumption can possibly be supported under certain decision-making processes, for example, if the individual decides whether to seek treatment without considering how much to spend during the course of treatment. That rules out the possibility that the individual decides not to seek care because of the anticipated cost of a course of treatment. In support of such a sequential model of decision making, it might be claimed that the patient delegates all treatment decisions to the doctor. Empirically, however, such a defense is weak because typically survey data span a period of calendar time and not the duration of an illness episode (Deb and Trivedi 1997). Even if it is accepted that medical care decisions are made in a sequential manner, correlation between unobservables would still arise in cases in which common variables are omitted from the two stages of the decision-making process (Jones 2000).

The expected level of medical expenditure is given by the following:

$$E\left[y_i \mid \mathbf{X}_i\right] = \Pr\left(y_i > 0 \mid \mathbf{X}_{1i}\right) E\left[y_i \mid y_i > 0, \mathbf{X}_{2i}\right]. \tag{11.5}$$

Unfortunately, this value cannot be estimated directly when the second part of the model is estimated in logs, as is usually the case. This is known as the retransformation problem; we have to get back from logs to levels. Assumption 11.4 is not sufficient to identify 11.5. For possible solutions to the problem, see Jones (2000) and Mullahy (1998). This rather weakens the argument that the 2PM is reasonable when one is interested only in estimating the conditional means.

Tobit model

Whereas the 2PM assumes that two independent decisions lie behind medical expenditures, the Tobit model, at the other extreme, assumes a single decision. The

individual chooses the level of medical expenditure that maximizes his or her welfare. Positive expenditures correspond to desired expenditures. Zero expenditure represents a corner solution, in which income and/or preferences for health are so low that spending nothing on health care is best for the individual. The model can be described using the concept of a latent, desired level of expenditure:

$$y_i^* = \mathbf{X}_i\beta + \varepsilon_i, \qquad \varepsilon_i \sim IN(0, \sigma^2). \tag{11.6}$$

Observed expenditure is assumed to be related to the latent value by the following:

$$y_i = \begin{cases} y_i^* \text{ if } y_i^* > 0 \\ 0 \text{ otherwise.} \end{cases} \tag{11.7}$$

The assumption of a single decision-making process is most probably strong. It requires that before making contact with the health services, the individual has full information on the costs of alternative courses of treatment. It also rules out the possibility that the initial decision to seek treatment is made solely by the individual, while both the patient and the doctor influence the decision about the amount of treatment.

The Tobit model is estimated by maximum likelihood (ML). As a rule of thumb, Tobit ML estimates may be approximated by the OLS estimates from the 2PM divided by the proportion of nonzero observations in the sample (Greene 2000). Predicted medical expenditure over the whole sample is still based on equation 11.5, but the second term in the product is no longer given by equation 11.4 but by the following:

$$E\left[y_i \mid y_i > 0, \mathbf{X}_i\right] = \mathbf{X}_i\beta + \sigma\lambda_i, \quad \lambda_i = \frac{\phi\left(\mathbf{X}_i\beta / \sigma\right)}{\Phi\left(\mathbf{X}_i\beta / \sigma\right)} \tag{11.8}$$

where $\phi(\ )$ and $\Phi(\ )$ are the standard normal probability density and cumulative density functions, respectively, and λ_i is known as the inverse Mill's ratio (IMR).

Sample selection model

The sample selection model (SSM), or generalized Tobit, can be considered, somewhat informally, to lie midway between the extremes of the Tobit and the 2PM. Whereas the Tobit assumes a single decision process and the 2PM two independent decisions, the SSM allows for two interdependent decisions. The decision to seek medical care and the choice of how much to spend can be influenced by distinct but correlated observable and unobservable factors. In latent variable form, the model is given by the following:

$$y_{ji}^* = \mathbf{X}_{ji}\beta_j + \varepsilon_{ji}, \qquad j = 1, 2 \tag{11.9}$$

$$y_i = \begin{cases} y_{2i}^* \text{ if } y_{1i}^* > 0 \\ 0 \text{ otherwise.} \end{cases} \tag{11.10}$$

Assuming the two error terms are jointly normally distributed, the model can be estimated either by the Heckman two-step procedure or by ML. The former involves estimating a probit for the probability of nonzero expenditure, using the results to estimate the IMR and then running OLS on the nonzeros with the estimated IMR

included to correct for selection bias. That is, in the second stage, the following is estimated:

$$y_i = \mathbf{X}_{2i}\beta + \rho\sigma_2 \frac{\phi\left(\mathbf{X}_{1i}\hat{\beta}_1\right)}{\Phi\left(\mathbf{X}_{1i}\hat{\beta}_1\right)} + e_{2i}, \tag{11.11}$$

where ρ is the correlation coefficient between the errors, and σ_2 is the standard deviation of ε_{2i} ($\sigma_1 = 1$). The *t*-ratio for the IMR provides a test for selection bias. Standard errors must be corrected for the inclusion of the estimated IMR among the regressors. Packages programmed for the Heckman estimator will make the correction automatically. Efficiency gains can be realized through ML estimation.

Although the SSM is, in an informal sense, more general, this comes at the cost of making greater demands on the data with respect to identification. Given the nonlinearity of the IMR, equation 11.11 is identified even if the regressor matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ are identical, but in this case the Mill's ratio will be closely correlated with the other regressors and, consequently, parameters will not be estimated with precision. It is therefore preferable to have a variable that influences the decision of whether to spend anything on health care but, conditional on this, does not influence the positive level of expenditure. Such variables, however, are few and far between.

The Tobit and 2PM avoid this problem but only by assumption. The bottom line is that it is difficult to make an a priori case for any one model of medical expenditures. One should probably be most skeptical of the Tobit model and its assumption of a single decision process driving both zero and positive expenditures. In choosing between the 2PM and the SSM, it is necessary to consider the purpose of the analysis (prediction or parameter estimation), the likely degree of selection bias, and the information available to identify it.

Box 11.2 *Example of Limited Dependent Variable Models—Medical Expenditure in Vietnam, 1998*

We examine correlates of annual out-of-pocket expenditures on health care in Vietnam. We use data from the 1998 VLSS. Almost one-fifth (18%) of the observations made no expenditures on medical care. In addition to this mass at zero expenditure, the distribution has a long right tail. Given such skewness, one would expect a log transformation of the dependent variable to be appropriate, and the results confirm this. We make two comparisons, the 2PM with the SSM taking logs of positive expenditures in each case and the 2PM with the Tobit leaving the dependent variable in levels (see the table below).

Results from the maximum likelihood estimator of the SSM are given. These do not differ substantially from estimates obtained using the Heckman two-step procedure. Estimates of the coefficients of the selection equation display no substantial differences across the estimators. There are no differences in levels of significance. Coefficient estimates for the continuous parts of the models do show some differences, with those from the SSM generally of greater magnitude. There are some differences in levels of significance.

There is a positive and large degree of correlation between the two equation errors (0.847). The null of no correlation, and therefore no selection bias, is firmly rejected. In the absence of any variable that can plausibly be argued to affect the probability of positive expenditure but not its level, the correlation parameter is being identified through functional form alone. Graphical analysis confirms that, in this case, the inverse Mill's ratio is sufficiently nonlinear in its argument to avoid severe collinearity problems.

(continued)

Box 11.2 *continued*

Comparison of Two-Part and Sample Selection Model Estimates of Medical Expenditure, Vietnam 1998

Dependent variables: Participation = 1 if medical expenditure positive; Continuous = log of (positive) expenditure

	Two-part model				*Sample selection model*			
	Participation (probit)		*Continuous (OLS)*		*Participation (MLE)*		*Continuous (MLE)*	
	Coeff.	*Rob. SE*	*Coeff.*	*Rob. SE*	*Coeff.*	*Rob. SE*	*Coeff.*	*Rob. SE*
body mass index	–0.1382***	*0.0332*	–0.0800***	*0.0254*	–0.1117***	*0.0297*	–0.1430***	*0.0283*
(body mass index)^2	0.2820***	*0.0820*	0.1212*	*0.0643*	0.2265***	*0.0728*	0.2488***	*0.0709*
log(rental value of house)	0.3079***	*0.0434*	0.5065***	*0.0264*	0.3393***	*0.0350*	0.6262***	*0.0378*
satisfactory sanitation	–0.2160***	*0.0775*	–0.2362***	*0.0434*	–0.2183***	*0.0713*	–0.3283***	*0.0605*
house not of solid materials	0.0900*	*0.0528*	0.1896***	*0.0363*	0.0831*	*0.0459*	0.2279***	*0.0428*
attended school, no diploma	0.0527	*0.1110*	–0.2522***	*0.0386*	0.0173	*0.1023*	–0.2240***	*0.0638*
attended school & diploma	0.0985	*0.1320*	–0.1335***	*0.0482*	0.0674	*0.1221*	–0.0839	*0.0774*
head of hhold has diploma	–0.0563	*0.0570*	–0.1557***	*0.0391*	–0.0684	*0.0526*	0.1761***	*0.0462*
head of hhold school grade	–0.0025	*0.0078*	–0.0112**	*0.0049*	–0.0029	*0.0070*	–0.0118**	*0.0059*
						Rho	0.8470	*0.0195*
Sample size	27,368		22,645			Wald (Rho=0)	324.6	*p*=0.0000
Test slope parameters all zero	Wald = 515	*p* = 0.0000	F = 134.2	*p* = 0.0000			Wald = 3448	*p* = 0.0000

Note: All models also include a 3rd-degree polynomial in age, gender dummy, head of household dummy, quadratic in household size and regional dummies.
MLE = maximum likelihood estimator; Rob. SE = robust to hetero. and clustering standard error;
Rho = coefficient of correlation of errors; Wald (rho = 0) = Wald test of null of rho = 0.
***, **, and * significant at 1%, 5%, and 10%, respectively.

Comparison of the 2PM with the Tobit is a little less comforting (see following table). First, it is apparent that estimation in levels is less appropriate. The coefficient estimates differ substantially between the estimators and the scaling of the OLS coefficients, that is, dividing by the proportion of "positives" does not get us particularly close to the Tobit estimates. Mean predicted expenditure (over the full sample) from the Tobit model, at 374.2, is well above the actual mean of 157.2.

Box 11.2 *continued*

Comparison of Two-Part and Tobit Model Estimates of Medical Expenditures Vietnam 1998

Dependent variable: ***Level*** *of annual medical expenditure*

	Two-part model (OLS part)			*Tobit (MLE)*	
	Coeff.	*SE*	*Scaled coeff.*	*Coeff.*	*SE*
Body mass index	–19.21	*15.94*	–23.18	-47.53***	*15.11*
(body mass index)^2	42.11	*41.42*	50.81	100.69***	*37.64*
Log (rental value of house)	211.66***	*24.01*	255.43	249.15***	*8.86*
Satisfactory sanitation	–73.78***	*17.87*	–89.04	–111.55***	*16.05*
House not of solid materials	37.18***	*11.89*	44.87	51.34***	*12.36*
Attended school, no diploma	–54.36**	*22.93*	–65.61	–38.04**	*19.16*
Attended school & diploma	–31.23	*25.35*	–37.69	–6.14	*23.78*
Head of hhold has diploma	–10.14	*27.97*	–12.24	–21.09	*18.30*
Head of hhold school grade	–9.85**	*4.13*	–11.88	–9.03***	*2.13*
Sample size	22,645			27,335	
Test of all slope parameters zero	F = 14.29	$p = 0.0000$		LR = 1887	$p = 0.0000$

Note: All models also include a 3rd degree polynomial in age, gender dummy, head of household dummy, quadratic in household size and regional dummies.

Scaled coeff. = OLS coefficient divided by sample proportion with positive expenditure.

MLE = maximum likelihood estimator; SE = standard error; LR = Likelihood ratio test.

***, **, and * significant at 1%, 5%, and 10%, respectively.

Source: Authors.

Computation

Computation of the 2PM is straightforward. Run a `probit` for the probability of positive expenditure followed by OLS `(regr)` for the log of expenditure on the selected sample. Stata has a preprogrammed routine `heckman` for the SSM. For the (consistent) two-step estimator, use the following:

```
heckman depvar varlist, sel(depvar_s = varlist_s) twostep ///
   mills(imr)
```

where `depvar` is the continuous dependent variable (e.g., expenditures) and *`varlist`* associated regressors; `depvar_s` is a binary variable identifying the selected sample (those with positive expenditures) and *`varlist_s`* associated regressors; `mills(imr)` saves the inverse Mill's ratio and calls it `imr`. Omitting the `twostep` option gives the MLE, and with this `robust` and `cluster` adjusted standard errors can be requested.

To examine whether the Mill's ratio is nonlinear over its sample range, the following can be used:

```
predict xbsel if depvar_s==1, xbsel
twoway (scatter xbsel imr if depvar_s==1)
```

To estimate a Tobit model with censoring at zero, as in the example, use the following:

```
tobit depvar varlist, ll(0)
```

Count dependent variables

Many of the variables of interest in the health sector are nonnegative counts of events. For example, visits to the doctor, drugs dispensed, days ill, and so on. A count is a variable that can take only integer-values. Often, as with most health count variables, negative values are not possible. Typically, the distribution of such variables tends to be right skewed, often comprising a large proportion of zeros and a long right-hand tail. The discrete nature of a nonnegative count dependent variable and the shape of its distribution demand the use of particular estimators. For example, least squares would not guarantee that predicted values are nonnegative.

The most basic approach is to assume a Poisson process to describe the probability of observing a specific count of events over a fixed interval. That is, the probability of observing a count of y_i, conditional on a set of explanatory variables, $\mathbf{X}_i$, is assumed to be given by

$$\Pr(y_i \mid \mathbf{X}_i) = \exp(-\lambda_i)\lambda_i^{y_i} / y_i! \tag{11.12}$$

where exp() is the exponential function, $y_i!$ indicates y_i factorial, and λ_i is the conditional mean of the count and is usually specified as

$$\lambda_i = E[y_i \mid \mathbf{X}_i] = \exp(\mathbf{X}_i\beta). \tag{11.13}$$

A peculiarity of the Poisson distribution is that its mean and its variance are both equal to its one parameter, λ. This is often restrictive. In health applications, for example, the conditional mean is usually less than the conditional variance. In jargon, there is *overdispersion*. One consequence can be underprediction of the number of observations with zero counts; again, an empirical feature of many health care applications. Overdispersion can be allowed for, or rather imposed, through alternative distributional assumptions. For example, a negative binomial specification maintains the Poisson process (equation 11.12) but extends equation 11.13 to include an error term, for which a (gamma) distribution is assumed. As a result, the (conditional) variance of the count is restricted to be greater than its mean (Cameron and Trivedi 1986). The difference between the variance and mean, that is, the dispersion, can be specified as proportional to the mean (NegBin I) or a quadratic function of the mean (NegBin II) (Cameron and Trivedi 1986). The model can be further generalized by allowing the dispersion to vary across observations with a set of regressors.

Overdispersion is not the only reason a simple Poisson model may underpredict the number of zero counts. There may be a particular process responsible for generating zeros that is distinct from that generating other values of the count variable. One possibility, in the context of health care utilization, is a sequential decision-making process, as discussed in the previous section. This takes us back to the 2PM. In a count framework, the 2PM consists of a probit/logit (or Poisson/NegBin) to model the probability of a nonzero count followed by a count regression, such as Poisson or NegBin, applied to observations with positive counts only and allowing for the truncation at zero (Pohlmeier and Ulrich 1995). Independence is assumed between the two processes. Other possibilities are "zero-inflated" models and latent class models (Jones 2000, 318–24).

Unobservable heterogeneity, deriving time-invariant individual effects in a panel data context or community effects in a cross section, can be taken into account

in estimation of the Poisson model through a random-effects specification. Alternatively, with a fixed-effects specification of the Poisson, individual/community effects are eliminated. These are somewhat analogous to the random- and fixed-effects specification in a linear context discussed in chapter 10. The random effects specification is more efficient but requires an assumption that individual/community effects are independent of the regressors. A fixed-effects specification relaxes the assumption. Apart from taking unobservable heterogeneity into account, these methods have the further important advantage of relaxing the equi-dispersion restriction of the Poisson model (Wooldridge 2002).

Box 11.3 *Example of Count Data Models—Pharmacy Visits in Vietnam, 1998*

Annual visits to a pharmacy or drug peddler in Vietnam provides a good example of a distribution suited to the application of count data models. That is, there are a large number of zeros and a long right tail.

No. of pharmacy visits	*Frequency*
0	20865
1	3980
2	1899
3	846
4	434
5	197
6	79
7	52
8	25
9	5
10+	124

It should be acknowledged that we chose this distribution on the basis of its suitability for count analysis. With many count variables encountered in health applications, the dominance of zero values is much greater than in this example and the best option is simply to dichotomize the variable and use probit or logit to model the probability of a nonzero count.

Estimates and robust standard errors from a NegBin II model of pharmacy visits are given in the first two columns of the following table. NegBin II was chosen over NegBin I by comparison of the log-likelihood values. There is strong evidence of overdispersion as indicated by the magnitude of the dispersion parameter and the LR test, which decisively rejects the Poisson (equi-dispersion) specification.

Moving to a 2PM, there is some loss of significance, with significant effects in the first stage probit only. Restricting the count regression to positive values is not sufficient to remove overdispersion—a Poisson specification is still strongly rejected. Finally, we estimate a fixed-effects Poisson on all observations. The fixed effects are those of 194 communes. Point estimates from the FE Poisson are somewhat similar to those from NegBin II on the full sample, but there are large differences in levels of significance for some interesting variables. In particular, the household consumption effect becomes strongly significant. Apparently, the commune effects had initially confounded this (negative) income effect.

(continued)

Box 11.3 *continued*

Count Models for Annual Pharmacy Visits, Vietnam 1998

			Two-part model					
	NegBin II (all observations)		*Probit*		*Truncated NegBin II*		*Fixed-effects Poisson*	
	Coeff.	*Rob. SE*	*Coeff.*	*Rob. SE*	*Coeff.*	*Rob. SE*	*Coeff.*	*Rob. SE*
Log hhold. consumption per capita	–0.0314	*0.0648*	–0.0451	*0.0432*	0.0559	0.1303	–0.0710***	*0.0221*
Attended school, no diploma	–0.1696**	*0.0734*	–0.0771	*0.0547*	–0.1038	0.3238	–0.1422***	*0.0263*
Attended school & diploma	–0.1486	*0.0976*	–0.0760	*0.0701*	–0.0957	0.2928	–0.1171***	*0.0327*
Body mass index	–0.1640***	*0.0401*	–0.1006***	*0.0242*	–0.0818	0.1015	–0.1462***	*0.0207*
Body mass index^2/100	0.3582***	*0.1019*	0.2081***	*0.0593*	0.2117	0.2973	0.3215***	*0.0500*
Satisfactory sanitation	–0.1792**	*0.0748*	–0.1065**	*0.0443*	–0.1445	0.1076	–0.1347***	*0.0276*
House not built of solid materials	0.1394***	*0.0535*	0.0399	*0.0365*	0.1900	0.1807	0.0786***	*0.0221*
Head of household	0.1187**	*0.0485*	0.0662***	*0.0239*	0.0850	0.3022	0.1147***	*0.0233*
Household size	–0.0401***	*0.0119*	–0.0352***	*0.0080*	0.0017	0.0227	–0.0525***	*0.0048*
Dispersion parameter (alpha)	2.6387	0.1547	n.a.		2.5372		n.a.	
LR test of equidispersion	10,685	$p = 0.0000$	n.a.		41,656	$p = 0.0000$	n.a.	
Sample size	27,365		27,368		7,441		27,176	
Log-likelihood	–25,661.4		–15,287.4		–10,176.5		–28,132.8	

Note: All models also include a 3rd-degree polynomial in age and gender dummy. All models except FE Poisson include region dummies. Rob. SE = robust to hetero. and clustering standard error; Log-L = log likelihood. LR test of equidispersion is NegBin against Poisson ($p = p$-value). ***, **, and * significant at 1%, 5%, and 10%, respectively.

Source: Authors.

Computation

The Stata programmed routine for the Poisson model is `poisson`. For NegBin models it is

```
nbreg depvar varlist, dispersion(constant) cluster(commune)
```

where `dispersion(constant)` is optional and requests NegBin I; the default is NegBin II . Here the `cluster` option is used to correct standard errors for within-commune correlation. Note that an LR test against Poisson is generated only if the options `robust` or `cluster` are not specified. For the second part of a 2PM, the truncated Poisson or negative binomial can be computed by using the commands

`trnpois0` and `trnbin0`, respectively, which can be downloaded from the Stata Web site. In the latter case,

```
trnbin0 depvar varlist if depvar>0, cluster(commune)
```

Random- and fixed-effects Poisson models are obtained from the following:

```
xtpois depvar varlist, fe i(commune)
```

where `i(commune)` specifies common effects for all observations with the same values of the variable `commune`, and the option `fe` requests the fixed-effects model. The default is random effects.

Further reading

For a comprehensive review of econometric analyses of health and health care data, see Jones (2000). A more concise review, along with applications, can be found in Jones and O'Donnell (2002). For a practical guide to health econometrics containing many worked examples and Stata code, see Jones (2007). More generally, Wooldridge (2002) and Cameron and Trivedi (2005) are both excellent microeconometrics textbooks.

References

Amemiya, T. 1981. "Qualitative Response Models: A Survey." *Journal of Economic Literature* 19(4): 481–536.

Cameron, A. C., and P. K. Trivedi. 1986. "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests." *Journal of Applied Econometrics* 1: 29–53.

Cameron, A., and R. Trivedi. 2005. *Microeconometrics: Methods and Applications.* New York: Cambridge University Press.

Deb, P., and P. K. Trivedi. 1997. "Demand for Medical Care by the Elderly: A Finite Mixture Approach." *Journal of Applied Econometrics* 12(3): 313–36.

Duan, N., W. G. Manning, C. N. Morris, and J. P. Newhouse. 1983. "A Comparison of Alternative Models of the Demand for Health Care." *Journal of Business and Economic Statistics* 1: 115–26.

Greene, W. 2000. *Econometric Analysis.* Upper Saddle River, NJ: Prentice-Hall Inc.

Horrace, W. C., and R. L. Oaxaca. 2006. "Results on the Bias and Inconsistency of Ordinary Least Squares for the Linear Probability Model." *Economics Letters* 90: 321–27.

Jones, A. M. 2000. Health Econometrics. In *Handbook of Health Economics,* ed. A. J. Culyer and J. P. Newhouse, 265–346. Amsterdam, Netherlands: Elsevier North-Holland.

——— 2007. *Applied Econometrics for Health Economists—A Practical Guide.* Oxford, England: Radcliffe Medical Publishing.

Jones, A. M., and O. O'Donnell. 2002. *Econometric Analysis of Health Data.* Chichester, UK: John Wiley and Sons.

Mullahy, J. 1998. "Much Ado About Two: Reconsidering Retransformation and the Two-Part Model in Health Econometrics." *Journal of Health Economics* 17(3): 247–81.

Pohlmeier, W., and V. Ulrich. 1995. "An Econometric Model of the Two-Part Decision-Making Process in the Demand for Health Care." *Journal of Human Resources* 30(2): 339–61.

Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.