18-42



## 18.3.4 ~~23.10.3~~ PANEL DATA APPLICATIONS

The ordered probit model is used to model discrete scales that represent indicators of a continuous underlying variable such as strength of preference, performance, or level of attainment. Many of the recently assembled national panel data sets contain survey questions that ask about subjective assessments of health, satisfaction, or well-being, all of which are applications of this interpretation. Examples include: the following

- The European Community Household Panel (ECHP) includes questions about job satisfaction [see D'Addio (2004)].
- The British Household Panel Survey (BHPS) includes questions about health status [see Contoyannis et al. (2004)].
- The German Socioeconomic Household Panel (GSOEP) includes questions about subjective well being [see Winkelmann (2004)] and subjective assessment of health satisfaction [see Riphahn et al. (2003) and Example ~~23.19~~ 18.4.]

Ostensibly, the applications would fit well into the ordered probit frameworks already described. However, given the panel nature of the data, it will be desirable to augment the model with some accommodation of the individual heterogeneity that is likely to be present. The two standard models, fixed and random effects, have both been applied to the analyses of these survey data.

### 18.3.4.a ~~23.10.3.a.~~ Ordered Probit Models with Fixed Effects

D'Addio et al. (2003), using methodology developed by Frijters et al. (2004) and Ferrer-i-Carbonel et al. (2004), analyzed survey data on job satisfaction using the Danish component of the European Community Household Panel. Their estimator for an ordered logit model is built around the logic of Chamberlain's estimator for the binary logit model. [See Section 23.5.2.] 17.4.4 Because the approach is robust to individual specific threshold parameters and allows time-invariant variables, it differs sharply from the fixed effects models we have considered thus far as well as from the ordered probit model of Section 23.10.1.[54] 10 FN 10 Unlike Chamberlain's estimator for the binary logit model, however, their conditional estimator is not a function of minimal sufficient statistics. As such, the incidental parameters problem remains an issue.

Das and van Soest (2000) proposed a somewhat simpler approach. [See, as well, Long's (1997) discussion of the "parallel regressions assumption," which employs this device in a cross-section framework]. Consider the base case ordered logit model with fixed effects,

$$y_{it}^* = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it},\ \varepsilon_{it}\,|\,\mathbf{X}_i \sim N[0, 1],$$

$$y_{it} = j \quad \text{if} \quad \mu_{j-1} < y_{it}^* < \mu_j,\ j = 0, 1, \ldots, J \quad \text{and} \quad \mu_{-1} = -\infty, \mu_0 = 0, \mu_J = +\infty.$$

The model assumptions imply that

$$\text{Prob}(y_{it} = j\,|\,\mathbf{X}_i) = \Lambda(\mu_j - \alpha_i - \mathbf{x}_{it}'\boldsymbol{\beta}) - \Lambda(\mu_{j-1} - \alpha_i - \mathbf{x}_{it}'\boldsymbol{\beta}),$$

where $\Lambda(t)$ is the cdf of the logistic distribution. Now, define a binary variable

$$w_{it,j} = 1 \text{ if } y_{it} > j, \quad j = 0, \ldots, J - 1.$$

10 ~~54~~ Cross-section versions of the ordered probit model with individual specific thresholds appear in Terza (1985a), Pudney and Shields (2000), and Greene (2007).

18-43



It follows that

$$\begin{aligned}\text{Prob}[w_{it,j} = 1 \mid \mathbf{X}_i] &= \Lambda(\alpha_i - \mu_j + \mathbf{x}_{it}'\boldsymbol{\beta})\\ &= \Lambda(\theta_i + \mathbf{x}_{it}'\boldsymbol{\beta}).\end{aligned}$$

The "$j$" specific constant, which is the same for all individuals, is absorbed in $\theta_i$. Thus, a fixed effects binary logit model applies to each of the $J-1$ binary random variables, $w_{it,j}$. The method in Section 23.5.2 can now be applied to each of the $J-1$ random samples. This provides $J-1$ estimators of the parameter vector $\boldsymbol{\beta}$ (but no estimator of the threshold parameters). The authors propose to reconcile these different estimators by using a minimum distance estimator of the common true $\boldsymbol{\beta}$. (See Section ~~15.3~~ 13.3.) The minimum distance estimator at the second step is chosen to minimize

17.4.4

$$q = \sum_{j=0}^{J-1}\sum_{m=0}^{J-1}(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta})'[\mathbf{V}_{jm}^{-1}](\hat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}),$$

where $[\mathbf{V}_{jm}^{-1}]$ is the $j, m$ block of the inverse of the $(J-1)K \times (J-1)K$ partitioned matrix $\mathbf{V}$ that contains Asy. Cov$[\hat{\boldsymbol{\beta}}_j, \hat{\boldsymbol{\beta}}_m]$. The appropriate form of this matrix for a set of cross-section estimators is given in Brant (1990). Das and van Soest (2000) used the counterpart for Chamberlain's fixed effects estimator but do not provide the specifics for computing the off-diagonal blocks in $\mathbf{V}$.

The full ordered probit model with fixed effects, including the individual specific constants, can be estimated by unconditional maximum likelihood using the results in Section 16.9.6.c. The likelihood function is concave [see Pratt (1981)], so despite its superficial complexity, the estimation is straightforward. (In the following application, with more than 27,000 observations and 7,293 individual effects, estimation of the full model required roughly five seconds of computation.) No theoretical counterpart to the Hsiao (1986, 2003) and Abrevaya (1997) results on the small $T$ bias (incidental parameters problem) of the MLE in the presence of fixed effects has been derived for the ordered probit model. The Monte Carlo results in Greene (2004) (see, as well, Chapter ~~17~~ 15), suggest that biases comparable to those in the binary choice models persist in the ordered probit model as well. As in the binary choice case, the complication of the fixed effects model is the small sample bias, not the computation. The Das and van Soest approach finesses this problem—their estimator is consistent—but at the cost of losing the information needed to compute partial effects or predicted probabilities.

Au: x-ref to Section 16.9.6.c OK?

### ~~23.10.3.b~~ 18.3.4.b Ordered Probit Models with Random Effects

The random effects ordered probit model model has been much more widely used than the fixed effects model. Applications include Groot and van den Brink (2003), who studied training levels of employees, with firm effects; Winkelmann (2003b), who examined subjective measures of well being with individual and family effects; Contoyannis et al. (2004), who analyzed self-reported measures of health status; and numerous others. In the simplest case, the method of the Butler and Moffitt (1982) quadrature method (Section 16.9.6.b) can be extended to this model.

Au: x-ref to Section 16.9.6.c OK?

***Example ~~23.19~~ 18-16 Health Satisfaction***

The GSOEP German Health Care data that we have used in Examples ~~11.11~~, 16.16, and others includes a self-reported measure of health satisfaction, *HSAT*, that takes values

17.4

18-44



FN 11

0, 1, . . . , 10.~~[52]~~ [11] This is a typical application of a scale variable that reflects an underlying continuous variable, "health." The frequencies and sample proportions for the reported values are as follows:

| *HSAT* | *Frequency* | *Proportion* |
|---|---|---|
| 0 | 447 | 1.6% |
| 1 | 255 | 0.9% |
| 2 | 642 | 2.3% |
| 3 | 1173 | 4.2% |
| 4 | 1390 | 5.0% |
| 5 | 4233 | 15.4% |
| 6 | 2530 | 9.2% |
| 7 | 4231 | 15.4% |
| 8 | 6172 | 22.5% |
| 9 | 3061 | 11.2% |
| 10 | 3192 | 11.6% |

We have fit pooled and panel data versions of the ordered probit model to these data. The model used is

$\beta_4 Kids_{it}$

$$y^*_{it} = \beta_1 + \beta_2\,Age_{it} + \beta_3\,Income_{it} + \beta_4\,Education_{it} + \beta_5\,Married_{it} + \beta_6\,Working_{it} + \varepsilon_{it} + c_i,$$

where $c_i$ will be the common fixed or random effect. (We are interested in comparing the fixed and random effects estimators, so we have not included any time-invariant variables such as gender in the equation.) Table 23.21 lists five estimated models. (Standard errors for the estimated threshold parameters are omitted.) The first is the pooled ordered probit model. The second and third are fixed effects. Column 2 shows the unconditional fixed effects estimates using the results of Section 16.9.6.c. Column 3 shows the Das and van Soest estimator. For the minimum distance estimator, we used an inefficient weighting matrix, the block-diagonal matrix in which the $j$th block is the inverse of the $j$th asymptotic covariance matrix for the individual logit estimators. With this weighting matrix, the estimator is

TB 18.12 18.12

14.9.6.d

$$\hat{\beta}_{MDE} = \left[\sum_{j=0}^{9} \mathbf{V}_j^{-1}\right]^{-1} \sum_{j=0}^{9} \mathbf{V}_j^{-1}\hat{\beta}_j,$$

and the estimator of the asymptotic covariance matrix is approximately equal to the bracketed inverse matrix. The fourth set of results is the random effects estimator computed using the maximum simulated likelihood method. This model can be estimated using Butler and Moffitt's quadrature method; however, we found that even with a large number of nodes, the quadrature estimator converged to a point where the log-likelihood was far lower than the MSL estimator, and at parameter values that were implausibly different from the other estimates. Using different starting values and different numbers of quadrature points did not change this outcome. The MSL estimator for a random constant term (see Section 17.5) is considerably slower, but produces more reasonable results. The fifth set of results is the Mundlak form of the random effects model, which includes the group means in the models as controls to accommodate possible correlation between the latent heterogeneity and the included variables. As noted in Example 23.17, the components of the ordered choice model must be interpreted with some care. By construction, the partial effects of the variables on the probabilities of the outcomes must change sign, so the simple coefficients do not show the complete picture implied by the estimated model. Table ~~23.22~~ 18.13 shows the partial effects for the pooled model to illustrate the computations.

Au: x-ref to Section 17.5 OK?

18.2

TB 18.13

~~[52]~~ 11 In the original data set, 40 (of 27,326) observations on this variable were coded with noninteger values between 6 and 7. For purposes of our example, we have recoded all 40 observations to 7.

18-45



TABLE ~~23.21~~ 18.12 Estimated Ordered Probit Models for Health Satisfaction

| Variable | (1) Pooled | (2) Fixed Effects Unconditional | (3) Fixed Effects Conditional | (4) Random Effects | (5) Random Effects Mundlak Controls: Variables | (5) Random Effects Mundlak Controls: Means |
|---|---|---|---|---|---|---|
| Constant | 2.4739 | | | 3.8577 | 3.2603 | |
| | (0.04669) | | | (0.05072) | (0.05323) | |
| Age | −0.01913 | −0.07162 | −0.1011 | −0.03319 | −0.06282 | 0.03940 |
| | (0.00064) | (0.002743) | (0.002878) | (0.00065) | (0.00234) | (0.002442) |
| Income | 0.1811 | 0.2992 | 0.4353 | 0.09436 | 0.2618 | 0.1461 |
| | (0.03774) | (0.07058) | (0.07462) | (0.03632) | (0.06156) | (0.07695) |
| Kids | 0.06081 | −0.06385 | −0.1170 | 0.01410 | −0.05458 | 0.1854 |
| | (0.01459) | (0.02837) | (0.03041) | (0.01421) | (0.02566) | (0.03129) |
| Education | 0.03421 | 0.02590 | 0.06013 | 0.04728 | 0.02296 | 0.02257 |
| | (0.002828) | (0.02677) | (0.02819) | (0.002863) | (0.02793) | (0.02807) |
| Married | 0.02574 | 0.05157 | 0.08505 | 0.07327 | 0.04605 | −0.04829 |
| | (0.01623) | (0.04030) | (0.04181) | (0.01575) | (0.03506) | (0.03963) |
| Working | 0.1292 | −0.02659 | −0.007969 | 0.07108 | −0.02383 | 0.2702 |
| | (0.01403) | (0.02758) | (0.02830) | (0.01338) | (0.02311) | (0.02856) |
| $\mu_1$ | 0.1949 | 0.3249 | | 0.2726 | 0.2752 | |
| $\mu_2$ | 0.5029 | 0.8449 | | 0.7060 | 0.7119 | |
| $\mu_3$ | 0.8411 | 1.3940 | | 1.1778 | 1.1867 | |
| $\mu_4$ | 1.111 | 1.8230 | | 1.5512 | 1.5623 | |
| $\mu_5$ | 1.6700 | 2.6992 | | 2.3244 | 2.3379 | |
| $\mu_6$ | 1.9350 | 3.1272 | | 2.6957 | 2.7097 | |
| $\mu_7$ | 2.3468 | 3.7923 | | 3.2757 | 3.2911 | |
| $\mu_8$ | 3.0023 | 4.8436 | | 4.1967 | 4.2168 | |
| $\mu_9$ | 3.4615 | 5.5727 | | 4.8308 | 4.8569 | |
| $\sigma_u$ | 0.0000 | 0.0000 | | 1.0078 | 0.9936 | |
| ln L | −56813.52 | −41875.63 | | −53215.54 | −53070.43 | |

KT Winkelmann (2003b) used the random effects approach to analyze the subjective well being (SWB) question (also coded 0 to 10) in the German Socioeconomic Panel (GSOEP) data set. The ordered probit model in this study is based on the latent regression KT

$$y^*_{imt} = \mathbf{x}'_{imt}\beta + \varepsilon_{imt} + u_{im} + v_i.$$

TABLE ~~23.22~~ 18.13 Estimated Marginal Effects: Pooled Model

| HSAT | Age | Income | Kids | Education | Married | Working |
|---|---|---|---|---|---|---|
| 0 | 0.0006 | −0.0061 | −0.0020 | −0.0012 | −0.0009 | −0.0046 |
| 1 | 0.0003 | −0.0031 | −0.0010 | −0.0006 | −0.0004 | −0.0023 |
| 2 | 0.0008 | −0.0072 | −0.0024 | −0.0014 | −0.0010 | −0.0053 |
| 3 | 0.0012 | −0.0113 | −0.0038 | −0.0021 | −0.0016 | −0.0083 |
| 4 | 0.0012 | −0.0111 | −0.0037 | −0.0021 | −0.0016 | −0.0080 |
| 5 | 0.0024 | −0.0231 | −0.0078 | −0.0044 | −0.0033 | −0.0163 |
| 6 | 0.0008 | −0.0073 | −0.0025 | −0.0014 | −0.0010 | −0.0050 |
| 7 | 0.0003 | −0.0024 | −0.0009 | −0.0005 | −0.0003 | −0.0012 |
| 8 | −0.0019 | 0.0184 | 0.0061 | 0.0035 | 0.0026 | 0.0136 |
| 9 | −0.0021 | 0.0198 | 0.0066 | 0.0037 | 0.0028 | 0.0141 |
| 10 | −0.0035 | 0.0336 | 0.0114 | 0.0063 | 0.0047 | 0.0233 |

18-46



The independent variables include age, gender, employment status, income, family size, and an indicator for good health. An unusual feature of the model is the nested random
14.9.6.b effects (see Section 9.7.1), which include a family effect, $v_i$, as well as the individual family member ($i$ in family $m$) effect, $u_{im}$. The GLS/MLE approach we applied to
14.9.6.b the linear regression model in Section 9.7.1 is unavailable in this nonlinear setting. Winkelmann instead employed a Hermite quadrature procedure to maximize the log-likelihood function.

Contoyannis, Jones, and Rice (2004) analyzed a self-assessed health scale that ranged from 1 (very poor) to 5 (excellent) in the British Household Panel Survey. Their model accommodated a variety of complications in survey data. The latent regression underlying their ordered probit model is

$$h_{it}^* = \mathbf{x}_{it}'\beta + \mathbf{H}_{i,t-1}'\gamma + \alpha_i + \varepsilon_{it},$$

where $\mathbf{x}_{it}$ includes marital status, race, education, household size, age, income, and number of children in the household. The lagged value, $\mathbf{H}_{i,t-1}$, is a set of binary variables for the observed health status in the previous period. (This is the same device that
18.3 was used by Butler et al. in Example 23.18.) In this case, the lagged values capture state dependence—the assumption that the health outcome is redrawn randomly in each period is inconsistent with evident runs in the data. The initial formulation of the regression is a fixed effects model. To control for the possible correlation between the effects, $\alpha_i$, and the regressors, and the initial conditions problem that helps to explain the state dependence, they use a hybrid of Mundlak's (1978) correction and a suggestion by Wooldridge (2002a) for modeling the initial conditions,

$$\alpha_i = \alpha_0 + \bar{\mathbf{x}}'\alpha_1 + \mathbf{H}_{i,1}'\delta + u_i,$$

where $u_i$ is exogenous. Inserting the second equation into the first produces a random effects model that can be fit using the quadrature method we considered earlier.

## 23.11 MODELS FOR UNORDERED MULTIPLE CHOICES

Some studies of multiple-choice settings include the following:

1. Hensher (1986, 1991), McFadden (1974), and many others have analyzed the travel mode of urban commuters.
2. Schmidt and Strauss (1975a,b) and Boskin (1974) have analyzed occupational choice among multiple alternatives.
3. Terza (1985a) has studied the assignment of bond ratings to corporate bonds as a choice among multiple alternatives.
4. Rossi and Allenby (1999, 2003) studied consumer brand choices in a repeated choice (panel data) model.
5. Train (2003) studied the choice of electricity supplier by a sample of California electricity customers.
6. Hensher, Rose, and Greene (2006) analyzed choices of automobile models by a sample of consumers offered a hypothetical menu of features.

### 18.3.5 Extensions of the Ordered Probit Model

The basic specification of the ordered probit model can be extended in the same directions as we considered in constructing models for binary choice in Chapter 17. These include heteroscedasticity in the random utility function (see Section 17.3.7.b, Keele and Park (2005) and Wang and Kockelman (2005) for an application) and heterogeneity in the preferences (i.e., random parameters and latent classes). [An extensive study of heterogeneity in health satisfaction based on 22 waves of the GSOEP is Jones and Schurer (2010).] Two specification issues that are specific to the ordered choice model are accommodating heterogeneity in the threshold parameters and reconciling differences in the meaning of the preference scale across different groups. We will sketch the model extensions in this section. Further details are given in Chapters 6 and 7 of Hensher and Greene (2010).

#### 18.3.5.a Threshold Models – Generalized Ordered Choice Models

The model analyzed thus far assumes that the thresholds $\mu_j$ are the same for every individual in the sample. Terza (1985a), Pudney and Shields (2000), King, Murray, Salomon and Tandon (KMST, 2004), Boes and Winkelmann (2006a), Greene, Harris, Hollingsworth and Maitra (2008), and Greene and Hensher (2009), all present applications that include individual variation in the thresholds of the ordered choice model.

In his analysis of bond ratings, Terza (1985) suggested the generalization,

$$\mu_{ij} = \mu_j + \mathbf{x}_i'\boldsymbol{\delta}.$$

With three outcomes, the probabilities are

$$
\begin{aligned}
& y_i^* = \alpha + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \\
\text{and}\quad & y_i = 0 \text{ if } y_i^* \le 0, \\
& \quad\;\; 1 \text{ if } 0 < y_i^* \le \mu + \mathbf{x}_i'\boldsymbol{\delta}, \\
& \quad\;\; 2 \text{ if } y_i^* > \mu + \mathbf{x}_i'\boldsymbol{\delta}.
\end{aligned}
$$

For three outcomes, the model has two thresholds, $\mu_0 = 0$ and $\mu_1 = \mu + x_i'\delta$. The three probabilities can be written

$$
\begin{aligned}
P_0 &= \text{Prob}(y_i = 0|\mathbf{x}_i) = \Phi[-(\alpha + \mathbf{x}_i'\boldsymbol{\beta})] \\
P_1 &= \text{Prob}(y_i = 1|\mathbf{x}_i) = \Phi[(\mu + \mathbf{x}_i'\boldsymbol{\delta}) - (\alpha + \mathbf{x}_i'\boldsymbol{\beta})] - \Phi[-(\alpha + \mathbf{x}_i'\boldsymbol{\beta})] \\
P_2 &= \text{Prob}(y_i = 2|\mathbf{x}_i) = 1 - \Phi[(\mu + \mathbf{x}_i'\boldsymbol{\delta}) - (\alpha + \mathbf{x}_i'\boldsymbol{\beta})].
\end{aligned}
$$

AU: OK to spell out "e.g."?

for example

For applications of this approach, see, e.g., Kerkhofs and Lindeboom (1995), Groot and van den Brink (2003) and Lindeboom and van Doorslayer (2003). Note that if $\boldsymbol{\delta}$ is unrestricted, then Prob($y_i$ =1|$\mathbf{x}_i$) can be negative. This is a shortcoming of the model when specified in this form. Subsequent development of the generalized model involves specifications that avoid this internal inconsistency. Note, as well, that if the model is recast in terms of $\mu$ and $\boldsymbol{\gamma} = [\alpha,(\boldsymbol{\beta} - \boldsymbol{\delta})]$, then the model is not distinguished from the original ordered probit model with a constant threshold parameter. This identification issue emerges prominently in Pudney and Shield's (2000) continued development of this model.

Pudney and Shields's (2000) "Generalized Ordered Probit Model," was also formulated to accommodate *observable* individual heterogeneity in the threshold parameters. Their application was in the context of job promotion for UK nurses in which the steps on the promotion ladder are individual specific. In their setting, in contrast to Terza's, some of the

18 -48

variables in the threshold equations are explicitly different from those in the regression. The authors constructed a generalized model and a test of "threshold constancy" by defining $\mathbf{q}_i$ to include a constant term and those variables that are unique to the threshold model Variables that are common to both the thresholds and the regression are placed in $\mathbf{x}_i$, and the model is reparameterized as

minus minus

$$\Pr(y_i = g|\mathbf{x}_i,\mathbf{q}_i) = \Phi[\mathbf{q}_i'\boldsymbol{\delta}_g - \mathbf{x}_i'(\boldsymbol{\beta} - \boldsymbol{\delta}_g)] - \Phi[\mathbf{q}_i'\boldsymbol{\delta}_{g-1} - \mathbf{x}_i'(\boldsymbol{\beta} - \boldsymbol{\delta}_{g-1})].$$

An important point noted by the authors is that the same model results if these common variables are placed in the thresholds instead. This is a minor algebraic result, but it exposes an ambiguity in the interpretation of the model – whether a particular variable affects the regression or the thresholds is one of the issues that was developed in the original model specification.

As will be evident in the application in the next section, the specification of the threshold parameters is a crucial feature of the ordered choice model. KMST (2004), Greene (2007a), Eluru, Bhat and Hensher (2008) and Greene and Hensher (2009) employ a "hierarchical ordered probit," or HOPIT model,

$$
\begin{aligned}
& y_i^* = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i, \\
& y_i = j \text{ if } \mu_{i,j-1} \le y_i^* < \mu_{ij}, \\
& \mu_0 = 0, \\
& \mu_{i,j} = \exp(\lambda_j + \boldsymbol{\gamma}'\mathbf{z}_i) \quad \text{(case 1)}, \\
\text{or}\quad & \mu_{i,j} = \exp(\lambda_j + \boldsymbol{\gamma}_j'\mathbf{z}_i) \quad \text{(case 2)}.
\end{aligned}
$$

Case 2 is the Terza (1985) and Pudney and Shields (2000) model with an exponential rather than linear function for the thresholds. This formulation addresses two problems: (i) the thresholds are mathematically distinct from the regression; (ii) by this construction, the threshold parameters must be positive. With a slight modification, the ordering of the thresholds can also be imposed. In case 1, 1 2

$$\mu_{i,j} = [\exp(\lambda_1) + \exp(\lambda_2) + \ldots + \exp(\lambda_j)] \times \exp(\boldsymbol{\gamma}'\mathbf{z}_i),$$

and in case 2,

minus

$$\mu_{i,j} = \mu_{i,j-1} + \exp(\lambda_j + \boldsymbol{\gamma}_j'\mathbf{z}_i).$$

In practical terms, the model can now be fit with the constraint that all predicted probabilities are greater than zero. This is a numerical solution to the problem of ordering the thresholds for all data vectors.

KT

KT This extension of the ordered choice model shows a case of **identification through functional form**. As we saw in the previous two models, the parameters $(\lambda_j, \gamma_j, \beta)$ would not be separately identified if all the functions were linear. The contemporary literature views models that are unidentified without a change in functional form with some skepticism. However, the underlying theory of this model does not insist on linearity of the thresholds (or the utility function, for that matter), but it *does* insist on the ordering of the thresholds, and one might equally criticize the original model for being unidentified *because the model builder insists on a linear form*. That is, there is no obvious reason that the threshold parameters must be linear functions of the variables, or that linearity enjoys some claim to first precedence in the utility function. This is a methodological issue that cannot be resolved here. The nonlinearity of the preceding specification, or others that resemble it, does provide the benefit of a simple way to achieve other fundamental results, e.g., coherency of the model (all positive probabilities).

for example

AU: OK to spell out "e.g."?

18-49

## 18.3.4.b Thresholds and Heterogeneity – Anchoring Vignettes

The introduction of observed heterogeneity into the threshold parameters attempts to deal with a fundamentally restrictive assumption of the ordered choice model. Survey respondents rarely view the survey questions exactly the same way. This is certainly true in surveys of health satisfaction or subjective well being. [See Boes and Winkelmann (2006b) and Ferrer-i-Carbonell and Frijters (2004).] KMST (2004) identify two very basic features of survey data that will make this problematic; first, they often measure concepts that are definable only with reference to examples, such as freedom, health, satisfaction, etc. Second, individuals do, in fact, often understand survey questions very differently, particularly with respect to answers at the extremes. A widely used term for this interpersonal incomparability is **differential item functioning** (DIF). Kapteyn, Smith and Van Soest (KSV, 2007) and Van Soest, Delaney, Harmon, Kapteyn and Smith (2007) suggest the results in Figure 18.5 to describe the implications of DIF. The figure shows the distribution of Health (or drinking behavior in the latter study) in two hypothetical countries. The density for country A (the upper figure) is to the left of that for country B implying that on average, people in country A are less healthy than those in country B. But, the people in the two countries culturally offer very different response scales if asked to report their health on a five point scale, as shown. In the figure, those in country A have a much more positive view of a given, objective health status than those in country B. A person in country A with health status indicated by the dotted line would report that they are in "Very Good" health while a person in country B with the same health status would report only "Fair." A simple frequency of the distribution of self-assessments of health status in the two countries would suggest that people in country A are much healthier than those in country B when, in fact, the opposite is true. Correcting for the influences of DIF in such a situation would be essential to obtaining a meaningful comparison of the two countries. The impact of DIF is an accepted feature of the model within a population, but could be strongly distortionary when comparing very disparate groups, such as across countries, as in KMST (political groups), Murray, Tandon, Mathers and Sudana (2002) (health outcomes), Tandon et al. (2004) and KSV (work disability), Sirven, Santos-Egglmann and Spagnoli (2008) and Gupta, Kristensens and Possoli (2008) (health), Angelini et al. (2008) (life satisfaction), Kristensen and Johansson (2008) and Bago d'Uva et al. (2008), all of whom used the ordered probit model to make cross group comparisons..

KMST proposed the use of *anchoring vignettes* to resolve this difference in perceptions across groups. The essential approach is to use a series of examples that, it is believed, all respondents will agree on to estimate each respondent's DIF and correct for it. The idea of using vignettes to anchor perceptions in survey questions is not itself new; KMST cite a number of earlier uses. The innovation is their method for incorporating the approach in a formal model for the ordered choices. The bivariate and multivariate probit models that they develp combine the elements described in Sectons 18.3.1 – 18.3.3 and the HOPIT model in Section 18.3.4.a.

Au: OK to spell out "etc."?

and so on

FIG 18.5

18-50

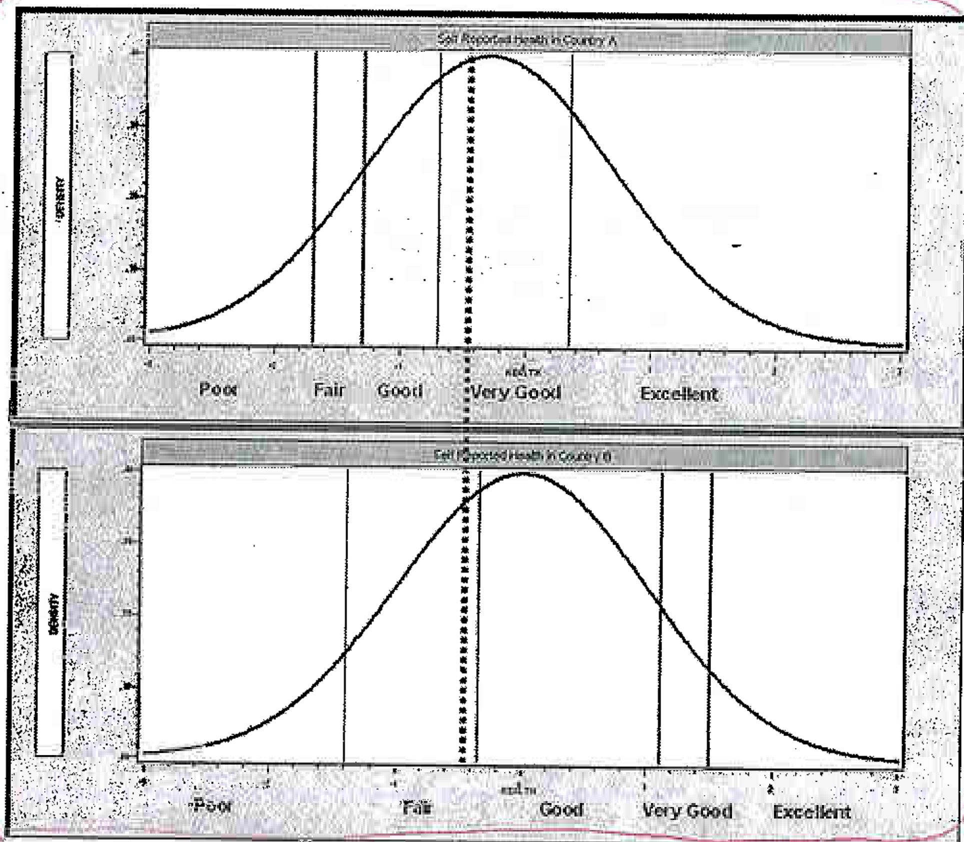


**Figure 18.5 Differential Item Functioning in Ordered Choices**

## 18.4 MODELS FOR COUNTS OF EVENTS

We have encountered behavioral variables that involve counts of events at several points in this text. In Examples 14.10 and 17.20, we examined the number of times an individual visited the physician using the GSOEP data. The credit default data that we used in Examples 7.10 and 17.22 also includes another behavioral variable, the number of derogatory reports in an individual's credit history. Finally, in Example 17.23, we analyzed data on firm innovation. Innovation is often analyzed [for example, by Hausman, Hall, and Griliches (1984) and many others] in terms of the number of patents that the firm obtains (or applies for). In each of these cases, the variable of interest is a count of events. This obviously differs from the discrete dependent variables we analyzed in the previous two sections. A count is a quantitative measure that is, at least in principle, amenable to analysis using multiple linear regression. However, the typical preponderance of zeros and small values, and the discrete nature of the outcome variable suggest that the regression approach can be improved by a method that explicitly accounts for these aspects.

Like the basic multinomial logit model for unordered data in Section 18.2 and the simple probit and logit models for binary and ordered data in Sections 17.2 and 18.3, the Poisson regression model is the fundamental starting point for the analysis of count data. We will develop the elements of modeling for count data in this framework in Sections 18.4.1 – 18.4.3, then turn to more elaborate, flexible specifications in subsequent sections. Sections 18.4.4 and 18.4.5 will present the negative binomial and other alternatives to the Poisson functional form. Section 18.4.6 will describe the implications for the model specification of some complicating features of observed data, truncation and censoring. Truncation arises when certain values, such as zero, are absent from the observed data because of the sampling mechanism, not as a function of the data geneating process. Data on recreation site visitation that are gathered at the site, for example, will, by construction, not contain any zeros. Censoring arises when certain ranges of outcomes are all coded with the same value. In the example analyzed below, the response variable is censored at 12, though values larger than 12 are possible "in the field." As we have done in the several earlier treatments, in Section 18.4.7, we will examine extensions of the count data models that are made possible when the analysis is based on panel data. Finally, Section 18.4.8 discusses some behavioral models that involve more than one equation. For an example, based on the large number of zeros in the observed data, it appears that our count of doctor visits might be generated by a two part process, a first step in which the individual decides whether or not to visit the physician at all, and a second decision, given the first, how many times to do so. The "hurdle model" that applies here, and some related variants are discussed in Section 18.4.8.

and

next

18-52

# 18.4.1 The Poisson Regression Model



This chapter will describe modeling approaches for events.[1] As suggested, the two measures, counts of events and duration between events, are usually studied with different techniques, and for different purposes. Sections ~~25.2 to 25.5 will describe some~~ 19 19 of the econometric approaches to modeling event counts. The application of duration models to economic data is somewhat less frequent, but conversely, far more frequent in the other fields mentioned. The chapter, and this text, end in Section 25.6 with a discussion of models for duration. 19

## 19 25.2 MODELS FOR COUNTS OF EVENTS

Data on patents suggested in Section 23.2 are typical of **count data.** In principle, we could analyze these data using multiple linear regression. But the preponderance of zeros and the small values and clearly discrete nature of the dependent variable suggest that we can improve on least squares and the linear model with a specification that accounts for these characteristics. The **Poisson regression model** has been widely used to study such data.

The Poisson regression model specifies that each $y_i$ is drawn from a Poisson distribution with parameter $\lambda_i$, which is related to the regressors $\mathbf{x}_i$. The primary equation of the model is

$$\text{Prob}(Y = y_i \mid \mathbf{x}_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \quad y_i = 0, 1, 2, \ldots. \tag{25-1}$$

19 (18-17)

The most common formulation for $\lambda_i$ is the **loglinear model,**

$$\ln \lambda_i = \mathbf{x}_i'\boldsymbol{\beta}.$$

It is easily shown that the expected number of events *per period* is given by

$$E[y_i \mid \mathbf{x}_i] = \text{Var}[y_i \mid \mathbf{x}_i] = \lambda_i = e^{\mathbf{x}_i'\boldsymbol{\beta}},$$

so

$$\frac{\partial E[y_i \mid \mathbf{x}_i]}{\partial \mathbf{x}_i} = \lambda_i \boldsymbol{\beta}.$$

With the parameter estimates in hand, this vector can be computed using any data vector desired.

In principle, the Poisson model is simply a nonlinear regression. But it is far easier to estimate the parameters with maximum likelihood techniques. The log-likelihood function is

$$\ln L = \sum_{i=1}^{n} [-\lambda_i + y_i \mathbf{x}_i'\boldsymbol{\beta} - \ln y_i!].$$

[1] Some particularly rich surveys of these topics (among dozens available) are Cameron and Trivedi (1986, ~~1998, 2005), Winkelmann (2003), and Kalbfleisch and Prentice (2002).~~

[2] ~~We have estimated a Poisson regression model using two-step nonlinear least squares in Example 16.6.~~

18-53



The likelihood equations are

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} (y_i - \lambda_i)\mathbf{x}_i = \mathbf{0}.$$

The Hessian is

$$\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = -\sum_{i=1}^{n} \lambda_i \mathbf{x}_i \mathbf{x}_i'.$$

The Hessian is negative definite for all $\mathbf{x}$ and $\boldsymbol{\beta}$. Newton's method is a simple algorithm for this model and will usually converge rapidly. At convergence, $[\sum_{i=1}^{n} \hat{\lambda}_i \mathbf{x}_i \mathbf{x}_i']^{-1}$ provides an estimator of the asymptotic covariance matrix for the parameter estimates. Given the estimates, the prediction for observation $i$ is $\hat{\lambda}_i = \exp(\mathbf{x}_i'\hat{\boldsymbol{\beta}})$. A standard error for the prediction interval can be formed by using a linear Taylor series approximation. The estimated variance of the prediction will be $\hat{\lambda}_i^2 \mathbf{x}_i' \mathbf{V} \mathbf{x}_i$, where $\mathbf{V}$ is the estimated asymptotic covariance matrix for $\hat{\boldsymbol{\beta}}$. $\mathbf{x}_i$

For testing hypotheses, the three standard tests are very convenient in this model. The Wald statistic is computed as usual. As in any discrete choice model, the likelihood ratio test has the intuitive form

$$\text{LR} = 2\sum_{i=1}^{n} \ln\left(\frac{\hat{P}_i}{\hat{P}_{\text{restricted},i}}\right),$$

where the probabilities in the denominator are computed with using the restricted model. Using the BHHH estimator for the asymptotic covariance matrix, the LM statistic is simply

delete prime

$$\text{LM} = \left[\sum_{i=1}^{n} \cancel{\mathbf{x}_i'}(y_i - \hat{\lambda}_i)\right]' \left[\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' (y_i - \hat{\lambda}_i)^2\right]^{-1} \left[\sum_{i=1}^{n} \mathbf{x}_i (y_i - \hat{\lambda}_i)\right] = \mathbf{i}'\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{i}, \quad \text{(~~25-2~~ 18-18)}$$

where each row of $\mathbf{G}$ is simply the corresponding row of $\mathbf{X}$ multiplied by $e_i = (y_i - \hat{\lambda}_i)$, $\hat{\lambda}_i$ is computed using the restricted coefficient vector, and $\mathbf{i}$ is a column of ones.

### 18.4.2 ~~25.2.1~~ MEASURING GOODNESS OF FIT

The Poisson model produces no natural counterpart to the $R^2$ in a linear regression model, as usual, because the conditional mean function is nonlinear and, moreover, because the regression is heteroscedastic. But many alternatives have been suggested.[12] A measure based on the standardized residuals is

FN 12

$$R_p^2 = 1 - \frac{\sum_{i=1}^{n} \left[\frac{y_i - \hat{\lambda}_i}{\sqrt{\hat{\lambda}_i}}\right]^2}{\sum_{i=1}^{n} \left[\frac{y_i - \bar{y}}{\sqrt{\bar{y}}}\right]^2}.$$

12 ~~3~~ See the surveys by Cameron and Windmeijer (1993), Gurmu and Trivedi (1994), and Greene (1995b).

18-54



This measure has the virtue that it compares the fit of the model with that provided by a model with only a constant term. But it can be negative, and it can rise when a variable is dropped from the model. For an individual observation, the **deviance** is

$$d_i = 2[y_i \ln(y_i/\hat{\lambda}_i) - (y_i - \hat{\lambda}_i)] = 2[y_i \ln(y_i/\hat{\lambda}_i) - e_i],$$

where, by convention, $0\ln(0) = 0$. If the model contains a constant term, then $\sum_{i=1}^{n} e_i = 0$. The sum of the deviances,

$$G^2 = \sum_{i=1}^{n} d_i = 2\sum_{i=1}^{n} y_i \ln(y_i/\hat{\lambda}_i),$$

is reported as an alternative fit measure by some computer programs. This statistic will equal 0.0 for a model that produces a perfect fit. (Note that because $y_i$ is an integer while the prediction is continuous, it could not happen.) Cameron and Windmeijer (1993) suggest that the fit measure based on the deviances,

$$R_d^2 = 1 - \frac{\sum_{i=1}^{n}\left[y_i \log\left(\frac{y_i}{\hat{\lambda}_i}\right) - (y_i - \hat{\lambda}_i)\right]}{\sum_{i=1}^{n}\left[y_i \log\left(\frac{y_i}{\bar{y}}\right)\right]},$$

has a number of desirable properties. First, denote the log-likelihood function for the model in which $\psi_i$ is used as the prediction (e.g., the mean) of $y_i$ as $\ell(\psi_i, y_i)$. The Poisson model fit by MLE is, then, $\ell(\hat{\lambda}_i, y_i)$, the model with only a constant term is $\ell(\bar{y}, y_i)$, and a model that achieves a perfect fit (by predicting $y_i$ with itself) is $l(y_i, y_i)$. Then

$$R_d^2 = \frac{\ell(\hat{\lambda}, y_i) - \ell(\bar{y}, y_i)}{\ell(y_i, y_i) - \ell(\bar{y}, y_i)}.$$

FN 13

Both numerator and denominator measure the improvement of the model over one with only a constant term. The denominator measures the maximum improvement, since one cannot improve on a perfect fit. Hence, the measure is bounded by zero and one and increases as regressors are added to the model.[13] We note, finally, the passing resemblance of $R_d^2$ to the "pseudo-$R^2$," or "likelihood ratio index" reported by some statistical packages (e.g., Stata),

$$R_{\mathrm{LRI}}^2 = 1 - \frac{\ell(\hat{\lambda}_i, y_i)}{\ell(\bar{y}, y_i)}.$$

Many modifications of the Poisson model have been analyzed by economists. In this and the next few sections, we briefly examine a few of them.

## 18.4.3 ~~25.3.2~~ TESTING FOR OVERDISPERSION

The Poisson model has been criticized because of its implicit assumption that the variance of $y_i$ equals its mean. Many extensions of the Poisson model that relax this assumption have been proposed by Hausman, Hall, and Griliches (1984), McCullagh and Nelder (1983), and Cameron and Trivedi (1986), to name but a few.

13 ~~9~~ Note that multiplying both numerator and denominator by 2 produces the ratio of two likelihood ratio statistics, each of which is distributed as chi-squared.

18.55



The first step in this extended analysis is usually a test for overdispersion in the context of the simple model. A number of authors have devised tests for "overdispersion" within the context of the Poisson model. [See Cameron and Trivedi (1990), Gurmu (1991), and Lee (1986).] We will consider three of the common tests, one based on a regression approach, one a conditional moment test, and a third, a **Lagrange multiplier test,** based on an alternative model. KI KI

Cameron and Trivedi (1990) offer several different tests for overdispersion. A simple regression-based procedure used for testing the hypothesis

$$H_0\colon \mathrm{Var}[y_i] = E[y_i],$$
$$H_1\colon \mathrm{Var}[y_i] = E[y_i] + \alpha g(E[y_i]),$$

is carried out by regressing

$$z_i = \frac{(y_i - \hat{\lambda}_i)^2 - y_i}{\hat{\lambda}_i\sqrt{2}},$$

where $\hat{\lambda}_i$ is the predicted value from the regression, on either a constant term or $\hat{\lambda}_i$ without a constant term. A simple $t$ test of whether the coefficient is significantly different from zero tests $H_0$ versus $H_1$.

Cameron and Trivedi's regression based test for overdispersion is formulated around the alternative $\mathrm{Var}[y_i] = E[y_i] + g(E[y_i])$. This is a very specific type of overdispersion. Consider the more general hypothesis that $\mathrm{Var}[y_i]$ is completely given by $E[y_i]$. The alternative is that the variance is systematically related to the regressors in a way that is not completely accounted for by $E[y_i]$. Formally, we have $E[y_i] = \exp(\boldsymbol{\beta}'\mathbf{x}_i) = \lambda_i$. The null hypothesis is that $\mathrm{Var}[y_i] = \lambda_i$ as well. We can test the hypothesis using a **conditional moment test.** The expected first derivatives and the moment restriction are

$$E[\mathbf{x}_i(y_i - \lambda_i)] = \mathbf{0} \quad \text{and} \quad E\{\mathbf{z}_i[(y_i - \lambda_i)^2 - \lambda_i]\} = \mathbf{0}.$$

To carry out the test, we do the following. Let $e_i = y_i - \hat{\lambda}_i$ and $\mathbf{z}_i = \mathbf{x}_i$ without the constant term.

1. Compute the Poisson regression by maximum likelihood.
2. Compute $\mathbf{r} = \sum_{i=1}^n \mathbf{z}_i[e_i^2 - \hat{\lambda}_i] = \sum_{i=1}^n \mathbf{z}v_i$ based on the maximum likelihood estimates.
3. Compute $\mathbf{M'M} = \sum_{i=1}^n \mathbf{z}_i\mathbf{z}_i'v_i^2$, $\mathbf{D'D} = \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i'e_i^2$, and $\mathbf{M'D} = \sum_{i=1}^n \mathbf{z}_i\mathbf{x}_i'v_ie_i$.
4. Compute $\mathbf{S} = \mathbf{M'M} - \mathbf{M'D}(\mathbf{D'D})^{-1}\mathbf{D'M}$.
5. $C = \mathbf{r}'\mathbf{S}^{-1}\mathbf{r}$ is the chi-squared statistic. It has degrees of freedom equal to the number of variables in $\mathbf{z}_i$.

KI

The next section presents the **negative binomial model.** This model relaxes the Poisson assumption that the mean equals the variance. The Poisson model is obtained as a parametric restriction on the negative binomial model, so a Lagrange multiplier test can be computed. In general, if an alternative distribution for which the Poisson model is obtained as a parametric restriction, such as the negative binomial model, can be specified, then a Lagrange multiplier statistic can be computed. [See Cameron and

Trivedi (1986, p. 41).] The LM statistic is

$$\text{LM} = \left[\frac{\sum_{i=1}^{n} \hat{w}_i[(y_i - \hat{\lambda}_i)^2 - y_i]}{\sqrt{2\sum_{i=1}^{n} \hat{w}_i \hat{\lambda}_i^2}}\right]^2. \tag{25-3}$$

(18-19)

The weight, $\hat{w}_i$, depends on the assumed alternative distribution. For the negative binomial model discussed later, $\hat{w}_i$ equals 1.0. Thus, under this alternative, the statistic is particularly simple to compute:

$$\text{LM} = \frac{(\mathbf{e}'\mathbf{e} - n\bar{y})^2}{2\hat{\boldsymbol{\lambda}}'\hat{\boldsymbol{\lambda}}}. \tag{25-4}$$

(18-20)

The main advantage of this test statistic is that one need only estimate the Poisson model to compute it. Under the hypothesis of the Poisson model, the limiting distribution of the LM statistic is chi-squared with one degree of freedom.

### 18.4.4 ~~25.2.3~~ HETEROGENEITY AND THE NEGATIVE BINOMIAL REGRESSION MODEL

The assumed equality of the conditional mean and variance functions is typically taken to be the major shortcoming of the Poisson regression model. Many alternatives have been suggested [see Hausman, Hall, and Griliches (1984), Cameron and Trivedi (1986, 1998), Gurmu and Trivedi (1994), Johnson and Kotz (1993), and Winkelmann (2003) for discussion]. The most common is the negative binomial model, which arises from a natural formulation of cross-section heterogeneity. [See Hilbe (2007).] We generalize the Poisson model by introducing an individual, unobserved effect into the conditional mean,

$$\ln \mu_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i = \ln \lambda_i + \ln u_i,$$

where the disturbance $\varepsilon_i$ reflects either **specification error,** as in the classical regression model, or the kind of cross-sectional heterogeneity that normally characterizes microeconomic data. Then, the distribution of $y_i$ conditioned on $\mathbf{x}_i$ *and* $u_i$ (i.e., $\varepsilon_i$) remains Poisson with conditional mean and variance $\mu_i$:

$$f(y_i \mid \mathbf{x}_i, u_i) = \frac{e^{-\lambda_i u_i}(\lambda_i u_i)^{y_i}}{y_i!}.$$

The unconditional distribution $f(y_i \mid \mathbf{x}_i)$ is the expected value (over $u_i$) of $f(y_i \mid \mathbf{x}_i, u_i)$,

$$f(y_i \mid \mathbf{x}_i) = \int_0^{\infty} \frac{e^{-\lambda_i u_i}(\lambda_i u_i)^{y_i}}{y_i!} g(u_i)\, du_i.$$

FN 14

The choice of a density for $u_i$ defines the unconditional distribution. For mathematical convenience, a gamma distribution is usually assumed for $u_i = \exp(\varepsilon_i)$.[14] As in other models of heterogeneity, the mean of the distribution is unidentified if the model contains a constant term (because the disturbance enters multiplicatively) so $E[\exp(\varepsilon_i)]$ is

14 ~~5~~ An alternative approach based on the normal distribution is suggested in Terza (1998), Greene (1995a, 1997a, 2007d), ~~and~~ Winkelmann (1997), and Riphahn, Wambach, and Million (2003). The normal-Poisson mixture is also easily extended to the random effects model discussed in the next section. There is no closed form for the normal-Poisson mixture model, but it can be easily approximated by using Hermite quadrature or simulation. See Sections ~~16~~ 14.9.6.b and ~~23.5.1~~ 17.4.8.

18-57



assumed to be 1.0. With this normalization,

$$g(u_i) = \frac{\theta^\theta}{\Gamma(\theta)} e^{-\theta u_i} u_i^{\theta-1}.$$

The density for $y_i$ is then

$$\begin{aligned} f(y_i \mid \mathbf{x}_i) &= \int_0^\infty \frac{e^{-\lambda_i u_i}(\lambda_i u_i)^{y_i}}{y_i!} \frac{\theta^\theta u_i^{\theta-1} e^{-\theta u_i}}{\Gamma(\theta)} du_i \\ &= \frac{\theta^\theta \lambda_i^{y_i}}{\Gamma(y_i+1)\Gamma(\theta)} \int_0^\infty e^{-(\lambda_i+\theta)u_i} u_i^{\theta+y_i-1} du_i \\ &= \frac{\theta^\theta \lambda_i^{y_i} \Gamma(\theta+y_i)}{\Gamma(y_i+1)\Gamma(\theta)(\lambda_i+\theta)^{\theta+y_i}} \\ &= \frac{\Gamma(\theta+y_i)}{\Gamma(y_i+1)\Gamma(\theta)} r_i^{y_i}(1-r_i)^\theta, \quad \text{where } r_i = \frac{\lambda_i}{\lambda_i+\theta}, \end{aligned}$$

which is one form of the negative binomial distribution. The distribution has conditional mean $\lambda_i$ and conditional variance $\lambda_i(1 + (1/\theta)\lambda_i)$. [This model is Negbin 2 in Cameron and Trivedi's (1986) presentation.] The negative binomial model can be estimated by maximum likelihood without much difficulty. A test of the Poisson distribution is often carried out by testing the hypothesis $\alpha = 1/\theta = 0$ using the Wald or likelihood ratio test.

### 18.4.5 ~~19 25.2.4~~ FUNCTIONAL FORMS FOR COUNT DATA MODELS

The equidispersion assumption of the Poisson regression model, $E[y_i \mid \mathbf{x}_i] = \text{Var}[y_i \mid \mathbf{x}_i]$, is a major shortcoming. Observed data rarely, if ever, display this feature. The very large amount of research activity on functional forms for count models is often focused on testing for equidispersion and building functional forms that relax this assumption. In practice, the Poisson model is typically only the departure point for an extended specification search.

One easily remedied minor issue concerns the units of measurement of the data. In the Poisson and negative binomial models, the parameter $\lambda_i$ is the expected number of events *per unit of time*. Thus, there is a presumption in the model formulation, for example, the Poisson, that the same amount of time is observed for each $i$. In a spatial context, such as measurements of the incidence of a disease per group of $N_i$ persons, or the number of bomb craters per square mile (London, 1940), the assumption would be that the same physical area or the same size of population applies to each observation. Where this differs by individual, it will introduce a type of heteroscedasticity in the model. The simple remedy is to modify the model to account for the **exposure**, $T_i$, of the observation as follows:

$$\text{Prob}(y_i = j \mid \mathbf{x}_i, T_i) = \frac{\exp(-T_i\phi_i)(T_i\phi_i)^j}{j!}, \quad \phi_i = \exp(\mathbf{x}_i'\boldsymbol{\beta}),\ j = 0, 1, \ldots.$$

The original model is returned if we write $\lambda_i = \exp(\mathbf{x}_i'\boldsymbol{\beta} + \ln T_i)$. Thus, when the exposure differs by observation, the appropriate accommodation is to include the log of exposure in the regression part of the model with a coefficient of 1.0. (For less than

18-58



obvious reasons, the term "offset variable" is commonly associated with the exposure variable $T_i$.) Note that if $T_i$ is the same for all $i$, $\ln T_i$ will simply vanish into the constant term of the model (assuming one is included in $\mathbf{x}_i$).

The recent literature, mostly associating the result with Cameron and Trivedi's (1986, 1998) work, defines two familiar forms of the negative binomial model. The **Negbin 2** (NB2) **form** of the probability is

KT

$$\begin{aligned}\text{Prob}(Y = y_i \mid \mathbf{x}_i) &= \frac{\Gamma(\theta + y_i)}{\Gamma(y_i + 1)\Gamma(\theta)} r_i^{y_i}(1 - r_i)^{\theta},\\ \lambda_i &= \exp(\mathbf{x}_i'\boldsymbol{\beta}),\\ r_i &= \lambda_i/(\theta + \lambda_i).\end{aligned} \qquad (\cancel{25}\text{-}5)\ (18\text{-}21)$$

This is the default form of the model in the received econometrics packages that provide an estimator for this model. The **Negbin 1** (NB1) **form** of the model results if $\theta$ in the preceding is replaced with $\theta_i = \theta\lambda_i$. Then, $r_i$ reduces to $r = 1/(1 + \theta)$, and the density becomes

$$\text{Prob}(Y = y_i \mid \mathbf{x}_i) = \frac{\Gamma(\theta\lambda_i + y_i)}{\Gamma(y_i + 1)\Gamma(\theta\lambda_i)} r^{y_i}(1 - r)^{\theta\lambda_i}. \qquad (\cancel{25}\text{-}6)\ (18\text{-}22)$$

AU: X-ref to Exm 19.1 OK?

This is not a simple reparameterization of the model. The results in Example ~~25~~ 19.1 demonstrate that the log-likelihood functions are not equal at the maxima, and the parameters are not simple transformations in one model versus the other. We are not aware of a theory that justifies using one form or the other for the negative binomial model. Neither is a restricted version of the other, so we cannot carry out a likelihood ratio test of one versus the other. The more general **Negbin *P*** (NBP) family does nest both of them, so this may provide a more general, encompassing approach to finding the right specification. [See Greene (~~2005~~ 2008).] The Negbin $P$ model is obtained by replacing $\theta$ in the Negbin 2 form with $\theta\lambda_i^{2-P}$. We have examined the cases of $P = 1$ and $P = 2$ in (~~25~~ 19-5) and (~~25~~ 19-6). The full model is

KT

$$\text{Prob}(Y = y_i \mid \mathbf{x}_i) = \frac{\Gamma(\theta\lambda_i^Q + y_i)}{\Gamma(y_i + 1)\Gamma(\theta\lambda_i^Q)}\left(\frac{\lambda_i}{\theta\lambda_i^Q + \lambda_i}\right)^{y_i}\left(\frac{\theta\lambda_i^Q}{\theta\lambda_i^Q + \lambda_i}\right)^{\theta\lambda_i^Q}, \quad Q = 2 - P.$$

The conditional mean function for the three cases considered is

$$E[y_i \mid \mathbf{x}_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta}) = \lambda_i.$$

The parameter $P$ is picking up the scaling. A general result is that for all three variants of the model,

$$\text{Var}[y_i \mid \mathbf{x}_i] = \lambda_i\left(1 + \alpha\lambda_i^{P-1}\right), \quad \text{where } \alpha = 1/\theta.$$

Thus, the NB2 form has a variance function that is quadratic in the mean while the NB1 form's variance is a simple multiple of the mean. There have been many other functional forms proposed for count data models, including the generalized Poisson, gamma, and Polya-Aeppli forms described in Winkelmann (2003) and Greene (2007a, Chapter 24).

The heteroscedasticity in the count models is induced by the relationship between the variance and the mean. The single parameter $\theta$ picks up an implicit overall scaling, so it does not contribute to this aspect of the model. As in the linear model, microeconomic data are likely to induce heterogeneity in both the mean and variance of the response

18-59



variable. A specification that allows independent variation of both will be of some virtue. The result

$$\mathrm{Var}[y_i \mid \mathbf{x}_i] = \lambda_i\left(1 + (1/\theta)\lambda_i^{P-1}\right)$$

suggests that a natural platform for separately modeling heteroscedasticity will be the dispersion parameter, $\theta$, which we now parameterize as

$$\theta_i = \theta \exp(\mathbf{z}_i'\delta).$$

Operationally, this is a relatively minor extension of the model. But, it is likely to introduce quite a substantial increase in the flexibility of the specification. Indeed, a heterogeneous Negbin P model is likely to be sufficiently parameterized to accommodate the behavior of most data sets. (Of course, the specialized models discussed in Section ~~25.4,~~ 18.4.8 + 9 for example, the zero inflation models, may yet be more appropriate for a given situation.)

18.7 delete hyphen in two-count

***Example ~~25.1~~ Count Data Models for Doctor Visits***

The study by Riphahn et al. (2003) that provided the data we have used in numerous earlier examples analyzed the two-count variables DocVis (visits to the doctor) and HospVis (visits to the hospital). The authors were interested in the joint determination of these two count variables. One of the issues considered in the study was whether the data contained evidence of moral hazard, that is, whether health care utilization as measured by these two outcomes was influenced by the subscription to health insurance. The data contain indicators of two levels of insurance coverage, PUBLIC, which is the main source of insurance, and ADDON, which is a secondary optional insurance. In the sample of 27,326 observations (family/years), 24,203 individuals held the public insurance. (There is quite a lot of within group variation in this. Individuals did not routinely obtain the insurance for all periods.) Of these 24,203, 23,689 had only public insurance and 514 had both types. (One could not have only the ADDON insurance.) To explore the issue, we have analyzed the DocVis variable with the count data models described in this section. The exogenous variables in our model are

$$\mathbf{x}_{it} = (1, Age, Education, Income, Kids, Public).$$

(Variables are described in ~~Example 11.11~~ Appendix Table F7.1.)

Table ~~25.1~~ 18.14 presents the estimates of the several count models. In all specifications, the coefficient on PUBLIC is positive, large, and highly statistically significant, which is consistent with the results in the authors' study. The various test statistics strongly reject the hypothesis of equidispersion. Cameron and Trivedi's (1990) semiparametric tests from the Poisson model (see Section ~~23.2.2~~ 18.4.3) have $t$ statistics of 22.147 for $g_i = \mu_i$ and 22.504 for $g_i = \mu_i^2$. Both of these are far larger than the critical value of 1.96. The LM statistic is 972,714.48, which is also larger than the (any) critical value. On these bases, we would reject the hypothesis of equidispersion. The Wald and likelihood ratio tests based on the negative binomial models produce the same conclusion. For comparing the different negative binomial models, note that Negbin 2 is the worst of the three by the likelihood function, although NB1 and NB2 are not directly comparable. On the other hand, note that in the NBP model, the estimate of $P$ is more than 10 standard errors from 1.0000 or 2.000, so both NB1 and NB2 are rejected in favor of the unrestricted NBP form of the model. The NBP and the heterogeneous NB2 model are not nested either, but comparing the log-likelihoods, it does appear that the heterogeneous model is substantially superior. We computed the Vuong statistic based on the individual contributions to the log-likelihoods, with $v_i = \ln L_i(\text{NBP}) = \ln L_i(\text{NB2-H})$. (See Section ~~7.8.4~~ 14.6.6.) The value of the statistic is −3.27. On this basis, we would reject NBP in favor of NB2-H. Finally, with regard to the original question, the coefficient on PUBLIC is larger than 10 times the estimated standard error in every specification. We would conclude that the results are consistent with the proposition that there is evidence of moral hazard.

minus sign

18-60



**TABLE ~~25.1~~ 18.14** Estimated Models for DOCVIS (standard errors in parentheses)

| *Variable* | *Poisson* | *Negbin 2* | *Negbin 2 Heterogeneous* | *Negbin 1* | *Negbin P* |
|---|---|---|---|---|---|
| Constant | 0.7162 | 0.7628 | 0.7928 | 0.6848 | 0.6517 |
| | (0.03287) | (0.07247) | (0.07459) | (0.06807) | (0.07759) |
| Age | 0.01844 | 0.01803 | 0.01704 | 0.01585 | 0.01907 |
| | (0.0003316) | (0.0007915) | (0.0008146) | (0.0007042) | (0.0008078) |
| Education | −0.03429 | −0.03839 | −0.03581 | −0.02381 | −0.03388 |
| | (0.001797) | (0.003965) | (0.004036) | (0.003702) | (0.004308) |
| Income | −0.4751 | −0.4206 | −0.4108 | −0.1892 | −0.3337 |
| | (0.02198) | (0.04700) | (0.04752) | (0.04452) | (0.05161) |
| Kids | −0.1582 | −0.1513 | −0.1568 | −0.1342 | −0.1622 |
| | (0.007956) | (0.01738) | (0.01773) | (0.01647) | (0.01856) |
| Public | 0.2364 | 0.2324 | 0.2411 | 0.1616 | 0.2195 |
| | (0.01328) | (0.02900) | (0.03006) | (0.02678) | (0.03155) |
| $P$ | 0.0000 | 2.0000 | 2.0000 | 1.0000 | 1.5473 |
| | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.03444) |
| $\theta$ | 0.0000 | 1.9242 | 2.6060 | 6.1865 | 3.2470 |
| | (0.0000) | (0.02008) | (0.05954) | (0.06861) | (0.1346) |
| $\delta$ (Female) | 0.0000 | 0.0000 | −0.3838 | 0.0000 | 0.0000 |
| | (0.0000) | (0.0000) | (0.02046) | (0.0000) | (0.0000) |
| $\delta$ (Married) | 0.0000 | 0.0000 | −0.1359 | 0.0000 | 0.0000 |
| | (0.0000) | (0.0000) | (0.02307) | (0.0000) | (0.0000) |
| ln $L$ | −104440.3 | −60265.49 | −60121.77 | −60260.68 | −60197.15 |

19

## ~~25~~.3 PANEL DATA MODELS

The familiar approaches to accommodating heterogeneity in panel data have fairly straightforward extensions in the count data setting. [Hausman, Hall, and Griliches (1984) give full details for these models.] We will examine them for the Poisson model. The authors [and Allison (2000)] also give results for the negative binomial model.

### 19 ~~25~~.3.1 ROBUST COVARIANCE MATRICES

The standard asymptotic covariance matrix estimator for the Poisson model is

$$\text{Est. Asy. Var}[\hat{\beta}] = \left[-\frac{\partial^2 \ln L}{\partial\hat{\beta}\,\partial\hat{\beta}'}\right]^{-1} = \left[\sum_{i=1}^{n} \hat{\lambda}_i \mathbf{x}_i \mathbf{x}_i'\right]^{-1} = [\mathbf{X}'\hat{\mathbf{\Lambda}}\mathbf{X}]^{-1},$$

where $\hat{\Lambda}$ is a diagonal matrix of predicted values. The BHHH estimator is

prime

$$\text{Est. Asy. Var}[\hat{\beta}] = \left[\sum_{i=1}^{n}\left(\frac{\partial \ln P_i}{\partial\hat{\beta}}\right)\left(\frac{\partial \ln P_i}{\partial\hat{\beta}}\right)\right]^{-1}$$

$$= \left[\sum_{i=1}^{n}(y_i - \hat{\lambda}_i')^2 \mathbf{x}_i \mathbf{x}_i'\right]^{-1} = [\mathbf{X}'\hat{\mathbf{E}}^2\mathbf{X}]^{-1},$$

18-61

## 18.4.6 Truncation and Censoring in Models for Counts

Truncation and censoring are relatively common in applications of models for counts. Truncation arises as a consequence of discarding what appear to be unusable data, such as the zero values in survey data on the number of uses of recreation facilities [Shaw (1988), Bockstael et al. (1990)]. In this setting, a more common case which also gives rise to truncation is on site sampling. When one is interested in visitation by the entire population, which will naturally include zero visits, but one draws their sample "on site," the distribution of visits is truncated at zero by construction. Every visitor has visited at least once. Shaw (1988), Englin and Shonkwiler (1995), Grogger and Carson (1991), Creel and Loomis (1990), Egan and Herriges (2006) and Martinez-Espinera and Amoako-Tuffour (2008) are among a number of studies that have treated truncation due to on-site sampling in environmental and recreation applications. Truncation will also arise when data are trimmed to remove what appear to be unusual values. Figure 18.6 displays a histogram for the number of doctor visits in the 1988 wave of the GSOEP data THAT we have used in several examples. There is a suspiciously large spike at zero and an extremely long right tail of what might seem to be atypical observations. For modeling purposes, it might be tempting to remove these "non-Poisson" appearing observations in these tails. (Other models might be a better solution.) The distribution that characterizes what remains in the sample is a truncated distribution. Truncation is not innocent. If the entire population is of interest, then conventional statistical inference (such as estimation) on the truncated sample produces a systematic bias known as (of course) "truncation bias." This would arise, for example, if an ordinary Poisson model intended to characterize the full population is fit to the sample from a truncated population.

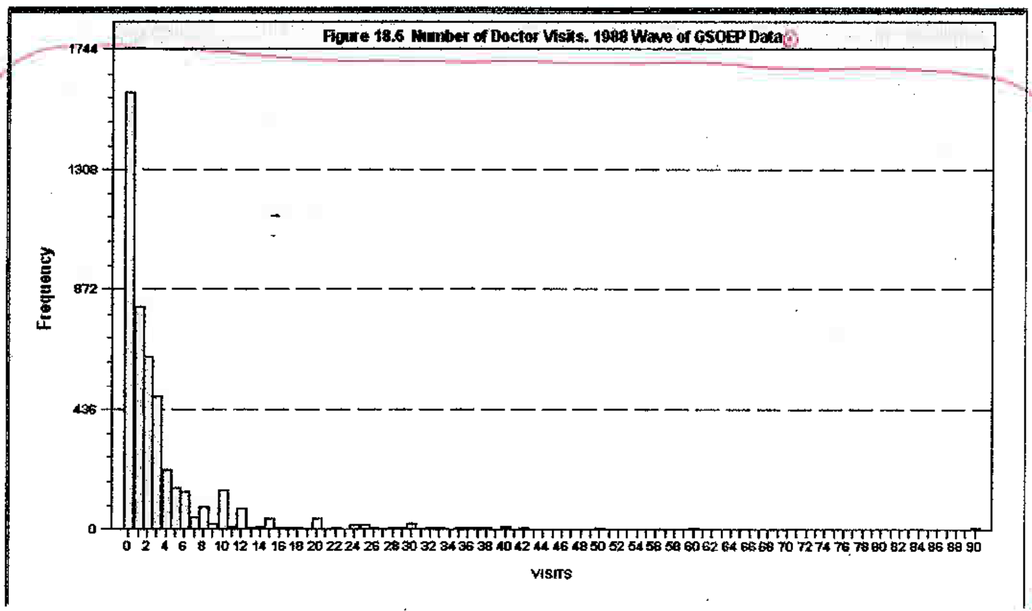


Figure 18.6 Number of Doctor Visits. 1988 Wave of GSOEP Data

Censoring, in contrast, is generally a feature of the sampling design. In the application in Example 18.9, the dependent variable is the self reported number of extramarital affairs in a survey taken by the magazine *Psychology Today*. The possible answers are 0,1,2,3, 4-10 (coded as 7) and "monthly, weekly or daily" coded as 12. The two upper categories are censored. Similarly, in the doctor visits data in the previous paragraph, recognizing the possibility of truncation bias due to data trimming, we might, instead, simply censor the distribution of values at 15. The resulting variable would take values 0,...,14, "15 or more." In both cases, applying conventional estimation methods leads to predictable biases. However, it is

18-62

also possible to reconstruct the estimators specifically to account for the truncation or censoring in the data.

Truncation and censoring produce similar effects on the distribution of the random variable and on the features of the population such as the mean. For the truncation case, suppose that the original random variable has a Poisson distribution (all these results can be directly extended to the negative binomial or any of the other models considered earlier),

minus

$$P(y_i = j|\mathbf{x}_i) = \exp(-\lambda_i)\lambda_i^j/j! = P_{i,j}.$$

If the distribution is truncated at value $C$ – that is, only values $C$+1,... are observed, then the resulting random variable has probability distribution

$$P(y_i = j\,|\,\mathbf{x}_i, y_i > C) = \frac{P(y_i = j\,|\,\mathbf{x}_i)}{P(y_i > C\,|\,\mathbf{x}_i)} = \frac{P(y_i = j\,|\,\mathbf{x}_i)}{1 - P(y_i \le C\,|\,\mathbf{x}_i)}.$$

Au: ok to spell out i.e.?

The original distribution must be scaled up so that it sums to one for the cells that remain in the truncated distribution. The leading case is truncation at zero, i.e., "left truncation", which, for the Poisson model produces

that is,

$$P(y_i = j\,|\,\mathbf{x}_i, y_i > 0) = \frac{\exp(-\lambda_i)\lambda_i^j}{j![1 - \exp(-\lambda_i)]} = \frac{P_{i,j}}{1 - P_{i,0}}, \; j = 1, \ldots$$

[See, e.g., Mullahy (1986), Shaw (1988), Grogger and Carson (1991), Greene (1998), and Winkelmann (1987).] The conditional mean function is

$$E(y_i\,|\,\mathbf{x}_i, y_i > 0) = \frac{1}{[1 - \exp(-\lambda_i)]}\sum_{j=1}^{\infty} \frac{j\exp(-\lambda_i)\lambda_i^j}{j!} = \frac{\lambda_i}{[1 - \exp(-\lambda_i)]} > \lambda_i.$$

The second equality results because the sum can be started at zero – the first term is zero – and this produces the expected value of the original variable. As might be expected, truncation "from below" has the effect of increasing the expected value. It can be shown that it decreases the conditional variance however. The partial effects are

$$\delta_i = \frac{\partial E[y_i\,|\,\mathbf{x}_i, y_i > 0]}{\partial \mathbf{x}_i} = \left[\frac{1 - P_{i,0} - \lambda_i P_{i,0}}{(1 - P_{i,0})^2}\right]\lambda_i \boldsymbol{\beta}. \qquad (18\text{-}23)$$

The term outside the brackets is the partial effects in the absence of the truncation while the bracketed term rises from slighter greater than 0.5 to 1.0 as $\lambda_i$ increases from just above zero.

***EXAMPLE 18.8 Major Derogatory Reports***

In Section 17.5.6 and Examples 17.9 and 17.22, we examined a binary choice model for the accept/reject decision for a sample of applicants for a major credit card. Among the variables in that model is "Major Derogatory Reports" (MDRs). This is an interesting behavioral variable in its own right that should be appropriately modeled using the count data specifications in this chapter. In the sample of 13,444 individuals, 10,833 had zero MDRs while the values for the remaining 2561 ranged from 1 to 22. This preponderance of zeros exceeds by far what one would anticipate in a Poisson model that was dispersed enough to

18-63

TB 18.15

produce the distribution of remaining individuals. As we will pursue an Example 18.11, a natural approach for these data is to treat the extremely large block of zeros explicitly in an extended model. For present purposes, we will consider the nonzero observations apart from the zeros, and examine the effect of accounting for left truncation at zero on the estimated models. Estimation results are shown in Table 18.15. The first column of results compared to the second shows the suspected impact of incorrectly including the zero observations. The coefficients change only slightly, but the partial effects are far smaller when the zeros are included in the estimation. It was not possible to fit the truncated negative binomial with these data.

**TABLE 18.15 Estimated ~~Censored~~ Truncated Poison Regression Model (t ratios in parentheses)**

| | Poisson Full Sample | Poisson | Truncated Poisson |
|---|---|---|---|
| **Constant** | 0.8756 (17.10) | 0.8698 (16.78) | 0.7400 (11.99) |
| **Age** | 0.0036 (2.38) | 0.0035 (2.32) | 0.0049 (2.75) |
| **Income** | -0.0039 (-4.78) | -0.0036 (-3.83) | -0.0051 (-4.51) |
| **Own-Rent** | -0.1005 (-3.52) | -0.1020 (-3.56) | -0.1415 (-4.18) |
| **Self Employed** | -0.0325 (-0.62) | -0.0345 (-0.66) | -0.0515 (-0.82) |
| **Dependents** | 0.0445 (4.69) | 0.0440 (4.62) | 0.0606 (5.48) |
| **MthsCurAdr** | 0.00004 (0.23) | 0.00005 (0.25) | 0.00007 (0.30) |
| **ln *L*** | -5379.30 | -5378.79 | -5097.08 |
| | **Average Partial Effects** | | |
| **Age** | 0.0017 | 0.0085 | 0.0084 |
| **Income** | -0.0018 | -0.0087 | -0.0089 |
| **Own-Rent** | -0.0465 | -0.2477 | -0.2460 |
| **Self-Employed** | -0.0150 | -0.0837 | -0.0895 |
| **Dependents** | 0.0206 | 0.1068 | 0.1054 |
| **MthsCurAdr** | 0.00002 | 0.00012 | 0.00013 |
| **Cond'l. Mean** | 0.4628 | 2.4295 | 2.4295 |
| **Scale factor** | 0.4628 | 2.4295 | 1.7381 |

Censoring is handled similarly. The usual case is "right censoring," in which realized values greater than or equal to $C$ are all given the value $C$. In this case, we have a two part distribution [see Terza (1985b)]. The observed random variable, $y_i$ is constructed from an underlying random variable, $y_i^*$ by

$$y_i = \text{Min}(y_i^*, C).$$

Probabilities are constructed using the axioms of probability. This produces

$$\text{Prob}(y_i = j|\mathbf{x}_i) = P_{i,j},\ j = 0,1,\ldots,C-1,$$

$$\text{Prob}(y_i = C|\mathbf{x}_i) = \sum_{j=C}^{\infty} P_{i,j} = 1 - \sum_{j=0}^{C-1} P_{i,j}.$$

18-64

In this case, the conditional mean function is

$$\begin{aligned}E[y_i \mid \mathbf{x}_i] &= \sum_{j=0}^{C-1} jP_{i,j} + \sum_{j=C}^{\infty} CP_{i,j} \\ &= \sum_{j=0}^{\infty} jP_{i,j} - \sum_{j=C}^{\infty} (j-C)P_{i,j} \\ &= \lambda_i - \sum_{j=C}^{\infty} (j-C)P_{i,j} \\ &< \lambda_i.\end{aligned}$$

The infinite sum is computed by using the complement. Thus,

$$\begin{aligned}E[y_i \mid \mathbf{x}_i] &= \lambda_i - \left[\sum_{j=0}^{\infty} (j-C)P_{i,j} - \sum_{j=0}^{C-1} (j-C)P_{i,j}\right] \\ &= \lambda_i - (\lambda_i - C) + \sum_{j=0}^{C-1} (j-C)P_{i,j} \\ &= C - \sum_{j=0}^{C-1} (C-j)P_{i,j}.\end{aligned}$$

***EXAMPLE 18.9 Extramarital Affairs***

In 1969, the popular magazine *Psychology Today* published a 101-question survey on sex and asked its readers to mail in their answers. The results of the survey were discussed in the July 1970 issue. From the approximately 2,000 replies that were collected in electronic form (of about 20,000 received), Professor Ray Fair (1978) extracted a sample of 601 observations on men and women then currently married for the first time and analyzed their responses to a question about extramarital affairs. Fair's analysis in this frequently cited study suggests several interesting econometric questions. [In addition, his 1977 companion paper in *Econometrica* on estimation of the tobit model contributed to the development of the EM algorithm, which was published by and is usually associated with Dempster, Laird, and Rubin (1977).]

Fair used the tobit model that we discuss in Chapter 19 as a platform The nonexperimental nature of the data (which can be downloaded from the Internet at http://fairmodel.econ.yale.edu/rayfair/work.ss.htm and are given in Appendix Table F18.1). provides a laboratory case that we can use to examine the relationships among the tobit, truncated regression, and probit models. Although the tobit model seems to be a natural choice for the model for these data, given the clusterof zeros, the fact that the behavioral outcome variable is a count that typically takes a small value suggests that the models for counts that we have examined in this chapter might be yet a better choice. Finally, the preponderance of zeros in the data that initially motivated the tobit model suggests that even the standard Poisson model, although an improvement, might still be inadequate. We will pursue that aspect of the data later. In this example, we will focus on just the censoring issue. Other features of the models and data are reconsidered in the exercises.

The study was based on 601 observations on the following variables (full details on data coding are given in the data file and Appendix Table F18.1):

$y$ = number of affairs in the past year, 0, 1, 2, 3, 4–10 coded as 7, "monthly, weekly, or daily," coded as 12. Sample mean=1.46. Frequencies=(451, 34, 17, 19, 42, 38).
$z1$ = sex=0 for female, 1 for male. Sample mean=0.476.
$z2$ = age. Sample mean=32.5.
$z3$ = number of years married. Sample mean=8.18.
$z4$ = children, 0=no, 1=yes. Sample mean=0.715.
$z5$ = religiousness, 1=anti, . . . , 5=very. Sample mean=3.12.

18-65

$z6$ = education, years, 9=grade school, 12=high school, . . . , 20=Ph.D or other. Sample mean=16.2.
$z7$ = occupation, "Hollingshead scale," 1–7. Sample mean=4.19.
$z8$ = self-rating of marriage, 1=very unhappy, . . . , 5=very happy. Sample mean=3.93.

AU: Confirm TB 18.16 is correct

The tobit model was fit to $y$ using a constant term and all eight variables. A restricted model was fit by excluding $z1$, $z4$, and $z6$, none of which was individually statistically significant in the model. We are able to match exactly Fair's results for both equations. The tobit model should only be viewed as an approximation for these data. The dependent variable is a count, not a continuous measurement. The Poisson regression model, or perhaps one of the many variants of it, should be a preferable modeling framework. Table 18.15 presents estimates of the Poisson and negative binomial regression models. There is ample evidence of overdispersion in these data; the $t$ ratio on the estimated overdispersion parameter is 7.014/0.945 = 7.42, which is strongly suggestive. The large absolute value of the coefficient is likewise suggestive.

TB 18.16

16

Responses of 7 and 12 do not represent the actual counts. It is unclear what the effect of the first recoding would be, because it might well be the mean of the observations in this group. But the second is clearly a censored observation. To remove both of these effects, we have recoded both the values 7 and 12 as 4 and treated this observation (appropriately) as a censored observation, with 4 denoting "4 or more." As shown in the third and fourth sets of results in Table 18.16, the effect of this treatment of the data is greatly to reduce the measured effects. Although this step does remove a deficiency in the data, it does not remove the overdispersion; at this point, the negative binomial model is still the preferred specification.

**TABLE 18.16 Censored Poisson and Negative Binomial Distributions**

| | *Poisson Regression* | | | *Negative Binomial Regression* | | |
|---|---|---|---|---|---|---|
| *Variable* | *Estimate* | *Standard Error* | *Marginal Effect* | *Estimate* | *Standard Error* | *Marginal Effect* |
| | | | *Based on Uncensored Poisson Distribution* | | | |
| Constant | 2.53 | 0.197 | — | 2.19 | 0.664 | — |
| $z_2$ | −0.0322 | 0.00585 | −0.0470 | −0.0262 | 0.0192 | −0.00393 |
| $z_3$ | 0.116 | 0.00991 | 0.168 | 0.0848 | 0.0350 | 0.127 |
| $z_5$ | −0.354 | 0.0309 | −0.515 | −0.422 | 0.111 | −0.632 |
| $z_7$ | 0.0798 | 0.0194 | 0.116 | 0.0604 | 0.0702 | 0.0906 |
| $z_8$ | −0.409 | 0.0274 | −0.0596 | −0.431 | 0.111 | −0.646 |
| $\alpha$ | | | | 7.01 | 0.786 | |
| ln $L$ | −1427.037 | | | −728.2441 | | |
| | | | *Based on Poisson Distribution Right Censored at $y = 4$* | | | |
| Constant | 1.90 | 0.283 | — | 4.79 | 1.16 | — |
| $z_2$ | −0.0328 | 0.00838 | −0.0235 | −0.0166 | 0.0250 | −0.00428 |
| $z_3$ | 0.105 | 0.0140 | 0.0754 | 0.174 | 0.0568 | 0.045 |
| $z_5$ | −0.323 | 0.0437 | −0.232 | −0.723 | 0.198 | −0.186 |
| $z_7$ | 0.0798 | 0.0275 | 0.0521 | 0.0900 | 0.116 | 0.0232 |
| $z_8$ | −0.390 | 0.0391 | −0.279 | −0.854 | 0.216 | −0.220 |
| $\alpha$ | | | | 9.40 | 1.35 | |
| ln $L$ | −747.7541 | | | −482.0505 | | |

18-11



**TABLE 25.1** Estimated Models for DOCVIS (standard errors in parentheses)

| *Variable* | *Poisson* | *Negbin 2* | *Negbin 2 Heterogeneous* | *Negbin 1* | *Negbin P* |
|---|---|---|---|---|---|
| Constant | 0.7162<br>(0.03287) | 0.7628<br>(0.07247) | 0.7928<br>(0.07459) | 0.6848<br>(0.06807) | 0.6577<br>(0.07759) |
| Age | 0.01844<br>(0.0003316) | 0.01803<br>(0.0007915) | 0.01704<br>(0.0008146) | 0.01585<br>(0.0007042) | 0.01907<br>(0.0008078) |
| Education | −0.03429<br>(0.001797) | −0.03839<br>(0.003965) | −0.03581<br>(0.004036) | −0.02381<br>(0.003702) | −0.03388<br>(0.004308) |
| Income | −0.4751<br>(0.02198) | −0.4206<br>(0.04700) | −0.4108<br>(0.04752) | −0.1892<br>(0.04452) | −0.3337<br>(0.05161) |
| Kids | −0.1582<br>(0.007956) | −0.1513<br>(0.01738) | −0.1568<br>(0.01773) | −0.1342<br>(0.01647) | −0.1622<br>(0.01856) |
| Public | 0.2364<br>(0.01328) | 0.2324<br>(0.02900) | 0.2411<br>(0.03006) | 0.1616<br>(0.02678) | 0.2195<br>(0.03155) |
| $P$ | 0.0000<br>(0.0000) | 2.0000<br>(0.0000) | 2.0000<br>(0.0000) | 1.0000<br>(0.0000) | 1.5473<br>(0.03444) |
| $\theta$ | 0.0000<br>(0.0000) | 1.9242<br>(0.02008) | 2.6060<br>(0.05954) | 6.1865<br>(0.06861) | 3.2470<br>(0.1346) |
| $\delta$ (Female) | 0.0000<br>(0.0000) | 0.0000<br>(0.0000) | −0.3838<br>(0.02046) | 0.0000<br>(0.0000) | 0.0000<br>(0.0000) |
| $\delta$ (Married) | 0.0000<br>(0.0000) | 0.0000<br>(0.0000) | −0.1359<br>(0.02307) | 0.0000<br>(0.0000) | 0.0000<br>(0.0000) |
| ln $L$ | −104440.3 | −60265.49 | −60121.77 | −60260.68 | −60197.15 |

## 18.4.7 ~~25.3~~ PANEL DATA MODELS

The familiar approaches to accommodating heterogeneity in panel data have fairly straightforward extensions in the count data setting. [Hausman, Hall, and Griliches (1984) give full details for these models.] We will examine them for the Poisson model. The authors [and Allison (2000)] also give results for the negative binomial model.

### 18.4.7.a ~~25.3.1~~ ROBUST COVARIANCE MATRICES FOR POOLED ESTIMATORS

The standard asymptotic covariance matrix estimator for the Poisson model is

$$\text{Est. Asy. Var}[\hat{\beta}] = \left[-\frac{\partial^2 \ln L}{\partial\hat{\beta}\,\partial\hat{\beta}'}\right]^{-1} = \left[\sum_{i=1}^{n}\hat{\lambda}_i \mathbf{x}_i\mathbf{x}_i'\right]^{-1} = [\mathbf{X}'\hat{\mathbf{\Lambda}}\mathbf{X}]^{-1},$$

where $\hat{\Lambda}$ is a diagonal matrix of predicted values. The BHHH estimator is

$$\text{Est. Asy. Var}[\hat{\beta}] = \left[\sum_{i=1}^{n}\left(\frac{\partial \ln P_i}{\partial\hat{\beta}}\right)\left(\frac{\partial \ln P_i}{\partial\hat{\beta}}\right)'\right]^{-1}$$

prime

$$= \left[\sum_{i=1}^{n}(y_i - \hat{\lambda}_i)^2\mathbf{x}_i\mathbf{x}_i'\right]^{-1} = [\mathbf{X}'\hat{\mathbf{E}}^2\mathbf{X}]^{-1},$$

18-67



where $\hat{\mathbf{E}}$ is a diagonal matrix of residuals. The Poisson model is one in which the MLE is robust to certain misspecifications of the model, such as the failure to incorporate latent heterogeneity in the mean (i.e., one fits the Poisson model when the negative binomial is appropriate). In this case, a robust covariance matrix is the "sandwich" estimator,

$$\text{Robust Est. Asy. Var}[\hat{\beta}] = [\mathbf{X}'\hat{\mathbf{\Lambda}}\mathbf{X}]^{-1}[\mathbf{X}'\hat{\mathbf{E}}^2\mathbf{X}][\mathbf{X}'\hat{\mathbf{\Lambda}}\mathbf{X}]^{-1},$$

which is appropriate to accommodate this failure of the model. It has become common to employ this estimator with all specifications, including the negative binomial. One might question the virtue of this. Because the negative binomial model already accounts for the latent heterogeneity, it is unclear what *additional* failure of the assumptions of the model this estimator would be robust to. The questions raised in Section ~~16~~.8.3 and 14
14 ~~16~~.8.4 about robust covariance matrices would be relevant here.

A related calculation is used when observations occur in groups that may be correlated. This would include a random effects setting in a panel in which observations have a common latent heterogeneity as well as more general, stratified, and clustered data sets. The parameter estimator is unchanged in this case (and an assumption is made that the estimator is still consistent), but an adjustment is made to the estimated asymptotic covariance matrix. The calculation is done as follows: Suppose the $n$ observations are assembled in $G$ clusters of observations, in which the number of observations in the $i$th cluster is $n_i$. Thus, $\sum_{i=1}^{G} n_i = n$. Denote by $\beta$ the full set of model parameters in whatever variant of the model is being estimated. Let the observation-specific gradients and Hessians be $\mathbf{g}_{ij} = \partial \ln L_{ij}/\partial\beta = (y_{ij} - \lambda_{ij})\mathbf{x}_{ij}$ and $\mathbf{H}_{ij} = \partial^2 \ln L_{ij}/\partial\beta\partial\beta' = -\lambda_{ij}\mathbf{x}_{ij}\mathbf{x}_{ij}'$. The uncorrected estimator of the asymptotic covariance matrix based on the Hessian is

$$\mathbf{V}_H = -\mathbf{H}^{-1} = \left(-\sum_{i=1}^{G}\sum_{j=1}^{n_i}\mathbf{H}_{ij}\right)^{-1}.$$

The corrected asymptotic covariance matrix is

$$\text{Est. Asy. Var}[\hat{\beta}] = \mathbf{V}_H\left(\frac{G}{G-1}\right)\left[\sum_{i=1}^{G}\left(\sum_{j=1}^{n_i}\mathbf{g}_{ij}\right)\left(\sum_{j=1}^{n_i}\mathbf{g}_{ij}\right)'\right]\mathbf{V}_H.$$

Note that if there is exactly one observation per cluster, then this is $G/(G-1)$ times the sandwich (robust) estimator.

### 18.4.7.b ~~25~~.3.2 FIXED EFFECTS

Consider first a fixed effects approach. The Poisson distribution is assumed to have conditional mean

$$\log \lambda_{it} = \beta'\mathbf{x}_{it} + \alpha_i, \qquad \text{(18-24)} \; (\text{~~25-7~~})$$

where now, $\mathbf{x}_{it}$ has been redefined to exclude the constant term. The approach used in the linear model of transforming $y_{it}$ to group mean deviations does not remove the heterogeneity, nor does it leave a Poisson distribution for the transformed variable. However, the Poisson model with fixed effects can be fit using the methods described for the probit model in Section ~~23.5.2~~. The extension to the Poisson model requires only the minor modifications, $g_{it} = (y_{it} - \lambda_{it})$ and $h_{it} = -\lambda_{it}$. Everything else in that derivation applies with only a simple change in the notation. The first-order conditions

17.4.3

18-68



for maximizing the log-likelihood function for the Poisson model will include

$$\frac{\partial \ln L}{\partial \alpha_i} = \sum_{t=1}^{T_i}(y_{it} - e^{\alpha_i}\mu_{it}) = 0 \quad \text{where } \mu_{it} = e^{\mathbf{x}_{it}'\boldsymbol{\beta}}.$$

This implies an explicit solution for $\alpha_i$ in terms of $\boldsymbol{\beta}$ in this model,

$$\hat{\alpha}_i = \ln\left(\frac{(1/T_i)\sum_{t=1}^{T_i} y_{it}}{(1/T_i)\sum_{t=1}^{T_i}\hat{\mu}_{it}}\right) = \ln\left(\frac{\bar{y}_i}{\bar{\hat{\mu}}_i}\right). \qquad (25\text{-}8)$$

(18-25)

Unlike the regression or the probit model, this does not require that there be within-group variation in $y_{it}$—all the values can be the same. It does require that at least one observation for individual $i$ be nonzero, however. The rest of the solution for the fixed effects estimator follows the same lines as that for the probit model. An alternative approach, albeit with little practical gain, would be to concentrate the log-likelihood function by inserting this solution for $\alpha_i$ back into the original log-likelihood, then maximizing the resulting function of $\boldsymbol{\beta}$. While logically this makes sense, the approach suggested earlier for the probit model is simpler to implement. and

An estimator that is not a function of the fixed effects is found by obtaining the joint distribution of $(y_{i1}, \ldots, y_{iT_i})$ conditional on their sum. For the Poisson model, a close cousin to the multinomial logit model discussed earlier is produced:

$$p\left(y_{i1}, y_{i2}, \ldots, y_{iT_i}\,\middle|\,\sum_{t=1}^{T_i} y_{it}\right) = \frac{\left(\sum_{t=1}^{T_i} y_{it}\right)!}{\left(\prod_{t=1}^{T_i} y_{it}!\right)}\prod_{t=1}^{T_i} p_{it}^{y_{it}}, \qquad (25\text{-}9)$$

(18-26)

where

$$p_{it} = \frac{e^{\mathbf{x}_{it}'\boldsymbol{\beta}+\alpha_i}}{\sum_{t=1}^{T_i} e^{\mathbf{x}_{it}'\boldsymbol{\beta}+\alpha_i}} = \frac{e^{\mathbf{x}_{it}'\boldsymbol{\beta}}}{\sum_{t=1}^{T_i} e^{\mathbf{x}_{it}'\boldsymbol{\beta}}}. \qquad (25\text{-}10)$$

(18-27)

The contribution of group $i$ to the conditional log-likelihood is

$$\ln L_i = \sum_{t=1}^{T_i} y_{it}\ln p_{it}.$$

Note, once again, that the contribution to $\ln L$ of a group in which $y_{it} = 0$ in every period is zero. Cameron and Trivedi (1998) have shown that these two approaches give identical results.

Hausman, Hall, and Griliches (1984) (HHG) report the following conditional density for the fixed effects negative binomial (FENB) model:

$$p\left(y_{i1}, y_{i2}, \ldots, y_{iT_i}\,\middle|\,\sum_{t=1}^{T_i} y_{it}\right) = \frac{\Gamma\left(1+\sum_{t=1}^{T_i} y_{it}\right)\Gamma\left(\sum_{t=1}^{T_i}\lambda_{it}\right)}{\Gamma\left(\sum_{t=1}^{T_i} y_{it} + \sum_{t=1}^{T_i}\lambda_{it}\right)}\prod_{t=1}^{T_i}\frac{\Gamma(y_{it}+\lambda_{it})}{\Gamma(1+y_{it})\Gamma(\lambda_{it})},$$

which is free of the fixed effects. This is the default FENB formulation used in popular software packages such as SAS, Stata, and LIMDEP. Researchers accustomed to the admonishments that fixed effects models cannot contain overall constants or time-invariant covariates are sometimes surprised to find (perhaps accidentally) that this fixed effects model allows both. [This issue is explored at length in Allison (2000) and Allison and Waterman (2002).] The resolution of this apparent contradiction is that the

18-69



HHG FENB model is not obtained by shifting the conditional mean function by the fixed effect, $\ln \lambda_{it} = \mathbf{x}_{it}'\beta + \alpha_i$, as it is in the Poisson model. Rather, the HHG model is obtained by building the fixed effect into the model as an individual specific $\theta_i$ in the Negbin 1 form in (~~25~~-6). The conditional mean functions in the models are as follows (we have changed the notation slightly to conform to our earlier formulation):

18-22

$$\text{NB1(HHG): } E[y_{it} \mid \mathbf{x}_{it}] = \theta_i \phi_{it} = \theta_i \exp(\mathbf{x}_{it}'\beta),$$
$$\text{NB2: } E[y_{it} \mid \mathbf{x}_{it}] = \exp(\alpha_i)\phi_{it} = \lambda_{it} = \exp(\mathbf{x}_{it}'\beta + \alpha_i).$$

The conditional variances are

$$\text{NB1(HHG): } \text{Var}[y_{it} \mid \mathbf{x}_{it}] = \theta_i \phi_{it}[1 + \theta_i],$$
$$\text{NB2: } \text{Var}[y_{it} \mid \mathbf{x}_{it}] = \lambda_{it}[1 + \theta\lambda_{it}].$$

Letting $\mu_i = \ln \theta_i$, it appears that the HHG formulation does provide a fixed effect in the mean, as now, $E[y_{it} \mid \mathbf{x}_{it}] = \exp(\mathbf{x}_{it}'\beta + \mu_i)$. Indeed, by this construction, it appears (as the authors suggest) that there are separate effects in both the mean and the variance. They make this explicit by writing $\theta_i = \exp(\mu_i)\gamma_i$ so that in their model,

$$E[y_{it} \mid \mathbf{x}_{it}] = \gamma_i \exp(\mathbf{x}_{it}'\beta + \mu_i),$$
$$\text{Var}[y_{it} \mid \mathbf{x}_{it}] = \gamma_i \exp(\mathbf{x}_{it}'\beta + \mu_i)/[1 + \gamma_i \exp(\mu_i)].$$

The contradiction arises because the authors assert that $\mu_i$ and $\gamma_i$ are separate parameters. In fact, they cannot vary separately; only $\theta_i$ can vary autonomously. The firm-specific effect in the HHG model is still isolated in the scaling parameter, which falls out of the conditional density. The mean is homogeneous, which explains why a separate constant, or a time invariant regressor (or another set of firm-specific effects) can reside there. [See Greene (2007d) and Allison and Waterman (2002) for further discussion.]

### 18.4.7.c ~~25~~.3.3 RANDOM EFFECTS

The fixed effects approach has the same flaws and virtues in this setting as in the probit case. It is not necessary to assume that the heterogeneity is uncorrelated with the included, exogenous variables. If the uncorrelatedness of the regressors and the heterogeneity can be maintained, then the random effects model is an attractive alternative model. Once again, the approach used in the linear regression model, partial deviations from the group means followed by generalized least squares (see Chapter ~~9~~), is not usable here. The approach used is to formulate the joint probability conditioned upon the heterogeneity, then integrate it out of the joint distribution. Thus, we form

$$p(y_{i1}, \ldots, y_{iT_i} \mid u_i) = \prod_{t=1}^{T_i} p(y_{it} \mid u_i).$$

Then the random effect is swept out by obtaining

$$p(y_{i1}, \ldots, y_{iT_i}) = \int_{u_i} p(y_{i1}, \ldots, y_{iT_i}, u_i)\, du_i$$
$$= \int_{u_i} p(y_{i1}, \ldots, y_{iT_i} \mid u_i) g(u_i)\, du_i$$
$$= E_{u_i}[p(y_{i1}, \ldots, y_{iT_i} \mid u_i)].$$

18-70



This is exactly the approach used earlier to condition the heterogeneity out of the Poisson model to produce the negative binomial model. If, as before, we take $p(y_{it} \mid u_i)$ to be Poisson with mean $\lambda_{it} = \exp(\mathbf{x}_{it}'\beta + u_i)$ in which $\exp(u_i)$ is distributed as gamma with mean 1.0 and variance $1/\alpha$, then the preceding steps produce a negative binomial distribution,

$$p(y_{i1}, \ldots, y_{iT_i}) = \frac{\left[\prod_{t=1}^{T_i} \lambda_{it}^{y_{it}}\right] \Gamma\left(\theta + \sum_{t=1}^{T_i} y_{it}\right)}{\left[\Gamma(\theta) \prod_{t=1}^{T_i} y_{it}!\right] \left[\left(\sum_{t=1}^{T_i} \lambda_{it}\right)^{\sum_{t=1}^{T_i} y_{it}}\right]} Q_i^{\theta}(1 - Q_i)^{\sum_{t=1}^{T_i} y_{it}}, \qquad (25\text{-}14)$$

(18-28)

where

$$Q_i = \frac{\theta}{\theta + \sum_{t=1}^{T_i} \lambda_{it}}.$$

For estimation purposes, we have a negative binomial distribution for $Y_i = \Sigma_t y_{it}$ with mean $\Lambda_i = \Sigma_t \lambda_{it}$.

Like the fixed effects model, introducing random effects into the negative binomial model adds some additional complexity. We do note, because the negative binomial model derives from the Poisson model by adding latent heterogeneity to the conditional mean, adding a random effect to the negative binomial model might well amount to introducing the heterogeneity a second time. However, one might prefer to interpret the negative binomial as the density for $y_{it}$ in its own right, and treat the common effects in the familiar fashion. Hausman et al.'s (1984) random effects negative binomial (RENB) model is a hierarchical model that is constructed as follows. The heterogeneity is assumed to enter $\lambda_{it}$ additively with a gamma distribution with mean 1, $\Gamma(\theta_i, \theta_i)$. Then, $\theta_i/(1+\theta_i)$ is assumed to have a beta distribution with parameters $a$ and $b$ [see Appendix B.4.6)]. The resulting unconditional density after the heterogeneity is integrated out is

$$p(y_{i1}, y_{i2}, \ldots, y_{iT_i}) = \frac{\Gamma(a+b)\Gamma\left(a + \sum_{t=1}^{T_i} \lambda_{it}\right)\Gamma\left(b + \sum_{t=1}^{T_i} y_{it}\right)}{\Gamma(a)\Gamma(b)\Gamma\left(a + \sum_{t=1}^{T_i} \lambda_{it} + b + \sum_{t=1}^{T_i} y_{it}\right)}.$$

As before, the relationship between the heterogeneity and the conditional mean function is unclear, because the random effect impacts the parameter of the scedastic function. An alternative approach that maintains the essential flavor of the Poisson model (and other random effects models) is to augment the NB2 form with the random effect,

$$\text{Prob}(Y = y_{it} \mid \mathbf{x}_{it}, \varepsilon_i) = \frac{\Gamma(\theta + y_{it})}{\Gamma(y_{it}+1)\Gamma(\theta)} r_{it}^{y_{it}} (1 - r_{it})^{\theta},$$
$$\lambda_{it} = \exp(\mathbf{x}_{it}'\beta + \varepsilon_i),$$
$$r_{it} = \lambda_{it}/(\theta + \lambda_{it}).$$

We then estimate the parameters by forming the conditional (on $\varepsilon_i$) log-likelihood and integrating $\varepsilon_i$ out either by quadrature or simulation. The parameters are simpler to interpret by this construction. Estimates of the two forms of the random effects model are presented in Example 25.2 for a comparison.

18.10 There is a mild preference in the received literature for the fixed effects estimators over the random effects estimators. The virtue of dispensing with the assumption of uncorrelatedness of the regressors and the group specific effects is substantial. On the

18-71



other hand, the assumption does come at a cost. To compute the probabilities or the marginal effects, it is necessary to estimate the constants, $\alpha_i$. The unscaled coefficients in these models are of limited usefulness because of the nonlinearity of the conditional mean functions.

Other approaches to the random effects model have been proposed. Greene (1994, 1995a), Riphahn et al. (2003), and Terza (1995) specify a normally distributed heterogeneity, on the assumption that this is a more natural distribution for the aggregate of small independent effects. Brannas and Johanssen (1994) have suggested a semiparametric approach based on the GMM estimator by superimposing a very general form of heterogeneity on the Poisson model. They assume that conditioned on a random effect $\varepsilon_{it}$, $y_{it}$ is distributed as Poisson with mean $\varepsilon_{it}\lambda_{it}$. The covariance structure of $\varepsilon_{it}$ is allowed to be fully general. For $t, s = 1, \ldots, T$, $\text{Var}[\varepsilon_{it}] = \sigma_i^2$, $\text{Cov}[\varepsilon_{it}, \varepsilon_{js}] = \gamma_{ij}(|t-s|)$. For long time series, this model is likely to have far too many parameters to be identified without some restrictions, such as first-order homogeneity ($\beta_i = \beta\,\forall i$), uncorrelatedness across groups, [$\gamma_{ij}(.) = 0$ for $i \neq j$], groupwise homoscedasticity ($\sigma_i^2 = \sigma^2\,\forall i$), and nonautocorrelatedness [$\gamma(r) = 0\,\forall r \neq 0$]. With these assumptions, the estimation procedure they propose is similar to the procedures suggested earlier. If the model imposes enough restrictions, then the parameters can be estimated by the method of moments. The authors discuss estimation of the model in its full generality. Finally, the latent class model discussed in Section ~~16.9.7~~ 14.10 and the random parameters model in Section 17.5 15.9 extend naturally to the Poisson model. Indeed, most of the received applications of the latent class structure have been in the Poisson regression framework. [See Greene (2001) for a survey.]

18.10

*Example ~~25.2~~ Panel Data Models for Doctor Visits*

The German health care panel data set contains 7,293 individuals with group sizes ranging from 1 to 7. Table ~~25.2~~ 18.16 presents the fixed and random effects estimates of the equation for DocVis. The pooled estimates are also shown for comparison. Overall, the panel data treatments bring large changes in the estimates compared to the pooled estimates. There is also a considerable amount of variation across the specifications. With respect to the parameter of interest, *Public*, we find that the size of the coefficient falls substantially with all panel data treatments. Whether using the pooled, fixed, or random effects specifications, the test statistics (Wald, LR) all reject the Poisson model in favor of the negative binomial. Similarly, either common effects specification is preferred to the pooled estimator. There is no simple basis for choosing between the fixed and random effects models, and we have further blurred the distinction by suggesting two formulations of each of them. We do note that the two random effects estimators are producing similar results, which one might hope for. But, the two fixed effects estimators are producing very different estimates. The NB1 estimates include two coefficients, *Income* and *Education*, that are positive, but negative in every other case. Moreover, the coefficient on *Public*, which is large and significant throughout the table, has become small and less significant with the fixed effects estimators.

TB 18.17

We also fit a three-class latent class model for these data. (See Section ~~16.9.7~~ 14.10.) The three class probabilities were modeled as functions of *Married* and *Female*, which appear from the results to be significant determinants of the class sorting. The average prior probabilities for the three classes are 0.09212, 0.49361, and 0.41427. The coefficients on *Public* in the three classes, with associated *t* ratios are 0.3388 (11.541), 0.1907 (3.987), and 0.1084 (4.282). The qualitative result concerning evidence of moral hazard suggested at the outset of Example ~~25.1~~ 18.7 appears to be supported in a variety of specifications (with FE-NB1 the sole exception).

18-72

~~17~~ 18.16

TABLE ~~25.2~~ Estimated Panel Data Models for Doctor Visits (standard errors in parentheses)

| | Poisson | | | | Negative Binomial | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Fixed Effects | | Random Effects | |
| *Variable* | *Pooled (Robust S.E.)* | *Fixed Effects* | *Random Effects* | *Pooled NB2* | *FE-NB1* | *FE-NB2* | *HHG-Gamma* | *Normal* |
| Constant | 0.7162 (0.1319) | 0.0000 | 0.4957 (0.05463) | 0.7628 (0.07247) | −1.2354 (0.1079) | 0.0000 | −0.6343 (0.07328) | 0.1169 (0.06612) |
| Age | 0.01844 (0.001336) | 0.03115 (0.001443) | 0.02329 (0.0004458) | 0.01803 (0.0007916) | 0.02389 (0.001188) | 0.04479 (0.002769) | 0.01899 (0.0007820) | 0.02231 (0.0006969) |
| Educ | −0.03429 (0.007255) | −0.03803 (0.01733) | −0.03427 (0.004352) | −0.03839 (0.003965) | 0.01652 (0.006501) | −0.04589 (0.02967) | −0.01779 (0.004056) | −0.03773 (0.003595) |
| Income | −0.4751 (.08212) | −0.3030 (0.04104) | −0.2646 (0.01520) | −0.4206 (0.04700) | 0.02373 (0.05530) | −0.1968 (0.07320) | −0.08126 (0.04565) | −0.1743 (0.04273) |
| Kids | −0.1582 (0.03115) | −0.001927 (0.01546) | −0.03854 (0.005272) | −0.1513 (0.01738) | −0.03381 (0.02116) | −0.001274 (0.02920) | −0.1103 (0.01675) | −0.1187 (0.01582) |
| Public | 0.2365 (0.04307) | 0.1015 (0.02980) | 0.1535 (0.01268) | 0.2324 (0.02900) | 0.05837 (0.03896) | 0.09700 (0.05334) | 0.1486 (0.02834) | 0.1940 (0.02574) |
| $\theta$ | 0.0000 | 0.0000 | 1.1646 (0.01940) | 1.9242 (0.02008) | 0.0000 | 1.9199 (0.02994) | 0.0000 | 1.0808 (0.01203) |
| $a$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 2.1463 (0.05955) | 0.0000 |
| $b$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 3.8011 (0.1145) | 0.0000 |
| $\sigma$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.9737 (0.008235) |
| ln $L$ | −104440.3 | ~~−60337.13~~ −47703.34 | −71763.13 | −60265.49 | −34016.16 | −49476.36 | −58182.52 | −58177.66 |

18-73

### 18.4.8 Two Part Models: Zero Inflation and Hurdle Models

Mullahy (1986), Heilbron (1989), Lambert (1992), Johnson and Kotz (1993), and Greene (1994) have analyzed an extension of the hurdle model in which the zero outcome can arise from one of two regimes.[15] In one regime, the outcome is always zero. In the other, the usual Poisson process is at work, which can produce the zero outcome or some other. In Lambert's application, she analyzes the number of defective items produced by a manufacturing process in a given time interval. If the process is under control, then the outcome is always zero (by definition). If it is not under control, then the number of defective items is distributed as Poisson and may be zero or positive in any period. The model at work is therefore

FN 15

$$\text{Prob}(y_i = 0 \mid \mathbf{x}_i) = \text{Prob(regime 1)} + \text{Prob}(y_i = 0 \mid \mathbf{x}_i, \text{ regime 2})\text{Prob(regime 2)},$$

$$\text{Prob}(y_i = j \mid \mathbf{x}_i) = \text{Prob}(y_i = j \mid \mathbf{x}_i, \text{ regime 2})\text{Prob(regime 2)}, \; j = 1, 2, \ldots.$$

Let $z$ denote a binary indicator of regime 1 ($z = 0$) or regime 2 ($z = 1$), and let $y^*$ denote the outcome of the Poisson process in regime 2. Then the observed $y$ is $z \times y^*$. A natural extension of the splitting model is to allow $z$ to be determined by a set of covariates. These covariates need not be the same as those that determine the conditional probabilities in the Poisson process. Thus, the model is

$$\text{Prob}(z_i = 0 \mid \mathbf{w}_i) = F(\mathbf{w}_i, \gamma), \qquad \text{(Regime 1: } y \text{ will equal zero.)}$$

$$\text{Prob}(y_i = j \mid \mathbf{x}_i, z_i = 1) = \frac{\exp(-\lambda_i)\lambda_i^j}{j!}. \qquad \text{(Regime 2: } y \text{ will be a count outcome.)}$$

The zero inflation model can also be viewed as a type of latent class model. The two class probabilities are $F(\mathbf{w}_i, \gamma)$ and $1 - F(\mathbf{w}_i, \gamma)$, and the two regimes are $y = 0$ and the Poisson or negative binomial data generating process.[16] The extension of the ZIP formulation to the negative binomial model is widely labeled the ZINB model.[17] [See Zaninotti and Falischetti (2010) for an application.]

minus

FN 16

FN 17

The mean of this random variable in the Poisson case is

minus

$$E[y_i \mid \mathbf{x}_i, \mathbf{w}_i] = F_i \times 0 + (1 - F_i) \times E[y_i^* \mid \mathbf{x}_i, z_i = 1] \;=\; (1 - F_i)\lambda_i.$$

Lambert (1992) and Greene (1994) consider a number of alternative formulations, including logit and probit models discussed in Sections 17.2 and 17.3, for the probability of the two regimes.

KT

KT

[15] The model is variously labeled the "With Zeros," or WZ, model [Mullahy (1986)], the **Zero Inflated Poisson,** or ZIP, model [Lambert (1992)], and "Zero-Altered Poisson," or ZAP, model [Greene (1994)]

[16] Harris and Zhao (2007) applied this approach to a survey of teenage smokers and nonsmokers in Australia, using an ordered probit model. (See Section 18.3.)

[17] Greene (2005) presents a survey of two part models, including the zero inflation models.

It might be of interest to test simply whether there is a regime splitting mechanism at work or not. Unfortunately, the basic model and the zero-inflated model are not nested. Setting the parameters of the splitting model to zero, for example, does not produce Prob[$z = 0$] = 0. In the probit case, this probability becomes 0.5, which maintains the regime split. The preceding tests for over- or underdispersion would be rather indirect. What is desired is a test of non-Poissonness. An alternative distribution may (but need not) produce a systematically different proportion of zeros than the Poisson. Testing for a different distribution, as opposed to a different set of parameters, is a difficult procedure. Because the hypotheses are necessarily nonnested, the power of any test is a function of the alternative hypothesis and may, under some, be small. Vuong (1989) has proposed a test statistic for **nonnested models** that is well suited for this setting when the alternative distribution can be specified. (See Section 14.6.6.) Let $f_j(y_i|\mathbf{x}_i)$ denote the predicted probability that the random variable $Y$ equals $y_i$ under the assumption that the distribution is $f_j(y_i|\mathbf{x}_i)$, for $j = 1, 2$, and let

$$m_i = \ln\left(\frac{f_1(y_i \mid \mathbf{x}_i)}{f_2(y_i \mid \mathbf{x}_i)}\right).$$

Then Vuong's statistic for testing the nonnested hypothesis of model 1 versus model 2 is

$$v = \frac{\sqrt{n}\left[\frac{1}{n}\Sigma_{i=1}^{n} m_i\right]}{\sqrt{\frac{1}{n}\Sigma_{i=1}^{n}\left(m_i - \bar{m}\right)^2}} = \frac{\sqrt{n}\ \bar{m}}{s_m}.$$

This is the standard statistic for testing the hypothesis that $E[m_i]$ equals zero. Vuong shows that $v$ has a limiting standard normal distribution. As he notes, the statistic is bidirectional. If $|v|$ is less than two, then the test does not favor one model or the other. Otherwise, large values favor model 1 whereas small (negative) values favor model 2. Carrying out the test requires estimation of both models and computation of both sets of predicted probabilities. In Greene (1994), it is shown that the Vuong test has some power to discern the zero inflation phenomenon. The logic of the testing procedure is to allow for overdispersion by specifying a negative binomial count data process, and then examine whether, *even allowing for the overdispersion,* there still appear to be excess zeros. In his application, that appears to be the case.

***Example 18.~~10~~ 11 Zero Inflation Models for Major Derogatory Reports***

In Example 18.8, we examined the counts of major derogatory reports for a sample of 13,444 credit card applicants. It was noted that there are over 10,800 zeros in the counts. One might guess that among credit card users, there are a certain (probably large) proportion of individuals who would never generate an MDR, and some other proportion who might or might not, depending on circumstances. We propose to extend the count models in Example 10.8 to accommodate the zeros. The extensions to the ZIP and ZINB models are shown in Table 18.17. Only the coefficients are shown for purpose of the comparisons. Vuong's diagnostic statistic appears to confirm intuition that the Poisson model does not adequately describe the data; the value is 20.6981. Using the model parameters to compute a prediction of the number of zeros, it is clear that the splitting model does perform better than the basic Poisson regression. For the simple Poisson model, the average probability of zero times the sample size gives a prediction of 8609. For the ZIP model, the value is 10914.8, which is a dramatic improvement. ~~The~~ by the likelihood ratio test, the negative binomial is clearly preferred; comparing the two zero inflation models, the difference in the log likelihood functions is over 1,000. As might be expected, the Vuong statistic falls considerably, to 4.5943. However, the simple model with no zero inflation is still rejected by the test.

18-75

TABLE 18.17 Estimated Zero Inflated Count Models

| | Poisson | | | Negative Binomial | | |
|---|---|---|---|---|---|---|
| | Poisson Regression | Zero Inflation | | Negative Binomial | Zero Inflation | |
| | | Regression | Zero Regime | | Regression | Zero Regime |
| **Constant** | -1.33276 | 0.75483 | 2.06919 | -1.54536 | -0:39628 | 4.18910 |
| **Age** | 0.01286 | 0.00358 | -0.01741 | 0.01807 | -0.00280 | -0.14339 |
| **Income** | -0.02577 | -0.05127 | -0.03023 | -0.02482 | -0.05502 | -0.33903 |
| **OwnRent** | -0.17801 | -0.15593 | -0.01738 | -0.18985 | -0.28591 | -0.50026 |
| **Self Empl** | 0.04691 | -0.01257 | | 0.07920 | 0.06817 | |
| **Dependents** | 0.13760 | 0.06038 | -0.09098 | 0.14054 | 0.08599 | -0.32897 |
| **Cur. Add.** | 0.00195 | 0.00046 | | 0.00245 | 0.00257 | |
| **α** | | | | 6.41435 | 4.85653 | |
| **ln L** | -15467.71 | -11569.74 | | -10582.88 | -10516.46 | |
| **Vuong** | | 20.6981 | | | 4.5943 | |

coyment/

In some settings, the zero outcome of the data-generating process is qualitatively different from the positive ones. The zero or nonzero value of the outcome is the result of a separate decision whether or not to "participate" in the activity. On deciding to participate, the individual decides separately how much, that is, how intensively. Mullahy (1986) argues that this fact constitutes a shortcoming of the Poisson (or negative binomial) model and suggests a **hurdle model** as an alternative.[18] In his formulation, a binary probability model determines whether a zero or a nonzero outcome occurs, then, in the latter case, a (truncated) Poisson distribution describes the positive outcomes. The model is

KT FN 18 KT and/

$$\text{Prob}(y_i = 0 \mid \mathbf{x}_i) = e^{-\theta}$$

$$\text{Prob}(y_i = j \mid \mathbf{x}_i) = (1 - e^{-\theta})\frac{\exp(-\lambda_i)\lambda_i^j}{j![1-\exp(-\lambda_i)]}, \; j = 1, 2, \ldots.$$

they/

This formulation changes the probability of the zero outcome and scales the remaining probabilities so that the sum to one. Mullahy suggests some formulations and applies the model to a sample of observations on daily beverage consumption. Mullahy's formulation adds a new restriction that $\text{Prob}(y_i = 0 \mid \mathbf{x}_i)$ no longer depends on the covariates, however. The natural next step is to parameterize this probability. This extension of the hurdle model would combine a binary choice model like those in Section 17.2 and 17.3 with a truncated count model as shown in Section 18.4.6. This would produce, for example, for a logit participation equation and a Poisson intensity equation,

$$\text{Prob}(y_i = 0 \mid \mathbf{w}_i) = \Lambda(\mathbf{w}_i'\boldsymbol{\gamma})$$

$$\text{Prob}(y_i = j \mid \mathbf{x}_i, \mathbf{w}_i, y_i > 0) = \frac{[1-\Lambda(\mathbf{w}_i'\boldsymbol{\gamma})]\exp(-\lambda_i)\lambda_i^j}{j![1-\exp(-\lambda_i)]}.$$

[18] For a similar treatment in continuous data application, see Cragg (1971).

18-76

The conditional mean function in the hurdle model is

$$E[y_i \mid \mathbf{x}_i, \mathbf{w}_i] = \frac{[1 - F(\mathbf{w}_i'\boldsymbol{\gamma})]\lambda_i}{[1 - \exp(-\lambda_i)]}, \lambda_i = \exp(\mathbf{x}_i'\boldsymbol{\beta}),$$

where $F(.)$ is the probability model used for the participation equation (probit or logit). The partial effects are obtained by differentiating with respect to the two sets of variables separately,

$$\frac{\partial E[y_i \mid \mathbf{x}_i, \mathbf{w}_i]}{\partial \mathbf{x}_i} = [1 - F(\mathbf{w}_i'\boldsymbol{\gamma})]\boldsymbol{\delta}_i,$$

$$\frac{\partial E[y_i \mid \mathbf{x}_i, \mathbf{w}_i]}{\partial \mathbf{w}_i} = \left\{ \frac{-f(\mathbf{w}_i'\boldsymbol{\gamma})\lambda_i}{[1 - \exp(-\lambda_i)]} \right\} \boldsymbol{\gamma},$$

where $\boldsymbol{\delta}_i$ is defined in (18-23) and $f(.)$ is the density corresponding to $F(.)$. For variables that appear in both $\mathbf{x}_i$ and $\mathbf{w}_i$, the effects are added. For dummy variables, the preceding would be an approximation; the appropriate result would be obtained by taking the difference of the conditional mean with the variable fixed at one and zero.

It might be of interest to test for hurdle effects. The hurdle model is similar to the zero inflation model in that a model without hurdle effects is not nested within the hurdle model; setting $\boldsymbol{\gamma} = \mathbf{0}$ produces either $F = \alpha$, a constant, or $F = ½$ if the constant term is also set to zero. Neither serves the purpose. Nor does forcing $\boldsymbol{\gamma} = \boldsymbol{\beta}$ in a model with $\mathbf{w}_i = \mathbf{x}_i$ and $F = \Lambda$ with a Poisson intensity equation, which might be intuitively appealing. A complementary log log model with

minus

$$\text{Prob}(y_i = 0 \mid \mathbf{w}_i) = \exp[-\exp(\mathbf{w}_i'\boldsymbol{\gamma})]$$

does produce the desired result if $\mathbf{w}_i = \mathbf{x}_i$. In this case, "hurdle effects" are absent if $\boldsymbol{\gamma} = \boldsymbol{\beta}$. The strategy in this case, then, would be a test of this restriction. But, this formulation is otherwise restrictive, first in the choice of variables and second in its unconventional functional form. The more general approach to this test would be the Vuong test used earlier to test the zero inflation model against the simpler Poisson or negative binomial model.

The hurdle model bears some similarity to the zero inflation model; however, the behavioral implications are different. The zero inflation model can usefully be viewed as a latent class model. The splitting probability defines a regime determination. In the hurdle model, the splitting equation represents a behavioral outcome on the same level as the intensity (count) equation. Both of these modifications substantially alter the Poisson formulation. First, note that the equality of the mean and variance of the distribution no longer follows; both modifications induce overdispersion. On the other hand, the overdispersion does not arise from heterogeneity; it arises from the nature of the process generating the zeros. As such, an interesting identification problem arises in this model. If the data do appear to be characterized by overdispersion, then it seems less than obvious whether it should be attributed to heterogeneity or to the regime splitting mechanism. Mullahy (1986) argues the point more strongly. He demonstrates that overdispersion will always induce excess zeros. As such, in a splitting model, we may misinterpret the excess zeros as due to the splitting process instead of the heterogeneity.

18-77

*Example 18.~~11~~ 12 Hurdle Model for Doctor Visits*

TB 18.19

The hurdle model is a natural specification for models of utilization of the health care system, and has been used in a number of studies. Table 18.~~18~~ 19 shows the parameter estimates for a hurdle model for doctor visits baswed on the entire pooled sample of 27,326 observations. The decomposition of the partial effects shows that the participation and intensity decisions each contribute substantively to the effects of Age, Income, and Public insurance. The value of the Vuong statistic is 51.16, strongly in favor of the hurdle model compared to the pooled Poisson model with no hurdle effects. The effect of the hurdle model on the partial effects ~~are~~ is shown in the last column where the results for the Poisson model are shown in parentheses. c19

**TABLE 18.~~18~~ 19 Estimated Hurdle Model for Doctor Visits**

| | Participation Equation | | Intensity Equation | | Total Partial Effect (Poisson Model) |
|---|---|---|---|---|---|
| | Parameter | Partial Effect | Parameter | Partial Effect | |
| Constant | -0.0598 | | 1.1203 | | |
| Age | 0.0221 | 0.0244 | 0.0113 | 0.0538 | 0.0782 ( 0.0625) |
| Income | 0.0725 | 0.0800 | -0.5152 | -2.4470 | -2.3670 (-1.8130) |
| Kids | | | -0.0842 | -0.4000 | -0.4000 (-0.4836) |
| Public | 0.2411 | 0.2663 | 0.1966 | 0.9338 | 1.2001 ( 0.9744) |
| Education | -0.0291 | -0.0321 | | | -0.0321 |
| Married | -0.0233 | -0.0258 | | | -0.0258 |
| Working | -0.3624 | -0.4003 | | | -0.4003 |

18-78

## 18.4.9 ENDOGENOUS VARIABLES AND ENDOGENOUS PARTICIPATION

As in other situations, one would expect to find endogenous variables in models for counts. For example, in the study on which we have relied for our examples of health care utilization, Riphahn, Wambach and Million (RWM, 2003), the authors were interested in the role of insurance (specifically the *Add-On* insurance) in the usage variable. One might expect the choice to buy insurance to be at least partly influenced by some of the same factors that motivate usage of the health care system. Insurance purchase might well be endogenous in a model such as the hurdle model in Example 18.12.

The Poisson model presents a complication for modeling endogeneity that arises in some other cases as well. For simplicity, consider a continuous variable, such as *Income*, to continue our ongoing example. A model of income determination and doctor visits might appear

Ail: OK to spell out "e.g."?

minus

$$Income = \mathbf{z}_i'\gamma + u_i,$$

$$\text{Prob}(DocVis_i = j|\mathbf{x}_i, Income_i) = \exp(-\lambda_i)\lambda_i^j/j!,\ \ \lambda_i = \exp(\mathbf{x}_i'\beta + \delta Income_i).$$

for example,

Endogeneity as we have analyzed it, e.g., in Chapter 8 and Sections 17.3.5 and 17.5.5, arises through correlation between the endogenous variable and the unobserved, omitted factors in the main equation. But, the Poisson model does not contain any unobservables. This is a major shortcoming of the specification as a "regression" model; all of the regression variation of the dependent variable arises through variation of the observables. There is no accommodation for unobserved heterogeneity or omitted factors. This is the compelling motivation for the negative binomial model or, in RWM's case, the Poisson-normal mixture model. [See Terza (2010, pp. 555-556) for discussion of this issue.] If the model is reformulated to accommodate heterogeneity, as in

$$\lambda_i = \exp(\mathbf{x}_i'\beta + \delta Income_i + \varepsilon_i),$$

then $Income_i$ will be endogenous if $u_i$ and $\varepsilon_i$ are correlated.

A bivariate normal model for $(u_i, \varepsilon_i)$ with zero means, variances $\sigma_u^2$ and $\sigma_\varepsilon^2$ and correlation $\rho$ provides a convenient (and the usual) platform to operationalize this idea. By projecting $\varepsilon_i$ on $u_i$, we have

$$\varepsilon_i = (\rho\sigma_\varepsilon/\sigma_u)u_i + v_i,$$

minus

where $v_i$ is normally distributed with mean zero and variance $\sigma_\varepsilon^2(1 - \rho^2)$. It will prove convenient to parameterize these based on the regression and the specific parameters as follows:

$$\begin{aligned}\varepsilon_i &= \rho\sigma_\varepsilon\,(Income_i - \mathbf{z}_i'\gamma)/\sigma_u + v_i,\\ &= \tau\,[(Income_i - \mathbf{z}_i'\gamma)/\sigma_u] + \theta w_i.\end{aligned}$$

where $w_i$ will be normally distributed with mean zero and variance one while $\tau = \rho\sigma_\varepsilon$ and $\theta^2 = \sigma_\varepsilon^2(1 - \rho^2)$. Then, combining terms,

$$\varepsilon_i = \tau\, u_i^* + \theta w_i.$$

With this parameterization, the conditional mean function in the Poisson regression model is

$$\lambda_i = \exp(\mathbf{x}_i'\beta + \delta Income_i + \tau u_i^* + \theta w_i).$$

18-79

The parameters to be estimated are $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\delta$, $\sigma_\varepsilon$, $\sigma_u$ and $\rho$. There are two ways to proceed. A two step method can be based on the fact that $\boldsymbol{\gamma}$ and $\sigma_u$ can be consistently estimated by linear regression of *Income* on $\mathbf{z}$. After this first step, we can compute values of $u_i^*$ and formulate the Poisson regression model in terms of

$$\hat{\lambda}_i(w_i) = \exp[\mathbf{x}_i'\boldsymbol{\beta} + \delta Income_i + \tau \hat{u}_i + \theta w_i].$$

The log likelihood to be maximized at the second step is

$$\ln L(\boldsymbol{\beta},\delta,\tau,\theta|\mathbf{w}) = \sum_{i=1}^{n} -\hat{\lambda}_i(w_i) + y_i \ln \hat{\lambda}_i(w_i) - \ln y_i!.$$

A remaining complication is that the unobserved heterogeneity, $w_i$ remains in the equation so it must be integrated out of the log likelihood function. The unconditional log likelihood function is obtained by integrating the standard normally distributed $w_i$ out of the conditional densities.

$$\ln L(\boldsymbol{\beta},\boldsymbol{\gamma},\tau,\theta) = \sum_{i=1}^{n} \ln\left\{\int_{-\infty}^{\infty} \left[\frac{\exp\left(-\hat{\lambda}_i(w_i)\right)\left(\hat{\lambda}_i(w_i)\right)^{y_i}}{y_i!}\right]\phi(w_i)dw_i\right\}..$$

The method of Butler and Moffitt or maximum simulated likelihood that we used to fit a probit model in Section 17.4.2 can be used to estimate $\boldsymbol{\beta}$, $\delta$, $\tau$ and $\theta$. Estimates of $\rho$ and $\sigma_\varepsilon$ can be deduced from the last two of these; $\sigma_\varepsilon^2 = \theta^2 + \tau^2$ and $\rho = \tau/\sigma_\varepsilon$. This is the control function method discussed in Section 17.3.5 and is also the "residual inclusion" method discussed by Terza, Basu and Rathouz (2008).

AU: This term was a KT on msp 18-12. Here also?

The full set of parameters can be estimated in a single step using **full information maximum likelihood**. To estimate all parameters simultaneously and efficiently, we would form the log likelihood from joint density of *DocVis* and *Income* as *P*(*DocVis*|*Income*)*f*(*Income*). Thus,

$$f(DocVis, Income) = \frac{\exp[-\lambda_i(w_i)][\lambda_i(w_i)]^{y_i}}{y_i!}\frac{1}{\sigma_u}\phi\left(\frac{Income - \mathbf{z}_i'\boldsymbol{\gamma}}{\sigma_u}\right)$$

$$\lambda_i(w_i) = \exp\left(\mathbf{x}_i'\boldsymbol{\beta} + \delta Income_i + \tau(Income_i - \mathbf{z}_i'\boldsymbol{\gamma})/\sigma_u + \theta w_i\right)$$

As before, the unobserved $w_i$ must be integrated out of the log likelihood function. Either quadrature or simulation can be used. The parameters to be estimated by maximizing the full log likelihood are $(\boldsymbol{\beta},\boldsymbol{\gamma},\delta,\sigma_u,\sigma_\varepsilon,\rho)$. The invariance principle has been used to simplify the estimation a bit by parameterizing the log likelihood function in terms of $\tau$ and $\theta$. Some additional simplification can also be obtained by using the Olsen (1978) [and Tobin (1958)] transformations, $\eta = 1/\sigma_u$ and $\alpha = (1/\sigma_u)\boldsymbol{\gamma}$.

An endogenous binary variable, such as *Public* or *AddOn* in our *DocVis* example is handled similarly, but is a bit simpler. The structural equations of the model are

$$
\begin{aligned}
T^* &= \mathbf{z}_i'\boldsymbol{\gamma} + u_i, \qquad && u \sim N[0,1],\\
T &= 1(T^* > 0), && \\
\lambda &= \exp(\mathbf{x}'\boldsymbol{\beta} + \delta T + \varepsilon) && \varepsilon \sim N[0,\sigma_\varepsilon^2],
\end{aligned}
$$

with $\text{Cov}(u,\varepsilon) = \rho\sigma_\varepsilon$. The endogeneity of $T$ is implied by a nonzero $\rho$. We use the bivariate normal result

$$u = (\rho/\sigma_\varepsilon)\varepsilon + v$$

where $v$ is normally distributed with mean zero and variance $1 - \rho^2$. Then, using our earlier results for the probit model (Section 17.2),

$$P(T|\varepsilon) = \Phi\left[(2T-1)\left(\frac{\mathbf{z}'\boldsymbol{\gamma} + (\rho/\sigma_\varepsilon)\varepsilon}{\sqrt{1-\rho^2}}\right)\right], T = 0,1$$

It will be convenient once again to write $\varepsilon = \sigma_\varepsilon w$ where $w \sim N[0,1]$. Making the substitution, we have

$$P(T|w) = \Phi\left[(2T-1)\left(\frac{\mathbf{z}'\boldsymbol{\gamma} + \rho w}{\sqrt{1-\rho^2}}\right)\right], T = 0,1.$$

The probability density function for $y|T,w$ is Poisson with $\lambda(w) = \exp(\mathbf{x}'\boldsymbol{\beta} + \delta T + \sigma_\varepsilon w)$. Combining terms,

$$P(y,T|w) = \frac{\exp[-\lambda(w)][\lambda(w)]^y}{y!}\Phi\left[(2T-1)\left(\frac{\mathbf{z}'\boldsymbol{\gamma} + \rho w}{\sqrt{1-\rho^2}}\right)\right].$$

This last result provides the terms that enter the log likelihood for $(\boldsymbol{\beta},\boldsymbol{\gamma},\delta,\rho,\sigma_\varepsilon)$. As before, the unobserved heterogeneity, $w$, must be integrated out of the log likelihood, so either the quadrature or simulation method discussed in Chapter 17 are used to obtain the parameter estimates. Note that this model may also be estimated in two steps, with $\boldsymbol{\gamma}$ obtained in the first step probit. The two step method will not be appreciably simpler, since the second term in the density must remain to identify $\rho$. The residual inclusion method is not fesible here since $T^*$ is not observed.

This same set of methods is used to allow for endogeneity of the participation equation in the hurdle model in Section 18.4.8. Mechanically, the hurdle model with endogenous participation is essentially the same as the endogenous binary variable. [See Greene (2005,2007).]

18-81

## 18.5 SUMMARY AND CONCLUSIONS

The analysis of individual decisions in microeconometrics is largely about discrete decisions such as whether to participate in an activity or not, whether to make a purchase or not, or what brand of product to buy. This chapter and Chapter 17 have developed the four essential models used in that type of analysis. Random utility, the binary choice model and regression-style modeling of probabilities developed in Chapter 17 are the three fundamental building blocks of discrete choice modeling. This chapter extended those tools into the three primary areas of choice modeling, unordered choice models, ordered choice models and models for counts. In each case, we developed a core modeling framework that provides the broad platform then developed a variety of extensions. and

In the analysis of unordered choice models, such as brand or location, the multinomial logit (MNL) model has provided the essential starting point. The MNL works well to provide a basic framework, but as a behavioral model in its own right it has some important shortcomings. Much of the recent research in this area has focused on relaxing these behavioral assumptions. The most recent research in this area, on the mixed logit model has produced broadly flexible functional forms that can match behavioral modeling to empirical specification and estimation.

The ordered choice model is a natural extension of the binary choice setting and also a convenient bridge between models of choice between two alternatives and more complex models of choice among multiple alternatives. We began this analysis with the ordered probit and logit model pioneered by Zavoina and McKelvey (1975). Recent developments of this model have produced the same sorts of extensions to panel data and modeling heterogeneity that we considered in Chapter 17 for binary choice. We also examined some multiple equation specifications.. For all its versatility, the familiar ordered choice models have an important shortcoming in the assumed constancy underlying preference behind the rating scale. The current work on differential item functioning, such as King et al. (2004) has produced significant progress on filling this gap in the theory.

Finally, we examined probability models for counts of events. Here, the Poisson regression model provides the broad framework for the analysis. The Poisson model has two shortcomings that have motivated the current stream of research. The functional form binds the mean of the random variable to its variance, producing an unrealistic regression specification. Second, the basic model has no component that accommodates unmeasured heterogeneity. (This second feature is what produces the first.) Current research has produced a rich variety of models for counts, such as two part behavioral models, that account for many different aspects of the decision making process and the mechanisms that generate the observed data.

18-82

AU: KTs with blue checkmarks were not bold KTs in chapter

## Key Terms and Concepts

Censored variable
Censoring
Choice based sample
Conditional logit model
Count data
Delta method
Deviance
Differential item functioning (DIF)
Event count
Exposure
Full information maximum likelihood (FIML)
Heterogeneity
Hurdle model
Identification through functional form
Independence from Irrelevant alternatives (IIA)
Lagrange multiplier test
Limited information
Loglinear model
Marginal effects
Method of simulated moments
Mixed logit model
Multinomial choice
Multinomial logit model
Multinomial probit model (MNP)
Negative binomial model
(NB1) Negbin 1 form
(NB2) Negbin 2 form
Negbin *P* model (NBP)
Nested logit model
Nonnested models
Ordered choice model
Overdispersion
Parallel regressions assumption
Poisson regression model
Random coefficients
Random parameters logit model (RPL)
Revealed preference data
Specification error
Stated choice experiment
Subjective well-being
Unordered choice model
Willingness to pay space
Zero inflated Poisson model (ZIP)

18-83

## Exercises

1. We are interested in the ordered probit model. Our data consist of 250 observations, of which the response are

| **y** | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| **n** | 50 | 40 | 45 | 80 | 35 |

Using the preceding data, obtain maximum likelihood estimates of the unknown parameters of the model. (Hint: Consider the probabilities as the unknown parameters.)

2. For the zero-inflated Poisson (ZIP) model in Section 18.4.8, we derived the conditional mean function, $E[y_i \,|\mathbf{x}_i,\mathbf{w}_i] = (1 - F_i)\lambda_i$.
   a. For the same model, now obtain $Var[y_i|\mathbf{x}_i,\mathbf{w}_i]$. Then, obtain $\tau_i = \text{Var}[y_i|\mathbf{x}_i,\mathbf{w}_i] \,/\, E[y_i|\mathbf{x}_i,\mathbf{w}_i]$. Does the zero inflation produce overdispersion? (That is, is the ratio greater than one?)
   b. Obtain the partial effect for a variable $z_i$ that appears in both $\mathbf{w}_i$ and $\mathbf{x}_i$.

3. Consider estimation of a Poisson regression model for $y_i|x_i$ . The data are truncated on the left—these are on-site observations at a recreasion site, so zeros do not appear in the data set. The data are censored on the right—any response greater than 5 is recorded as a 5. Construct the log-likelihood for a data set drawn under this sampling scheme.

## Applications

1. Appendix Table F18.1 provides Fair's (1978) *Redbook Magazine* survey on extramarital affairs. The variables in the data set are as follows:

*id* = an identification number,
*C* = constant, value = 1,
*yrb* = a constructed measure of time spent in extramarital affairs,
*v1* = a rating of the marriage, coded 1 to 5,
*v2* = age, in years, aggregated,
*v3* = number of years married,
*v4* = number of children, top coded at 5,
*v5* = religiosity, 1 to 4, 1 = not, 4 = very,
*v6* = education, coded 9, 12, 14, 16, 17, 20,
*v7* = occupation,
*v8* = husband's occupation,

and three other variables that are not used. The sample contains a survey of 6,366 married women. For this exercise,we will analyze, first, the binary variable $A = 1$ if $yrb > 0$, 0 otherwise. The regressors of interest are *v1* to *v8*; however, not necessarily all of them belong in your model. Use these data to build a binary choice model for *A*. Report all computed results for the model. Compute the marginal effects for the variables you choose. Compare the results you obtain for a probit model to those for a logit model. Are there any substantial differences in the results for the two models?

18-89

END 18

2. Continuing the analysis of the first application, we now consider the self-reported rating, y1. This is a natural candidate for an ordered choice model, because the simple four-item coding is a censored version of what would be a continuous scale on some subjective satisfaction variable. Analyze this variable using an ordered probit model. What variables appear to explain the response to this survey question? (Note, the variable is coded 1, 2, 3, 4, 5. Some programs accept data for ordered choice modeling in this form, e.g., *Stata*, while others require the variable to be coded 0, 1, 2, 3, 4, e.g., *LIMDEP*. Be sure to determine which is appropriate for the program you are using and transform the data if necessary.) Can you obtain the partial effects for your model? Report them as well. What do they suggest about the impact of the different independent variables on the reported ratings?

for example

Au: OK to spell out "e.g." 2x?

3. Several applications in the preceding chapters using the German health care data have examined the variable DocVis, the reported number of visits to the doctor. The data are described in Appendix Table F7.1. A second count variable in that data set that we have not examined is HospVis, the number of visits to hospital. For this application, we will examine this variable. To begin, we treat the full sample (27,326) observations as a cross section.
a. Begin by fitting a Poisson regression model to this variable. The exogenous variables are listed in the Appendix Table F7.1. Determine an appropriate specification for the right-hand side of your model. Report the regression results and the partial effects.
b. Estimate the model using ordinary least squares and compare your least squares results to the partial effects you computed in part a. What do you find?
c. Is there evidence of overdispersion in the data? Test for overdispersion. Now, reestimate the model using a negative binomial specification. What is the result? Do your results change? Use a likelihood ratio test to test the hypothesis of the negative binomial model against the Poisson.

4. The GSOEP data are an unbalanced panel, with 7,293 groups. Continue your analysis in Application 3 by fitting the Poisson model with fixed and with random effects and compare your results. (Recall, like the linear model, the Poisson fixed effects model may not contain any time invariant variables.) How do the panel data results compare to the pooled results?

5. Appendix Table F18.2 contains data on ship accidents reported in McCullagh and Nelder (1983). The data set contains 40 observations on the number of incidents of wave damage for oceangoing ships. Regressors include "aggregate months of service," and three sets of dummy variables, Type (1, . . . , 5), operation period (1960–1974 or 1975–1979), and construction period (1960–1964, 1965–1969, or 1970–1974). There are six missing values on the dependent variable, leaving 34 usable observations.
a. Fit a Poisson model for these data, using the log of service months, four types of dummy variables, two construction period variables, and one operation period dummy variable. Report your results.
b. The authors note that the rate of accidents is supposed to be per period, but the exposure (aggregate months) differs by ship. Reestimate your model constraining the coefficient on log of service months to equal one.
c. The authors take overdispersion as a given in these data. Do you find evidence of overdispersion? Show your results.