# 19 Count Data and Related Models

## 19.1 Why Count Data Models?

A **count variable** is a variable that takes on nonnegative integer values. Many variables that we would like to explain in terms of covariates come as counts. A few examples include the number of times someone is arrested during a given year, number of emergency room drug episodes during a given week, number of cigarettes smoked per day, and number of patents applied for by a firm during a year. These examples have two important characteristics in common: there is no natural a priori upper bound, and the outcome will be zero for at least some members of the population. Other count variables do have an upper bound. For example, for the number of children in a family who are high school graduates, the upper bound is number of children in the family.

If $y$ is the count variable and $\mathbf{x}$ is a vector of explanatory variables, we are often interested in the population regression, $\mathrm{E}(y \mid \mathbf{x})$. Throughout this book we have discussed various models for conditional expectations, and we have discussed different methods of estimation. The most straightforward approach is a linear model, $\mathrm{E}(y \mid \mathbf{x}) = \mathbf{x}\boldsymbol{\beta}$, estimated by OLS. For count data, linear models have shortcomings very similar to those for binary responses or corner solution responses: because $y \geq 0$, we know that $\mathrm{E}(y \mid \mathbf{x})$ should be nonnegative for all $\mathbf{x}$. If $\hat{\boldsymbol{\beta}}$ is the OLS estimator, there usually will be values of $\mathbf{x}$ such that $\mathbf{x}\hat{\boldsymbol{\beta}} < 0$—so that the predicted value of $y$ is negative.

For strictly positive variables, we often use the natural log transformation, $\log(y)$, and use a linear model. This approach is not possible in interesting count data applications, where $y$ takes on the value zero for a nontrivial fraction of the population. Transformations could be applied that are defined for all $y \geq 0$—for example, $\log(1 + y)$—but $\log(1 + y)$ itself is nonnegative, and it is not obvious how to recover $\mathrm{E}(y \mid \mathbf{x})$ from a linear model for $\mathrm{E}[\log(1 + y) \mid \mathbf{x}]$. With count data, it is better to model $\mathrm{E}(y \mid \mathbf{x})$ directly and to choose functional forms that ensure positivity for any value of $\mathbf{x}$ and any parameter values. When $y$ has no upper bound, the most popular of these is the exponential function, $\mathrm{E}(y \mid \mathbf{x}) = \exp(\mathbf{x}\boldsymbol{\beta})$.

In Chapter 12 we discussed nonlinear least squares (NLS) as a general method for estimating nonlinear models of conditional means. NLS can certainly be applied to count data models, but it is not ideal: NLS is relatively inefficient unless $\mathrm{Var}(y \mid \mathbf{x})$ is constant (see Chapter 12), and all of the standard distributions for count data imply heteroskedasticity.

In Section 19.2 we discuss the most popular model for count data, the Poisson regression model. As we will see, the Poisson regression model has some nice features. First, if $y$ given $\mathbf{x}$ has a Poisson distribution—which used to be the maintained assumption in count data contexts—then the conditional maximum likelihood estimators are fully efficient. Second, the Poisson assumption turns out to be unnecessary for consistent estimation of the conditional mean parameters. As we will see in Section 19.2, the Poisson *quasi*–maximum likelihood estimator is fully robust to distributional misspecification. It also maintains certain efficiency properties even when the distribution is not Poisson.

In Section 19.3 we discuss other count data models, and in Section 19.4 we cover quasi-MLEs for other nonnegative response variables. In Section 19.5 we cover multiplicative panel data models, which are motivated by unobserved effects count data models but can also be used for other nonnegative responses.

## 19.2 Poisson Regression Models with Cross Section Data

In Chapter 13 we used the basic Poisson regression model to illustrate maximum likelihood estimation. Here, we study Poisson regression in much more detail, emphasizing the properties of the estimator when the Poisson distributional assumption is incorrect.

### 19.2.1 Assumptions Used for Poisson Regression

The basic Poisson regression model assumes that $y$ given $\mathbf{x} \equiv (x_1, \ldots, x_K)$ has a Poisson distribution, as in El Sayyad (1973) and Maddala (1983, Section 2.15). The density of $y$ given $\mathbf{x}$ under the Poisson assumption is completely determined by the conditional mean $\mu(\mathbf{x}) \equiv \mathrm{E}(y \mid \mathbf{x})$:

$$f(y \mid \mathbf{x}) = \exp[-\mu(\mathbf{x})][\mu(\mathbf{x})]^y / y!, \qquad y = 0, 1, \ldots \tag{19.1}$$

where $y!$ is $y$ factorial. Given a parametric model for $\mu(\mathbf{x})$ [such as $\mu(\mathbf{x}) = \exp(\mathbf{x}\boldsymbol{\beta})$] and a random sample $\{(\mathbf{x}_i, y_i)\colon i = 1, 2, \ldots, N\}$ on $(\mathbf{x}, y)$, it is fairly straightforward to obtain the conditional MLEs of the parameters. The statistical properties then follow from our treatment of CMLE in Chapter 13.

It has long been recognized that the Poisson distributional assumption imposes restrictions on the conditional moments of $y$ that are often violated in applications. The most important of these is equality of the conditional variance and mean:

$$\mathrm{Var}(y \mid \mathbf{x}) = \mathrm{E}(y \mid \mathbf{x}) \tag{19.2}$$

The variance-mean equality has been rejected in numerous applications, and later we show that assumption (19.2) is violated for fairly simple departures from the Poisson model. Importantly, whether or not assumption (19.2) holds has implications for how we carry out statistical inference. In fact, as we will see, it is assumption (19.2), not the Poisson assumption per se, that is important for large-sample inference; this point will become clear in Section 19.2.2. In what follows we refer to assumption (19.2) as the **Poisson variance assumption**.

A weaker assumption allows the variance-mean ratio to be any positive constant:

$$\mathrm{Var}(y \mid \mathbf{x}) = \sigma^2 \mathrm{E}(y \mid \mathbf{x}) \tag{19.3}$$

where $\sigma^2 > 0$ is the variance-mean ratio. This assumption is used in the **generalized linear models (GLM)** literature, and so we will refer to assumption (19.3) as the **Poisson GLM variance assumption**. The GLM literature is concerned with quasi-maximum likelihood estimation of a class of nonlinear models that contains Poisson regression as a special case. We do not need to introduce the full GLM apparatus and terminology to analyze Poisson regression. See McCullagh and Nelder (1989).

The case $\sigma^2 > 1$ is empirically relevant because it implies that the variance is greater than the mean; this situation is called **overdispersion** (relative to the Poisson case). One distribution for $y$ given $\mathbf{x}$ where assumption (19.3) holds with overdispersion is what Cameron and Trivedi (1986) call NegBin I—a particular parameterization of the negative binomial distribution. When $\sigma^2 < 1$ we say there is **underdispersion**. Underdispersion is less common than overdispersion, but underdispersion has been found in some applications.

There are plenty of count distributions for which assumption (19.3) does not hold—for example, the NegBin II model in Cameron and Trivedi (1986). Therefore, we are often interested in estimating the conditional mean parameters without specifying the conditional variance. As we will see, Poisson regression turns out to be well suited for this purpose.

Given a parametric model $m(\mathbf{x}, \boldsymbol{\beta})$ for $\mu(\mathbf{x})$, where $\boldsymbol{\beta}$ is a $P \times 1$ vector of parameters, the log likelihood for observation $i$ is

$$\ell_i(\boldsymbol{\beta}) = y_i \log[m(\mathbf{x}_i, \boldsymbol{\beta})] - m(\mathbf{x}_i, \boldsymbol{\beta}) \tag{19.4}$$

where we drop the term $\log(y_i!)$ because it does not depend on the parameters $\boldsymbol{\beta}$ (for computational reasons dropping this term is a good idea in practice, too, as $y_i!$ gets very large for even moderate $y_i$). We let $\mathscr{B} \subset \mathbb{R}^P$ denote the parameter space, which is needed for the theoretical development but is practically unimportant in most cases.

The most common mean function in applications is the exponential:

$$m(\mathbf{x}, \boldsymbol{\beta}) = \exp(\mathbf{x}\boldsymbol{\beta}) \tag{19.5}$$

where $\mathbf{x}$ is $1 \times K$ and contains unity as its first element, and $\boldsymbol{\beta}$ is $K \times 1$. Under assumption (19.5) the log likelihood is $\ell_i(\boldsymbol{\beta}) = y_i \mathbf{x}_i \boldsymbol{\beta} - \exp(\mathbf{x}_i \boldsymbol{\beta})$. The parameters in model (19.5) are easy to interpret. If $x_j$ is continuous, then

$$\frac{\partial \mathrm{E}(y \mid \mathbf{x})}{\partial x_j} = \exp(\mathbf{x}\boldsymbol{\beta})\beta_j$$

and so

$$\beta_j = \frac{\partial \mathrm{E}(y \mid \mathbf{x})}{\partial x_j} \cdot \frac{1}{\mathrm{E}(y \mid \mathbf{x})} = \frac{\partial \log[\mathrm{E}(y \mid \mathbf{x})]}{\partial x_j}$$

Therefore, $100\beta_j$ is the semielasticity of $\mathrm{E}(y \mid \mathbf{x})$ with respect to $x_j$: for small changes $\Delta x_j$, the percentage change in $\mathrm{E}(y \mid \mathbf{x})$ is roughly $(100\beta_j)\Delta x_j$. If we replace $x_j$ with $\log(x_j)$, $\beta_j$ is the elasticity of $\mathrm{E}(y \mid \mathbf{x})$ with respect to $x_j$. Using assumption (19.5) as the model for $\mathrm{E}(y \mid \mathbf{x})$ is analogous to using $\log(y)$ as the dependent variable in linear regression analysis.

Quadratic terms can be added with no additional effort, except in interpreting the parameters. In what follows, we will write the exponential function as in assumption (19.5), leaving transformations of $\mathbf{x}$—such as logs, quadratics, interaction terms, and so on—implicit. See Wooldridge (1997c) for a discussion of other functional forms.

### 19.2.2 Consistency of the Poisson QMLE

Once we have specified a conditional mean function, we are interested in cases where, other than the conditional mean, the Poisson distribution can be arbitrarily misspecified (subject to regularity conditions). When $y_i$ given $\mathbf{x}_i$ does *not* have a Poisson distribution, we call the estimator $\hat{\boldsymbol{\beta}}$ that solves

$$\max_{\boldsymbol{\beta} \in \mathscr{B}} \sum_{i=1}^{N} \ell_i(\boldsymbol{\beta}) \tag{19.6}$$

the **Poisson quasi–maximum likelihood estimator (QMLE)**. A careful discussion of the consistency of the Poisson QMLE requires introduction of the true value of the parameter, as in Chapters 12 and 13. That is, we assume that for some value $\boldsymbol{\beta}_{\mathrm{o}}$ in the parameter space $\mathscr{B}$,

$$\mathrm{E}(y \mid \mathbf{x}) = m(\mathbf{x}, \boldsymbol{\beta}_{\mathrm{o}}) \tag{19.7}$$

To prove consistency of the Poisson QMLE under assumption (19.5), the key is to show that $\boldsymbol{\beta}_{\mathrm{o}}$ is the unique solution to

$$\max_{\boldsymbol{\beta} \in \mathscr{B}} \mathrm{E}[\ell_i(\boldsymbol{\beta})] \tag{19.8}$$

Then, under the regularity conditions listed in Theorem 12.2, it follows from this theorem that the solution to equation (19.6) is weakly consistent for $\boldsymbol{\beta}_{\mathrm{o}}$.

Wooldridge (1997c) provides a simple proof that $\boldsymbol{\beta}_{\mathrm{o}}$ is a solution to equation (19.8) when assumption (19.7) holds (see also Problem 19.1). It also follows from the general results on quasi-MLE in the **linear exponential family (LEF)** by Gourieroux, Monfort, and Trognon (1984a) (hereafter, GMT, 1984a). Uniqueness of $\boldsymbol{\beta}_{\mathrm{o}}$ must be assumed separately, as it depends on the distribution of $\mathbf{x}_i$. That is, in addition to assumption (19.7), identification of $\boldsymbol{\beta}_{\mathrm{o}}$ requires some restrictions on the distribution of explanatory variables, and these depend on the nature of the regression function $m$. In the linear regression case, we require full rank of $\mathrm{E}(\mathbf{x}_i'\mathbf{x}_i)$. For Poisson QMLE with an exponential regression function $\exp(\mathbf{x}\boldsymbol{\beta})$, it can be shown that multiple solutions to equation (19.8) exist whenever there is perfect multicollinearity in $\mathbf{x}_i$, just as in the linear regression case. If we rule out perfect multicollinearity, we can usually conclude that $\boldsymbol{\beta}_{\mathrm{o}}$ is identified under assumption (19.7).

It is important to remember that consistency of the Poisson QMLE does not require any additional assumptions concerning the distribution of $y_i$ given $\mathbf{x}_i$. In particular, $\mathrm{Var}(y_i \mid \mathbf{x}_i)$ can be virtually anything (subject to regularity conditions needed to apply the results of Chapter 12).

### 19.2.3 Asymptotic Normality of the Poisson QMLE

If the Poission QMLE is consistent for $\boldsymbol{\beta}_{\mathrm{o}}$ without any assumptions beyond (19.7), why did we introduce assumptions (19.2) and (19.3)? It turns out that whether these assumptions hold determines which asymptotic variance matrix estimators and inference procedures are valid, as we now show.

The asymptotic normality of the Poisson QMLE follows from Theorem 12.3. The result is

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\mathrm{o}}) \xrightarrow{d} \mathrm{Normal}(0, \mathbf{A}_{\mathrm{o}}^{-1}\mathbf{B}_{\mathrm{o}}\mathbf{A}_{\mathrm{o}}^{-1}) \tag{19.9}$$

where

$$\mathbf{A}_{\mathrm{o}} \equiv \mathrm{E}[-\mathbf{H}_i(\boldsymbol{\beta}_{\mathrm{o}})] \tag{19.10}$$

and

$$\mathbf{B}_{\mathrm{o}} \equiv \mathrm{E}[\mathbf{s}_i(\boldsymbol{\beta}_{\mathrm{o}})\mathbf{s}_i(\boldsymbol{\beta}_{\mathrm{o}})'] = \mathrm{Var}[\mathbf{s}_i(\boldsymbol{\beta}_{\mathrm{o}})] \tag{19.11}$$

where we define $\mathbf{A}_{\mathrm{o}}$ in terms of minus the Hessian because the Poisson QMLE solves a maximization rather than a minimization problem. Taking the gradient of equation (19.4) and transposing gives the score for observation $i$ as

$$\mathbf{s}_i(\boldsymbol{\beta}) = \nabla_{\beta} m(\mathbf{x}_i, \boldsymbol{\beta})'[y_i - m(\mathbf{x}_i, \boldsymbol{\beta})]/m(\mathbf{x}_i, \boldsymbol{\beta}) \tag{19.12}$$

It is easily seen that, under assumption (19.7), $\mathbf{s}_i(\boldsymbol{\beta}_{\mathrm{o}})$ has a zero mean conditional on $\mathbf{x}_i$. The Hessian is more complicated but, under assumption (19.7), it can be shown that

$$-\mathrm{E}[\mathbf{H}_i(\boldsymbol{\beta}_{\mathrm{o}}) \mid \mathbf{x}_i] = \nabla_\beta m(\mathbf{x}_i, \boldsymbol{\beta}_{\mathrm{o}})' \nabla_\beta m(\mathbf{x}_i, \boldsymbol{\beta}_{\mathrm{o}}) / m(\mathbf{x}_i, \boldsymbol{\beta}_{\mathrm{o}}) \tag{19.13}$$

Then $\mathbf{A}_{\mathrm{o}}$ is the expected value of this expression (over the distribution of $\mathbf{x}_i$). A fully robust asymptotic variance matrix estimator for $\hat{\boldsymbol{\beta}}$ follows from equation (12.49):

$$\left(\sum_{i=1}^{N} \hat{\mathbf{A}}_i\right)^{-1} \left(\sum_{i=1}^{N} \hat{\mathbf{s}}_i \hat{\mathbf{s}}_i'\right) \left(\sum_{i=1}^{N} \hat{\mathbf{A}}_i\right)^{-1} \tag{19.14}$$

where $\hat{\mathbf{s}}_i$ is obtained from equation (19.12) with $\hat{\boldsymbol{\beta}}$ in place of $\boldsymbol{\beta}$, and $\hat{\mathbf{A}}_i$ is the right-hand side of equation (19.13) with $\hat{\boldsymbol{\beta}}$ in place of $\boldsymbol{\beta}_{\mathrm{o}}$. This is the fully robust variance matrix estimator in the sense that it requires only assumption (19.7) and the regularity conditions from Chapter 12.

The asymptotic variance of $\hat{\boldsymbol{\beta}}$ simplifies under the GLM assumption (19.3). Maintaining assumption (19.3) (where $\sigma_{\mathrm{o}}^2$ now denotes the true value of $\sigma^2$) and defining $u_i \equiv y_i - m(\mathbf{x}_i, \boldsymbol{\beta}_{\mathrm{o}})$, the law of iterated expectations implies that

$$\begin{aligned}\mathbf{B}_{\mathrm{o}} &= \mathrm{E}[u_i^2 \nabla_\beta m_i(\boldsymbol{\beta}_{\mathrm{o}})' \nabla_\beta m_i(\boldsymbol{\beta}_{\mathrm{o}}) / \{m_i(\boldsymbol{\beta}_{\mathrm{o}})\}^2] \\ &= \mathrm{E}[\mathrm{E}(u_i^2 \mid \mathbf{x}_i) \nabla_\beta m_i(\boldsymbol{\beta}_{\mathrm{o}})' \nabla_\beta m_i(\boldsymbol{\beta}_{\mathrm{o}}) / \{m_i(\boldsymbol{\beta}_{\mathrm{o}})\}^2] = \sigma_{\mathrm{o}}^2 \mathbf{A}_{\mathrm{o}}\end{aligned}$$

since $\mathrm{E}(u_i^2 \mid \mathbf{x}_i) = \sigma_{\mathrm{o}}^2 m_i(\boldsymbol{\beta}_{\mathrm{o}})$ under assumptions (19.3) and (19.7). Therefore, $\mathbf{A}_{\mathrm{o}}^{-1} \mathbf{B}_{\mathrm{o}} \mathbf{A}_{\mathrm{o}}^{-1} = \sigma_{\mathrm{o}}^2 \mathbf{A}_{\mathrm{o}}^{-1}$, so we only need to estimate $\sigma_{\mathrm{o}}^2$ in addition to obtaining $\hat{\mathbf{A}}$. A consistent estimator of $\sigma_{\mathrm{o}}^2$ is obtained from $\sigma_{\mathrm{o}}^2 = \mathrm{E}[u_i^2 / m_i(\boldsymbol{\beta}_{\mathrm{o}})]$, which follows from assumption (19.3) and iterated expectations. The usual analogy principle argument gives the estimator

$$\hat{\sigma}^2 = N^{-1} \sum_{i=1}^{N} \hat{u}_i^2 / \hat{m}_i = N^{-1} \sum_{i=1}^{N} (\hat{u}_i / \sqrt{\hat{m}_i})^2 \tag{19.15}$$

The last representation shows that $\hat{\sigma}^2$ is simply the average sum of squared weighted residuals, where the weights are the inverse of the estimated nominal standard deviations. (In the GLM literature, the weighted residuals $\tilde{u}_i \equiv \hat{u}_i / \sqrt{\hat{m}_i}$ are sometimes called the **Pearson residuals**. In earlier chapters we also called them standardized residuals.) In the GLM literature, a degrees-of-freedom adjustment is usually made by replacing $N^{-1}$ with $(N - P)^{-1}$ in equation (19.15).

Given $\hat{\sigma}^2$ and $\hat{\mathbf{A}}$, it is straightforward to obtain an estimate of $\mathrm{Avar}(\hat{\boldsymbol{\beta}})$ under assumption (19.3). In fact, we can write

$$\text{Av}\hat{\text{a}}\text{r}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2 \hat{\mathbf{A}}^{-1}/N = \hat{\sigma}^2 \left( \sum_{i=1}^{N} \nabla_\beta \hat{m}_i' \nabla_\beta \hat{m}_i / \hat{m}_i \right)^{-1} \tag{19.16}$$

Note that the matrix is always positive definite when the inverse exists, so it produces well-defined standard errors (given, as usual, by the square roots of the diagonal elements). We call these the **GLM standard errors**.

If the Poisson variance assumption (19.2) holds, things are even easier because $\sigma^2$ is known to be unity; the estimated asymptotic variance of $\hat{\boldsymbol{\beta}}$ is given in equation (19.16) but with $\hat{\sigma}^2 \equiv 1$. The same estimator can be derived from the MLE theory in Chapter 13 as the inverse of the estimated information matrix (conditional on the $\mathbf{x}_i$); see Section 13.5.2.

Under assumption (19.3) in the case of overdispersion ($\sigma^2 > 1$), standard errors of the $\hat{\beta}_j$ obtained from equation (19.16) with $\hat{\sigma}^2 = 1$ will systematically underestimate the asymptotic standard deviations, sometimes by a large factor. For example, if $\sigma^2 = 2$, the correct GLM standard errors are, in the limit, 41 percent larger than the incorrect, nominal Poisson standard errors. It is common to see very significant coefficients reported for Poisson regressions—a recent example is Model (1993)—but we must interpret the standard errors with caution when they are obtained under assumption (19.2). The GLM standard errors are easily obtained by multiplying the Poisson standard errors by $\hat{\sigma} \equiv \sqrt{\hat{\sigma}^2}$. The most robust standard errors are obtained from expression (19.14), as these are valid under *any* conditional variance assumption. In practice, it is a good idea to report the fully robust standard errors along with the GLM standard errors and $\hat{\sigma}$.

If $y$ given $\mathbf{x}$ has a Poisson distribution, it follows from the general efficiency of the conditional MLE—see Section 14.5.2—that the Poisson QMLE is fully efficient in the class of estimators that ignores information on the marginal distribution of $\mathbf{x}$.

A nice property of the Poisson QMLE is that it retains some efficiency for certain departures from the Poisson assumption. The efficiency results of GMT (1984a) can be applied here: if the GLM assumption (19.3) holds for some $\sigma^2 > 0$, the Poisson QMLE is efficient in the class of all QMLEs in the linear exponential family of distributions. In particular, the Poisson QMLE is more efficient than the nonlinear least squares estimator, as well as many other QMLEs in the LEF, some of which we cover in Sections 19.3 and 19.4.

Wooldridge (1997c) gives an example of Poisson regression to an economic model of crime, where the response variable is number of arrests of a young man living in California during 1986. Wooldridge finds overdispersion: $\hat{\sigma}$ is either 1.228 or 1.172, depending on the functional form for the conditional mean. The following example shows that underdispersion is possible.

**Table 19.1**
OLS and Poisson Estimates of a Fertility Equation

| Dependent Variable: *children* | | |
|---|---|---|
| Independent Variable | Linear (OLS) | Exponential (Poisson QMLE) |
| *educ* | −.0644<br>(.0063) | −.0217<br>(.0025) |
| *age* | .272<br>(.017) | .337<br>(.009) |
| *age*$^2$ | −.0019<br>(.0003) | −.0041<br>(.0001) |
| *evermarr* | .682<br>(.052) | .315<br>(.021) |
| *urban* | −.228<br>(.046) | −.086<br>(.019) |
| *electric* | −.262<br>(.076) | −.121<br>(.034) |
| *tv* | −.250<br>(.090) | −.145<br>(.041) |
| *constant* | −3.394<br>(.245) | −5.375<br>(.141) |
| Log-likelihood value | — | −6,497.060 |
| *R*-squared | .590 | .598 |
| $\hat{\sigma}$ | 1.424 | .867 |

*Example 19.1 (Effects of Education on Fertility):* We use the data in FERTIL2.RAW to estimate the effects of education on women's fertility in Botswana. The response variable, *children*, is number of living children. We use a standard exponential regression function, and the explanatory variables are years of schooling (*educ*), a quadratic in age, and binary indicators for ever married, living in an urban area, having electricity, and owning a television. The results are given in Table 19.1. A linear regression model is also included, with the usual OLS standard errors. For Poisson regression, the standard errors are the GLM standard errors. A total of 4,358 observations are used.

As expected, the signs of the coefficients agree in the linear and exponential models, but their interpretations differ. For Poisson regression, the coefficient on *educ* implies that another year of education reduces expected number of children by about 2.2 percent, and the effect is very statistically significant. The linear model estimate implies that another year of education reduces expected number of children by about .064. (So, if 100 women get another year of education, we estimate they will have about six fewer children.)

The estimate of $\sigma$ in the Poisson regression implies underdispersion: the variance is less than the mean. (Incidentally, the $\hat{\sigma}$'s for the linear and Poisson models are not comparable.) One implication is that the GLM standard errors are actually less than the corresponding Poisson MLE standard errors.

For the linear model, the $R$-squared is the usual one. For the exponential model, the $R$-squared is computed as the squared correlation coefficient between $children_i$ and $\widehat{children}_i = \exp(\mathbf{x}_i\hat{\boldsymbol{\beta}})$. The exponential regression function fits slightly better.

### 19.2.4 Hypothesis Testing

Classical hypothesis testing is fairly straightforward in a QMLE setting. Testing hypotheses about individual parameters is easily carried out using asymptotic $t$ statistics after computing the appropriate standard error, as we discussed in Section 19.2.3. Multiple hypotheses tests can be carried out using the Wald, quasi–likelihood ratio, or score test. We covered these generally in Sections 12.6 and 13.6, and they apply immediately to the Poisson QMLE.

The Wald statistic for testing nonlinear hypotheses is computed as in equation (12.63), where $\hat{\mathbf{V}}$ is chosen appropriately depending on the degree of robustness desired, with expression (19.14) being the most robust. The Wald statistic is convenient for testing multiple exclusion restrictions in a robust fashion.

When the GLM assumption (19.3) holds, the quasi–likelihood ratio statistic can be used. Let $\breve{\boldsymbol{\beta}}$ be the restricted estimator, where $Q$ restrictions of the form $\mathbf{c}(\breve{\boldsymbol{\beta}}) = \mathbf{0}$ have been imposed. Let $\hat{\boldsymbol{\beta}}$ be the unrestricted QMLE. Let $\mathscr{L}(\boldsymbol{\beta})$ be the quasi–log likelihood for the sample of size $N$, given in expression (19.6). Let $\hat{\sigma}^2$ be given in equation (19.15) (with or without the degrees-of-freedom adjustment), where the $\hat{u}_i$ are the residuals from the unconstrained maximization. The QLR statistic,

$$QLR \equiv 2[\mathscr{L}(\hat{\boldsymbol{\beta}}) - \mathscr{L}(\breve{\boldsymbol{\beta}})]/\hat{\sigma}^2 \tag{19.17}$$

converges in distribution to $\chi^2_Q$ under $\mathrm{H}_0$, under the conditions laid out in Section 12.6.3. The division of the usual likelihood ratio statistic by $\hat{\sigma}^2$ provides for some degree of robustness. If we set $\hat{\sigma}^2 = 1$, we obtain the usual LR statistic, which is valid only under assumption (19.2). There is no usable quasi-LR statistic when the GLM assumption (19.3) does not hold.

The score test can also be used to test multiple hypotheses. In this case we estimate only the restricted model. Partition $\boldsymbol{\beta}$ as $(\boldsymbol{\alpha}', \boldsymbol{\gamma}')'$, where $\boldsymbol{\alpha}$ is $P_1 \times 1$ and $\boldsymbol{\gamma}$ is $P_2 \times 1$, and assume that the null hypothesis is

$$\mathrm{H}_0\colon \boldsymbol{\gamma}_{\mathrm{o}} = \bar{\boldsymbol{\gamma}} \tag{19.18}$$

where $\bar{\boldsymbol{\gamma}}$ is a $P_2 \times 1$ vector of specified constants (often, $\bar{\boldsymbol{\gamma}} = \mathbf{0}$). Let $\breve{\boldsymbol{\beta}}$ be the estimator of $\boldsymbol{\beta}$ obtained under the restriction $\boldsymbol{\gamma} = \bar{\boldsymbol{\gamma}}$ [so $\breve{\boldsymbol{\beta}} \equiv (\breve{\boldsymbol{\alpha}}', \bar{\boldsymbol{\gamma}}')'$], and define quantities under the restricted estimation as $\check{m}_i \equiv m(\mathbf{x}_i, \check{\boldsymbol{\beta}})$, $\check{u}_i \equiv y_i - \check{m}_i$, and $\nabla_\beta \check{m}_i \equiv (\nabla_\alpha \check{m}_i, \nabla_\gamma \check{m}_i) \equiv \nabla_\beta m(\mathbf{x}_i, \check{\boldsymbol{\beta}})$. Now weight the residuals and gradient by the inverse of nominal Poisson standard deviation, estimated under the null, $1/\sqrt{\check{m}_i}$:

$$\tilde{u}_i \equiv \check{u}_i/\sqrt{\check{m}_i}, \qquad \nabla_\beta \tilde{m}_i \equiv \nabla_\beta \check{m}_i/\sqrt{\check{m}_i} \tag{19.19}$$

so that the $\tilde{u}_i$ here are the Pearson residuals obtained under the null. A form of the score statistic that is valid under the GLM assumption (19.3) [and therefore under assumption (19.2)] is $NR_u^2$ from the regression

$$\tilde{u}_i \text{ on } \nabla_\beta \tilde{m}_i, \qquad i = 1, 2, \ldots, N \tag{19.20}$$

where $R_u^2$ denotes the uncentered $R$-squared. Under $\mathrm{H}_0$ and assumption (19.3), $NR_u^2 \overset{a}{\sim} \chi^2_{P_2}$. This is identical to the score statistic in equation (12.68) but where we use $\tilde{\mathbf{B}} = \tilde{\sigma}^2 \tilde{\mathbf{A}}$, where the notation is self-explanatory. For more, see Wooldridge (1991a, 1997c).

Following our development for nonlinear regression in Section 12.6.2, it is easy to obtain a test that is completely robust to variance misspecification. Let $\tilde{\mathbf{r}}_i$ denote the $1 \times P_2$ residuals from the regression

$$\nabla_\gamma \tilde{m}_i \text{ on } \nabla_\alpha \tilde{m}_i \tag{19.21}$$

In other words, regress each element of the weighted gradient with respect to the restricted parameters on the weighted gradient with respect to the unrestricted parameters. The residuals are put into the $1 \times P_2$ vector $\tilde{\mathbf{r}}_i$. The robust score statistic is obtained as $N - \mathrm{SSR}$ from the regression

$$1 \text{ on } \tilde{u}_i \tilde{\mathbf{r}}_i, \qquad i = 1, 2, \ldots, N \tag{19.22}$$

where $\tilde{u}_i \tilde{\mathbf{r}}_i = (\tilde{u}_i \tilde{r}_{i1}, \tilde{u}_i \tilde{r}_{i2}, \ldots, \tilde{u}_i \tilde{r}_{iP_2})$ is a $1 \times P_2$ vector.

As an example, consider testing $\mathrm{H}_0$: $\boldsymbol{\gamma} = \mathbf{0}$ in the exponential model $\mathrm{E}(y \mid \mathbf{x}) = \exp(\mathbf{x}\boldsymbol{\beta}) = \exp(\mathbf{x}_1\boldsymbol{\alpha} + \mathbf{x}_2\boldsymbol{\gamma})$. Then $\nabla_\beta m(\mathbf{x}, \boldsymbol{\beta}) = \exp(\mathbf{x}\boldsymbol{\beta})\mathbf{x}$. Let $\check{\boldsymbol{\alpha}}$ be the Poisson QMLE obtained under $\boldsymbol{\gamma} = \mathbf{0}$, and define $\check{m}_i \equiv \exp(\mathbf{x}_{i1}\check{\boldsymbol{\alpha}})$, with $\check{u}_i$ the residuals. Now $\nabla_\alpha \check{m}_i = \exp(\mathbf{x}_{i1}\check{\boldsymbol{\alpha}})\mathbf{x}_{i1}$, $\nabla_\gamma \check{m}_i = \exp(\mathbf{x}_{i1}\check{\boldsymbol{\alpha}})\mathbf{x}_{i2}$, and $\nabla_\beta \tilde{m}_i = \check{m}_i \mathbf{x}_i/\sqrt{\check{m}_i} = \sqrt{\check{m}_i}\mathbf{x}_i$. Therefore, the test that is valid under the GLM variance assumption is $NR_u^2$ from the OLS regression $\tilde{u}_i$ on $\sqrt{\check{m}_i}\mathbf{x}_i$, where the $\tilde{u}_i$ are the weighted residuals. For the robust test, first obtain the $1 \times P_2$ residuals $\tilde{\mathbf{r}}_i$ from the regression $\sqrt{\check{m}_i}\mathbf{x}_{i2}$ on $\sqrt{\check{m}_i}\mathbf{x}_{i1}$; then obtain the statistic from regression (19.22).

### 19.2.5 Specification Testing

Various specification tests have been proposed in the context of Poisson regression. The two most important kinds are conditional mean specification tests and conditional variance specification tests. For conditional mean tests, we usually begin with a fairly simple model whose parameters are easy to interpret—such as $m(\mathbf{x}, \boldsymbol{\beta}) = \exp(\mathbf{x}\boldsymbol{\beta})$—and then test this against other alternatives. Once the set of conditioning variables $\mathbf{x}$ has been specified, all such tests are functional form tests.

A useful class of functional form tests can be obtained using the score principle, where the null model $m(\mathbf{x}, \boldsymbol{\beta})$ is nested in a more general model. Fully robust tests and less robust tests are obtained exactly as in the previous section. Wooldridge (1997c, Section 3.5) contains details and some examples, including an extension of RESET to exponential regression models.

Conditional variance tests are more difficult to compute, especially if we want to maintain only that the first two moments are correctly specified under $\mathrm{H}_0$. For example, it is very natural to test the GLM assumption (19.3) as a way of determining whether the Poisson QMLE is efficient in the class of estimators using only assumption (19.7). Cameron and Trivedi (1986) propose tests of the stronger assumption (19.2) and, in fact, take the null to be that the Poisson distribution is correct in its entirety. These tests are useful if we are interested in whether $y$ given $\mathbf{x}$ truly has a Poisson distribution. However, assumption (19.2) is not necessary for consistency or relative efficiency of the Poisson QMLE.

Wooldridge (1991b) proposes fully robust tests of conditional variances in the context of the linear exponential family, which contains Poisson regression as a special case. To test assumption (19.3), write $u_i = y_i - m(\mathbf{x}_i, \boldsymbol{\beta}_\mathrm{o})$ and note that, under assumptions (19.3) and (19.7), $u_i^2 - \sigma_\mathrm{o}^2 m(\mathbf{x}_i, \boldsymbol{\beta}_\mathrm{o})$ is uncorrelated with any function of $\mathbf{x}_i$. Let $\mathbf{h}(\mathbf{x}_i, \boldsymbol{\beta})$ be a $1 \times Q$ vector of functions of $\mathbf{x}_i$ and $\boldsymbol{\beta}$, and consider the alternative model

$$\mathrm{E}(u_i^2 \mid \mathbf{x}_i) = \sigma_\mathrm{o}^2 m(\mathbf{x}_i, \boldsymbol{\beta}_\mathrm{o}) + \mathbf{h}(\mathbf{x}_i, \boldsymbol{\beta}_\mathrm{o})\boldsymbol{\delta}_\mathrm{o} \tag{19.23}$$

For example, the elements of $\mathbf{h}(\mathbf{x}_i, \boldsymbol{\beta})$ can be powers of $m(\mathbf{x}_i, \boldsymbol{\beta})$. Popular choices are unity and $\{\mathrm{m}(\mathbf{x}_i, \boldsymbol{\beta})\}^2$. A test of $\mathrm{H}_0$: $\boldsymbol{\delta}_\mathrm{o} = \mathbf{0}$ is then a test of the GLM assumption. While there are several moment conditions that can be used, a fruitful one is to use the weighted residuals, as we did with the conditional mean tests. We base the test on

$$N^{-1} \sum_{i=1}^{N} (\hat{\mathbf{h}}_i / \hat{m}_i)' \{(\hat{u}_i^2 - \hat{\sigma}^2 \hat{m}_i)/\hat{m}_i\} = N^{-1} \sum_{i=1}^{N} \tilde{\mathbf{h}}_i' (\tilde{u}_i^2 - \hat{\sigma}^2) \tag{19.24}$$

where $\tilde{\mathbf{h}}_i = \hat{\mathbf{h}}_i / \hat{m}_i$ and $\tilde{u}_i = \hat{u}_i / \sqrt{\hat{m}_i}$. (Note that $\hat{\mathbf{h}}_i$ is weighted by $1/\hat{m}_i$, not $1/\sqrt{\hat{m}_i}$.) To turn this equation into a test statistic, we must confront the fact that its standardized limiting distribution depends on the limiting distributions of $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_\mathrm{o})$ and $\sqrt{N}(\hat{\sigma}^2 - \sigma_\mathrm{o}^2)$. To handle this problem, we use a trick suggested by Wooldridge

(1991b) that removes the dependence of the limiting distribution of the test statistic on that of $\sqrt{N}(\hat{\sigma}^2 - \sigma_o^2)$: replace $\tilde{\mathbf{h}}_i$ in equation (19.24) with its demeaned counterpart, $\tilde{\mathbf{r}}_i \equiv \tilde{\mathbf{h}}_i - \bar{\mathbf{h}}$, where $\bar{\mathbf{h}}$ is just the $1 \times Q$ vector of sample averages of each element of $\tilde{\mathbf{h}}_i$. There is an additional purging that then leads to a simple regression-based statistic. Let $\nabla_\beta \hat{m}_i$ be the unweighted gradient of the conditional mean function, evaluated at the Poisson QMLE $\hat{\boldsymbol{\beta}}$, and define $\nabla_\beta \tilde{m}_i \equiv \nabla_\beta \hat{m}_i / \sqrt{\hat{m}_i}$, as before. The following steps come from Wooldridge (1991b, Procedure 4.1):

1. Obtain $\hat{\sigma}^2$ as in equation (19.15) and $\hat{\mathbf{A}}$ as in equation (19.16), and define the $P \times Q$ matrix $\hat{\mathbf{J}} = \hat{\sigma}^2(N^{-1}\sum_{i=1}^N \nabla_\beta \hat{m}_i' \tilde{\mathbf{r}}_i / \hat{m}_i)$.
2. For each $i$, define the $1 \times Q$ vector

$$\hat{\mathbf{z}}_i \equiv (\tilde{u}_i^2 - \hat{\sigma}^2)\tilde{\mathbf{r}}_i - \hat{\mathbf{s}}_i' \hat{\mathbf{A}}^{-1} \hat{\mathbf{J}} \tag{19.25}$$

where $\hat{\mathbf{s}}_i \equiv \nabla_\beta \tilde{m}_i' \tilde{u}_i$ is the Poisson score for observation $i$.

3. Run the regression

$$1 \text{ on } \hat{\mathbf{z}}_i, \qquad i = 1, 2, \ldots, N \tag{19.26}$$

Under assumptions (19.3) and (19.7), $N - \text{SSR}$ from this regression is distributed asymptotically as $\chi_Q^2$.

The leading case occurs when $\hat{m}_i = \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}})$ and $\nabla_\beta \hat{m}_i = \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}})\mathbf{x}_i = \hat{m}_i \mathbf{x}_i$. The subtraction of $\hat{\mathbf{s}}_i' \hat{\mathbf{A}}^{-1} \hat{\mathbf{J}}$ in equation (19.25) is a simple way of handling the fact that the limiting distribution of $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o)$ affects the limiting distribution of the unadjusted statistic in equation (19.24). This particular adjustment ensures that the tests are just as efficient as any maximum-likelihood-based statistic if $\sigma_o^2 = 1$ and the Poisson assumption is correct. But this procedure is fully robust in the sense that only assumptions (19.3) and (19.7) are maintained under $\text{H}_0$. For further discussion the reader is referred to Wooldridge (1991b).

In practice, it is probably sufficient to choose the number of elements in $Q$ to be small. Setting $\hat{\mathbf{h}}_i = (1, \hat{m}_i^2)$, so that $\tilde{\mathbf{h}}_i = (1/\hat{m}_i, \hat{m}_i)$, is likely to produce a fairly powerful two-degrees-of-freedom test against a fairly broad class of alternatives.

The procedure is easily modified to test the more restrictive assumption (19.2). First, replace $\hat{\sigma}^2$ everywhere with unity. Second, there is no need to demean the auxiliary regressors $\tilde{\mathbf{h}}_i$ (so that now $\tilde{\mathbf{h}}_i$ can contain a constant); thus, wherever $\tilde{\mathbf{r}}_i$ appears, simply use $\tilde{\mathbf{h}}_i$. Everything else is the same. For the reasons discussed earlier, when the focus is on $\text{E}(y \mid \mathbf{x})$, we are more interested in testing assumption (19.3) than assumption (19.2).

## 19.3 Other Count Data Regression Models

### 19.3.1 Negative Binomial Regression Models

The Poisson regression model nominally maintains assumption (19.2) but retains some asymptotic efficiency under assumption (19.3). A popular alternative to the Poisson QMLE is full maximum likelihood analysis of the NegBin I model of Cameron and Trivedi (1986). NegBin I is a particular parameterization of the negative binomial distribution. An important restriction in the NegBin I model is that it implies assumption (19.3) with $\sigma^2 > 1$, so that there cannot be underdispersion. (We drop the "o" subscript in this section for notational simplicity.) Typically, NegBin I is parameterized through the mean parameters $\boldsymbol{\beta}$ and an additional parameter, $\eta^2 > 0$, where $\sigma^2 = 1 + \eta^2$. On the one hand, when $\boldsymbol{\beta}$ and $\eta^2$ are estimated jointly, the maximum likelihood estimators are generally inconsistent if the NegBin I assumption fails. On the other hand, if the NegBin I distribution holds, then the NegBin I MLE is more efficient than the Poisson QMLE (this conclusion follows from Section 14.5.2). Still, under assumption (19.3), the Poisson QMLE is more efficient than an estimator that requires only the conditional mean to be correctly specified for consistency. On balance, because of its robustness, the Poisson QMLE has the edge over NegBin I for estimating the parameters of the conditional mean. If conditional probabilities need to be estimated, then a more flexible model is probably warranted.

Other count data distributions imply a conditional variance other than assumption (19.3). A leading example is the NegBin II model of Cameron and Trivedi (1986). The NegBin II model can be derived from a model of unobserved heterogeneity in a Poisson model. Specifically, let $c_i > 0$ be unobserved heterogeneity, and assume that

$$y_i \mid \mathbf{x}_i, c_i \sim \text{Poisson}[c_i m(\mathbf{x}_i, \boldsymbol{\beta})]$$

If we further assume that $c_i$ is independent of $\mathbf{x}_i$ and has a gamma distribution with unit mean and $\text{Var}(c_i) = \eta^2$, then the distribution of $y_i$ given $\mathbf{x}_i$ can be shown to be negative binomial, with conditional mean and variance

$$\text{E}(y_i \mid \mathbf{x}_i) = m(\mathbf{x}_i, \boldsymbol{\beta}), \tag{19.27}$$

$$\begin{aligned} \text{Var}(y_i \mid \mathbf{x}_i) &= \text{E}[\text{Var}(y_i \mid \mathbf{x}_i, c_i) \mid \mathbf{x}_i] + \text{Var}[\text{E}(y_i \mid \mathbf{x}_i, c_i) \mid \mathbf{x}_i] \\ &= m(\mathbf{x}_i, \boldsymbol{\beta}) + \eta^2[m(\mathbf{x}_i, \boldsymbol{\beta})]^2 \end{aligned} \tag{19.28}$$

so that the conditional variance of $y_i$ given $\mathbf{x}_i$ is a quadratic in the conditional mean. Because we can write equation (19.28) as $\text{E}(y_i \mid \mathbf{x}_i)[1 + \eta^2 \text{E}(y_i \mid \mathbf{x}_i)]$, NegBin II also implies overdispersion, but where the amount of overdispersion increases with $\mathrm{E}(y_i \mid \mathbf{x}_i)$.

The log-likelihood function for observation $i$ is

$$\ell_i(\boldsymbol{\beta}, \eta^2) = \eta^{-2} \log\left[\frac{\eta^{-2}}{\eta^{-2} + m(\mathbf{x}_i, \boldsymbol{\beta})}\right] + y_i \log\left[\frac{m(\mathbf{x}_i, \boldsymbol{\beta})}{\eta^{-2} + m(\mathbf{x}_i, \boldsymbol{\beta})}\right]$$
$$+ \log[\Gamma(y_i + \eta^{-2})/\Gamma(\eta^{-2})] \qquad (19.29)$$

where $\Gamma(\cdot)$ is the gamma function defined for $r > 0$ by $\Gamma(r) = \int_0^\infty z^{r-1} \exp(-z)\, \mathrm{d}z$.

You are referred to Cameron and Trivedi (1986) for details. The parameters $\boldsymbol{\beta}$ and $\eta^2$ can be jointly estimated using standard maximum likelihood methods.

It turns out that, for *fixed* $\eta^2$, the log likelihood in equation (19.29) is in the linear exponential family; see GMT (1984a). Therefore, if we fix $\eta^2$ at any positive value, say $\bar{\eta}^2$, and estimate $\boldsymbol{\beta}$ by maximizing $\sum_{i=1}^N \ell_i(\boldsymbol{\beta}, \bar{\eta}^2)$ with respect to $\boldsymbol{\beta}$, then the resulting QMLE is consistent under the conditional mean assumption (19.27) *only*: for fixed $\eta^2$, the negative binomial QMLE has the same robustness properties as the Poisson QMLE. (Notice that when $\eta^2$ is fixed, the term involving the gamma function in equation (19.29) does not affect the QMLE.)

The structure of the asymptotic variance estimators and test statistics is very similar to the Poisson regression case. Let

$$\hat{v}_i = \hat{m}_i + \bar{\eta}^2 \hat{m}_i^2 \qquad (19.30)$$

be the estimated nominal variance for the given value $\bar{\eta}^2$. We simply weight the residuals $\hat{u}_i$ and gradient $\nabla_\beta \hat{m}_i$ by $1/\sqrt{\hat{v}_i}$:

$$\tilde{u}_i = \hat{u}_i/\sqrt{\hat{v}_i}, \qquad \nabla_\beta \tilde{m}_i = \nabla_\beta \hat{m}_i/\sqrt{\hat{v}_i} \qquad (19.31)$$

For example, under conditions (19.27) and (19.28), a valid estimator of $\mathrm{Avar}(\hat{\boldsymbol{\beta}})$ is

$$\left(\sum_{i=1}^N \nabla_\beta \hat{m}_i' \nabla_\beta \hat{m}_i / \hat{v}_i\right)^{-1}$$

If we drop condition (19.28), the estimator in expression (19.14) should be used but with the standardized residuals and gradients given by equation (19.31). Score statistics are modified in the same way.

When $\eta^2$ is set to unity, we obtain the **geometric QMLE**. A better approach is to replace $\eta^2$ by a first-stage estimate, say $\hat{\eta}^2$, and then estimate $\boldsymbol{\beta}$ by two-step QMLE. As we discussed in Chapters 12 and 13, sometimes the asymptotic distribution of the first-stage estimator needs to be taken into account. A nice feature of the two-step

QMLE in this context is that the key condition, assumption (12.37), can be shown to hold under assumption (19.27). Therefore, we can ignore the first-stage estimation of $\eta^2$.

Under assumption (19.28), a consistent estimator of $\eta^2$ is easy to obtain, given an initial estimator of $\boldsymbol{\beta}$ (such as the Poisson QMLE or the geometric QMLE). Given $\hat{\hat{\boldsymbol{\beta}}}$, form $\hat{\hat{m}}_i$ and $\hat{\hat{u}}_i$ as the usual fitted values and residuals. One consistent estimator of $\eta^2$ is the coefficient on $\hat{\hat{m}}_i^2$ in the regression (through the origin) of $\hat{\hat{u}}_i^2 - \hat{\hat{m}}_i$ on $\hat{\hat{m}}_i^2$; this is the estimator suggested by Gourieroux, Monfort, and Trognon (1984b) and Cameron and Trivedi (1986). An alternative estimator of $\eta^2$, which is closely related to the GLM estimator of $\sigma^2$ suggested in equation (19.15), is a weighted least squares estimate, which can be obtained from the OLS regression $\tilde{\tilde{u}}_i^2 - 1$ on $\hat{\hat{m}}_i$, where the $\tilde{\tilde{u}}_i$ are residuals $\hat{\hat{u}}_i$ weighted by $\hat{\hat{m}}_i^{-1/2}$. The resulting two-step estimator of $\boldsymbol{\beta}$ is consistent under assumption (19.7) only, so it is just as robust as the Poisson QMLE. It makes sense to use fully robust standard errors and test statistics. If assumption (19.3) holds, the Poisson QMLE is asymptotically more efficient; if assumption (19.28) holds, the two-step negative binomial estimator is more efficient. Notice that neither variance assumption contains the other as a special case for all parameter values; see Wooldridge (1997c) for additional discussion.

The variance specification tests discussed in Section 19.2.5 can be extended to the negative binomial QMLE; see Wooldridge (1991b).

### 19.3.2 Binomial Regression Models

Sometimes we wish to analyze count data conditional on a known upper bound. For example, Thomas, Strauss, and Henriques (1990) study child mortality within families conditional on number of children ever born. Another example takes the dependent variable, $y_i$, to be the number of adult children in family $i$ who are high school graduates; the known upper bound, $n_i$, is the number of children in family $i$. By conditioning on $n_i$ we are, presumably, treating it as exogenous.

Let $\mathbf{x}_i$ be a set of exogenous variables. A natural starting point is to assume that $y_i$ given $(n_i, \mathbf{x}_i)$ has a binomial distribution, denoted Binomial $[n_i, p(\mathbf{x}_i, \boldsymbol{\beta})]$, where $p(\mathbf{x}_i, \boldsymbol{\beta})$ is a function bounded between zero and one. Usually, $y_i$ is viewed as the sum of $n_i$ independent Bernoulli (zero-one) random variables, and $p(\mathbf{x}_i, \boldsymbol{\beta})$ is the (conditional) probability of success on each trial.

The binomial assumption is too restrictive for all applications. The presence of an unobserved effect would invalidate the binomial assumption (after the effect is integrated out). For example, when $y_i$ is the number of children in a family graduating from high school, unobserved family effects may play an important role.