Reading for Joseph Hilbe's online course, "Modeling Count Data" at statistics.com

REPLACEMENT CHAPTER 10

Hilbe, Joseph (2007), *Negative Binomial Regression*, Cambridge University Press

CHAPTER 10
Negative Binomial Panel Models

A basic assumption in the construction of models from likelihood theory is that observations in the model are independent. This is a reasonable assumption for perhaps the majority of studies. However, for longitudinal studies this assumption is not feasible; nor does it hold when data are clustered. For example, observations from a study on student drop-out can be clustered by the type of schools sampled. If the study is related to intervention strategies, schools in affluent suburban, middle class suburban, middle class urban, and below poverty level schools have more highly correlated strategies within the school type than between types or groups. Likewise, if we have study data taken on a group of individual patients over time (e.g., treatment results obtained once per month for a year), the data related to individuals in the various time periods are likely to be more highly correlated than are treatment results between patients. Any time the data can be grouped into clusters, or panels, of correlated groups, we must adjust the likelihood-based model (based on independent observations) to account for the extra-correlation.

We have previously employed robust variance estimators and bootstrapped standard errors when faced with overdispersed count data. Overdispersed Poisson models were adjusted by using different types of negative binomial models, or by extending the basic Poisson model by adjusting the variance or by designing a new log-likelihood function to account for the specific cause of the overdispersion. Examples we have previously discussed include zero-inflated models, zero-truncated models, hurdle models, and censored or truncated models.

In this chapter we shall describe a group of models that
1) add at least one extra parameter to the linear predictor, specifying how observations within panels are to be construed, and
2) derive new log-likelihoods based on panels of correlated observations.

The type of parameters that are added to the linear predictor, and the manner in which panels are treated will determine the type of panel model described. We shall first discuss fixed effects count models, differentiating between unconditional and conditional varieties. Following an examination of fixed-effects count models, we address random-effects models, followed by generalized estimating equations, or population-averaged models. Each of these types of panel models has software support in several commercial packages. Our final group of panel models has only recently found commercial software support, with only one package providing support for negative binomial models (LIMDEP). These regression models are commonly referred to as multilevel models. The two foremost members of this variety of panel model are random intercept and random coefficient or parameter models. More complex multilevel models are, for the most part, built on their basis. With respect to multilevel negative binomial models, current research has only begun in earnest within the last couple of years. They are still in the developmental stage.

10.1 Unconditional Fixed-Effects Negative Binomial Model

Fixed-effects count models may be estimated in two ways – unconditionally and conditionally. We begin with a consideration of the unconditional fixed-effects Poisson model since it is the basis on which we can understand the negative binomial parameterizations.

Unconditional estimation of the fixed effects Poisson model can be obtained using standard GLM software as well as the traditional maximum likelihood Poisson procedure. The model is specified by including a separate fixed effect for each defined panel in the data. The fixed effects are specified by indicator variables, just as is done when estimating factor or categorical predictors. We represent this relationship as:

$$\ln(\mu_{ik}) = \exp(\beta x_{ik} + \delta_i) \qquad \text{Eq. 10.1}$$

where δ is the fixed effect associated with individual, i, and subscript k indexes the observations associated with individual i. When a panel relates observations collected over a time period, it is customary to use the subscript t instead of k. We shall use k throughout our discussion, but with the knowledge that t is commonly used for longitudinal models.

The log-likelihood for the unconditional fixed effects Poisson takes the form of

$$\mathcal{L}(x\beta_i;y_i) == \sum_{i=1}^{n} \sum_{k=1}^{N} [\, y_{ik}\,(x_{ik}\beta + \delta_i) - \exp(x_{ik}\beta + \delta_i) - \ln\Gamma(y_{ik} + 1)] \qquad \text{Eq. 10.2}$$

Note the similarity to that of the standard Poisson log-likelihood function defined in Chapter 3 as:

$$\mathcal{L}(x\beta_i;y_i) = \sum_{i=1}^{n} \{\, y_i(x_i\beta) - \exp(x_i\beta) - \ln\Gamma(y_i+1) \,\}$$

I shall use the well known *ships* data set that was used in McCullagh & Nelder (1989), Hardin & Hilbe (2003), and other sources. The dataset contains values on the number of reported accidents for ships belonging to a company over a given time period. The variables are defined as:

```
accident  : number of accidents  (reponse)
ship      : ship identification (1-8)
op        : ship operated between the years 1975 and 1979 (1/0)
co65_69   : ship was in construction between 1965 and 1969 (1/0)
co70_74   : ship was in construction between 1970 and 1974 (1/0)
co75_79   : ship was in construction between 1975 and 1979 (1/0)
service   : months in service
```

With the natural log of the months of service specified as the offset, a basic Poisson model of the data is given as

```
. glm accident op co_65_69 - co_75_79 , nolog fam(poi) lnoffset(service)

Generalized linear models                          No. of obs      =        34
Optimization     : ML                              Residual df     =        29
                                                   Scale parameter =         1
Deviance         =  62.36534078                    (1/df) Deviance =  2.150529
Pearson          =  82.73714004                    (1/df) Pearson  =  2.853005
                                                   AIC             =  5.006819
Log likelihood   = -80.11591605                    BIC             = -39.89911
------------------------------------------------------------------------------
             |                 OIM
    accident |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          op |   .3874638    .118107     3.28   0.001     .1559783    .6189494
    co_65_69 |   .7542017   .1487697     5.07   0.000     .4626185    1.045785
    co_70_74 |    1.05087     .15757     6.67   0.000     .7420385    1.359701
    co_75_79 |   .7040507   .2203103     3.20   0.001     .2722504    1.135851
       _cons |   -6.94765   .1269363   -54.73   0.000    -7.196441    -6.69886
     service | (exposure)
------------------------------------------------------------------------------
```

Predictors appear to be significant; however, the model is clearly overdispersed. We have purposefully ignored the correlation of values within each panel of ship in the above model. A negative binomial model can be used to generically account for the overdispersion. After obtaining a value of α using *nbreg*, the GLM model appears as:
(NOTE: mid section of output excluded for space reasons)

```
. glm accident op co_65_69 - co_75_79 , fam(nb .1303451) lnoffset(service)

Generalized linear models                          No. of obs      =        34
Optimization     : ML                              Residual df     =        29
                                                   Scale parameter =         1
Deviance         =  36.84717336                    (1/df) Deviance =  1.270592
Pearson          =  42.24099154                    (1/df) Pearson  =  1.456586
Variance function: V(u) = u+(.1303451)u^2          [Neg. Binomial]
                                                   AIC             =  4.874952
Log likelihood   = -77.87418504                    BIC             = -65.41728
------------------------------------------------------------------------------
             |                 OIM
    accident |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          op |   .3536459   .2347302     1.51   0.132    -.1064169    .8137087
    co_65_69 |   1.012518    .329175     3.08   0.002     .3673472     1.65769
    co_70_74 |   1.255125   .3086897     4.07   0.000     .6501045    1.860146
    co_75_79 |   .7595303   .3854008     1.97   0.049     .0041585    1.514902
       _cons |  -6.933539   .2849396   -24.33   0.000    -7.492011   -6.375068
     service | (exposure)
------------------------------------------------------------------------------
```

Much of the overdispersion has been accommodated by the negative binomial model, but there is still evidence of extra correlation in the data. We also know from the data what may be causing overdispersion – the panel-specific effect of the individual ships.

We assign a specific indicator to each panel. Each ship will have a separate intercept. This type of model is called an unconditional fixed-effects model. As a Poisson model we have

```
. glm accident op co_65_69 - co_75_79 ship2-ship5 ,fam(poi) lnoffset(service)

Generalized linear models                          No. of obs      =        34
Optimization     : ML                              Residual df     =        25
                                                   Scale parameter =         1
Deviance         =  38.69505154                    (1/df) Deviance =  1.547802
Pearson          =  42.27525312                    (1/df) Pearson  =   1.69101
                                                   AIC             =  4.545928
Log likelihood   = -68.28077143                    BIC             = -49.46396
------------------------------------------------------------------------------
             |                 OIM
    accident |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          op |    .384467   .1182722     3.25   0.001     .1526578    .6162761
    co_65_69 |   .6971404   .1496414     4.66   0.000     .4038487    .9904322
    co_70_74 |   .8184266   .1697736     4.82   0.000     .4856763    1.151177
    co_75_79 |   .4534266   .2331705     1.94   0.052    -.0035791    .9104324
       ship2 |  -.5433443   .1775899    -3.06   0.002    -.8914141   -.1952745
       ship3 |  -.6874016   .3290472    -2.09   0.037    -1.332322    -.042481
       ship4 |  -.0759614   .2905787    -0.26   0.794    -.6454851    .4935623
       ship5 |   .3255795   .2358794     1.38   0.168    -.1367357    .7878946
       _cons |  -6.405902   .2174441   -29.46   0.000    -6.832084   -5.979719
     service | (exposure)
------------------------------------------------------------------------------
```

A substantial amount of the overdispersion present in the original Poisson model has been accounted for. However, the negative binomial handled the overdispersion better than the unconditional fixed effects Poisson.

It should be re-iterated at this point that the AIC and BIC statistics that appear with Stata’s *glm* command may give incorrect results. Values appearing under the table of estimates – when given -- are correct. These values are calculated from a Stata file written by the author called *abic.ado*. Remember as well that there are several different definitions for both the AIC and BIC statistic. I use Akaike’s (1973) original formula for AIC, which is found on page 27, Table 2.1. Other formulations are presented by the software as well.

We next attempt to model an unconditional fixed-effects negative binomial.

```
. glm accident op co_65_69-co_75_79 ship2-ship5, fam(nb 0.0000000253)
lnoffset(service)

Generalized linear models                          No. of obs      =        34
Optimization     : ML                              Residual df     =        25
                                                   Scale parameter =         1
Deviance         =  38.69504594                    (1/df) Deviance =  1.547802
Pearson          =  42.27517882                    (1/df) Pearson  =  1.691007
Variance function: V(u) = u+(0.0000000253)u^2      [Neg. Binomial]
                                                   AIC             =  4.545928
Log likelihood   = -68.28077281                    BIC             = -49.46397
------------------------------------------------------------------------------
             |                 OIM
    accident |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          op |    .384467   .1182722     3.25   0.001     .1526577    .6162762
    co_65_69 |   .6971404   .1496415     4.66   0.000     .4038485    .9904323
    co_70_74 |   .8184266   .1697737     4.82   0.000     .4856763    1.151177
    co_75_79 |   .4534266   .2331705     1.94   0.052    -.0035792    .9104324
       ship2 |  -.5433445   .1775899    -3.06   0.002    -.8914144   -.1952747
       ship3 |  -.6873984   .3290468    -2.09   0.037    -1.332318   -.0424786
       ship4 |  -.0759617   .2905787    -0.26   0.794    -.6454854     .493562
       ship5 |   .3255795   .2358794     1.38   0.168    -.1367356    .7878946
       _cons |  -6.405901   .2174441   -29.46   0.000    -6.832084   -5.979719
     service | (exposure)
------------------------------------------------------------------------------
```

From a previous maximum likelihood estimation of the model it is discovered that the negative binomial value of α is approximately 0.0. This indicates that the model is in fact Poisson, and that the extra overdispersion is likely from a source other than the fixed panel effect of the individual ships. Nevertheless, the fit, as indicated by the AIC and BIC statistics, appears to favor the unconditional fixed-effects Poisson model over a model not accounting for the panel effect of individual ships.

A caveat on using this form of fixed-effects regression: use it only if there are a relatively few number of panels in the data. If there are more than twenty panels, it is preferred to use the conditional fixed-effects model. The greater the number of panels, the greater the possible bias in parameter estimates for the levels or panels of the effect variable. This is called the 'incidental parameters problem', first defined by Neyman and Scott (1948). It is interesting that a number of econometricians have thought that the incidental parameters problem, which we shall refer to as the IP problem, affects the unconditional fixed-effects Poisson model. Woutersen (2002) attempted to ameliorate the IP problem with Poisson models by employing an integrated moment estimator. Other attempts include Lancaster (2002) and Vadeby (2002). Most of these "solutions" are based on separating the main model parameters from the array of fixed-effects parameters. However, it has been demonstrated by Greene (2006) and others that the IP problem is not real when applied to the Poisson model. This conclusion is based on the observation that the Poisson conditional fixed effects estimator is numerically equal to the unconditional estimator, which means that there is no IP problem. On the other hand, the IP problem does affect the unconditional fixed-effects negative binomial. But the fixed-effects negative binomial model has a different problem. It is intrinsically different from the Poisson. Recall that the Poisson fixed-effects has a mean, μ_{ik}, value of $\exp(\beta x_{ik}$

+ δ_i). This means that the fixed effect is built into the Poisson mean parameter. The negative binomial fixed-effects model, though, builds the fixed-effects into the distribution of the gamma heterogeneity, α, not the mean. This makes it rather difficult to interpret the IP problem with the negative binomial. One result is that the estimator is inconsistent in the presence of a large number of fixed-effects. But exactly how it is inconsistent is still a matter of debate.

There is good evidence that in the presence of a large number of fixed effects the unconditional negative binomial will underestimate standard errors, resulting in insufficient coverage of the confidence intervals. That is, negative binomial predictors appear to enter the model as significant when in fact they do not. Simulation studies (Greene, 2006) have demonstrated that scaling the unconditional fixed effects negative binomial model standard errors by the deviance-based dispersion statistic produces standard errors that are closer to the nominal values. This is not the case when using Pearson χ^2-based dispersion as the basis for scaling standard errors, as is the norm for non panel models. On the other hand, using the deviance-based dispersion statistic for scaling unconditional fixed effects Poisson models does not improve coverage, and the Pearson χ^2 dispersion should be used. These facts need to be kept clearly in mind when modeling unconditional fixed-effects count models.

10.2 Conditional Fixed-Effects Negative Binomial Model

Panel data models are constructed in order to control for all of the stable predictors in the model and to account for the correlation resulting from observations being associated within groups or panels. The value of conditional fixed-effects models is that a near infinite number of panels may be adjusted, while at the same time being conditioned out of the actual model itself. We do not have to deal with a host of dummy intercepts.

A conditional fixed-effects model is derived by conditioning out the fixed effects from the model estimation. Like unconditional fixed-effects models, there is an separate fixed effect, δ, specified in the linear predictor. Hence, $\eta = x\beta + \delta$. However, unlike the unconditional version, a revised log-likelihood function is derived to affect the conditioning out of the panel effects through a sufficient statistic.

The conditional log-likelihood is conditioned on the sum of the responses within each panel.

$$\Sigma \; y_{ik} \qquad \text{Eq. 10.3}$$

For the Poisson, this yields a conditional log-likelihood that is defined, without subscripts for ease of interpretation, as:

CONDITIONAL FIXED EFFECTS POISSON LOG-LIKELIHOOD

$$\mathcal{L}(\beta;y) = \Sigma\{ \ln\Gamma(\Sigma y+1) - \Sigma(\ln\Gamma(y+1)) + \Sigma[y^*x\beta - y^*\ln(\Sigma(\exp(x\beta)))]\} \quad \text{Eq. 10.4}$$

or

$$\mathcal{L}(\mu;y) = \Sigma\{ \ln\Gamma(\Sigma y+1) - \Sigma(\ln\Gamma(y+1)) + \Sigma[y^*x\beta - y^*\ln(\Sigma\mu)]\}$$

Following the derivation of the model as proposed by Hausman, Hall and Griliches (1984), the conditional fixed-effects negative binomial log-likelihood is shown as:

CONDITIONAL FIXED EFFECTS NEGATIVE BINOMIAL LOG-LIKELIHOOD

$$\mathcal{L}(\beta;y) = \ln\Gamma(\Sigma(\exp(x\beta))) + \ln\Gamma(\Sigma y+1) - \Sigma(\ln\Gamma(y+1)) - \ln\Gamma(\Sigma y + \Sigma(\exp(x\beta))) + \Sigma(\ln\Gamma(\exp(x\beta)+y)) - \Sigma(\ln\Gamma(\exp(x\beta))) \quad \text{Eq. 10.5}$$

or

$$\mathcal{L}(\mu;y) = \ln\Gamma(\Sigma\mu) + \ln\Gamma(\Sigma y+1) - \Sigma(\ln\Gamma(y+1)) - \ln\Gamma(\Sigma y + \Sigma(\mu) + \Sigma(\ln\Gamma(\mu+y)) - \Sigma(\ln\Gamma(\mu)) \quad \text{Eq. 10.6}$$

Note that the heterogeneity parameter, δ, does not appear in the log likelihood. It does not, as a result, appear in the model output.

Another model that we should mention, but that has had very little application and currently has no commercial software support, is the NB-1 conditional fixed effects model. Its log-likelihood can be derived as:

CONDITIONAL FIXED EFFECTS NB-1 LOG-LIKELIHOOD

$$\mathcal{L}(\beta;y) = \Sigma\{\ln\Gamma(\exp(x\beta)+y) - \ln\Gamma(\exp(x\beta)) - \ln\Gamma(y+1)\} + \ln\Gamma(\Sigma(x\beta) + \ln\Gamma(\Sigma(y)+1) - \ln\Gamma(\Sigma(x\beta) + \Sigma(y)) \quad \text{Eq. 10.7}$$

or

$$\mathcal{L}(\mu;y) = \Sigma\{\ln\Gamma(\mu+y) - \ln\Gamma(\mu) - \ln\Gamma(y+1)\} + \ln\Gamma(\Sigma(x\beta) + \ln\Gamma(\Sigma(y)+1) - \ln\Gamma(\Sigma(x\beta) + \Sigma(y)) \quad \text{Eq. 10.8}$$

Complete derivations of both the Poisson and negative binomial log likelihood functions can be found in Cameron & Trevedi (1998), and Hardin and Hilbe (2003).

We next model the same data using conditional fixed-effects as we did with unconditional fixed-effects.

```
. xtpoisson accident op co_65_69 - co_75_79, nolog i(ship) exposure(service) fe

Conditional fixed-effects Poisson regression    Number of obs      =        34
Group variable (i): ship                        Number of groups   =         5
                                                Obs per group: min =         6
                                                               avg =       6.8
                                                               max =         7
                                                Wald chi2(4)       =     48.44
Log likelihood  = -54.641859                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
    accident |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          op |    .384467   .1182722     3.25   0.001     .1526578    .6162761
    co_65_69 |   .6971405   .1496414     4.66   0.000     .4038487    .9904322
    co_70_74 |   .8184266   .1697737     4.82   0.000     .4856764    1.151177
    co_75_79 |   .4534267   .2331705     1.94   0.052    -.0035791    .9104324
     service | (exposure)
------------------------------------------------------------------------------
AIC = 3.450
```

The associated AIC statistic is 3.450, which is a full one unit lower in value than the unconditional model. Compare the above list of parameter estimates and standard errors output with that of the unconditional results:

```
. glm accident op co_65_69 - co_75_79 ship2-ship5 ,fam(poi) lnoffset(service)

------------------------------------------------------------------------------
             |                 OIM
    accident |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          op |    .384467   .1182722     3.25   0.001     .1526577    .6162762
    co_65_69 |   .6971404   .1496415     4.66   0.000     .4038485    .9904323
    co_70_74 |   .8184266   .1697737     4.82   0.000     .4856763    1.151177
    co_75_79 |   .4534266   .2331705     1.94   0.052    -.0035792    .9104324
       ship2 |  -.5433445   .1775899    -3.06   0.002    -.8914144   -.1952747
       ship3 |  -.6873984   .3290468    -2.09   0.037    -1.332318   -.0424786
       ship4 |  -.0759617   .2905787    -0.26   0.794    -.6454854     .493562
       ship5 |   .3255795   .2358794     1.38   0.168    -.1367356    .7878946
       _cons |  -6.405901   .2174441   -29.46   0.000    -6.832084   -5.979719
     service | (exposure)
------------------------------------------------------------------------------
```

The first thing that can be noticed is that the parameter estimates and standard errors for the unconditional and conditional fixed-effects Poisson models are identical, though this equality is not an equivalence of the two approaches. The conditional fixed-effects Poisson model does not include a constant, whereas the unconditional does. Of interest to note as well is the fact that the respective log likelihoods differ (-54.64 to -68.28). We previously pointed out that the AIC statistics differ as well (3.45 to 4.55). We may conclude from this that, although the estimates and standard errors are the same, the two models intrinsically differ, with the preferred fit being that of the conditional fixed-effects Poisson model.

We now turn to modeling the same data using the conditional fixed-effects negative binomial model. It can be displayed as:

```
. xtnbreg accident op co_65_69 - co_75_79, nolog i(ship) exposure(service) fe

Conditional FE negative binomial regression     Number of obs      =        34
Group variable (i): ship                        Number of groups   =         5
                                                Obs per group: min =         6
                                                               avg =       6.8
                                                               max =         7
                                                Wald chi2(4)       =     34.81
Log likelihood  =  -53.08425                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
    accident |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          op |   .3835077   .1423269     2.69   0.007      .104552    .6624634
    co_65_69 |   .6772064   .1769717     3.83   0.000     .3303482    1.024065
    co_70_74 |   .8186325   .2009683     4.07   0.000      .424742    1.212523
    co_75_79 |   .4774847   .2773654     1.72   0.085    -.0661414    1.021111
       _cons |  -5.995012   .8205518    -7.31   0.000    -7.603264    -4.38676
     service | (exposure)
------------------------------------------------------------------------------
AIC Statistic =        3.417
```

Compare the above with the unconditional model:

```
. glm accident op co_65_69-co_75_79 ship2-ship5, fam(nb 0.0000000253)
lnoffset(service)

Log likelihood   = -68.28077281                             AIC  =  4.545928
------------------------------------------------------------------------------
             |                 OIM
    accident |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          op |    .384467   .1182722     3.25   0.001     .1526577    .6162762
    co_65_69 |   .6971404   .1496415     4.66   0.000     .4038485    .9904323
    co_70_74 |   .8184266   .1697737     4.82   0.000     .4856763    1.151177
    co_75_79 |   .4534266   .2331705     1.94   0.052    -.0035792    .9104324
       ship2 |  -.5433445   .1775899    -3.06   0.002    -.8914144   -.1952747
       ship3 |  -.6873984   .3290468    -2.09   0.037    -1.332318   -.0424786
       ship4 |  -.0759617   .2905787    -0.26   0.794    -.6454854     .493562
       ship5 |   .3255795   .2358794     1.38   0.168    -.1367356    .7878946
       _cons |  -6.405901   .2174441   -29.46   0.000    -6.832084   -5.979719
     service | (exposure)
------------------------------------------------------------------------------
```

The conditional and unconditional fixed-effects negative binomial models do not normally have the same parameter estimates. However, in this case the data are very close to Poisson, so the results will not be too dissimilar. Note though that, unlike the Poisson, both the conditional and unconditional negative binomial models have a constant in the model. And, as previously indicated, the conditional version does not have a value for α. For a discussion on unconditional fixed effects constants see Greene (2003).

Unfortunately it has been discovered that the conditional fixed-effects negative binomial model is not a true fixed-effects model since it fails to control for all of its predictors. In addition, the α parameter that is conditioned out of the loglikelihood does not correspond to the different intercepts in the decomposition of μ. Allison and Waterman (2002) provide a full discussion, together with alternative models. The negative multinomial model has been suggested as an alternative for the conditional negative binomial. However, the negative multinomial produces the same estimators as a conditional Poisson, so does not provide any additional capability for handling overdispersion over what is available with Poisson options. The other foremost alternative is to revert to the unconditional negative binomial model. In fact, they recommend that the unconditional negative binomial be used rather then the conditional. But, as previously discussed, it should also be accompanied by scaling the standard errors by the Pearson Chi2 dispersion. If this strategy is unsatisfactory, then one should consider using other panel models; e.g. random-effects models or GEE models.

10.3 Random Effects Negative Binomial

Random-effects models begin with the same notation as fixed-effects models in that a heterogeneity parameter is added to the linear predictor. Moreover, the fixed-effects parameter, δ, is now considered to be an iid random parameter rather than a fixed parameter. It is derived from a known probability distribution. In the case of Poisson, the random parameter can follow the usual Gaussian distribution, the gamma distribution, or the inverse Gaussian distribution. Gamma is the preferred random distribution to use since it is the conjugate prior to Poisson. The gamma distribution also allows an analytic solution of the integral in the likelihood. Other random distributions do not have these favorable features.

We shall use the term ν rather than δ for depicting the random parameter for random-effects count models. In so doing we shall be consistent with common terminology. We shall also use the standard GLM term μ rather than λ for the Poisson and negative binomial fitted value. λ is commonly found in the literature on count response models. But as with our choice of using ν, we shall use the term μ to maintain consistency for all count models that in some respect emanate from a GLM background.

The framework for the random-effects Poisson is

$$\ln(\mu_{ik}) = \beta x_{ik} + \nu_i \qquad \text{Eq. 10.9}$$

with $\nu_i = \nu + \epsilon_i$

Following the derivation of the random gamma effects Poisson model by Hausman, Hall and Griliches (1984), we assume a random multiplicative effect on μ specified as:

$$\begin{aligned}\Pr(y_{ik};\nu_i,x) &= \{\Pi(\mu_{ik}\nu_i)^y/y_{ik}!\}\ \exp(-\Sigma\mu_{ik}\nu_i)\\ &= (\Pi(\mu_{ik}^{\ y})/y_{ik}!)\ \exp(-\nu_i\Sigma\mu_{ik})\nu^{\Sigma y}\end{aligned} \qquad \text{Eq. 10.10}$$

Summing of subject-specific observations are over panels with a mean given for each panel of

$$\mu_{ik} = \exp(x_{ik}\beta) \qquad \text{Eq. 10.11}$$

where each panel has separately defined means given as

$$\nu_i \mu_{ik} = \exp(x_{ik}\beta + \eta_{ik}) \qquad \text{Eq. 10.12}$$

With ν following a gamma distribution with a mean of one and a variance of θ, we have the mixture

$$\Pr(\nu_i, \mu_{ik}) = \theta^{\theta}/\Gamma(\theta)\ \nu_i^{\theta-1} \exp(-\theta\nu_i) \quad \Pi \exp(-\nu_i\mu_{ik})(\nu_i\mu_{ik})^{y} / y_{ik}!) \qquad \text{Eq. 10.13}$$

where the terms prior to the product sign specify the gamma distributed random component and the terms from the product sign to the right provide the Poisson probability function. This mixture is of the same structural form as we derived for the NB-1 probability function in chapter 5.

Each panel is independent of one another, with their joint density combined as the product of the individual panels. The log-likelihood for the gamma distributed Poisson random effects model can be calculated by integrating over ν_i. The result is:

RANDOM EFFECTS POISSON WITH GAMMA EFFECT

$$\mathcal{L}(\beta;y) = \sum_{i=1}^{n} \{ \ln\Gamma(\theta+\Sigma y_{ik}) - \ln\Gamma(\theta) - \Sigma(\ln\Gamma(y_{ik}+1) + \theta\ln(u_i) + (\Sigma y_{ik})\ln(1-u_i) - (\Sigma y_{ik})\ln(\Sigma(\exp(x_{ik}\beta))) + \Sigma(y^{*}x_{ik}\beta)\} \qquad \text{Eq. 10.14}$$

where $\theta = 1/\nu$

and $u_i = \theta / (\theta + \Sigma(\exp(x_{ik}\beta))$

and all Σ within the braces are defined as

$$\sum_{k=1}^{n}$$

We shall use the same data that were used for examining fixed-effects models for the examples of random-effects Poisson and negative binomial. Random effects Poisson, with a gamma effect, is shown below as:

```
. xtpoisson accident op co_65_69-co_75_79, nolog exposure(service) i(ship) re

Random-effects Poisson regression               Number of obs      =        34
Group variable (i): ship                        Number of groups   =         5
Random effects u_i ~ Gamma                      Obs per group: min =         6
                                                               avg =       6.8
                                                               max =         7
                                                Wald chi2(4)       =     50.90
Log likelihood  = -74.811217                    Prob > chi2        =    0.0000
------------------------------------------------------------------------------
    accident |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          op |   .3827453   .1182568     3.24   0.001     .1509662    .6145244
    co_65_69 |   .7092879   .1496072     4.74   0.000     .4160633    1.002513
    co_70_74 |   .8573273   .1696864     5.05   0.000     .5247481    1.189906
    co_75_79 |   .4958618   .2321316     2.14   0.033     .0408922    .9508313
       _cons |  -6.591175   .2179892   -30.24   0.000    -7.018426   -6.163924
     service | (exposure)
-------------+----------------------------------------------------------------
    /lnalpha |  -2.368406   .8474597                     -4.029397   -.7074155
-------------+----------------------------------------------------------------
       alpha |   .0936298   .0793475                      .0177851    .4929165
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0: chibar2(01) =    10.61 Prob>=chibar2 = 0.001

AIC Statistic =        4.754
```

The likelihood-ratio tests whether the data are better modeled using a panel structure or whether a pooled structure is preferred. Here we find that the random effects (panel) parameterization is preferred over the pooled, or standard, Poisson model.

It is interesting to compare this output with that of a NB-1 model on the same data. Recall that mixing the gamma random parameter with the Poisson probability function resulted in a NB-1 PDF. Of course the NB-1 PDF does not account for the panel structure of the data as does the gamma distributed random effects Poisson. However, because of the base similarity of the two models, we should expect that the outputs of the respective models will be similar, but not identical. This suspicion is indeed confirmed. Note also that the output above specifies "alpha" as the heterogeneity parameter. It is the same as what we have referred to as ν. Interpret δ in the same manner.

CONSTANT NEGATIVE BINOMIAL (NB-1)

```
. nbreg accident op co_65_69-co_75_79, exposure(service) cluster(ship)
disp(constant)

Negative binomial regression                      Number of obs   =         34
Dispersion           = constant                   Wald chi2(2)    =          .
Log pseudolikelihood = -74.801716                 Prob > chi2     =          .
                                   (Std. Err. adjusted for 5 clusters in ship)
------------------------------------------------------------------------------
             |               Robust
    accident |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          op |   .3824838   .0903809     4.23   0.000     .2053404    .5596272
    co_65_69 |   .7174666   .0996523     7.20   0.000     .5221517    .9127814
    co_70_74 |   1.025627   .2156908     4.76   0.000      .602881    1.448373
    co_75_79 |   .7266669   .1996568     3.64   0.000     .3353468    1.117987
       _cons |  -6.924931   .0522819  -132.45   0.000    -7.027402   -6.822461
     service | (exposure)
-------------+----------------------------------------------------------------
    /lndelta |  -.1042511   .4995717                     -1.083394    .8748916
-------------+----------------------------------------------------------------
       delta |   .9009991   .4501137                       .338445    2.398615
------------------------------------------------------------------------------
AIC Statistic =        4.753
```

We next compare the above with the standard NB-2.

NEGATIVE BINOMIAL (NB-2)

```
. nbreg accident op co_65_69-co_75_79, exposure(service) cluster(ship)

Negative binomial regression                      Number of obs   =         34
Dispersion           = mean                       Wald chi2(2)    =          .
Log pseudolikelihood = -77.874185                 Prob > chi2     =          .
                                   (Std. Err. adjusted for 5 clusters in ship)
------------------------------------------------------------------------------
             |               Robust
    accident |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          op |   .3536459   .2161704     1.64   0.102    -.0700402     .777332
    co_65_69 |   1.012518   .6365455     1.59   0.112    -.2350879    2.260125
    co_70_74 |   1.255125   .3774548     3.33   0.001     .5153274    1.994923
    co_75_79 |   .7595303   .2988691     2.54   0.011     .1737576    1.345303
       _cons |  -6.933539   .0955349   -72.58   0.000    -7.120784   -6.746294
     service | (exposure)
-------------+----------------------------------------------------------------
    /lnalpha |  -2.037569   1.517455                     -5.011727    .9365884
-------------+----------------------------------------------------------------
       alpha |   .1303451   .1977929                      .0066594    2.551263
------------------------------------------------------------------------------
AIC Statistic =        4.934
```

When deriving the random effects negative binomial, we begin with the same Poisson-gamma mixture as Equation 10.13. By rearranging terms and not integrating out ν as we did for the Poisson, we have:

$$\Pr(y;\mu\nu) = \frac{\Gamma(\mu_{ik} + y_{ik})}{\Gamma(\mu_{ik})\Gamma(y_{ik}+1)} \left(\frac{1}{1+\nu_i}\right)^{\mu_{ik}} \left(\frac{\nu_i}{1+\nu_i}\right)^{y_{ik}}$$

Eq. 10.15a

which is the panel structure form of the NB-1 model. For the random effect we select the beta distribution, which is the conjugate prior of the negative binomial, as gamma was the conjugate prior of the Poisson. With the dispersion defined as the variance divided by the mean, or $1 + \nu$, it is stipulated that the inverse dispersion is distributed following a Beta distribution. We have, therefore,

$\nu_i /(1+\nu_i) \sim \text{Beta}(a,b)$

which layers the random panel effect onto the negative binomial model. Deriving the log-likelihood function results in the following form of the function:

RANDOM EFFECTS NEGATIVE BINOMIAL WITH BETA EFFECT PDF

$$f(y_{it};\beta,a,b) = \frac{\Gamma(a+b) + \Gamma(a+\Sigma(\exp(x_{ik}\beta))) + \Gamma(b+\Sigma y_{ik})}{\Gamma(a) - \Gamma(b) - \Gamma(a+b+\Sigma(\exp(x_{ik}\beta)) + \Sigma y_{ik})}$$

$$\prod_{i=1}^{n_i} \frac{\Gamma((\exp(x_{it}\beta))+ y_{it})}{\Gamma(\exp(x_{it}\beta)))\,\Gamma(y_{it}+1)}$$

Eq. 10.15b

and

RANDOM EFFECTS NEGATIVE BINOMIAL WITH BETA EFFECT

$$\mathcal{L}(\beta;y_{it},a,b) = \Sigma\ln\Gamma(a+b) + \ln\Gamma(a+\Sigma(\exp(x_{ik}\beta))) + \ln\Gamma(b+\Sigma y_{ik}) - \ln\Gamma(a) - \ln\Gamma(b)-\ln\Gamma(a+b+\Sigma(\exp(x_{ik}\beta)) + \Sigma y_{ik}) + \Sigma(\ln\Gamma((\exp(x_{it}\beta))+ y_{it}) - \ln\Gamma(y_{it}+1) - \ln\Gamma(\exp(x_{it}\beta)))$$

Eq. 10.16

Where the first Σ is summed from i=1 to $n_{i.}$ and the second through fifth Σ is summed from k=1 to n_k, and the final Σ is summed from t=1 to $n_{t.}$ Derivatives of the conditional fixed-effects and random-effects Poisson and negative binomial models are given in Greene (2006).

Output of the beta distributed random effect negative binomial is shown below for the data we have used in this chapter.

```
. xtnbreg accident op co_65_69-co_75_79, exposure(service) i(ship) re

Random-effects negative binomial regression     Number of obs      =        34
Group variable (i): ship                        Number of groups   =         5
Random effects u_i ~ Beta                       Obs per group: min =         6
                                                               avg =       6.8
                                                               max =         7
                                                Wald chi2(4)       =     37.15
Log likelihood  = -73.222498                    Prob > chi2        =    0.0000
------------------------------------------------------------------------------
    accident |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          op |   .3815599   .1426935     2.67   0.007     .1018857     .661234
    co_65_69 |   .6935116   .1777982     3.90   0.000     .3450335     1.04199
    co_70_74 |   .8766236   .2024878     4.33   0.000     .4797548    1.273492
    co_75_79 |   .5452717    .277925     1.96   0.050     .0005488    1.089995
       _cons |  -6.039451   .8228179    -7.34   0.000    -7.652145   -4.426758
     service | (exposure)
-------------+----------------------------------------------------------------
       /ln_r |   3.641897   1.097047                      1.491725    5.792069
       /ln_s |   3.029242   1.108291                      .8570312    5.201453
-------------+----------------------------------------------------------------
           r |   38.16417   41.86788                      4.444755    327.6904
           s |   20.68155   22.92118                      2.356155    181.5358
------------------------------------------------------------------------------
Likelihood-ratio test vs. pooled: chibar2(01) =     3.16 Prob>=chibar2 = 0.038

AIC Statistic =        4.719
```

A likelihood ratio test accompanies the output, testing the random-effects panel estimator with the pooled NB-1, or constant dispersion, estimator. Here the random-effects model is preferred. r and s refer to the Beta distribution values for the a and b parameters respectively. Note the extremely wide confidence intervals.

We shall use another example of a random effects model. It is not only a good random effects example in its own right, but it will prove useful when discussing generalized estimating equations.

The data come from Thall and Vail (1990) and is used in Hardin & Hilbe (2003). Called the *Progabide* data set, the data are from a panel study of seizures in patients with epilepsy. Four successive two-week counts of seizures were taken for each patient. The response is *seizure*, with explanatory predictors consisting of the *progabide* treatment (1/0), a follow-up indicator called *time* (1/0), and an interaction of the two, called *timeXprog*. An offset, called *Period*, is given for weeks in the study, which are either 2 or 8. Since *Period* is converted by a natural log, the two values of *lnPeriod* are 2.079442 and 0.6931472. There are 295 observations on 59 epileptic patients (panels), with 5 observations, *t*, each.

Results of modeling the data are

GAMMA DISTRIBUTED RANDOM EFFECTS POISSON

```
. xtpoisson seizures time progabide timeXprog, nolog offset(lnPeriod) re i(id)

Random-effects Poisson regression               Number of obs      =       295
Group variable: id                              Number of groups   =        59
Random effects u_i ~ Gamma                      Obs per group: min =         5
                                                               avg =       5.0
                                                               max =         5
                                                Wald chi2(3)       =      5.73
Log likelihood  = -1017.3826                    Prob > chi2        =    0.1253
------------------------------------------------------------------------------
    seizures |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |    .111836   .0468768     2.39   0.017     .0199591    .2037129
   progabide |   .0275345   .2108952     0.13   0.896    -.3858125    .4408815
   timeXprog |  -.1047258   .0650304    -1.61   0.107     -.232183    .0227314
       _cons |   1.347609   .1529187     8.81   0.000     1.047894    1.647324
    lnPeriod |   (offset)
-------------+----------------------------------------------------------------
    /lnalpha |   -.474377   .1731544                     -.8137534   -.1350007
-------------+----------------------------------------------------------------
       alpha |   .6222726   .1077492                      .4431915    .8737153
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0: chibar2(01) =  2602.24 Prob>=chibar2 = 0.000
```

The model appears to favor the panel specification of the data (Prob>=chibar2=0.000). Note that alpha is 0.6222.

BETA DISTRIBUTED RANDOM EFFECTS NEGATIVE BINOMIAL

```
. xtnbreg seizures time progabide timeXprog, nolog offset(lnPeriod) re i(id)

Random-effects negative binomial regression     Number of obs      =       295
Group variable: id                              Number of groups   =        59

Random effects u_i ~ Beta                       Obs per group: min =         5
                                                               avg =       5.0
                                                               max =         5

                                                Wald chi2(3)       =      6.88
Log likelihood  = -891.81046                    Prob > chi2        =    0.0759
------------------------------------------------------------------------------
    seizures |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   .0637907   .0821828     0.78   0.438    -.0972846     .224866
   progabide |    .025827   .1822767     0.14   0.887    -.3314287    .3830827
   timeXprog |  -.2619935   .1158202    -2.26   0.024    -.4889969   -.0349901
       _cons |   .7961454   .1908792     4.17   0.000     .4220289    1.170262
    lnPeriod |   (offset)
-------------+----------------------------------------------------------------
       /ln_r |   1.261781   .1973829                      .8749176    1.648644
       /ln_s |    1.51837    .220681                      1.085843    1.950897
-------------+----------------------------------------------------------------
           r |   3.531705   .6970981                      2.398678    5.199925
           s |   4.564778    1.00736                      2.961936    7.034993
------------------------------------------------------------------------------
Likelihood-ratio test vs. pooled: chibar2(01) =   228.19 Prob>=chibar2 = 0.000
```

Again, the panel structure of the data is confirmed by the model (Prob>=chibar2=0.000). It is important to remember that this conclusion holds with respect to the subject specific parameterization of the data. A margin or population averaged model may yield contrary results, however, I believe that this would occur in very few instances. A thorough discussion of the derivation of the random-effects Poisson and negative binomial models can be found in Frees (2004) and Greene (2006).

A disadvantage of a random-effects model is that it assumes that the subject-specific effects are uncorrelated with other predictors. The Hausman test is commonly used to evaluate whether data should be modeled using a fixed- or a random-effects model. The test is based on work done by Mundlak (1978) who argued that the fixed-effects model is more robust than the random-effects model to important predictors left out of the model. That the subject-specific effects are not correlated highly with model predictors is specified as the null hypothesis. See Freese, p 247, for additional discussion.

Random effects estimators are more efficient than fixed-effects estimators when the data come from within a larger population of observations, as well as when there are more panels in the data. Data coming from a smaller complete data set, with relatively few panels, prefer the fixed-effects estimator.

Random-effects models are subject-specific models in that the log-likelihood models individual observations rather than the average of panels, or marginal distribution. GEE's are population averaging models, and care must be taken when interpreting GEE against random effects model output, as we shall turn to next.

10.4 Generalized Estimating Equations

10.4.1 The GEE Algorithm

Generalized estimating equations (GEE) refers to a population averaging panel method first proposed by Liang and Zeger in 1986. It is developed as an extension to the standard generalized linear models algorithm. Unlike the random-effects model, which is subject-specific, GEE is a population-averaged approach in which the marginal effects of the model are averaged across individuals. Essentially, GEE models the average response of individuals sharing the same predictors across all of the panels.

The GEE algorithm is structured such that observations are grouped in panels, in a similar manner to fixed-effects and random-effects models. At the heart of the model specification, the variance function is factored to include an identity matrix operating as a within-panel correlation structure. This panel form of the variance function appears as:

$$V(\mu_i) = [D(V(\mu_{ik}))^{1/2} \, I_{(nxn)} \, D(V(\mu_{ik}))^{1/2}]_{nxn} \qquad \text{Eq. 10.17}$$

Where $V(\mu_{ik})$ is the GLM variance function defined from the family being modeled. For example, the variance function of the Poisson family is μ; the variance of the NB-2 model is $\mu + \alpha\mu^2$. This structure (represented by the identity matrix) is called the independent correlation structure.

The benefit of the GEE approach is that the identity matrix, which is sandwiched between the factored GLM variance functions, can be replaced by a parameterized matrix containing values other than one and zero. The structure of values that are substituted into this alternative matrix define the various GEE correlation structures. These include:

Foremost GEE Correlation Structures

1: Independent	2: Exchangeable	3: Unstructured
4: Autoregressive	5: Stationary	6: Non-stationary

The most commonly used correlation structure is the exchangeable, which we shall later describe in more detail. All of the structures define constraints on the values to be estimated. Those values are estimated from Pearson residuals obtained using the regression parameters. Pearson residuals are defined, in panel format, as

$$r_{ik} = \Sigma(y_{ik}-\mu_{ik})^2/V(\mu_{ik}). \qquad \text{Eq. 10.18}$$

The Poisson Pearson residual is defined as $\Sigma(y_{ik}-\mu_{ik})^2/\mu_{ik}$. The exchangeable correlation is then defined as, without subscripts (see Hardin & Hilbe, 2002),

$$\alpha = \frac{1}{\phi} \Sigma_{i=1}\left\{ \frac{\Sigma_{j=1}\Sigma_{k=1} r_{ij}r_{ik} - \Sigma_{j=1}r_{ij}^2}{n_i(n_i-1)} \right\}$$

Eq. 10.19

where the dispersion parameter, ϕ, is calculated as

$$\phi = 1/n_i \ \Sigma_{i=1}\Sigma_{j=1} \ r_{ij}^2$$

Eq. 10.19a

The exchangeable correlation structure has also been referred to as the compound symmetry matrix and the equal correlation structure. All off-diagonal values have a single scalar constant.

Other correlation matrices are defined in different manners, depending on the purport of the structure. However, all are inserted into the variance function as

$$V(\mu_i) = [D(V(\mu_{ik}))^{1/2} R(a)_{(nxn)} D(V(\mu_{ik}))^{1/2}]_{nxn} \qquad \text{Eq. 10.20}$$

The GEE algorithm begins by estimating a model from the member families; e.g. Poisson. After the initial iteration the Pearson residuals are calculated (Eq 10.18) and put into the formula for calculating R(a) (Eq. 10.19). R(a) is then inserted into Eq. 10.20 in place of the identity matrix. The updated variance function is then used as such in the second iteration. Again, another updated variance function is calculated, and so on until the algorithm converges as does any GLM model.
Since the resulting GEE model is not based on a pure probability function, the method is called a quasi-likelihood model. Recall that we used a similar appellation for instance when an otherwise GLM variance function is multiplied by either a constant, or by another non-constant variable. In either case the working likelihood function is not based on a probability function. We therefore use a robust or sandwich variance estimator to adjust standard errors. Models using the independence correlation structure may find that such an adjustment is unnecessary, and if there is indeed no extra correlation in the data, model based standard errors are preferable. For out purposes though, I'll use empirical standard errors as the default, allowing for a more consistent measure across models.

VARIANCE ESTIMATORS

> Empirical (aka sandwich or robust/semi-robust). Consistent when the mean Model is correctly specified (if no missing data)
>
> Model-based (aka naïve). Consistent when both the mean model and the Covariance model are correctly specified.

For examples we shall use the German Health data, years from 1984-1988. This modification of the data is stored in the file *rwm_1980*.

The number of visits to the physician (*docvis*) are modeled on gender (*female*) and educational level. We wish to see if gender and education level have a bearing on the numbers of doctor visits during 1984-1988.

```
Response   =  docvis    : count from 0-121

Predictor  =  female    :(1=female;0=male)
              edlevel1 : Not HS grad   (reference)
              edlevel2 : HS grad
              edlevel3 : Univ/Coll
              edlevel4 : Grad school

panel id   =  id        : individual
time       =  year      : 1984-1988
```

DATA

```
. tab docvis

     docvis |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      7,572       38.61       38.61
          1 |      2,582       13.17       51.78
          2 |      2,357       12.02       63.80
          3 |      1,858        9.48       73.28
          4 |      1,137        5.80       79.08
          5 |        792        4.04       83.11
          6 |        683        3.48       86.60
          7 |        393        2.00       88.60
          8 |        370        1.89       90.49
          9 |        196        1.00       91.49
         10 |        347        1.77       93.26
         11 |        130        0.66       93.92
         12 |        250        1.27       95.20
         13 |         93        0.47       95.67
         14 |        109        0.56       96.23
         15 |        116        0.59       96.82
         16 |         62        0.32       97.13
         17 |         50        0.25       97.39
         18 |         51        0.26       97.65
         19 |         20        0.10       97.75
         20 |         69        0.35       98.10
         21 |         26        0.13       98.24
         22 |         30        0.15       98.39
         23 |         15        0.08       98.46
         24 |         37        0.19       98.65
         25 |         30        0.15       98.81

. tab female

     female |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |     10,187       51.95       51.95
          1 |      9,422       48.05      100.00
------------+-----------------------------------
      Total |     19,609      100.00

. tab year

       year |      Freq.     Percent        Cum.
------------+-----------------------------------
       1984 |      3,874       19.76       19.76
       1985 |      3,794       19.35       39.10
       1986 |      3,792       19.34       58.44
       1987 |      3,666       18.70       77.14
       1988 |      4,483       22.86      100.00
------------+-----------------------------------
      Total |     19,609      100.00

. tab edlevel

    edlevel |      Freq.     Percent        Cum.
------------+-----------------------------------
Not HS grad |     15,433       78.70       78.70
    HS grad |      1,153        5.88       84.58
  Coll/Univ |      1,733        8.84       93.42
Grad School |      1,290        6.58      100.00
------------+-----------------------------------
      Total |     19,609      100.00
```

10.4.2 Correlation Structures

Although GEE models are robust to the use of incorrect correlation structures, it is nevertheless preferable to select the structure most appropriate to the data or to the goal of the study. One may check the observed correlation matrix if there is no known reason to select a specific matrix based on previous clinical studies. This might not provide a definitive solution as to which is the best correlation structure for the data, but it can nevertheless inform you about which type of structure is not appropriate.

The QIC statistic can be used to quantitatively decide on the preferred correlation structure. The statistic, created by Pan (2001), is called the *quasi-likelihood under the independence model information criterion*. It is similar to the AIC statistic, but tests correlation structures within the scope of generalized estimating equations. The QICu statistic, also developed by Pan (2001a), helps the user to decide on the best subset of model predictors for a particular correlation structure. Later research has shown, however, that it is preferable to use the QIC for evaluating both situations – deciding between families and links, and deciding the best model subset. The QIC statistic is discussed at more length in Section 10.4.3.

If two of more correlation structures result in nearly the same QIC statistic, and no other factor can be used to help decide which structure to use, the preferred choice is to employ the simplest structure – one with the least parameters -- which fits the data. Hardin & Hilbe (2002) provide complete details regarding the use of both statistics.

There are a few summary guidelines that may be helpful to deciding which correlation structure to use.

1: Independent - if number of panels are small.

2: Exchangeable or Unstructured CS – if data relates to first-level clustered data

3: Autoregressive, Non-Stationary, and Stationary or m-dependent CS - when panel data relate to measurements over time periods.

4: Autoregressive CS - usually associated with longitudinal time-series data.

5: Unstructured - if panels are small and data are balanced and complete.

A listing of the major correlation structures follows. 5x5 matrix schematics of the respective correlation structures are displayed, together with representative Poisson and negative binomial GEE models. Only the lower half of the symmetric matrix is completed. I also provide additional guidelines on when each structure should be used.

INDEPENDENT CORRELATION STRUCTURE

SCHEMATIC

```
1
0    1
0    0     1
0    0     0     1
0    0     0     0     1
```

```
. xtgee docvis female edlevel2-edlevel4 , eform fam(poi) corr(indep) i(id)
robust

GEE population-averaged model                   Number of obs      =     19609
Group variable:                         id      Number of groups   =      6127
Link:                                  log      Obs per group: min =         1
Family:                            Poisson                     avg =       3.2
Correlation:                   independent                     max =         5
                                                Wald chi2(4)       =    173.84
Scale parameter:                         1      Prob > chi2        =    0.0000
Pearson chi2(19609):             204367.06      Deviance           = 119319.59
Dispersion (Pearson):             10.42211      Dispersion         =   6.08494
                                      (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |             Semi-robust
      docvis |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   1.381491   .0519136     8.60   0.000     1.283399     1.48708
    edlevel2 |   .8399478   .0548827    -2.67   0.008     .7389827    .9547075
    edlevel3 |   .7348061   .0422671    -5.36   0.000     .6564633    .8224984
    edlevel4 |   .6093514   .0460152    -6.56   0.000       .52552    .7065555
------------------------------------------------------------------------------

. xtcorr

Estimated within-id correlation matrix R:

        c1      c2      c3      c4      c5
r1  1.0000
r2  0.0000  1.0000
r3  0.0000  0.0000  1.0000
r4  0.0000  0.0000  0.0000  1.0000
r5  0.0000  0.0000  0.0000  0.0000  1.0000
```

The independent CS imposes the same structure on the GEE model as the standard variance-covariance matrix of a GLM. Observations are considered to be independent of one another. The use of this model is to set a base for evaluation of other GEE correlation structures. The structure assumes a zero correlation between subsequent measures of a subject within panels. Use this CS if the size of panels are small and if there is evidently no panel effect in the data. Again, I suggest adjusting standard errors by a robust or sandwich variance estimator.

We see that females have an approximate mean 38% increase in risk of going to the doctor than do males. Those patients with less education visit doctor more often.

For comparative purposes, I show the unexponentiated parameter coefficients.

POISSON GEE, INDEPENDENCE CORRELATION STRUCTURE, COEFFICIENTS

(Std. Err. adjusted for clustering on id)

docvis	Coef.	Semi-robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
female	.3231631	.0375779	8.60	0.000	.2495117	.3968145
edlevel2	-.1744155	.0653406	-2.67	0.008	-.3024807	-.0463503
edlevel3	-.3081486	.0575214	-5.36	0.000	-.4208884	-.1954087
edlevel4	-.4953602	.075515	-6.56	0.000	-.643367	-.3473535
_cons	1.045432	.0303437	34.45	0.000	.9859596	1.104905

NEGATIVE BINOMIAL

When modeling negative binomial (NB2) GEE models - the heterogeneity parameter is not estimated as a separate parameter. It is apportioned across panels in a manner similar to that of the conditional fixed-effects negative binomial.

Recall that GLM applications require that the negative binomial ancillary or heterogeneity parameter be entered into the algorithm as a constant. When modeling a negative binomial within the GLM framework, I have suggested earlier in this book that one should first estimate the model using a maximum likelihood negative binomial, with the resultant estimated ancillary parameter being used as the constant in the subsequent GLM-based algorithm. The parameter estimates and standard errors are identical, as we have previously observed.

The same logic maintains for GEE models. We first model a maximum likelihood negative binomial, then use the estimated ancillary parameter in the GEE algorithm. Of course, the standard NB-2 model assumes an independence correlation structure. With the GEE independence structure model, the results are straightforward. When using other correlation structures though, the independence-based ancillary parameter may not be quite as appropriate; but there is no alternative, unless we use a method similar to that employed in Table 5.4. Our methodology will be to use the NB-2 ancillary parameter for the negative binomial GEE α constant regardless of which correlation structure is used.

Stata's GEE procedure is called *xtgee*. The command works in a counter-intuitive manner when used for the negative binomial family. The value for α, the ancillary or heterogeneity parameter, when entered into the algorithm by the user, is inverted by the algorithm, producing a model with a displayed α of $1/\alpha$. However, the log-likelihood function is itself parameterized with an inverted α. Recall the discussion of the parameterization of the NB-2 PDF with $\alpha = 1/k$. In any case, the inversion in effect cancels the second inversion, maintaining the PDF that shown as Equation 5.29. The confusing part of this rests with the displayed α in the GEE output. The examples below should clarify how to handle this situation. Unfortunately there is no documentation in the Stata manuals or help files regarding this apparent inversion.

The above apparent inconsistency does not appear in other major commercial GEE implementations; e.g. SAS (an option of GENMOD), SPSS (an option of GENLIN), or S-Plus, Genstat, Sudaan, or R.

We model the same data as above for the Poisson GEE, using the negative binomial family with independence correlation structure. First, however, we model the data using a full maximum likelihood negative binomial, with a robust clustering effect on *id*.

ML NB-2 ESTIMATION – TO OBTAIN VALUE OF α

```
. nbreg docvis female edlevel2-edlevel4 , nolog  irr robust cluster(id) robust

Negative binomial regression                      Number of obs   =      19609
Dispersion          = mean                        Wald chi2(4)    =     186.16
Log pseudolikelihood = -43110.7                   Prob > chi2     =     0.0000

                                  (Std. Err. adjusted for 6127 clusters in id)
------------------------------------------------------------------------------
             |               Robust
      docvis |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   1.391372   .0513088     8.96   0.000     1.294357    1.495659
    edlevel2 |   .8331946   .0546531    -2.78   0.005     .7326763    .9475032
    edlevel3 |   .7276139   .0420571    -5.50   0.000     .6496813    .8148949
    edlevel4 |   .6001595   .0435391    -7.04   0.000     .5206136    .6918595
-------------+----------------------------------------------------------------
    /lnalpha |   .7558711   .0200549                      .7165642    .7951779
-------------+----------------------------------------------------------------
       alpha |   2.129466   .0427062                      2.047387    2.214835
------------------------------------------------------------------------------
```

Now use the estimated value of α, 2.129466, as a constant in the GEE model.

GEE NB-2 ESTIMATION – WITH α ENTERED AS A CONSTANT

```
. xtgee docvis female edlevel2-edlevel4 , i(id) corr(indep) fam(nb 2.129466)
eform robust

GEE population-averaged model                   Number of obs      =     19609
Group variable:                         id      Number of groups   =      6127
Link:                                  log      Obs per group: min =         1
Family:        negative binomial(k=.4696) ←                    avg =       3.2
Correlation:                   independent                     max =         5
                                                Wald chi2(4)       =    196.15
Scale parameter:                         1      Prob > chi2        =    0.0000

Pearson chi2(19609):              82189.19      Deviance           =  48468.79
Dispersion (Pearson):             4.191401      Dispersion         =  2.471763

                                      (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |             Semi-robust
      docvis |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   1.391372   .0517178     8.89   0.000     1.293612    1.496521
    edlevel2 |   .8331946   .0546379    -2.78   0.005     .7327026    .9474693
    edlevel3 |   .7276139   .0422568    -5.48   0.000     .6493319    .8153334
    edlevel4 |   .6001595   .0439389    -6.97   0.000     .5199344    .6927633
------------------------------------------------------------------------------
```

The two models are nearly identical, and would be if it were not for rounding error for the entered constant. Notice the inverted value of α.

```
. di 1/2.129466
.4696013              /*  INVERTED VALUE OF ALPHA */
```

that is produced in the output. If we used the inverted value as a constant instead, the parameter estimates would be mistaken:

The correlation structure is the same as for any independence model.

```
. xtcorr

Estimated within-id correlation matrix R:

        c1      c2      c3      c4      c5
r1  1.0000
r2  0.0000  1.0000
r3  0.0000  0.0000  1.0000
r4  0.0000  0.0000  0.0000  1.0000
r5  0.0000  0.0000  0.0000  0.0000  1.0000
```

The unexponentiated parameter estimates for the above independence negative binomial model are displayed as:

NB GEE, INDEPENDENCE CORRELATION STRUCTURE, COEFFICIENTS

```
                                    (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |             Semi-robust
      docvis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   .3302906   .0371703     8.89   0.000     .2574381    .4031431
    edlevel2 |  -.1824881   .0655764    -2.78   0.005    -.3110154   -.0539607
    edlevel3 |  -.3179847   .0580758    -5.48   0.000    -.4318113   -.2041582
    edlevel4 |  -.5105598    .073212    -6.97   0.000    -.6540527   -.3670669
       _cons |   1.043704   .0303465    34.39   0.000     .9842261    1.103182
------------------------------------------------------------------------------
```

The independent correlation structure imposes on the GEE model the same structure as the standard variance-covariance matrix of a generalized linear model. The observations are considered to be independent of one another. The use of this model is to set a base for evaluation of other GEE correlation structures. The structure assumes a zero correlation between subsequent measures of a subject within panels.

Use this correlation structure if the size of panels are small and if there is evidently no panel effect in the data. Theoretically we need not adjust the standard errors of independence GEE models by a robust or sandwich variance estimator. However, I suggest doing so regardless since they are normally used to compare with other correlation structures – ones which do require use of robust techniques. Rarely would an independence structure GEE be itself used for modeling data.

EXCHANGEABLE CORRELATION STRUCTURE

SCHEMATIC

```
1
a     1
a     A      1
a     A      a      1
a     A      a      A      1
```

EXAMPLE

```
 1
.29    1
.29   .29    1
.29   .29   .29    1
.29   .29   .29   .29    1
```

```
. xtgee docvis female edlevel2-edlevel4 , i(id) c(exch) fam(nb 2.129466)  eform
robust


GEE population-averaged model                   Number of obs      =     19609
Group variable:                         id      Number of groups   =      6127
Link:                                  log      Obs per group: min =         1
Family:         negative binomial(k=.4696)                     avg =       3.2
Correlation:                  exchangeable                     max =         5
                                                Wald chi2(4)       =    187.17
Scale parameter:                         1      Prob > chi2        =    0.0000

                                     (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |             Semi-robust
      docvis |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |     1.3822   .0505856     8.84   0.000     1.286527    1.484989
    edlevel2 |   .8513858   .0549046    -2.49   0.013     .7502978    .9660934
    edlevel3 |   .7263763   .0403438    -5.76   0.000     .6514556    .8099131
    edlevel4 |   .6341802   .0459056    -6.29   0.000     .5502978    .7308488
------------------------------------------------------------------------------

. xtcorr

Estimated within-id correlation matrix R:

        c1      c2      c3      c4      c5
r1  1.0000
r2  0.3324  1.0000
r3  0.3324  0.3324  1.0000
r4  0.3324  0.3324  0.3324  1.0000
r5  0.3324  0.3324  0.3324  0.3324  1.0000
```

The exchangeable correlation structure is the most commonly used structure among GEE models. It is the default for several of the major commercial software implementations and it is generally the appropriate model to use with clustered or first-level nested data. It is generally not to be used with longitudinal data.

The exchangeable correlation structure, defined in Equation 10.19, assumes that the correlations between subsequent measurements within a panel are the same, irrespective of any time interval. The value of *a*, displayed in the schematic matrix above, is a scalar. It does not vary between panels. This is why it is not appropriate for assessing longitudinal effects. Its primary use is for use with clustered data. Use this correlation structure when the observations are clustered and not collected over time.

UNSTRUCTURED CORRELATION STRUCTURE

SCHEMATIC

```
 1
C1     1
C2    C5     1
C3    C6    C8     1
C4    C7    C9    C10    1
```

EXAMPLE

```
 1
.34    1
.29   .28    1
.33   .14   .24    1
.21   .07   .11   .23    1
```

```
. xtgee docvis female edlevel2-edlevel4 , i(id) c(unst) fam(nb 2.129466)
t(year) eform robust

GEE population-averaged model                   Number of obs      =     19609
Group and time vars:                id year     Number of groups   =      6127
Link:                                   log     Obs per group: min =         1
Family:         negative binomial(k=.4696)                     avg =       3.2
Correlation:                   unstructured                    max =         5
                                                Wald chi2(4)       =    184.02
Scale parameter:                          1     Prob > chi2        =    0.0000

                                    (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |             Semi-robust
      docvis |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   1.374117   .0501441     8.71   0.000     1.279269    1.475998
    edlevel2 |   .8558149    .055595    -2.40   0.017     .7535022    .9720199
    edlevel3 |   .7259285   .0400149    -5.81   0.000     .6515887    .8087498
    edlevel4 |   .6360158   .0460806    -6.25   0.000      .551819    .7330592
------------------------------------------------------------------------------
```

```
. xtcorr

Estimated within-id correlation matrix R:

        c1      c2      c3      c4      c5
r1  1.0000
r2  0.3683  1.0000
r3  0.3390  0.3576  1.0000
r4  0.3001  0.3218  0.5045  1.0000
r5  0.3245  0.2490  0.2655  0.2988  1.0000
```

Compare the correlation structure with that of the Poisson model on the same data. Notice the similarity of the two matrices.

UNSTRUCTURED CORRELATION STRUCTURE WITH POISSON MODEL

```
. xtcorr

Estimated within-id correlation matrix R:

        c1      c2      c3      c4      c5
r1  1.0000
r2  0.3716  1.0000
r3  0.3474  0.3583  1.0000
r4  0.3071  0.3193  0.5092  1.0000
r5  0.3249  0.2519  0.2689  0.3082  1.0000
```

We would suspect that the negative binomial model itself differs little from the Poisson.

POISSON ESTIMATES

docvis	IRR	Semi-robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
female	1.364511	.0502951	8.43	0.000	1.269411	1.466736
edlevel2	.8607318	.0563454	-2.29	0.022	.7570879	.9785642
edlevel3	.7327953	.0401485	-5.67	0.000	.6581834	.8158652
edlevel4	.6463942	.0481763	-5.85	0.000	.5585431	.7480631

We do find that the two tables of parameter estimates are similar. We shall later learn how we can evaluate between the two using a QIC statistic.

In the unstructured CS all correlations are assumed to be different; correlations are freely estimated from the data. This can result in the calculation of a great many correlation coefficients for large matrices.

Because it (can have) has a different coefficient for each cell, the CS optimally fits the data. However, it loses efficiency, and hence interpretability, when models have more than about three predictors.

The number of coefficients to be estimated is based on the size of the largest panel of observations.

coefficients = p(p-1)/2

where p is the number of observations in the largest panel.

Use this CS when the size of the panels is small, there are relatively few predictors, and there are no missing values.

AUTOREGRESSIVE CORRELATION STRUCTURE

SCHEMATIC

```
 1
C^1    1
C^2   C^1    1
C^3   C^2   C^1    1
C^4   C^3   C^2   C^1    1
```

EXAMPLE

```
 1
.48    1
.23   .48     1
.11   .23    .48    1
.05   .11    .23   .48    1
```

```
. xtgee docvis female edlevel2-edlevel4, i(id) t(year) corr(ar 1) force eform
fam(nb 2.129466) robust
note:  some groups have fewer than 2 observations
       not possible to estimate correlations for those groups
       1150 groups omitted from estimation


GEE population-averaged model                      Number of obs      =     18459
Group and time vars:                 id year       Number of groups   =      4977
Link:                                    log       Obs per group: min =         2
Family:         negative binomial(k=.4696)                        avg =       3.7
Correlation:                           AR(1)                      max =         5
                                                   Wald chi2(4)       =    187.86
Scale parameter:                           1       Prob > chi2        =    0.0000

                                     (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |             Semi-robust
      docvis |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   1.402319   .0539399     8.79   0.000     1.300486    1.512127
    edlevel2 |   .8248012   .0561665    -2.83   0.005     .7217471    .9425699
    edlevel3 |   .7200313   .0439616    -5.38   0.000      .638824    .8115618
    edlevel4 |   .5970729     .04618    -6.67   0.000     .5130881    .6948046
------------------------------------------------------------------------------
```

```
. xtcorr

Estimated within-id correlation matrix R:

        c1      c2      c3      c4      c5
r1  1.0000
r2  0.3755  1.0000
r3  0.1410  0.3755  1.0000
r4  0.0529  0.1410  0.3755  1.0000
r5  0.0199  0.0529  0.1410  0.3755  1.0000
```

The autoregressive correlation structure assumes that there is a marked decrease in correlation coefficient values with the corresponding increase in measurements within panel time intervals. Each off-diagonal from the main diagonal decreases by the square of the previous diagonal. One might consider the decrease in values to be inceasing powers of the first off diagonal.

```
. di (.3755)^2
.14100025

. di ((.3755)^2)^2
.01988107
```

Large matrices produce very small coefficient values. The depiction here is the case for AR(1) models, as it is for this example, but not for all AR levels. Note also that the use of the *force* option is required when time intervals are not equal. The intervals are generally equal in this model, except when missing values exclude a particular time interval. In any case, it is best to use this option with time-based GEE models are being estimated, in case there are non-equal intervals in the data. If there are not, the results will not be affected.

Use this correlation structure when the panels are collections of data over time for the same person.

STATIONARY OR m-DEPENDENT CORRELATION STRUCTURE

SCHEMATIC

```
 1
C1     1
C2     C1     1
 0     C2     C1     1
 0     0      C2     C1     1
```

EXAMPLE

```
 1
.27     1
.18   .27     1
 0    .18    .27     1
 0     0     .18    .27     1
```

```
. xtgee docvis female edlevel2-edlevel4, i(id) t(year) corr(sta) force eform
fam(nb 2.129466) robust
note:  some groups have fewer than 2 observations
       not possible to estimate correlations for those groups
       1150 groups omitted from estimation


GEE population-averaged model                   Number of obs      =     18459
Group and time vars:               id year      Number of groups   =      4977
Link:                                  log      Obs per group: min =         2
Family:         negative binomial(k=.4696)                     avg =       3.7
Correlation:                 stationary(1)                     max =         5
                                                Wald chi2(4)       =    189.02
Scale parameter:                         1      Prob > chi2        =    0.0000

                                     (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |             Semi-robust
      docvis |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   1.402276   .0543889     8.72   0.000     1.299627    1.513033
    edlevel2 |   .8149429   .0555179    -3.00   0.003     .7130815    .9313549
    edlevel3 |   .7180757   .0438845    -5.42   0.000     .6370154     .809451
    edlevel4 |   .5936009    .046362    -6.68   0.000     .5093462    .6917927
------------------------------------------------------------------------------


. xtcorr

Estimated within-id correlation matrix R:

        c1      c2      c3      c4      c5
r1  1.0000
r2  0.3754  1.0000
r3  0.0000  0.3754  1.0000
r4  0.0000  0.0000  0.3754  1.0000
r5  0.0000  0.0000  0.0000  0.3754  1.0000
```

The stationary correlation structure specifies a constant correlation for each off-diagonal. The diagonals are then interpreted as lags or measurements. Correlations *c* lags apart are equal in value to one another, *c*+1 lags apart are also equal to one another, and so forth until a defined stop, *m*, is reached. Correlations greater than *m* are defined as zero, hence the meaning of *m*-dependent. In larger matrices the correlation structure appears as a band. Since m=1 in the above stationary model, correlations greater than the first off-diagonal have values of zero. The correlation structure for the same stationary model, but with a rank of 2, appears as:

```
        c1      c2      c3      c4      c5
r1  1.0000
r2  0.3738  1.0000
r3  0.3250  0.3738  1.0000
r4  0.0000  0.3250  0.3738  1.0000
r5  0.0000  0.0000  0.3250  0.3738  1.0000
```

The stationary correlation structure is primarily used when the off-diagonals, or lags, are thought of as time intervals.

NONSTATIONARY CORRELATION STRUCTURE

SCHEMATIC

```
 1
C1    1
C5    C2    1
 0    C6    C3    1
 0    0     C7    C4    1
```

EXAMPLE

```
 1
.99   1
.71  .84    1
 0   .78   .73    1
 0    0    .56   .70    1
```

```
. xtgee docvis female edlevel2-edlevel4, nolog i(id) t(year) corr(non 1) force
eform fam(nb 2.129466) robust
note:  some groups have fewer than 2 observations
       not possible to estimate correlations for those groups
       1150 groups omitted from estimation

GEE population-averaged model                   Number of obs      =     18459
Group and time vars:                id year     Number of groups   =      4977
Link:                                   log     Obs per group: min =         2
Family:         negative binomial(k=.4696)                     avg =       3.7
Correlation:                          nonst                    max =         5
                                                Wald chi2(4)       =    175.35
Scale parameter:                          1     Prob > chi2        =    0.0000

                                     (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |             Semi-robust
      docvis |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   1.395383   .0541659     8.58   0.000     1.293158    1.505689
    edlevel2 |   .8390004    .058883    -2.50   0.012     .7311775    .9627234
    edlevel3 |     .73923   .0457923    -4.88   0.000     .6547132    .8346571
    edlevel4 |   .5997748   .0472779    -6.49   0.000     .5139149    .6999792
------------------------------------------------------------------------------

. xtcorr

Estimated within-id correlation matrix R:

        c1      c2      c3      c4      c5
r1  1.0000
r2  0.3749  1.0000
r3  0.0000  0.3645  1.0000
r4  0.0000  0.0000  0.5121  1.0000
r5  0.0000  0.0000  0.0000  0.3024  1.0000
```

The nonstationary correlation structure is the same as the stationary except that the values of each lag or off-diagonal are not constant. Of course, correlation values beyond *m* are all 0.

Some statisticians use the nonstationary correlation structure when they have ruled out the others, but still have a limit to the range of measurement error or lags in the data.

10.4.3 GEE Goodness-of-Fit

The QIC and QICu statistics were mentioned at the outset of Section 10.4.2. We will here apply QIC evaluations to model results, testing which model is best, as defined by QIC criteria.

The QIC/QICu statistics are used to compare quasi-likelihood GEE models, in a similar manner as to how the AIC statistic is used to compare likelihood models. The QIC statistic is designed to evaluate correlation structures; eg which correlation structure best fits the data, whereas the QICu statistic is designed to evaluate which predictors best explain the response within the model framework.

The QIC statistic, expressed as QIC(R) to represent its correlation structure form, is defined as:

$$QIC(R) = -2Q(g^{-1}(x\beta_R)) + 2\text{trace}(A_I^{-1}V_{MS,R})$$

with

$2Q(g^{-1}(x\beta_R))$ = value of the computed quasi-likelihood using the model coefficients with hypothesized correlation structure R. When evaluating the quasi-likelihood, $\mu = g^{-1}(x\beta_R)$ is substituted in place of $\mu= g^{-1}()$, where $g^{-1}()$ is the model inverse link function.

A_I is the independence structure variance matrix for the model.

$V_{MS,R}$) is the modified robust or sandwich variance estimator of the model, with hypothesized structure R.

The AIC penalty term, 2p, is expressed as $2\text{trace}(A_I^{-1}V_{MS,R})$ for the QIC statistic. See Hardin & Hilbe (2003) for more details.

QICu approximates QIC when the GEE model is correctly specified. As mentioned earlier, recent research has indicated that use of the QIC alone is adequate for comparing models – comparisons of both correlation structures and predictors with the same correlation structure. The QICu should never be used for selecting working correlations. In either case, the preferred model is the one having the smaller QIC or QICu statistic.

Implementations of the QIC statistic are rare at this writing. The QIC/QICu statistics were first used to evaluate data in Hardin & Hilbe (2003), however, the code was designed to be used for a particular model. A generic SAS macro was developed in 2005, but it was found to not work well with certain types of models. James Cui developed the first general QIC/QICu procedure, publishing it together with Cui (2007). Written in Stata, he has subsequently developed the application in R. We will use his Stata implementation for the *rwm_1980* data.

Unfortunately, Cui's procedure, when used for modeling QIC negative binomial (NB-2) models, allows only an alpha of 1; ie the QIC is capable of modeling only a geometric model (NB-2; α=1). We cannot therefore use it compare the GEE negative binomial models we developed in this section. I shall therefore show how the procedure can be used to compare Poisson models of the data.

I use the *nodisplay* option to suppress a display of the model header and coefficients, showing only the relevant summary statistics. The commands for the independent and exchangeable models are shown. A summary of all relevant correlation structures follows.

INDEPENDENCE CORRELATION STRUCTURE

```
. qic docvis female edlevel2-edlevel4 , nolog i(id)  corr(indep) fam(poi) force
eform robust nodisp

              QIC and QIC_u
___________________________________________
Corr =                 indep
Family =                 poi
Link =                   log
p =                        5
Trace =               81.319
QIC =              73322.026
QIC_u =            73169.388
___________________________________________
```

EXCHANGEABLE CORRELATION STRUCTURE

```
. qic docvis female edlevel2-edlevel4 , nolog i(id)  corr(exch) fam(poi) force
eform robust nodispl

              QIC and QIC_u
___________________________________________
Corr =                  exch
Family =                 poi
Link =                   log
p =                        5
Trace =               78.498
QIC =              73211.119
QIC_u =            73064.122
___________________________________________
```

QIC/QICu SUMMARY

```
POISSON
Note: smallest value preferred model or structure

INDEPENDENT        QIC = 73322.026     QICu = 73169.338
EXCHANGEABLE       QIC = 73211.119 <-> QICu = 73064.112
UNSTRUCTURED       QIC = 73503.857     QICu = 73357.037
AUTOREGRESSIVE 1   QIC = 73732.923     QICu = 73565.748
AUTOREGRESSIVE 2   QIC = 73984.353     QICu = 73773.997
STATIONARY 1       QIC = 73710.195     QICu = 73540.892
STATIONARY 2       QIC = 74411.345     QICu = 74193.740
NONSTATIONARY 1    QIC = 73794.576     QICu = 73621.983
NONSTATIONARY 2    QIC = 74860.042     QICu = 74636.289
```

The preferred correlation structure, using QIC, is the exchangeable. This implies that the temporal structure of the data adds no insight into explaining the response. For additional comparative purposes, I modeled the negative binomial, with α=1, for the various correlation structures. The nonstationary 2 correlation structure was preferred over others, with a QIC value of 212928.774. The values for the two statistics appear as:

```
NEGATIVE BINOMIAL (α=1)
NONSTATIONARY 2     QIC = 212928.774    QICu = 212878.481
```

The values are near 3-fold higher than those of the corresponding Poisson. However, to determine whether GEE appropriately accommodates any extra correlation in the data, we must compare the above Poisson results with a negative binomial model with α=2.1295. At the current time, we cannot determine this with a general purpose procedure.

10.4.4 QUASI-LIKELIHOOD LEAST SQUARES

Many times when modeling a GEE we find that the model fails to converge. The usual reason is that the correlation matrix is not positive definite. When this occurs, most software provides the user with an error message, and stops the converging process. Justine Shults and fellow collaborators have been building on the previous work of Crowder (1995), Chaganty (1997), and Chaganty & Shults (1999) to provide those who run into such convergence problems with a way to appropriately model data within the population averaging framework. The models being developed are called *quasi-least squares*, and have been developed to be robust to modeling with non-positive definite matrices. At this writing, several of the standard GEE families with canonical links and correlation structures have been incorporated into the quasi-least squares methodology:

FAMILIES	CORRELATION STRUCTURES
Gaussian (link=identity)	Autoregresive 1 (AR 1)
Bernoulli/Binomial (link=logit)	Stationary 1 (sta 1) [tridiagonal]
Poisson (link-log)	Exchangeable (exc) [equicorrelated]
	Markov (Markov)

Current software implementations exist in Stata, with R being developed at this writing. Other matrices are being developed, as well as extended options.

Again quasi-least squares (QLS) models are alternatives to GEE when the model matrix is not positive definite, or when a Markov correlation structure is desired for the GEE-like model. The Markov correlation structure is not available in any current GEE software. The QLS algorithm calls an underlying GEE procedure, making adjustments to the covariance matrix at each iteration such that it provides positive definite results.

A QLS model using the Markov correlation structure is given below. The software does not allow exponentiated coefficients, but does provide for robust, jackknifed and bootstrapped standard errors.

MARKOV CORRELATION STRUCTURE

```
. xtqls docvis female edlevel2-edlevel4 , i(id) t(year) c(Markov) f(poi)
vce(robust)                                             =========

GEE population-averaged model                   Number of obs      =     19609
Group and time vars:               id year      Number of groups   =      6127
Link:                                  log      Obs per group: min =         1
Family:                            Poisson                     avg =       3.2
Correlation:             fixed (specified)                     max =         5
                                                Wald chi2(4)       =    162.91
Scale parameter:                         1      Prob > chi2        =    0.0000

                                     (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |             Semi-robust
      docvis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   .3157038   .0372515     8.47   0.000     .2426922    .3887153
    edlevel2 |  -.1651521    .065258    -2.53   0.011    -.2930555   -.0372487
    edlevel3 |  -.3030402   .0553477    -5.48   0.000    -.4115197   -.1945607
    edlevel4 |  -.4391883   .0768606    -5.71   0.000    -.5898323   -.2885444
       _cons |    1.03804   .0304523    34.09   0.000     .9783551    1.097726
------------------------------------------------------------------------------

. xtcorr

Estimated within-id correlation matrix R:

        c1      c2      c3      c4      c5
r1  1.0000
r2  0.3883  1.0000
r3  0.1508  0.3883  1.0000
r4  0.0586  0.1508  0.3883  1.0000
r5  0.0227  0.0586  0.1508  0.3883  1.0000
```

Several points can be observed with the above Markov model. First, the correlation structure is nearly identical to the Poisson Autoregressive 1 model.

```
POISSON AR 1
        c1      c2      c3      c4      c5
r1  1.0000
r2  0.3791  1.0000
r3  0.1438  0.3791  1.0000
r4  0.0545  0.1438  0.3791  1.0000
r5  0.0207  0.0545  0.1438  0.3791  1.0000
```

Second, each exponentiated coefficient is consistent with a Poisson GEE model with an exchangeable correlation structure.

```
 MARKOV                  EXCHANGEABLE
. di exp(.3157038)
1.371224                    1.373005

. di exp(-.1651521)
.84776475                   .8558109

. di exp(-.3030402)
.73856941                   .7330806

. di exp(-.4391883)
.6445594                    .6445709
```

FEASIBILITY

GEE models fail when the correlation matrix is not positive definite. This occurs when values of the matrix exceed structural assumptions. We use the term *feasible* to designate values that are within these assumptions.

The feasible region of α (top-most non-diagonal value) for specified correlation structures that can be adjusted by QLS are defined as:

EXCHANGEABLE: $(-1/n_m-1)$; 1 , where n_m is the maximum value of n_i over i=1,2,…, m. That is, it is the value of the panel with the greatest number of observations.

For the German health data (*rwm_1980*), we can calculate the highest value for a panel by the following:

```
. sort id
. by id: egen cnt=count(id)
. su cnt

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         cnt |     19609    3.865725    1.218388          1          5
```

n_m = 5. Therefore the left-side range is -1/(5-1) = -.25. The range of feasibility is then (-.25; 1). Since α = .3362, α is feasible.

AR-1: -1; 1
Note: a negative α, eg -0.90, produces within subject correlations that alternate in sign. EX: the correlation between the 1st and 2nd measures of a subject will be

$$(-0.90)^{|2-1|} = -0.90;$$

correlation between the 1st and 3rd measures will be $(-0.90)^{|3-1|} = 0.81$.

STATIONARY (TRIDIAGONAL): $(-1/c_m; 1/c_m)$ where $c_m = 2\sin[(\pi[n_m-1])/(2[n_m+1])]$. For large n, the interval is approximately (-0.5;0.5). For the German data,

$$c_m = 2*\sin((\_pi*4)/(2*6)) = 1.73205$$

or 1.73. Range = -1/1.73; 1/1.73 or -0.578; 0.578

EXAMPLE

```
xtqls docvis female edlevel2-edlevel4 , i(id) t(year) c(sta 1) f(poi)
vce(robust)

   . xtcorr
   Estimated within-idcode correlation matrix R:

           c1      c2      c3  .  .  .
   r1  1.0000
   r2  0.5571  1.0000
   r3  0.0000  0.5571  1.0000
    .   .   .
```

which is outside the range of feasibility, but only by a little, producing a correlation matrix that is not positive definite. Hence is fails if modeled using GEE.

MARKOV: (-1; 1)

10.4.5 SUMMARY COMMENT ON GEE MODELS

The advantage of GEE over random-effects models relates to the ability of GEE to allow specific correlation structures to be assumed within panels. Parameter estimates are calculated without having to specify the joint distribution of the repeated observations. In random-effects models a between-subject effect represents the difference between subjects conditional on them having the same random effect. Such models are thus termed conditional or subject-specific models. GEE parameter estimates represent the average difference between subjects, thus are known as marginal, or population-averaged models. Which to use depends on the context; i.e. on the goals of the study.

Our final section will discuss an emerging area of study – multilevel models. Two of the basic multilevel models are random intercept and random coefficient models. Since other more complex multilevel models are built on their bases, we shall restrict our discussion to these two models.

10.5 Multilivel Negative Binomial Models

Multilevel models are sometimes called hierarchical models, particularly in educational and social science research. However, the majority of statisticians now tend to draw a distinction between multilevel and hierarchical models, primarily because of the manner in which the methods define order of levels, or nesting. Regardless, the idea behind multilevel models is to model the dependence that exists between nested levels in the data. For instance, we may model visits to the doctor within groups of different hospitals. Unlike GEE models, multilevel models are not based on the framework of generalized linear models. Rather, they are an extension to the random-effects models we discussed in the previous section.

Until recently, nearly all discussion, and application, of multilevel models have been of continuous response models. Binary response models, especially logistic models, were introduced about ten years ago. Only in the last few years have Poisson models been discussed within the domain of multilevel regression. Negative binomial models have been largely ignored. As of this writing, only LIMDEP provides the capability of modeling negative binomial random coefficient models. We shall use it in this section to examine both random intercept and random coefficient models. Random intercept models are considered to be the most elementary, and fundamental, of multilevel models.

10.5.1 Random Intercept Negative Binomial Models

Suppose that we are studying student performance on statewide exit examinations. Schools are funded, and school administrators retained, on the basis of how their students perform on the examination. When studying such data it is evident that student performance within schools is more highly correlated than between schools. Likewise, average school performance within various school districts is likely correlated. Here we have student performance nested in schools, which itself is nested within school districts. Correlation effects exist for students within schools, and correlation effects exist for schools within school districts. We may add other levels, such as types of school programs within schools, but three levels are sufficient to see the problem. The levels of dependency must be adjusted by the models if the resulting levels of overdispersion are to be accommodated.

Multilevel models handle nested levels of dependency by allowing the regression coefficients to vary within levels. Because the multilevel algorithm permits the coefficients to vary, statisticians have come to use the more specific term, *random coefficient model*, for this type of multilevel model.

The most basic random coefficient model is one in which only the regression intercept is allowed to vary. Such a model is called a *random intercept model.* It is a subset of random coefficient models.

The random intercept model may be expressed as an equation:

$$y_{ik} = \beta_{0i} + \beta_1 X + \epsilon_{ik} \qquad \text{Eq. 10.21}$$

y_{it} is the response for individual *i* in group *k*, or at time *k* (which we would change to *t).* β_{0i} refers to the regression intercept, varying over the individual, *i*. $\beta_1 X$ is the coefficient for predictor *x*, and ϵ_{ik} is the error term, varying over both individuals and groups.

The following example comes from the German Health data set. It is from 1996, prior to the later Reform data that we have used in previous discussions. Model variables include the following:

Response:

docvis	The number of visits to the doctor by a patient recorded over seven time periods.

Predictors include:

age	Age (25-64)
female	1=Female; 0=Male
educ	Years of schooling (7-18)
married	1=Married; 0 = Not married
hhninc	Net monthly house income in Marks/10000 (0-30.67)
hsat	Health satisfaction evaluation (0-10)
_groupti	periods in which data were recorded (1-7)

The majority of LIMDEP procedures are available in a point-and-select format. Random coefficient models, however, require the use of the command line. The code for the model is, followed by output,

```
--> Negb;lhs=docvis;rhs=one,age,female,educ,married,hhninc,hsat;pds=_groupti
    ;rpm;fcn=one(n);pts=20;halton $
```

```
+---------------------------------------------+
| Random Coefficients  NegBnReg Model         |
| Maximum Likelihood Estimates                |
| Model estimated: May 29, 2006 at 05:35:36PM.|
| Dependent variable               DOCVIS     |
| Weighting variable                 None     |
| Number of observations             6209     |
| Iterations completed                 24     |
| Log likelihood function       -12723.32     |
| Number of parameters                  9     |
| Info. Criterion: AIC =          4.10125     |
|   Finite Sample: AIC =          4.10125     |
| Info. Criterion: BIC =          4.11101     |
| Info. Criterion:HQIC =          4.10463     |
| Restricted log likelihood     -46669.83     |
```

```
| Chi squared                      67893.02     |
| Degrees of freedom                      1     |
| Prob[ChiSqd > value] =         .0000000     |
| Unbalanced panel has     887 individuals.   |
| Negative binomial regression model          |
| Simulation based on  20 Halton draws        |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Nonrandom parameters
 AGE            .01798893      .00163076    11.031   .0000    44.3351586
 FEMALE         .39665379      .03117120    12.725   .0000     .42277339
 EDUC         -.04380941      .00829631    -5.281   .0000    10.9408707
 MARRIED        .09475447      .04169113     2.273   .0230     .84538573
 HHNINC       -.00497977      .08748656     -.057   .9546     .34929630
 HSAT         -.21528106      .00596570   -36.087   .0000    6.69640844
          Means for random parameters
 Constant      1.52593656      .13369248    11.414   .0000
          Scale parameters for dists. of random parameters
 Constant       .80989785      .01676004    48.323   .0000
          Dispersion parameter for NegBin distribution
 ScalParm      1.18747048      .02671618    44.448   .0000


Implied standard deviations of random parameters

Matrix S.D_Beta has  1 rows and  1 columns.
               1
        +--------------
       1|     .80990
```

ScalParm is the negative binomial heterogeneity parameter. At 1.19, overdispersion still appears to remain in the data.

It appears that all predictors but *hhninc* significantly contribute to the model.

10.5.2 Random Coefficient Negative Binomial Models

We previously provided the formula for a random intercept model, where the intercept varies over periods. The random coefficient model expands this analysis to allow the regression coefficients to vary. The equation for a model in which the coefficient, β_{1i}, varies, but not the intercept, can be expressed as:

$$y_{ik} = \beta_0 + \beta_{1i} x + \epsilon_{ik} \qquad \text{Eq. 10.22}$$

We may allow both the intercept and coefficient to vary, giving us:

$$y_{ik} = \beta_{0i} + \beta_{1i} x + \epsilon_{ik} \qquad \text{Eq. 10.23}$$

Both of the above models are random coefficient models. They are also referred to as random parameter and random intercept models. More complex models can exist depending on the number of nested levels in the data.

We now use a random coefficient model on the same data, allowing the coefficient on health satisfaction, *hsat*, to vary. *hsat* has eleven levels.

```
--> Negb;lhs=docvis;rhs=one,age,female,educ,married,hhninc,hsat;pds=_groupti
    ;rpm;fcn=one(n),hsat(n);cor;pts=20;halton $

+---------------------------------------------+
+---------------------------------------------+
| Random Coefficients  NegBnReg Model         |
| Maximum Likelihood Estimates                |
| Model estimated: May 29, 2006 at 05:38:24PM.|
| Dependent variable              DOCVIS      |
| Weighting variable                None      |
| Number of observations            6209      |
| Iterations completed                28      |
| Log likelihood function      -12694.26      |
| Number of parameters                11      |
| Info. Criterion: AIC =         4.09253      |
|   Finite Sample: AIC =         4.09254      |
| Info. Criterion: BIC =         4.10446      |
| Info. Criterion:HQIC =         4.09666      |
| Restricted log likelihood    -46669.83      |
| Chi squared                   67951.15      |
| Degrees of freedom                   3      |
| Prob[ChiSqd > value] =        .0000000      |
| Unbalanced panel has     887 individuals.   |
| Negative binomial regression model          |
| Simulation based on  20 Halton draws        |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Nonrandom parameters
 AGE           .01680197      .00159059    10.563   .0000    44.3351586
 FEMALE        .37039825      .02990835    12.384   .0000     .42277339
 EDUC         -.03147513      .00813808    -3.868   .0001    10.9408707
 MARRIED       .09847528      .04141046     2.378   .0174     .84538573
 HHNINC        .01194585      .08915694      .134   .8934     .34929630
          Means for random parameters
 Constant     1.65242649      .12955696    12.754   .0000
 HSAT         -.24630006      .00653340   -37.699   .0000    6.69640844
          Diagonal elements of Cholesky matrix
 Constant      .78345690      .04466106    17.542   .0000
 HSAT          .11609985      .00249648    46.505   .0000
          Below diagonal elements of Cholesky matrix
 lHSA_ONE     -.08790569      .00670183   -13.117   .0000
          Dispersion parameter for NegBin distribution
 ScalParm     1.24095398      .02905699    42.708   .0000
Implied covariance matrix of random parameters
Matrix Var_Beta has  2 rows and  2 columns.
               1             2
        +---------------------------
       1|     .61380      -.06887
       2|    -.06887       .02121
```

```
Implied standard deviations of random parameters
Matrix S.D_Beta has  2 rows and  1 columns.
               1
        +--------------
       1|     .78346
       2|     .14562
Implied correlation matrix of random parameters
Matrix Cor_Beta has  2 rows and  2 columns.
               1             2
        +---------------------------
       1|    1.00000      -.60364
       2|    -.60364      1.00000
```

The negative binomial by construction already picks up some heterogeneity that manifests itself in the overdispersion. The random coefficients formulation is an extension that gathers together other time invariant heterogeneity across individuals. Random coefficient models allow the randomness of the coefficients to explain the heterogeneity across individuals as well as the heterogeneity across groups. This heterogeneity, in sum, results in the differences found in the responses, y_{ik}, due to changes in the predictors. The fact that levels of nesting, or of additional unexplained heterogeneity, can be explained by these models, make them attractive to those who wish to model count data having such a structure.

The only caveat to keep in mind when using negative binomial random coefficient models is to be careful of over-specification. That is, multiple adjustments are being given to the otherwise Poisson counts. Care must be taken to assure that our model does not make too much adjustment.

10.6 Summary

Panel data, consisting of data in clustered and longitudinal format, violates the basic maximum likelihood assumption of the independence of observations. Whether the data is clustered by groups or recorded by observations over a periods of time, the same methods are used to estimate the respective panel models.

In this chapter I presented overviews of the foremost panel models: unconditional and conditional fixed effects models, random effects models, and generalized estimating equations. Multilevel mixed models is a comparatively new area of research, with multilevel count models being the most recent. LIMDEP is the only commercial software supporting negative binomial linear mixed models, with its initial application in 2006. The software limits use to random intercept and the more detailed random coefficient negative binomial models. Examples of both models are provided in this chapter.

Hierarchical GLMs, called HGLMs, and double HGLMs have recently been developed, primarily by John Nelder and Yougjo Lee. However, they have not employed HGLM theory to the negative binomial, and we do not discuss them here. HGLMs are supported only by GENSTAT software.

REFERENCES

Akaike, H. (1973), Information theory and extension of the maximum likelihood principle, in *Second international Symposium on Information Theory*, Ed B.N. Petrov and F. Csaki, 267-281, Budapest, Hungary: Akademiai Kiado.

Allison, P.D. and R Waterman (2002). Fixed-Effects Negative Binomial Regression Models, unpublished manuscript.

Amemiya, T. (1984). Tobit Models: A Survey, *Journal of Econometrics*, **24**: 3-61.

Anscombe, F.J. (1948). The transformations of Poisson, binomial, and negative binomial data, *Biometrika* **35**: 246-254.

Anscombe, F.J. (1949). The statistical analysis of insect counts based on the negative binomial distribution, *Biometrics* **5**: 165-173.

Anscombe, F.J. (1972). Contribution to the discussion of H. Hotelling's paper, *Journal of the Royal Statistical Society – Series B* **15(1)**: 229-230.

Bartlett, M.S. (1947). The use of transformations, *Biometrics* **3**: 39-52.

Beall, G. (1942). The Transformation of data from entomological field experiments so that that analysis of variance becomes applicable, *Biometrika* **29**: 243-262.

Blom, G. (1954). Transformations of the binomial, negative binomial, Poisson, and χ^2 distributions, *Biometrika* **41**: 302-316.

Breslow, N.E. (1984). Extra-Poisson Variation in Log-linear Models, *Applied Statistics*, **33, 1**: 38-44.

Cameron, A C. and P.K. Trevedi (1986). Econometric Models based on Count Data: Comparisons and Applications of Some Estimators, *Journal of Applied Econometrics*, **1**: 29-53.

Cameron, A C. and P.K. Trevedi (1990). Regression-based Tests for Overdispersion in the Poisson Model, *Journal of Econometrics*, **46**: 347-364.

Cameron, A C. and P.K. Trevedi (1998). *Regression Analysis of Count Data*, New York: Cambridge University Press.

Collett, D. (1989). *Modelling Binary Data*, London:Chapman & Hall

Consul, P. and G Jain (1973). A generalization of the Poisson distribution, *Technometrics*, **15**: 791-799.

Consul, P.C. and R.C. Gupta (1980). The generalized binomial distribution and its characterization by zero regression, *SIAM Journal of Applied Mathematics*, **39(2)**: 231-237.

Consul, P. and F. Famoye (1992). Generalized Poisson regression model, *Communications in statistics – Theory and Method*, **21**: 89-109.

Cui, J. (2007), QIC program and model selection in GEE analysis, *Stata Journal* Vol 7: Num 2, pp: 209-220.

Drescher, D. (2005). Alternative distributions for observation driven count series models, *Economics Working Paper No 2005-11*, Christian-Albrechts-Universitat, Kiel, Germany.

Edwards, A. W. F. (1972). *Likelihood*, Baltimore, MD: John Hopkins University Press.

Eggenberger F., and G. Polya (1923). "Uber die Statistik Verketteter Vorgange," Journal of Applied Mathematics and Mechanics, 1:279-289.

Englin, J. and J. Shonkwiler (1995). Estimating social welfare using count data models: An application under conditions of endogenous stratification and truncation. *Review of Economics and Statistics* **77**: 104-112.

Fair, R. (1978). A Theory of Extramarital Affairs, Journal of Political Economy, **86**: 45-61.

Famoye, F (1995). “Generalized Binomial Regression Model,” Biometrical Journal 37, 5:581-594.

Famoye, F. and K. Singh (2006). Zero-Truncated Generalized Poisson Regression Model with an Application to Domestic Violence, *Journal of Data Science* **4**: 117-130.

Faraway, J. (2006). *Extending the Linear Model with R*, Boca Raton, FL: Chapman & Hall/CRC Press.

Frees, E. (2004). *Longitudinal and Panel Data*, Cambridge,U.K.: Cambridge University Press.

Goldberger, A.S. (1983). Abnormal selection bias. In *Studies in Econometrics, Time Series, and Multivariate Statistics*, ed. Karlin, S., T. Amemiya, and L.A. Goodman, New York: Academic Press, pp. 67-85.

Gould, W., J. Pitblado, and W. Scribney (2006). *Maximum likelihood estimation with Stata*, Third Edition, College Station: Stata Press.

Greene, W.H. (1992). Statistical Models for Credit Scoring, Working Paper, Department of Economics, Stern School of Business, New York University.

Greene, W.H. (1994). Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models, **EC-94-10**, Department of Economics, Stern School of Business, New York University.

Greene, W.H. (2003). *Econometric Analysis*, Fifth Edition, New York: Macmillan.

Greene, W.H. (2006). *LIMDEP Econometric Modeling Guide*, Version 9, Plainview, NY: Econometric Software Inc.

Greene, W.H. (2006a). A General Approach to Incorporating 'Selectivity' in a Model, Working Paper, Department of Economics, Stern School of Business, New York University.

Greenwood, M., and G.U. Yule (1920). "An Inquiry into the Nature of Frequency Distributions of Multiple Happenings, with Particular Reference to the Occurrence of Multiple Attacks of Disease or Repeated Accidents," *Journal of the Royal Statistical Society* A, 83:255-279.

Gurmu, S. and P. K. Trivedi (1992). Overdispersion tests for truncated Poisson regression models, *Journal of Econometrics*, **54**: 347-370.

Hardin, J.W. and J.M. Hilbe (2001). *Generalized Linear Models & Extensions*, College Station, TX: Stata Press

Hardin, J.W. and J.M. Hilbe (2002). *Generalized Estimating Equations*, Boca Raton, FL: Chapman & Hall/CRC Press.

Hardin, J.W. and J.M. Hilbe (2007). *Generalized Linear Models & Extensions*, Second Edition, College Station, TX: Stata Press

Hardin, J.W. (2003). The Sandwich Estimate of Variance, in *Advances in Econometrics: Maximum Likelihood of Misspecified Models: Twenty Years Later*, ed. Fomby, T. and C. Hill, Elsevier, **17**: 45-73.

Hausman, J., B. Hall, and Z. Griliches (1984). Econometric Models for Count Data with an Application to the Patents – R&D Relationship, *Econometrica*, **52**: 909-938.

Heckman, J. (1979). Sample Selection Bias as a Specification Error, *Econometrica*, **47**: 153-161.

Heilbron, D. (1989). Generalized Linear Models for Altered Zero Probabilities and Overdispersion in Count Data, Technical Report, Department of Epidemiology and Biostatistics, University of California, San Francisco.

Hilbe, J. M. (1993). Log Negative Binomial Regression as a Generalized Linear Model, Technical Report COS **93/94-5-26**, Department of Sociology, Arizona State University.

Hilbe, J. (1993a). Generalized linear models, *Stata Technical Bulletin*, STB-11, sg16.

Hilbe, J. (1993). Generalized linear models using power links, *Stata Technical Bulletin*, STB-12, sg16.1

Hilbe, J. (1994). Negative binomial regression, *Stata Technical Bulletin*, STB-18, sg16.5

Hilbe, J. M. (1994). Generalized linear models, *The American Statistician*, **48(3)**: 255-265.

Hilbe, J. and B. Turlach (1995). Generalized linear models. In *XploRe: An Interactive Statistical Computing Environment*, ed. W. Hardle, S Klinke, and B.Tulach, pp. 195-222. New York: Springer-Verlag.

Hilbe, J. and D. Judson (1999). Right, Left, and uncensored Poisson regression, in Stata Technical Bulletin Reprints, 186-189. College Station, TX: Stata Press

Hilbe, J. and W. Greene (2007). Count Response Regression Models, in ed. C.R. Rao, J.P Miller, and D.C.Rao, *Epidemiology and Medical Statistics,* Elsevier Handbook of Statistics Series. London, UK: Elsevier.

Hilbe, J. (2000). Two-parameter log-gamma and log-invese Gaussian models, in *Stata Technical Bulletin Reprints*, 118-121. College Station,TX: Stata Press.

Hilbe, J. (2005). CPOISSON: Stata module to estimate censored Poisson regression, Boston College of Economics, Statistical Software Components, http://ideas.repec.org/c/boc/bocode/s456411.html

Hilbe, J. (2005), CENSORNB: Stata module to estimate censored negative binomial regression as survival model, Boston College of Economics, Statistical Software Components, http://ideas.repec.org/c/boc/bocode/s456508.html

Hoffman, J. (2004). *Generalized Linear Models*, Boston, MA: Allyn and Bacon.

Hosmer, D. and S. Lemeshow (2003), *Applied Logistic Regression*, second edition, New York:Wiley

Huber, P.J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 221-233.

Jain, G.C. and P.C. Consul (1971). A generalized negative binomial distibution, *SIAM Journal of Applied Mathematics*, **21(4)**: 501-513.

Karim, M.R. and S. Zeger (1989). A SAS macro for longitudinal data analysis, Department of Biostatistics, The John Hopkins University: Technical Report **674**.

Katz, E. (2001). Bias in conditional and Unconditional Fixed Effects Logit Estimation, *Political Analysis*, **9:4**: 379-384.

King, G. (1988). Statistical models for political science event counts: Bias in conventional procedures and evidence for the exponential Poisson regression model. *American Journal of Political Science*, **32**: 838-863.

King, G. (1989). Event count models for international relations: generalizations and applications. *International Studies Quarterly*, **33**: 123-147.

Lambert, D. (1992). Zero-inflated Poisson Regression with an Application to Defects in Manufacturing, *Technometrics*, **34**: 1-14.

Lancaster, T. (2002). Orthogonal Parameter and Panel Data, *Review of Economic Studies* **69**: 647-666.

Lawless, J. F. (1987). Negative binomial and mixed Poisson regression, *The Canadian Journal of Statistics*, **15, 3**: 209-225.

Lee, Y, J. Nelder, and Y. Pawitan (2006). *Generalized Linear Models with Random Effects*, Boca Raton, FL.: Chapman & Hall/CRC Press.

Liang, K.-Y. and S.Zeger (1986). Longitudinal data analysis using generalized linear models, *Biometrika*, **73**: 13-22.

Long, J.S., (1997). *Regression Models for Categorical and Limited Dependent Variables*, Thousand Oaks, CA: Sage Pub.

Long, J. S. and J. Freese (2003, 2006). *Regression Models for Categorical Dependent Variables using Stata*, Second Edition, College Station, TX: Stata Press.

Loomis, J. B. (2003). Travel cost demand model based river recreation benefit estimates with on-site and household surveys: Comparative results and a correction procedure, *Water Resources Research*, **39(4)**: 1105.

Marquardt, D.W. (1963). An algorithm for least-squares estimation of nonlinear parameters, *Journal of the Society for Industrial and Applied Mathematics*, **11**: 431-441.

McCullagh P. (1983). Quasi-likelihood functions, *Annals of Statistics* **11**: 59-67.

McCullagh, P. and J.A. Nelder (1989). *Generalized Linear Models*, Second Edition, New York: Chapman & Hall.

Mullahy, J. (1986). Specification and Testing of Some Modified Count Data Models, *Journal of Econometrics*, **33**: 341-365.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data, *Econometrica*, **46**: 69-85.

Mwalili, S., E. Lesaffre, and D. DeClerk (2005). The Zero-Inflated Negative Binomial Regresson Model with Correction for Misclassification: An example in Caries Research, Technical Report 0462, LAP Statistics Network Interuniversity Attraction Pole. Catholic University of Louvain la Neuve, Belgium. www.stat.ucl.ac.be/IAP.

Nelder, J.A. and R.W.M. Wedderburn (1972). Generalized linear models, *Journal of the Royal Statistical Society*, Series A, **135(3)**: 370-384.

Nelder, J. and D. Pregibon (1987). "An Extended Quasi-likelihood Function," *Biometrika* 74:221-232.

Nelder, J.A. and Y. Lee (1992). Likelihood, Quasi-likelihood, and Pseudo-likelihood: Some Comparisons, *Journal of the Royal Statistical Society* Series B, **54**: 273-284.

Nelder, J.A. (1994). Generalized Linear Models with Negative Binomial or Beta-Binomial Errors, unpublished manuscript.

Nelson, D. L. (1975). Some remarks on generalized of negative binomial and Poisson distributions, *Technometrics*, **17**: 135-136.

Neyman, J, and E. L. Scott (1948). Consistent Estimation from Partially Consistent Observations, *Econometrica*, **16, 1**: 1-32

Pan, W. (2001). Akaike's information criterion in generalized estimating equations, *Biometrics* **57**: 120-125.

Pan, W. (2001a). On the robust variance estimator in generalized estimating equations, Biometrika **88(3)**: 901-906.

Rabe-Hesketh, S. and A. Skrondal (2004). *Generalized Latent Variable Modeling*, Boca Raton, FL.: Chapman & Hall/CRC Press.

Rabe-Hesketh, S. and A. Skrondal (2005). *Multilevel and Longitudinal Modeling Using Stata*, College Station: Stata Press.

Shaw, D. (1988). On-Site Samples' Regression, *Journal of Econometrics* **37**: 211-223.

Shults, J., S.J. Ratcliffe, and M. Leonard (2007), Improved generalized estimating equation analysis via xtqls for quasi-least squares in Stata, *Stata Journal*, Vol 7, Num 2, pp: 147-166.

Terza, J.V. (1998). A Tobit-Type Estimator for the Censored Poisson Regression Model, *Econometric Letters*, **18**: 361-365.

Thall, P. and S Vail (1990). "Some Covariance Models for Longitudinal Count Data and Overdispersion,' Biometrika, 46: 657-671.

Twist, J.W.R. (2003). *Applied Longitudinal Data Analysis for Epidemiology*, Cambridge, UK.: Cambridge University Press.

Vadeby, A. (2002). Estimation in a Model with Incidental Parameters, LiTH-MAT-R-2002-02 working paper.

Venables, W. and B. Ripley (2002). *Modern Applied Statistics with S* , Fourth Edition, New York: Springer-Verlag.

Vuong, Q.H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, **57**: 307-333.

Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika* **61**: 439-47.

White H. (1980). A heteroskestasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48(4)**: 817-838.

Winkelmann, R (1995). "Duration Dependence and Dispersion in Count-Data Models," *Journal of Business and Economic Statistics*, 13:467-474.

Winkelmann, R. (2003). *Econometric Analysis of Count Data*. Fourth Edition, Heidelberg, Germany: Springer-Verlag.

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

Woutersen, T. (2002). Robustness Against Incidental Parameters , University of Western Ontario, Department of Economics Working papers, 20028.

Yang, Z., J. Hardin, and C. Abby (2006). A Score Test for Overdispersion Based on Generalized Poisson Model. Unpublished manuscript.

Zeger, S.L., K.-Y. Liang, and P.S. Albert (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44**: 1049-1060.