**Example of two-part modeling: “Negative Binomial Hurdle”**

**data** e;

**Step 1: Separate zero-inflated outcome into 2 new outcomes:**  
  
**DVp** = prevalence, where 0=yes and 1=no (backwards because default predicts 0, so to predict the 1 instead, we recode the data)  
  
**DVf** = frequency, where . = no and amount if present

set d;

if DV=**0** then do; DVp=**1**; DVf=**.**; end;

if DV ge **1** then do; DVp=**0**; DVf=DV; end;

**run**;

\*change back DVp from 1 to 0 and 0 to 1;

**data** f;

**Step 2: Stack 2 new outcomes into single outcome and create necessary identifier variables for which is which**  
**DVpf** = new combined outcome for prevalence and frequency  
  
**INTp/f** = indicator variables to use in specifying fixed and random effects  
  
**dist** = which distribution to use for each (from PROC GLIMMIX manual)

set e;

CYear=Year-**1**;

DVpf=Dvp; dist=**4**; INTp=**1**; INTf=**0**; output;

DVpf=DVf; dist=**7**; INTp=**0**; INTf=**1**; output;

**run**;

\*4=logit distribution;

\*7=negative binomial distribution;

**Step 3: Fit multivariate model 🡪 Here, fixed linear random intercept for both DVs  
DVp uses logit link, binary distribution, and DVf uses log link and negative binomial distribution**

**PROC** **GLIMMIX** DATA=f NOCLPRINT NOITPRINT NAMELEN=**100** METHOD=QUAD;

CLASS nysid year;

MODEL DVpf = INTp INTp\*cyear INTf INTf\*cyear/ NOINT SOLUTION DIST=BYOBS (dist);

RANDOM INTp INTf / TYPE=UN SUBJECT=nysid;

Contrast "2 DF test of change" INTp\*cyear **1**, INTf\*cyear **1** /chisq;

**RUN**;

The GLIMMIX Procedure

| **Model Information** | |
| --- | --- |
| **Data Set** | WORK.F |
| **Response Variable** | DVpf |
| **Response Distribution** | Multivariate |
| **Link Function** | Multiple |
| **Variance Function** | Default |
| **Variance Matrix Blocked By** | nysid |
| **Estimation Technique** | Maximum Likelihood |
| **Likelihood Approximation** | Gauss-Hermite Quadrature |
| **Degrees of Freedom Method** | Containment |

|  |  |
| --- | --- |
| **Number of Observations Read** | 8316 |
| **Number of Observations Used** | 5250 |

| **Response Profile** | | |
| --- | --- | --- |
| **Ordered Value** | **DVpf** | **Total Frequency** |
| **1** | **0** | 2264 |
| **2** | **1** | 722 |
| **The GLIMMIX procedure is modeling the probability that DVpf='0'.** | | |

| **Dimensions** | |
| --- | --- |
| **G-side Cov. Parameters** | 3 |
| **R-side Cov. Parameters** | 1 |
| **Columns in X** | 4 |
| **Columns in Z per Subject** | 2 |
| **Subjects (Blocks in V)** | 681 |
| **Max Obs per Subject** | 12 |

| **Optimization Information** | |
| --- | --- |
| **Optimization Technique** | Dual Quasi-Newton |
| **Parameters in Optimization** | 8 |
| **Lower Boundaries** | 3 |
| **Upper Boundaries** | 0 |
| **Fixed Effects** | Not Profiled |
| **Starting From** | GLM estimates |
| **Quadrature Points** | 5 |

|  |
| --- |
| Convergence criterion (GCONV=1E-8) satisfied. |

| **Fit Statistics** | |
| --- | --- |
| **-2 Log Likelihood** | 16761.16 |
| **AIC (smaller is better)** | 16777.16 |
| **AICC (smaller is better)** | 16777.19 |
| **BIC (smaller is better)** | 16813.35 |
| **CAIC (smaller is better)** | 16821.35 |
| **HQIC (smaller is better)** | 16791.17 |

| **Fit Statistics for Conditional Distributions** | | | |
| --- | --- | --- | --- |
| **Description** | **Binary** | **Negative Binomial** | **Total** |
| **-2 log L(DVpf | r. effects)** | 1874.44 | 12873.86 | 14748.30 |
| **Pearson Chi-Square** | 1954.81 | 1862.61 | 3817.41 |
| **Pearson Chi-Square / DF** | 0.65 | 0.82 | 0.73 |

| **Covariance Parameter Estimates** | | | |
| --- | --- | --- | --- |
| **Cov Parm** | **Subject** | **Estimate** | **Standard Error** |
| **UN(1,1)** | **nysid** | 4.2771 | 0.5256 |
| **UN(2,1)** | **nysid** | 1.2192 | 0.1102 |
| **UN(2,2)** | **nysid** | 0.4340 | 0.03715 |
| **Scale** |  | 0.2470 | 0.01224 |

| **Solutions for Fixed Effects** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Effect** | **Estimate** | **Standard Error** | **DF** | **t Value** | **Pr > |t|** |
| **INTp** | 2.5782 | 0.1564 | 680 | 16.48 | <.0001 |
| **INTp\*CYear** | -0.05970 | 0.008349 | 3932 | -7.15 | <.0001 |
| **INTf** | 2.2805 | 0.03624 | 634 | 62.93 | <.0001 |
| **CYear\*INTf** | -0.03370 | 0.002061 | 3932 | -16.36 | <.0001 |

| **Type III Tests of Fixed Effects** | | | | |
| --- | --- | --- | --- | --- |
| **Effect** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **INTp** | 1 | 680 | 271.58 | <.0001 |
| **INTp\*CYear** | 1 | 3932 | 51.12 | <.0001 |
| **INTf** | 1 | 634 | 3960.40 | <.0001 |
| **CYear\*INTf** | 1 | 3932 | 267.50 | <.0001 |

| **Contrasts** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Label** | **Num DF** | **Den DF** | **Chi-Square** | **F Value** | **Pr > ChiSq** | **Pr > F** |
| **2 DF test of change** | 2 | 3932 | 312.32 | 156.16 | <.0001 | <.0001 |