NYU STERN
NEW YORK UNIVERSITY LEONARD N. STERN SCHOOL OF BUSINESS

Department of Economics

# Econometric Analysis of Panel Data

# Nonlinear Models

## Part I. Weibull Regression Model

In class, we examined a 'loglinear,' exponential regression model,

$$f(y_i \mid \mathbf{x_i},1) = \frac{1}{\theta_i}\exp\left(-\frac{y_i}{\theta_i}\right), \ \theta_i = \exp(\mathbf{x_i}'\boldsymbol{\beta}) = E[y_i|\mathbf{x}_i]$$

The Weibull model is an extension of the exponential model which adds a shape parameter, $\gamma$;

$$f(y_i \mid \mathbf{x_i},\gamma) = \frac{\gamma y_i^{\gamma-1}}{\theta_i^{\gamma}}\exp\left(-\left[\frac{y_i}{\theta_i}\right]^{\gamma}\right) \quad E[y_i|\mathbf{x}_i]=\Gamma[(\gamma+1)/2]\ \theta_i \ = \ .5*\text{sqr}(\pi) \ \text{ if } \gamma = 2.$$

The exponential model results when $\gamma = 1$. (This distribution looks like, but is not the gamma distribution we discussed in class.) An interesting special case is the Rayleigh distribution, which has $\gamma = 2$. The resulting density is

$$f(y_i \mid \mathbf{x_i},2) = \frac{2y_i}{\theta_i^2}\exp\left(-\left[\frac{y_i}{\theta_i}\right]^2\right)$$

One of the interesting things about the Rayleigh distribution is that $E[y|\mathbf{x_i}] = .5\sqrt{\pi}\,\theta_i$ (compared to $\theta_i$ for the exponential. $.5\sqrt{\pi}$ is approximately equal to 0.866.) One difference is the variance. The variance of the exponential variable is $\theta_i^2$. The variance of the Rayleigh variable is $[\Gamma(2) - \Gamma^2(1.5)]\theta_i^2$. Since $\Gamma(t)$ = t-1! for integer t, $\Gamma(2) = 1$. When t = an integer + .5, we can use the recurrence $\Gamma(t) = (t-1)\Gamma(t-1)$ until we reach $\Gamma(.5)$ which equals $\sqrt{\pi}$. Combining terms, then, the variance of the Rayleigh variable is $[1-(.5\sqrt{\pi})^2]\theta_i^2 = 0.2146\theta_i^2$.

a. The parameters $\boldsymbol{\beta}$ in the Rayleigh model could be estimated either by nonlinear least squares or by maximum likelihood. Which would be more efficient? Explain.

b. Form the log likelihood and derive the expressions for the first order conditions for maximizing the log likelihood for the Weibull model.

c. How would you test the null hypothesis of the Rayleigh model ($\gamma$=2) against the more general null of the Weibull model ($\gamma$ unrestricted)?

d. How would you test the null hypothesis of the Rayleigh model ($\gamma$=2) against the alternative of the Exponential model ($\gamma$ = 1)?

e. Maximum likelihood estimates of the parameters of the three models based on the German health data discussed in class appear below. Carry out the test in part c. Which of the three do you think is the appropriate model given the results below.

f. In the Rayleigh model, show how to obtain the three available estimators of the asymptotic covariance matrix of the MLE of $\boldsymbol{\beta}$. Remember, you are not estimating $\gamma$ (it equals 2), and the expected value of $y_i$ is $.5\sqrt{\pi}\,\theta_i$.

```
+---------------------------------------------+
| Weibull (Loglinear) Regression Model        |
| Dependent variable               HHNINC     |
| Number of observations            27322     |
| Log likelihood function        12033.50     |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Parameters in conditional mean function
 Constant      3.44054643      .02266279   151.815   .0000
 EDUC          -.10914142      .00147212   -74.139   .0000    11.3201838
 MARRIED       -.31230818      .00750583   -41.609   .0000     .75869263
 AGE            .00053144      .00044049     1.206   .2276    43.5271942
          Shape parameter for Weibull model
 P_scale       2.12853619      .00466881   455.905   .0000

+---------------------------------------------+
| Exponential (Loglinear) Regression Model    |
| Log likelihood function        1539.191     |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Parameters in conditional mean function
 Constant      1.82555590      .04219675    43.263   .0000
 EDUC          -.05545277      .00267224   -20.751   .0000    11.3201838
 MARRIED       -.23664845      .01460746   -16.201   .0000     .75869263
 AGE            .00087436      .00057331     1.525   .1272    43.5271942

+---------------------------------------------+
| Weibull (Loglinear) Regression Model        |
| Log likelihood function        11918.69     |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Parameters in conditional mean function
 Constant      3.28524659      .02586426   127.019   .0000
 EDUC          -.10377049      .00172163   -60.275   .0000    11.3201838
 MARRIED       -.31371176      .00871996   -35.976   .0000     .75869263
 AGE            .00064343      .00048739     1.320   .1868    43.5271942
          Shape parameter for Weibull model
 P_scale       1.99999964   ......(Fixed Parameter).......
```

## Part II. Marginal Effects in a Heteroscedastic Probit Model

Consider the following extension of the probit model. We make the disturbance heteroscedastic:

$$y_i^* = \alpha + x_{i1}\beta_1 + x_{i2}\beta_2 + \varepsilon_i$$
$$\varepsilon_i \sim N[0,\sigma_i^2] \text{ where } \sigma_i = \exp(\gamma_1 x_{i1} + \gamma_2 x_{i3})$$

This extension produces the probability model

$$\text{Prob}[y_i = 1 \mid x_{i1}, x_{i2}, x_{i3}] = \Phi\left(\frac{\alpha + x_{i1}\beta_1 + x_{i2}\beta_2}{\exp(x_{i1}\gamma_1 + x_{i3}\gamma_3)}\right)$$

Derive the partial (marginal) effects for this model, $\partial\text{Prob}(y_i=1)/\partial x_{i1}$, $\partial\text{Prob}(y_i=1)/\partial x_{i2}$, and $\partial\text{Prob}(y_i=1)/\partial x_{i3}$. It's worth noting that the partial effect for $x_{i3}$ has the opposite sign from the coefficient.

## Part III. Binomial Loglinear Model

Theory "Z" states that the age and education of the mother have an influence on the probability that a child will be female. Theory "Not Z" says that these two variables are irrelevant. Theory "There is no Theory" goes even further and states that the probability is always exactly one half. Consider modeling the number of female children, *Girls$_i$* in a sample of families; the number of children is *Kids$_i$*. The model in question is

$$\text{Kids}_i = \text{total number of children} = 0,1,...$$
$$\text{Girls}_i = \text{ number of female children} = 0,1,...,K_i$$
$$\text{Prob}(\text{GIRLS} = \text{Girls}_i \mid \mathbf{x}_i, \text{Kids}_i) = \binom{\text{Kids}_i}{\text{Girls}_i}\theta_i^{\text{Girls}_i}(1-\theta_i)^{\text{Kids}_i - \text{Girls}_i}$$
$$0 < \theta_i < 1,\ \theta_i = \text{ probability of a female child}$$
$$\theta_i = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta})}{1+\exp(\mathbf{x}_i'\boldsymbol{\beta})},\ \mathbf{x}_i = (1, \text{Age}_i, \text{Educ}_i), \beta = (\beta_0, \beta_1, \beta_2)$$

(Note that if $\text{Kids}_i = 0$, the probability that $\text{Girls}_i$ equals zero is 1.).

The three theories are: Z = all three coefficients nonzero
Not Z = $\beta_1 = \beta_2 = 0$, $\beta_0$ unrestricted
No Theory = $\beta_0 = \beta_1 = \beta_2 = 0$

1. Derive the log likelihood for estimation of the three unknown parameters. (Note, the factorial term at the beginning of the probabilities does not involve the parameters, so it can be ignored. This is often labeled 'an irrelevant constant.')
2. Derive the first order conditions for maximizing your log likelihood function.
3. Discuss exactly how you will test the hypothesis of theory "Not Z" against the alternative of theory "Z." How will you test the hypothesis of "No Theory" against theory "Z." What statistics will you use.
4. The data you need to do your estimation and carry out your tests are placed in two formats on the course website, .xls for a spreadsheet and .csv is an ascii text file. The files contain 500 observations on Age, Educ, Kids, Girls. Use these data to estimate your model and test the hypotheses.

**http://people.stern.nyu.edu/wgreene/Econometrics/BinomialData.xls**
**http://people.stern.nyu.edu/wgreene/Econometrics/BinomialData.csv**

(Disclaimer: The data are completely synthetic – simulated with a random number generator. This is a numerical example, not a study based on actual outcomes.)
Tip: Once you have read the data into NLOGIT, you can compute your estimates with

```
maximize
; labels=beta0,beta1,beta2 ; start = 0,0,0
; fcn = bx = beta0+beta1*educ+beta2*age |
        ti = exp(bx)/(1+exp(bx)) |
        girls * log(ti) + (kids-girls)*log(1-ti) $
```

To fix certain coefficients to zero, one convenient way is to use ;FIX=list. For example, to force $\beta_2$ to equal zero in the results, you would add ;Fix=beta2 to the command. (This forces the estimate to equal the starting value(s).) Also, note that in your results, what NLOGIT reports as the "Log Likelihood" in its results is actually the negative of the log likelihood.
5. Using your results for for Theory Z, compute the probabilities that are predicted for the data set, and show the distribution with a kernel density estimator.

Create ; Probi = Lgp(b(1)+b(2)*educ+b(3)*age) $
Kernel ; Rhs = Probi $

6. The expected number of Girls in a family with $\text{Kids}_i$ children is

$$E[\text{Girls}_i|\text{Kids}_i,x_i] = \theta_i \times \text{Kids}_i.$$

What is the partial effect with respect to Age? I.e., $\partial E[\text{Girls}_i|\text{Kids}_i,x_i]/\partial \text{Age}_i$ computed at the mean of age and education. Hint: $\theta_i$, the probability, is the logit probability, $\Lambda(\beta'x)$. The derivative of $\Lambda(t)$ with respect to t is $d\Lambda(t)/dt = \Lambda(t)[1 - \Lambda(t)]$.

## Part IV. Odds Ratio in the Logit Model

The results below present logit estimates of a model of whether the number of doctor visits is greater than zero based on the health care data discussed in class. (We used this example in class.)

```
+---------------------------------------------+
| Logit Model                                 |
| Dependent variable              DOCTOR      |
| Number of observations           27326      |
| Log likelihood function      -17407.69      |
| Restricted log likelihood    -18016.64      |
| Chi squared                   1217.911      |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Characteristics in numerator of Prob[Y = 1]
 HHNINC       -.13813513      .07764383    -1.779   .0752     .35213516
 HHKIDS       -.25400914      .02984645    -8.511   .0000     .40271576
 EDUC         -.02375730      .00578666    -4.106   .0000    11.3201838
 MARRIED       .11799754      .03374477     3.497   .0005     .75869263
 AGE           .01811793      .00132457    13.678   .0000    43.5271942
 FEMALE        .53279823      .02817810    18.908   .0000     .47880829
 WORKING      -.15388095      .03185320    -4.831   .0000     .67714662
 Constant     -.05351200      .09905516     -.540   .5890
```

a. The results given are estimates of the coefficients, $\boldsymbol{\beta}$. Researchers are sometimes interested in 'odds ratios,' which are computed as exp($\boldsymbol{\beta}$). (See, for example, the Stata manual.) How would the results in the table above change if we reported these, instead? Show explicitly.

b. The restricted log likelihood in a binary choice model is computed for a model which contains only a constant term. This, in turn, is ultimately a function of the proportion of ones in the sample. Given the value above, deduce the number of observations for which DOCTOR equals 1 in the sample of 27,326. (Hint: there are two solutions – the problem is symmetric in P and (1-P). The correct solution is the larger one.)

## Part V. The Poisson Regression Model

The following is based on the health care data used in several previous examples. We consider fitting a Poisson regression model to the variable DOCVIS which is the number of visits to the doctor by the individual in the given period. The model is as follows:

$$\text{Prob}[\text{DocVis}_i = y_i \mid \mathbf{x}_i] = \frac{\exp(-\theta_i)\theta_i^{y_i}}{y_i\,!}, y_i = 0,1,\ldots,\ \theta_i = \exp(\mathbf{x}_i'\boldsymbol{\beta})$$

a. Derive the log likelihood function for estimating $\boldsymbol{\beta}$ from a sample of n observations on $y_i$ and $\mathbf{x}_i$.

b. This is yet another log linear model in which E[$y_i$] = $\theta_i$. Use this result to show that the first derivatives of the log likelihood function have expectation zero.

c. Derive the forms of the three estimators of the asymptotic covariance matrix.

d. Show that the restricted log likelihood in which $\mathbf{x}_i$ contains only a constant term is a function only of the sample mean of $y_i$s.

e. Using the health care data set, estimate a Poisson model for DOCVIS in which

$\mathbf{x}_i$=[1, `female,age,hhninc,hhkids,educ,married]`.

f. Using your estimator, test the hypothesis that all coefficients in the model except the constant term are zero. The easiest test to use will be the likelihood ratio test. Show how to do the Lagrange multiplier test. (It has a particularly simple form in this model.) If you have access to the necessary matrix computations, carry out the LM test.

Estimating the Poisson Model.

All programs that you might use these days, Stata, SAS, SPSS, NLOGIT, EViews, have a pushbutton estimator for the Poisson model. But, this one, like the probit or logit models, is exceedingly simple to estimate, and you can program Newton's method and see how it works close up. The following shows how you can do this with NLOGIT. The annotations show what each command does. You should just put these commands on your editing screen, and execute them as shown below. (The lines with leading question marks are comments that can be ignored.) Based on part III, you should also be able to write a MAXIMIZE command to do the estimation. You might try this as well.

```
? (1) You have to load the Healthcare.lpj data set. I assume this id
? done.  The next line defines the variables in the equation as
? specified in the assignment.  Note, though that this also defines a
? matrix named X

     namelist ; x=one,female,age,hhninc,hhkids,educ,married$

? This next line shows you what you will be doing with your program.
? It fits the Poisson model using the internal estimator. We will
? replicate these results

     poisson ; lhs=docvis;rhs=x$

? Now, we obtain starting values for the iterations.  If all the slopes
? were zero, then E[y] would equal exp(α), so we can estimate the
? constant term with the log of the mean of the dependent variable.
? Then start the other coefficients at zero. The matrix command defines
? a column vector of this form.

     calc   ; list ; a0=log(xbr(docvis))$
     matrix ; beta = [a0/0/0/0/0/0/0] $

? This small set of commands does the iterations.  Note, the function
? involves the log of yi!.  We use Gamma(y+1) = y! and a special version,
? the log of the gamma function, lgm(y+1) = logy!
?************************************************************************
? To do the iterations, highlight and execute these commands. When done,
? the calc command shows you g’H^-1g. Execute the commands several times.
? You will see this go toward zero very quickly. When it gets very small,
? you are done iterating.  Then just display the results. Did you replicate
? the “real” results above?

     procedure $
     create ; ey = exp(beta'x)                 ? Mean
            ; logli = -ey + docvis*log(ey)     ? logL(i)
                 - lgm(docvis+1)               ? logL(i)
            ; gi = docvis - ey                 ? first derivative
            ; hi = ey $                        ? second derivative
? Matrix manipulations do the update of Newton’s method.
     matrix ; score      = X'gi
            ; Hessian    = X'[hi]X
            ; update     = <Hessian>*score
            ; beta       = beta + update $
     calc   ; list ; ghg = score'update $      ?
     endproc$
     execute ; n = 5 $
? Display results
     matrix ; stat(beta,<Hessian>,x)$
```