

Econ 673: Microeconometrics

Chapter 9: Count Data Models

Count Data Models

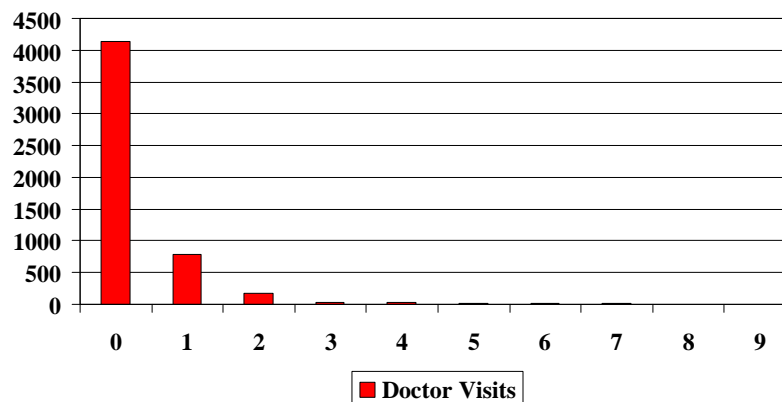
- Count Data models are used to characterize realizations of a non-negative integer-valued random variable
- Typically concerned with low probability events and many zeroes
- The dual to count data models are duration models, focusing on the length of time between the occurrence of an event (e.g., a recreation trip or the visit to a doctor)

Examples of Count Data Models

- Examples abound in the literature (145 hits in EconLit), including
 - Frequency of doctor visits
 - Patents
 - Recreation demand
 - Takeover bids
 - Bank failures
 - Accident frequency
 - Number of loan defaults
 - Presidential appointments to the Supreme Court
 - Number of criminal offenses
 - Manufacturing Defects

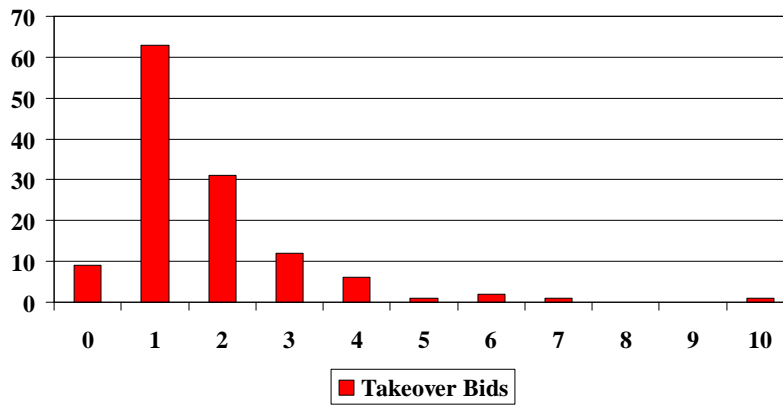
Example #1: Doctors Visits

(Cameron, Trivedi, Milne, and Piggott, 1988)



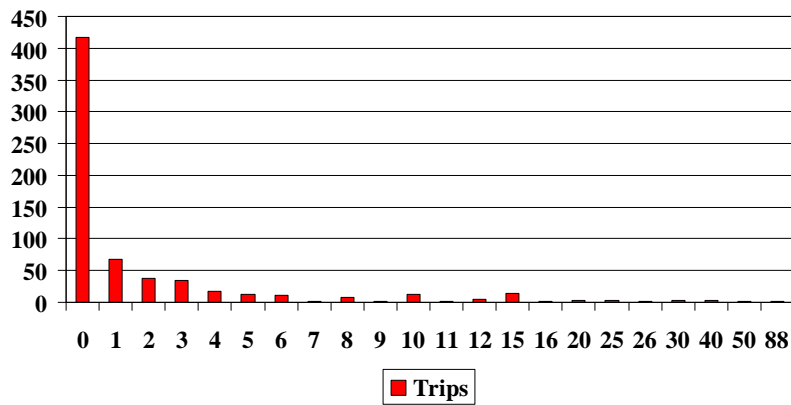
Example #2: Takeover Bids

(Jaggia and Thosar, 1993)



Example #3: Recreational Trips to Lake Somerville, Texas

(Ozuna and Gomez, 1995)



Outline

- Single Equation Models
 - Poisson
 - Distributional Assumptions
 - Estimation
 - Interpretation
 - Prediction
 - Limitations
 - Overdispersion
 - Excess zeros
 - Generalizations
 - Accounting for Overdispersion
 - The Negative Binomial Specification
 - Alternative Mixture Models
- Multivariate Count Data Models
- Comparisons of Continuous and Count Data Demand Systems

Sources – Univariate Setting

- *Cameron, A., and P. Trivedi (1998), *Regression Analysis of Count Data*, New York: Cambridge University Press, Chapter 1, Sections 3.1 to 3.5.
- Greene, W. H., (2000) *Econometric Analysis*, 4th edition, Upper Saddle River, New Jersey: Prentice-Hall, Inc., Section 19.9.
- Ruud, P., (2000) *An Introduction to Classical Econometric Theory*, New York: Oxford University Press, Section 27.2.2.
- *von Haefen, R., and D. Phaneuf (2003), “Estimating Preferences for Outdoor Recreation: A Comparison of Continuous and Count Data Demand Systems,” *Journal of Environmental Economics and Management*, **45**(3): 612-30.
- Englin, J., and J. Shonkwiler (1995), “Estimating Social Welfare Using Count Data Models: An Application to Long-run Recreation Demand Under Conditions of Endogenous Stratification and Truncation,” *Review of Economics and Statistics* **77**: 104-112.

The Poisson Distribution

- The most basic of the count data models is based on the Poisson distribution, developed in 1837 by Poisson
- Classic early study by Bortkiewicz (1898) analyzed the number of annual deaths in Prussian army from mule kicks
- If Y is a discrete random variable that is distributed Poisson, then

$$\Pr[Y = y] = \frac{e^{-\mu t} (\mu t)^y}{y!}, \quad y = 0, 1, 2, \dots$$

where

μ denotes the intensity or rate parameter, $\mu > 0$

t denotes the exposure period

The Poisson Distribution (cont'd)

- Usually normalize $t=1$, so that

$$\Pr[Y = y] = \frac{e^{-\mu} (\mu)^y}{y!}, \quad y = 0, 1, 2, \dots$$

- A key feature of the Poisson distribution is that

$$E[Y] = \text{Var}[Y] = \mu$$

referred to as the *equidispersion* property.

- *Additivity* property also holds; i.e., if $Y_i \sim i.i.d. P[\mu_i]$, then

$$S_Y \equiv \sum_i Y_i \sim P\left[\sum_i \mu_i\right]$$

“Rare Events” Interpretation

- Consider a sequence of n independent Bernoulli trials, each with success probability of π
- Let $Y_{n,\pi}$ denote the total number of successes

$$P_{n\pi}(k) = P[Y_{n\pi} = k] = \binom{n}{k} \pi^k (1-\pi)^{n-k}$$

and

$$\mu \equiv n\pi > 0$$

then, for a fixed μ ,

$$\lim_{\substack{n \rightarrow \infty \\ \pi \rightarrow 0 \\ \mu \text{ fixed}}} P_{n\pi}(k) = \lim_{\substack{n \rightarrow \infty \\ \pi \rightarrow 0 \\ \mu \text{ fixed}}} P_{\mu}(k) = \frac{\mu^k e^{-\mu}}{k!}$$

The Poisson Regression Model

- The standard Poisson regression model results if the intensity parameter is assumed to be a function of observed characteristics
- Let y_i denote the observed number of occurrences of the event of interest. The Poisson regression model assumes that

$$f(y_i | x_i) = \frac{e^{-\mu(x_i; \beta)} \mu(x_i; \beta)^{y_i}}{y_i!}$$

so that

$$E[y_i | x_i] = \text{Var}[y_i | x_i] = \mu(x_i; \beta)$$

The Log-Linear Poisson Regression Model

- The most frequently used functional form assumes that

$$\mu(x_i; \beta) = \exp(\beta' x_i)$$

insuring that $\mu > 0$, as required.

- The resulting nonlinear regression model is heteroskedastic, with

$$\begin{aligned} y_i &= E[y_i | x_i] + (y_i - E[y_i | x_i]) \\ &= \exp(\beta' x_i) + \varepsilon_i \end{aligned}$$

and

$$\text{Var}[\varepsilon_i | x_i] = \exp(\beta' x_i)$$

Estimation

- While either LS or WLS can be applied to the nonlinear regression model, ML is typically employed
- The log-likelihood function is given by

$$LL(y, x; \beta) = \sum_{i=1}^N \{y_i \beta' x_i - \exp(\beta' x_i) - \ln(y_i!)\}$$

with first order conditions

$$\sum_{i=1}^N [y_i - \exp(\beta' x_i)] x_i = \mathbf{0}$$

and

$$\frac{\partial^2 LL}{\partial \beta \partial \beta'} = -\sum_{i=1}^N \exp(\beta' x_i) x_i x_i' \quad \text{i.e., } LL \text{ is globally concave}$$

Properties of MLE

- Consistency of the $\hat{\beta}_{ML}$ requires only that conditional mean of y_i is correctly specified; i.e., it need not be Poisson distributed, we need only

$$E[y_i | x_i] = \exp(\beta' x_i)$$

however, the resulting ML standard errors will be incorrect unless y_i is Poisson.

- The standard errors can be corrected
- More efficient estimates can be obtained if equidispersion does not hold

Properties of MLE (cont'd)

- If the counts do follow a Poisson process, then

$$\hat{\beta}_{ML} \stackrel{a}{\sim} N(\beta, \Omega_{ML})$$

$$\Omega_{ML} = -E \left[\frac{\partial^2 LL}{\partial \beta \partial \beta'} \right] = \left(\sum_{i=1}^N \exp(\beta' x_i) x_i x_i' \right)^{-1}$$

Alternatively, one can replace Ω_{ML} with

$$\Omega_{MLOP} = \left(\sum_{i=1}^N [y - \exp(\beta' x_i)]^2 x_i x_i' \right)^{-1}$$

Overdispersion

- One of the chief criticisms of the Poisson specification is that it assumes equidispersion; i.e.,

$$E[y_i | x_i] = Var[y_i | x_i]$$

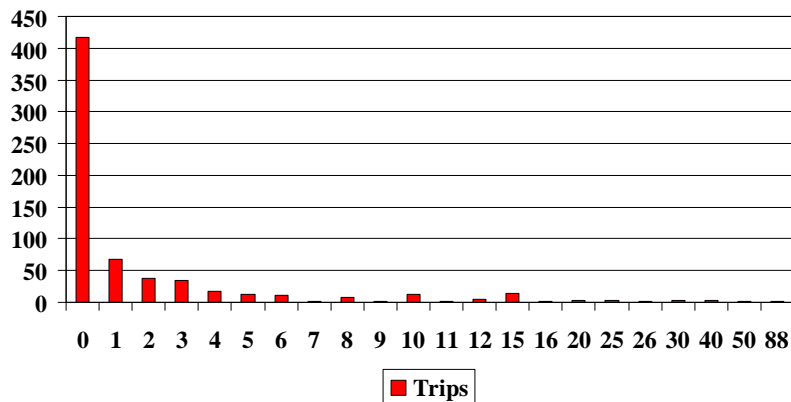
- In practice, overdispersion tends to hold; i.e.,

$$E[y_i | x_i] < Var[y_i | x_i]$$

- This manifests itself in terms of fatter tails and a greater number of zeros than would characterize a distribution with equidispersion

Example #3: Recreational Trips to Lake Somerville, Texas

(Ozuna and Gomez, 1995)



Solutions to Overdispersion

- Specify an alternative functional form for the conditional variance; e.g.,

$$\begin{aligned}V[y_i | x_i] &= \omega_i \\ &= \omega(\mu_i, \alpha)\end{aligned}$$

Given that this specification being correct, the appropriate asymptotic standard errors for $\hat{\beta}_{ML}$ can be constructed

- Specify an alternative distributional assumption (e.g., negative binomial) and apply ML estimation procedures

Poisson Pseudo-ML

- Given that the mean and variance have been correctly specified, and $\hat{\beta}_{ML}$ derived from the Poisson model, it can be shown that

$$\hat{\beta}_{ML} \stackrel{a}{\sim} N(\beta, \Omega_{PML})$$

where

$$\Omega_{PML} = \left(\sum_{i=1}^N \mu_i x_i x_i' \right)^{-1} \left(\sum_{i=1}^N \omega_i x_i x_i' \right) \left(\sum_{i=1}^N \mu_i x_i x_i' \right)^{-1}$$

NB1 Model

- One common specification is:

$$\begin{aligned}V[y_i | x_i] &= \omega(\mu_i, \alpha) \\ &= (1 + \alpha)\mu_i \\ &= \phi_i E[y_i | x_i]\end{aligned}$$

- $\hat{\Omega}_{PML}$ is then constructed using

$$\hat{\phi}_{NB1} = \frac{1}{n-k} \sum_{i=1}^N \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}$$

NB2 Model

- A second specification assumes that:

$$\begin{aligned}V[y_i | x_i] &= \omega(\mu_i, \alpha) \\ &= \mu_i + \alpha\mu_i^2\end{aligned}$$

- $\hat{\Omega}_{PML}$ is then constructed using

$$\hat{\alpha}_{NB2} = \frac{1}{n-k} \sum_{i=1}^N \frac{\left\{ (y_i - \hat{\mu}_i)^2 - \hat{\mu}_i \right\}}{\hat{\mu}_i^2}$$

RS (robust sandwich) Model

- Finally, we can proceed without specifying an exact functional form for the variance
- $\hat{\Omega}_{PML}$ is then constructed using

$$\Omega_{PML} = \left(\sum_{i=1}^N \mu_i x_i x_i' \right)^{-1} \left(\sum_{i=1}^N (y_i - \hat{\mu}_i)^2 x_i x_i' \right) \left(\sum_{i=1}^N \mu_i x_i x_i' \right)^{-1}$$

Example: Doctor Visits

- Cameron, Trivedi, Milne, and Piggott (1998) studied the number of visits to doctors during a 2 week period
- Explanatory variables include
 - Gender (Sex=1 for females)
 - Age (divided by 100)
 - Private health insurance dummy variable (LEVYPLUS)
 - Low income free government health insurance (FREEPOOR)
 - Old age/disability free government health insurance (FREEREPA)
 - Number of recent illnesses (ILLNESS)
 - Number of reduced activity days (ACTDAYS)
 - Health score (HSCORE)
 - Chronic conditions (CHCOND1 and CHCOND2)
- The mean number of visits is 0.302, whereas the variance is 0.637, suggesting overdispersion

Parameter Estimates

Variable	Estimate	Standard Errors				
		ML	MLOP	NB1	NB2	RS
Intercept	-2.224	0.190	0.144	0.219	0.207	0.254
Sex	0.157	0.056	0.041	0.065	0.062	0.079
Age	1.056	1.001	0.750	1.153	1.112	1.364
Age ²	-0.849	1.078	0.809	1.242	1.210	1.460
Income	-0.205	0.088	0.062	0.102	0.096	0.129
LevyPlus	0.123	0.072	0.056	0.083	0.077	0.095
FreePoor	-0.440	0.180	0.116	0.207	0.188	0.290
FreeREPA	0.080	0.092	0.070	0.106	0.102	0.126
Illness	0.187	0.018	0.014	0.021	0.021	0.024
Actdays	0.127	0.005	0.004	0.006	0.006	0.008
Hscore	0.030	0.010	0.007	0.012	0.012	0.014
ChCond1	0.114	0.066	0.051	0.077	0.071	0.091
ChCond2	0.141	0.083	0.059	0.096	0.092	0.122

Negative Binomial Distribution

- The previous methods rely upon the Poisson-based ML estimates, correcting for overdispersion
- However, they are inefficient in that they do not use the overdispersion characteristic in estimation
- An alternative is to rely upon a parametric distribution without equidispersion; e.g., the Negative Binomial

$$f(y_i | \mu_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i} \right)^{\alpha^{-1}} \left(\frac{\mu_i}{\alpha^{-1} + \mu_i} \right)^{y_i}$$

This reduces to the Poisson distribution if $\alpha_i = 0$

ML Estimation for NB Specification

- The corresponding log-likelihood in this case is given by

$$LL(y, x; \beta, \alpha) = \sum_{i=1}^N \left\{ \sum_{j=0}^{y_i-1} \ln(j + \alpha^{-1}) - \ln(y_i!) \right. \\ \left. + (y_i + \alpha^{-1}) \ln(1 + \alpha \exp(\beta' x_i)) + y_i \ln \alpha + y_i \beta' x_i \right\}$$

The corresponding first order conditions are straightforward and the resulting variance-covariance matrix is block diagonal

Parameter Estimates

Variable	Poisson		NB	
	Estimate	Std.Err.	Estimate	Std.Err.
Intercept	-2.224	0.207	-2.190	0.222
Sex	0.157	0.062	0.217	0.066
Age	1.056	1.112	-0.216	1.233
Age ²	-0.849	1.210	0.609	1.380
Income	-0.205	0.096	-0.142	0.098
LevyPlus	0.123	0.077	0.118	0.085
FreePoor	-0.440	0.188	-0.497	0.175
FreeREPA	0.080	0.102	0.145	0.117
Illness	0.187	0.021	0.214	0.026
Actdays	0.127	0.006	0.144	0.008
Hscore	0.030	0.012	0.038	0.014
ChCond1	0.114	0.071	0.099	0.077
ChCond2	0.141	0.092	0.190	0.095
LL	-3355.5		-3198.7	

Indications of Overdispersion

- Comparing the unconditional mean and variance of the count data will give an indication of overdispersion, but will tend to overstate its degree, since

$$\frac{\text{Var}[y_i]}{E[y_i]} > \frac{\text{Var}[y_i | x_i]}{E[y_i | x_i]}$$

- Cameron and Trivedi (1998) suggest a good rule of thumb is that overdispersion is an issue if

$$\frac{\text{Var}[y_i]}{E[y_i]} > 2$$

Tests of Overdispersion

- One can estimate both the Negative Binomial and Poisson regression models and test for overdispersion
- The distribution of the test statistics are not standard due to the restriction that $\alpha > 0$
 - LR test statistic for a level of δ is $\chi^2_{1-2\delta}(1)$
 - Wald test is implemented as a one-sided t test
- One can also test for overdispersion by estimating the Poisson model and running the auxiliary regression

$$\frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} = \alpha \hat{\mu}_i + u_i \quad \hat{\mu}_i = \exp(\hat{\beta}' x_i)$$

Model Interpretation

- As with all nonlinear regression models, the parameters do not indicate the marginal impact of an explanatory variable
- For the case of an exponential conditional mean; i.e.,

$$E[y_i | x_{ik}] = \exp(\beta' x_i)$$

We have

$$\frac{\partial E[y_i | x_{ik}]}{\partial x_{ik}} = \beta_k \exp(\beta' x_i)$$

or in elasticity form

$$\frac{\partial \ln E[y_i | x_{ik}]}{\partial \ln x_{ik}} = \beta_k x_{ik}$$

Overall Response

- One is often interested in the overall response in the population
- One overall measure of response for the same level change is

$$\frac{1}{N} \sum_{i=1}^N \frac{\partial E[y_i | x_i]}{\partial x_{ik}} = \frac{1}{N} \sum_{i=1}^N \beta_k \exp(\beta' x_i) = \beta_k \left[\frac{1}{N} \sum_{i=1}^N \exp(\beta' x_i) \right]$$

For the Poisson ML estimates, when there is a constant in the model, this reduces considerable, since

$$\frac{1}{N} \sum_{i=1}^N \exp(\beta' x_i) = \frac{1}{N} \sum_{i=1}^N y_i = \bar{y} \Rightarrow \frac{1}{N} \sum_{i=1}^N \frac{\partial E[y_i | x_i]}{\partial x_{ik}} = \beta_k \bar{y}$$

Overall Response (cont'd)

- Alternatively, if you are looking at the same percentage change in the explanatory variable, then

$$\frac{1}{N} \sum_{i=1}^N \frac{\partial \ln E[y_i | x_i]}{\partial \ln x_{ik}} = \frac{1}{N} \sum_{i=1}^N \beta_k x_{ik} = \beta_k \bar{x}_{ik}$$

- As we have seen in other nonlinear regression models, one does not want to rely on the response as the average characteristic

$$\left. \frac{\partial E[y_i | x_i]}{\partial x_{ik}} \right|_{\bar{x}} = \beta_k \exp(\beta' \bar{x}) < \frac{1}{N} \sum_{i=1}^N \frac{\partial E[y_i | x_i]}{\partial x_{ik}}$$

Miscellaneous Notes on Response Measures

- While the parameters do not indicate the marginal impact, their relative sizes do indicate the relative strength of each variable's effect; i.e.,

$$\frac{\left(\frac{\partial E[y_i | x_i]}{\partial x_{ik}} \right)}{\left(\frac{\partial E[y_i | x_i]}{\partial x_{ij}} \right)} = \frac{\beta_k}{\beta_j}$$

- Finally, while the marginal effects are nonlinear functions of the estimated parameters, their standard deviation can be constructed by simulation or bootstrapping

Variable Effects: Doctor Visits

Variable	Coeff.	Average	At Avg.	Elasticity
Intercept	-2.224			
Sex	0.157	0.047	0.035	0.082
Age	1.056	0.319	0.241	0.430
Age ²	-0.849	-0.256	-0.193	-0.176
Income	-0.205	-0.062	-0.047	-0.120
LevyPlus	0.123	0.037	0.028	0.055
FreePoor	-0.440	-0.133	-0.100	-0.019
FreeREPA	0.080	0.024	-0.018	0.017
Illness	0.187	0.056	0.043	0.268
Actdays	0.127	0.038	0.029	0.109
Hscore	0.030	0.009	0.007	0.037
ChCond1	0.114	0.034	0.026	0.046
ChCond2	0.141	0.043	0.032	0.016

Problems with the Poisson Regression Model

- Overdispersion
 - Typically attributed to omitted and/or unobserved sources of heterogeneity
 - Can use Poisson ML estimates with corrected standard errors
 - Alternatively, can use models without equidispersion
 - Negative Binomial
 - mixed Poisson
- Truncation (especially zero truncation)
- Excess zeros
- Serial correlation

Mixed Poisson

- Mixture Models in the context of count data models are much like in the mixed logit context – capturing unobserved heterogeneity
- Suppose

$$E[y_i | x_i, v_i] = \mu_i v_i$$

where

$$\mu_i = \exp(\beta' x_i)$$

and

$$v_i = \exp(\varepsilon_i)$$

with ε_i capturing unobserved omitted explanatory variables

Mixed Poisson (cont'd)

- If, for example, one assumes that ε_i is iid with

$$E[v_i | x_i] = 1$$

and

$$\text{Var}[\varepsilon_i | x_i] = \sigma^2$$

with y_i distributed according to a Poisson distribution, conditional on x_i and v_i ; i.e.,

$$E[y_i | x_i, v_i] = \text{Var}[y_i | x_i, v_i] = \mu_i$$

then

$$E[y_i | x_i] = E_{v_i}[\mu_i v_i] = \mu_i$$

and

$$\text{Var}[y_i | x_i] = E_{v_i}(\text{Var}[y_i | x_i, v_i]) + \text{Var}_{v_i}(E[y_i | x_i, v_i]) = \mu_i [1 + \sigma_v^2 \mu_i]$$

As in NB2



Mixture Models Properties

- In general, mixture models are characterized by
 - Overdispersion
 - A greater number of zeros relative to equidispersion model
 - Thicker tails than the parent distribution (in this case Poisson)
- If v_i is drawn from a gamma distribution, then the resulting mixed Poisson model is in fact the Negative Binomial model
- Other mixing distributions are, of course, feasible

Truncation and Censored in Count Data Models

- It is not uncommon to find truncation and censoring problems in count data settings.
- Truncation frequently occurs with “zeros” being unobserved, resulting the “Positive Poisson” model
- Censoring can arise when “large” realizations are pooled into a single category (e.g., over 10).
- The fundamental steps required to correct of these data problems are similar to those seen for continuous variables, resulting in a corrected pdf or cdf

Truncation

- In general, suppose that y_i had a discrete pdf of $h(y_i, \Lambda)$ where Λ denotes the parameters of the distribution.

- Let

$$H(y_i, \Lambda) = \sum_{i=0}^{y_i} h(y_i, \Lambda) \quad y_i = 0, \dots$$

denote the corresponding cdf

- For a sample truncated from below, excluding values of y_i less than r , then

$$f(y_i, \Lambda | y_i \geq r) = \begin{cases} \frac{h(y_i, \Lambda)}{1 - H(r-1, \Lambda)} & y_i = r, r+1, \dots \\ 0 & \text{otherwise} \end{cases}$$

Truncated Poisson

- Consider a situation in which only positive counts are observed; e.g.,
 - On site surveys (recreational, doctor visits, etc.)
 - Accident frequencies from police reports, etc.
- If the counts follow a Poisson process, then

$$H(0, \Lambda) = \Pr[y_i = 0] = \exp(-\mu_i)$$

and

$$f(y_i, \Lambda | y_i \geq 1) = \begin{cases} \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i! (1 - e^{-\mu_i})} & y_i = 1, 2, \dots \\ 0 & \text{otherwise} \end{cases}$$

Truncated Poisson (cont'd)

- The central moments of the truncated Poisson are given by

$$E[y_i | y_i > 0] = \frac{\mu_i}{1 - e^{-\mu_i}}$$

$$Var[y_i | y_i > 0] = \frac{\mu_i}{1 - e^{-\mu_i}} \left[1 - \frac{\mu_i e^{-\mu_i}}{1 - e^{-\mu_i}} \right] < \mu_i < E[y_i | y_i > 0]$$

Truncated Poisson (cont'd)

For the Truncated Poisson regression model, the corresponding log-likelihood function is then given by

$$LL(y, x; \beta) = \sum_{i=1}^N \left\{ y_i \beta' x_i - \exp(\beta' x_i) - \ln(y_i!) - \ln(1 - \exp[-\exp(\beta' x_i)]) \right\}$$

- Without the truncation correction, the parameter estimates will be biased
- Unlike its untruncated counterpart, the ML parameters estimates will be biased, even after correcting for truncation *if* overdispersion exists
- This latter results suggests using a model which allows for overdispersion (e.g., negative binomial)

Truncated Negative Binomial

- If we have a NB distribution truncated at zero, then

$$h(y_i | \mu_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i} \right)^{\alpha^{-1}} \left(\frac{\mu_i}{\alpha^{-1} + \mu_i} \right)^{y_i}$$

$$H(0, \Lambda) = \Pr[y_i = 0] = \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i} \right)^{\alpha^{-1}}$$

Which can in turn be used to derive truncated distribution and log-likelihood for estimation

Example:

Creel and Loomis (1990), *AJAE*

- Analyzed deer hunter trips in California in 1987
- Total number of trips were analyzed using
 - OLS
 - NLS
 - Truncated NLS (truncated at 0.5)
 - Poisson
 - Truncated Poisson
 - Negative Binomial
 - Truncated Negative Binomial

Example: Creel and Loomis (1990) (cont'd)

- 2223 observations with a mean number of trips of 2.76
- Three subsets of the data were created
 - Specification (707)
 - Estimation (764)
 - Prediction (752)
- Explanatory Variables
 - Travel Cost (TC)
 - Travel Time (TIME)
 - Average Trip Length (DAYS)
 - Number of prior trips deer hunting (YEARS)
 - Success last year (BAG, dummy)
 - Number of prior passed up opportunities to bag a deer (PASNO)
 - Number of deer seen on past trip (DEERSN)
 - Household Income (INCOME)
 - Zonal hunting season length (SEASON)

Results

Variable	OLS	NLS	TNLS	POIS	TPOIS	NB	TNB
ONE	4.674	1.827	2.190	1.560	1.603	1.514	1.332
TC	-0.012	-0.006	-0.027	-0.006	-0.013	-0.006	-0.014
TIME	-0.052	-0.098	-0.175	-0.024	-0.033	-0.017	-0.017
DAYS	-0.154	-0.055	-0.057	-0.044	-0.048	-0.038	-0.050
YEARS	0.037	0.012	0.019	0.010	0.011	0.009	0.011
BAG	-0.252	-0.104	-0.189	-0.078	-0.104	-0.0747	-0.147
PASNO	0.226	0.027	0.033	0.034	0.030	0.040	0.062
DEERSN	0.001	0.001	0.001	0.001	0.001	0.001	0.001
INCOME	-0.019	-0.008	-0.014	-0.006	-0.008	-0.006	-0.011
SEASON	0.030	0.024	0.026	0.021	0.033	0.018	0.036
log-L	-1892.9	-1810.4	-1523.5	-1539.1	-1295.8	-1443.2	-1145.6

Out-of-Sample Predictions

	Statistical Model						
	OLS	NLS	TNLS	POIS	TPOIS	NB	TNB*
R ²	0.233	0.297	0.027	0.346	0.334	0.328	0.301
Act-Pred	-121.9	50.5	-775.0	16.0	-132.8	-99.1	-74.1
%Err	-6.6	2.7	-40.9	0.9	-7.2	-5.4	-4.0
CS/pred trips	117.25	172.82	36.72	153.62	74.71	163.05	70.07

Systems of Count Models

- Count data models are often applied to the estimation of consumer demand
- In these settings, a system of counts is needed to allow for substitution possibilities among commodities, each good taking the form of counts
- Approaches
 - Independent Poisson
 - Seemingly Unrelated Poisson Regression Model (SUPREME)
 - Multivariate Poisson Log-normal

Sources

- Ozuna, T., and R. Gomez (1990), “Estimating a System of Recreation Demand Functions Using a Seemingly Unrelated Poisson Regression Approach,” *The Review of Economics and Statistics* **76**: 356-360.
- Shonkwiler, J. (1999), “Recreation Demand Systems for Multiple Site Count Data Travel Cost Models,” in Herriges, J., and C. Kling, *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice*, Cheltenham: Edward Elgar, pp. 253-270.
- *von Haefen, R., and D. Phaneuf (2003), “Estimating Preferences for Outdoor Recreation: A Comparison of Continuous and Count Data Demand Systems,” *Journal of Environmental Economics and Management*, **45**(3): 612-30.

The SUPREME Specification

- The Supreme specification is fundamentally a mixed Poisson specification, where the mixing distribution is another independent Poisson
- Consider a two good example, with

$$\left. \begin{array}{l} y_i^* \sim \text{Poisson}[\mu_i] \quad i = 1, 2 \\ \omega \sim \text{Poisson}[\xi] \end{array} \right\} \begin{array}{l} \text{assumed} \\ \text{independent} \end{array}$$

Then

$$y_i = y_i^* + \omega \sim \text{Poisson}[\mu_i + \xi] \quad i = 1, 2$$

$$E[y_i] = \text{Var}[y_i] = \mu_i + \xi = \theta_i$$

$$\text{Cov}[y_1, y_2] = \xi$$

The SUPREME Specification (cont'd)

The resulting bivariate Poisson distribution for (y_1, y_2) becomes

$$P(y_1, y_2 | \theta_1, \theta_2, \xi) = \exp(\xi - \theta_1 - \theta_2) \times \sum_{j=0}^{\min(y_1, y_2)} \left\{ \frac{\xi^j}{j!} \left[\frac{(\theta_1 - \xi)^{(y_1-j)}}{(y_1-j)!} \right] \left[\frac{(\theta_2 - \xi)^{(y_2-j)}}{(y_2-j)!} \right] \right\}$$

The SUPREME Specification (cont'd)

- The “regression” version of this model comes about by specifying that

$$\theta_i = E[y_i | x_i, \beta_i] \quad i = 1, 2$$

Typically, one uses a linear exponential form, with

$$\theta_i = \exp(\beta_i' x_i) \quad i = 1, 2$$

where x_i ($i=1,2$) are $k_i \times 1$ vectors of exogenous explanatory variables and the β 's are unknown parameter vectors.

Estimation the Supreme Model

- The SUPREME specification is estimated using maximum likelihood estimation, with the log-likelihood function given by:

$$LL(y_1, y_2; x_1, x_2, \beta_1, \beta_2, \xi) = \sum_{i=1}^N \left[\xi - \exp(\beta_1' x_{1i}) - \exp(\beta_2' x_{2i}) + \ln \left(\sum_{j=0}^{\min(y_{1i}, y_{2i})} A_{ij} \right) \right]$$

where

$$A_{ij} = \frac{\xi^j}{j!} \left[\frac{(e^{\beta_1' x_{1i}} - \xi)^{(y_{1i}-j)}}{(y_{1i}-j)!} \right] \left[\frac{(e^{\beta_2' x_{2i}} - \xi)^{(y_{2i}-j)}}{(y_{2i}-j)!} \right]$$

Merits of the SUPREME Model

- Advantages
 - Relatively easy to implement
 - Efficiency gain over independent Poisson, since it allows for covariance (even when explanatory variables are the same in the two equations)
 - With simultaneous estimation, one can test cross equation restrictions
- Disadvantages
 - Cannot accommodate overdispersion
 - Cannot capture negative correlations between equations
 - Typically restricted to bivariate setting, though can be theoretically extended to more than two

Example: Ozuna and Gomez (1990)

- Examined recreation trips to two Texas Lakes
 - Conroe
 - Somerville
- 659 observations from a survey of boat owners
- Explanatory Variables
 - Income
 - Travel costs to each site (and two substitute sites)
 - Quality ratings for four sites
 - Dummy variables for
 - water skiing
 - annual pass to Lake Somerville
 - overnight stay

Results

	SUR		SUPREME		Independent Poisson	
	Conroe	Somerville	Conroe	Somerville	Conroe	Somerville
Intercept	0.492	0.725	0.257	0.013	0.245	0.003
C_{conroe}	-0.111	0.033	-0.103	-0.032	-0.104	0.032
$C_{livingston}$	0.010	-0.003	0.064	0.043	0.064	0.004
$C_{somerville}$	0.033	-0.104	0.028	-0.067	0.028	-0.068
$C_{houston}$	-0.021	0.070	0.005	0.046	0.005	0.046
Income	-0.065	-0.120	-0.044	-0.067	-0.044	-0.066
Q_{conroe}	1.390		0.507		0.507	
$Q_{somerville}$		0.833		0.487		0.488
Waterski		1.019		0.486		0.484
$Q_{livingston}$	-0.151	-0.065	-0.064	-0.102	-0.064	-0.101
Overnight	-1.453		-0.571		-0.566	
Fee		5.904		0.647		-0.647
ξ				0.023		

Mixed Multivariate Poisson Models

- The concept underlying the SUPREME specification can be extended to a multivariate setting much like Mixed Logit extends the MNL model
- Suppose that

$$y_{ij} | x_{ij}, \beta_{ij} \sim \text{Poisson}[\mu_{ij}] \quad j = 1, \dots, J$$

where

$$\mu_{ij} = \exp(\beta'_{ij} x_{ij})$$

$$\beta_{i.} \sim f(\beta | \theta)$$

then

$$\Pr[y_{ij} = k | x_{ij}, \beta_{ij}] = \frac{e^{-\mu_{ij}} \mu_{ij}^k}{k!}$$

Mixed Multivariate Poisson Models (cont'd)

- The unconditional choice probabilities are then obtained by integrating over the chosen parametric distributions, with

$$\Pr[y_{i.} = k. | x_{i.}] = \int \left[\prod_{j=1}^J \frac{e^{-\mu_{ij}} \mu_{ij}^k}{k!} \right] f(\beta | \theta) d\beta$$

- Simulation methods can be used in ML estimation
- The distributional assumptions associated with the parameters can be used to induce
 - correlation between equations
 - overdispersion

Comparing Continuous and Count Data Demand Systems

- von Haefen and Phaneuf (2003) provide a comparison of demand systems based on
 - Kuhn-Tucker framework
 - Count data models
- They focus on both
 - conceptual differences in the modeling approaches and
 - the empirical performance of each model in analyzing wetland recreation usage in Iowa

Review of Kuhn-Tucker Model

- Recall that the KT model starts with a direct utility function with an integrated error term capturing unobserved factors causing preference heterogeneity; i.e.,

$$U(x, q, z, \varepsilon)$$

where

- x denotes the $M \times 1$ vector of commodities,
- q denotes an $M \times K$ vector of commodity characteristics,
- z is the numeraire good and
- ε denotes the vector capturing unobserved preference factors

Review of Kuhn-Tucker Model (cont'd)

- The individual is assumed to solve

$$\underset{x}{\text{Max}} U(x, q, y - p'x, \varepsilon) \quad \text{s.t. } x \geq 0$$

yielding first order conditions

$$\begin{aligned} p_i &\geq \frac{U_{x_i}(x^*, q, y - p'x^*, \varepsilon)}{U_z(x^*, q, y - p'x^*, \varepsilon)} \\ &\equiv \xi_{x_i}(p, q, y, \varepsilon) \quad \forall i = 1, \dots, M \end{aligned}$$

where ξ_{x_i} represents a “virtual price” for commodity i

Review of Kuhn-Tucker Model (cont'd)

- Thus,

$$x_i \begin{cases} > 0 & \text{if } p_i = \xi_{x_i}(p, q, y, \varepsilon) \\ = 0 & \text{if } p_i > \xi_{x_i}(p, q, y, \varepsilon) \end{cases}$$

- The individual's derived demands can be written as

$$x^* = f[\xi_x^*(p, q, y, \varepsilon), q, y, \varepsilon]$$

- The KT conditions are used to derive log-likelihood functions used in estimation

Count Data Demand Systems

- Ideally a count data model proceeds from the maximization problem

$$\text{Max}_x U(x, q, y - p'x, \varepsilon) \quad \text{s.t. } x_i \in \{0, 1, \dots\} \forall i$$

- However, the count nature of the choice set precludes the use of differential calculus
- The solution relies on comparison across a large choice set
- Count data models instead rely on utility maximization in the specification of the expected number of trips and the count distributional assumptions to generate trips

Count Data Demand Systems (cont'd)

- Formally, expected trips are specified, with

$$E[x] = f(p, q, y)$$

- Utility theory is employed at this stage by requiring that these *expected trip demand* equations satisfy standard integrability restrictions; i.e.,
 - adding up
 - homogeneity of degree zero in prices and income
 - symmetry and negative semi-definite Slutsky matrix
- Note: These assumptions *are not* imposed on the actual trip demands, just their expectations

Count Data Demand Systems (cont'd)

- Welfare analysis proceeds by using the expected demands to integrate back to an indirect utility function

$$v(p, q, y)$$

- It is not clear whose preferences this indirect utility represents
 - It is derived from average or expected trip demands, not any one individual's demands
 - All prices are likely to enter these average demands, so they will enter the counterpart *representative indirect utility function*,
 - This representative utility function is in turn used for welfare analysis for all agents so all price changes matter for all agents
 - This contrast with the KT approach, for which only consumed commodity price enter an individual indirect utility function

Model Specification - KT

- The authors use a Stone-Geary utility function

$$\sum_{i=1}^M \Psi_i \ln(\phi_i x_i + \theta_i) + \ln(y - p'x)$$

where

Ψ_i is a quality index for good i

ϕ_i is a *repackaging* parameter for good i

θ_i is a translation parameter for good i

Model Specification – KT (cont'd)

- Two specifications are considered

Specification #1

$$\Psi_i = \exp(\delta' s + \gamma' q_i + \varepsilon_i)$$

$$\phi_i = 1$$

where s denotes individual socio-demographic characteristics and q_i denotes attributes of good i

Specification #2

$$\Psi_i = \exp(\delta' s + \varepsilon_i)$$

$$\phi_i = \exp(\gamma' q_i)$$

} authors refer to this as
weak complementary
model

Model Specification – KT (cont'd)

- The resulting KT conditions are given by

$$\varepsilon_i \leq \ln p_i + \ln \left(x_i + \frac{\theta_i}{\phi_i} \right) - \ln(y - p'x) - \ln \bar{\Psi}_i = g_i$$

where

$$\bar{\Psi}_i = \exp(\ln \Psi_i - \varepsilon_i)$$

Assuming iid extreme value errors, the likelihood function becomes

$$L = \text{abs} |J| \left(\prod_{i=1}^M 1_{x_i > 0} \frac{\exp(-g_i / v_i)}{v_i} \right) \exp \left(- \sum_{i=1}^M \exp(-g_i / v_i) \right)$$

Model Specification – Counts

- The Count Data model starts with a specification of expected counts for each commodity. Shonkwiler (1999), started with

$$E[x_i] = \exp\left(\alpha_j + \sum_{j=1}^M \beta_{ij} p_j + \lambda_i \ln y\right) \quad \forall i = 1, \dots, M$$

Integrability restrictions can be imposed on this system of incomplete demand equations by assuming that:

$$\begin{aligned} \alpha_i &> 0 & \lambda_i &= \lambda \quad \forall i \\ \beta_{ii} &< 0 & \beta_{ij} &= 0 \quad \forall j \neq i \end{aligned}$$

Model Specification – Counts (cont'd)

- The resulting system of expected demands becomes

$$E[x_i] = \exp(\alpha_i + \beta_{ii} p_i + \lambda \ln y) \quad \forall i = 1, \dots, M$$

- Notice that we now have
 - no cross price effects
 - a common income effect
- vonHaefen and Phaneuf use a slight modification of this functional form

$$E[x_i] = \frac{1}{\phi_i} \exp\left(\alpha_i + \beta_{ii} \frac{p_i - \omega_i}{\phi_i} + \lambda \ln y\right) \quad \forall i = 1, \dots, M$$

Model Specification – Counts (cont'd)

- Two specifications are considered
Specification #1: “Simple repackaging”

$$\omega_i = 0$$

$$\phi_i = \gamma' q_i$$

where ω_i is a “cross-product” parameter and ϕ_i is the “repackaging” parameter

Specification #2

$$\omega_i = \gamma' q_i$$

$$\phi_i = 1$$

Accounting for Excess Zeros

- To complete the model specification, the authors note that the data are characterized by excess zeros.
- One way to handle this problem is to use a “zero-inflated” count model, which assumes that

$$\Pr[x_i = 0] = \pi_i + (1 - \pi_i)e^{-\mu_i}$$

$$\Pr[x_i = r] = (1 - \pi_i) \frac{e^{-\mu_i} \mu_i^r}{r!}$$

This is a *finite mixture model* in which one of the distributions is degenerate at zero and the other is Poisson

Accounting for Excess Zeros (cont'd)

- Notice that for such models

$$\text{Var}[x_i] = (1 - \pi_i)(\mu_i + \pi_i \mu_i^2) > (1 - \pi_i)\mu_i = E[x_i]$$

thus excess zeros imply overdispersion

- One can then specify a functional form for spike at zero; e.g., logistic with

$$\pi_i = \frac{\exp(\tau'z)}{1 + \exp(\tau'z)}$$

Error Specification – Count Model

- von Haefen and Phaneuf take this model one step further, assuming that the counts are independent zero inflated negative binomial counts with

$$\mu_i = \frac{1}{\phi_i} \exp\left(\alpha_i + \beta_{ii} \frac{p_i - \omega_i}{\phi_i} + \lambda \ln y\right) \quad \forall i = 1, \dots, M$$

$$\pi_i = \frac{\exp(\rho_i + \delta's)}{1 + \exp(\rho_i + \delta's)}$$

Note: This allows the spike probability, π_i , to vary by commodity, but requires the marginal impact of socio-demographic characteristics to be the same

Count Model Log-Likelihood

The resulting log-likelihood function is given by

$$L = \prod_{i=1}^M \left\{ \mathbf{1}_{x_i=0} \left[\pi_i + (1 - \pi_i) \left(\frac{1}{1 + v_i \eta_i} \right)^{\frac{1}{v_i}} \right] \right. \\ \left. + \mathbf{1}_{x_i > 0} (1 - \pi_i) \frac{\Gamma\left(\frac{1}{v_i} + x_i\right)}{\Gamma\left(\frac{1}{v_i}\right) x!} \left(\frac{1}{1 + v_i \eta_i} \right)^{\frac{1}{v_i}} \left(\frac{v_i \eta_i}{1 + v_i \eta_i} \right)^{x_i} \right\}$$

Application – Iowa Wetlands

- Analysis based on 1997 recreational trips to wetlands in Iowa
- 2891 observations
- vonHaefen and Phaneuf aggregated destinations to 5 “mega-zones”
- explanatory variables include
 - travel costs
 - pheasant counts in each zone
 - socio-demographic characteristics
 - gender
 - hunting license
 - age
 - college education dummy variable

Descriptive Statistics Average Trip Data

	Site					
	1	2	3	4	5	All
Trips	0.81	1.18	3.60	1.44	1.47	8.51
Range	[0,40]	[0,40]	[0,49]	[0,48]	[0,45]	[0,49]
%0's	88	84	60	79	78	31?
%≥40	0.03	0.07	0.45	0.28	0.21	2.56
Cost	156	119	76	118	106	115
Phsnt.	18	51	56	28	40	38

Other Summary Statistics

% Visiting	
0 Sites	31.96
1 Site	37.22
2 Sites	21.20
3 Sites	7.26
4 Sites	2.01
5 Sites	0.35

Demographics	
Mean Income	\$43,266
% Male	73.64
Mean Age	49.08
% License	68.35
% 4yr college	28.02

KT Results

	Specification 1	Specification 2
Log-Likelihood	-14,919	-14,940
θ_1	5.522	6.963
θ_2	6.915	
θ_3	6.656	
θ_4	5.475	
θ_5	5.843	
$\delta_{constant}$	-5.508	-5.065
δ_{permit}	0.344	0.341
δ_{male}	-0.035	-0.034
δ_{age}	-0.005	-0.004
$\delta_{college}$	-0.095	-0.092
$\gamma_{pheasants}$	0.011	0.003

Count Data Model Results Spike Parameters

	Specification 1	Specification 2
Log-Likelihood	-15,370	-15,400
ρ_1	-2.34	-1.96
ρ_2	-0.97	-0.44
ρ_3	-1.74	-1.67
ρ_4	-1.63	-1.60
ρ_5	-1.12	-1.42
δ_{permit}	-2.34	-2.18
δ_{male}	-0.63	-0.60
δ_{age}	0.04	0.03
$\delta_{college}$	-1.04	-1.07

Count Data Model Results Count Parameters

	Specification 1	Specification 2
α_1	-3.19	-5.55
α_2	-1.26	
α_3	0.02	
α_4	-1.80	
α_5	-0.39	
β_{11}	-0.22	-0.015
β_{22}	-1.18	-0.026
β_{33}	-2.14	-0.038
β_{44}	-0.65	-0.023
β_{55}	-1.48	-0.032
$\gamma_{pheasants}$	1 (not estimated)	0.808
λ	0.71	0.699

Elasticity Implications

		KT 1	KT 2	Count 1	Count 2
Own Price	1	-2.60	-3.00	-1.91	-2.31
	2	-2.69	-2.94	-2.79	-3.08
	3	-2.11	-1.94	-2.90	-2.89
	4	-2.66	-2.66	-2.78	-2.70
	5	-2.36	-2.57	-3.86	-3.36
Income	1	2.61	2.92	0.71	0.70
	2	2.85	2.75	0.71	0.70
	3	2.01	1.89	0.71	0.70
	4	2.75	2.70	0.71	0.70
	5	2.39	2.53	0.71	0.70
Own Quality	1	0.49	0.12	0.91	0.21
	2	1.63	0.31	1.79	1.06
	3	1.14	0.15	1.90	1.72
	4	0.79	0.16	1.78	0.51
	5	1.04	0.23	2.86	1.04

Welfare Estimates (\$'s)

Scenario	KT Models		Count Models	
	Spec. 1	Spec. 2	Spec. 1	Spec. 2
\$50 fee at Sites 1 and 2	-76.67 (2.89)	-84.28 (3.11)	-60.04 (4.53)	-65.29 (5.51)
20% increase in Pheasants at Site 2	16.47 (2.38)	1.98 (0.53)	11.81 (0.92)	9.57 (0.97)