Lecture 7
Count Data Models

Count Data Models

• Counts are non-negative integers. They represent the number of occurrences of an event within a fixed period.

• Examples:
- Number of "jumps" (higher than 2*σ) in stock returns per day.
- Number of trades in a time interval.
- Number of a given disaster –i.e., default- per month.
- Number of crimes on campus per semester.

Note: We have *rare events*, in general, far from normal distributed data.

• The Poisson distribution is often used for these type of data.

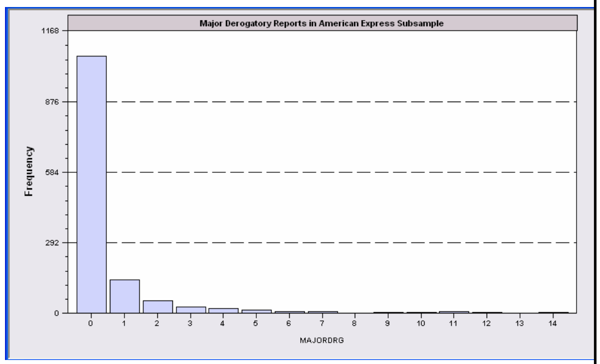
• Goal: Model count data as a function of covariates, X.

Count Data Models – Data (Greene)

AmEx Credit Card Holders

N = 13,777

Number of major derogatory reports in 1 year



- Issues:
  - Nonrandom selection
  - Excess zeros

Note: In general, far from normal distributed data.

Count Data Models – Data (Greene)

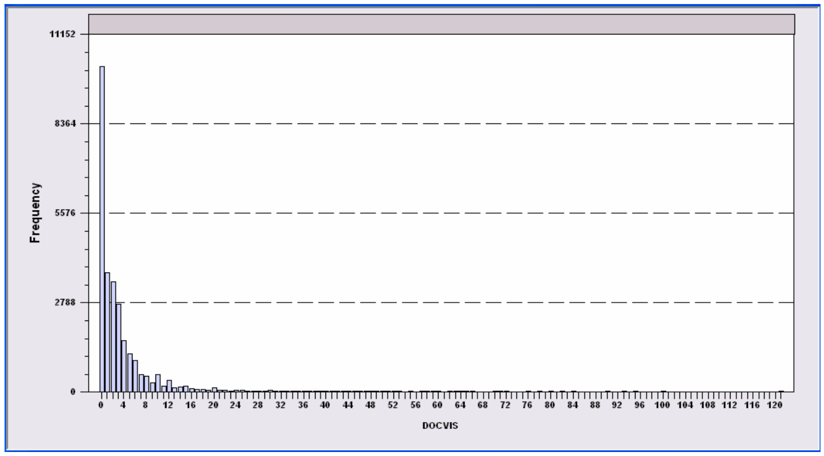
- Histogram for Credit Data

Histogram for MAJORDRG NOBS= 1310, Too low: 0, Too high: 0

Bin	Lower limit	Upper limit	Frequency	Cumulative Freq
0	.000	1.000	1053 (.8038)	1053(.8038)
1	1.000	2.000	136 (.1038)	1189(.9076)
2	2.000	3.000	50 (.0382)	1239(.9458)
3	3.000	4.000	24 (.0183)	1263(.9641)
4	4.000	5.000	17 (.0130)	1280(.9771)
5	5.000	6.000	10 (.0076)	1290(.9847)
6	6.000	7.000	5 (.0038)	1295(.9885)
7	7.000	8.000	6 (.0046)	1301(.9931)
8	8.000	9.000	0 (.0000)	1301(.9931)
9	9.000	10.000	2 (.0015)	1303(.9947)
10	10.000	11.000	1 (.0008)	1304(.9954)
11	11.000	12.000	4 (.0031)	1308(.9985)
12	12.000	13.000	1 (.0008)	1309(.9992)
13	13.000	14.000	0 (.0000)	1309(.9992)
14	14.000	15.000	1 (.0008)	1310(1.0000)

Count Data Models – Data (Greene)

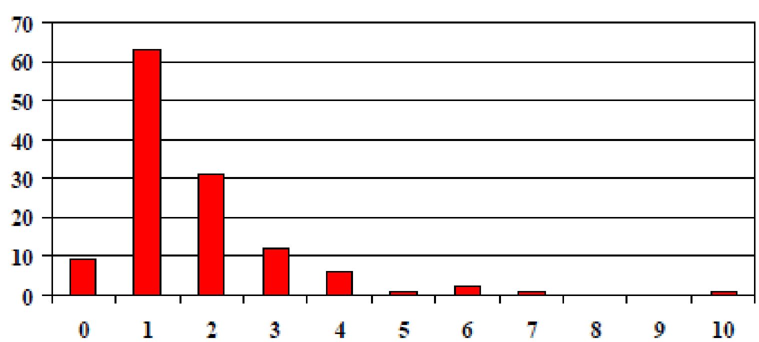
• Histogram for Credit Data



• Usual feature: Lots of zeros.

Count Data Models – Data

• Histogram for Takeover Bids –from Jaggia and Thosar (1993).



• Usual feature: Fat tails, far from normal.

Review: The Poisson Distribution

- Suppose events are occurring randomly and uniformly in time.
- The events occur with a known average.
- Let X be the number of events occurring (arrivals) in a fixed period of time (time-interval of given length).
- Typical example: X = number of crime cases coming before a criminal court per year (original Poisson's application in 1838.)
- Then, X will have a **Poisson distribution** with parameter λ.

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 0,1,2,3,...$$

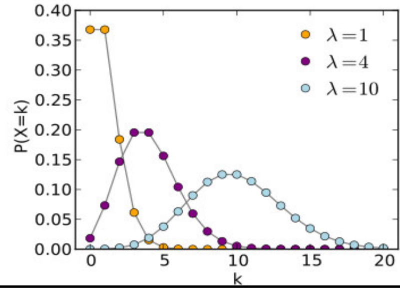
- The *intensity parameter,* λ, represents the expected number of occurrences in a fixed period of time –i.e., λ=E[X].
- It is also the variance of the count: λ=Var[X] => $\lambda>0$.
- Additive property holds.

Review: The Poisson Distribution

- **Example**:

  On average, a trade occurs every 15 seconds. Suppose trades are independent. We are interested in the probability of observing 10 trades in a minute (X=10). A Poisson distribution can be used with λ=4 (4 trades per minute).

- **Poisson probability function**



Note: As λ increases, the Poisson distribution approximates a normal distribution.

Review: The Poisson Distribution

• We can come up with the Poisson model by thinking of events counts as counts of rare events.

• Specifically, a Poisson RV approximates a binomial RV when the binomial parameter N (number of trials) is large and p (probability of a success) is small.

• The *Law of Rare Events*.

Count Data Models & Duration Models

• From the Math Review: There is a relation between counts and durations (or waiting time between events).

• If for every $t > 0$ the number of arrivals in the time interval $[0,t]$ follows the Poisson distribution with mean λt, then the sequence of inter-arrival times are *i.i.d.* exponential RVs having mean $1/\lambda$.

• We can also model duration data as a function of covariates, X. Many times which approach to use depends on the data available.

Poisson Regression Model

• Goal: Model count data as a function of covariates, X. The benchmark model is the Poisson model.

Q: Why do we need special models? What is wrong with OLS?
Like in probit and logit models, the dependent variable has restricted support. OLS regression can/will predict values that are negative and will also predict non-integer values. Nonsense results.

• Given the Poisson distribution, we model the mean –i.e., λ- as a function of covariates. This creates the Poisson regression model:

$$P(Y_i = j \mid X_i) = \frac{\lambda_i^{\,j} e^{-\lambda_i}}{j!}, \qquad j = 0,1,2,3,...$$

$\lambda_i = E[Y_i \mid X_i] = Var[Y_i \mid X_i] = \exp(X_i'\beta)$ => we make sure λ_i>0.

Poisson Regression Model - Estimation

• We usually model $\lambda_i = \exp(\mathbf{X}_i'\boldsymbol{\beta})>0$, but other formulations OK.

=> $y_i = \exp(\mathbf{X}_i'\boldsymbol{\beta}) + \varepsilon_i$

• We have a non-linear regression model, but with heteroscedasticity –i.e., $Var[\varepsilon_i|\mathbf{X}_i] = \lambda_i = \exp(\mathbf{X}_i'\boldsymbol{\beta})$ => G-NLLS is possible.

• ML is typically done. The log likelihood is given by:

$$LogL(\beta) = \sum_{i=1}^{N} y_i X_i'\beta + \exp(X_i'\beta) - \ln(y_i!)$$

• The f.o.c.'s are:

$$\frac{\delta LogL\ (\beta)}{\delta \beta'} = \sum_{i=1}^{N} \{ y_i - \exp(X_i'\beta)\} X_i = 0$$

Poisson Regression Model – Estimation

• The s.o.c.'s are:

$$\frac{\delta LogL\ (\beta)}{\delta\beta\delta\beta\ '} = \sum_{i=1}^{N} -\exp(X_i{}'\beta) X_i X_i{}'$$

The LogL is globally concave => a unique maximum. Likely, fast convergence.

• The usual ML theory yields β_{MLE} asymptotically normal with mean β and variance given by the inverse of the information matrix:

$$Var[\beta \mid X] = \left(\sum_{i=1}^{N} \exp(X_i{}'\beta) X_i X_i{}' \right)^{-1}$$

Note: For consistency of the MLE, we only require that conditional mean of y_i is correctly specified; -i.e., it need not be Poisson distributed. But, the ML standard errors will be incorrect.

Poisson Regression Model – Partial Effects

• As usual, to interpret the coefficients, we calculate partial effects (delta method or bootstrapping for standard errors):

$$\frac{\delta\{\lambda_i = E[Y_i \mid X_i]\}}{\delta X_{i,k}} = \lambda_i \beta_k$$

• We estimate the partial effects at the mean of the **X** or at average.

• While the parameters do not indicate the marginal impact, their relative sizes indicate the relative strength of each variable's effect:

$$\frac{\dfrac{\delta\{\lambda_i = E[Y_i \mid X_i]\}}{\delta X_{i,k}}}{\dfrac{\delta\{\lambda_i = E[Y_i \mid X_i]\}}{\delta X_{i,j}}} = \frac{\lambda_i \beta_k}{\lambda_i \beta_j} = \beta_k / \beta_j$$

Poisson Regression Model - Evaluation

- LR test to compare restricted and unrestricted models
- AIC, BIC
- McFadden pseudo-R^2.= 1 – LogL(β)/LogL(0)
- Predicted probabilities
- G^2 (Sum of model deviances):

$$G^2 = 2\sum_{i=1}^{N} y_i \ln(y_i / \bar{\lambda})$$

=> equal to zero for a model with perfect fit.

- One implication of the Poisson assumption:

$\text{Var}[y_i|\mathbf{x}_i] = E[y_i|\mathbf{x}_i]$ (*equi-dispersion*)

=> check this assumption, if it does not hold, Poisson model is inappropriate.

Poisson Regression Model – Example (Greene)

```
----------------------------------------------------------------------
Poisson Regression
Dependent variable                 DOCVIS
Log likelihood function   -103727.29625
Restricted log likelihood -108662.13583
Chi squared [   6 d.f.]      9869.67916
Significance level               .00000
McFadden Pseudo R-squared      .0454145
Estimation based on N =  27326, K =   7
Information Criteria: Normalization=1/N
              Normalized   Unnormalized
AIC              7.59235   207468.59251
Chi- squared =255127.59573  RsqP= .0818
G  - squared =154416.01169  RsqD= .0601
Overdispersion tests: g=mu(i)  : 20.974
Overdispersion tests: g=mu(i)^2: 20.943
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er.  P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
Constant|      .77267***       .02814       27.463   .0000
     AGE|      .01763***       .00035       50.894   .0000      43.5257
    EDUC|     -.02981***       .00175      -17.075   .0000      11.3206
  FEMALE|      .29287***       .00702       41.731   .0000       .47877
 MARRIED|      .00964          .00874        1.103   .2702       .75862
  HHNINC|     -.52229***       .02259      -23.121   .0000       .35208
  HHKIDS|     -.16032***       .00840      -19.081   .0000       .40273
--------+-------------------------------------------------------------
```

Poisson Regression Model – Example (Greene)

• Alternative Covariance Matrices

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
	Standard - Negative Inverse of Second Derivatives				
Constant	.77267***	.02814	27.463	.0000	
AGE	.01763***	.00035	50.894	.0000	43.5257
EDUC	-.02981***	.00175	-17.075	.0000	11.3206
FEMALE	.29287***	.00702	41.731	.0000	.47877
MARRIED	.00964	.00874	1.103	.2702	.75862
HHNINC	-.52229***	.02259	-23.121	.0000	.35208
HHKIDS	-.16032***	.00840	-19.081	.0000	.40273
	Robust - Sandwich				
Constant	.77267***	.08529	9.059	.0000	
AGE	.01763***	.00105	16.773	.0000	43.5257
EDUC	-.02981***	.00487	-6.123	.0000	11.3206
FEMALE	.29287***	.02250	13.015	.0000	.47877
MARRIED	.00964	.02906	.332	.7401	.75862
HHNINC	-.52229***	.06674	-7.825	.0000	.35208
HHKIDS	-.16032***	.02657	-6.034	.0000	.40273
	Cluster Correction				
Constant	.77267***	.11628	6.645	.0000	
AGE	.01763***	.00142	12.440	.0000	43.5257
EDUC	-.02981***	.00685	-4.355	.0000	11.3206
FEMALE	.29287***	.03213	9.116	.0000	.47877
MARRIED	.00964	.03851	.250	.8023	.75862
HHNINC	-.52229***	.08295	-6.297	.0000	.35208
HHKIDS	-.16032***	.03455	-4.640	.0000	.40273

Poisson Regression Model – Example (Greene)

• Partial Effects $\dfrac{\partial E[y_i \mid \mathbf{x}_i]}{\partial \mathbf{x}_i} = \lambda_i \boldsymbol{\beta}$

```
Partial derivatives of expected val. with
respect to the vector of characteristics.
Effects are averaged over individuals.
Observations used for means are All Obs.
Conditional Mean at Sample Point    3.1835
Scale Factor for Marginal Effects   3.1835
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
AGE	.05613***	.00131	42.991	.0000	43.5257
EDUC	-.09490***	.00596	-15.923	.0000	11.3206
FEMALE	.93237***	.02555	36.491	.0000	.47877
MARRIED	.03069	.02945	1.042	.2973	.75862
HHNINC	-1.66271***	.07803	-21.308	.0000	.35208
HHKIDS	-.51037***	.02879	-17.730	.0000	.40273

<u>Note</u>: With dummies, partial effects are calculated as differences.

Poisson Model: Issues

• The Poisson model has several restrictive assumptions
- All events are independent
- Constant arrival rate, λ.
- No limit on the number of occurrences
- In the Binomial formulation, N goes to infinity.

• Herding behavior violates independence. We see an IPO (or a zebra), it is very likely we will see more. This is called *positive contagion*. It increases the variance of the count.

• Uneven (arbitrary) time periods can create contagion and thus increase the variance.

Poisson Model: Issues

• Heterogeneity can violate the constant arrival rate assumption. For example, a CEO is more likely to reject a hostile bid early in her tenure (the Board that elected the CEO will be more supportive) than later. Unobserved heterogeneity increases the count's variance.

• In many cases, there is an upper limit to the number of possible events, M_i. A CEO can only reject a hostile bid, if there is a hostile bid. Thus, the maximum number of hostile bid rejections is 10 if there are 10 hostile bids.

• This maximum number is called an observation's *exposure*. It can be incorporated as $E[y_i|\mathbf{x}_i] = \lambda_i = \exp(\mathbf{X}_i'\boldsymbol{\beta}) * M_i = \exp[\mathbf{X}_i'\boldsymbol{\beta}+\ln(M_i)]$

Poisson Model: Overdispersion

• One implication of the Poisson model is equi-dispersion. That is, the mean and variance are equal: $\text{Var}[y_i|\mathbf{x}_i] = E[y_i|\mathbf{x}_i]$

• But, the first three cases (herding, uneven periods, heterogeneity) tend to cause *overdisperion*. That is,

$$\text{Var}[y_i|\mathbf{x}_i] > E[y_i|\mathbf{x}_i].$$

• It is not rare to see overdispersion ('extra' heterogeneity) in the data:

- A few traders will do many trades, many traders will do a few.
- A few assets will have many jumps, many assets will have few.

• Under overdispersion: Standard errors and p-values are too small.

Poisson Model: Overdispersion - Testing

• Check for overdisperion:

- Check overdispersion rate: $\text{Var}[y_i]/E[y_i]$ (in general, relative to *df.)*
- Cameron and Trivedi's (CT) (1990) test.

• CT's (1990) test. It is based on the assumption that under the Poisson model $\{(y\text{-}E[y])^2 - E[y]\}$ has zero mean:

H_0 (Poisson Model correct): $\text{Var}[y_i] = E[y_i]$

H_A: $\text{Var}[y_i] = E[y_i] + \alpha\, g(E[y_i])$

Simple linear regression: $\{(y\text{-}E[\hat{y}])^2 - y\}/\{E[\hat{y}]\ \text{sqrt}(2)\}$ against some $g(E[y_i])$, usually a linear –*g*=mu(i)- or quadratic function –*g*=mu(i)^2 .

• CT's rule of thumb: If $\text{Var}[y_i]/E[y_i] > 2$ => overdispersion

Poisson Model: Dealing with Overdispersion

- When overdispersion occurs, we modify the model:

- Keep Poisson model, but add ad-hoc models for the variance. For example,

$$\text{Var}[y_i] = \phi\, \lambda_i \,,$$

where $\hat{\phi} = \frac{1}{N-k} \sum_i \frac{(y_i - \hat{\lambda}_i)^2}{\hat{\lambda}_i}$ (NB-1 Model)

Then, use ML estimation. If the mean and variance are correctly specified, β_{ML} will have the usual good properties.

- Specify an alternative distribution that can generate overdispersion.

Poisson Model: Dealing with Overdispersion

• Specify an alternative distribution that can generate overdispersion. Usual alternative distributions:

(1) Assume the overdispersion is gamma distributed across means—resulting in a negative binomial model (or Poisson-gamma model)

(2) Assume the overdispersion is normally distributed (Poisson-normal model).

Poisson Model: Summary of Issues

• Overdispersion
 – Usually attributed to omitted and/or unobserved heterogeneity
 – Can use Poisson ML estimates with corrected standard errors
 – Alternatively, can use models without equidispersion
  - Negative Binomial
  - Mixed Poisson

• Truncation (especially, zero-truncation) –number of mergers and acquisitions. We only sample from M&A's. A Poisson model would falsely allow Prob[Y_i=0]>0.

• Excess zeros –data generated by two process: one for the "true zeros," and one for the "excess zeros."

• Correlated counts –i.e., *i.i.d.* assumption does not hold. Big count today is likely to be followed by a big count tomorrow.

Poisson Model: Omitted Heterogeneity

• An elegant solution to overdispersion, is the omitted (latent) heterogeneity. We model heterogeneity, by introducing a random effect on the expected mean:

$$\lambda_i^* = \exp(\mathbf{X}_i'\boldsymbol{\beta} + u_i) = \exp(\mathbf{X}_i'\boldsymbol{\beta} + u_i) = \lambda_i\, h_i,$$

where $h_i = \exp(u_i)$ follows a one parameter gamma distribution G(θ, θ), with mean=1 (same mean as in the Poisson model = λ_i) and variance=1/θ= α. Then,

RV

$$\text{Prob}[y=j|x,u] = \frac{\exp(-\lambda)\lambda^j}{j!}, \lambda = \exp(x'\beta + u)$$

$$\text{Prob}[y=j|x] = \int_u \text{Prob}[y=j|x,u]f(u)du$$ we need *f*(u) to integrate

$$\text{If } f(\exp(u)) = \frac{\alpha^\alpha \exp(-\alpha u)u^{\alpha-1}}{\Gamma(\alpha)} \text{ (Gamma with mean 1)}$$ α>0

Then Prob[y=j|x] is negative binomial.

• Note: when α=0, we are back to the Poisson model.

Negative Binomial Model

• The Negative Binomial Distribution

$$P(Y_i = j \mid X_i) = \frac{\Gamma(\theta + y_i)}{\Gamma(y_i + 1)\Gamma(\theta)} r_i^{y_i} (1 - r_i)^{\theta}, \qquad j = 0,1,2,3,\ldots$$

$$\lambda_i = \exp(X_i{}'\beta) \qquad \& \qquad r_i = \frac{\lambda_i}{\lambda_i + \theta}$$

• Characteristics:
- $\text{Prob}(Y_i=j|x_i)$ has greater mass to the right and left of the mean.
- Conditional mean function is the same as the Poisson:
  $E[y_i|x_i] = \lambda_i = \text{Exp}(\mathbf{X}_i{}'\boldsymbol{\beta})$, => same partial effects.
  $\text{Var}[y_i|x_i] = \lambda_i\,(1 + \alpha\,\lambda_i) > \lambda_i$ (a squared in the Var[.])

• The Negative Binomial (NegBin) Model can accommodate over- and under-dispersion; at the cost of an additional parameter ($\alpha=1/\theta$).

Negative Binomial Model

• The Negative Binomial (NegBin) Model can accommodate overdispersion. The model has an additional parameter ($\alpha=1/\theta$).
- $\text{Var}[y_i]/E[y_i] = \{1+ \alpha\, E[y_i]\}$ ($\alpha=0$ => Poisson model, again)

• There are alternative parameterizations of the negative binomial, with different variance functions. The one above is called the Negbin-2 (NB-2) model by Cameron and Trivedi (1986).

• Different models can be generated by specifying different distributions for u_i. For example, u_i follows an inverse Gaussian distribution -Dean et al. (1989). This Poisson-Inverse Gaussian model has heavier tail than the NegBin model.

NegBin Model – NB-P

• Without the heterogeneity argument, we could have introduced directly the NegBin distribution as Prob(Y_i=j|x_i) is the NegBin pdf.

• Along this line of thinking, Cameron and Trivedi (1998) make a generalization, the NB-P model, where $\theta = \theta_i \lambda_i^{2-P}$

• Then, we have the Negbin P (NB-P) model:

$$P(Y_i = j \mid X_i) = \frac{\Gamma(\theta_i \lambda_i^{2-P} + y_i)}{\Gamma(y_i + 1)\Gamma(\theta_i \lambda_i^{2-P})} r_i^{y_i} (1 - r_i)^{\theta_i \lambda_i^{2-P}}, \qquad j = 0,1,2,3,\ldots$$

• NB-2 is a special case, P=2. The conditional mean is still λ_i and the conditional variance is:

$$\text{Var}[y_i \mid x_i] = \lambda_i \, [1 + (1/\theta_i)\, \lambda_i^{2-P}]$$

where θ_i can be modeled as a function of some driving variables, $\mathbf{z}_i$

NegBin Model – NB-P

• By letting $\theta_i = f(\mathbf{z}_i)$, we generalize the NegBin model. For example,

$\theta_i = \exp(\mathbf{z}_i'\gamma)$ => we are modeling the variance.

• These models are called *Generalized Negative Binomial Model.*

• The NB-1 and NB-2 models are non-nested. Vuong (1989) test is a possibility:

$$V = [\text{sqrt}(N) \text{ mean}(m_i)]/s_m, \xrightarrow{d} N(0,1)$$

where $m_i = LogL(\text{NB-2}) - LogL(\text{NB-1})$

• Large values favor the NB-2 model. In applications, Greene (2007) finds that this statistic is rarely outside the inconclusive region (-1.96 to +1.96).

NegBin Model – Estimation

• Estimation: Maximum Likelihood
- For the NB-2, we have

$$LogL(\beta,\theta)=\sum_{i=1}^{N}\ln\left(\frac{\Gamma(\theta+y_i)}{\Gamma(y_i+1)\Gamma(\theta)}\right)+y_i\ln(\frac{\lambda_i}{\lambda_i+\theta})+\theta\ln(\frac{\theta}{\lambda_i+\theta})$$

The f.o.c.'s are straightforward and the resulting variance-covariance matrix is block diagonal.

Note: Poisson is consistent when NegBin is appropriate. Therefore, this is a case for the Robust covariance matrix estimator. (Neglected heterogeneity that is uncorrelated with x_i.)

NegBin Model – Model Evaluation

• Model Evaluation as usual:
- LR, W, and LM tests
- AIC, BIC
- pseudo-R^2

• Testing the NegBing Model.
- Relative to the Poisson model, we have an extra parameter in the NegBin model, α.

- We can use a LR-test to test H_0: α=0. This tests the NegBin model.
- A Wald test will also work

• For non-nested models (NB-1 vs. NB-2), use Vuong test.

Negative Binomial Model – Example (Greene)

```
----------------------------------------------------------------------
Negative Binomial Regression
Dependent variable                DOCVIS
Log likelihood function    -60134.50735  NegBin  LogL
Restricted log likelihood -103727.29625  Poisson LogL
Chi squared [   1 d.f.]     87185.57782  Reject Poisson model
Significance level               .00000
McFadden Pseudo R-squared      .4202634
Estimation based on N =  27326, K =   8
Information Criteria: Normalization=1/N
              Normalized   Unnormalized
AIC              4.40185   120285.01469
NegBin form 2; Psi(i) = theta
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
Constant|     .80825***       .05955       13.572   .0000
     AGE|     .01806***       .00079       22.780   .0000      43.5257
    EDUC|    -.03717***       .00386       -9.622   .0000      11.3206
  FEMALE|     .32596***       .01586       20.556   .0000       .47877
 MARRIED|    -.00605          .01880        -.322   .7477       .75862
  HHNINC|    -.46768***       .04663      -10.029   .0000       .35208
  HHKIDS|    -.15274***       .01729       -8.832   .0000       .40273
        |Dispersion parameter for count data model
   Alpha|    1.89679***       .01981       95.747   .0000
--------+-------------------------------------------------------------
```

Negative Binomial Model – Example (Greene)

- Partial Effects Should Be the Same

```
+---------------------------------------------------------------------
Scale Factor for Marginal Effects  3.1835   POISSON
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
     AGE|     .05613***       .00131       42.991   .0000      43.5257
    EDUC|    -.09490***       .00596      -15.923   .0000      11.3206
  FEMALE|     .93237***       .02555       36.491   .0000       .47877
 MARRIED|     .03069          .02945        1.042   .2973       .75862
  HHNINC|   -1.66271***       .07803      -21.308   .0000       .35208
  HHKIDS|    -.51037***       .02879      -17.730   .0000       .40273
--------+-------------------------------------------------------------
Scale Factor for Marginal Effects  3.1924  NEGATIVE BINOMIAL
--------+-------------------------------------------------------------
     AGE|     .05767***       .00317       18.202   .0000      43.5257
    EDUC|    -.11867***       .01348       -8.804   .0000      11.3206
  FEMALE|    1.04058***       .06212       16.751   .0000       .47877
 MARRIED|    -.01931          .06382        -.302   .7623       .75862
  HHNINC|   -1.49301***       .16272       -9.176   .0000       .35208
  HHKIDS|    -.48759***       .06022       -8.097   .0000       .40273
--------+-------------------------------------------------------------
```

Formulations for Negative Binomial (Greene)

Poisson

$$\text{Prob}[Y = y_i \mid \mathbf{x}_i] = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{\Gamma(1+y_i)},$$

$$\lambda_i = \exp(\alpha + \mathbf{x}_i'\boldsymbol{\beta}),\ y_i = 0,1,\ldots,i = 1,\ldots,N$$

$$E[y \mid \mathbf{x}_i] = Var[y \mid \mathbf{x}_i] = \lambda_i$$

$$\text{E}[\text{y}_\text{i} \mid \mathbf{x}_\text{i}] = \lambda_\text{i}$$

Negative Binomial – 2

$$E[y_i|\mathbf{x}_i,\varepsilon_i] = \exp(\alpha+\mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i) = h_i\lambda_i,$$

$$\text{Prob}[Y = y_i|\mathbf{x}_i] = \frac{\Gamma(\theta + y_i)r_i^{\theta}(1-r_i)^{y_i}}{\Gamma(1+y_i)\Gamma(\theta)},$$

$$y_i = 0,1,\ldots,\theta > 0,$$

$$r_i = \theta/(\theta+\lambda_i)$$

$$E[y_i|\mathbf{x}_i] = \lambda_i, \quad \text{Var}[y_i|\mathbf{x}_i] = \lambda_i\,[1 + (1/\theta)\lambda_i]$$

$$= \lambda_i\,[1 + \kappa\,\lambda_i]$$

$$\kappa = \text{Var}[h_i].$$

Negative Binomial – 1

$$\text{Var}[y_i|\mathbf{x}_i] = \lambda_i + \kappa\,\lambda_i^{1} = \lambda_i\,[1 + \kappa],$$

Replace θ with $\theta\lambda_i$ in NB-2.

$$\text{Prob}[Y = y_i \mid \mathbf{x}_i] = \frac{\Gamma(\theta\lambda_i + y_i)q^{\theta\lambda_i}(1-q)^{y_i}}{\Gamma(y_i+1)\Gamma(\theta\lambda_i)}$$

$$y_i = 0,1,\ldots;\ q = 1/1(1+\theta).$$

$$\text{E}[y_i|\mathbf{x}_i] = \lambda_i$$

Replace θ with $\theta\lambda_i^{2-P}$ in NB-1

$$\text{Prob}[Y = y_i \mid \mathbf{x}_i] = \frac{\Gamma(\theta\lambda_i^{2-P} + y_i)s_i^{\theta\lambda_i^{2-P}}(1-s_i)^{y_i}}{\Gamma(y_i+1)\Gamma(\theta\lambda_i^{2-P})}$$

$$s_i = \frac{\lambda_i}{\lambda_i + \theta\lambda_i^{2-P}}.$$

$$\text{E}[y_i|\mathbf{x}_i] = \lambda_i$$

$$\text{Var}[y_i|\mathbf{x}_i] = \lambda_i\,[1 + (1/\theta)\lambda_i^{P-1}].$$

Formulations: NegBin-1 Model (Greene)

```
----------------------------------------------------------------------
Negative Binomial Regression
Dependent variable                 DOCVIS
Log likelihood function     -60025.78734
Restricted log likelihood -103727.29625
NegBin form 1; Psi(i) = theta*exp[bx(i)]
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
Constant|      .62584***        .05816       10.761   .0000
     AGE|      .01428***        .00073       19.462   .0000     43.5257
    EDUC|     -.01549***        .00359       -4.314   .0000     11.3206
  FEMALE|      .33028***        .01479       22.328   .0000      .47877
 MARRIED|      .04324**         .01852        2.335   .0196      .75862
  HHNINC|     -.24543***        .04540       -5.406   .0000      .35208
  HHKIDS|     -.14877***        .01745       -8.526   .0000      .40273
        |Dispersion parameter for count data model
   Alpha|     6.09246***        .06694       91.018   .0000
--------+-------------------------------------------------------------
```

Formulations: NegBin-P Model (Greene)

```
------------------------------------------------
Negative Binomial (P) Model
Dependent variable                 DOCVIS
Log likelihood function    -59992.32903
Restricted log likelihood -103727.29625
Chi squared [   1 d.f.]     87469.93445
--------+----------------------------------------
-
Variable| Coefficient    Standard Error  b/St.Er.
--------+----------------------------------------
-
Constant|    .60840***        .06452        9.429
     AGE|    .01710***        .00082       20.782
    EDUC|   -.02313***        .00414       -5.581
  FEMALE|    .36386***        .01640       22.187
 MARRIED|    .03670*          .02030        1.808
  HHNINC|   -.35093***        .05146       -6.819
  HHKIDS|   -.16902***        .01911       -8.843
        |Dispersion parameter for count data
model
   Alpha|   3.85713***        .14581       26.453
        |Negative Binomial. General form, NegBin
P
       P|   1.38693***        .03142       44.140
--------+----------------------------------------
```

NB-2	NB-1	Poisson
.80825***	.62584***	.77267***
.01806***	.01428***	.01763***
-.03717***	-.01549***	-.02981***
.32596***	.33028***	.29287***
-.00605	.04324**	.00964
-.46768***	-.24543***	-.52229***
-.15274***	-.14877***	-.16032***
Dispersion	Dispersion	
1.89679***	6.09246***	

Partial Effects for Different Models (Greene)

```
Scale Factor for Marginal Effects  3.1835   POISSON
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
     AGE|     .05613***       .00131       42.991   .0000      43.5257
    EDUC|    -.09490***       .00596      -15.923   .0000      11.3206
  FEMALE|     .93237***       .02555       36.491   .0000       .47877
 MARRIED|     .03069          .02945        1.042   .2973       .75862
  HHNINC|   -1.66271***       .07803      -21.308   .0000       .35208
  HHKIDS|    -.51037***       .02879      -17.730   .0000       .40273
--------+-------------------------------------------------------------
Scale Factor for Marginal Effects  3.1924  NEGATIVE BINOMIAL - 2
     AGE|     .05767***       .00317       18.202   .0000      43.5257
    EDUC|    -.11867***       .01348       -8.804   .0000      11.3206
  FEMALE|    1.04058***       .06212       16.751   .0000       .47877
 MARRIED|    -.01931          .06382        -.302   .7623       .75862
  HHNINC|   -1.49301***       .16272       -9.176   .0000       .35208
  HHKIDS|    -.48759***       .06022       -8.097   .0000       .40273
--------+-------------------------------------------------------------
Scale Factor for Marginal Effects  3.0077  NEGATIVE BINOMIAL - P
     AGE|     .05143***       .00246       20.934   .0000      43.5257
    EDUC|    -.06957***       .01241       -5.605   .0000      11.3206
  FEMALE|    1.09436***       .04968       22.027   .0000       .47877
 MARRIED|     .11038*         .06109        1.807   .0708       .75862
  HHNINC|   -1.05547***       .15411       -6.849   .0000       .35208
  HHKIDS|    -.50835***       .05753       -8.836   .0000       .40273
```

Issues: Truncation

• Often, because of the way we collect data, we only observe $y_i \geq 1$. For example, we study M&A. We collect data on actual M&A offers.

• Good sample to get information on the decision to go for a M&A, but we get no information on the M&A offers that do no go through.

• Our data is truncated at 0.

• If we use a Poisson/NB model, we need to incorporate this fact. We need to use the zero-truncated Poisson/NB model.

$$P(y_i = j \mid y_i > 0, X_i) = \frac{P(y_i = j \,\&\, y_i > 0 \mid X_i)}{P(y_i > 0 \mid X_i)} = \frac{P(y_i = j \mid X_i)}{[1 - P(y_i = 0 \mid X_i)]} = \frac{\exp(X_i'\beta)^j e^{-(X_i'\beta)}}{j![1 - e^{-(X_i'\beta)}]}, \quad j = 1,2,3,...$$

• We increase each unconditional probability by the factor [1-$f(0)$], so the probability mass of the truncated distribution adds up to 1.

Issues: Truncation

• The unconditional count is:
$E[y_i | X_i] = \lambda_i$

• The conditional count is:
$E[y_i | y_i > 0, X_i] = \lambda_i / [1 - \exp(-\lambda_i)]$

• ML estimation is straightforward.

Issues: Excess zeros

• Often the numbers of zeros in the sample cannot be accommodated properly by a Poisson or Negative Binomial model. Both models would underpredict them.

• There is said to be an "*excess zeros*" problem. New models are needed to deal with these type of data.

• These models, called *Two-part models,* allow for two different process: one drives whether the value is 0 or positive (participation part), and the other one drives the value of the strictly positive count (amount part).

• Proposed models:
- *Zero inflated*
- *Hurdle models*

Zero Inflation – ZIP Models

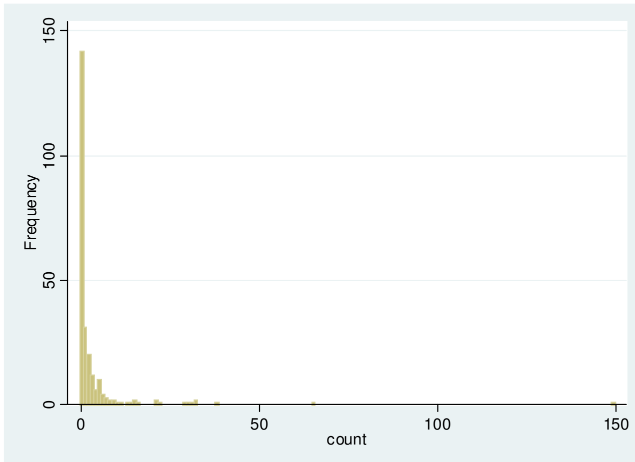
• Zero-inflated model have two kinds of zeros: "true zeros" and "excess zeros."

• Two groups of people: Always Zero & Not Always Zero
Example: Investors (traders) who sometime just did not trade that week versus investors who never ever do.

• Two models: (1) for the count and (2) for excess zeros. The key difference is that the count model allows zeros now. It is not a truncated count model, but allows for "corner solutions."

• If we are interested in modeling trading, the zeros from investors who will never trade are not relevant. But, we only observe the zero, not the type of investor. This is the *excess zeros problem.*

Zero Inflation – ZIP Models



• Note: lots of zeros.

Zero Inflation Poisson (ZIP) Models

• We are interested in a stock trading per week model for investors. Two regimes (distributions) for the two types of investors (or zeros):
(1) Degenerate at zero (prob[0]=1). (For investors that never trade.)
(2) Poisson (For traders, 0 is possible)

• We convert this problem into a latent variable model.

$d_i^* = \mathbf{w}_i'\boldsymbol{\delta} + u_i$, $\quad u_i \sim N(0, \sigma^2)$

$d_i = I[d_i^*>0]$, $\quad i$ trader if $d_i^* > 0$.

- Participation part (Always Zero or Not Always Zero):
  - $\text{Prob}[d_i = 0 \,|\, \mathbf{w}_i] = \Pi(\mathbf{w}_i'\boldsymbol{\delta})$
    $\text{Prob}[d_i = 1 \,|\, \mathbf{w}_i] = 1 - \Pi(\mathbf{w}_i'\boldsymbol{\delta})$

=> we can use a logit or a probit to model $\Pi(\mathbf{w}_i'\boldsymbol{\delta})$.

Zero Inflation Poisson (ZIP) Models

- Amount part:
  - $y_i^* | \mathbf{x}_i \sim f_P$ =Poisson (latent Poisson, NB also possible)
    $\lambda_i = \exp(\mathbf{x}_i'\boldsymbol{\beta})$
  - $y_i = d_i\, y_i^*$ (y_i is observed, along with $\mathbf{x}_i, \mathbf{w}_i$)

$$P(y_i = j \mid X_i, d_i = 1) = \frac{\exp(X_i'\beta)^j e^{-(X_i'\beta)}}{j!}$$

- Mixing groups (d_i=0 (Always Zero) & d_i=1 (Not Always Zero)):
  - Conditional probability of 0 --i.e., Prob[Y_i = 0|w_i,x_i,d_i]
    - Prob[Y_i = 0|w_i,x_i,d_i=0] = 1 (no trade, if no participation)
    - Prob[Y_i = y_i|w_i,x_i,d_i=1] = Prob[y_i^*|x_i] = $f_P(y_i)$ (Poisson)

Zero Inflation Poisson (ZIP) Models

- Mixing groups (continuation)
- Unconditional probability of 0 --i.e., Prob[Y_i = 0|w_i,x_i,d_i]
  - Prob[Y_i=0|w_i,x_i] = 1*Prob[d_i=0] + Pr[0|d_i=1] * Prob[d_i=1]
    = $\Pi(w_i'\delta)$ + f_P (Y_i=0) * [1-$\Pi(w_i'\delta)$]
    = $\Pi(w_i'\delta)$ + exp(-λ_i) [1-$\Pi(w_i'\delta)$]
  - Prob[Y_i=j|w_i,x_i] = 0*Prob[d_i=0] + Pr[Y_i=j|d_i=1] * Prob[d_i=1]
    = f_P (Y_i=j) * [1-$\Pi(w_i'\delta)$]
    = [exp(- λ_i) λ_i^j/j!] [1-$\Pi(w_i'\delta)$]

- Expectation & Variance of counts:
  - E[Y_i=j|w_i,x_i] = 0*Prob[d_i=0] + λ_i*Prob[d_i=1] = λ_i * [1-$\Pi(w_i'\delta)$]
  - Var[Y_i=j|w_i,x_i] = λ_i * [1-$\Pi(w_i'\delta)$] * [1+$\lambda_i\Pi(w_i'\delta)$]

Zero Inflation Poisson (ZIP) Models

• Overdispersion

$Var[Y_i=j|w_i,x_i] / E[Y_i=j|w_i,x_i] = 1 + \lambda_i \Pi(w_i'\delta)$

- The more likely the Always Zero regime, the greater the overdispersion.

• Partial effects

- $\delta E[Y_i=j|w_i,x_i]/\delta x_{ik} = \lambda_i * [1-\Pi(w_i'\delta)] \beta_k$
- $\delta E[Y_i=j|w_i,x_i]/\delta w_{ik} = \lambda_i * [\delta\Pi(w_i'\delta)/\delta w_{ik}] \delta_k$

• Similar results are obtained for the Zero-inflation NegBin model (ZINB).

Two Forms of Zero Inflation Models

• Different ways of thinking of w_i (determinants of Π) and x_i (determinants of the amount *j*), generate different models. The ZIP-tau model, allows for the same determinants, but scales the β's in the Π model.

ZIP - tau = ZIP(τ)

$$\text{Prob}(y_i = j | x_i) = \frac{\exp(-\lambda_i)\lambda_i^j}{j!}, \boxed{\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i)}$$

$$\text{Prob}(0 \text{ regime}) = \boxed{F(\tau\boldsymbol{\beta}'\mathbf{x}_i)}$$

Zero Inflation = ZIP

$$\text{Prob}(y_i = j | x_i) = \frac{\exp(-\lambda_i)\lambda_i^j}{j!}, \boxed{\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i)}$$

$$\text{Prob}(0 \text{ regime}) = \boxed{F(\boldsymbol{\gamma}'\mathbf{z}_i)}$$

Notes on Zero Inflation Models (Greene)

- Poisson is not nested in ZIP. tau = 0 in ZIP(tau) or $\gamma = 0$ in ZIP does not produce Poisson; it produces ZIP with P(regime 0) = ½.
  - Standard tests are not appropriate
  - Use Vuong statistic. ZIP model almost always wins.

- Zero Inflation models extend to NB models – ZINB(tau) and ZINB are standard models
  - Creates two sources of overdispersion
  - Generally difficult to estimate

ZIP(τ) Model

```
----------------------------------------------------------------------
Zero Altered Poisson      Regression Model
Logistic distribution used for splitting model.
ZAP term in probability is F[tau x ln LAMBDA]
Comparison of estimated models
            Pr[0|means]       Number of zeros        Log-likelihood
Poisson          .04933   Act.= 10135 Prd.=  1347.9   -103727.29625
Z.I.Poisson      .35944   Act.= 10135 Prd.=  9822.1    -84012.30960
Note, the ZIP log-likelihood is not directly comparable.
ZIP model with nonzero Q does not encompass the others.
Vuong statistic for testing ZIP vs. unaltered model is     44.5723
Distributed as standard normal. A value greater than
+1.96 favors the zero altered Z.I.Poisson model.
A value less than -1.96 rejects the ZIP model.
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
        |Poisson/NB/Gamma regression model
Constant|    1.45145***        .01121      129.498   .0000
     AGE|     .01140***        .00013       86.245   .0000     43.5257
    EDUC|    -.02306***        .00075      -30.829   .0000     11.3206
  FEMALE|     .13129***        .00256       51.357   .0000      .47877
 MARRIED|    -.02270***        .00317       -7.151   .0000      .75862
  HHNINC|    -.41799***        .00898      -46.527   .0000      .35208
  HHKIDS|    -.08750***        .00322      -27.189   .0000      .40273
        |Zero inflation model
     Tau|    -.38910***        .00836      -46.550   .0000
--------+-------------------------------------------------------------
```

ZIP Model

```
----------------------------------------------------------------------
Zero Altered Poisson      Regression Model
Logistic distribution used for splitting model.
ZAP term in probability is F[tau x Z(i)     ]
Comparison of estimated models
            Pr[0|means]       Number of zeros        Log-likelihood
Poisson           .04933   Act.= 10135 Prd.=  1347.9   -103727.29625
Z.I.Poisson       .36565   Act.= 10135 Prd.=  9991.8    -83843.36088
Vuong statistic for testing ZIP vs. unaltered model is     44.6739
Distributed as standard normal. A value greater than
+1.96 favors the zero altered Z.I.Poisson model.
A value less than -1.96 rejects the ZIP model.
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
        |Poisson/NB/Gamma regression model
Constant|    1.47301***       .01123      131.119   .0000
     AGE|     .01100***       .00013       83.038   .0000      43.5257
    EDUC|    -.02164***       .00075      -28.864   .0000      11.3206
  FEMALE|     .10943***       .00256       42.728   .0000       .47877
 MARRIED|    -.02774***       .00318       -8.723   .0000       .75862
  HHNINC|    -.42240***       .00902      -46.838   .0000       .35208
  HHKIDS|    -.08182***       .00323      -25.370   .0000       .40273
        |Zero inflation model
Constant|    -.75828***       .06803      -11.146   .0000
  FEMALE|    -.59011***       .02652      -22.250   .0000       .47877
    EDUC|     .04114***       .00561        7.336   .0000      11.3206
--------+-------------------------------------------------------------
```

Partial Effects for Different Models

```
Scale Factor for Marginal Effects  3.1835   POISSON
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
     AGE|     .05613***       .00131       42.991   .0000      43.5257
    EDUC|    -.09490***       .00596      -15.923   .0000      11.3206
  FEMALE|     .93237***       .02555       36.491   .0000       .47877
 MARRIED|     .03069          .02945        1.042   .2973       .75862
  HHNINC|   -1.66271***       .07803      -21.308   .0000       .35208
  HHKIDS|    -.51037***       .02879      -17.730   .0000       .40273
--------+-------------------------------------------------------------
Scale Factor for Marginal Effects  3.1924  NEGATIVE BINOMIAL - 2
     AGE|     .05767***       .00317       18.202   .0000      43.5257
    EDUC|    -.11867***       .01348       -8.804   .0000      11.3206
  FEMALE|    1.04058***       .06212       16.751   .0000       .47877
 MARRIED|    -.01931          .06382        -.302   .7623       .75862
  HHNINC|   -1.49301***       .16272       -9.176   .0000       .35208
  HHKIDS|    -.48759***       .06022       -8.097   .0000       .40273
--------+-------------------------------------------------------------
Scale Factor for Marginal Effects  3.1149  ZERO INFLATED POISSON
     AGE|     .03427***       .00052       66.157   .0000      43.5257
    EDUC|    -.11192***       .00662      -16.901   .0000      11.3206
  FEMALE|     .97958***       .02917       33.577   .0000       .47877
 MARRIED|    -.08639***       .01031       -8.379   .0000       .75862
  HHNINC|   -1.31573***       .03112      -42.278   .0000       .35208
  HHKIDS|    -.25486***       .01064      -23.958   .0000       .40273
--------+-------------------------------------------------------------
```

Vuong Statistic for Nonnested Models (Greene)

Model 0: $\log L_{i,0} = \log f_0(y_i \mid x_i, \theta_0) = m_{i,0}$

Model 0 is the Zero Inflation Model

Model 1: $\log L_{i,1} = \log f_1(y_i \mid x_i, \theta_1) = m_{i,1}$

Model 1 is the Poisson model

(Not nested. α=0 implies the splitting probability is 1/2, not 1)

Define $a_i = m_{i,0} - m_{i,1} = \log \dfrac{f_0(y_i \mid x_i, \theta_0)}{f_1(y_i \mid x_i, \theta_1)}$

$$V = \frac{[\bar{a}]}{s_a / \sqrt{n}} = \frac{\sqrt{n}\left[\frac{1}{n}\Sigma_{i=1}^{n}\left(\log\frac{f_0(y_i \mid x_i, \theta_0)}{f_1(y_i \mid x_i, \theta_1)}\right)\right]}{\sqrt{\frac{1}{n-1}\Sigma_{i=1}^{n}\left[\log\frac{f_0(y_i \mid x_i, \theta_0)}{f_1(y_i \mid x_i, \theta_1)} - \overline{\log\frac{f_0(y_i \mid x_i, \theta_0)}{f_1(y_i \mid x_i, \theta_1)}}\right]^2}}$$

Limiting distribution is standard normal. Large + favors model 0, large - favors model 1, -1.96 < V < 1.96 is inconclusive.

```
Vuong statistic for testing ZIP vs. unaltered model is      44.6739
Distributed as standard normal. A value greater than
+1.96 favors the zero altered Z.I.Poisson model.
A value less than -1.96 rejects the ZIP model.
```

Hurdle Models

• A hurdle model is also a modified count model with two parts:
  - one generating the zeros
  - one generating the positive values.

- The models are not constrained to be the same.

• A binomial probability model governs the binary outcome of whether a count variable has a zero or a positive value.
- If y_i>0, the "hurdle is crossed," the conditional distribution of the positive values is governed by a zero-truncated count model.
=> Difference with ZI models: The amount part does not allow zeros.

• Popular models in health economics (use of health care facilities, counselling, drugs, alcohol, etc.).

A Hurdle Model

• Two part model:

- Participation part: Probability model for more than zero occurrences. For example, a logit model:

$$P(y_i = 0 \mid W_i) = \frac{\exp(W_i'\gamma)}{1+\exp(W_i'\gamma)} = \pi_i$$

- Amount part: Model for number of occurrences given that the number is greater than zero.

For example, a (zero-truncated) Poisson model:

$$P(y_i = j \mid y_i > 0, X_i) = \frac{P(y_i = j \,\&\, y_i > 0 \mid X_i)}{P(y_i > 0 \mid X_i)} = \frac{\exp(X_i'\beta)^j e^{-(X_i'\beta)}}{j![1-e^{-(X_i'\beta)}]}, \quad j = 1,2,....$$

$[1-f_P(0)]$

A Hurdle Model

• Now, we can calculate the expected value of y_i. Then,

$E[y_i|X_i] = \pi_i *0+(1 - \pi_i) * E[y_i|y_i >0, X_i] = (1 - \pi_i) * \{\lambda_i/[1-\exp(-\lambda_i)]\}$

-The last terms comes from the mean of a zero-truncated Poisson.

• Partial effects will involve both parts of the model.

Note: The estimates of the parameters and choice probabilities from a truncated Poisson model will be biased and inconsistent in the presence of overdispersion. (Correct specification of the conditional mean of the truncated dependent variable requires the correct specification of all the moments of the underlying CDF.)

=> NegBin can help. In this case, $E[y_i|X_i] = (1 - \pi_i) * \{\lambda_i/[1-f_{NB}(0)]\}$

A Hurdle Model – Application (Greene)

• Doctor Visits

```
----------------------------------------------------------------------
Poisson hurdle model for counts
Dependent variable              DOCVIS
Log likelihood function    -84211.96961
Restricted log likelihood -103727.29625
Chi squared [   1 d.f.]     39030.65329
Significance level               .00000
McFadden Pseudo R-squared      .1881407
Estimation based on N =  27326, K =  10
LOGIT  hurdle equation
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
        |Parameters of count model equation
Constant|    1.53350***       .01053      145.596   .0000
     AGE|     .01088***       .00013       85.292   .0000      43.5257
    EDUC|    -.02387***       .00072      -32.957   .0000      11.3206
  FEMALE|     .10244***       .00243       42.128   .0000       .47877
 MARRIED|    -.03463***       .00294      -11.787   .0000       .75862
  HHNINC|    -.46142***       .00873      -52.842   .0000       .35208
  HHKIDS|    -.07842***       .00301      -26.022   .0000       .40273
        |Parameters of binary hurdle equation
Constant|     .77475***       .06634       11.678   .0000
  FEMALE|     .59389***       .02597       22.865   .0000       .47877
    EDUC|    -.04562***       .00546       -8.357   .0000      11.3206
```

A Hurdle Model – Application (Greene)

• Partial Effects

```
----------------------------------------------------------------------
Partial derivatives of expected val. with
respect to the vector of characteristics.
Effects are averaged over individuals.
Observations used for means are All Obs.
Conditional Mean at Sample Point    .0109
Scale Factor for Marginal Effects  3.0118
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
        |Effects in Count Model Equation
Constant|    4.61864         2.84230        1.625   .1042
     AGE|     .03278          .02018        1.625   .1042      43.5257
    EDUC|    -.07189          .04429       -1.623   .1045      11.3206
  FEMALE|     .30854          .19000        1.624   .1044       .47877
 MARRIED|    -.10431          .06479       -1.610   .1074       .75862
  HHNINC|   -1.38971          .85557       -1.624   .1043       .35208
  HHKIDS|    -.23620          .14563       -1.622   .1048       .40273
        |Effects in Binary Hurdle Equation
Constant|     .86178***       .07379       11.678   .0000
  FEMALE|     .66060***       .02889       22.865   .0000       .47877
    EDUC|    -.05074***       .00607       -8.357   .0000      11.3206
        |Combined effect is the sum of the two parts
Constant|    5.48042*        2.85728        1.918   .0551
    EDUC|    -.12264***       .04479       -2.738   .0062      11.3206
  FEMALE|     .96915***       .19441        4.985   .0000       .47877
```

Panel Data Models

• We have repeated measures on individuals, *i*, over time, *t*: $\{(y_{it}, x_{it})$ for $i = 1, ..., N$ and $t = 1, ..., T\}$. For count data models (and DCM), y_{it} are nonnegative integer-valued outcomes.

• Typical issues for count data panels:
- Conditional on x_{it}, the y_{it}'s are likely to be serially correlated for a given *i*, partly because of state dependence and partly because of serial correlation in shocks.
=> Each additional year of data is not independent of previous years.

- Cross-sectional dependence between observations is also to be expected given emphasis on stratified clustered sampling designs.

Panel Data Models: Basic Models

• Pooled model (or population-averaged)

$$y_{it} = \alpha + x_{it}' \beta + \varepsilon_{it}$$

• Two-way effects model allows intercept to vary over *i* and *t*

$$y_{it} = \alpha_i + \gamma_t + x_{it}' \beta + \varepsilon_{it}$$

• Individual-specific effects model

$$y_{it} = \alpha_i + x_{it}' \beta + \varepsilon_{it} \qquad \alpha_i\text{: fixed effect or random effect}$$

• Mixed model or random coefficients model allows β to vary over *i*

$$y_{it} = \alpha_i + x_{it}' \beta_i + \varepsilon_{it}$$

Panel Data Models: Basic Models

• Individual-specific effects model

$$y_{it} = \alpha_i + x_{it}'\beta + \varepsilon_{it} = x_{it}'\beta + (\alpha_i + \varepsilon_{it})$$

• Fixed effects (FE):

- α_i is a random variable possibly correlated with x_{it} (endogenous), but not ε_{it}. For example, education is correlated with time-invariant ability.

=> pooled OLS, pooled GLS, RE are inconsistent for β

=> within (FE) and FD estimators are consistent.

• Random effects (RE) or population-averaged (PA):

- α_i is purely random (usually, *i.i.d.* $(0, \sigma^2)$ unrelated to x_{it}

=> appropriate FE and RE estimators are consistent for β.

Panel Data Models: Non-linear Models

• In contrast to linear models, solutions for nonlinear models tend to lack generality and are model-specific. Standard count models include: Poisson and negative binomial.

• Count models involve discreteness, nonlinearity and intrinsic heteroskedasticity. Endogeneity may be an issue.

• General approaches are similar to those for the linear case: Pooled (PA), RE and FE

• Pooled or population-averaged (PA) model: Apply as usual.

- This is the same model as in cross-section case, with adjustment for correlation over time for a given individual.

Panel Data Models: Non-linear Models

• RE and FE have some complications:
- RE often not tractable. Numerical integration needed.
- FE models complicated for short panels (small T, large N).

• A fully parametric model may be specified, with *separable heterogeneity* and conditional density

$$f(y_{it}|\alpha_i,x_{it}) = f(y_{it}|\alpha_i + x_{it}'\beta,\gamma) \qquad t=1,2,..,T; \quad i=1,2...,N$$

or *nonseparable heterogeneity*

$$f(y_{it}|\alpha_i,x_{it}) = f(y_{it}|\alpha_i + x_{it}'\beta_i,\gamma) \qquad t=1,2,..,T; \quad i=1,2...,N$$

where γ denotes additional model parameters such as variance parameters and α_i is an individual effects.

Panel Data Models: Non-linear Models

• Random Parameters: Mixed models, latent class models, hiererchical – all extended to Poisson and NB.

• Standard errors: clustered-robust, bootstrapping are OK.

Panel Data Models: Pooled (Trivedi)

• Pooled estimation:

$$y_{it} | x_{it} \sim f[\alpha_i \lambda_{it}] = f[\exp(\mathbf{x}_{it}'\boldsymbol{\beta})]$$

• We can assume a correlated error structure.

• Specify an f. For example, Poisson:

$$y_{it} | x_{it} \sim \text{Poisson}[\exp(\mathbf{x}_{it}'\boldsymbol{\beta})]$$

• Pooled Poisson of y_{it} on intercept and xit gives consistent $\boldsymbol{\beta}$.

- Use cluster-robust standard errors where cluster on the individual.
- These control for both overdispersion and correlation over t for a given i.

Panel Data Models: Pooled (Trivedi)

```
. * Pooled Poisson estimator with cluster-robust standard errors
. poisson mdu lcoins ndisease female age lfam child, vce(cluster id)

Iteration 0:   log pseudolikelihood = -62580.248
Iteration 1:   log pseudolikelihood = -62579.401
Iteration 2:   log pseudolikelihood = -62579.401

Poisson regression                                Number of obs   =      20186
                                                  Wald chi2(6)    =     476.93
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -62579.401                 Pseudo R2       =     0.0609

                                  (Std. Err. adjusted for 5908 clusters in id)
------------------------------------------------------------------------------
             |               Robust
         mdu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lcoins |  -.0808023   .0080013   -10.10   0.000    -.0964846   -.0651199
    ndisease |   .0339334   .0026024    13.04   0.000     .0288328     .039034
      female |   .1717862   .0342551     5.01   0.000     .1046473    .2389251
         age |   .0040585   .0016891     2.40   0.016      .000748    .0073691
        lfam |  -.1481981   .0323434    -4.58   0.000     -.21159    -.0848062
       child |   .1030453   .0506901     2.03   0.042     .0036944    .2023961
       _cons |    .748789   .0785738     9.53   0.000     .5947872    .9027907
------------------------------------------------------------------------------
```

By comparison, the default (non cluster-robust) s.e.'s are 1/4 as large.
=> The default (non cluster-robust) t-statistics are 4 times as large.

Panel Data Models: PA (Trivedi)

- Assume that for the i^{th} observation moments are like for GLM Poisson

$$\begin{aligned} \mathrm{E}[y_{it}|\mathbf{x}_{it}] &= \exp(\mathbf{x}_{it}'\beta) \\ \mathrm{V}[y_{it}|\mathbf{x}_{it}] &= \phi \times \exp(\mathbf{x}_{it}'\boldsymbol{\beta}). \end{aligned}$$

- Stack the conditional means for the i^{th} individual:

$$\mathrm{E}[\mathbf{y}_i|\mathbf{X}_i] = \mathbf{m}_i(\boldsymbol{\beta}) = \begin{bmatrix} \exp(\mathbf{x}_{i1}'\boldsymbol{\beta}) \\ \vdots \\ \exp(\mathbf{x}_{iT}'\boldsymbol{\beta}) \end{bmatrix}.$$

where $\mathbf{y}_i = [\mathbf{y}_{i1}, ..., \mathbf{y}_{iT}]'$ and $\mathbf{X}_i = [\mathbf{x}_{i1}, ..., \mathbf{x}_{iT}]'$.

- Stack the conditional variances for the i^{th} individual.
  - With no correlation

$$\mathrm{V}[y_i|\mathbf{X}_i] = \phi\mathbf{H}_i(\boldsymbol{\beta}) = \phi \times \mathrm{Diag}[\exp(\mathbf{x}_{it}'\boldsymbol{\beta})].$$

Panel Data Models: PA (Trivedi)

- Assume a pattern $\mathbf{R}(\rho)$ for autocorrelation over t for given i so

$$\mathrm{V}[y_i|\mathbf{X}_i] = \phi\mathbf{H}_i(\boldsymbol{\beta})^{1/2}\mathbf{R}(\rho)\mathbf{H}_i(\boldsymbol{\beta})^{1/2}$$

- This is called a working matrix.
  - Example: $\mathbf{R}(\rho) = \mathbf{I}$ if there is no correlation
  - Example: $\mathbf{R}(\rho) = \mathbf{R}(\rho)$ has diagonal entries 1 and off diagonal entries ρ if there is equicorrelation.
  - Example: $\mathbf{R}(\rho) = \mathbf{R}$ where diagonal entries 1 and off-diagonals unrestricted (< 1).

Panel Data Models: PA (Trivedi)

- The GLM estimator solves: $\sum_{i=1}^{N} \frac{\partial \mathbf{m}_i'(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \mathbf{H}_i(\boldsymbol{\beta})^{-1}(\mathbf{y}_i - \mathbf{m}_i(\boldsymbol{\theta})) = \mathbf{0}$.
- Generalized estimating equations (GEE) estimator or population-averaged estimator (PA) of Liang and Zeger (1986) solves

$$\sum_{i=1}^{N} \frac{\partial \mathbf{m}_i'(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \widehat{\Omega}_i^{-1}(\mathbf{y}_i - \mathbf{m}_i(\boldsymbol{\beta})) = \mathbf{0},$$

where $\widehat{\Omega}_i$ equals Ω_i in with $\mathrm{R}(\boldsymbol{\alpha})$ replaced by $\mathrm{R}(\widehat{\boldsymbol{\alpha}})$ where $\mathrm{plim}\,\widehat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}$.

- Cluster-robust estimate of the variance matrix of the GEE estimator is

$$\widehat{\mathrm{V}}[\widehat{\boldsymbol{\beta}}_{\mathrm{GEE}}] = \left(\widehat{\mathrm{D}}'\widehat{\Omega}^{-1}\widehat{\mathrm{D}}\right)^{-1} \left(\sum_{g=1}^{G} \mathbf{D}_g'\widehat{\Omega}_g^{-1}\widehat{\mathbf{u}}_g\widehat{\mathbf{u}}_g'\widehat{\Omega}_g^{-1}\mathbf{D}_g\right) \left(\mathrm{D}'\widehat{\Omega}^{-1}\mathrm{D}\right)^{-1}$$

where $\widehat{\mathrm{D}}_g = \partial \mathbf{m}_g'(\boldsymbol{\beta})/\partial\boldsymbol{\beta}|_{\widehat{\boldsymbol{\beta}}}$, $\widehat{\mathrm{D}} = [\widehat{\mathrm{D}}_1, ..., \widehat{\mathrm{D}}_G]'$, $\widehat{\mathbf{u}}_g = \mathbf{y}_g - \mathbf{m}_g(\widehat{\boldsymbol{\beta}})$, and now $\widehat{\Omega}_g = \mathrm{H}_g(\widehat{\boldsymbol{\beta}})^{1/2}\mathrm{R}(\widehat{\boldsymbol{\rho}})\mathrm{H}_g(\widehat{\boldsymbol{\beta}})^{1/2}$.

  - The asymptotic theory requires that $G \to \infty$.

Panel Data Models: PA (Trivedi)

```
GEE population-averaged model                   Number of obs      =     20186
Group and time vars:                 id year    Number of groups   =      5908
Link:                                    log    Obs per group: min =         1
Family:                              Poisson                   avg =       3.4
Correlation:                    unstructured                   max =         5
                                                Wald chi2(6)       =    508.61
Scale parameter:                           1    Prob > chi2        =    0.0000

                                    (Std. Err. adjusted for clustering on id)
```

mdu	Coef.	Semi-robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
lcoins	-.0804454	.0077782	-10.34	0.000	-.0956904	-.0652004
ndisease	.0346067	.0024238	14.28	0.000	.0298561	.0393573
female	.1585075	.0334407	4.74	0.000	.0929649	.2240502
age	.0030901	.0015356	2.01	0.044	.0000803	.0060999
lfam	-.1406549	.0293672	-4.79	0.000	-.1982135	-.0830962
child	.1013677	.04301	2.36	0.018	.0170696	.1856658
_cons	.7764626	.0717221	10.83	0.000	.6358897	.9170354

• In general, SE's are within 10% of pooled Poisson cluster-robust SE's.
• The default (non cluster-robust) t-statistics are 3.5 to 4 times larger.
• No control for overdispersion.

Panel Data Models: PA (Trivedi)

• The correlations $Cor[y_{it},y_{is}|x_i]$ for PA (unstructured) are not equal. But they are not declining as fast as AR(1).

```
. matrix list e(R)

symmetric e(R)[5,5]
           c1          c2          c3          c4          c5
r1          1
r2  .53143297           1
r3  .40817495   .58547795           1
r4  .32357326   .35321716   .54321752           1
r5  .34152288   .29803555   .43767583   .61948751           1
```

Panel Data Models: FE

• Fixed Effects:

$$y_{it} \,|\, x_{it} \sim f[\alpha_i\, \lambda_{it}] = f[\alpha_i \exp(\mathbf{x}_{it}'\boldsymbol{\beta})]$$

- In general, estimation is not possible in short panels.
- Incidental parameters problem:
  - N fixed effects α_i plus K regressors means $(N + K)$ parameters
  - But $(N+K) \rightarrow\infty$ as $N \rightarrow\infty$
  - Need to eliminate α_i by some sort of differencing, or concentrated likelihood argument.

• Fixed effects extensions to hurdle, finite mixture, zero-inflated models are currently not available.

Panel Data Models: FE Poisson (Trivedi)

• Derivation of fixed effects estimator for the Poisson panel

- Poisson MLE simultaneously estimates β and $\alpha_1, \ldots, \alpha_{iN}$. The log-likelihood is

$$\begin{aligned}\ln L(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \ln\left[\prod_i \prod_t \{\exp(-\alpha_i \lambda_{it})\,(\alpha_i \lambda_{it})^{y_{it}} / y_{it}!\}\right] \\ &= \sum_i \left[-\alpha_i \sum_t \lambda_{it} + \ln \alpha_i \sum_t y_{it} + \sum_t y_{it} \ln \lambda_{it} - \sum_t \ln y_{it}\right.\end{aligned}$$

where $\lambda_{it} = \exp(\mathbf{x}_{it}'\beta)$.

- f.o.c.'s w.r.t. α_i yields $\hat{\alpha}_i = \Sigma_t y_{it} / \Sigma_t \lambda_{it}$ (a sufficient statistic for α_i).

- Substituting $\hat{\alpha}_i$ into lnL yields the concentrated likelihood function.

- Dropping terms not involving β:

$$\ln L_{\text{conc}}(\boldsymbol{\beta}) \propto \sum_i \sum_t \left[y_{it} \ln \lambda_{it} - y_{it} \ln\left(\sum_s \lambda_{is}\right)\right]$$

Panel Data Models: FE Poisson (Trivedi)

- There is no incidental parameters problem

- Consistent estimates of β for fixed T and $N \to \infty$ can be obtained by maximization of $\ln L_{conc}(\beta)$

- f.o.c. with respect to β yields first-order conditions:

$$\sum_i \sum_t \left[y_{it}\mathbf{x}_{it} - y_{it}\left[\sum_s \lambda_{is}\mathbf{x}_{is}\right] / \left[\sum_s \lambda_{is}\right]\right] = \mathbf{0}$$

that can be re-expressed as

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{x}_{it}\left(y_{it} - \frac{\lambda_{it}}{\bar{\lambda}_i}\bar{y}_i\right) = \mathbf{0}$$

<u>Note</u>: $\lambda_{it}/(\Sigma_t \lambda_{it})$ =. Time-invariant X_i's dissappear!

Panel Data Models: FE Poisson – Pros & Cons

• Time-invariant regressors will be eliminated also by the differencing transformation. Some marginal effects not identified.

• May substitute individual specific dummy variables, though this raises some computational issues.

• Poisson and linear panel model special in that simultaneous estimation of β and α provides consistent estimates of β in short panels, so there is no incidental parameters problem.

• The above assumes strict exogeneity of regressors.

• We can handle endogenous regressors under weak exogeneity assumption. A moment condition estimator can be defined using the previous f.o.c.'s.

• This FE approach does not extend to several empirically important models: hurdle, finite mixture models, and zip.

PDM: FE-Poisson with panel bootsrapped SE's (Trivedi)

```
. xtpoisson mdu lcoins ndisease female age lfam child, fe vce(boot, reps(100) seed(10
(running xtpoisson on estimation sample)

Bootstrap replications (100)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100

Conditional fixed-effects Poisson regression    Number of obs      =     17791
Group variable: id                              Number of groups   =      4977

                                                Obs per group: min =         2
                                                               avg =       3.6
                                                               max =         5

                                                Wald chi2(3)       =      4.64
Log likelihood  = -24173.211                    Prob > chi2        =    0.2002

                                   (Replications based on 4977 clusters in id)
```

mdu	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf. Interval]	
age	-.0112009	.0095077	-1.18	0.239	-.0298356	.0074339
lfam	.0877134	.1125783	0.78	0.436	-.132936	.3083627
child	.1059867	.0738452	1.44	0.151	-.0387472	.2507206

• The default (non cluster-robust) t-statistics are 2 times larger.

Panel Data Models: RE (Trivedi)

• Random Effects:

$y_{it} | x_{it} \sim f[\alpha_i \exp(\mathbf{x}_{it}'\beta)] = f[\alpha_i \exp(\ln \alpha_i + \mathbf{x}_{it}'\beta)]$

α_i is unobserved but is not correlated with $\mathbf{x}_{it}$.

- Poisson: Two treatments:
  - (1) α_i is gamma distributed.
    - It becomes a NegBin model (analytical solution!).
    - $E[y_{it} | x_{it}, \beta] = \lambda_{it} = \exp(\mathbf{X}_{it}'\beta)$.

  - (2) Contemporary treatments are assuming $\ln \alpha_i \sim N(0,\sigma^2)$

    => analytical (closed form) solution does not exist (one-dimensional integral, done with simulation or quadrature based estimators).

Panel Data Models: RE (Trivedi)

- Contemporary treatments are assuming $\ln \alpha_i \sim N(0,\sigma^2)$

  => analytical (closed form) solution does not exist (one-dimensional integral, done with simulation or quadrature based estimators.
  - It can extend to slope coefficients (higher-dimensional integral)
  - $E[y_{it} | x_{it}, \beta] = \lambda_{it} = \exp(\mathbf{X}_{it}'\beta)$.

- NB with random effects is equivalent to two "effects" one time varying, one time invariant. The model is probably overspecified.

• Note: It is common to find similar results for RE models (1) and (2).

PDM: RE-gamma with panel bootsrapped SE's (Trivedi)

```
Random-effects Poisson regression               Number of obs      =     20186
Group variable: id                              Number of groups   =      5908

Random effects u_i ~ Gamma                      Obs per group: min =         1
                                                               avg =       3.4
                                                               max =         5

                                                Wald chi2(6)       =    529.10
Log likelihood  = -43240.556                    Prob > chi2        =    0.0000

                                   (Replications based on 5908 clusters in id)
```

mdu	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf. Interval]	
lcoins	-.0878258	.0086097	-10.20	0.000	-.1047004	-.0709511
ndisease	.0387629	.0026904	14.41	0.000	.0334899	.0440359
female	.1667192	.0379216	4.40	0.000	.0923942	.2410442
age	.0019159	.0016242	1.18	0.238	-.0012675	.0050994
lfam	-.1351786	.0308529	-4.38	0.000	-.1956492	-.0747079
child	.1082678	.0495487	2.19	0.029	.0111541	.2053816
_cons	.7574177	.0754536	10.04	0.000	.6095314	.905304
/lnalpha	.0251256	.0270297			-.0278516	.0781029
alpha	1.025444	.0277175			.9725326	1.081234

```
Likelihood-ratio test of alpha=0: chibar2(01) =  3.9e+04 Prob>=chibar2 = 0.000
```

Panel Poisson: Estimator comparison (Trivedi)

- Compare following estimators
  - Pooled Poisson with cluster-robust SE.'s
  - Pooled population averaged Poisson with unstructured correlations and cluster-robust SE's
  - RE Poisson with gamma random effect and cluster-robust SE'.s.
  - RE Poisson with normal random effect and default SE.'s
  - FE Poisson and cluster-robust SE's

- Find that
  - Similar results for all RE models
  - Note that these data are not good to illustrate FE as regressors have little within variation.

Panel Poisson: Estimator comparison (Trivedi)

Variable	POOLED	POPAVE	RE_GAMMA	RE_NOR~L	FIXED
#1					
lcoins	-0.0808	-0.0804	-0.0878	-0.1145	
	0.0080	0.0078	0.0086	0.0073	
ndisease	0.0339	0.0346	0.0388	0.0409	
	0.0026	0.0024	0.0027	0.0023	
female	0.1718	0.1585	0.1667	0.2084	
	0.0343	0.0334	0.0379	0.0305	
age	0.0041	0.0031	0.0019	0.0027	-0.0112
	0.0017	0.0015	0.0016	0.0012	0.0095
lfam	-0.1482	-0.1407	-0.1352	-0.1443	0.0877
	0.0323	0.0294	0.0309	0.0265	0.1126
child	0.1030	0.1014	0.1083	0.0737	0.1060
	0.0507	0.0430	0.0495	0.0345	0.0738
_cons	0.7488	0.7765	0.7574	0.2873	
	0.0786	0.0717	0.0755	0.0642	
lnalpha					
_cons			0.0251		
			0.0270		
lnsig2u					
_cons				0.0550	
				0.0255	

Panel Poisson: FE vs RE (Trivedi)

- Strength of fixed effects versus random effects
  - Allows α_i to be correlated with $\mathbf{x}_{it}$.
  - So consistent estimates if regressors are correlated with the error provided regressors are correlated only with the time-invariant component of the error
  - An alternative to IV to get causal estimates.
- Limitations:
  - Coefficients of time-invariant regressors are not identified
  - For identified regressors standard errors can be much larger
  - Marginal effect in a nonlinear model depend on α_i

$$\mathrm{ME}_j = \partial \mathrm{E}[y_{it}]/\partial \mathrm{x}_{it,j} = \alpha_i \exp(\mathbf{x}_{it}'\boldsymbol{\beta})\beta_j$$

    and α_i is unknown.

A Peculiarity of the FE-NB Model (Greene)

- 'True' FE model has $\lambda_i=\exp(c_i+x_{it}'\beta)$. Cannot be fit if there are time invariant variables.
- Hausman, Hall and Griliches (Econometrica, 1984) has c_i appearing in θ (variance).
  - Produces different results
  - Implies that the FEM can contain time invariant variables.

Panel Data Models - Application (Greene)

```
+----------------------------------------------+
| Panel Model with Group Effects               |
| Log likelihood function        -33576.74     | Hausman et al. version.
| Unbalanced panel has     7293 individuals.   | FENB turns into a logit
| Neg.Binomial Regression -- Fixed Effects     | model.
+----------------------------------------------+
+---------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient   | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+---------+--------------+----------------+--------+--------+----------+
 HHNINC   |     .23681421       .05317660     4.453   .0000    .35208362
 EDUC     |     .08097026       .00267695    30.247   .0000   11.3206310
 HSAT     |    -.13764986       .00336492   -40.907   .0000   6.78542607
+----------------------------------------------+
| FIXED EFFECTS NegBin Model                   |
| Log likelihood function        -51020.09     | 'True' FE model. Estimated
| Bypassed 1153 groups with inestimable a(i).  | by 'brute force.'
| Negative binomial regression model           |
+----------------------------------------------+
+---------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient   | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+---------+--------------+----------------+--------+--------+----------+
---------+Index function for probability
 HHNINC   |     .14058502       .04799217     2.929   .0034    .35040228
 EDUC     |    -.01688381       .02135354     -.791   .4291   11.2596731
 HSAT     |    -.15775644       .00304539   -51.802   .0000   6.66405976
---------+Overdispersion parameter
 Alpha    |    7.58363763       .01432940   529.236   .0000
```

PDM - Moment based Estimation (Trivedi)

- Predetrmined means regressor correlated with current and past shoocks but not future shocks: $E[u_{it}x_{is}] = 0$ for $s \geq t$, but $\neq 0$ for $S < t$.
- Two specifications are considered:

$$
\begin{aligned}
y_{it} &= \exp(\mathbf{x}'_{it}\beta)\nu_i w_{it} \\
y_{it} &= \exp(\mathbf{x}'_{it}\beta)\nu_i + w_{it}
\end{aligned}
$$

- A quasi-differencing transformation is used to eliminate the fixed effect.
- Then a moment condition is constructed for estimation.
- Depending upon which specification is used different moment conditions obtain.
- Chamberlain and Wooldridge derive quasi-differencing transformations that are shown in the table below.

PDM - Moment based Estimation (Trivedi)

- Relies on a number of ways of eliminating the fixed effects
- Error may enter additively or multiplicatively
- Estimating equations are orthogonality conditions after quasi-differencing which eliminates the fixed effect

Model	**Moment spec.**	**Estimating equations**
Strict exog.	$\mathrm{E}[\mathbf{x}_{it}u_{it+j}] = 0$, $j \geq 0$	
Predetermined regressors	$\mathrm{E}[\mathbf{x}_{it}u_{it-s}] \neq 0$, $s \geq 1$	
GMM	Chamberlain	$\mathrm{E}\left[y_{it}\frac{\lambda_{it-1}}{\lambda_{it}} - y_{it-1} \mid \mathbf{x}_i^{t-1})\right] = 0$
	Wooldridge	$\mathrm{E}\left[\frac{y_{it}}{\lambda_{it}} - \frac{y_{it-1}}{\lambda_{it-1}} \mid \mathbf{x}_i^{t-1})\right] = 0$
GMM/endog	Wooldridge	$\mathrm{E}\left[\frac{y_{it}}{\lambda_{it}} - \frac{y_{it-1}}{\lambda_{it-1}} \mid \mathbf{x}_i^{t-2})\right] = 0$

PDM - Moment based Estimation (Trivedi)

• <u>Example</u>: Fixed Effects GMM in Stata 11

```
. program gmm_poi2
      version 11
      syntax varlist if, at(name) myrhs(varlist) ///
>    mylhs(varlist) myidvar(varlist)
      quietly {
      tempvar mu mubar ybar
      gen double `mu' = 0 `if'
      local j = 1
      foreach var of varlist `myrhs' {
         replace `mu' = `mu' + `var'*`at'[1,`j'] `if'
         local j = `j' + 1
         }
      replace `mu' = exp(`mu')
      egen double `mubar' = mean(`mu') `if', by(`myidvar')
      egen double `ybar' = mean(`mylhs') `if', by(`myidvar')
      replace `varlist' = `mylhs' - `mu'*`ybar'/`mubar' `if'
      }
  end
```

PDM - Moment based Estimation (Trivedi)

• Implementing FE GMM in Stata 11

```
. gmm gmm_poi2, mylhs(officevis) myrhs(insprv age income totchr)   ///
>  myidvar(dupersid) nequations(1) parameters(insprv age income totchr)   ///
>  instruments(insprv age income totchr, noconstant) onestep

Step 1
Iteration 0:   GMM criterion Q(b) =  .00140916
Iteration 1:   GMM criterion Q(b) =  1.487e-07
Iteration 2:   GMM criterion Q(b) =  1.583e-14
Iteration 3:   GMM criterion Q(b) =  1.843e-28

GMM estimation

Number of parameters =   4
Number of moments    =   4
Initial weight matrix: Unadjusted                     Number of obs  =    78888
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf. Interval]	
/insprv	-.0080549	.5460749	-0.01	0.988	-1.078342	1.062232
/age	-.5125841	13.1682	-0.04	0.969	-26.32178	25.29662
/income	.001128	.0013911	0.81	0.417	-.0015984	.0038545
/totchr	.2211125	.3354182	0.66	0.510	-.4362951	.8785201

```
Instruments for equation 1: insprv age income totchr

. estimates store PFEGMM
```

PDM - Moment based Estimation (Trivedi)

• Standard FE with robust SE (with xtpqml add-on) in Stata 11

```
. * Add-on xtpqml gives panel robust se's
. xtpqml officevis insprv age income totchr, fe i(dupersid)
note: 1900 groups (15200 obs) dropped because of all zero outcomes

Iteration 0:   log likelihood = -84468.435
Iteration 1:   log likelihood =  -84154.68
Iteration 2:   log likelihood = -84154.647
Iteration 3:   log likelihood = -84154.647

Conditional fixed-effects Poisson regression    Number of obs      =     63688
Group variable: dupersid                        Number of groups   =      7961

                                                Obs per group: min =         8
                                                               avg =       8.0
                                                               max =         8

                                                Wald chi2(4)       =    618.20
Log likelihood  = -84154.647                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
   officevis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      insprv |  -.0080549    .027985    -0.29   0.773    -.0629046    .0467947
         age |  -.5125841   .0629145    -8.15   0.000    -.6358943   -.3892739
      income |    .001128    .000258     4.37   0.000     .0006224    .0016336
      totchr |   .2211125   .0091051    24.28   0.000     .2032669    .2389582
------------------------------------------------------------------------------
Calculating Robust Standard Errors...
------------------------------------------------------------------------------
   officevis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
officevis    |
      insprv |  -.0080549   .0715881    -0.11   0.910    -.1483651    .1322552
         age |  -.5125841   .1804831    -2.84   0.005    -.8663245   -.1588438
      income |    .001128   .0007661     1.47   0.141    -.0003734    .0026295
      totchr |   .2211125   .0250814     8.82   0.000     .1719539    .2702712
------------------------------------------------------------------------------
```

PDM – Dynamics (Trivedi)

• Individual effects model allows for time series persistence via unobserved heterogeneity, α_i. For example, high α_i means high IPOs each period.

• Alternative time series persistence is via true state dependence, y_{t-1}. For example, a lot of IPOs last period lead to a lot of IPOs this period.

• Linear model:

$$y_{it} = [\alpha_i + \rho y_{it-1} + \alpha_i + \mathbf{x}_{it}'\beta + \varepsilon_{it}$$

• Poisson model with exponetial feedback: One possibility (designed to confront the zero problem) is

$$\mu_{it} = \alpha_i \boldsymbol{\lambda}_{it-1} = \alpha_i \exp(\rho y^*_{it-1} + \mathbf{x}_{it}'\beta), \qquad y^*_{it-1} = \min(c, y_{it-1}).$$

PDM – Dynamics (Trivedi)

- In fixed effects case, the Poisson FE estimator is now inconsistent. Instead assume weak exogeneity

$$E[y_{it} | y_{it-1}, y_{it-2}, \dots, \mathbf{x}_{it}, \mathbf{x}_{it-1}, \dots] = \alpha_i \boldsymbol{\lambda}_{it-1}$$

- Use an alternative quasi-difference

$$E[y_{i,t} - (\boldsymbol{\lambda}_{i,t}/\boldsymbol{\lambda}_{i,t-1}) y_{i,t-1} | y_{it-1}, y_{it-2}, \dots, \mathbf{x}_{it}, \mathbf{x}_{it-1}, \dots] = 0$$

- Then, MM or GMM based on:

$$E[\mathbf{z}_{i,t} \{y_{i,t} - (\boldsymbol{\lambda}_{it}/\boldsymbol{\lambda}_{i,,t-1}) y_{i,t-1}\}] = 0$$

where $\mathbf{z}_{i,t}$ is a vector of instruments. For example, in the just-identified case: $(y_{i,t-1}, \mathbf{x}_{it})$.

- Windmeijer (2008) has a discussion of this topic.

PDM – Dynamics – GMM Example (Trivedi)

- Just Identified (JI) GMM: Ignoring individual specific effects

```
. gmm (officevis - exp({xb:L.officevis insprv educ age income totchr}+{b0})),   ///
>  instruments(L.officevis insprv educ age income totchr) onestep vce(cluster dupersid)

Step 1
Iteration 0:   GMM criterion Q(b) =  4.9539327
Iteration 1:   GMM criterion Q(b) =  4.7296297
Iteration 2:   GMM criterion Q(b) =  1.4832673
Iteration 3:   GMM criterion Q(b) =  .01045573
Iteration 4:   GMM criterion Q(b) =  6.508e-06
Iteration 5:   GMM criterion Q(b) =  3.032e-12
Iteration 6:   GMM criterion Q(b) =  7.264e-25

GMM estimation

Number of parameters =   7
Number of moments    =   7
Initial weight matrix: Unadjusted                     Number of obs  =    69027

                           (Std. Err. adjusted for 9861 clusters in dupersid)
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/xb_L_offi~s	.064072	.0041069	15.60	0.000	.0560228	.0721213
/xb_insprv	.2152153	.0331676	6.49	0.000	.1502079	.2802227
/xb_educ	.0404143	.0065808	6.14	0.000	.0275162	.0533124
/xb_age	.1221278	.0134542	9.08	0.000	.0957581	.1484976
/xb_income	-.0003585	.0004981	-0.72	0.472	-.0013347	.0006178
/xb_totchr	.3027348	.0141805	21.35	0.000	.2749415	.330528
/b0	-1.447292	.0952543	-15.19	0.000	-1.633987	-1.260597

PDM – Dynamics – GMM Example (Trivedi)

• Over Identified (OI) GMM

```
. gmm (officevis - exp({xb:L.officevis insprv educ age income totchr}+{b0})),   ///
>  instruments(L.officevis educ age income totchr female white hispanic married employed) /
>  onestep vce(cluster dupersid)

Step 1
Iteration 0:   GMM criterion Q(b) =  4.9696148
Iteration 1:   GMM criterion Q(b) =  3.7545442
Iteration 2:   GMM criterion Q(b) =  .86353039
Iteration 3:   GMM criterion Q(b) =  .25844389
Iteration 4:   GMM criterion Q(b) =  .07248002
Iteration 5:   GMM criterion Q(b) =  .07235453
Iteration 6:   GMM criterion Q(b) =  .07235443

GMM estimation

Number of parameters =   7
Number of moments    =  11
Initial weight matrix: Unadjusted                    Number of obs  =    69027

                             (Std. Err. adjusted for 9861 clusters in dupersid)
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/xb_L_offi~s	.0631186	.0042901	14.71	0.000	.0547101	.071527
/xb_insprv	.0468067	.1154105	0.41	0.685	-.1793937	.273007
/xb_educ	.0422612	.0074362	5.68	0.000	.0276866	.0568359
/xb_age	.1208516	.0136986	8.82	0.000	.0940028	.1477003
/xb_income	.0004412	.0007107	0.62	0.535	-.0009518	.0018341
/xb_totchr	.2988192	.0144326	20.70	0.000	.2705318	.3271066
/b0	-1.361726	.0972536	-14.00	0.000	-1.55234	-1.171113

PDM - Dynamics – Poisson Extension (Trivedi)

- A different ML approach to dynamic specification

$$
\begin{aligned}
y_{i,t} &\sim P(\lambda_{it}),\ i = 1, ..., N; t = 1, ..., T \\
f(y_{i,t}|\lambda_{it}) &= \frac{e^{-\lambda_{it}}\lambda_{it}^{y_{it}}}{y_{it}!} \\
\lambda_{it} &= \nu_{it}\mu_{it} = \mathrm{E}[y_{it}|y_{i,t-1}, \mathbf{x}_{it}, \alpha_i] = g(y_{i,t-1}, \mathbf{x}_{it}, \alpha_i)
\end{aligned}
$$

- Initial conditions problem in dynamic model. In a short panel bias induced by neglect of dependence on initial condition.
- The lagged dependent variable on the right hand side a source of bias because the lagged dependent variable and individual-specific effect are correlated.
- Wooldridge's method (2005) integrates out the individual-specific random effect after conditioning on the initial value and covariates. Random effect model used to accommodate the initial conditions.

PDM - Dynamics – Poisson Extension (Trivedi)

$$E[y_{it}|\mathbf{x}_{it}, y_{it-1}, \alpha_i] = h(y_{it}, \mathbf{x}_{it}, \boldsymbol{\alpha}_i)$$

where α_i is the individual-specific effect.

- 1st alternative: Autoregressive dependence through the exponential mean.

$$E[y_{it}|\mathbf{x}_{it}, y_{it-1}, \alpha_i] = \exp(\rho y_{it-1} + \mathbf{x}'_{it}\boldsymbol{\beta} + \boldsymbol{\alpha}_i)$$

- If the α_i are uncorrelated with the regressors, and further if parametric assumptions are to be avoided, then this model can be estimated using either the nonlinear least squares or pooled Poisson MLE. In either case it is desirable to use the robust variance formula.
- Limitation: Potentially explosive if large values of y_{it} are realized.

PDM - Dynamics – Initial Conditions (Trivedi)

- Dynamic panel model requires additional assumptions about the relationship between the initial observations ("initial conditions") $\mathbf{y}_0$ and the α_i.
- Effect of initial value on the future events is important in a short panel. The initial-value effect might be a part of individual-specific effect
- Wooldridge's method requires a specification of the conditional distribution of α_i given $\mathbf{y}_0$ and $\mathbf{z}_i$, with the latter entering separably.
- Under the assumption that the initial conditions are nonrandom, the standard random effects conditional maximum likelihood approach identifies the parameters of interest.
- For a class of nonlinear dynamic panel models, including the Poisson model, Wooldridge (2005) analyzes this model which **conditions** the joint distribution on the initial conditions.

PDM – Conditionally correlated RE (Trivedi)

- **Where parametric FE models are not feasible, the conditionally correlated random (CCR) effects model** (Mundlak (1978) and Chamberlain (1984)) **provides a compromise between FE and RE models.**
- Standard RE panel model assumes that α_i and $\mathbf{x}_{it}$ are uncorrelated. Making α_i a function of $\mathbf{x}_{i1}, ..., \mathbf{x}_{iT}$ allows for possible correlation:

$$\alpha_i = \mathbf{z}_i'\boldsymbol{\lambda} + \varepsilon_i$$

- Mundlak's (more parsimonious) method allows the individual-specific effect to be determined by time averages of covariates, denoted $\mathbf{z}_i$; Chamberlain's method suggests a richer model with a weighted sum of the covariates for the random effect.

PDM – Conditionally correlated RE (Trivedi)

- We can further allow for initial condition effect by including $\mathbf{y}_0$ thus:

$$\alpha_i = \mathbf{y}_0'\boldsymbol{\eta} + \mathbf{z}_i'\boldsymbol{\lambda} + \varepsilon_i$$

  where $\mathbf{y}_0$ is a vector of initial conditions, $\mathbf{z}_i = \overline{\mathbf{x}}_i$ denotes the time-average of the exogenous variables and ε_i may be interpreted as unobserved heterogeneity.
- The formulation essentially introduces no additional problems though the averages change when new data are added. Estimation and inference in the pooled Poisson or NLS model can proceed as before.
- Formulation can also be used when no dynamics are present in the model. In this case ε_i can be integrated out using a distributional assumption about $f(\varepsilon)$.

Dynamic GMM without initial condition (Trivedi)

• Here individual specific effect is captured by the initial condition

```
. gmm (officevis - exp({xb:L.officevis insprv educ age income totchr}+{b0})),   ///
>  instruments(L.officevis insprv educ age income totchr) onestep vce(cluster dupersid)

Step 1
Iteration 0:   GMM criterion Q(b) =  4.9539327
Iteration 1:   GMM criterion Q(b) =  4.7296297
Iteration 2:   GMM criterion Q(b) =  1.4832673
Iteration 3:   GMM criterion Q(b) =  .01045573
Iteration 4:   GMM criterion Q(b) =  6.508e-06
Iteration 5:   GMM criterion Q(b) =  3.032e-12
Iteration 6:   GMM criterion Q(b) =  7.264e-25

GMM estimation

Number of parameters =   7
Number of moments    =   7
Initial weight matrix: Unadjusted                     Number of obs   =    69027

                               (Std. Err. adjusted for 9861 clusters in dupersid)
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/xb_L_offi~s	.064072	.0041069	15.60	0.000	.0560228	.0721213
/xb_insprv	.2152153	.0331676	6.49	0.000	.1502079	.2802227
/xb_educ	.0404143	.0065808	6.14	0.000	.0275162	.0533124
/xb_age	.1221278	.0134542	9.08	0.000	.0957581	.1484976
/xb_income	-.0003585	.0004981	-0.72	0.472	-.0013347	.0006178
/xb_totchr	.3027348	.0141805	21.35	0.000	.2749415	.330528

Overidentified dynamic GMM with initial condition

```
. gmm (officevis - exp({xb:L.officevis insprv educ age income totchr}+{b0})),   ///
>  instruments(L.officevis educ age income totchr female white hispanic married empl
>  onestep vce(cluster dupersid)

Step 1
Iteration 0:   GMM criterion Q(b) =  4.9696148
Iteration 1:   GMM criterion Q(b) =  3.7545442
Iteration 2:   GMM criterion Q(b) =  .86353039
Iteration 3:   GMM criterion Q(b) =  .25844389
Iteration 4:   GMM criterion Q(b) =  .07248002
Iteration 5:   GMM criterion Q(b) =  .07235453
Iteration 6:   GMM criterion Q(b) =  .07235443

GMM estimation

Number of parameters =   7
Number of moments    =  11
Initial weight matrix: Unadjusted                     Number of obs   =    69027

                               (Std. Err. adjusted for 9861 clusters in dupersid)
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/xb_L_offi~s	.0631186	.0042901	14.71	0.000	.0547101	.071527
/xb_insprv	.0468067	.1154105	0.41	0.685	-.1793937	.273007
/xb_educ	.0422612	.0074362	5.68	0.000	.0276866	.0568359
/xb_age	.1208516	.0136986	8.82	0.000	.0940028	.1477003
/xb_income	.0004412	.0007107	0.62	0.535	-.0009518	.0018341

Dynamic JI GMM with Initial Conditions

```
. gmm (officevis - exp({xb:L.officevis insprv educ age income totchr}+{b0})),   ///
>  instruments(L.officevis educ age income totchr female white hispanic married empl
>  onestep vce(cluster dupersid)

Step 1
Iteration 0:   GMM criterion Q(b) =  4.9696148
Iteration 1:   GMM criterion Q(b) =  3.7545442
Iteration 2:   GMM criterion Q(b) =  .86353039
Iteration 3:   GMM criterion Q(b) =  .25844389
Iteration 4:   GMM criterion Q(b) =  .07248002
Iteration 5:   GMM criterion Q(b) =  .07235453
Iteration 6:   GMM criterion Q(b) =  .07235443

GMM estimation

Number of parameters =   7
Number of moments    =  11
Initial weight matrix: Unadjusted                    Number of obs   =    69027

                          (Std. Err. adjusted for 9861 clusters in dupersid)
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/xb_L_offi~s	.0631186	.0042901	14.71	0.000	.0547101	.071527
/xb_insprv	.0468067	.1154105	0.41	0.685	-.1793937	.273007
/xb_educ	.0422612	.0074362	5.68	0.000	.0276866	.0568359
/xb_age	.1208516	.0136986	8.82	0.000	.0940028	.1477003
/xb_income	.0004412	.0007107	0.62	0.535	-.0009518	.0018341

Dynamic OI GMM with Initial Conditions

```
. gmm (officevis - exp({xb:L.officevis T0officevis insprv educ age income totchr}+{
>  instruments(L.officevis T0officevis educ age income totchr female white hispanic
>  onestep vce(cluster dupersid) nolog

Final GMM criterion Q(b) =  .0685762

GMM estimation

Number of parameters =   8
Number of moments    =  12
Initial weight matrix: Unadjusted                    Number of obs   =    69027

                          (Std. Err. adjusted for 9861 clusters in dupersid)
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/xb_L_offi~s	.0490201	.0046062	10.64	0.000	.039992	.0580481
/xb_T0offi~s	.0305356	.0044538	6.86	0.000	.0218063	.0392648
/xb_insprv	.0565968	.1135886	0.50	0.618	-.1660328	.2792264
/xb_educ	.0402952	.0059253	6.80	0.000	.0286819	.0519085
/xb_age	.1299791	.0098075	13.25	0.000	.1107567	.1492014
/xb_income	.0004368	.000703	0.62	0.534	-.0009411	.0018148
/xb_totchr	.2805608	.0101571	27.62	0.000	.2606532	.3004684
/b0	-1.408679	.0607941	-23.17	0.000	-1.527833	-1.289525

```
Instruments for equation 1: L.officevis T0officevis educ age income totchr female w
    married employed _cons
```

PDM: Remarks (Trivedi)

• Much progress in estimating panel count models, especially in dealing with endogeneity and nonseprable heterogeneity.

• Great progress in variance estimation.

• RE models pose fewer problems.

• For FE models moment-based/IV methods seem more tractable for handling endogeneity and dynamics. Stata's new suite of GMM commands are very helpful in this regard.

• Because FE models do not currently handle important cases, and have other limitations, CCR panel model with initial conditions, is an attractive alternative, at least for balanced panels.