Outline of all Lectures

I. Basic cross-section methods:
   - Poisson, GLM, negative binomial
II. More advanced cross-section methods:
   - Hurdle, zero-inflated, finite mixtures, endogeneity
III. Time series and panel methods
IV. Further Topics:
   - multivariate, maximum simulated likelihood, Bayesian

Outline of additional cross-section count methods

- Introduction
- Censored and truncated data
- Richer parametric models
  - hurdle model
  - zero-inflated model
  - continuous mixtures
  - hierarchical models
  - model comparison
- Finite mixtures model
- Endogenous regressors
- Quantile regression

Counts left-truncated at zero

- Sampling rule is such that observe only $y$ and $x$ for $y \geq 1$
  i.e. only those who participate at least once are in sample.
- Truncated density (given untruncated density $f(y|x, \theta)$) is
  \[
  f(y|x, \theta, y \geq 0) = \frac{f(y|x, \theta)}{\Pr[y \geq 0|x, \theta]} = \frac{f(y|x, \theta)}{[1 - f(0|x, \theta)]}.
  \]
- MLE is inconsistent if any aspect of the parametric model is misspecified.
- Need to assume that the process for nonzeros is the same as zeroes.
  - e.g. If data are on annual number of hunting trips for only those who hunted this year, then a missing 0 is interpreted as being for a hunter who did not hunt this year (rather than for all people).
Counts right-censored

- Sampling rule is that observe only 0, 1, 2, ..., \( c - 1 \), \( c \) or more
  i.e. Only record counts up to \( c \) and then any value above \( c \).
- Censored density (given uncensored density \( f(y|x, \theta) \) and cdf is \( F(y|x, \theta) \))
  \[ \begin{align*}
  f(y|x, \theta) \\
  1 - F(c - 1|x, \theta) = 1 - \sum_{i=0}^{c-1} f(j|x, \theta)
  \end{align*} \]
- Log-likelihood (where \( d_i = 1 \) if uncensored and \( d_i = 0 \) if censored)
  \[ L(\theta) = \sum_{i=1}^{N} \{ d_i \ln f(y_i|x_i, \theta) + (1 - d_i) \ln (1 - \sum_{j=0}^{c-1} f(j|x_i, \theta)) \} \]
- MLE is inconsistent if any aspect of the parametric model is misspecified
  > So pick a good density - at least negative binomial.

Left-truncated at 0 (11% truncated) & right-censored at 10 (26% censored) are less efficient than NB on the complete data.

### Table

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>ZNE</th>
<th>ZTE</th>
<th>NECESS10</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1_private</td>
<td>0.1641</td>
<td>0.1096</td>
<td>0.1238</td>
</tr>
<tr>
<td>medicaid</td>
<td>0.0312</td>
<td>0.0345</td>
<td>0.0409</td>
</tr>
<tr>
<td>age</td>
<td>0.0503</td>
<td>0.0572</td>
<td>0.0613</td>
</tr>
<tr>
<td>age2</td>
<td>0.0454</td>
<td>0.0470</td>
<td>0.0539</td>
</tr>
<tr>
<td>educyr</td>
<td>0.0621</td>
<td>0.0625</td>
<td>0.0735</td>
</tr>
<tr>
<td>secthm</td>
<td>0.0604</td>
<td>0.0604</td>
<td>0.0605</td>
</tr>
<tr>
<td>secthr</td>
<td>0.0642</td>
<td>0.0644</td>
<td>0.0652</td>
</tr>
<tr>
<td>ctechr</td>
<td>0.0319</td>
<td>0.0315</td>
<td>0.0377</td>
</tr>
<tr>
<td>_cons</td>
<td>-0.0030</td>
<td>-0.0038</td>
<td>-0.0035</td>
</tr>
<tr>
<td>Tnatives</td>
<td>0.0049</td>
<td>0.0050</td>
<td>0.0051</td>
</tr>
<tr>
<td>ctechr</td>
<td>2.2474</td>
<td>2.3376</td>
<td>2.7462</td>
</tr>
</tbody>
</table>

Counts recorded in intervals

- Sampling rule is that observe only counts in ranges.
  e.g. 0, 1-4, 5-9, 10 and above.
- Interval density is simply
  \[ \Pr[a \leq y \leq b] = \sum_{j=a}^{b} f(j|x, \theta). \]
- Let interval ranges by \([a_0, a_1 - 1], [a_1, a_2 - 1], \ldots, [a_m, a_{m+1}]\), where \( a_0 = 0, a_{m+1} = \infty \).
  Let \( d_k \) be binary indicators for whether in interval \( k \) (\( k = 0, \ldots, m \)).
  Then
  \[ \ln L(\theta) = \sum_{i=1}^{N} \left[ \sum_{k=0}^{m} d_{ik} \ln \left( \sum_{j=a_{k+1}}^{a_k} f(j|x, \theta) \right) \right]. \]
- MLE is inconsistent if any aspect of the parametric model is misspecified.
- For convenience could instead use ordered logit or probit here.

Richer parametric models

- Data frequently exhibit “non-Poisson” features:
  > Overdispersion: conditional variance exceeds conditional mean whereas Poisson imposes equality.
  > Excess zeros: higher frequency of zeros than predicted by Poisson.
- This provides motivation for richer parametric models than basic Poisson.
- Some models still have \( \text{E}[y|x] = \exp(x'\beta) \)
  > Then richer model can provide more efficient estimates.
- Other models imply \( \text{E}[y|x] \neq \exp(x'\beta) \)
  > Then Poisson QMLE is inconsistent
  > And marginal effects and coefficient interpretation more difficult.
Hurdle model or two-part model

- Suppose zero counts are determined by a different process to positive counts.
  - Zeros: density $f_1(y|x_1, \theta_1)$ so $Pr[y = 0] = f_1(0)$ and $Pr[y > 0] = 1 - f_1(0)$.
  - Positives: density $f_2(y|x_2, \theta_2)$ so truncated density $f_2(y)/(1 - f_2(0))$.
- e.g. First - do I hunt this year or not? Second - given I chose to hunt, how many times ($\geq 1$)?
- Combined density is
  $$f(y|x_1, x_2, \theta_1, \theta_2) = \begin{cases} f_1(y|x_1, \theta_1) & y = 0 \\ \frac{1 - f_1(0|x_1, \theta_1)}{1 - f_2(0|x_2, \theta_2)} \times f_2(y|x_2, \theta_2) & y \geq 1 \end{cases}$$
- MLE is inconsistent if any aspect of model misspecified.

Conditional mean is now
$$E[y|x] = Pr[y_1 = 0|x_1] \times E_{y_2 > 0}[y_2|y_2 > 0, x_2].$$
- This makes marginal effects more complicated.
- Example: $f_1(\cdot)$ is logit and $f_2(\cdot)$ is negative binomial.
- Then
  $$E[y|x] = \Lambda(x_1 \theta) \times \exp(x_2 \beta) /[1 - (1 + \alpha_2 \exp(x_2 \beta))^{-1/\alpha_2}],$$
  where $\Lambda(z) = e^z/(1 + e^z)$.

Hurdle model - logit and negative binomial

```
. hnllogit docvis x1list, nolog
Negative Binomial-Logit Hurdle Regression
  Number of obs = 367
  Wald chi2(7) = 309.90
  Log likelihood = -10843.225
  Prob > chi2 = .0000

                      Coef.  Std. Err.      z    P>|z|     [95% Conf. Interval]
------------- -------- -------- -------- -------- --------------------------
Logit
  private    .658678   .1246408   5.21    0.000     .410893 -- .906563
   medicaid  .051825   .1726694   0.32    0.748    -.2830032 -- .395843
  age        .042878   .2289464   1.88    0.059   -.1190724 -- .816877
    yrs      -.003404   .0149067  -2.34    0.019   -.0068400 -- .000037
  educyr     .047035   .0855706   0.55    0.583     .0065571 -- .075529
  actlim     .1627927  .1527743   1.07    0.287  -.1392254 -- .464048
   age2      1.050562   .0619222  16.64    0.000     .9188676 -- 1.182236
   sex       -20.9463   .8433138  -25.11    0.000   -.3727827 -- -4.60508
  neighro
    neigh   -.1995866    .045429   3.17    0.002     .041891 -- .277222
     med    .0772038   .0470388   1.63    0.104    -.006981 -- .161403
     age     -.001192   .0004162  -4.62    0.000   -.0021112 -- .000769
    educyr  .0026974   .0043937   0.60    0.545     .0016855 -- .003709
  actlim    -.035384   .0351616   0.51    0.607   -.1056278 -- .034860
  age2      -9.69065   .9293992   10.34    0.000   -.2039481 -- .203232
  sex       -.019466   .0197326    -1.00    0.313     -.058232 -- .029299
  neigh2    -.025967   .0418671  -0.62    0.536    -.1068028 -- .054868
------------- -------- -------- -------- -------- --------------------------
AIC Statistic =  5.712
```

Zero-inflated model (or with-zeroes model)

- Suppose there is an additional reason for zero counts
  - Extra model for 0: density $f_1(y|x_1, \theta_1)$
  - Usual model for 0: realization of 0 from density $f_2(y|x_2, \theta_2)$.
- e.g. Some zeroes are mismeasurement and some are true zeros.
- Zero-inflated model has density
  $$f(y|x_1, x_2, \theta_1, \theta_2) = \begin{cases} f_1(0|x_1, \theta_1) + (1 - f_1(0|x_1, \theta_1)) \times f_2(0|x_2, \theta_2) & y = 0 \\ (1 - f_1(0|x_1, \theta_1)) \times f_2(y|x_2, \theta_2) & y \geq 1 \end{cases}$$
- MLE is inconsistent if any aspect of model misspecified.
- Not used much in econometrics - hurdle model more popular.
Zero-inflated negative binomial

Continuous mixture models

- Mixture motivation for negative binomial assumes \( y | \theta \sim \text{Poisson}(\theta) \)
  where \( \theta = \lambda \nu \) is the product of two components:
  - observed individual heterogeneity \( \lambda = \exp(x' \beta) \)
  - unobserved individual heterogeneity \( \nu \sim \text{Gamma}[1, \alpha] \).

- Integrating out
  \[
  h(y|\lambda) = \int f(y|\lambda, \nu) g(\nu) d\nu = \int [e^{-\lambda \nu} (\lambda \nu)^y / y!] \times g(\nu) d\nu
  \]
  gives \( y|\lambda \sim \text{NB} [\lambda, \lambda + \alpha \lambda^2] \) if \( \nu \sim \text{Gamma}[1, \alpha] \).

- Different distributions of \( \nu \) lead to different models
  - e.g. Poisson-lognormal mixture (random effects model)
  - e.g. Poisson-Inverse Gaussian.

- Even if no closed form solution can estimate using
  - numerical integration (one-dimensional) e.g. Gaussian quadrature.
  - Monte Carlo integration e.g. maximum simulated likelihood.

Hierarchical models

- For multi-level surveys cross-section data individuals \( i \) may be in cluster \( j \)
  - e.g. patient \( i \) in hospital \( j \)
  - e.g. individual \( i \) in household \( j \) or village \( j \)

- Hierarchical model or generalized linear mixed model example
  \[
  y_i \sim \text{Poisson}[\mu_{ij} = \exp(x_i' \beta_j + \epsilon_{ij})]
  
  \beta_j = W_j \gamma + v_j
  
  \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2_\epsilon)
  
  v_j \sim \mathcal{N}(0, \text{Diag}(\sigma^2_v))
  \]

- Estimate by MLE or by Bayesian methods.

Model comparison for fully parametric models

- Choice between nested models using likelihood ratio tests
  - e.g. Poisson versus negative binomial.

- Choice between non-nested models using Vuong’s (1989) likelihood ratio test
  - e.g. Zero-inflated NB versus NB

- Choice between non-nested mixture models using penalized log-likelihood
  - Akaike’s information criterion (AIC) and extensions (\( q = \# \) parameters)
    \[
    AIC = -2 \ln L + 2q
    
    BIC = -2 \ln L + q \ln N
    
    CAIC = -2 \ln L + q(1 + \ln N)
    \]

- Prefer model with small AIC or BIC.
- AIC penalty for larger model too small. Bayesian IC (BIC) better.
- Compare predicted means: \( \bar{E}[y|x, \hat{\theta}] \).
- Compare observed frequencies \( \bar{p}_j \) to average predicted frequencies

\[
\bar{p}_j = N^{-1} \sum_{i=1}^{N} \hat{p}_{ij},
\]

where \( \hat{p}_{ij} = \hat{P}(y_i = j) \).

Compare AIC, BIC for regular NB, hurdle logit/NB and zero-inflated NB.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>NBREG</th>
<th>HURDLENB</th>
<th>ZINB</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>3677</td>
<td>3677</td>
<td>3677</td>
</tr>
<tr>
<td>ll</td>
<td>-10593.2</td>
<td>-10942.9</td>
<td>-10492.9</td>
</tr>
<tr>
<td>aic</td>
<td>21196.7</td>
<td>21020.4</td>
<td>21034.8</td>
</tr>
<tr>
<td>bic</td>
<td>21252.6</td>
<td>21126.0</td>
<td>21125.3</td>
</tr>
</tbody>
</table>

Hurdle NB and ZINB are big improvement on regular NB
- \( \ln L \) is approximately 100 higher than for NB
- AIC and BIC is much smaller (with only 9 extra parameters)
Little difference between Hurdle NB and ZINB.

The conditional means from the three models are similar.

```stata
summarize docvis dvnreg dvhurdle dvzinh
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>docvis</td>
<td>3677</td>
<td>6.8252</td>
<td>7.394937</td>
<td>0</td>
<td>144</td>
</tr>
<tr>
<td>dvnreg</td>
<td>3677</td>
<td>6.890034</td>
<td>3.486562</td>
<td>2.07925</td>
<td>41.31503</td>
</tr>
<tr>
<td>dvhurdle</td>
<td>3677</td>
<td>6.840676</td>
<td>3.134925</td>
<td>1.35431</td>
<td>31.867874</td>
</tr>
<tr>
<td>dvzinh</td>
<td>3677</td>
<td>2.301923</td>
<td>2.139132</td>
<td>0.472877</td>
<td>39.48132</td>
</tr>
</tbody>
</table>

```stata
corrdocvis dvnreg dvhurdle dvzinh
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>docvis</th>
<th>dvnreg</th>
<th>dvhurdle</th>
<th>dvzinh</th>
</tr>
</thead>
<tbody>
<tr>
<td>docvis</td>
<td>1.0000</td>
<td>0.3870</td>
<td>0.3990</td>
<td>0.3983</td>
</tr>
<tr>
<td>dvnreg</td>
<td>0.3870</td>
<td>1.0000</td>
<td>0.9894</td>
<td>0.9882</td>
</tr>
<tr>
<td>dvhurdle</td>
<td>0.3990</td>
<td>0.9894</td>
<td>1.0000</td>
<td>0.9982</td>
</tr>
<tr>
<td>dvzinh</td>
<td>0.3983</td>
<td>0.9882</td>
<td>0.9982</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Finite mixtures model

- Density is weighted sum of two (or more) densities
  - Permits flexible models e.g. bimodal from Poissons.

- For an \( m \)-component model

\[
f(y|x, \theta, \pi) = \sum_{j=1}^{m} \pi_j f_j(y|x, \theta_j), \quad 0 \leq \pi_j \leq 1, \quad \sum_{j=1}^{m} \pi_j = 1.
\]

- For a 2-component model

\[
f(y|x, \theta_1, \theta_2, \pi) = \pi f_1(y|x, \theta_1) + (1 - \pi) f_2(y|x, \theta_2)
\]

- MLE maximizes

\[
\ln L(\theta) = \sum_{i=1}^{N} \ln(\pi f_1(y_i|x, \theta_1) + (1 - \pi) f_2(y_i|x, \theta_2)).
\]
  - Can restrict some parameters to be the same. e.g. only intercept differs
  - EM algorithm often used rather than Newton-Raphson.
Latent class model

- Finite mixture model can be interpreted as a latent class model.
- There are two types of people (given observables $X$)
  - e.g. "sick" type and "healthy" type
  - there is a probability of being drawn from either type.
- Similar to unobserved heterogeneity in duration data models.

$d(\theta) = \sum_{k=0}^{\infty} \left( \frac{1}{N} \sum_{i=1}^{N} f(Y_i = k; \theta, \pi) \right)^{1/2}$

where $P_k$ equals fraction of observations with $Y_i = k$.

attraction is that it is less influenced by outlying observations

estimate using an iterative method (HELMIX)

Component 1 occurs with probability 0.71 and is low use.
Component 2 occurs with probability 0.29 and is high use.

Component 1

Component 2

<table>
<thead>
<tr>
<th>Component</th>
<th>Poisson regression</th>
<th>Log-predictiveic $\beta_{1,0}$</th>
<th>$\beta_{1,2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.627</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.000</td>
<td>-0.000</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component</th>
<th>Poisson regression</th>
<th>Log-predictiveic $\beta_{1,0}$</th>
<th>$\beta_{1,2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>586.486</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-1.514</td>
<td>-1.514</td>
<td></td>
</tr>
</tbody>
</table>

A. Coia Cameron | Univ. of Calif. - Davis | Data: Count Data I | Sect. 1

March 30, 2000
Endogenous regressor: linear model

- Begin with review of the linear regression model: $y_i = x_i'\beta + u_i$.
- If regressors are correlated with error then OLS is inconsistent.
  - Reason: OLS $\hat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'u$ so
    $\lim\hat{\beta} = \beta + (\lim N^{-1}X'X)^{-1} \lim N^{-1}X'u$
    $\neq \beta$ if $\lim N^{-1}X'u \neq 0$.
- Solution: Assume the existence of an instrument $z$ where
  - changes in $z$ are associated with changes in $x$
  - but changes in $z$ do not lead to change in $y$ (aside from indirectly via $x$)

$$z \rightarrow x \rightarrow y \xrightarrow{\uparrow} y \leftarrow u$$

- Leads to instrumental variables (IV) estimator and two-stage least squares (2SLS) estimator.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>dvfit1</td>
<td>3677</td>
<td>3.801692</td>
<td>2.176922</td>
<td>9815556</td>
<td>27.28715</td>
</tr>
<tr>
<td>dvfit2</td>
<td>3677</td>
<td>13.91943</td>
<td>5.07463</td>
<td>5.615558</td>
<td>55.13366</td>
</tr>
<tr>
<td>dvcombined-t</td>
<td>3677</td>
<td>6.7385555</td>
<td>3.013985</td>
<td>2.342815</td>
<td>35.46714</td>
</tr>
<tr>
<td>docvis</td>
<td>3677</td>
<td>6.826282</td>
<td>7.394937</td>
<td>0</td>
<td>144</td>
</tr>
</tbody>
</table>

Log-likelihood comparison across models:
Poisson -15019; 2-component Poisson -11052; 2-component NB2 -10534; 2-component NB1 -10493.
Last is almost exactly same as hurdle NB and ZINB (-10493).

- Formally key assumption is:
  $E[u_i|z_i] = 0$
- Just-identified case (# instruments = # endogenous)
  $\hat{\beta}_{IV} = (Z'X)^{-1}Z'y$.
- Over-identified case (# instruments > # endogenous)
  $\hat{\beta}_{2SLS} = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y$.
- Example: log-earnings ($y$) regressed on years of school ($x$)
  - ability is an omitted regressor so part of error ($u$) and clearly correlated with $x$
  - instrument $z$ is correlated with years of school but not directly with earnings
  - example of $z$ may be distance from school or college.

Linear IV interpretation 1: method of moments

- We have $E[u_i|z_i] = 0 \Rightarrow E[z_iu_i] = 0 \Rightarrow E[z_i(y_i - x_i'\beta)] = 0$.
- The IV estimator solves the corresponding sample moment condition
  $\sum_{i=1}^N z_i(y_i - x_i'\beta) = 0$.
- Just-identified case can solve
  $\hat{\beta}_{IV} = (\sum_{i} z_i x_i) (\sum_{i} z_i y_i)^{-1} (\sum_{i} z_i y_i) = (Z'X)^{-1}Z'y.$
- Over-identified case cannot solve so minimize the quadratic form:
  $Q(\beta) = (\sum_{i} z_i(y_i - x_i'\beta))^TW(\sum_{i} z_i(y_i - x_i'\beta))$
  leads to generalized method of moments (GMM) estimator
  $\hat{\beta}_{GMM} = (Z'ZWZ'X)^{-1}Z'ZWZ'y$
  - 2SLS is special case $W = (Z'Z)^{-1}$.
- Essentially method of moments based on $E[z_iu_i] = 0$.
  - This generalizes to nonlinear models such as Poisson.
Linear IV interpretation 2: two-stage least squares

- Replace endogenous regressor by its predicted value.
- Specify structural equation for $y_1$ and reduced form equation for $y_2$
  - Split $x$ into endogenous regressor $y_2$ and exogenous regressors $z_1$
  - Split $z$ into instrument $z_2$ for $y_2$ and other exogenous regressors $z_1$
  
  Structural eqn: $y_{1i} = \beta_1 y_{2i} + z_{1i}^\prime \beta_2 + u_{1i}$

  Reduced-form eqn: $y_{2i} = \gamma_1 z_{2i} + z_{1i}^\prime \gamma_2 + v_{2i}$

- Two-stage least squares
  - OLS of $y_2$ on $z_2$ and $z_1$ gives prediction $\hat{y}_{2i} = \hat{\gamma}_1 z_{1i} + \hat{z}_{2i}^\prime \hat{\gamma}_2$.
  - OLS of $y_{1i}$ on $\hat{y}_{2i}$ and $z_{1i}$ gives estimates equal to IV/2SLS.
- Essentially OLS with $y_{2i}$ replaced by $\hat{y}_{2i}$
  - This does not generalize to nonlinear models such as Poisson.
  - In particular, it leads to inconsistent estimates.

Poisson endogenous method 1: nonlinear GMM

- Problem is $E[(y_i - \exp(x_i^\prime \beta)) | x_i] \neq 0$.
- Assume existence of instruments $z_i$ such that
  $$E[(y_i - \exp(x_i^\prime \beta)) | z_i] = 0$$
  $$\Rightarrow E[z_i (y_i - \exp(x_i^\prime \beta))] = 0$$

- Just-identified case: $\hat{\beta}_{MM}$ solves
  $$\sum_{i=1}^n (y_i - \exp(x_i^\prime \beta)) z_i = 0.$$ 

- Over-identified case $\hat{\beta}_{GMM}$ minimizes
  $$\left(\sum_{i=1}^n (y_i - \exp(x_i^\prime \beta)) z_i\right)^\prime W \left(\sum_{i=1}^n (y_i - \exp(x_i^\prime \beta)) z_i\right)$$
  - usually $W = (Z'Z)^{-1}$ (called nonlinear 2SLS).

Linear IV interpretation 3: control function

- Add predicted residual to control for endogeneity.
- Model relationship between structural model error and reduced form error
  $$u_{1i} = \alpha v_{2i} + \varepsilon_i$$
  where $\varepsilon_i$ is independent of $v_{2i}$, $y_{2i}$ and $z_{1i}$.
- Then
  $$y_{1i} = \beta_1 y_{2i} + \hat{z}_{1i}^\prime \beta_2 + \alpha v_{2i} + \varepsilon_i$$
- Control function approach
  - OLS of $y_2$ on $z_2$ and $z_1$ gives residual $\hat{v}_{2i} = y_{2i} - \hat{\gamma}_1 z_{1i} - \hat{z}_{2i}^\prime \hat{\gamma}_2$.
  - OLS of $y_{1i}$ on $\hat{y}_{2i}$, $z_{1i}$ and $\hat{v}_{2i}$ gives estimates equal to IV/2SLS.
- Essentially OLS with $y_{2i}$ augmented by the control for endogeneity $\hat{v}_{2i}$
  - This generalizes to nonlinear models such as Poisson.
Poisson endogenous: method 2 control function

- Add error in Poisson model (allows for overdispersion and endogeneity)

Structural eqn:
\[ y_{1i} \sim \text{Poisson}[\mu_i = \exp(\beta_1 y_{2i} + z_1^i \beta_2 + u_{1i})] \]

Reduced-form eqn:
\[ y_{2i} = \gamma_1 z_{2i} + z_1^i \gamma_2 + v_{2i} \]

Error model:
\[ u_{1i} = \alpha v_{2i} + \epsilon_i \]

- Then
\[ \mu_i | y_{2i}, z_{1i}, v_{2i}, \epsilon_i = \exp(\beta_1 y_{2i} + z_1^i \beta_2 + \alpha v_{2i} + \epsilon_i) \]
\[ = \exp(\epsilon_i) \exp(\beta_1 y_{2i} + z_1^i \beta_2 + \alpha v_{2i}) \]
\[ = E[\exp(\epsilon_i)] \exp(\beta_1 y_{2i} + z_1^i \beta_2 + \alpha v_{2i}) \]
\[ = \exp(\beta_1 y_{2i} + z_1^i \beta_2 + \alpha v_{2i}) \]

where if \( \epsilon_i \) is i.i.d. then \( E[\exp(\epsilon_i)] \) is a constant that is absorbed in \( \beta_2 \).

Control function approach

1. OLS of \( y_2 \) on \( z_2 \) and \( z_1 \) gives residual \( \hat{v}_{2i} = y_{2i} - \hat{\gamma}_1 z_{1i} - \hat{\gamma}_2 z_{2i} \).
2. Poisson of \( y_1 \) on \( y_{2i}, z_{1i} \) and \( \hat{v}_{2i} \) gives IV estimate.

Control function approach for same example.
First-stage: OLS for reduced form

```
. global xlist2 medicaid age2 educyr actlim totchr
. regress private xlist2 income ssiratio, vce(robust)
```

Linear regression

|          | Coef. | Robust Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|----------|-------|------------------|-------|------|----------------------|
| private  | -3.93447 | 0.137623        | -22.66 | 0.000 | -4.27844 - -3.594071 |
| medicaid | -0.08121 | 0.0293734 | -2.83  | 0.004 | -0.1407688 - -0.0215303 |
| age      | 0.005257 | 0.0001959 | 2.68   | 0.007 | 0.000411 - 0.0004908 |
| age2     | 0.002123 | 0.000204942 | 10.37  | 0.000 | 0.0012345 - 0.002527 |
| educyr   | -0.030896 | 0.0167674 | -1.70  | 0.089 | -0.064718 - 0.004545 |
| actlim   | 0.085063 | 0.0067743 | 3.22   | 0.001 | 0.0072965 - 0.0029769 |
| income   | 0.027416 | 0.0047986 | 5.79   | 0.000 | 0.0031813 - 0.00002652 |
| ssiratio | 0.064767 | 0.0011178 | -3.07  | 0.002 | -0.106675 - -0.223599 |
| _cons    | 3.551058 | 1.09581 | 3.22  | 0.001 | 1.3826 - 5.769516 |

```

Second stage: Poisson with first-stage predicted residual as regressor

```
. predict lpihat, residual
. * Second-stage Poisson with robust SEs
. poisson docsiv private xlist2 lpihat, vce(robust) nolog

Poisson regression
|          | Coef. | Robust Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|----------|-------|------------------|-------|------|----------------------|
| private  | 1.359554 | 0.2953175 | 2.24  | 0.025 | 0.897407 - 1.821768 |
| medicaid | -0.262043 | 0.0907642 | 2.91  | 0.004 | -0.442788 - -0.081316 |
| age      | 0.035604 | 0.0096904 | 3.47  | 0.000 | 0.016644 - 0.054564 |
| age2     | -0.000476 | 0.0000757 | -1.70  | 0.089 | -0.001955 - 0.000007 |
| educyr   | 0.080606 | 0.0065461 | 1.21  | 0.229 | 0.000000 - 0.161213 |
| actlim   | 0.076341 | 0.0041248 | 1.84  | 0.067 | 0.000000 - 0.153681 |
| totchr   | 0.24147 | 0.029175 | 8.22  | 0.000 | 0.20475 - 0.278195 |
| ssiratio | -0.156838 | 0.29497 | -5.27  | 0.000 | -0.640394 - -0.073272 |
| _cons    | -1.090647 | 2.661445 | -0.44  | 0.660 | -2.21744 - 0.036153 |

```

Private is 0.551 (0.245) compared to 0.340 for NL2SLS
Should bootstrap to get correct s.e.'s (1puhat is a generated regressor)

**Poisson endogenous method 3: structural approach**

- Example with binary endogenous regressor $y_{2i}$ is

\[
\text{Outcome eqn: } \quad y_{1i} \sim \text{Poisson}[\mu_i = \exp(\beta_1 y_{2i} + z_{1i}^T \beta_2 + \delta_1 u_i)]
\]

\[
\text{Participation eqn: } \quad \Pr[y_{2i} = 1] = \Lambda(z_{2i}^T \beta_2 + \lambda_1 u_i)
\]

\[
\text{Error model: } \quad u_i \sim \mathcal{N}[0, 1]
\]

- Estimate by simulated maximum likelihood.
- Deb and Trivedi (2006).
- Can also extend the two-part (hurdle) model to incorporate selection

- This allows for correlation due to unobservables between process for $y = 0$ or not and process for positives.

Here little change in standard errors.

**Quantile regression**

- The $q^{th}$ quantile regression estimator $\hat{\beta}_q$ minimizes over $\beta_q$

\[
Q(\beta_q) = \sum_{i:y_i \geq x_i^T \beta_q} q |y_i - x_i^T \beta_q| + \sum_{i:y_i < x_i^T \beta_q} (1-q) |y_i - x_i^T \beta_q|,
\]

$0 < q < 1$.

- Example: median regression with $q = 0.5$.
- For count $y$ adapt standard methods for continuous $y$ by:
  - Replace count $y$ by continuous variable $z = y + u$ where $u \sim \text{Uniform}[0, 1]$.
  - Then reconvert predicted $z$-quantile to $y$-quantile using ceiling function.

**References**

- Censored, truncated, hurdle and zero-inflated is in standard tests.
- Finite mixtures:
Endogenous regressors:

Quantile regression: