Advances in Count Data Regression:
IV. Further Topics

A. Colin Cameron
Univ. of Calif. - Davis

28th Annual Workshop in Applied Statistics
Southern California Chapter of the American Statistical Association
Held at University of California - Los Angeles

March 28 2009

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 Further Topics

- Multivariate data
- Maximum simulated likelihood
- Bayesian methods (very brief)
- Summary

Multivariate data

- Example is number of doctor visits and number of hospital stays.
- Multivariate versions of Poisson and negative binomial exist but are too restrictive
  - e.g. may permit only positive correlation with bound much less than 1.
- Instead approaches are
  - moment-based approach that generalizes seemingly unrelated equations
  - parametric approach that incorporates correlated latent variables
  - parametric approach that uses copulas to introduce correlation given specified marginal distributions
Multivariate data: bivariate Poisson

- Define count variables $y_1$ and $y_2$ (Kocherlakota and Kocherlakota, 1993)

$$
\begin{align*}
y_1 &= u_1 + u_3, \\
y_2 &= u_2 + u_3, \\
u_j &\sim \text{Poisson}[\mu_j], \ j = 1, 2, 3.
\end{align*}
$$

- Then joint frequency distribution is

$$
\Pr[y_1 = r, y_2 = s] = \exp[\mu_1 + \mu_2 + \mu_3] \sum_{l=0}^{\min(r,s)} \frac{\mu_1^r \mu_2^s \mu_3^l}{(r-l)!(s-l)!l!}.
$$

  - with marginals $y_1 \sim P[\mu_1 + \mu_3]$ and $y_2 \sim P[\mu_2 + \mu_3]$.

Multivariate data: moment-based approach

- Extend seemingly unrelated regressions for linear models.
- Specify

$$
\begin{align*}
E[y_1|x_1] &= \exp(x_1' \beta_1) \\
E[y_2|x_2] &= \exp(x_2' \beta_2)
\end{align*}
$$

- For variances and covariance suppose

$$
\begin{align*}
\text{Var}[y_1|x_1] &= \alpha_1 \exp(x_1' \beta_1) \\
\text{Var}[y_2|x_2] &= \alpha_2 \exp(x_1' \beta_1) \\
\text{Cov}[y_1, y_2|x_1, x_2] &= \rho \exp(x_1' \beta_1)^{1/2} \exp(x_2' \beta_2)^{1/2}
\end{align*}
$$

- Then estimate by GEE type estimator.
- Similar to univariate Poisson quasi-likelihood except improved efficiency is possible.

Multivariate data: correlated latent variables

- Squared correlation coefficient is bounded

$$
\rho_{12}^2 = \mu_3^2 / [(\mu_1 + \mu_3)(\mu_2 + \mu_3)]
$$

$$
0 \leq \rho_{12} \leq 3 / (\min(\mu_1, \mu_2))
$$

- Regression model specifies $\mu_j = \exp(x_j' \beta_j), \ j = 1, 2, 3$.

  - The marginal conditional means $E[y_1] = \mu_1 + \mu_3$ and $E[y_2] = \mu_2 + \mu_3$

  are no longer of simple exponential form.

- Introduce correlated latent variables (Marshall and Olkin, 1990).
- For example, for Poisson random variables

$$
\begin{align*}
y_1|x_1, v_1 &\sim P[\exp(v_1 + x_1' \beta_1)] \\
y_2|x_2, v_2 &\sim P[\exp(v_2 + x_2' \beta_2)] \\
(v_1, v_2) &\sim \text{bivariate normal}
\end{align*}
$$

  - This parametric model incorporates both overdispersion and correlation.

- Integrate out

$$
f(y_1, y_2|x_1, x_2, v_1, v_2) = \int f_1(y_1|x_1, v_1)f_2(y_2|x_2, v_2)g(v_1, v_2)dv_1dv_2.
$$

- Estimation is by maximum simulated likelihood.
Multivariate data: copulas

- Specify appropriate marginals and introduce correlation via copulas (Sklar, 1973).
- We begin with the Gaussian copula.
- Convert \( Y_1 \) and \( Y_2 \) to \( Y_1^* \) and \( Y_2^* \) with standard normal marginals as follows
  - For \( j = 1, 2 \) convert \( Y_j \) with c.d.f. \( F_j(\cdot) \) into \( U_j \sim \text{Uniform}(0,1) \)
    \[ U_j = F_j(Y_j), \quad j = 1, 2. \]
  - For \( j = 1, 2 \) convert \( U_j \) to \( Y_j^* \) with standard normal c.d.f. \( G_j(\cdot) \)
    \[ Y_j^* = G_j^{-1}(U_j), \quad j = 1, 2. \]
- Then model \( Y_1^* \) and \( Y_2^* \) as regular bivariate normal with zero means, unit variances and covariance \( \rho \).
- More generally for other copulas (Sklar, 1973)
  - A bivariate copula \( C(u_1, u_2) \) is a bivariate c.d.f. with marginals that are \( \text{Uniform}(0,1) \).
  - There are many copulas other than the Gaussian that can be applied to \( U_1, U_2 \).
  - Extension to multivariate case is immediate for the Gaussian copula.
- For regression case distribution of \( Y_j \) depends on regressors and parameters to be estimated.
- For Gaussian copula can do two-stage estimator
  - Estimate parameters of marginal distribution of \( y_j | x_j \), \( j = 1, 2 \).
  - Estimate parameters of the copula function \( \rho \) for bivariate case.
- For Gaussian copula FIML is difficult.

Multivariate data: copulas for counts

- The preceding implicitly assumed that \( Y_1 \) and \( Y_2 \) are continuous.
- For counts there is added complication that \( Y \) is discrete.
  - See for example Cameron, Li, Trivedi and Zimmer (2004).
- Also can use copulas to jointly model discrete and continuous random variables.

Maximum simulated likelihood: motivation

- Several models lead to low-dimensional integrals that have no closed-form solution.
- Cross-section example
  - Poisson with normally distributed intercept to capture overdispersion
- Panel example
  - Poisson with normally distributed intercept to capture correlation over time for a given individual
- Multivariate example
  - Poisson with multivariate normally distributed intercepts to capture correlation across variables.
- Can approximate integral by Gaussian quadrature or by Monte Carlo integration.
• Numerical integration using Gaussian quadrature:

\[ I = \int_a^b f(x) \, dx \]

\[ = \int_c^d w(x) r(x) \, dx \text{ where } r(x) = f(x) / w(x) \]

\[ \simeq \sum_{j=1}^m w_j r(x_j), \]

• Here three common choices of \( w(x) \) are:
  - \( w(x) = e^{-x^2} \) for \([c, d] = (-\infty, \infty)\) Gauss-Hermite
  - \( w(x) = e^{-x} \) for \([c, d] = (0, \infty)\) Gauss-Laguerre
  - \( w(x) = 1 \) for \([c, d] = [-1, 1]\) Gauss-Legendre

• Tables give good choices (for given \( m \)) of
  - weights \( w_j \)
  - evaluation points \( x_j \).

• Monte Carlo integration when \( g(x) \) is a density:

\[ E[h(x)] = \int_a^b h(x) g(x) \, dx \]

\[ \simeq S^{-1} \sum_{s=1}^S h(x^s) \]

• Here \( \{x^s, s = 1, \ldots, S\} \) is a Monte Carlo sample of \( S \) pseudo-random numbers from the density \( g(x) \).

• So to compute \( E[h(x)] \) when \( x \) has density \( g(x) \):
  - make many draws of \( x \)
  - compute \( h(x) \) each time
  - average these.

Maximum simulated likelihood

• Consider Poisson with normally distributed intercept

\[ y \sim \text{Poisson}[\exp(x' \beta + \nu)] \] with density \( g(y|x, \beta, \nu) \)

\[ \nu \sim \mathcal{N}[0, \sigma^2] \] with density \( h(\sigma) \).

• Then a typical observation has density

\[ f(y|x, \beta, \sigma) = \int g(y|x, \beta, \nu) h(\nu|\sigma) \]

\[ = \int_{-\infty}^{\infty} \frac{\exp(-x' \beta + \nu) (x' \beta + \nu)^y}{y!} \frac{1}{\sqrt{2\pi\sigma}} e^{-\nu^2/2\sigma^2} d\nu \]

\[ \simeq S^{-1} \sum_{s=1}^S \frac{\exp(-x' \beta + \nu_s^s) (x' \beta + \nu_s^s)^y}{y!} \]

where \( \nu_s^s, s = 1, \ldots, S \) are \( S \) draws from \( \mathcal{N}[0, 1] \) and \( S \to \infty \).

• The MLE of \((\beta, \sigma)\) maximizes

\[ \sum_{i=1}^N \ln f(y_i|x_i, \beta, \sigma). \]

• The maximum simulated likelihood estimator (MSL) maximizes

\[ \sum_{i=1}^N \ln \left( S^{-1} \sum_{s=1}^S g(y_i|x_i, \beta, \nu_s^s) \right) \]

\[ = \sum_{i=1}^N \ln \left( S^{-1} \sum_{s=1}^S \exp(-x' \beta + \nu_s^s) (x' \beta + \nu_s^s)^y/y! \right) \]

• Same distribution as MLE as \( S \to \infty \).

• Biased for finite \( S \) as unbiased for \( f(y_i) \) but not for \( \ln f(y_i) \)
  - Make \( S \) large but then computationally expensive as need to do this for each round of Newton’s iterative method.
Bayesian methods

1. Bayesian methods
   - Density or likelihood \( f(y|\theta) \)
   - Prior density \( \pi(\theta) \)
   - Posterior density
     \[
     f(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{f(y)}
     \]

2. Problem is that closed form solution for \( f(\theta|y) \) is rarely available.

Bayesian example

3. Recent computational advances make it possible to do Bayesian analysis when there is no closed form solution for the posterior density.
   - Importance sampling
   - Markov chain Monte Carlo (MCMC) methods
   - data augmentation


5. Bayesian methods can be used to enable Bayesian inference
   - different from classical statistical inference
   - can incorporate prior beliefs on parameters \( \theta \).

6. Bayesian methods can be used to enable classical inference
   - MLE is the Bayesian posterior mode if prior is diffuse
   - computationally faster then MSL in high-dimensional problems

7. Consider Poisson with normally distributed intercept
   \[
   y \sim \text{Poisson}[\exp(x_i^T\beta + v_i)]
   \]
   \[
   v_i \sim \mathcal{N}[0, \sigma^2].
   \]

8. Problem is that because \( v_i \) is not observed we need to integrate it out.

9. A Bayesian solution is data augmentation
   - generate values of the latent variables \( v_i \) and treat as data (like \( y_i \))

10. Actually more convenient to do this for
    \[
    z_i = \exp(x_i^T\beta + v_i).
    \]
The Bayesian procedure comprises

1. Joint density of data: \( y_i \) and \( z_i \) (given \( x_i, \beta, \sigma^2 \)) is
\[
y_i, z_i | x_i, \beta, \sigma^2 \sim \text{Poisson}[\mu_i = e^{z_i}] \times N[z_i, \sigma^2].
\]

2. Priors for the parameters: \( \beta \) and \( \sigma^2 \)
\[
\beta \sim N[0, kI]; \\
\sigma^{-2} \sim \text{Gamma}\left[\frac{b}{2} \times \left(\frac{c}{2}\right)^{-1}\right]
\]
where \( k = 10 \) and \( b = 5 \) and \( c = 10 \) give diffuse priors.

3. Posterior density for \( \beta, \sigma^2, z_i | y_i, x_i \) is computed recursively using MCMC algorithm.

MCMC for this example

- MCMC algorithm blocks as \( z_i, \beta, \sigma^2 \)
  - \( p(z_i | \beta, \sigma^2, x_i, y_i) \) needs Metropolis Hastings
  - \( p(\beta | -) \) is \( \beta \sim N[\overline{\beta}, \overline{H}_\beta^{-1}] \)
  - \( p(1/\sigma^2 | -) \) is \( \sigma^{-2} \sim \text{Gamma}\left[? \times (?)^{-1}\right] \)

Summary of count regression

- Cross-section count data basic approaches:
  - Moment-based
    * \( E[y | x] = \exp(x' \beta) \) and do Poisson QMLE with robust s.e.'s
  - Fully parametric
    * MLE of models richer than Poisson.

- Time series count data:
  - No standard preferred model.
- Panel count data
  - Econometricians focus on multiplicative individual specific effect.
- Fixed effects
  - Use quasi-difference $E[(y_{it} - (\lambda_{it} / \lambda_i) \bar{y}_i) | x_{it}, ..., x_{iT}] = 0$ with robust s.e.'s
  - For dynamic models use related quasi first-difference.
- Random effects
  - Population-averaged approach
  - Fully parametric MLE of models such as Poisson-gamma.

The cross-section and static panel count models can be estimated in STATA, LIMDEP and TSP.

- Count methods also exist (though no off-the-shelf programs) for the usual complications
  - Endogenous regressors
  - Sample selection
  - Multivariate data
  - Measurement error
  - Bayesian methods
  - Semiparametric methods

References

- Materials for these lectures are at http://cameron.econ.ucdavis.edu/racd/count.html
  - Stata program, dataset and output used for talk
  - Pravin Trivedi’s website (he will post final version of his paper soon).
- Multivariate data: Copulas

- Simulated maximum likelihood
- Bayesian methods