

Fitting the bivariate mixed Poisson regression model by maximum simulated likelihood

Stephen P. Jenkins

Professor of Economic and Social Policy, LSE

Email: s.jenkins@lse.ac.uk

Joint work with Fernando Rios-Avila

UK Stata Conference 2026

Introduction & motivations

- Poisson regression is a popular method for modelling an outcome in the form of a discrete integer count, even if data over/underdispersed (robustness results)
 - It's also increasingly popular for modelling non-negative outcomes, especially with zeros: Poisson Pseudo-Maximum Likelihood (PPML)
 - E.g. trade flows: Santos-Silva & Tenreyro, 'The log of gravity', *REStat*, 2006
- But what if you have data on **2 outcomes, each a discrete count**, for the same observation units?
 - E.g., (1) number of hospital visits per year by an individual, and (2) number of visits per year to a doctor by the same individual
 - You might expect there to be a non-zero correlation between the two outcomes – how large is it? ... even after adjusting using covariates ...
 - Answering the question requires a *bivariate* model in which **dependence** is modelled

Introduction & motivations (2)

Bivariate count data regression models with dependence are of 2 types:

1. Copula approach

- Stata implementation of several models: `bivcnto` by Xu & Hardin (SJ, 2016, 'XH')

2. Correlated latent factor (RE) approach

- One `bivcnto` option fits the Marshall-Olkin (1985) negbin model: 2 count outcomes with a common random effect (2 Poissons + 1 Gamma RE), i.e., correlation = 1
- This presentation is about `bimpoisson`, our new command for fitting 2 count outcomes in which the linear indices of predictors include a **bivariate normal RE**, thereby introducing over-dispersion
 - The model we implement is that of Munkin and Trivedi (*Econometrics Journal*, 1999, 'MT') employing optional first-order bias correction, antithetic acceleration, and with 2 types of Maximum Simulated Likelihood (MSL) estimator:
 - a) 'sampling function' MSL approach for case when $|\rho|$ close to 1 (default, as in MT)
 - b) Standard MSL estimator
 - We extend MT: allow Halton draws as well as standard uniform draws as in Jumamyradov & Munkin (*J Econometric Methods*, 2022, 'JM'; Matlab; discovered after we began our project)
 - We extend MT: use a bivariate normal sampling function as per JM (MT use standard bivariate normal)

- We're extending the portfolio of estimators available to Stata users
 - Not claiming that the MT model is the 'best' approach; that depends on research goals and context

Why MSL?

- Bivariate model for **2 binary outcomes** as in **biprobit**
 - Outcomes each modelled as linear indices of bivariate normal RE
 - You can write the log-likelihood as a closed form expression
 - So, you don't need MSL; standard ML is fine
- Bivariate model for **2 integer count outcomes**
 - Outcomes are non-linear functions of predictors including the bivariate normal RE
 - You can't write the log-likelihood as a closed form expression
 - So, you do need MSL
 - But we do 'borrow' a lot of syntax parsing code and help-file structure from **biprobit**
- MSL classics: Gouriéroux & Monfort (1997) and Train (2003), plus textbook discussions by chapters in, e.g., Greene and Cameron & Trivedi
- MSL in Stata: from **mvprobit** (Cappellari & Jenkins, *SJ* 2003) onwards, especially special issue of *SJ* 2006 with multiple papers on, e.g., Halton draws, GHK simulator, and implementations of MSL in various types of regression model; several *SJ* articles since
- MSL is tricky to implement for the MT model with sampling function approach, we found!

Outline

- This talk is based on a paper currently R&R at the *SJ*, with preprint available as IZA [Discussion Paper 18606](#); program and example dta- and do-files via `ssc d bimpoisson`
- Outline of the paper, also this talk:
 1. Introduction (done)
 2. MT's bivariate mixed Poisson regression model, and its estimation by MSL
 3. Introduction to `bimpoisson`, including estimation and post-estimation options
 4. Statistical performance of the estimators: our Monte Carlo analysis and review of results of earlier MC research by MT and JM
 5. Two empirical applications to datasets previously used
 - a) XH data: middle-range ρ
 - b) MT data: very high ρ
 6. Conclusions
- We provide guidance to users about empirical implementation choices such as type of MSL estimator, and type and number of draws, based on MC analysis and examples

MT's bivariate mixed Poisson regression model

MT model

For each obs:

Count variables y_1, y_2 ; vectors of observed predictors x_1, x_2 ;
unobserved REs $\varepsilon_1, \varepsilon_2$

$$y_j | x_j, \varepsilon_j \sim \text{Poisson}(\mu_j), \mu_j > 0, j = 1, 2$$

where

$$\mu_j = \exp(x_j' \beta_j + \varepsilon_j)$$

$$(\varepsilon_1, \varepsilon_2) \sim N[(0,0), \sigma_1^2, \sigma_2^2, \rho] \text{ with } 0 \leq |\rho| \leq 1$$

- $\Pr(y_1, y_2 | x_1, x_2)$ doesn't have closed form; hence use of MSL

MT model estimation by MSL

- Exploit Cholesky decomposition so that can sample from standard normal distribution

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 \\ \sigma_2 \rho & \sigma_2 \sqrt{1 - \rho^2} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}$$

- Standard MSL sample log-likelihood is (MT eq. 11)

$$\sum_{i=1}^N \log \left\{ \frac{1}{S} \sum_{s=1}^S \Pr(y_{1i} | x_{1i}, \sigma_1 \nu_{1i}^s) \Pr(y_{2i} | x_{2i}, \sigma_2 \rho \nu_{1i}^s + \sigma_2 \sqrt{1 - \rho^2} \nu_{2i}^s) \right\}$$

where $i = 1, \dots, N$ indexes observations; $S = \#$ draws

- Maximization of this expression produces estimates asymptotically equivalent to exact ML estimates if $S, N \rightarrow \infty$ with $\sqrt{N/S} \rightarrow 0$
 - Implement standard MSL estimator using `bimpoisson`'s `alt` option
- Caveat: if $|\rho|$ very close to 1, then log-likelihood 1st and 2nd derivatives w.r.t. ρ are unbounded $\Rightarrow$ you may get unreliable or infeasible estimates using standard MSL

MT: MSL using sampling function (SF) approach

- Sampling function ($\sim$ importance sampling) approach to address unboundedness problem
- Desirable properties of sampling function $g(.,.)$ (MT, p. 34):
 - Easy to take draws from; same support as density $f(.,.)$ and integrates to one; straightforward to calculate the inner sum in modified likelihood expression; bounded and smooth in parameters
- Modified sample log-likelihood:

$$\sum_{i=1}^N \log \left\{ \frac{1}{S} \sum_{s=1}^S \Pr(y_{1i}|x_{1i}, \varepsilon_{1i}^s) \Pr(y_{2i}|x_{2i}, \varepsilon_{2i}^s) \frac{f(\varepsilon_{1i}^s, \varepsilon_{2i}^s)}{g(\varepsilon_{1i}^s, \varepsilon_{2i}^s)} \right\}$$

where ε_{1i}^s and ε_{2i}^s are drawn from $g(\varepsilon_{1i}^s, \varepsilon_{2i}^s)$ for each observation i . The neat thing

- MT: nice statistical properties given standard MSL conditions on S, N ; but should use `vce(robust)`
- MT implement sampling function $g(.,.)$ as bivariate standard normal
- Following JM, we specify $g(.,.)$ as bivariate normal, i.e., unrestricted std. deviations, motivated by wanting to address cases in which true std. deviations of REs are > 1
 - Default MSL method in `bimpoisson`
- We also allow users to (optionally) set the sampling function correlation at a fixed value with the motivation that it makes draws more effective in terms of coverage and thence calculation of underlying double integral of bivariate Poisson density
 - We examine statistical performance using Monte Carlo analysis; not discussed here (see paper)

Additional tweaks to MSL estimation

- MT propose and justify a first-order bias correction method: use our **bias** option
- Derivation of bivariate distributions of draws
 - MT: pseudorandom numbers drawn from standard uniform distribution, also allowing antithetic acceleration (induces negative correlated draws for variance reduction)
 - **bimpoisson**: default is uniform draws, with **antithetic** option (in which case total number of draws = $2S$)
 - JM: Halton draws
 - **bimpoisson: halton** option, also with **burn** option (default = 5, given use primes 2, 3)
 - According to Train (2003, p. 231), “[t]he superior coverage and the negative correlation over observations that are obtained with Halton draws combine to make Halton draws far more effective than random [uniform] draws for simulation” ... but there is a caveat that this may be context dependent
- Number of simulations: default $S = 10$
 - Deliberately set very low, e.g., for syntax checking, but should always be larger ... much larger, we find from our MC analysis and examples (see later)
 - MT rule of thumb based on their MC analysis: for a sample with N^* observations, use

$$S = S^* \sqrt{N^*/n}$$

where S^* is # draws that MC analysis suggests yields unbiased parameter estimates
and n = sample size used in MC analysis (MT’s $n = 1000$)
- Our recommendation: fit model with some initial S , and then systematically increase S to check whether estimates change much
- NB we fit mixture distribution standard deviations in log metric and rho in atanh metric; and then transform back to natural metrics for reporting in estimates table and saved results

bimpoisson

Estimation and post-estimation

Syntax of **bimpoisson** (abbreviated)

bimpoisson *depvar1 depvar2 [varlist] [if] [in] [weight] [, options]*

Options

- **nsim(#)** specifies the number of simulations, *S*
- **antithetic** specifies that antithetic draws be used
- **halton** specifies use of Halton draws (default is standard uniform pseudorandom draws)
- **burn(#)** specifies the number of elements dropped when creating Halton draws (default is 5)
- **bias** specifies that the first-order bias correction be used
- **seed(#)** sets the seed for reproducibility; relevant if using standard uniform pseudorandom draws
- **rho sf(#)** sets the sampling function correlation, $|\#| < 1$ (default = 0)
- **alt** specifies use of the standard MSL estimator as alternative to the default estimator which uses a sampling function
- **vce(vcetype)** sets the vcetype; **robust** is recommended
- **rng0** sets the random number generator state back to its pre-estimation value (relevant if seed not set and uniform draws are used)
- **eform** reports exponentiated regression coefficients
- **display_options**
- **maximization_options** seldom used except **from()** when refitting a model using a greater number of draws than an initial model

bimpoisson Saved Results

In addition to the usual results saved after `m1`, `bimpoisson` also saves the following:

Scalars

- `e(nsims)` number of simulations, S
- `e(sigma1)` estimate of σ_1
- `e(sigma1_se)` estimate of $SE(\sigma_1)$
- `e(sigma2)` estimate of σ_2
- `e(sigma2_se)` estimate of $SE(\sigma_2)$
- `e(rho)` estimate of ρ
- `e(rho_se)` estimate of $SE(\rho)$
- `e(antithetic)` 1 if antithetic option used, 0 otherwise
- `e(halton)` 1 if halton option used, 0 otherwise
- `e(burn)` value of # set by `burn(#)`
- `e(bias)` 1 if bias option used, 0 otherwise
- `e(rho_sf)` value of # set by `rho_sf(#)` option
- `e(rng0)` 1 if `rng0` option used, 0 otherwise

Macros

- `e(seed)` contains the seed if set; “seed not set” otherwise

Post-estimation calculations

Count frequencies (marginal, joint); $\Pr(\text{count}_j = k)$
 Calculation formulae are in our paper (taken from MT)

predict *type newvar if in , statistic nooffset options*

Statistics

- **xb1** linear prediction for equation 1
- **xb2** linear prediction for equation 2
- **stdp1** standard error of the linear prediction for equation 1
- **stdp2** standard error of the linear prediction for equation 2
- **n1** count prediction for equation 1 (with random effect integrated out)
- **n2** count prediction for equation 2 (with random effect integrated out)
- **pr1 (#)** predicted prob(outcome 1 = #); # is integer-valued
- **pr2 (#)** predicted prob(outcome 2 = #); # is integer-valued
- **pr12 (#1 #2)** predicted prob(outcome 1 = #1, outcome 2 = #2); #1, #2 are integer-valued
 - Predicted probabilities are calculated using simulation; you can choose the number of sims and the seed

Monte Carlo analysis (MT's design + tweaks)

MT focused on

1. How many (standard uniform) draws – how does statistical performance estimates vary with S ?
 - MT's preferred case: unbiased estimates with $S = 50$ (plus 50 antithetic draws)
 - MT's results are conditional on their MC design of course (including $\sigma_1 = \sigma_2 = 0.5$, $\rho = -0.6, 0.6$)
2. First-order bias correction: how much does it help?
 - A lot – MT recommend using it
3. Antithetic acceleration: how much does it help?
 - A lot – MT recommend using it with standard uniform draws ('UA')
- Model parameters other than ρ generally unbiased regardless of S

JM used the same DGP as MT, but

- Halton draws throughout; and a modified sampling function (see above)
- Greater variation in parameter combinations, especially mixture distribution parameters (σ_1 , σ_2 , and ρ)
- Emphasise possible biases, especially for $\hat{\rho}$ when $\rho < 0.5$ or for σ_1, σ_2 when $\sigma_1, \sigma_2 \gg 1$

Our Monte Carlo analysis

Aims

- To check our code against MT's published results
- To examine statistical performance for a case with ρ close to 1
- How does the standard MSL method perform if ρ not close to 1?
- To compare performance of Halton versus UA draws (standard uniform + antithetics)
- To examine impacts of using larger and larger values of S

Hence, same DGP for non-mixture components as MT (and JM); plus $\sigma_1 = \sigma_2 = 0.5$ (as MT), and $\rho = 0.9$ (close to 1!) as well as $\rho = 0.6$

- **DGP**: 1000 observations on two predictors x_1 and x_2 drawn independently from $N(0, 1/16)$; two REs $(\varepsilon_1, \varepsilon_2) \sim N[(0,0), \sigma_1^2, \sigma_2^2, \rho]$; two count outcome variables $y_j \sim \text{Poisson}(\mu_j)$ with $\mu_j = \exp(\alpha_j + x_j\beta_j + \varepsilon_j)$; and the true values of the regression parameters are taken to be $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 1$ throughout
- **250 MC reps** rather than 100 (MT) or 200 (JM) – less simulation bias
- **Reporting**: (i) summarize simulation results using **simsum** (White, SJ 2010); (ii) report only $\hat{\rho}$ performance as RE variance and regression parameters unbiased

Our Monte Carlo analysis of $\hat{\rho}$

Focus on “mean_rho” estimates

Some headlines

- We couldn't replicate MT's results: e.g., for SF, UA, $S = 50$, MT estimate mean_rho = 0.607 for $\rho = 0.6$ case
- For $\rho = 0.6$, we need at least 200 draws
- Standard MSL ('Alt') performs well for $\rho = 0.6$
- Halton draws generally perform well relative to UA (and fitting faster since S draws, not $2S$)
- When $\rho = 0.9$, we still see a slight downward bias with 400 Halton draws
- All in all, appear to need many more draws than MT suggest, and it's better to use Halton draws

Table 1: Monte Carlo analysis of ρ parameter, by ρ , MSL method, draws type, and number of draws

	$\rho = 0.6$				$\rho = 0.9$	
	Alt		SF		SF	
	UA	H	UA	H	UA	H
$S = 50$						
mean_rho	0.623 (0.005)	0.588 (0.005)	0.518 (0.005)	0.572 (0.005)	0.748 (0.004)	0.805 (0.004)
SE_empirical	0.076 (0.003)	0.073 (0.003)	0.071 (0.003)	0.073 (0.003)	0.058 (0.003)	0.059 (0.003)
SE_model	0.088 (0.004)	0.075 (0.001)	0.071 (0.001)	0.080 (0.002)	0.057 (0.001)	0.063 (0.002)
$S = 100$						
mean_rho	0.610 (0.005)	0.591 (0.004)	0.546 (0.004)	0.591 (0.005)	0.823 (0.003)	0.829 (0.003)
SE_empirical	0.072 (0.003)	0.065 (0.003)	0.065 (0.003)	0.072 (0.003)	0.048 (0.002)	0.055 (0.002)
SE_model	0.075 (0.001)	0.071 (0.001)	0.070 (0.001)	0.074 (0.001)	0.054 (0.001)	0.064 (0.007)
$S = 200$						
mean_rho	0.605 (0.004)	0.596 (0.004)	0.568 (0.004)	0.599 (0.004)	0.823 (0.003)	0.853 (0.003)
SE_empirical	0.068 (0.003)	0.067 (0.003)	0.070 (0.003)	0.069 (0.003)	0.048 (0.002)	0.050 (0.002)
SE_model	0.071 (0.001)	0.070 (0.001)	0.069 (0.001)	0.073 (0.001)	0.054 (0.001)	0.060 (0.004)
$S = 400$						
mean_rho	0.603 (0.004)	0.596 (0.004)	0.581 (0.004)	0.601 (0.004)	0.850 (0.003)	0.880 (0.004)
SE_empirical	0.065 (0.003)	0.066 (0.003)	0.069 (0.003)	0.067 (0.003)	0.053 (0.002)	0.056 (0.003)
SE_model	0.070 (0.001)	0.067 (0.000)	0.069 (0.001)	0.071 (0.001)	0.053 (0.002)	0.057 (0.002)

NOTES: rho_sf = 0. UA = “uniform draws + antithetic draws” (total number of draws = $2S$). H = Halton draws. Alt: MSL. SF: MSL with sampling function. First-order bias correction applied. Robust standard errors. 250 Monte Carlo repetitions. Numbers in parentheses are Monte Carlo standard errors. Mean_rho: mean of estimated ρ . SE_empirical: standard deviation of estimated rho. SE_model: average of model-estimated SEs.

Application (1): XH data

Dataset, outcome and predictor variables exactly the same as used by Xu & Hardin (SJ, 2016)

3,874 individuals, 1989 German health registry data on number of doctor visits and number of days in hospital per year and predictor variables (no missing values)

Rho is mid-range

XH dataset

```
. use rwm1984.dta, clear
```

(German health data for 1984; Hardin & Hilbe, GLM and Extensions, 3rd ed)

```
. correlate docvis hospvis
```

```
(obs=3,874)
```

```

                | docvis hospvis
-----+-----
docvis         | 1.0000
hospvis        | 0.1458 1.0000

```

Table 3: Variables in the Xu-Hardin data

Variable name	Description	Mean
docvis	MD visits/year	3.163
hospvis	Days in hospital/year	0.121
MarF	1=married woman (reference category)	0.368
MarM	1=married man	0.421
SinM	1=single man	0.100
SinF	1=single woman	0.111
kids	1=have children; 0=no children	0.449
outwork	1=not working; 0=working	0.367
postHS	1=High school graduate or higher	0.186

Example output `bimpoisson` with XH data

Standard MSL (alt), UA draws with S = 400 (800 in total), seed set, plus bias correction, robust standard errors:

```
. bimpoisson docvis hospvis $xvars, robust nsim(400) bias alt ///
> antithetic seed(13159)

Generating pseudorandom uniform draws ...

Iteration 0: Log pseudolikelihood = -13419.598 (not concave)
Iteration 1: Log pseudolikelihood = -12022.94 (not concave)
Iteration 2: Log pseudolikelihood = -11450.799 (not concave)
Iteration 3: Log pseudolikelihood = -11071.132 (not concave)
Iteration 4: Log pseudolikelihood = -10353.703 (not concave)
Iteration 5: Log pseudolikelihood = -9825.9706 (not concave)
Iteration 6: Log pseudolikelihood = -9726.7421 (not concave)
Iteration 7: Log pseudolikelihood = -9725.1224 (not concave)
Iteration 8: Log pseudolikelihood = -9647.9827 (not concave)
Iteration 9: Log pseudolikelihood = -9614.5337
Iteration 10: Log pseudolikelihood = -9592.4443
Iteration 11: Log pseudolikelihood = -9584.8933
Iteration 12: Log pseudolikelihood = -9583.4684
Iteration 13: Log pseudolikelihood = -9583.3006
Iteration 14: Log pseudolikelihood = -9583.3003
```

Bivariate mixed Poisson regression (MSL)

Number of obs = 3,874

Wald chi2(6) = 140.91

Log pseudolikelihood = -9583.3003

Prob > chi2 = 0.0000

		Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
docvis							
	MarM	-.2648561	.0778074	-3.40	0.001	-.4173559	-.1123564
	SinM	-.4691794	.0950494	-4.94	0.000	-.6554728	-.282886
	SinF	.0742453	.1320539	0.56	0.574	-.1845756	.3330661
	kids	-.4425165	.062366	-7.10	0.000	-.5647515	-.3202815
	outwork	.5105739	.0926181	5.51	0.000	.3290458	.692102
	postHS	-.2619842	.0769608	-3.40	0.001	-.4128246	-.1111438
	_cons	.3326537	.0733339	4.54	0.000	.188922	.4763855
hospvis							
	MarM	.4414582	.1752062	2.52	0.012	.0980604	.784856
	SinM	-.2212477	.2987271	-0.74	0.459	-.806742	.3642467
	SinF	.1689398	.2508584	0.67	0.501	-.3227337	.6606133
	kids	-.1766938	.1586167	-1.11	0.265	-.4875769	.1341893
	outwork	.4563236	.1616185	2.82	0.005	.1395573	.77309
	postHS	-.4952979	.1829073	-2.71	0.007	-.8537896	-.1368062
	_cons	-4.371475	.2342582	-18.66	0.000	-4.830613	-3.912338
	/ln_sig1	.4447017	.0193587	22.97	0.000	.4067592	.4826441
	/ln_sig2	.8123536	.0531459	15.29	0.000	.7081896	.9165177
	/arho	.7033847	.0560584	12.55	0.000	.5935122	.8132572
	sigma1	1.560025	.0302001	51.66	0.000	1.501942	1.620353
	sigma2	2.253205	.1197486	18.82	0.000	2.030312	2.500568
	rho	.6065118	.035437	17.12	0.000	.5324169	.6713832

Note: 400 draws plus 400 antithetic draws, with first-order bias correction

XH data: vary S and draw type

- Model 1 is as shown on previous slide
- Increasing S substantially does make a difference especially to mixture parameter $\hat{\rho}$
- But it appears that ρ not close to 1, regardless
- Our preferred specification is Model 4

	(1)	(2)	(3)	(4)	(5)
	Alt UA	Alt H	Alt UA	Alt H	SF H
S	400	400	2,000	2,000	2,000
LogPL	-9,583	-9,575	-9,564	-9,551	-9,559
AIC	19,201	19,185	19,162	19,136	19,153

NOTES: Alt: MSL. SF: MSL with sampling function. First-order bias correction applied. UA: S uniform draws + S antithetic draws. H: S Halton draws. Robust standard errors in parentheses. $N = 3,874$.

	(1)	(2)	(3)	(4)	(5)
	Alt UA	Alt H	Alt UA	Alt H	SF H
S	400	400	2,000	2,000	2,000
<i>docvis</i>					
MarM	-0.265 (0.078)	-0.149 (0.089)	-0.304 (0.077)	-0.338 (0.066)	-0.376 (0.060)
SinM	-0.469 (0.095)	-0.588 (0.090)	-0.434 (0.103)	-0.718 (0.079)	-0.740 (0.069)
SinF	0.074 (0.132)	0.152 (0.110)	0.136 (0.087)	0.044 (0.090)	0.044 (0.078)
kids	-0.443 (0.062)	-0.494 (0.078)	-0.377 (0.057)	-0.477 (0.050)	-0.500 (0.050)
outwork	0.511 (0.093)	0.574 (0.062)	0.537 (0.067)	0.430 (0.058)	0.424 (0.049)
postHS	-0.262 (0.077)	-0.144 (0.081)	-0.209 (0.071)	-0.189 (0.068)	-0.153 (0.072)
_cons	0.333 (0.073)	0.278 (0.068)	0.282 (0.073)	0.421 (0.059)	0.455 (0.051)
<i>hospvis</i>					
MarM	0.441 (0.175)	0.563 (0.213)	0.342 (0.206)	0.461 (0.199)	0.350 (0.173)
SinM	-0.221 (0.299)	-0.254 (0.311)	-0.301 (0.276)	-0.295 (0.275)	-0.439 (0.292)
SinF	0.169 (0.251)	0.234 (0.248)	0.174 (0.225)	0.231 (0.227)	0.151 (0.241)
kids	-0.177 (0.159)	-0.068 (0.212)	-0.185 (0.145)	-0.116 (0.131)	-0.108 (0.164)
outwork	0.456 (0.162)	0.714 (0.180)	0.611 (0.181)	0.533 (0.176)	0.497 (0.156)
postHS	-0.495 (0.183)	-0.096 (0.351)	-0.283 (0.192)	-0.254 (0.195)	-0.237 (0.267)
_cons	-4.371 (0.234)	-4.522 (0.305)	-4.400 (0.253)	-4.415 (0.252)	-4.486 (0.291)
<i>mixture</i>					
sigma1	1.560 (0.031)	1.401 (0.030)	1.497 (0.031)	1.456 (0.020)	1.432 (0.020)
sigma2	2.253 (0.120)	2.048 (0.115)	2.130 (0.092)	2.081 (0.075)	2.233 (0.136)
rho	0.607 (0.035)	0.461 (0.050)	0.532 (0.040)	0.538 (0.039)	0.549 (0.043)

bimpoisson postestimation with XH data

```

quietly {
predict n1, n1
predict n2, n2
predict pr10, pr1(0)
predict pr11, pr1(1)
predict pr12, pr1(2)
predict pr13, pr1(3)
predict pr14, pr1(4)
predict pr15, pr1(5)
predict pr20, pr2(0)
predict pr21, pr2(1)
predict pr22, pr2(2)
predict pr23, pr2(3)
predict pr24, pr2(4)
predict pr25, pr2(5)
predict pr12_00, pr12(0 0)
predict pr12_10, pr12(1 0)
predict pr12_01, pr12(0 1)
predict pr12_11, pr12(1 1)
}

su docvis hospvis
su n1 n2
ta docvis
su pr1? // analogue to MT Table 5
ta hospvis
su pr2? // analogue to MT Table 5
ta docvis hospvis, nokey freq
cell
su pr12_?? // analogue to MT
Table 7

```

**Predictions based on Model 4
estimates (previous table)**

Table 5: Postestimation predictions from Xu-Hardin data (column %)

Counts	<i>docvis</i>		<i>hospvis</i>	
	empirical	fitted	empirical	fitted
0	41.6	37.5	92.3	92.3
1	11.6	20.0	6.1	5.4
2	11.4	11.3	1.1	1.2
3	9.1	7.1	0.2	0.4
4	5.5	4.8	0.1	0.2
5	4.3	3.4	0.1	0.2
other	16.5		0.1	
mean	3.16	3.60	0.12	0.05
<i>(docvis,hospvis)</i>				
(0,0)	40.1	37.0		
(1,0)	11.2	19.4		
(0,1)	1.3	1.0		
(1,1)	0.3	1.0		
other	47.1			

NOTES: Predictions from Model 4 in Table 4. Table entries show column percentages, except for 'mean' (count).

Application (2): MT data

Dataset, outcome and predictor variables exactly the same as used by Munkin & Trivedi
(*Econometrics J*, 1999)

4,406 elderly respondents to 1987/88 US National Medical Expenditure Survey with data
on

number of emergency room visits and number of hospital stays in reference period
and predictor variables (no missing values)

Rho is large (MT report $\hat{\rho} = 0.92$)

MT dataset

```
. use debtrivedi.dta, clear
```

```
. correlate emr hosp
```

```
(obs=4,406)
```

```
                | emr hosp
```

```
-----+-----
```

```
emr                | 1.0000
```

```
hosp                | 0.4761 1.0000
```

Compared to the XH data:

- Larger raw bivariate correlation
- More predictors, including health status ones
- Larger sample size

Table 6: Variables in the Munkin-Trivedi data

Variable name	Description	Mean
emr	# emergency room visits	0.264
hosp	# hospital stays	0.296
exclhlth	1= self-perceived health is excellent	0.078
poorhlth	1= self-perceived health is poor	0.126
numchron	# chronic health conditions	1.542
adldiff	1=has a condition that limits ADLs	0.204
noreast	1=lives in northeastern US	0.190
midwest	1=lives in the midwestern US	0.263
west	1=lives in the western US	0.181
age	age (years/10)	7.402
black	1=African American	0.117
male	1=male	0.404
married	1=married	0.546
school	# years of education	10.290
faminc	family income (\$10,000)	2.527
employed	1=employed	0.103
privins	1=covered by private health insurance	0.776
medicaid	1=covered by Medicaid	0.091

MT data: vary S and draw type (H vs. UA)

- Regression parameters not shown (for brevity; see paper): don't vary much for different models
- Focus on mixture parameter estimates
- 'MT' column: what MT reported
- Model 1: our version of 'MT'
- Other models switch to Halton draws and increase S substantially
- For these specifications, as S increases, so does $\hat{\rho}$
- For Model 6, $\hat{\rho} \approx 1$!
 - I.e., morphing into model like a Marshall-Olkin single common RE model (albeit normal RE rather than Gamma)

Table 7: Bivariate mixed Poisson regression model estimates, Munkin-Trivedi data

	MT	(1)	(2)	(3)	(4)	(5)	(6)
	UA	UA	H	H	H	H	H
S	150	150	150	850	850	2,500	5,000
$\rho_{\text{sf}}(\#)$	0	0	0	0	0.2	0	0
<i>mixture</i>							
sigma1	1.39 (0.06)	1.324 (0.053)	1.419 (0.066)	1.332 (0.057)	1.349 (0.054)	1.354 (0.059)	1.326 (0.052)
sigma2	1.36 (0.06)	1.323 (0.058)	1.369 (0.053)	1.301 (0.046)	1.338 (0.058)	1.363 (0.057)	1.331 (0.062)
rho	0.92 (0.02)	0.891 (0.021)	0.929 (0.018)	0.935 (0.013)	0.993 (0.004)	0.986 (0.010)	0.996 (0.003)
LogPL	-5,290	-5,305	-5,252	-5,241	-5,224	-5,251	-5,243
AIC	10,658	10,685	10,577	10,557	10,522	10,577	10,561

NOTES: MSL with sampling function. First-order bias correction applied. UA: S uniform draws + S antithetic draws. H: S Halton draws. Robust standard errors in parentheses. $N = 4,406$. 'MT' column estimates: [Munkin and Trivedi \(1999\)](#) Table 3).

bimpoisson postestimation with MT data

- Postestimation code is virtually identical to that used in the XH data example
- Predictions based on Model 6 estimates (previous table)
- Correspondences between empirical and predicted are generally quite close
- But MT report correspondences at least as close for a Marshall-Olkin model

Table 8: Postestimation predictions from the Munkin-Trivedi data (column %)

Counts	<i>emr</i>		<i>hosp</i>	
	empirical	fitted	empirical	fitted
0	81.8	82.4	80.4	80.9
1	13.4	12.3	13.6	13.0
2	3.1	3.0	4.0	3.3
3	1.2	1.1	1.1	1.2
4	0.3	0.5	0.5	0.6
5	0.2	0.2	0.1	0.3
other	0.0		0.3	
mean	2.26	0.29	0.30	0.33
<i>(emr,hosp)</i>				
(0,0)	72.4	72.1		
(1,0)	6.4	8.2		
(0,1)	7.3	9.2		
(1,1)	5.0	2.8		
other	8.9			

NOTES: Predictions from Model 6 in Table 7. Table entries show column percentages, except for 'mean' (count).

Conclusions

Conclusions

- **bimpoisson** enlarges the portfolio of models for bivariate count data models available to Stata users, coding up MT's mixture model
 - Whether to use copula or RE mixture type approach is unclear – former is harder to interpret but model fitting is easier (closed form likelihoods) and easier to model dependence as function of covariates, if need be (cf., Marra & Radice, 2025)
 - Heckman-Singer 'mass point' approaches to bivariate unobserved heterogeneity are another potential avenue for future research
 - Can **bimpoisson** be used for modelling dependence in non-negative variables?
 - What are the properties of bivariate PPML (cf. robustness of univariate PPML)?
- Although MT emphasize their sampling function approach, standard MSL works quite well unless ρ very high
- But do use Halton draws rather than UA: for given S , accuracy at least as good but shorter run-times
- It seems that good estimates require more draws than some might expect, especially when ρ is high
 - And remember to check estimates' robustness by fitting models with different S



References

References (1)

- Cameron, A. C., and P. K. Trivedi. 2005. *Microeconometrics. Methods and Applications*. 13th ed. Cambridge: Cambridge University Press.
- . 2013. *Regression Analysis of Count Data*. 2nd ed. Cambridge: Cambridge University Press.
- Cappellari, L., and S. P. Jenkins. 2003. Multivariate probit regression using simulated maximum likelihood. *The Stata Journal* 3(3): 278–294.
- . 2006. Calculation of multivariate normal probabilities by simulation, with applications to maximum simulated likelihood estimation. *The Stata Journal* 6(2): 156–189.
- Deb, P., and P. K. Trivedi. 1997. Demand for medical care by the elderly: a finite mixture approach. *Journal of Applied Econometrics* 12(3): 313–336.
- . 2006. Maximum simulated likelihood estimation of a negative binomial regression model with multinomial endogenous treatment. *The Stata Journal* 6(2): 246–255.
- Dhaene, G., and J. M. C. Santos Silva. 2012. Specification and testing of models estimated by quadrature. *Journal of Applied Econometrics* 27(2): 322–332.

Continued overleaf

References (2)

- Gouriéroux, C., and A. Monfort. 1991. Simulation based inference in models with heterogeneity. *Annales d'Économie et de Statistique* (20–21): 69–107.
- . 1997. *Simulation-based Econometric Methods*. Oxford: Oxford University Press.
- Greene, W. H. 2003. *Econometric Analysis*. 5th ed. Harlow: Longman.
- Haan, P., and A. Uhlenborff. 2006. Estimation of multinomial logit models with unobserved heterogeneity using maximum simulated likelihood. *The Stata Journal* 6(2): 229–245.
- Heckman, J. J., and B. Singer. 1984. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52(2): 271–320.
- Hole, A. R. 2007. Fitting mixed logit models by using maximum simulated likelihood. *The Stata Journal* 7(3): 388–401.
- Jumamyradov, M., and M. K. Munkin. 2022. Biases in maximum simulated likelihood estimation of bivariate models. *Journal of Econometric Methods* 11(1): 55–70.
- Marra, G., and R. Radice. 2025. Bivariate copula-based regression for joint modeling of healthcare visits. *Health Economics* 35(2): 332–345.
- Marshall, A. W., and I. Olkin. 1985. A family of bivariate distributions generated by the bivariate Bernoulli distribution. *Journal of the American Statistical Association* 80(390): 332–338.
- Munkin, M. K., and P. K. Trivedi. 1999. Simulated maximum likelihood estimation of multivariate mixed-Poisson regression models, with application. *The Econometrics Journal* 2(1): 29–48.
- Plum, A. 2016. bireprob: An estimator for bivariate random-effects probit models. *The Stata Journal* 16(1): 96–111.
- Sándor, Z., and P. András. 2004. Alternative sampling methods for estimating multivariate normal probabilities. *Journal of Econometrics* 120(2): 207–234.
- Stewart, M. B. 2006. Maximum simulated likelihood estimation of random-effects dynamic probit models with autocorrelated errors. *The Stata Journal* 6(2): 256–272.
- Train, K. E. 2003. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.
- White, I. R. 2010. simsum: Analyses of simulation studies including Monte Carlo error. *The Stata Journal* 10(3): 369–385.
- Xu, X., and J. W. Hardin. 2016. Regression models for bivariate count outcomes. *The Stata Journal* 16(2): 301–315.