

**VALIDATING A USER-DEVELOPED
BIVARIATE PSEUDO-RANDOM VECTOR GENERATOR**

by

**Joseph V. Terza
Golnoush Kazeminezhad
Department of Economics
Indiana University School of Liberal Arts at IUPUI
Indianapolis, IN 46202**

and

**Abbie Zhang
Department of Economics
School of Economics, Henan University
Kaifeng, 475004, China**

Context and Objective

Suppose you:

- have a new estimator for the parameters of a data generating process that has two outcomes (bivariate DGP)**
- seek to validate/assess the performance of the estimator using simulated data: statistical consistency; statistical precision; computational efficiency -- etc.**
- know the relevant aspects of the bivariate DGP [e.g. the pmf/pdf and, therefore, the cdf] conditional on regressors (presumed cause and controls)**
- developed a data simulator for the relevant bivariate DGP.**
- We offer some Mata coding tips on how to validate your bivariate data simulator by comparing generated relative frequencies of the outcome values to the corresponding true probabilities.**

Outline of Presentation

In this presentation, we will:

- Provide two examples in which the underlying data generating processes (DGPs) correspond with cases involving:**
 - (1) a Generalized Gamma Outcome w/ an Endogenous Treatment**
 - (2) Seemingly Unrelated Regressions (SUR) w/ Bivariate Dispersion-Flexible Count (DC) Outcomes**
- Overview the data simulation steps for each of these examples**
- Discuss our general approach for validating a bivariate data simulator**
- Discuss and demonstrate the fundamentals of Stata/Mata implementation for validation in the context of the second example (SUR-DC)**

Example: Generalized Gamma (GG) Outcome w/ Endogenous Treatment (ET)

-- Definitions, Notation, Assumptions

-- The relevant conditional potential outcomes model (CPOM) [see Terza, 2020]

$$\text{pdf}(Y_{X^*} | X_o, X_u) = \text{gg}(Y_{X^*}; X_o\beta_o + X^*\beta_X + X_u\beta_u, \sigma, \kappa) \quad (1)$$

where

X^* \equiv counterfactually mandated version of the treatment variable (binary)

Y_{X^*} \equiv potential outcome (is continuous but bounded from below at 0)

X_o \equiv the vector of observable control variables

X_u \equiv the scalar comprising the unobservable regressors

$\text{gg}(R; b, c, d)$ denotes the pdf of the GG random variable R with location, scale and shape parameters b, c and d , respectively, and the β s, σ and κ are parameters (for more information on the GG distribution see Manning et al., 2005).

Example: GG Outcome w/ ET -- Definitions, Notation, Assumptions (cont'd)

-- Suppose requisite conditions establishing the legitimacy (causal interpretability) of following aspect of the DGP specification are satisfied (see Terza, 2020)

$$\text{pdf}(Y | X_0, X, X_u) = \text{gg}(Y; X_0\beta_0 + X\beta_X + X_u\beta_u, \sigma, \kappa) \quad (2)$$

where

Y \equiv the observable version of the outcome

X \equiv the observable version of the treatment variable (binary)

Terza, J.V. (2020): “Regression-Based Causal Analysis from the Potential Outcomes Perspective,” *Journal of Econometric Methods*, published online ahead of print, DOI: <https://doi.org/10.1515/jem-2018-0030>.

Manning, W.G, Basu, A. and Mullahy, L. (2005): “Generalized Modeling Approaches to Risk Adjustment of Skewed Outcomes Data,” *Journal of Health Economics*, 24, 465-488.

Example: GG Outcome w/ ET -- Definitions, Notation, Assumptions (cont'd)

-- Moreover, suppose that

$$\mathbf{X} = \mathbf{I}(\mathbf{W}\boldsymbol{\delta} + \mathbf{X}_u > 0) \tag{3}$$

where

$$\mathbf{W} = [\mathbf{X}_0 \quad \mathbf{W}^+]$$

\mathbf{W}^+ is a vector of identifying instrumental variables

$(\mathbf{X}_u \mid \mathbf{W})$ is standard normally distributed

$\mathbf{I}(\mathbf{C})$ denotes the indicator function whose value is 1 if condition \mathbf{C} holds

and $\boldsymbol{\delta}$ is a vector of parameters.

Example: GG Outcome w/ ET -- Definitions, Notation, Assumptions (cont'd)

-- Under these assumptions, following Terza (2009), we can show that the joint pdf of Y and X conditional on W is

$$\text{pdf}(Y, X | W) = \left(\int_{-W\delta}^{\infty} \text{gg}(Y; X_0\beta_0 + X\beta_X + X_u\beta_u, \sigma, \kappa) \varphi(X_u) dX_u \right)^X \times \left(\int_{-\infty}^{-W\delta} \text{gg}(Y; X_0\beta_0 + X\beta_X + X_u\beta_u, \sigma, \kappa) \varphi(X_u) dX_u \right)^{1-X} \quad (4)$$

-- The relevant bivariate outcome in this case is $[Y \quad X]$.

-- Model parameters can be estimated by the full information maximum likelihood (FIML method based on (4)).

Terza, J.V. (2009): "Parametric Nonlinear Regression with Endogenous Switching," *Econometric Reviews*, 28, 555-580.

Example: GG Outcome w/ ET – Simulator and its Validation

- Because this estimator is novel and somewhat demanding (e.g. involves integrals) we would like to use simulated data to assess not only its consistency and precision properties but also its computational feasibility and efficiency.**
- The data simulator for assessing the FIML estimator is also based on (4).**
- This model is similar to Terza (2019), presented at this conference in 2019.**

Terza, J. (2019): “Mata Implementation of Gauss-Legendre Quadrature in the M-Estimation Context: Correcting for Sample-Selection Bias in a Generic Nonlinear Setting,”
https://www.stata.com/meeting/chicago19/slides/chicago19_Terza.pdf

Example: GG Outcome w/ ET – Simulator and its Validation

-- The relevant data simulation protocol comprises the following steps:

- i) Specify the *sampling design* comprising: a) values for δ , the β s, σ and κ ; and b) simple univariate and independent data generators for X_0 and W^+ .
- ii) Generate values of X_0 and W^+ for a given simulated sample size, according to the specified sampling design.
- iii) Generate standard normal values of X_u .
- iv) Generate values of X based on (3), X_u in (iii) and the design value of δ .
- v) Generate values of Y based on (4), the values of X_0 , W^+ and X from steps (ii), (iii) and (iv) and the sampling design values of the β s, σ and κ .

-- Before using data simulated via this protocol to assess the proposed estimator, it would be helpful to first assess the validity of the data simulator itself!

Example: Seemingly Unrelated Regression (SUR) w/ Bivariate Dispersion-Flexible

Count (DC) Outcomes

-- Definitions:

X^* \equiv counterfactually mandated version of the causal variable (any type)

$[Y_{1X^*} \quad Y_{2X^*}] \equiv$ bivariate vector of count-valued potential outcomes

$X_0 \equiv$ the vector of observable control variables

Example: SUR w/ Bivariate DC Outcomes (cont'd)

-- The relevant conditional potential outcomes model (CPOM) [see Terza, 2020 and Zhang and Terza, 2021]

$$\begin{aligned} \text{pmf}(Y_{1X^*}, Y_{2X^*} | X_0) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{cmp}_1(Y_{1X^*}; \lambda_1^*, \omega_1) \times \text{cmp}_2(Y_{2X^*}; \lambda_2^*, \omega_2) \\ &\quad \times \varphi_2(\mathbf{h}_1, \mathbf{h}_2; \rho_{12})] d\mathbf{h}_1 d\mathbf{h}_2 \quad (5) \end{aligned}$$

where

$\text{cmp}_r(Y_{rX^*}; \lambda_r^*, \omega_r) \equiv$ the pmf of the Conway-Maxwell-Poisson (CMP) distributed random variable Y_{rX^*} with parameters λ_r^* and ω_r ($r = 1, 2$)

$$\lambda_r^* \equiv \exp(X_0 \beta_{r0} + X^* \beta_{rX} + h_r)$$

Zhang, A. and Terza, J.V (2021): "Causal Inference Based on Correlated Dispersion-Flexible Count Regression," Unpublished Manuscript, Department of Economics, Indiana University School of Liberal Arts at IUPUI.

Example: SUR w/ Bivariate DC Outcomes (cont'd)

$\phi_2(\mathbf{h}_1, \mathbf{h}_2; \rho_{12}) \equiv$ the pdf of the standard bivariate normal vector $[\mathbf{h}_1 \quad \mathbf{h}_2]$ with correlation ρ_{12}

and the parameters to be estimated are β_{r0} , β_{rX} , ω_r (for $r = 1, 2$), and ρ_{12} .

-- For details on the CMP distribution see Conway & Maxwell (1962), Shmueli et al. (2005), Lord et al. (2008), Sellers & Shmueli (2010) and Huang (2017).

Conway, R. W., & Maxwell, W. L. (1962): "A Queuing Model with State Dependent Service Rates," *Journal of Industrial Engineering*, 12: 132–136

Huang, A. (2017): "Mean-parametrized Conway–Maxwell–Poisson regression models for dispersed counts," *Statistical Modelling*, 17(6), 359-380.

Lord, D., Guikema, S. D., & Geedipally, S. R. (2008): "Application of the Conway–Maxwell–Poisson generalized linear model for analyzing motor vehicle crashes," *Accident Analysis & Prevention*, 40(3), 1123-1134.

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., & Boatwright, P. (2005): "A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1), 127-142.

Sellers, K. F., & Shmueli, G. (2010): "A flexible regression model for count data," *The Annals of Applied Statistics*, 4(2), 943-961, 919.

Example: SUR w/ Bivariate DC Outcomes (cont'd)

-- Suppose that the requisite conditions establishing the legitimacy of following aspect of the DGP specification are satisfied (see Terza, 2020)

$$\text{pmf}(Y_1, Y_2 | X_0, X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{cmp}_1[(Y_1; \lambda_1, \omega_1) \times \text{cmp}_2(Y_2; \lambda_2, \omega_2) \times \varphi_2(\mathbf{h}_1, \mathbf{h}_2; \rho_{12})] d\mathbf{h}_1 d\mathbf{h}_2 \quad (6)$$

$[Y_1 \quad Y_2] \equiv$ the observable version of the outcome vector

$X \equiv$ the observable version of the causal variable

$\lambda_r \equiv \exp(X_0\beta_{r0} + X\beta_{rX} + \mathbf{h}_r)$ for $r = 1, 2$

-- Model parameters can be estimated by the full information maximum likelihood (FIML method based on (6)).

Example: SUR w/ Bivariate DC Outcomes (cont'd)

-- Details for (5) and (6)

$$\text{cmp}_r(Y_r; \lambda_r, \omega_r) = \frac{\lambda_r^{Y_r}}{Y_r! \exp(\omega_r) Z(\lambda_r, \exp(\omega_r))}$$

where

$$Z(\lambda_r, \exp(\omega_r)) = \sum_{j=0}^{\infty} \frac{\lambda_r^j}{(j!) \exp(\omega_r)}$$

Example: SUR w/ Bivariate DC Outcomes (cont'd)

-- This model is designed to:

-- Exploit possible statistical efficiency in estimation by taking explicit (parametric) account of cross-equation correlation through the bivariate normal mixture component (essentially ρ_{12}).

-- Avoid bias and efficiency loss by circumventing the equi-dispersion restriction imposed by the conventional Poisson regression model.

-- For details on the CMP distribution see the references listed in the posted version of the slides for this presentation.

Example: SUR w/ Bivariate DC Outcomes (cont'd)

- Because this estimator is novel and somewhat demanding (e.g. involves bivariate integrals) we would like to use simulated data to assess not only its consistency and precision properties but also its computational feasibility and efficiency.**
- The data simulator for assessing the FIML estimator is also based on (6).**
- The bivariate quadrature method implemented for the FIML estimator based on (6) is detailed Terza and Zhang (2020), presented at this conference in 2020.**

Terza, J. and Zhang, A. (2020): “Two-Dimensional Gauss-Legendre Quadrature: Seemingly Unrelated Dispersion-Flexible Count Regressions,” Presentation slides for Stata Conference published at: https://www.stata.com/meeting/us20/slides/us20_Terza.pdf

Example: SUR w/ Bivariate DC Outcomes -- Data Generator for $(Y_1, Y_2 | X, X_0)$

-- The relevant data simulation protocol comprises the following steps:

i) Specify the *sampling design* comprising: a) values for β s, ω s and ρ_{12} ; and b)

simple univariate and independent data generators for X_0 and X .

ii) Generate values of X_0 and X for a given simulated sample size, according to the specified sampling design.

iii) Generate joint standard normally distributed values for the vector $[h_1 \ h_2]$ using the design value of ρ_{12} .

iv) Generate values of the vector $[Y_1 \ Y_2]$ based on the values of X_0 , X and $[h_1 \ h_2]$ generated in steps (ii) and (iii) and the design values of the β s, ω s and ρ_{12} .

-- As in the previous example, we need to assess the validity of the data simulator.

SUR w/ Bivariate Outcomes: Data Generator for $(Y_1, Y_2 | X, X_0)$ (cont'd)

-- We represent the generic data simulator in this context as

$$\mathbf{Y1Y2} = \mathbf{Y1Y2datagen}(\mathbf{XB1}, \mathbf{XB2}, \text{additional parameters}) \quad (7)$$

where

$\mathbf{Y1Y2}$ is the $\text{sampsize} \times 2$ matrix comprising the generated (Y_1, Y_2) pairs with sampsize being the size of the simulated sample.

$\mathbf{XB1}$ is the $\text{sampsize} \times 1$ vector comprising the values of $\mathcal{X}^\dagger \beta_1$ with $\mathcal{X}^\dagger = [\mathcal{X} \ \mathcal{X}_0]$

being a $\text{sampsize} \times (K + 1)$ matrix of generated values for $\mathbf{X}^\dagger = [\mathbf{X} \ \mathbf{X}_0]$ and

$\beta_1' = [\beta_{1X} \ \beta_{10}']$ is the chosen (sampling design) value of the coefficient

parameters for the first index (note that, as always, \mathbf{X}_0 includes a constant term).

SUR with DC Outcomes: Data Generator for $(Y_1, Y_2 | X, X_0)$ (cont'd)

and

X_{B2} is similarly defined.

In our SUR with DC outcomes (CMP/normal mixture) model, additional parameters include ω_1 , ω_2 and ρ_{12} .

SUR with DC Outcomes: Validating the Data Generator for $(Y_1, Y_2 | X, X_0)$

- The validation will be conducted by comparing the table of joint cumulative relative frequencies (crf) for a generated sample of (Y_1, Y_2) pairs, obtained using the data generator in (7), with the true joint probabilities based on cdf values for (Y_1, Y_2) .**
- The joint crf and cdf values will be calculated for cells in the support of $[Y_1 \quad Y_2]$. defined by vectors of cell threshold values for Y_1 and Y_2 (say T_1 and T_2 , respectively; whose typical elements are t_1 and t_2 , respectively).**

SUR with DC Outcomes: Validation Sampling Design

-- In your Mata code specify:

sampsize = size of the sample to be generated

xb1 = a fixed value corresponding to $X^\dagger \beta_1$ in your data simulator

xb2 = a fixed value corresponding to $X^\dagger \beta_2$ in your data simulator

values for the additional requisite parameters

and

T1 and **T2** \equiv vectors of threshold values for Y_1 and Y_2 , respectively.

NOTE THAT THE FIRST AND LAST ELEMENTS OF BOTH T1 and T2 ARE VIRTUAL $-\infty$ AND $+\infty$, RESPECTIVELY.

SUR with DC Outcomes: Simulated Data Generation for Validation

-- To validate the simulator, generate the data on (Y_1, Y_2) using the data simulation

Mata function in (7) such that:

```
Y1Y2=Y1Y2datagen(xb1:*J(sampsize,1,1),xb2:*  
                  J(sampsize, 1,1),additional parameter values)  
                  (8)
```

-- Note that $\mathbf{xb1} : *J(\text{sampsize}, 1, 1)$ and $\mathbf{xb2} : *J(\text{sampsize}, 1, 1)$ are $\text{sampsize} \times 1$ vectors of constants.

SUR with DC Outcomes: Distributing Simulated (Y_1, Y_2) Pairs to Cells

-- Next, use Mata code to distribute the generated pairs of values of Y_1 and Y_2 in $Y1Y2$ into cells defined by ordered pairs of elements from T_1 and T_2 , respectively

[e.g., (t_1, t_2)]. For instance

$$T_1 = -1, 1, 2, 3, 4, 100$$

$$T_2 = -1, 1, 2, 3, 4, 100$$

-- To do this, first set up threshold subvectors for cell lower limits and upper limits

for Y_1 and Y_2 . For example, use the following mata code:

```
threshL1=T1[1..cols(T1)-1] e.g., = -1, 1, 2, 3, 4
threshU1=T1[2..cols(T1)] e.g., = 1, 2, 3, 4, 100
threshL2=T2[1..cols(T2)-1] e.g., = -1, 1, 2, 3, 4
threshU2=T2[2..cols(T2)] e.g., = 1, 2, 3, 4, 100
```

SUR with DC Outcomes: Distributing Simulated (Y_1 , Y_2) Pairs to Cells (cont'd)

-- Expand Y_1 [Y_2] into two matrices of order

`sampsize×cols(threshL1)` and `sampsize×cols(threshU1)`

e.g., 5

e.g., 5

`[sampsize×cols(threshL2) and sampsize×cols(threshU2)]`

e.g., 5

e.g., 5

using the following mata code

```
BIGY1L1=J(1,cols(threshL1),1)#Y1 e.g. sampsize×5
```

```
BIGY1U1=J(1,cols(threshU1),1)#Y1
```

```
BIGY2L2=J(1,cols(threshL2),1)#Y2 e.g. sampsize×5
```

```
BIGY2U2=J(1,cols(threshU2),1)#Y2
```


SUR with DC Outcomes: Distributing Simulated (Y_1 , Y_2) Pairs to Cells (cont'd)

-- Using the following mata code, construct the matrices of 1/0's that indicate the cell in which a particular pair of Y_1 and Y_2 values resides relative to the threshold values.

An element of this matrix will be 1 if the corresponding value of Y_1 (Y_2) lies between the relevant pair of given threshold values for a particular cell.

```
indicthreshL1=BIGY1L1:>threshL1  
indicthreshU1=BIGY1U1:<=threshU1  
indicthresh1=indicthreshL1:*indicthreshU1  
indicthreshL2=BIGY2L2:>threshL2  
indicthreshU2=BIGY2U2:<=threshU2  
indicthresh2=indicthreshL2:*indicthreshU2
```

SUR with DC Outcomes: Calculating the Cell Relative Frequencies

-- The matrix product of the indicator matrices (`indicthresh1` and `indicthresh2`) yields a $[\text{cols}(T1) - 1] \times [\text{cols}(T2) - 1]$ matrix of cell frequencies for the simulated (Y_1, Y_2) pairs. The relevant mata statement is

```
cumfreqY1Y2=indicthresh1'indicthresh2
```

-- Divide by the sample size to get the cell relative frequencies. The relevant mata statement is

```
relcumfreqY1Y2=cumfreqY1Y2:/samsize
```

SUR with DC Outcomes: Calculating the True Cell Probabilities

- Because we know the pmf in (6) we also know the corresponding cdf.**
- Suppose that the corresponding meta function for this cdf is**

$$\text{cdf}_{Y_1 Y_2} = \text{cdf}_{Y_1 Y_2} \text{fun}(Y_1, Y_2, X_{B1}, X_{B2}, \text{additional parameters}) \quad (9)$$

where

$Y_1, Y_2 \equiv$ column vectors of Y_1 and Y_2 values identical row dimensions

and

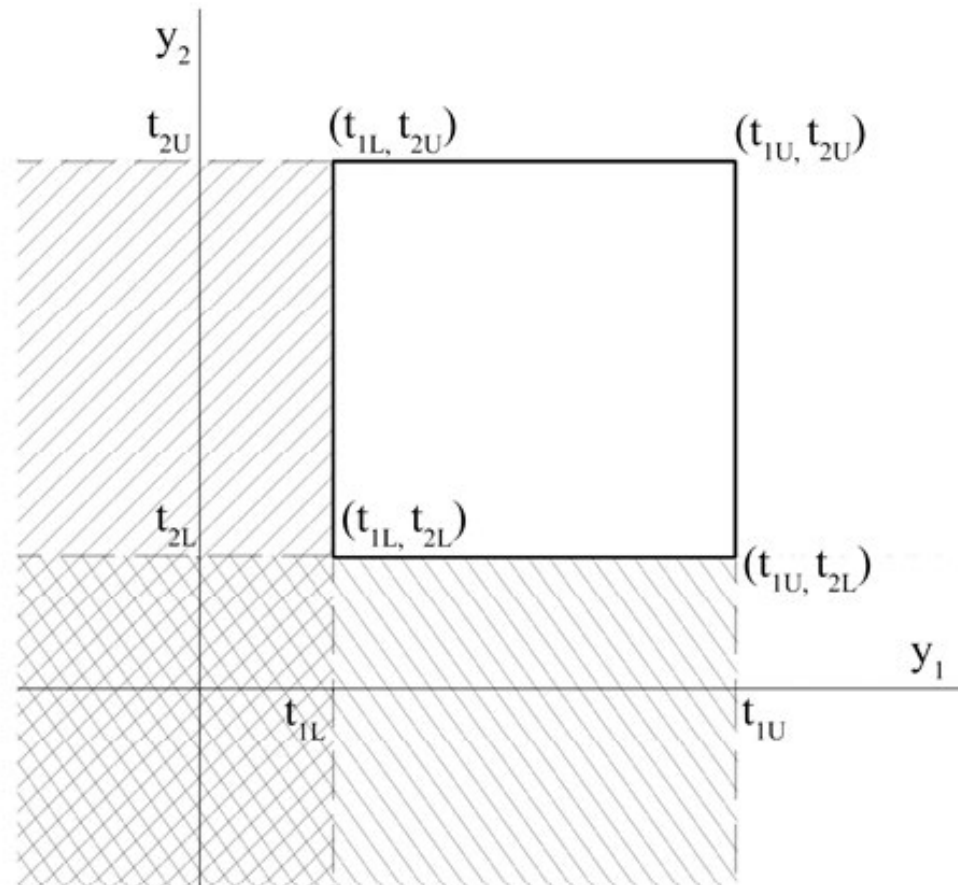
X_{B1}, X_{B2} are defined as in (7).

SUR with DC Outcomes: Calculating the True Cell Probabilities (cont'd)

-- Consider the cell defined by the threshold values t_{1L} and t_{1U} for Y_1 and the threshold values t_{2L} and t_{2U} for Y_2 as pictured in the figure. In terms of the joint cdf of Y_1 and Y_2 [$F(Y_1, Y_2)$], the probability measure of the rectangle with corners (t_{1L}, t_{2L}) , (t_{1U}, t_{2L}) , (t_{1L}, t_{2U}) and (t_{1U}, t_{2U}) is

$$F(t_{1U}, t_{2U}) - F(t_{1U}, t_{2L}) - F(t_{1L}, t_{2U}) + F(t_{1L}, t_{2L}) \quad (10)$$

$$\mathbf{F}(t_{1U}, t_{2U}) - \mathbf{F}(t_{1U}, t_{2L}) - \mathbf{F}(t_{1L}, t_{2U}) + \mathbf{F}(t_{1L}, t_{2L}) \quad (10)$$



$$\mathbf{F}(t_{1U}, t_{2U}) = \square$$

$$\mathbf{F}(t_{1L}, t_{2U}) = \text{diagonal hatching (top-left to bottom-right)}$$

$$\mathbf{F}(t_{1U}, t_{2L}) = \text{diagonal hatching (top-right to bottom-left)}$$

$$\mathbf{F}(t_{1L}, t_{2L}) = \text{cross-hatching}$$

SUR with DC Outcomes: Calculating the True Cell Probabilities (cont'd)

-- We first construct the 2-column matrix of all possible contiguous pairs of threshold values for Y_1 using the following mata statement

```
contigpairs1=T1[1..cols(T1)-1]',T1[2..cols(T1)]'
```

and do the same for Y_2

```
contigpairs2=T2[1..cols(T2)-1]',T2[2..cols(T2)]'
```

SUR with DC Outcomes: Calculating the True Cell Probabilities (cont'd)

-- Now construct the 4-column matrix of all possible combinations of contiguous pairs of threshold values for Y1 and Y2

```
pairsofpairs = contigpairs1#J(cols(T2)-1,1,1), J(cols(T1)-1,1,1)#contigpairs2
```

e.g. 25×2 horizontally concatenated with 25×2

result is a 25×4 matrix

-- Each row of this matrix corresponds to a cell defined by the relevant upper and lower thresholds for Y1 and Y2, respectively. We have

column of “pairsofpairs”

corresponding threshold

1, 2, 3, 4

$t_{1L}, t_{1U}, t_{2L}, t_{2U}$

SUR with DC Outcomes: Calculating the True Cell Probabilities (cont'd)

-- Accordingly, based on (10), the mata statement (in stylized form) that produces the vector of cell probabilities corresponding with the rows of the matrix “pairsofpairs”

is

```
cellprobs=A :- B :- C: + D
```

where

```
A=cdfY1Y2fun(pairsofpairs[.,2],pairsofpairs[.,4],XB1,XB2,additional params)
```

```
B=cdfY1Y2fun(pairsofpairs[.,1],pairsofpairs[.,4],XB1,XB2,additional params)
```

```
C=cdfY1Y2fun(pairsofpairs[.,2],pairsofpairs[.,3],XB1,XB2,additional params)
```

```
D=cdfY1Y2fun(pairsofpairs[.,1],pairsofpairs[.,1],XB1,XB2,additional params)
```

and `cdfY1Y2fun ()` is given in (9).

SUR with DC Outcomes: Comparing Rel Freq with True Probabilities (cont'd)

-- To facilitate comparison between relative frequencies and true probabilities reshape the vector of true cell probabilities to conform to the order of the matrix of cell relative frequencies using the following mata statement

```
tabcellprobs=colshape (cellprobs , cols (T2) -1)
```

-- Calculate the matrix of absolute value cell differences between the relative frequencies and the true probabilities

```
abscelldiff=abs (relcumfreqY1Y2 : -tabcellprobs)
```

-- As a metric for validating the data simulator, calculate the maximum absolute cell difference

```
max (abscelldiff)
```

SUR with DC Outcomes: Example for a Specific Sampling Design

-- Details of the sampling design:

`sampsize = 2,000,000`

`xb1 = 3`

`xb2 = 3`

`omega1=1`

`omega2=1`

`rho12=.75`

`T1 = -1, 1, 2, 3, 4, 100`

`T2 = -1, 1, 2, 3, 4, 100`

SUR with DC Outcomes: Example for a Specific Sampling Design (cont'd)

Table 1: Cell Relative Frequencies for the Simulated Sample of Size 2M

Relevant Cell Interval		Y ₂				
		Y ₂ ≤ 1	1 < Y ₂ ≤ 2	2 < Y ₂ ≤ 3	3 < Y ₂ ≤ 4	4 < Y ₂
Y ₁	Y ₁ ≤ 1	0.06814	0.06347	0.03803	0.01599	0.00772
	1 < Y ₁ ≤ 2	0.06343	0.08322	0.06469	0.03453	0.02188
	2 < Y ₁ ≤ 3	0.03801	0.06450	0.06194	0.03965	0.03281
	3 < Y ₁ ≤ 4	0.01594	0.03441	0.03962	0.03012	0.03197
	4 < Y ₁	0.00772	0.02190	0.03283	0.03192	0.05558

Sum of all elements in the table = 1

`sum(relcumfreqY1Y2) = 1`

SUR with DC Outcomes: Example for a Specific Sampling Design (cont'd)

Table 2: True Cell Probabilities Based on the CDF

Relevant Cell Interval		Y ₂				
		Y ₂ ≤ 1	1 < Y ₂ ≤ 2	2 < Y ₂ ≤ 3	3 < Y ₂ ≤ 4	4 < Y ₂
Y ₁	Y ₁ ≤ 1	0.06795	0.06330	0.03800	0.01601	0.00773
	1 < Y ₁ ≤ 2	0.06330	0.08325	0.06468	0.03429	0.02203
	2 < Y ₁ ≤ 3	0.03800	0.06468	0.06181	0.03956	0.03288
	3 < Y ₁ ≤ 4	0.01601	0.03429	0.03956	0.03007	0.03217
	4 < Y ₁	0.00773	0.02203	0.03288	0.03217	0.05560

Sum of all elements in the table = 1

`sum(tabcellprobs) = 1`

Maximum absolute difference between elements of Table 1 and Table 2

`max(abscelldiff)`
`.0002526852`