Biases in Maximum Simulated Likelihood Estimation of Bivariate Models

MAKSAT JUMAMYRADOV[†] AND MURAT K. MUNKIN[‡]

[†]*Department of Economics, CMC, University of South Florida, Tampa, FL 33620-5500, U.S.A.*
E-mail: maksat@usf.edu, mmunkin@usf.edu

Summary This paper finds that the maximum simulated likelihood (MSL) estimator produces substantial biases when applied to the bivariate normal distribution. A specification of the random parameter bivariate normal model is considered, in which a direct comparison between the MSL and maximum likelihood (ML) estimators is feasible. The analysis shows that MSL produces biased results for the correlation parameter. This paper also finds that the MSL estimator is biased for the bivariate Poisson-lognormal model, developed by Munkin and Trivedi (1999). A simulation study is conducted, which shows that MSL leads to serious inferential biases, especially large when variance parameters in the true data generating process are small. The MSL estimator produces biases in the estimated marginal effects, conditional means and probabilities of count outcomes.

Keywords: *Maximum Simulated Likelihood; Simulation Biases; Bivariate Models.*

1. INTRODUCTION

The maximum simulated likelihood (MSL) inferential framework has become a widely accepted tool utilised by applied economists to estimate models in which closed-form solutions are not available, and therefore, maximum likelihood (ML) estimation is not feasible. However, MSL has its limitations, some of which are known and some are open topics of research. We consider a random coefficient specification of the bivariate normal model, for which the ML approach is computationally feasible, and a direct comparison between MSL and ML is possible. We find that MSL fails to produce unbiased estimates for a range of parameter values in the true data generating process. However, these problems are not restricted to the bivariate normal models alone. This paper examines performance of the MSL estimator in the context of the bivariate Poisson-lognormal (BPLN) model, a continuous mixture built on the bivariate normal distribution, introduced by Munkin and Trivedi (1999), who use it to study the demand for health care by the US elderly. We conduct a simulation study and find problems with the MSL estimator producing results leading to serious inferential biases.

The MSL estimator was introduced by Lerman and Manski (1981). Its formal definition can be found in many studies, including Gourieroux and Monfort (1993), Lee (1995), Hajivassiliou and Ruud (1994). Applications of the MSL estimator include, for example, studies that numerically approximate multi-dimensional integrals (Geweke et al., 1997), models with unobserved heterogeneity (Filippini and Greene, 2016), models in which latent variables are integrated out (McFadden and Train, 2000). The MSL estimator is biased when the number of simulations is finite, which is always the case in practice. The bias could be partially corrected by adjusting the objective function using second-order Taylor expansion as discussed, for example, in Gourieroux et al. (1996). We use this technique of first-order bias correction in our analysis, following Munkin and Trivedi

(1999). They also utilise antithetic sampling, however, instead we choose Halton sequences (Halton, 1960) since we find them to be more efficient, producing more accurate approximations with fewer draws.

Our simulation experiments show that the size of the MSL bias depends on the values of the correlation and variance parameters and the number of simulation draws, both in the bivariate normal and BPLN models. In the case of the BPLN distribution, we find various biases for different parameter values in the estimated conditional means, marginal effects and probabilities of outcomes. The rest of the paper is organized as follows. Section 2 defines the MSL estimator, outlines its asymptotic properties and briefly describes the simulation bias reduction techniques used in this paper. Section 3 presents our specifications of the bivariate normal and BPLN models, presents estimation results and discusses them. Finally, Section 4 concludes.

2. MAXIMUM SIMULATED LIKELIHOOD ESTIMATOR

The existence of a closed form density $f(y_i|x_i, \theta)$ for dependent random variable y_i conditional on the vector of independent variables x_i and parameter vector θ is necessary for the definition of the ML estimator as

$$\hat{\theta}_N = \arg\max_{\theta} \sum_{i=1}^{N} \log f(y_i|x_i, \theta)$$

where (y_i, x_i) is a set of independent observations for $i = 1, ..., N$. However, when econometric models are specified conditional on latent variables which cannot be integrated out in closed-forms, the MSL estimator is a possible alternative. We define the MSL estimator and its properties following Gourieroux and Monfort (1990) and Gourieroux et al. (1996). Define $\tilde{f}(y_i, x_i, u, \theta)$ to be an unbiased simulator of the conditional density $f(y_i|x_i, \theta)$ if

$$f(y_i|x_i, \theta) = E_u[f(y_i, x_i, u, \theta)|y_i, x_i]$$

where the distribution of u is known and independent of y_i and x_i. Then the MSL estimator of θ is defined as

$$\hat{\theta}_{SN} = \arg\max_{\theta} \sum_{i=1}^{N} \log \left[\frac{1}{S} \sum_{s=1}^{S} \tilde{f}(y_i, x_i, u_i^s, \theta) \right],$$

where u_i^s $(s = 1, ..., S)$ are drawn independently for each individual i from the distribution of u_i. It is obtained by replacing the intractable conditional p.d.f. $f(y_i|x_i, \theta)$ with its unbiased approximation based on the simulator $\tilde{f}(y_i, x_i, u_i, \theta)$. The following propositions state asymptotic properties of the MSL estimator, which are relevant to our discussion.

PROPOSITION 2.1. *(a) The MSL estimator is inconsistent if S is fixed and N tends to infinity. (b) The MSL estimator is consistent if S and N tend to infinity. (c) If $S, N \to \infty$ and $N/S \to 0$ or $\sqrt{N}/S \to 0$, the MSL estimator will be asymptotically equivalent to the ML estimator such that*

$$\sqrt{N}(\hat{\theta}_{SN} - \theta_0) \xrightarrow{d} N\{0, I^{-1}(\theta_0)\}$$

where $I^{-1}(\theta_0)$ is the information matrix and θ_0 is the true value of θ.

Therefore, as long as $S, N \to \infty$, $\hat{\theta}_{SN}$ is consistent, however, it is asymptotically efficient only when $N/S \to 0$ or $\sqrt{N}/S \to 0$.

PROPOSITION 2.2. *When S tends to infinity, the asymptotic bias is equivalent to*

$$E_0\hat{\theta}_{SN} - \theta_0 \sim \frac{1}{S} I^{-1}(\theta_0) E_0[a(y, x, \theta_0)]$$

where E_0 is the expectation taken with respect to the true distribution of y_i given x_i and θ_0, and

$$a(y, x, \theta_0) = \frac{E_u(\frac{\partial \tilde{f}}{\partial \theta}) Var_u(\tilde{f})}{(E_u\tilde{f})^3} - \frac{cov_u(\frac{\partial \tilde{f}}{\partial \theta}, \tilde{f})}{(E_u\tilde{f})^2}.$$

Denote $\tilde{f}(y_i, x_i, u_i^s, \theta) = \tilde{f}_s$. The inconsistency of the MSL estimator for fixed S comes from the fact that $\log(\frac{1}{S}\sum_{s=1}^{S} \tilde{f}_s)$ is a biased simulator of $\log f$, even though $\tilde{f}_s$ is an unbiased simulator of f. Taking second-order Taylor expansion of $\log(\frac{1}{S}\sum_{s=1}^{S} \tilde{f}_s)$ around $\log f$ gives the following approximation

$$\log\left(\frac{1}{S}\sum_{s=1}^{S} \tilde{f}_s\right) \approx \log f + \frac{1}{f}\left(\frac{1}{S}\sum_{s=1}^{S} \tilde{f}_s - f\right) - \frac{1}{2f^2}\left(\frac{1}{S}\sum_{s=1}^{S} \tilde{f}_s - f\right)^2.$$

Since $E_u\tilde{f} = f$, then

$$E_u \log\left(\frac{1}{S}\sum_{s=1}^{S} \tilde{f}_s\right) \approx \log f - \frac{1}{2f^2} E_u\left(\frac{1}{S}\sum_{s=1}^{S} \tilde{f}_s - f\right)^2.$$

Then in general

$$\log\left(\frac{1}{S}\sum_{s=1}^{S} \tilde{f}_s\right) + \frac{1}{2S^2 f^2}\left(\sum_{s=1}^{S} \tilde{f}_s - f\right)^2 \tag{2.1}$$

is a simulator of $\log f$ with a smaller bias than $\log(\frac{1}{S}\sum_{s=1}^{S} \tilde{f})$. Simplifying the last term in (2.1) the MSL estimator with first-order asymptotic bias correction can be defined as

$$\tilde{\theta}_{SN} = \arg\max_{\theta} \left[\sum_{i=1}^{N} \log\left[\frac{1}{S}\sum_{s=1}^{S} \tilde{f}(y_i, x_i, u_i^s, \theta)\right] \right.$$
$$\left. + \frac{1}{2}\sum_{i=1}^{N} \frac{\sum_{s=1}^{S}\left[\tilde{f}(y_i, x_i, u_i^s, \theta) - \frac{1}{S}\sum_{s=1}^{S} \tilde{f}(y_i, x_i, u_i^s, \theta)\right]^2}{\left[\sum_{s=1}^{S} \tilde{f}(y_i, x_i, u_i^s, \theta)\right]^2}\right]$$

Next, we briefly describe how Halton sequences are constructed. For a more detailed presentation of this topic see, for example, Bhat (2003), Greene (2012) and Train (2009). Consider an example with the prime number 3. A Halton sequence is generated by dividing the unit interval (0, 1) into three equal segments. The breakpoints of this first division cycle (1/3, 2/3) are going to be the first two numbers in the sequence. Then, these three segments are divided into thirds again and the breakpoints of this second division cycle (1/9, 4/9, 7/9, 2/9, 5/9, 8/9) are added to the sequence. The sequence now becomes 1/3, 2/3, 1/9, 4/9, 7/9, 2/9, 5/9, 8/9. This process continues until the desired number of points are obtained. Each division cycle in the Halton sequence covers the areas on the unit interval that were not covered by the previous cycle, providing better coverage. Additionally, this filling-in property of the Halton sequence induces a negative correlation over observations, which reduces the simulation errors.

In the case of any prime number r any integer g can be expressed in terms of the base r as

$$g = \sum_{i=0}^{m} b_i r^i$$

where $0 \leq b_i \leq r-1$ and $r^m \leq g \leq r^{m+1}$. Then the Halton sequence is generated such that for the integer g the corresponding Halton value is

$$H(g) = \sum_{i=0}^{m} b_i r^{-i-1}$$

For instance, 8^{th} element of the Halton sequence values for base prime number 3 has $b_0 = 2$ and $b_1 = 2$. Then $H_3(8) = 2 \times 3^{-1} + 2 \times 3^{-2} = 8/9$.

Thus, a Halton sequence covers the support of the uniform distribution, segment $[0, 1]$, in a balanced way. Then pseudo-random numbers for any continuous distribution can be generated using the inverse c.d.f. method. For instance, if the desired density is standard normal, the sequence 1/3, 2/3, 1/9, 4/9, 7/9, 2/9, 5/9, 8/9 can be transformed as $\Phi^{-1}(1/3) = -0.43$, $\Phi^{-1}(2/3) = 0.43$, $\Phi^{-1}(1/9) = -1.2$, $\Phi^{-1}(4/9) = -0.14$, $\Phi^{-1}(7/9) = 0.76$, $\Phi^{-1}(2/9) = -0.76$, $\Phi^{-1}(5/9) = 0.14$, $\Phi^{-1}(8/9) = 1.2$.

The list of studies which use the MSL estimator and Halton sequences include, for example, Greene (2003), Hochguertel and Ohlsson (2009), Plum (2019). The number of Halton draws for each observation that these studies utilise vary. Greene (2003) suggests a new approach to the estimation of the normal-gamma stochastic frontier model based on the MSL estimation using 200 Halton draws for each observation. Hochguertel and Ohlsson (2009) use 100 Halton draws in their study of parents' *inter vivos*, "a transfer or gift made during one's lifetime", to their children by using censored regression models with nested multilevel error components. Plum (2019) study British labor market dynamics of unemployed by estimating a random-effects probit model using 75 Halton draws for each observation in the MSL framework. All these studies do not use first-order bias correction.

Bhat (2001) shows that 100 Halton draws provide better approximation results than 1000 pseudo-random draws for the mixed logit model. Train (2000) and Munizaga and Alvarez-Daziano (2001) have similar findings. Palma et al. (2020) report that around 93% of over 150 papers indexed in the Research Papers in Economics (RePEc) produced during 2008-2018 use fewer than 1000 Halton draws in their estimation of the mixed logit model. Furthermore, 72% of these papers use fewer than 500 Halton draws and 40% use fewer than 250 Halton draws. To our best knowledge, most papers that use Halton sequences as a variance reduction technique, do not use first-order bias correction.

3. APPLICATION OF THE MSL ESTIMATOR

This section investigates performance of the MSL estimator when it is applied to two bivariate distribution: normal and Poisson. First, we present a random parameter specification of the continuous bivariate normal model. This model can be a convenient framework in which MSL and ML approaches are directly compared. The bivariate Poisson-lognormal model does not have a closed-form, however, we analyse properties of the MSL estimator with Monte Carlo simulation experiments similarly to how it was done by Munkin and Trivedi (1999) but extending the range of possible parameter values.

3.1. Bivariate Normal Model

This section specifies a bivariate normal model with random slope coefficients which can be analysed both by ML and MSL procedures. Assume that continuous random variables y_{1i} and y_{2i} are bivariate normally distributed, conditional on independent explanatory variables x_{1i} and x_{2i}, for each observation $i = 1, ..., N$, such that

$$y_{1i} = \alpha_1 + x_{1i}\beta_{1i} + \varepsilon_{1i}$$
$$y_{2i} = \alpha_2 + x_{2i}\beta_{2i} + \varepsilon_{2i}$$

and ε_{1i} and ε_{2i} are bivariate normal errors

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma\right)$$

with

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

The conditional means of y_{1i} and y_{2i} include intercepts α_1, α_2, and slope coefficients β_{1i}, β_{2i} assumed to be random parameters, written as

$$\beta_{1i} = \beta_1 + u_{1i}$$
$$\beta_{2i} = \beta_2 + u_{2i}$$

where β_1 and β_2 are fixed, u_{1i} and u_{2i} are independent standard normal $N(0, 1)$. The correlation between y_{1i} and y_{2i} in this model is generated by the error terms ε_{1i} and ε_{2i}.

The specified model can be analysed either directly by the ML procedure, integrating out random variables u_{1i} and u_{2i} analytically, or using MSL, integrating them out numerically. In the ML case the bivariate normal model can be written as

$$y_{1i} = \alpha_1 + x_{1i}\beta_1 + v_{1i}$$
$$y_{2i} = \alpha_2 + x_{2i}\beta_2 + v_{2i}$$

where the error terms

$$v_{1i} = x_{1i}u_{1i} + \varepsilon_{1i}$$
$$v_{2i} = x_{2i}u_{2i} + \varepsilon_{2i}$$

are bivariate normal

$$\begin{pmatrix} v_{1i} \\ v_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_i\right)$$

where

$$\Sigma_i = \begin{pmatrix} x_{1i}^2 + \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & x_{2i}^2 + \sigma_2^2 \end{pmatrix}.$$

Then denoting $y_i' = (y_{1i}, y_{2i})$, $x_i' = (x_{1i}, x_{2i})$, $\alpha' = (\alpha_1, \alpha_2)$, $\beta' = (\beta_1, \beta_2)$ the log-likelihood function is

$$LL = \sum_{i=1}^{N} \log f(y_i|x_i)$$

where

$$f(y_i|x_i) = (2\pi)^{-1} |\Sigma_i|^{-1/2} \exp[-0.5(y_i - \alpha - x_i\beta)'\Sigma_i^{-1}(y_i - \alpha - x_i\beta)].$$

The MSL estimator of the parameters in the model can be defined when an approximation of $f(y_{1i}, y_{2i}|x_{1i}, x_{2i})$ is obtained from the joint conditional distribution of y_{1i} and y_{2i}

$$f(y_{1i}, y_{2i}|x_{1i}, x_{2i}, u_{1i}, u_{2i}) = (2\pi)^{-1} |\Sigma_i|^{-1/2} \times \exp\left[-0.5(y_i - \alpha - x_i(\beta + u_i))'\Sigma^{-1}(y_i - \alpha - x_i(\beta + u_i))\right]$$

integrating out random errors u_{1i} and u_{2i} numerically. Then the simulated log-likelihood function is

$$SLL = \sum_{i=1}^{N} \log \left[\frac{1}{S} \sum_{s=1}^{S} f(y_{1i}, y_{2i}|x_{1i}, x_{2i}, u_{1i}^s, u_{2i}^s) \right],$$

where u_{1i}^s and u_{2i}^s $(s = 1, ..., S)$ are drawn independently from the standard normal distribution.

We generate data according to the bivariate normal model with the following specification of the data generation process: $\alpha_1 = \alpha_2 = -1$, $\beta_1 = \beta_2 = 1$, $x_{1i} \sim N(0,1)$, $x_{2i} \sim N(0,1)$, $(\varepsilon_{1i}, \varepsilon_{2i}) \sim N\{(0,0), (\sigma_1^2, \rho\sigma_1\sigma_2, \sigma_2^2)\}$. That defines two normally distributed dependent variables y_{1i} and y_{2i} for each observation i $(i = 1, ..., N)$, where $N = 1000$. We choose three different values for σ_1 and σ_2 (0.25, 0.5 and 1), and two different values for ρ (0.96 and -0.96). Since we choose equal values for σ_1 and σ_2 in the data generation process, for the purpose of brevity in tables and figures presented below, we write $\sigma = (0.25, 0.5, 1)$ instead of $\sigma_1 = \sigma_2 = (0.25, 0.5, 1)$.

The number of Halton draws is chosen to be consistent with the levels used by leading MSL applications. In our simulation experiments we use $S = (250, 500, 1000)$ of Halton draws for generated 1000 observations. We repeat this experiment 200 times, generating a new data set in each case and calculating the MSL and ML estimates. Tables 1 and 2 present results based on the averages taken with respect to these 200 replications. The tables also report the standard errors of the estimated sample means.

Table 1 presents MSL results dividing them into two parts, with $\rho = 0.96$ (upper portion) and $\rho = -0.96$ (lower portion). The estimated parameters $\alpha_1, \alpha_2, \beta_1, \beta_2$ produce no sizable biases across different values of parameters σ_1 and σ_2. However, there are significant biases in the estimates of σ_1, σ_2 and especially ρ, which vary substantially with respect to true parameter values and number of draws. First, consider the case when $S = 250$. Specifically, when $\sigma = 1$, all parameter estimates are close to the true values irrespective of the value of ρ. However, when $\sigma = 0.5$, the estimates of the correlation parameter ρ are 0.894 (0.004) and -0.886 (0.004) when the true values are 0.96 and -0.96, respectively. The numbers in parenthesis are the standard errors of the sample means calculated based on 200 replications. The null hypothesis that parameter ρ is equal to the true values $\{0.96, -0.96\}$ is overwhelmingly rejected since the true values are separated from the estimated $\{0.894, -0.886\}$ by more than 15 standard errors. When the true values of the standard deviations become smaller, $\sigma = 0.25$, the estimated biases become even bigger. The estimates of ρ, 0.601 (0.011) and -0.606 (0.009), have much bigger standard errors but still well separated from the true values by at least 30 standard errors, which would reject the corresponding null hypotheses that $\rho = 0.96$ and $\rho = -0.96$. In addition, the null hypotheses, that σ_1 and σ_2 are equal to the true values 0.25, are rejected as well.

As expected, increasing the number of draws reduces the biases. Estimated σ_1, σ_2 and ρ get closer to their true values when $S = 500$ and $S = 1000$. However, even when $S = 1000$, the biases in ρ are significant. Particularly in the case when $\sigma = 0.25$, the estimates of ρ $\{0.816, -0.809\}$ are separated from their true values $\{0.96, -0.96\}$ by around 16 standard errors.

Table 1. Bivariate normal MSL

	$\rho = 0.96$								
	250 Halton draws			500 Halton draws			1000 Halton draws		
	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 1$
$\hat{\alpha}_1$	-0.999	-1.002	-1.001	-0.998	-1.001	-0.996	-1.000	-0.996	-1.002
	(0.001)	(0.002)	(0.003)	(0.001)	(0.002)	(0.003)	(0.001)	(0.002)	(0.003)
$\hat{\alpha}_2$	-0.999	-0.998	-1.002	-1.000	-0.999	-0.996	-0.997	-0.997	-1.005
	(0.001)	(0.002)	(0.003)	(0.001)	(0.002)	(0.003)	(0.001)	(0.002)	(0.003)
$\hat{\beta}_1$	0.996	0.999	0.998	0.996	1.002	1.003	0.997	0.994	0.996
	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.003)
$\hat{\beta}_2$	0.993	0.995	1.000	0.998	0.997	0.999	1.004	0.997	1.003
	(0.006)	(0.005)	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.004)
$\hat{\sigma}_1$	0.270	0.807	1.003	0.262	0.503	0.998	0.255	0.501	0.997
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
$\hat{\sigma}_2$	0.277	0.508	1.000	0.263	0.500	1.000	0.259	0.499	0.998
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
$\hat{\rho}$	0.601	0.894	0.955	0.719	0.920	0.960	0.816	0.950	0.963
	(0.011)	(0.004)	(0.001)	(0.009)	(0.004)	(0.001)	(0.009)	(0.003)	(0.001)
	ρ=-0.96								
	250 Halton draws			500 Halton draws			1000 Halton draws		
	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 1$
$\hat{\alpha}_1$	-1.000	-0.999	-1.005	-1.001	-1.000	-1.000	-0.999	-0.997	-1.000
	(0.001)	(0.002)	(0.003)	(0.001)	(0.002)	(0.003)	(0.001)	(0.002)	(0.003)
$\hat{\alpha}_2$	-1.001	-1.000	-0.997	-1.001	-0.999	-1.004	-1.001	-1.002	-0.998
	(0.001)	(0.002)	(0.003)	(0.001)	(0.002)	(0.003)	(0.001)	(0.002)	(0.003)
$\hat{\beta}_1$	0.999	1.002	0.998	1.007	0.992	1.003	0.994	0.991	1.001
	(0.006)	(0.005)	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.004)
$\hat{\beta}_2$	1.004	1.001	0.997	0.995	0.995	1.009	1.005	1.008	1.002
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.004)
$\hat{\sigma}_1$	0.275	0.51	0.005	0.263	0.502	1.000	0.257	0.500	0.999
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
$\hat{\sigma}_2$	0.274	0.505	1.001	0.265	0.503	0.998	0.255	0.498	1.001
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$\hat{\rho}$	-0.606	-0.886	-0.955	-0.720	-0.919	-0.958	-0.809	-0.950	-0.962
	(0.009)	(0.004)	(0.001)	(0.010)	(0.004)	(0.001)	(0.010)	(0.003)	(0.001)

Table 2 presents ML estimates of the model parameters for the same sets of true values. The ML estimator produces unbiased estimates, which make it clear that the biases presented in Table 1 are caused by computational difficulties of the MSL estimator in finding the maximum of the objective function. These biases are most likely caused by simulation imprecision in approximating the bivariate normal density. The fact, that the standard errors of ρ in the MSL case are 3 to 4 times bigger than those in the ML estimator for some parameter values, is an indicator that the additional variance caused by simulation is considerable.

Table 2. Bivariate normal ML

	ρ=0.96				ρ=-0.96		
	σ=0.25	σ=0.5	σ=1		σ=0.25	σ=0.5	σ=1
$\hat{\alpha}_1$	-0.999	-0.997	-0.997	$\hat{\alpha}_1$	-1.000	-1.001	-1.002
	(0.001)	(0.002)	(0.003)		(0.001)	(0.002)	(0.003)
$\hat{\alpha}_2$	-1.001	-0.998	-0.999	$\hat{\alpha}_2$	-1.002	-0.998	-1.000
	(0.001)	(0.002)	(0.003)		(0.001)	(0.002)	(0.003)
$\hat{\beta}_1$	0.999	0.999	0.997	$\hat{\beta}_1$	1.000	1.004	1.003
	(0.003)	(0.003)	(0.003)		(0.002)	(0.003)	(0.003)
$\hat{\beta}_2$	1.001	1.003	0.999	$\hat{\beta}_2$	0.999	0.995	1.002
	(0.003)	(0.003)	(0.003)		(0.003)	(0.003)	(0.003)
$\hat{\sigma}_1$	0.246	0.498	0.995	$\hat{\sigma}_1$	0.248	0.499	1.001
	(0.003)	(0.001)	(0.001)		(0.001)	(0.002)	(0.002)
$\hat{\sigma}_2$	0.246	0.501	0.999	$\hat{\sigma}_2$	0.248	0.499	0.999
	(0.003)	(0.001)	(0.001)		(0.001)	(0.001)	(0.002)
$\hat{\rho}$	0.960	0.959	0.96	$\hat{\rho}$	-0.956	-0.962	-0.961
	(0.003)	(0.001)	(0.001)		(0.003)	(0.001)	(0.001)

Figure 1 plots estimated $\hat{\rho}$ against their true values, where $\hat{\rho}$ is defined as averages of 200 MSL estimates of ρ obtained based on 200 samples generated for the same set of true values and estimated with $S = 250$ draws (Figures for the cases when $S = 500$ and $S = 1000$ are available in the online Appendix). The true values of ρ change from -0.96 to 0.96 in increments of 0.04. It can be seen from the figure that when $\sigma = 1$ the estimated $\hat{\rho}$ is close to the true values. However, when $\sigma = 0.5$ the correlation parameter is biased closer to the boundaries. The biases become stronger and spread over the entire support when $\sigma = 0.25$, such that the correlation parameter is underestimated for positive ρ and overestimated for negative.

Next, we fix ρ at three different values 0.25, 0.5 and 0.9 and estimate $\hat{\sigma}_1$, as averages of 200 MSL estimates of σ_1, calculated for the range of true values changing from 0.15 to 0.5 in increments of 0.05. When the true value of σ_1 is smaller than 0.15, the maximization procedure fails to converge. Figure 2 shows that $\hat{\sigma}_1$ is overestimated when σ_1 is small in value.

As can be seen from these results the MSL estimates can be seriously biased in the case of the bivariate normal model even if the number of Halton draws is 1000. Other than few exceptions, such as Deb and Trivedi (2006), who use 2000 Halton draws, the conventional literature is set on not exceeding 250 draws. The biases we obtained become even stronger if first-order bias correction is not applied, which most applications of MSL do not do.

3.2. Bivariate Poisson

Initially introduced by Aitchison and Ho (1989), the BPLN model accounts for overdispersion and allows for a full correlation structure. There are several specifications of bivariate count models in the literature, estimation of which are based on simulation-based methods. Munkin and Trivedi (1999) use the MSL estimator for the BPLN model to study the demand for health care by the US elderly. Hellström (2006) uses a truncated version of the same model to study households' number of leisure trips and a number of overnight stays. Chib and Winkelmann (2001) apply Bayesian Markov chain Monte Carlo (MCMC) methods to estimate the Poisson-lognormal model with six count dependent variables to study the demand for health care by the US elderly.

Next we investigate properties of the MSL estimator for the bivariate Poisson-lognormal model, analysed by Munkin and Trivedi (1999), however, they did it for just a limited range of parameter values. Assume that two count variables y_{1i} and y_{2i} $(i = 1, ..., N)$ are Poisson distributed

$$y_{1i}|x_{1i}, \varepsilon_{1i} \sim P(\mu_{1i})$$
$$y_{2i}|x_{2i}, \varepsilon_{2i} \sim P(\mu_{2i})$$

with conditional means μ_{1i} and μ_{2i} specified as

$$\mu_{1i} = \exp(\alpha_1 + x_{1i}\beta_1 + \varepsilon_{1i})$$
$$\mu_{2i} = \exp(\alpha_2 + x_{2i}\beta_2 + \varepsilon_{2i})$$

where x_{1i} and x_{2i} are sets of explanatory variables, and ε_{1i} and ε_{2i} are errors representing unobserved heterogeneity, assumed to be jointly normally distributed as

$$(\varepsilon_{1i}, \varepsilon_{2i}) \sim N((0, 0), \Sigma),$$

where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

and $0 \leqslant |\rho| \leqslant 1$. In derivation of the MSL estimator we follow Munkin and Trivedi (1999), making a small change which will potentially reduce the simulation bias in the choice of the importance sampling function. Based on the Cholesky decomposition of matrix Σ defined as $\Sigma = LL'$, where L is the lower triangular matrix

$$L = \begin{pmatrix} \sigma_1 & 0 \\ \sigma_2\rho & \sigma_2\sqrt{1-\rho^2} \end{pmatrix},$$

bivariate normal ε_1 and ε_2 can be written as

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = L \begin{pmatrix} u_1 \\ u_2 \end{pmatrix},$$

in terms of u_1 and u_2, two independent standard normal random variables. Then the simulated log-likelihood function can be defined as

$$\sum_{i=1}^{N} \log \left[\frac{1}{S} \sum_{s=1}^{S} f(y_{1i}|x_{1i}, \sigma_1 u_{1i}^s) f(y_{2i}|x_{2i}, \sigma_2 u_{1i}^s \rho + \sigma_2 u_{2i}^s \sqrt{1-\rho^2}) \right] \tag{3.2}$$

However, in this formulation, $\partial E(y_{2i}|x_{2i}, u_{1i}, u_{2i})/\partial\rho$ is not bounded at $\rho = \pm 1$ since

$$\frac{\partial E(y_{2i}|x_{2i}, u_{1i}, u_{2i})}{\partial\rho} = \exp(\alpha_2 + x_{2i}\beta_2 + \sigma_2 u_{1i}\rho + \sigma_2 u_{2i}\sqrt{1-\rho^2}) \left(\sigma_2 u_{1i} - \frac{\sigma_2 \rho u_{2i}}{\sqrt{1-\rho^2}} \right) \tag{3.3}$$

In addition, the first and second derivatives of the objective function (3.2) with respect to ρ are not bounded at $\rho = \pm 1$. Nevertheless, we test this objective function and find that the corresponding MSL estimator performs well for parameter values well separated from $\rho = \pm 1$.

To make sure that (3.3) is bounded for all values, Munkin and Trivedi (1999) introduce an importance sampling function

$$h(v_{1i}, v_{2i}) = \frac{1}{2\pi} exp\left[-0.5\left(v_{1i}^2 + v_{2i}^2\right)\right]$$

and write

$$\int_{v_{2i}} \int_{v_{1i}} f(y_{1i}|x_{1i}, \sigma_1 v_{1i}) f(y_{2i}|x_{2i}, \sigma_2 v_{2i}) \frac{g(v_{1i}, v_{2i})}{h(v_{1i}, v_{2i})} h(v_{1i}, v_{2i}) dv_{1i} dv_{2i}. \tag{3.4}$$

However, there is another potential problem with this specification. The chosen important sampling function draws more than 99% of the values inside the three standard deviation interval $[-3, 3]$. However, in order to approximate the integrals well, the drawn values should come from $[-3\sigma_1, 3\sigma_1]$ and $[-3\sigma_2, 3\sigma_2]$ intervals. If the true values of σ_1 and σ_2 are substantially different from 1 then MSL estimation becomes a challenge. Munkin and Trivedi (1999) do not address this problem because the true values for variance parameters in their simulation experiments are chosen to be $\sigma = 1$. However, we reparametrise the model as follows. Assume that $\varepsilon_1 = \sigma_1 v_1$ and $\varepsilon_2 = \sigma_2 v_2$. Then

$$\begin{aligned} SLL &= \sum_{i=1}^{N} \log\left[\frac{1}{S}\sum_{s=1}^{S} f(y_{1i}|x_{1i}, \sigma_1 v_{1i}^s) f(y_{2i}|x_{2i}, \sigma_2 v_{2i}^s) \frac{g(v_{1i}^s, v_{2i}^s)}{h(v_{1i}^s, v_{2i}^s)}\right] \qquad (3.5) \\ &= \sum_{i=1}^{N} \log\left[\frac{1}{S}\sum_{s=1}^{S} \frac{\exp(-\mu_{1i}^s)(\mu_{1i}^s)^{y_{1i}}}{y_{1i}!} \frac{\exp(-\mu_{2i}^s)(\mu_{2i}^s)^{y_{2i}}}{y_{2i}!} \frac{1}{\sqrt{1-\rho^2}}\right. \\ &\qquad \left.\times \exp\left(-\frac{(v_{1i}^s)^2 - 2\rho v_{1i}^s v_{2i}^s + (v_{2i}^s)^2}{2(1-\rho^2)} + \frac{(v_{1i}^s)^2 + (v_{2i}^s)^2}{2}\right)\right] \end{aligned}$$

where $\mu_{1is} = \exp(x_{1i}\beta_1 + \sigma_1 v_{1i}^s)$ and $\mu_{2is} = \exp(x_{2i}\beta_2 + \sigma_2 v_{2i}^s)$. Even though $1/\sqrt{1-\rho^2}$ is not bounded at $\rho = \pm 1$ the whole expression is, since factor

$$\exp\left[-\frac{1}{2(1-\rho^2)}\left(v_1^2 - 2\rho v_1 v_2 + v_2^2\right)\right]$$

converges to zero much faster than does $1/\sqrt{1-\rho^2}$ to infinity. The first and second derivatives of the SLL function with respect to ρ are bounded for the same reason.

In our simulation experiments, we test two specifications of the objective function: (3.2) and (3.5). In addition, we apply first-order bias correction to these objective function and compare results of the total four optimization procedures. In most of the cases the simulated log-likelihood (3.5) with first-order bias correction gives results with smallest biases which we report below.

Next we generate $N = 1000$ independent observations from the bivariate Poisson distribution with the following specifications of the data generation process: $\alpha_1 = \alpha_2 = -1$, $\beta_1 = \beta_2 = 1$, $x_{1i} \sim N(0, 1/16)$, $x_{2i} \sim N(0, 1/16)$ $(i = 1, ..., N)$. Unobserved heterogeneity variables are generated as $(\varepsilon_{1i}, \varepsilon_{2i}) \sim N\{(0, 0), (\sigma_1^2, \rho\sigma_1\sigma_2, \sigma_2^2)\}$, in which three different values for σ_1 and σ_2 $\{0.25, 0.5, 1\}$ are chosen, and two for ρ $\{0.4, -0.4\}$. These two values of ρ are consistent with sizable biases in estimation. We produce three sets of MSL estimates different with respect to the number of Halton draws utilised: $S = 250$, 500 and 1000.

Table 3 presents MSL results for the BPLN model based on 200 replications when the true values of the correlation parameter are chosen to be $\rho = 0.4$ (upper portion) and $\rho = -0.4$ (lower portion). The estimated values for parameters α_1, α_2, β_1,β_2 produce no sizable biases across all specifications. However, when $S = 250$ and $\sigma = 0.5$, the estimates of the correlation parameter ρ, 0.591 (0.031) and −0.588 (0.031), are more than 6 standard errors larger in magnitude than the true values 0.4 and −0.4, respectively. The biases become even larger for the case $\sigma = 0.25$, when the estimated ρ, 0.886 (0.021) and −0.854 (0.023), are separated from the true values by about 20 standard errors. Furthermore, there is a significant bias in the estimates of σ_1 and σ_2 when $\sigma = 0.25$ and $\rho = 0.4$. True and estimated values are separated from each other by more than 5 standard errors.

When we increase the number of Halton draws to $S = 500$, the results are mixed. For instance, when $\sigma = 0.25$, the estimated correlation parameter, 0.772 (0.036), gets closer to its true value $\rho = 0.4$ but the errors become large. However, when $\rho = -0.4$, the estimated correlation −0.900 (0.031) is further away from the true value. The biases in σ_1 and σ_1 become much larger.

Increasing the number of draws to $S = 1000$ moves estimated ρ closer to the true values although the errors become larger. However, for all three values of σ, there are still significant biases, especially when $\sigma = 0.25$. Moreover, it is clear that the biases in estimated σ_1 and σ_1 persist and in the case when $\rho = 0.4$ they become larger.

The biases in estimates of ρ and how the true values of σ_1 and σ_2 affect them could be better understood from Figure 3. We choose three values of the correlation parameter in the data generating process: $\rho = 0.25$, 0.5 and 0.9. Figure 3 plots estimated values of ρ against different true values of $\sigma = \sigma_1 = \sigma_2$ taken from the interval $[0.05, 1.25]$. Black lines represent the true values of ρ, and colour lines represent estimated ρ. It can be seen that the biases in ρ become visible as soon as $\sigma_1 < 0.75$. However, they reach their maxima for smallest σ.

Figure 4 plots $\hat{\rho}$, averages of the MSL estimates, against the true values for the range from −0.96 to 0.96 in increments of 0.04 for the bivariate Poisson model with $S = 250$ (Figures for $S = 500$ and $S = 1000$ are provided in the online Appendix). These calculations, once again, are based on 200 replications. It can be seen from the figure that biases in estimation of ρ are greatest when $\sigma = 0.25$ and when ρ is around ± 0.4.

Figure 5 shows how much the biases in estimating the variance parameter depend on the value of correlation ρ. This is done for the range of variance values from the interval $[0.05, 1.25]$ taken with increments of 0.05. For the high value $\rho = 0.9$, there is no bias for small σ_1, however, when $\rho = 0.5$ and $\rho = 0.25$ the bias is considerable, becoming larger for smaller σ_1 values. For larger variance values, there is a bias in the estimated σ_1 for all three lines, the biases being highest when $\rho = 0.9$ and smallest when $\rho = 0.25$.

Thus, the MSL estimator produces biased estimates for parameters σ_1, σ_2, ρ, and the size of the bias depends on parameters' true values. When first-order bias correction is not applied, the biases are even stronger. There are three important observations which can be taken from these results. First, the bias in $\rho$is larger when standard deviations are smaller. Second, estimates of σ_1 and σ_2 are also biased depending on the value of ρ. Finally, biases in these parameters decrease but do not disappear when we increase the number of Halton draws.

Another observation is that the biases in ρ in the bivariate normal model are monotonically increasing, being worst at the boundaries, and the biases in σ_1 and σ_2 follow similar patterns, irrespective of the values of ρ. However, in the bivariate Poisson-lognormal model the bias in ρ does not follow a monotonic pattern. Moreover, Figure 5 shows that

Table 3. Bivariate Poisson MSL

	$\rho = 0.4$								
	250 Halton draws			500 Halton draws			1000 Halton draws		
	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 1$
$\hat{\alpha}_1$	-1.028	-1.003	-1.004	-1.020	-0.991	-1.013	-1.006	-0.991	-1.005
	(0.006)	(0.005)	(0.006)	(0.006)	(0.007)	(0.008)	(0.007)	(0.008)	(0.007)
$\hat{\alpha}_2$	-1.035	-0.998	-1.011	-1.027	-0.983	-0.998	-1.003	-0.993	-0.993
	(0.006)	(0.005)	(0.006)	(0.005)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
$\hat{\beta}_1$	0.984	0.979	1.017	1.001	0.998	1.011	0.971	0.994	0.997
	(0.019)	(0.016)	(0.016)	(0.015)	(0.016)	(0.018)	(0.016)	(0.017)	(0.017)
$\hat{\beta}_2$	1.002	1.014	1.000	1.005	1.017	1.025	1.006	1.006	1.001
	(0.020)	(0.016)	(0.017)	(0.017)	(0.016)	(0.016)	(0.015)	(0.015)	(0.017)
$\hat{\sigma}_1$	0.311	0.481	1.025	0.182	0.464	1.024	0.193	0.472	1.007
	(0.010)	(0.010)	(0.006)	(0.017)	(0.010)	(0.008)	(0.017)	(0.010)	(0.007)
$\hat{\sigma}_2$	0.308	0.493	1.030	0.229	0.465	1.008	0.174	0.461	0.998
	(0.011)	(0.009)	(0.006)	(0.016)	(0.009)	(0.008)	(0.016)	(0.010)	(0.007)
$\hat{\rho}$	0.886	0.591	0.393	0.772	0.629	0.390	0.764	0.537	0.411
	(0.021)	(0.031)	(0.007)	(0.036)	(0.030)	(0.007)	(0.031)	(0.028)	(0.007)
	ρ=-0.4								
	250 Halton draws			500 Halton draws			1000 Halton draws		
	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 1$
$\hat{\alpha}_1$	-1.021	-0.996	-1.025	-1.019	-1.001	-1.012	-1.020	-0.996	-1.008
	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)	(0.008)	(0.005)	(0.007)	(0.008)
$\hat{\alpha}_2$	-1.025	-1.004	-1.016	-1.013	-0.990	-1.006	-1.014	-1.003	-0.992
	(0.006)	(0.005)	(0.006)	(0.007)	(0.008)	(0.008)	(0.005)	(0.007)	(0.008)
$\hat{\beta}_1$	0.995	0.994	0.991	0.999	0.988	0.977	1.034	1.000	1.018
	(0.016)	(0.014)	(0.018)	(0.018)	(0.016)	(0.019)	(0.016)	(0.016)	(0.017)
$\hat{\beta}_2$	1.016	1.006	1.012	0.957	0.988	0.995	0.996	1.005	0.990
	(0.018)	(0.016)	(0.017)	(0.018)	(0.016)	(0.018)	(0.015)	(0.016)	(0.019)
$\hat{\sigma}_1$	0.284	0.465	1.032	0.130	0.487	1.011	0.192	0.476	0.999
	(0.010)	(0.010)	(0.007)	(0.020)	(0.009)	(0.008)	(0.016)	(0.010)	(0.008)
$\hat{\sigma}_2$	0.275	0.481	1.027	0.157	0.466	1.007	0.170	0.480	0.992
	(0.011)	(0.009)	(0.006)	(0.019)	(0.011)	(0.008)	(0.017)	(0.012)	(0.008)
$\hat{\rho}$	-0.854	-0.588	-0.367	-0.900	-0.565	-0.385	-0.783	-0.532	-0.404
	(0.023)	(0.031)	(0.007)	(0.021)	(0.029)	(0.008)	(0.033)	(0.031)	(0.008)

the biases in σ_1 and σ_2 have different patterns depending on the value of ρ. It is likely that the additional functional complexity of the bivariate Poisson-lognormal model over bivariate normal affects performance of the MSL estimator.

Next we analyse how these biases can potentially affect inferential statements based on the corresponding parameter estimates. Integrating out u_{ji} from the mean gives

$$\lambda_{ji} = E(y_{ji}|x_{ji}) = exp(\alpha_j + x_{ji}\beta_j + \sigma_j^2/2),\ \ j = 1, 2.$$

Then the marginal effects for dependent variable y_{ji} in the bivariate Poisson model can be calculated as

$$\frac{\partial E(y_{ji}|x_{ji})}{\partial x_{ji}} = \lambda_{ji}\beta_j.$$

Define true marginal effects as

$$TrueME_{y_j} = \frac{1}{N}\sum_{i=1}^{N} exp(\alpha_j + x_{ji}\beta_j + \sigma_j^2/2)\beta_j,$$

calculated at the true parameter values, and define estimated marginal effects as

$$Est.ME_{y_j} = \frac{1}{N}\sum_{i=1}^{N} exp(\hat{\alpha}_j + x_{ji}\hat{\beta}_j + \hat{\sigma}_j^2/2)\hat{\beta}_j,$$

calculated at the parameter estimates. Figure 6 plots true and estimated marginal effects, averaged over 200 replications, for the bivariate Poisson model against true values of ρ taken from the $[-0.96, 0.96]$ interval. Black lines represent the true marginal effects. Colour lines represent estimated marginal effects for the specified values of σ_1 and σ_2. It can be seen from the figure that when $\sigma = 0.25$ and $\sigma = 0.5$ estimated marginal effects are close to their true values. However, when $\sigma = 1$, the estimated marginal effects are overestimated when there is a high negative and positive correlation, which is consistent with σ being overestimated near boundaries in Figure 5 for larger ρ. Particularly, the estimated marginal effect is 18 per cent higher than the true marginal effect when $\rho = 0.96$.

Figure 7 presents true and estimated conditional means, $E[y_1|x_1] = \lambda_1$ and $\hat{E}[y_1|x_1] = \hat{\lambda}_1$, for the bivariate Poisson model estimated as

$$\frac{1}{N}\sum_{i=1}^{N} exp(\alpha_j + x_{ji}\beta_j + \sigma_j^2/2),$$

and

$$\frac{1}{N}\sum_{i=1}^{N} exp(\hat{\alpha}_j + x_{ji}\hat{\beta}_j + \hat{\sigma}_j^2/2),$$

respectively. Black lines represent λ_1, and colour lines represent $\hat{\lambda}_1$. Estimated conditional means are close to their true values when $\sigma = 0.25$ and $\sigma = 0.5$. However, when $\sigma = 1$, $\hat{\lambda}_1$ is overestimated at the boundaries of ρ values. Particularly, $\hat{\lambda}_1$ is nearly 12 per cent higher than λ_1 when there is high positive correlation.

Figures 8, 9 and 10 plot actual and predicted cell frequencies for the dependent variable y_1, when $y_1 = 0$, $y_1 = 1$ and $y_1 \geqslant 2$, and denoted as $Pr(y_1 = 0)$, $Pr(y_1 = 1)$ and $Pr(y_1 \geqslant 2)$. These are calculated for the same range of ρ values between -0.96 and 0.96 in increments of 0.04. In all of these three figures, estimated cell frequencies are close to the actual values when $\sigma = 0.25$ and $\sigma = 0.5$. However, they deviate from the actual values when $\sigma = 1$. Specifically, $Pr(y_1 = 0)$ is overestimated when there is a high negative correlation and underestimated when there is a high positive correlation, $Pr(y_1 = 1)$ is underestimated at both ends, and $Pr(y_1 \geqslant 2)$ is overestimated by nearly 10 per cent when there is a high positive correlation.

4. CONCLUSION

Even though the MSL estimation approach has become a standard practical tool in dealing with distributions in which closed-form solutions do not exists, its performance in multivariate models is an open research question. This paper compares the ML and MSL estimators for a specification of the random parameter bivariate normal model and finds that MSL results in significant biases, while ML produces a reliable output. This problem is not restricted to the bivariate normal distribution alone, but it is present in the Poisson-lognormal continuous mixture distribution. The MSL estimator in that case produces significant biases the sizes of which depend on the true values of variance and correlation parameters. The biases are present in estimates of parameters, marginal effects, conditional means and probabilities of count outcomes. Our study uses the MSL estimator in combination with two simulation bias reduction techniques: Halton sequences and first-order bias correction. We find that the estimated biases become considerably larger if first-order bias correction is not utilised, as do most applications of MSL.

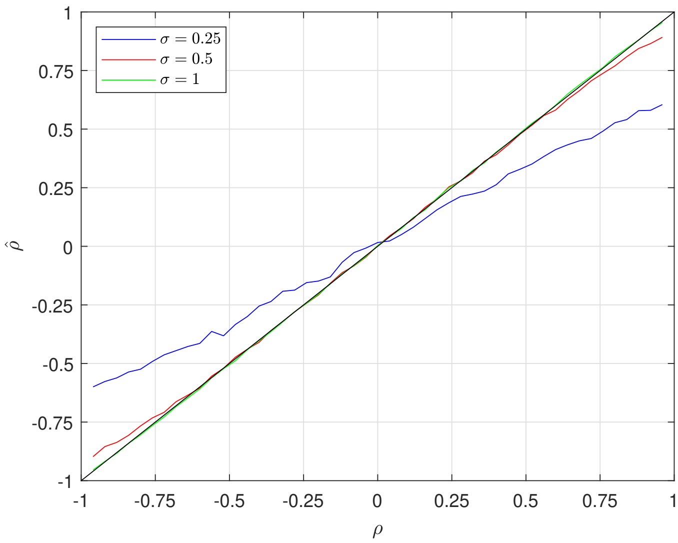

Figure 1. Plots of $\hat{\rho}$ Against the True Values: Bivariate Normal Model ($S = 250$).

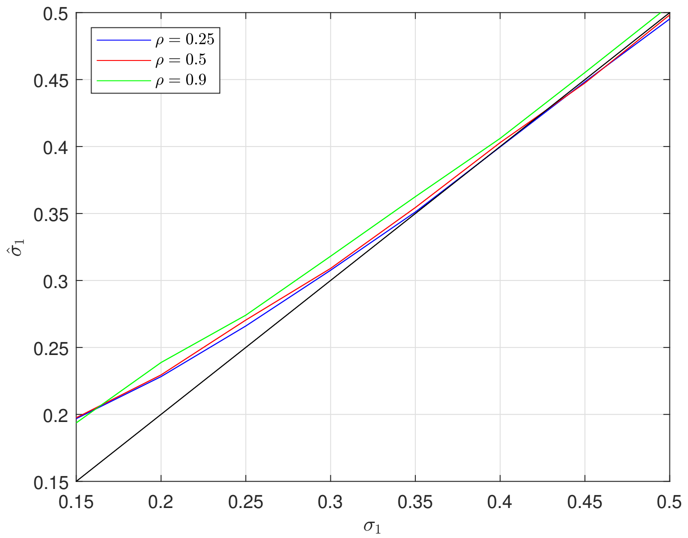

Figure 2. Plots of $\hat{\sigma}_1$ Against the True Values: Bivariate Normal Model ($S = 250$).

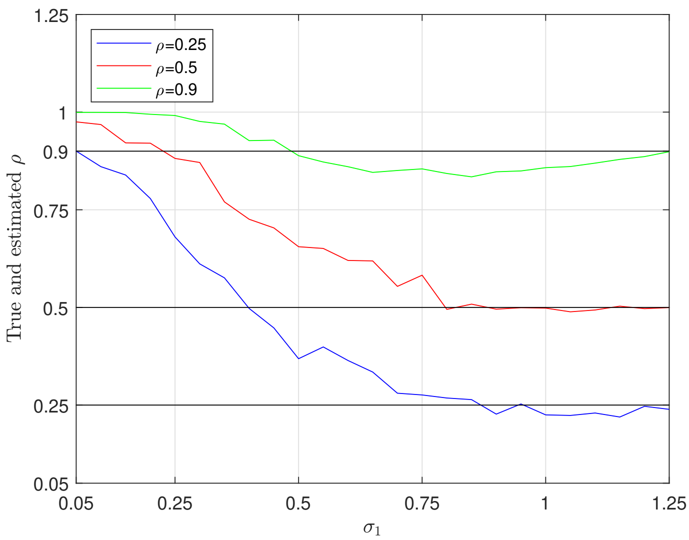

Figure 3. Plot of True and Estimated ρ against σ: Bivariate Poisson Model ($S = 250$)

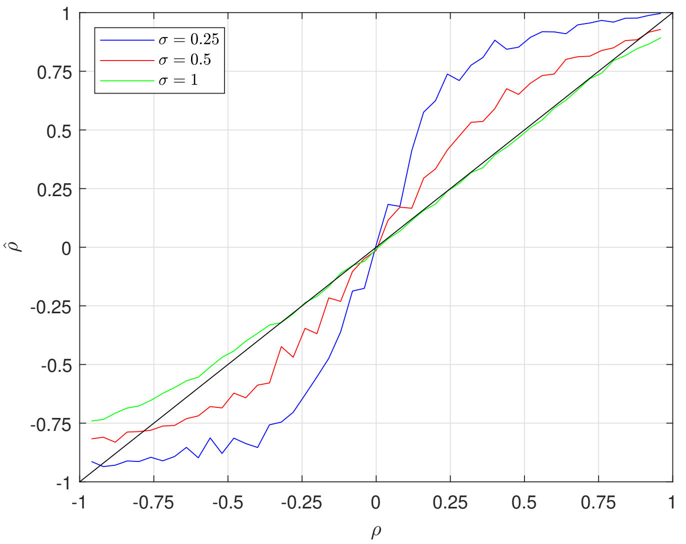

Figure 4. Plots of $\hat{\rho}$ Against the True Values: Bivariate Poisson Model ($S = 250$).

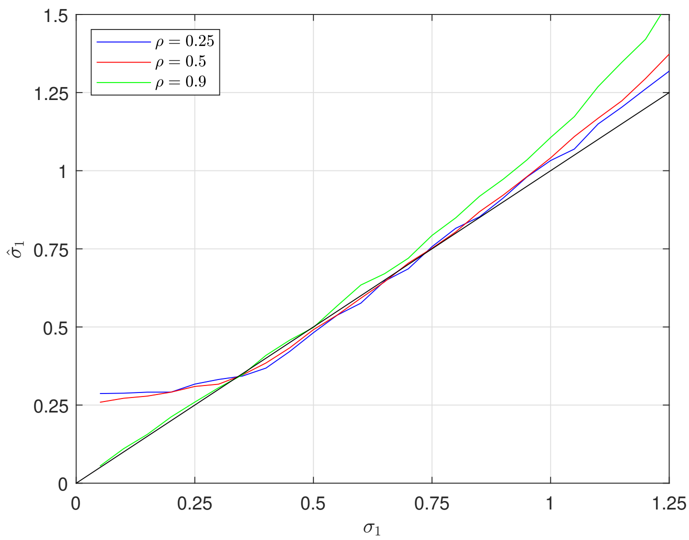

Figure 5. Plot of $\hat{\sigma}_1$ Against the True Values: Bivariate Poisson Model ($S = 250$).

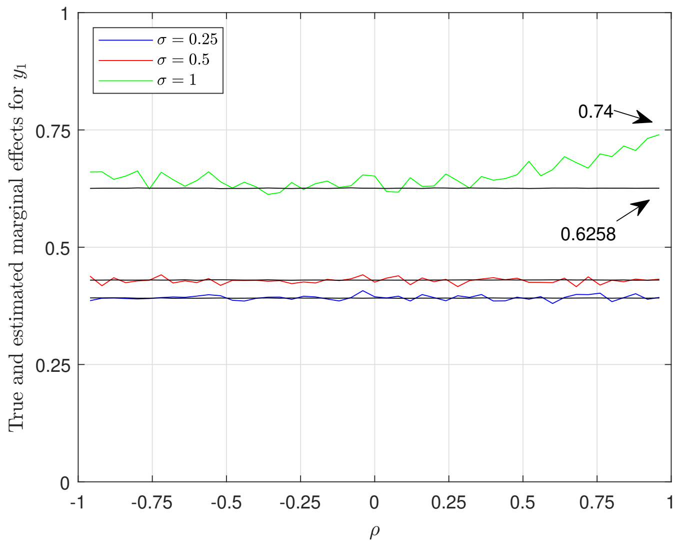

Figure 6. Plots of True and Estimated $\frac{\partial E[y_1|x_1]}{\partial x_1}$: Bivariate Poisson Model ($S = 250$).

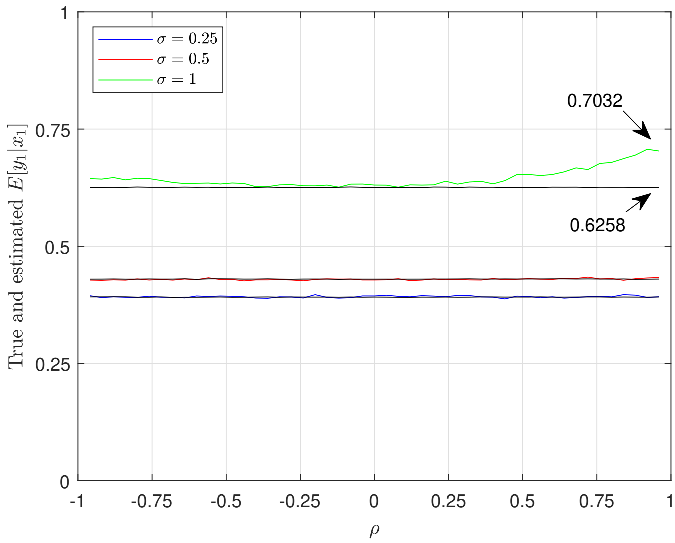

Figure 7. Plots of True and Estimated $E[y_1|x_1]$: Bivariate Poisson Model ($S = 250$).

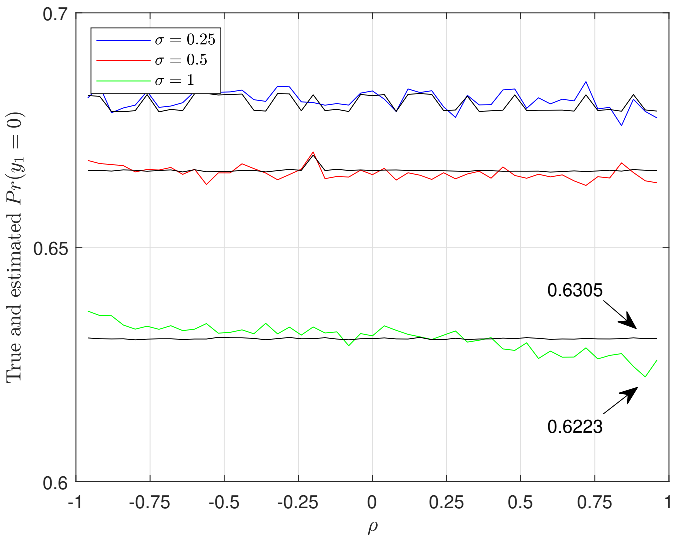

Figure 8. Plots of True and Estimated $Pr(y_1 = 0)$: Bivariate Poisson Model ($S = 250$).

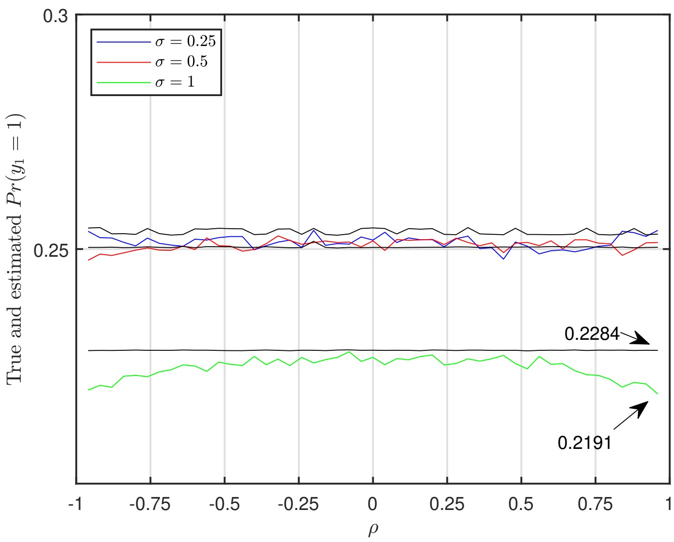

Figure 9. Plots of True and Estimated $Pr(y_1 = 1)$: Bivariate Poisson Model ($S = 250$).

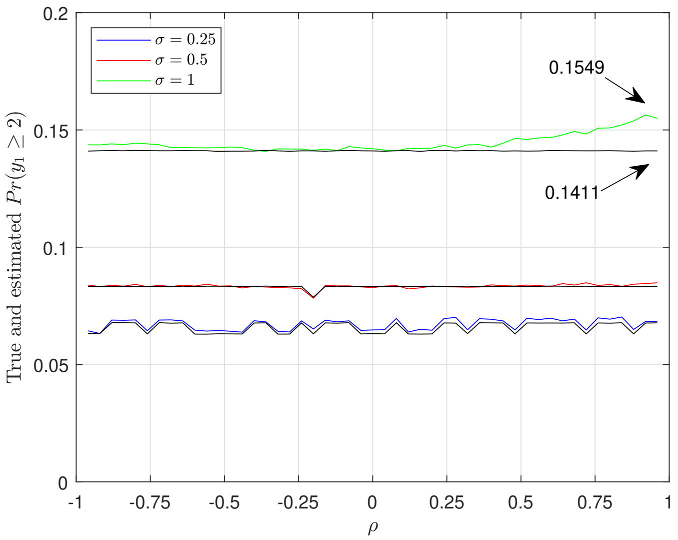

Figure 10. Plot of True and Estimated $Pr(y_1 \geqslant 2)$: Bivariate Poisson Model ($S = 250$).

5. APPENDIX

The online supplement includes figures in estimation of the bivariate normal and bivariate Poisson-lognormal models for the cases when $S = 500$ and $S = 1000$. The data generation processes in making these figures are the same as described in the paper manuscript. All the figures and tables presented in the paper manuscript, and this online supplement are made using MATLAB codes through cluster computing. We utilised 8GB memory and 12 workers on the cluster computer. Time spent to get the results for one of the lines in any of the figures (e.g. $S = 500$, $\sigma = 0.25$) varied from as low as 90 minutes to 7 hours depending on the model.

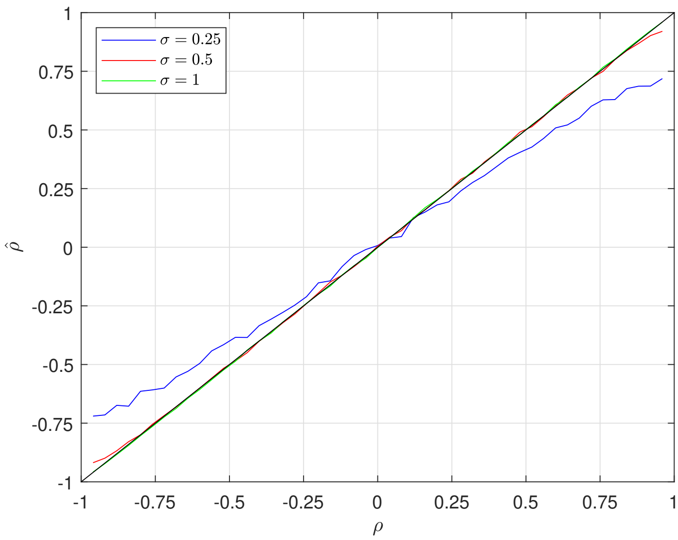

Figure 11. Plots of $\hat{\rho}$ Against the True Values: Bivariate Normal Model ($S = 500$).

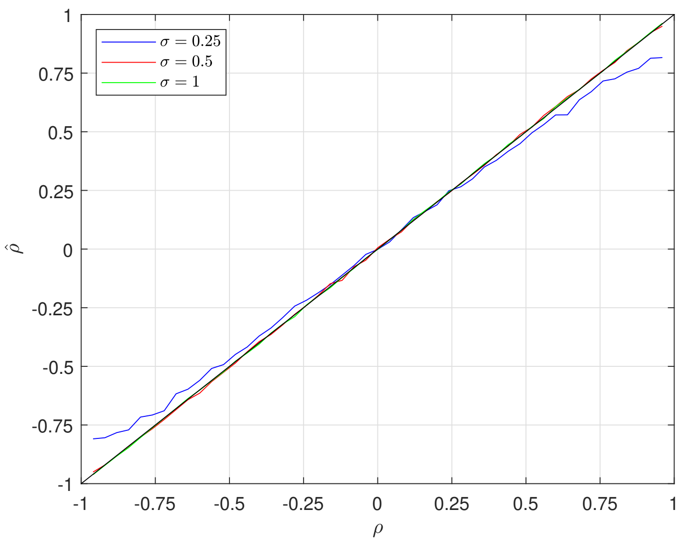

Figure 12. Plots of $\hat{\rho}$ Against the True Values: Bivariate Normal Model ($S = 1000$).

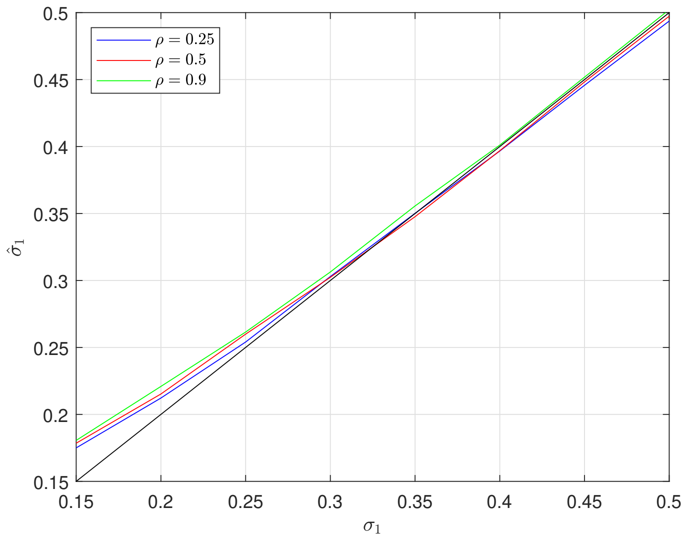

Figure 13. Plots of $\hat{\sigma}_1$ Against the True Values: Bivariate Normal Model ($S = 500$).

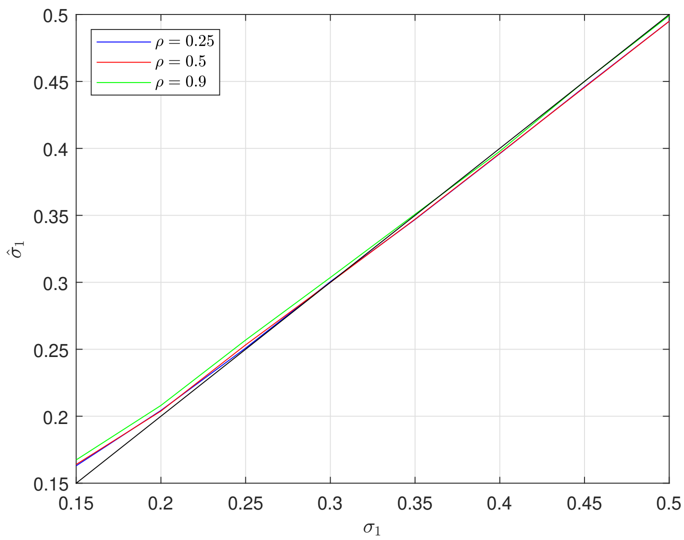

Figure 14. Plots of $\hat{\sigma}_1$ Against the True Values: Bivariate Normal Model ($S = 1000$).

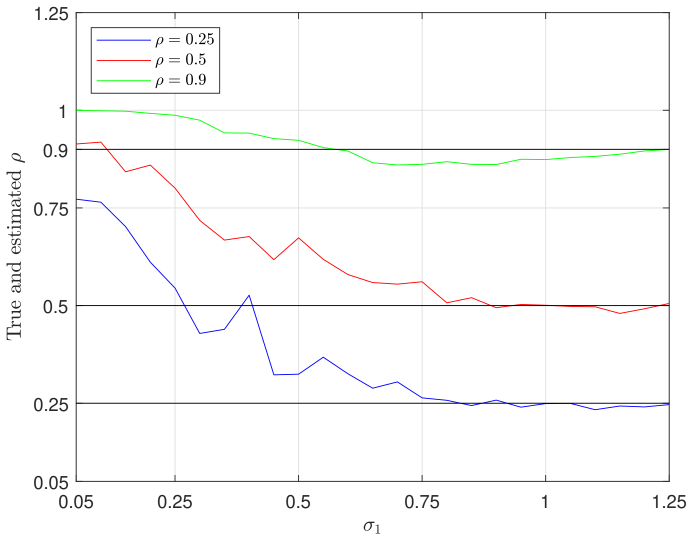

Figure 15. Plots of True and Estimated ρ Against σ_1: Bivariate Poisson Model ($S = 500$).

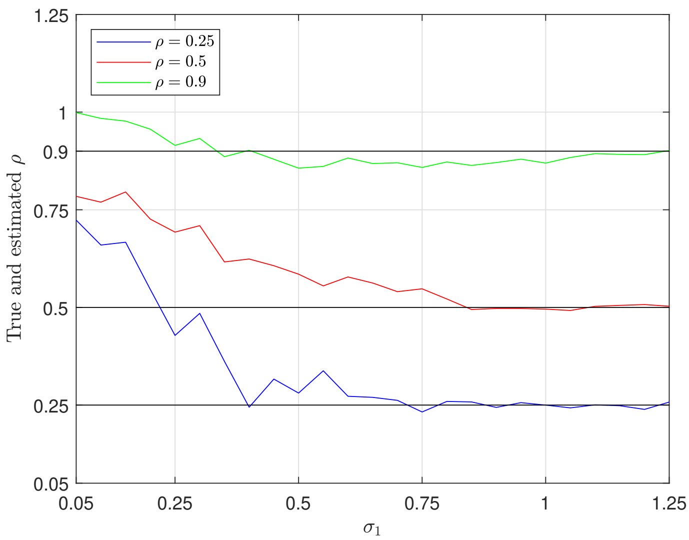

Figure 16. Plots of True and Estimated ρ Against σ_1: Bivariate Poisson Model ($S = 1000$).

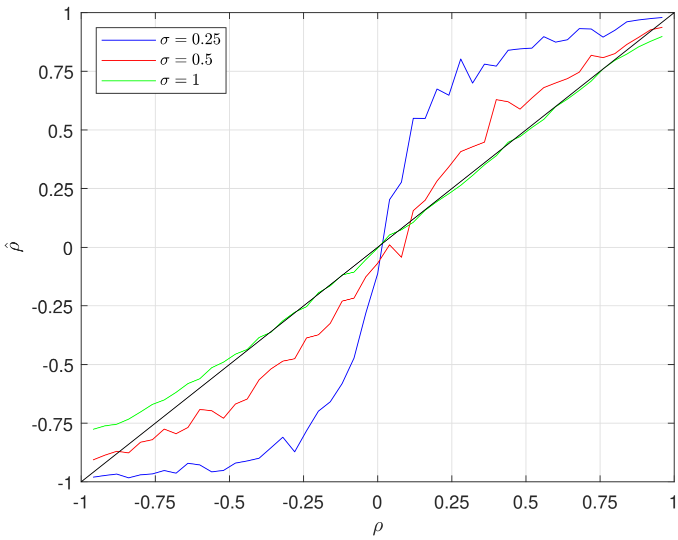

Figure 17. Plot of $\hat{\rho}$ Against the True Values: Bivariate Poisson Model ($S = 500$).

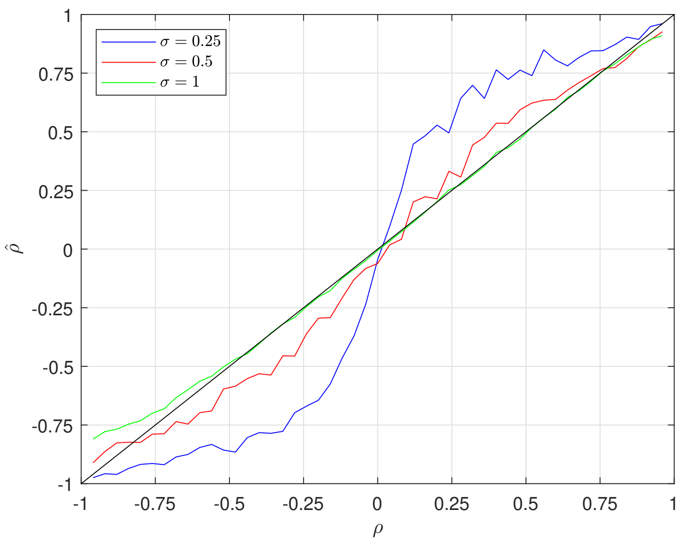

Figure 18. Plot of $\hat{\rho}$ Against the True Values: Bivariate Poisson Model ($S = 1000$).

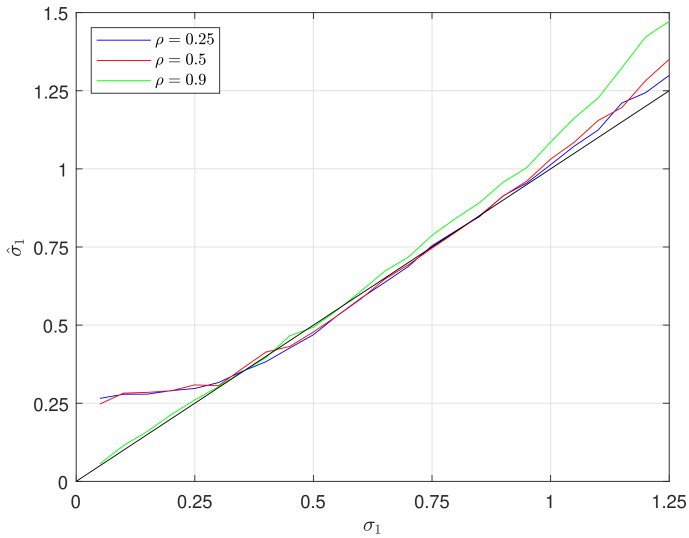

Figure 19. Plot of $\hat{\sigma}_1$ Against the True Values: Bivariate Poisson Model ($S = 500$).

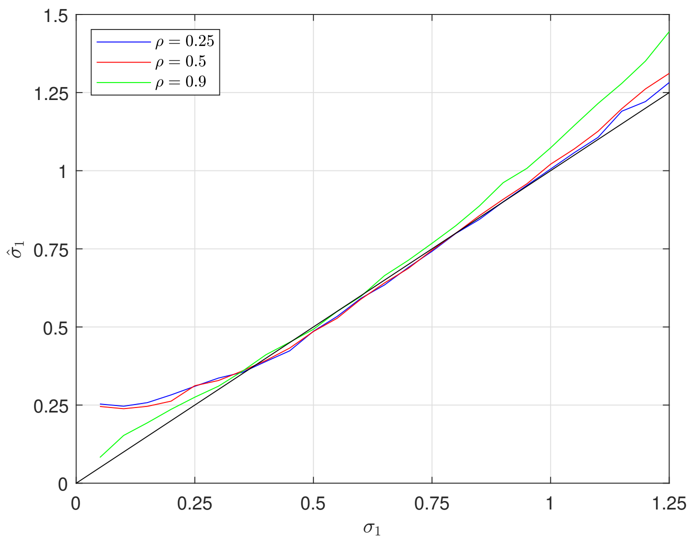

Figure 20. Plot of $\hat{\sigma}_1$ Against the True Values: Bivariate Poisson Model ($S = 1000$).

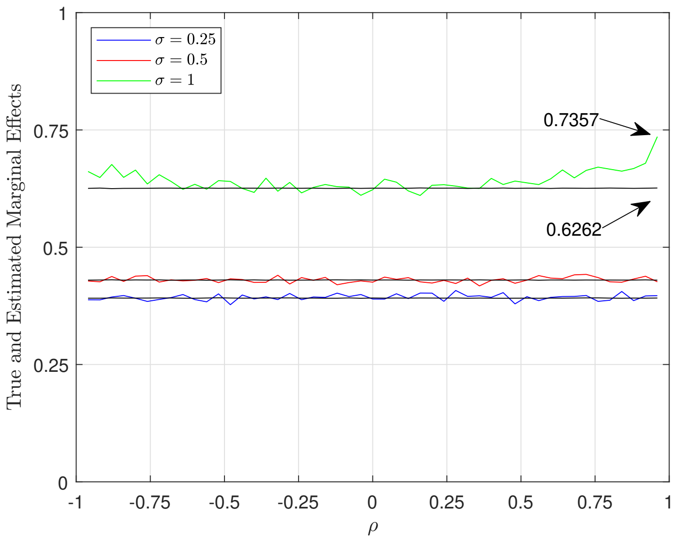

Figure 21. Plots of True and Estimated $\frac{\partial E[y_1|x_1]}{\partial x_1}$: Bivariate Poisson Model (S=500).

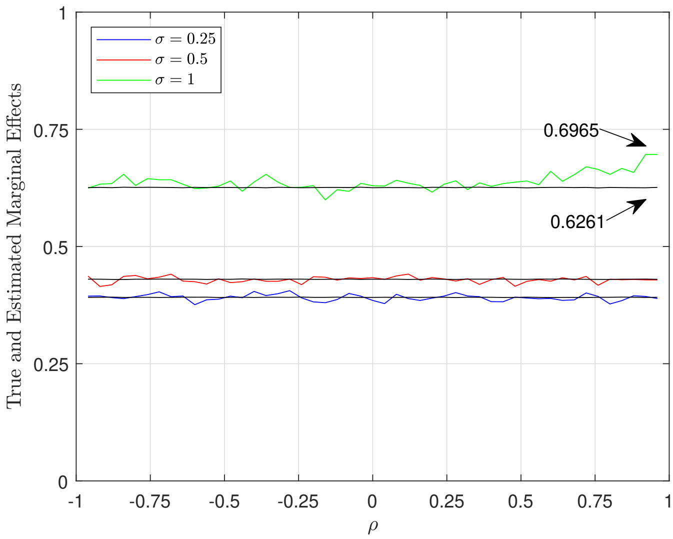

Figure 22. Plots of True and Estimated $\frac{\partial E[y_1|x_1]}{\partial x_1}$: Bivariate Poisson Model (S=1000).

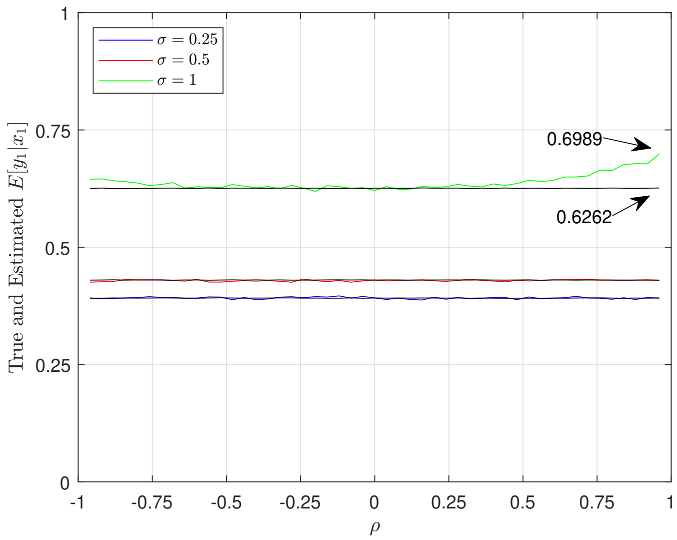

Figure 23. Plots of True and Estimated $E[y_1|x_1]$: Bivariate Poisson Model (S=500).

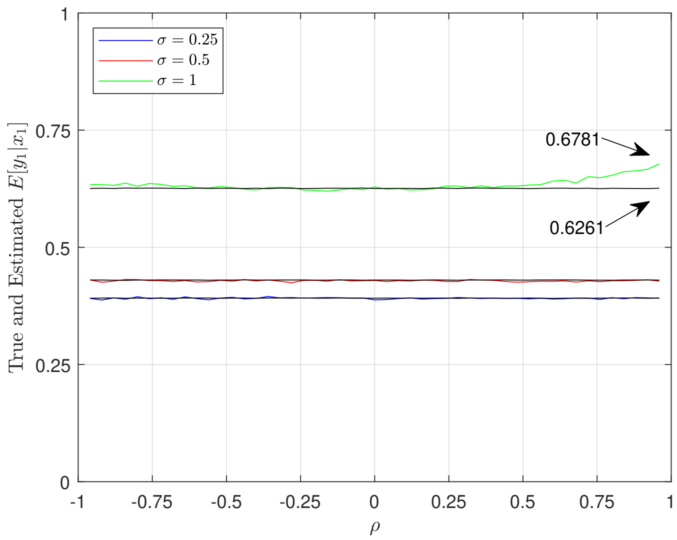

Figure 24. Plots of True and Estimated $E[y_1|x_1]$: Bivariate Poisson Model (S=1000).

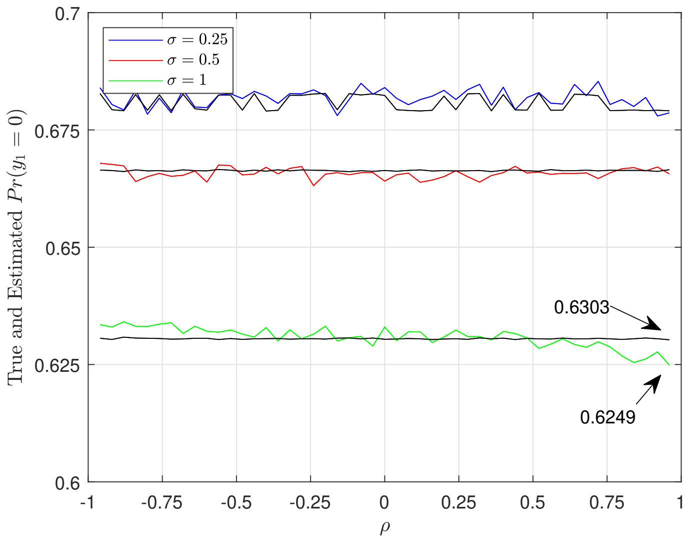

Figure 25. Plots of True and Estimated $Pr(y_1 = 0)$: Bivariate Poisson Model (S=500).

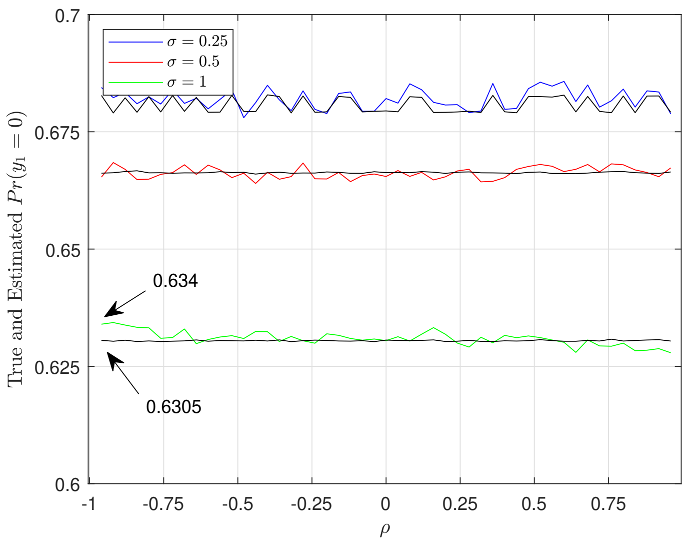

Figure 26. Plots of True and Estimated $Pr(y_1 = 0)$: Bivariate Poisson Model (S=1000).

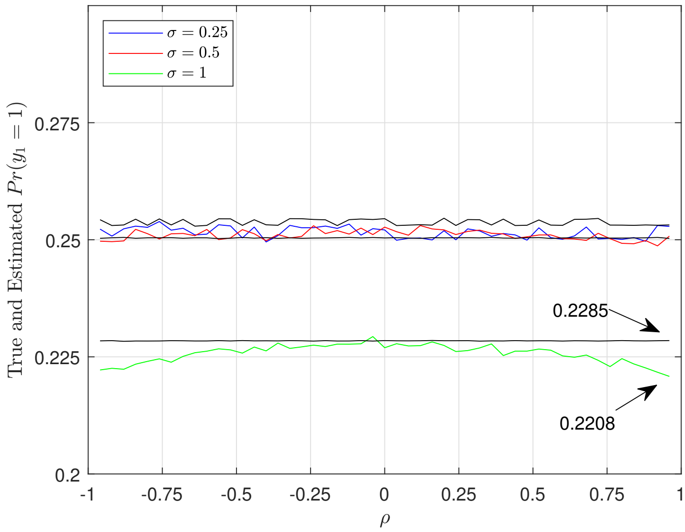

Figure 27. Plots of True and Estimated $Pr(y_1 = 1)$: Bivariate Poisson Model (S=500).

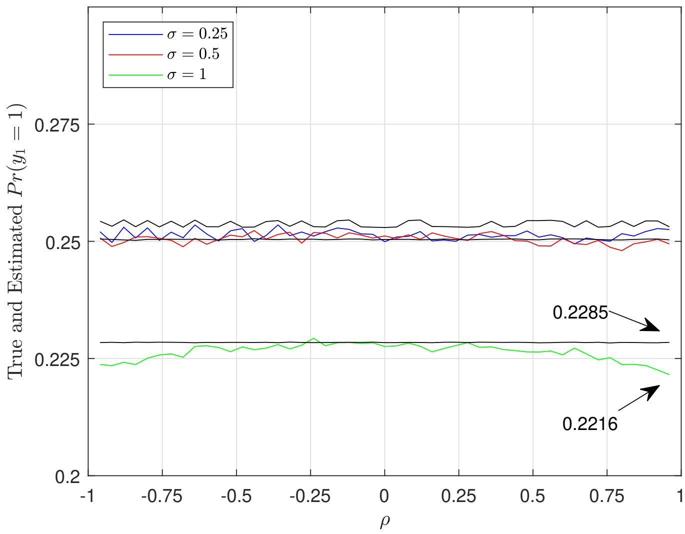

Figure 28. Plots of True and Estimated $Pr(y_1 = 1)$: Bivariate Poisson Model (S=1000).

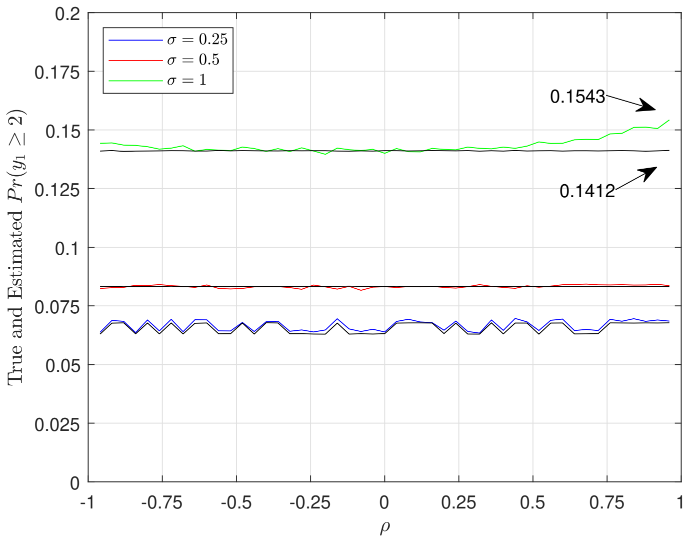

Figure 29. Plots of True and Estimated $Pr(y_1 \geq 2)$: Bivariate Poisson Model (S=500).

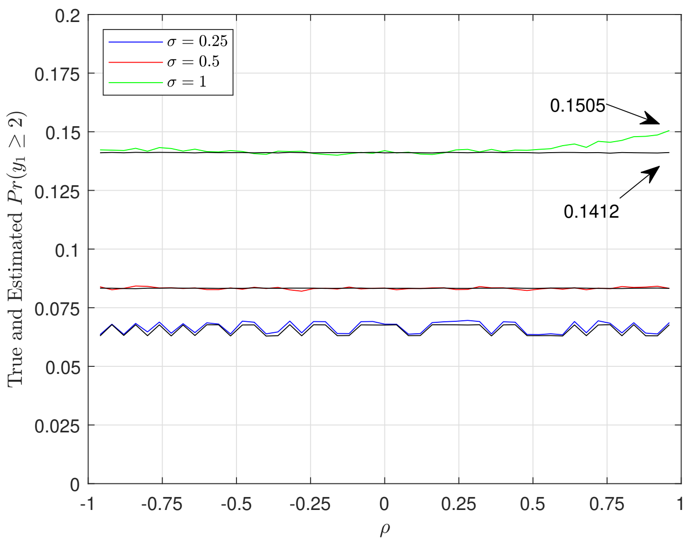

Figure 30. Plots of True and Estimated $Pr(y_1 \geq 2)$: Bivariate Poisson Model (S=1000).

REFERENCES

Aitchison, J. and C. Ho (1989). The multivariate poisson-log normal distribution. *Biometrika 76*(4), 643–653.

Bhat, C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B: Methodological 35*(7), 677–693.

Bhat, C. R. (2003). Simulation estimation of mixed discrete choice models using randomized and scrambled halton sequences. *Transportation Research Part B: Methodological 37*(9), 837–855.

Chib, S. and R. Winkelmann (2001). Markov chain monte carlo analysis of correlated count data. *Journal of Business and Economic Statistics 19*(4), 428–435.

Deb, P. and P. K. Trivedi (2006). Specification and simulated likelihood estimation of a non-normal treatment-outcome model with selection: Application to health care utilization. *The Econometrics Journal 9*(2), 307–331.

Filippini, M. and W. Greene (2016). Persistent and transient productive inefficiency: a maximum simulated likelihood approach. *Journal of Productivity Analysis 45*(2), 187–196.

Geweke, J. F., M. P. Keane, and D. E. Runkle (1997). Statistical inference in the multinomial multiperiod probit model. *Journal of Econometrics 80*(1), 125–165.

Gourieroux, C. and A. Monfort (1990). Simulation based inference in models with heterogeneity. *Annales d'Economie et de Statistique*, 69–107.

Gourieroux, C. and A. Monfort (1993). Simulation-based inference: A survey with special reference to panel data models. *Journal of Econometrics 59*(1-2), 5–33.

Gourieroux, M., C. Gourieroux, A. Monfort, D. A. Monfort, et al. (1996). *Simulation-based econometric methods*. Oxford university press.

Greene, W. (2012). H.(2012): Econometric analysis.

Greene, W. H. (2003). Simulated likelihood estimation of the normal-gamma stochastic frontier function. *Journal of Productivity Analysis 19*(2-3), 179–190.

Hajivassiliou, V. A. and P. A. Ruud (1994). Classical estimation methods for ldv models using simulation. *Handbook of econometrics 4*, 2383–2441.

Halton, J. H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik 2*(1), 84–90.

Hellström, J. (2006). A bivariate count data model for household tourism demand. *Journal of Applied Econometrics 21*(2), 213–226.

Hochguertel, S. and H. Ohlsson (2009). Compensatory inter vivos gifts. *Journal of applied econometrics 24*(6), 993–1023.

Lee, L.-F. (1995). Asymptotic bias in simulated maximum likelihood estimation of discrete choice models. *Econometric Theory*, 437–483.

Lerman, S. and C. Manski (1981). On the use of simulated frequencies to approximate choice probabilities. *Structural analysis of discrete data with econometric applications 10*, 305–319.

McFadden, D. and K. Train (2000). Mixed mnl models for discrete response. *Journal of applied Econometrics 15*(5), 447–470.

Munizaga, M. A. and R. Alvarez-Daziano (2001). Mixed logit vs. nested logit and probit models. *Departamento de Ingeniería Civil, Universidad de Chile, Santiago, Chile*.

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.

Palma, M. A., D. V. Vedenov, and D. Bessler (2020). The order of variables, simulation noise, and accuracy of mixed logit estimates. *Empirical Economics 58*(5), 2049–2083.

Plum, A. (2019). The british low-wage sector and the employment prospects of the unemployed. *Applied Economics 51*(13), 1411–1432.

Train, K. (2000). Halton sequences for mixed logit.

Train, K. E. (2009). *Discrete choice methods with simulation.* Cambridge university press.