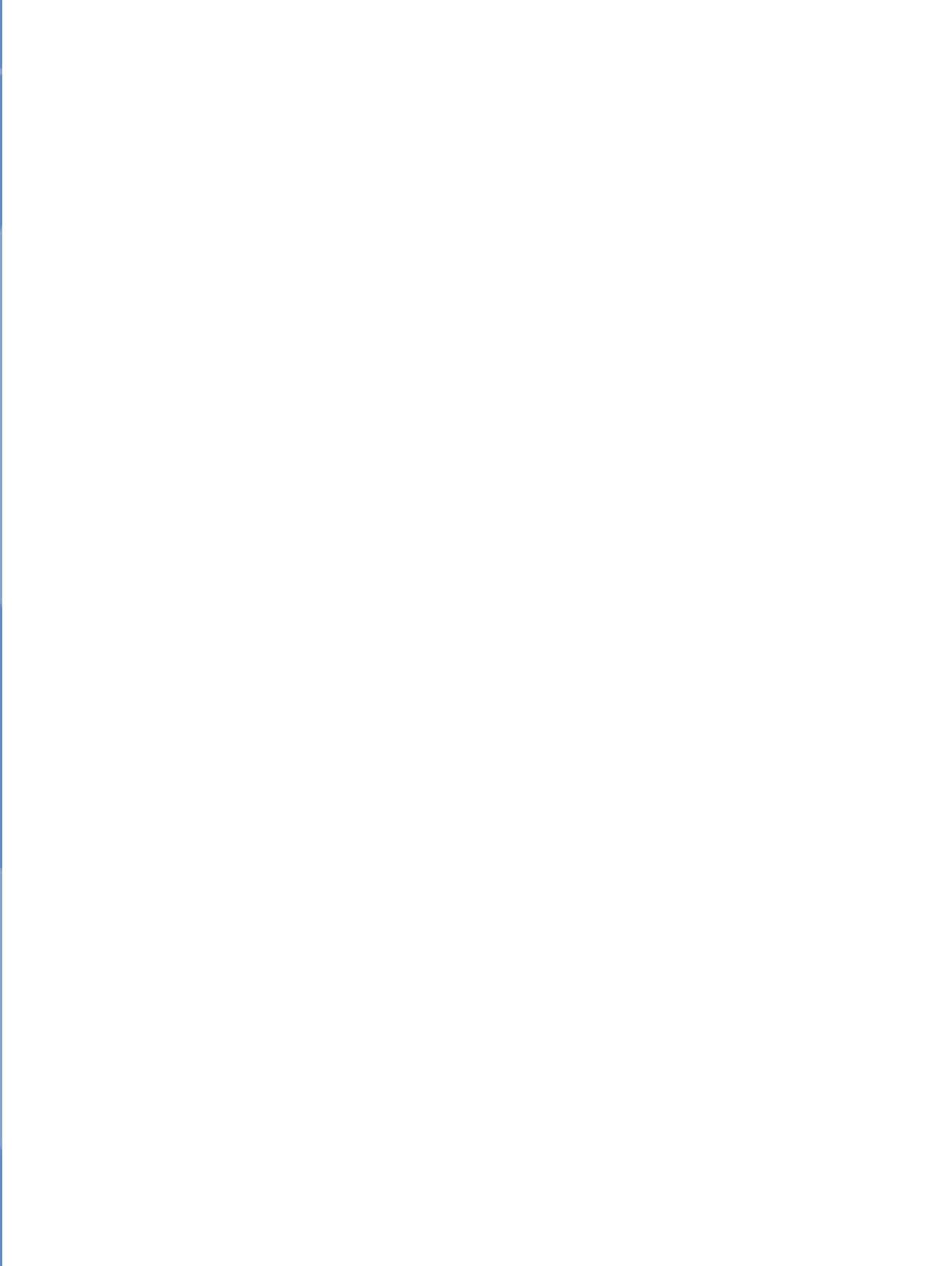


# ScholarWorks@GSU

## Flexible Bivariate Count Data Regression Models

Authors	Gurmu, Shiferaw;Elder, John
Download date	2026-03-13 18:01:36
Link to Item	<a href="https://hdl.handle.net/20.500.14694/15103">https://hdl.handle.net/20.500.14694/15103</a>



ANDREW YOUNG SCHOOL  
OF POLICY STUDIES



# Flexible Bivariate Count Data Regression Models

Shiferaw Gurmu\* and John Elder†

October 29, 2011

## Abstract

The paper develops a semiparametric estimation method for the bivariate count data regression model. We develop a series expansion approach in which dependence between count variables is introduced by means of stochastically related unobserved heterogeneity components, and in which, unlike existing commonly used models, positive as well as negative correlations are allowed. Extensions that accommodate excess zeros, censored data and multivariate generalizations are also given. Monte Carlo experiments and an empirical application to tobacco use confirms that the model performs well relative to existing bivariate models, in terms of various statistical criteria and in capturing the range of correlation among dependent variables. This article has supplementary materials online.

**KEYWORDS:** Series estimation; Negative correlation; Unobserved heterogeneity; Tobacco use; Censoring; Multivariate counts.

---

\*Corresponding author: Department of Economics, Andrew Young School of Policy Studies, P.O. Box 3992, Georgia State University, Atlanta, GA 30302. Phone: (404) 413-0161; E-mail: sgurmu@gsu.edu

†Department of Finance & Real Estate, 1272 Campus Delivery, Colorado State University, Fort Collins, CO 80523

# 1 INTRODUCTION

The main focus of this paper is on series estimation of bivariate count data regression models with more flexible correlation structure. Multivariate count data regression models arise in situations where several counts are correlated and joint estimation is required. For example, common measures of health-care utilization, such as the number of doctor consultations, the number of other ambulatory visits, and prescription drug utilizations, are likely to be jointly dependent. Other leading examples include the number of hospital admissions and the number of days spent in hospitals, the number of voluntary and involuntary job changes, the number of firms which enter and exit an industry and the number of patents made and papers published by scientists (Mayer and Chappell 1992; Cameron and Trivedi 1993; Jung and Winkelmann 1993; Hellstrom 2006; Stephan, Gurmu, Sumell and Black 2007). Gurmu and Trivedi (1994) and Cameron and Trivedi (1998) provide overviews of some of the earlier multivariate count data regression models. Kocherlakota and Kocherlakota (1992) and Johnson, Kotz and Balakrishnan (1997) discuss the statistical properties of multivariate discrete distributions.

Applications of these models in economics, statistics and related fields have largely been confined to the bivariate Poisson, bivariate negative binomial and bivariate Poisson-lognormal mixture regression models. The bivariate Poisson model imposes the restriction that the conditional mean of each count variable equals the conditional variance. This can result in misspecification of the marginals. For the common case of overdispersed counts, the bivariate (or multivariate) negative binomial model is potentially useful. However, the multivariate negative binomial model imposes *ad hoc* conditions on the form of the distribution of the common unobserved heterogeneity term affecting the counts. In estimation of truncated and censored models using a fully parametric approach, a misspecification of the distribution of unobservables leads to inconsistency. Since the conditional means in these models depend on the density, the form of the

density will affect the consistency of the regression parameters. A general flexible framework of estimation that does not require knowledge of the distribution of the unobservables is desirable.

Existing commonly used bivariate count models have another potential shortcoming: they accommodate only non-negative correlation between the counts. The statistics literature gives examples and general techniques on constructing negatively correlated multivariate Poisson distributions having Poisson marginals; see, for example, Kocherlakota and Kocherlakota (1992) for a review of this literature and Ophem (1999) for a related approach. In particular, Aitchison and Ho (1989) consider a multivariate log-normal mixture of independent Poisson distributions. Since the resulting mixture, the Poisson-log normal distribution, does not have a closed form solution, estimation of the model requires numerical integration (Munkin and Trivedi 1999; Hellstrom 2006). The approach also requires knowledge of the distribution of unobserved heterogeneity components. From the practical point of view, it is desirable to have flexible methods that allow for both positive and negative correlations between the dependent variables. This paper addresses these issues.

This paper develops semi-parametric methods for estimation of various mixture bivariate count models that allow for negative as well as positive correlations between the counts. In the context of models with two-factor unobserved heterogeneity components, this paper employs a series expansion approach to model the joint distribution of unobserved heterogeneity components in the context of a bivariate Poisson mixture. To this end, we begin by analyzing the role played by the distribution of unobserved heterogeneity in the sign and magnitude of the correlation between counts. We also present methods for estimating truncated and censored bivariate regression models as well as extensions to zero-inflation and multivariate models. These extensions can be found in a Supplemental Appendix. Simulation experiments and an application to tobacco consumption are presented. The application involves joint modeling of the use of smoking tobacco and chewing tobacco as reported in a survey of households.

The general approach is a multivariate generalization of the semiparametric techniques developed for univariate count regression models (Gurmu 1997; Gurmu, Rilstone, and Stern 1999). Semiparametric estimators based on squared polynomial series expansions provide consistent estimates (Gallant and Nychka 1987; Gurmu *et al.* 1999; Bierens 2008). As compared to flexible univariate methods, joint estimation of equations with correlated dependent count variables is expected to improve efficiency. In principle, if interest is only on estimation of the parameters of the conditional mean function in a multivariate count model, seemingly unrelated non-linear least squares estimation methods may be feasible, provided the underlying model allows for flexible variance and correlation structures. However, apart from estimation of parameters of economic interest, the focus of interest in most count regression models is consistent estimation of cell probabilities for different counts. The proposed series expansion approach provides flexible specifications for the joint distribution of the counts as well as the marginals. In addition to estimating the expected number of counts and the correlation between pairs of endogenous variable, the series method provides consistent estimates of joint probabilities of event counts, conditional on covariates. Correlations between jointly distributed variables are not restricted to be non-negative. Further, methods based on first and second order moments alone are not directly applicable to truncated and censored models. The semiparametric approach is particularly useful for truncated and censored regression models, where misspecification of the density leads to inconsistency. In a recent paper, Gurmu and Elder (2007) analyzed a special case of the bivariate model proposed in this paper based on first-order series expansion.

The remainder of the paper is organized as follows. The next section develops a method for flexible series estimation of a bivariate mixture count data regression model. Section 3 evaluates the feasibility and performance of the proposed method using Monte Carlo experiments. An application to individual tobacco consumption behavior is given in Section 4. Section 5 gives concluding remarks. Derivations of main results are given in Appendix A. Generalizations, detailed derivations, and further results from

simulations and applications are provided in Supplemental appendices.

## 2 TWO-FACTOR BIVARIATE MODEL

### 2.1 The Framework

A major shortcoming of standard models such as the bivariate Poisson and bivariate negative binomial models is that they do not allow for negative correlations between the count variables. This restriction arises from the assumption of a common unobserved heterogeneity component used in the underlying mixing or convolution technique, resulting in the correlation between the counts being introduced through a dispersion parameter. We consider a two-factor framework in which dependence between count variables is introduced through correlated unobserved heterogeneity components.

Consider two jointly distributed random variables,  $Y_1$  and  $Y_2$ , each denoting event counts. For observation  $i$  ( $i = 1, 2, \dots, N$ ), we observe  $\{y_{ji}, \mathbf{x}_{ji}\}_{j=1}^2$ , where  $\mathbf{x}_{ji}$  is a  $(k_j \times 1)$  vector of covariates. Without loss of generality, the mean parameter associated with  $y_{ji}$  can be parameterized as

$$\theta_{ji} = \exp(\mathbf{x}'_{ji}\boldsymbol{\beta}_j), \quad j = 1, 2 \quad (1)$$

where  $\boldsymbol{\beta}_j$  is a  $(k_j \times 1)$  vector of unknown parameters. We model the dependence between  $y_1$  and  $y_2$  by means of correlated unobserved heterogeneity components  $\nu_1$  and  $\nu_2$ . Each of the components is associated with only one of the event counts. Accordingly, for  $j = 1, 2$ , suppose  $(y_{ji} \mid \mathbf{x}_{ji}, \nu_{ji}) \sim \text{Poisson}(\theta_{ji}\nu_{ji})$  with  $(\nu_{1i}, \nu_{2i})$  having a bivariate distribution  $g(\nu_{1i}, \nu_{2i})$  in  $\mathfrak{R}_+^2$ . Then the ensuing mixture density can be expressed as

$$f(y_{1i}, y_{2i} \mid \mathbf{x}_i) = \int \int \left[ \prod_{j=1}^2 \frac{\exp(-\theta_{ji}\nu_{ji}) (\theta_{ji}\nu_{ji})^{y_{ji}}}{\Gamma(y_{ji} + 1)} \right] g(\nu_{1i}, \nu_{2i}) d\nu_{1i} d\nu_{2i}. \quad (2)$$

Let  $M(-\theta_{1i}, -\theta_{2i}) = E_\nu [\exp(-\theta_{1i}\nu_{1i} - \theta_{2i}\nu_{2i})]$  denote the bivariate moment generating function (MGF) of  $(\nu_{1i}, \nu_{2i})$  evaluated at  $(-\theta_{1i}, -\theta_{2i})$ . It can readily be seen that,

analogous to the one-factor error model (*e.g.*, Gurmu and Elder 2000), (2) takes the form

$$f(y_{1i}, y_{2i} | \mathbf{x}_i) = \left[ \prod_{j=1}^2 \frac{(\theta_{ji})^{y_{ji}}}{\Gamma(y_{ji} + 1)} \right] M^{(y_1, y_2)}(-\theta_{1i}, -\theta_{2i}), \quad (3)$$

where, suppressing  $i$ ,  $M^{(y_1, y_2)}(-\theta_1, -\theta_2) = \partial^y M(-\theta_1, -\theta_2) / (\partial(-\theta_1)^{y_1} \partial(-\theta_2)^{y_2})$  is the derivative of  $M(-\theta_1, -\theta_2)$  of order  $y = y_1 + y_2$ .

The sign of the correlation coefficient between  $y_1$  and  $y_2$  is determined by the sign of the covariance between the two unobserved variables,  $\text{Cov}(\nu_1, \nu_2)$ . This is summarized in the following proposition.

**Proposition 1** *In the mixture bivariate model (3), with the assumption of independence between  $\mathbf{x}$  and  $\nu$ , the covariance between event counts takes the form*

$$\text{cov}(y_{1i}, y_{2i} | \mathbf{x}_i) = \theta_{1i}\theta_{2i} [\text{Cov}(\nu_{1i}, \nu_{2i})].$$

where  $\text{cov}(\nu_{1i}, \nu_{2i}) = [M^{(1,1)}(0, 0) - M^{(1,0)}(0, 0)M^{(0,1)}(0, 0)]$ . Since  $\theta_{ji}$  is non-negative,  $\text{sign}(\text{Cov}(y_{1i}, y_{2i})) = \text{sign}(\text{Cov}(\nu_{1i}, \nu_{2i}))$ .

The above result can be obtained using iterated expectations. The intuition about the sign of the correlation is as follows. In the case of univariate mixing, the correlation between the counts is affected only by the variance of the common unobserved heterogeneity term. Hence correlation is non-negative. In the bivariate mixing, the variance of each unobserved components as well as the correlation between the components affect  $\text{Corr}(y_{1i}, y_{2i} | \mathbf{x}_i)$ . Hence, the sign of the correlation between the count variables is determined by the allowed signs for the correlation between the unobserved heterogeneity terms, which generally may or may not be restricted.

Using the law of iterated expectations, the other low-order moments of the outcome variables, conditional on  $\mathbf{x}_i$ , can be obtained as:

$$E(y_{ji} | \mathbf{x}_i) = \theta_{ji}E(\nu_{ji}), \quad j = 1, 2 \quad (4)$$

$$\text{Var}(y_{ji} \mid \mathbf{x}_i) = \theta_{ji} E(\nu_{ji}) + \theta_{ji}^2 \text{Var}(\nu_{ji}), \quad (5)$$

$$\text{Corr}(y_{1i}, y_{2i} \mid \mathbf{x}_i) = \frac{\text{Corr}(\nu_{1i}, \nu_{2i})}{\sqrt{\left(1 + \frac{E(\nu_{1i})}{\theta_{1i} \text{Var}(\nu_{1i})}\right) \left(1 + \frac{E(\nu_{2i})}{\theta_{2i} \text{Var}(\nu_{2i})}\right)}}. \quad (6)$$

Equation (6) shows that, since  $E(\nu_{ji})$ ,  $\text{Var}(\nu_{ji})$  and  $\theta_{ji}$  are strictly positive,

$$|\text{Corr}(y_{1i}, y_{2i} \mid \mathbf{x}_i)| < |\text{Corr}(\nu_{1i}, \nu_{2i})| \quad (7)$$

for any arbitrary mixing distribution for  $\boldsymbol{\nu}_i$ . While the sign of the correlation between the event counts is unrestricted, the magnitudes of these correlations are narrower than the corresponding correlations from the mixing distribution. However, as  $\theta_{1i}$  and  $\theta_{2i}$  get larger, the gap between the two correlations will be narrower, with  $\text{Corr}(y_{1i}, y_{2i} \mid \mathbf{x}_i)$  getting very close to  $\text{Corr}(\nu_{1i}, \nu_{2i})$ . Aitchison and Ho (1989) derive similar bounds for bivariate correlations between the count variables for multivariate Poisson-lognormal mixture model without regressors. The correlation bounds in (7) apply to any arbitrary Poisson-unobserved factor mixture model, including the one proposed in this paper.

The form of the density (3) depends upon the choice of the distribution of the unobservables,  $g(\nu_{1i}, \nu_{2i})$ . If  $g(\cdot)$  follows a bivariate (or generally a multivariate) log-normal distribution, we get the bivariate (or multivariate) Poisson log-normal distribution proposed by Aitchison and Ho (1989). However, since the Poisson-log normal mixture does not have a closed form, estimation of the unknown parameters requires numerical integration. Nevertheless, the low-order moments, including correlations, of the multivariate Poisson log-normal mixture distribution can easily be obtained (see Aitchison and Ho 1989). For example, Munkin and Trivedi (1999), Chib and Winkelmann (2001) and Hellstrom (2006) have applied the Poisson log-normal correlated model using simulated maximum likelihood or Markov Chain Monte Carlo estimation methods.

We develop an estimation approach for model (3) that does not require knowledge of the distribution of the unobserved heterogeneity components. Another attractive feature of the approach is that, in addition to allowing for negative correlations, we obtain a closed form for the resulting correlated mixture model. The strategy is to

approximate the density  $g(\nu_{1i}, \nu_{2i})$  using techniques in bivariate density expansion. We propose to take the approximate bivariate density to be of the general form

$$g_N(\nu_1, \nu_2) = w(\nu_1)w(\nu_2) \frac{1}{\varpi} \left[ \sum_{k=0}^K \sum_{r=0}^K \rho_{kr} P_k(\nu_1) P_r(\nu_2) \right]^2, \quad (8)$$

where  $w(\nu_1)$  and  $w(\nu_2)$  are the hypothesized baseline marginal densities of  $\nu_1$  and  $\nu_2$ , respectively,  $\{P_k(\nu_1), P_r(\nu_2)\}$  are orthonormal polynomials of degree  $k$  and  $r$ , respectively, defined on the respective marginal distributions, and

$$\varpi = \int \int w(\nu_1)w(\nu_2) \left[ \sum_{k=0}^K \sum_{r=0}^K \rho_{kr} P_k(\nu_1) P_r(\nu_2) \right]^2 dv_1 dv_2 \quad (9)$$

is a constant of proportionality. The polynomial degree  $K$  controls the extent to which the ensuing mixture model  $f(y_{1i}, y_{2i} \mid \mathbf{x}_i)$  deviates from the assumed baseline density. As discussed below, when  $K$  is set to zero in the proposed bivariate density of  $y_1$  and  $y_2$ , we get a density that is a product of two independent negative binomial distributions. The polynomial degree parameter is expected to increase with the sample size. In principle, the rule for increasing the size of series expansion can be data dependent (e.g., Gallant and Nychka 1987). The parameter  $\rho_{kr}$  may be considered as the coefficient of correlation between the  $(k, r)$  polynomials in the bivariate distribution  $g_N(\nu_1, \nu_2)$ .

The next step is to obtain the MGF of  $(\nu_{1i}, \nu_{2i})$  and its derivatives of order  $y = y_1 + y_2$ , say,  $M_N^{(y_1, y_2)}(-\theta_{1i}, -\theta_{2i})$ , and subsequently plug this in (3). Then  $\beta_1$ ,  $\beta_2$  and parameters in  $g_N(\nu_1, \nu_2)$  are estimated within the likelihood framework. This is discussed next.

## 2.2 Estimation

In our implementation, we use bivariate expansions based on generalized Laguerre polynomials. Let  $L_{kj}(\nu_{ji})$  denote relevant  $k$ -th order generalized Laguerre polynomial associated with the random variable  $\nu_{ji}$  with baseline gamma weight  $w(\nu_{ji})$  having parameters  $\alpha_j$  and  $\lambda_j$ ,  $j = 1, 2$ ; see equations (A.1) and (A.2) in Appendix A. The ensuing

$k$ -th order orthonormal Laguerre polynomial takes the form  $P_k(\nu_{ji}) = h_{kj}^{-1/2} L_{kj}(\nu_{ji})$ , where

$$h_{kj} = \frac{\Gamma(k + \alpha_j)}{\Gamma(\alpha_j)\Gamma(k + 1)}, \text{ for } j = 1, 2.$$

Given the prevalence of gamma mixing in count and related duration models, the system of Laguerre polynomials is a natural choice.

The key result we need now is the approximate MGF,  $M_N(-\theta_{1i}, -\theta_{2i})$ , and its derivative derived from the approximate joint density based on Laguerre series expansion. The derivation of  $M_N(-\theta_{1i}, -\theta_{2i})$  is given in Appendix A. The required derivative of this MGF is:

$$\begin{aligned} M_N^{(y_1, y_2)}(-\theta_{1i}, -\theta_{2i}) &= \frac{1}{\sum_{k=0}^K \sum_{l=0}^K \rho_{kr}^2} \left( \prod_{j=1}^2 \left( 1 + \frac{\theta_{ji}}{\lambda_j} \right)^{-\alpha_j} (\lambda_j + \theta_{ji})^{-y_{ji}} \Gamma(\alpha_j) \right) \\ &\quad \times \sum_{k,r,l,m=0}^K \rho_{kr} \rho_{lm} (h_{k1} h_{r2} h_{l1} h_{m2})^{1/2} \Psi_{1kl} \Psi_{2rm} \end{aligned} \quad (10)$$

where

$$\Psi_{1kl}(-\theta_{1i}, y_{1i}) = \sum_{k_1=0}^k \sum_{l_1=0}^l \binom{k}{k_1} \binom{l}{l_1} \frac{\Gamma(\alpha_1 + k_1 + l_1 + y_{1i})}{\Gamma(\alpha_1 + k_1)\Gamma(\alpha_1 + l_1)} \left( -1 - \frac{\theta_{1i}}{\lambda_1} \right)^{-(k_1+l_1)}, \quad (11)$$

and

$$\Psi_{2rm}(-\theta_{2i}, y_{2i}) = \sum_{r_2=0}^r \sum_{m_2=0}^m \binom{r}{r_2} \binom{m}{m_2} \frac{\Gamma(\alpha_2 + r_2 + m_2 + y_{2i})}{\Gamma(\alpha_2 + r_2)\Gamma(\alpha_2 + m_2)} \left( -1 - \frac{\theta_{2i}}{\lambda_2} \right)^{-(r_2+m_2)}, \quad (12)$$

To construct the log-likelihood based on (3) and (10), we need to restrict the mean of each unobserved heterogeneity component to unity; that is  $M_N^{(1,0)}(0,0) = 1$  and  $M_N^{(0,1)}(0,0) = 1$ . As shown in Appendix A (see A.6 and A.7), these conditions impose restrictions on  $\lambda_1$  and  $\lambda_2$ . Hence, the semiparametric (SP2) log-likelihood function

based on bivariate mixing is

$$\begin{aligned} \mathcal{L}_{sp2}(\boldsymbol{\varphi}) &= \sum_{i=1}^N [y_{1i} \ln \theta_{1i} + y_{2i} \ln \theta_{2i} - \ln(\Gamma(y_{1i} + 1)) - \ln(\Gamma(y_{2i} + 1))] \\ &\quad + \ln M_N^{(y_1, y_2)}(-\theta_{1i}, -\theta_{2i}) \end{aligned} \tag{13}$$

where  $\boldsymbol{\varphi}$  now consists of  $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \alpha_1, \alpha_2$ , and the  $\rho$ 's with normalization  $\rho_{00} = 1$ . SP2 denotes a bivariate semiparametric model with 2 unobserved heterogeneity components. Likewise, SPJ will denote a multivariate model with  $J$  unobserved heterogeneity components. SP will be used generically to denote the proposed semiparametric method for a class of multivariate count data models. As is conventional with orthogonal expansions (*e.g.*, Lancaster 1969),  $P_0(\nu_{ji}) = 1$  and  $\rho_{00} = 1$ , so that  $E[P_k(\nu_{ji})] = 0$  for  $k = 1, 2, \dots$ , implying that  $\rho_{kr} = 0$  if either  $k$  or  $r$  is zero, but not both. As explained in the paragraph following (8),  $\rho_{kr} = \text{Corr}(P_k(\nu_1), P_r(\nu_2))$  with respect to the approximate bivariate distribution  $g_N(\nu_1, \nu_2)$ . For example, if  $K = 1$ , the correlation parameter to be estimated is  $\rho_{11}$ , where  $\rho_{11} = \text{Corr}(P_1(\nu_1), P_1(\nu_2))$  and  $P_1(\nu_{ji}) = \left( \sqrt{\alpha_j} - \frac{\lambda_j}{\sqrt{\alpha_j}} \nu_{ji} \right)$ ,  $j = 1, 2$ . For  $K = 2$ , the correlation parameters are  $\rho_{11}, \rho_{12}, \rho_{21}$ , and  $\rho_{22}$ . For  $K = 2$ , for example,  $\rho_{11} = \text{Corr}(P_1(\nu_1), P_1(\nu_2))$ ,  $\rho_{12} = \text{Corr}(P_1(\nu_1), P_2(\nu_2))$ , and  $\rho_{22} = \text{Corr}(P_2(\nu_1), P_2(\nu_2))$ . Note that, for instance,  $P_2(\nu_1) = h_{21}^{-1/2}(\nu_1) L_{21}(\nu_1)$  is a quadratic polynomial in  $\nu_1$ . Here  $h_{21} = \Gamma(2 + \alpha_1)/2\Gamma(\alpha_1)$  and  $L_{kj}(\nu_j)$  is defined in equation (A.1) of Appendix A.

Due to the assumption of differential variances of the error components, the semiparametric approach provides a more flexible form for the variance of  $(y_{ji} \mid \mathbf{x}_i)$  as compared to standard one-factor bivariate count models based on mixing or convolutions. The correlation between  $y_1$  and  $y_2$  for any  $K$  is given in Appendix A, equation A.8. This correlation can take on zero, positive or negative values. The correlation between the count series depends on the mean parameters ( $\theta$ 's, hence  $\boldsymbol{\beta}$ 's), the dispersion parameters ( $\alpha$ 's), and the correlation parameters ( $\rho$ 's). By contrast, if the unobservables  $\log \nu_{1i}$  and  $\log \nu_{2i}$  are assumed to be distributed jointly as normally with mean vector 0,  $\text{Var}(\log \nu_{ji}) = \sigma_j^2$ , and  $\text{Corr}(\log \nu_{1i}, \log \nu_{2i}) = \tau$ , then  $\text{Corr}(y_1, y_2)$  implied by the

ensuing bivariate Poisson log-normal mixture density depends on  $\theta$ 's, the  $\sigma$ 's, and a single correlation parameter  $\tau$ . As discussed earlier following equation (7), for both the Poisson-Laguerre Polynomial and Poisson-lognormal mixture models, the range of values taken by the correlation between  $y_1$  and  $y_2$  is not as wide as that of the corresponding range of correlation values between  $\nu_1$  and  $\nu_2$ ; they can be close when  $\theta_j$ 's are large.

The SP2 approach is based on the assumption of a baseline gamma density. The gamma mixing is a popular approach in applications of count data and related models, including models of survival data. Since the SP2 density is based on gamma mixing, it nests the negative binomial class of models. When  $K = 0$  (note  $\rho_{00} = 1$ ), we get a density that is a product of two independent negative binomial distributions. If  $K > 0$ , we get a correlated bivariate density whose shape is modified by the presence of the  $\rho$ 's and degree of series truncation. The bivariate count data model proposed by Gurmu and Elder (2007) is a special case of (13) with  $K = 1$ . Analogous to the Poisson log-normal model, the SP2 density with  $K = 1$  may be thought of as a mixture of Poisson and bivariate gamma distributions. One of the main advantages of using the gamma baseline density is that it yields a computationally tractable closed form of the ensuing mixture model. As indicated by a referee, more polynomial terms ( $K$ ) increase the set of possible distributional shapes obtained. One might argue that the correct choice of the baseline density is important for low  $K$ , say  $K = 1$ . However, as explained below, this may not be an issue if no specification error is detected for low  $K$ .

The SP2 class of densities provides reasonable approximations to count models that include densities with long upper tails, relatively skewed densities, and densities with low means as well as those with high means. The consistency of the proposed estimator can be shown using the framework developed by Gallant and Nychka (1987) and Gurmu *et al.* (1999). The parameter space is a set  $\mathcal{B} \times \mathcal{M}$ , where  $\beta \in \mathcal{B} \subset \mathcal{R}^p$  and  $\mathcal{M}$  is the set of moment generating functions for positive random variables. Here  $\beta = (\beta'_1 \beta'_2)'$ . Note that the estimators of the moment generation function are random sequence based

on the approximating function  $g_N(\cdot)$ . This requires that  $K \rightarrow \infty$  as  $N \rightarrow \infty$ , among other things.

In general, the proposed series estimator may be considered as the quasi maximum likelihood estimator. Implementation of the estimator requires choosing the degree of the polynomial  $K$ . Following the literature on series estimation, one can use the Akaike Information Criterion to choose  $K$  or to compare different models. As argued by Gallant and Tauchen (1989), the truncated series expansion approach ( $K$  finite) may be viewed as finite-dimensional inference that has been subjected to a sensitivity analysis. That is, when specification error is detected, one increases the truncation point  $K$ . In implementation, we use finite-dimensional inference ( $K$  fixed) to compute standard errors. Comparison of bivariate models may also be based on how well they predict the cell probabilities,  $P(y_{1i} = r, y_{2i} = s)$ , for  $r, s = 0, 1, 2, \dots$ . These issues will be considered further in the application section.

### 3 MONTE CARLO EXPERIMENTS

This section reports the results of Monte Carlo experiments that illustrate the feasibility and finite sample performance of the proposed estimation approach. The data generating process (DGP) is based on the bivariate Poisson log-normal mixture model, which is the basis of the estimation approach considered by Munkin and Trivedi (1999) and others. We also evaluate its performance relative to the bivariate negative binomial maximum likelihood estimator, which is commonly used in the applied literature due to its flexibility in handling overdispersed data.

We define an empirically relevant DGP with low means and overdispersed bivariate counts with varying assumptions about the distribution of unobserved heterogeneity, allowing for positive as well as negative correlations between the response variables. Suppressing reference to observation  $i$ , the mean parameters of Poisson distributed

count variables  $y_1$  and  $y_2$  are specified as

$$\theta_j^* = \exp(\gamma_j + \beta_j x_j) \nu_j, \quad j = 1, 2. \quad (14)$$

The values of explanatory variables  $x_1$  and  $x_2$  are generated independently from normal  $N(0, 1/16)$ , with true values for the regression parameters set to  $\gamma_1 = \beta_1 = \gamma_2 = \beta_2 = 1$  in all experiments. The unobserved heterogeneity terms  $\nu_1$  and  $\nu_2$  are generated by bivariate uniform and bivariate lognormal distributions using the following specifications, along the implied average characteristics of the DGP.

1. Bivariate log-normal with positive correlation:  $\log(\nu_1) = \varepsilon_1$  and  $\log(\nu_2) = \varepsilon_2$  follow bivariate normal distribution with mean  $(0, 0)$ , variances  $\sigma_1^2 = \sigma_2^2 = 0.25$  and correlation parameter  $\rho_\varepsilon = 0.6$ . This gives  $\text{mean}(\nu_j) = 1.12$ ,  $\text{var}(\nu_j) = 0.37$  and  $\text{corr}(\nu_1, \nu_2) = 0.57$  and average moments:  $\text{mean}(y_j) = 3.08$ ,  $\text{var}(y_j) = 8.40$  and  $\text{corr}(y_1, y_2) = 0.24$ .
2. Bivariate log-normal with negative correlation:  $\log(\nu_1) = \varepsilon_1$  and  $\log(\nu_2) = \varepsilon_2$  follow bivariate normal distribution with mean  $(0, 0)$ , variances  $\sigma_1^2 = \sigma_2^2 = 0.25$  and correlation parameter  $\rho_\varepsilon = -0.6$ . This implies  $\text{mean}(\nu_j) = 1.12$ ,  $\text{var}(\nu_j) = 0.37$  and  $\text{corr}(\nu_1, \nu_2) = -0.49$  and the average moments:  $\text{mean}(y_j) = 3.08$ ,  $\text{var}(y_j) = 8.40$  and  $\text{corr}(y_1, y_2) = -0.20$ .
3. Bivariate uniform with correlation 0.6 with approximate moments:  $\text{mean}(\nu_j) = 1.0$ ,  $\text{var}(\nu_j) = 0.33$  and average moments  $\text{mean}(y_j) = 2.81$ ,  $\text{var}(y_j) = 6.10$  and  $\text{corr}(y_1, y_2) = 0.27$ .
4. Bivariate uniform with correlation  $-0.6$  with approximate moments:  $\text{mean}(\nu_j) = 1.0$ ,  $\text{var}(\nu_j) = 0.33$  and average moments  $\text{mean}(y_j) = 2.81$ ,  $\text{var}(y_j) = 6.10$  and  $\text{corr}(y_1, y_2) = -0.27$ .

Specifications 1 and 2 and the ensuing bivariate Poisson-lognormal mixture are based on the DGP used by Munkin and Trivedi (1999). Simulation experiments are based on sample size of 1000 with each experiment replicated 1000 times.

Table 1 summarizes the simulation results in terms of mean bias, root mean squared error (RMSE) and mean absolute bias based on bivariate lognormal DGP. Figure 1 gives kernel density estimates of the sampling distributions of  $\hat{\beta}_1$  (Beta1) and correlations between the count variables from DGPs 2 and 4; model estimated by SP2 with  $K = 2$ . Additional simulation results, including results using bivariate uniform DGP, kernel density estimates of the sampling distributions of coefficient estimates and estimated correlations, are given in the Supplemental Appendix online. The results show that the SP2 estimates have low and insignificant biases under each of the 4 DGPs, irrespective of the values of the underlying correlations. The bias associated with any of the coefficients is less than 1% for SP2 with  $K = 2$ . The RMSE is roughly about 2% for the intercept terms and about 10% for the slope coefficient estimates. Kernel density estimates demonstrate that, for a given  $K$ , the sampling distributions of  $\hat{\gamma}_j$  and  $\hat{\beta}_j$  from SP2 are approximately normal. As expected, the SP2 estimator performs much better than the commonly used bivariate negative binomial estimator. Unlike simulation results for univariate NB of Gurmu *et al.* (1999), the biases associated with slope parameters resulting from bivariate NB can be quite substantial (35% to 65%) even when data are uncensored. The variances of the slope estimates from bivariate NB associated with DGPs with negative correlations tend to be higher than those with positive correlations. The SP2 approach successfully predicts negative or positive correlations for all DGPs and even for each replication. However, the sampling distribution of the estimated correlations varies somewhat from the corresponding empirical distribution of correlations in all cases.

[Figure 1 about here]

Using the bias-corrected simulated ML estimator with  $N = 1000$ , 100 replications and 200 simulations, Munkin and Trivedi (1999, Table 1 Page 41) report mean estimates for parameters  $\gamma_1, \beta_1, \gamma_2$  and  $\beta_2$  in DGP 1 of 0.998, 1.012, 0.996 and 1.009, respectively with the corresponding standard errors of 0.028, 0.1, 0.028 and 0.097. By comparison,

Table 1: Monte Carlo results on parameter estimates by bivariate count models

DGP: Bivar	Est. Method	Parameter	Mean	RMSE	Mean
heterogeneity			bias		abs. bias
1. Lognormal, $\text{Corr}(\nu_1, \nu_2) = 0.57$					
	Bivar NB	$\gamma_1$	0.0116	0.0853	0.0634
		$\beta_1$	0.5890	0.6355	0.5893
		$\gamma_2$	0.0120	0.0862	0.0643
		$\beta_2$	0.5905	0.6297	0.5909
	Bivar SP2	$\gamma_1$	-0.0008	0.0246	0.0196
		$\beta_1$	0.0009	0.0972	0.0772
		$\gamma_2$	-0.0007	0.0253	0.0203
		$\beta_2$	0.0017	0.0960	0.0756
2. Lognormal, $\text{Corr}(\nu_1, \nu_2) = -0.49$					
	Bivar NB	$\gamma_1$	0.0307	0.0530	0.0418
		$\beta_1$	0.4198	0.4640	0.4201
		$\gamma_2$	0.0310	0.0535	0.0421
		$\beta_2$	0.4285	0.4686	0.4288
	Bivar SP2	$\gamma_1$	-0.0008	0.0248	0.0196
		$\beta_1$	-0.0029	0.0968	0.0774
		$\gamma_2$	-0.0007	0.0252	0.0202
		$\beta_2$	0.0028	0.0929	0.0742

NOTE: The SP2 results are for  $K = 2$ . The true values for the parameters are  $\gamma_1 = \beta_1 = \gamma_2 = \beta_2 = 1$ .

using the same design (DGP 1 with 100 replications), our SP2 parameter estimates for  $\gamma_1, \beta_1, \gamma_2$  and  $\beta_2$  are 0.999, 0.997, 0.998 and 1.014 with the corresponding standard deviations of 0.028, 0.086, 0.029 and 0.110. The results for DGP 2 (lognormal heterogeneity with negative correlation) are qualitatively similar across both estimation methods; results for SP2 but based on 1000 replications are given in Table 1. Even though DGPs 1 and 2 based on lognormal heterogeneity favor the Poisson-Lognormal model employed by Munkin and Trivedi (1999), the proposed SP2 approach performs very well.

We use the means of the dependent count variables of between 2.8 and 3.1 in DGPs 1 through 4 to mimic reasonable applications of count data in the existing literature. We also carried out additional simulations to explore examples of correlation bounds that might be encountered in practice. This is accomplished in two ways. First, we run two additional simulations to evaluate (i) the performance of the SP2 approach and (ii) the bounds on the estimated correlation when the conditional means are very low - in the neighborhood of 0.7. Second, we calculate the correlation bounds from the two-factor bivariate mixture count models with varying heterogeneity assumptions and means of the counts, based on Proposition 1 and equations 4 to 6. The means of the counts in this numerical exercise, which is not a Monte Carlo experiment, vary from 0.5 to about 40. This range captures most of the empirical applications.

The designs of the Monte Carlo simulation and calibration exercises as well as the results are provided in a Supplemental Appendix. The SP2 model again performs well in terms of bias and RMSE when the mean of the event counts is very low. However, as the mean becomes very small, the variance of the slope estimate tends to increase. The results also confirm that the magnitude of the correlation between the counts is narrower than the corresponding correlation from the distribution of unobserved heterogeneity components. As the means of the response variables increase, the correlation between the counts approaches the correlation between the unobservables. On the other hand, as the means of the response variables decrease, the gap between the two correlations

widens. This is potentially disconcerting since in many applications the means of the response variables may be low. However, the proposed series expansion approach does a good job of estimating the magnitude and sign of the correlations from the underlying data generating process. Next, we provide empirical applications involving event count variables with means ranging from as low as 0.3 to 3.2.

## 4 APPLICATION TO TOBACCO CONSUMPTION

As an example of an empirical application with negative correlation between counts, we consider an application to tobacco consumption behavior of individuals based on the 2001 household Tobacco Prevalence Survey data from Bangladesh. The Survey was conducted in two administrative districts of paramount interest for tobacco production and consumption in the country. Data on daily consumption of smoking- and chewing-tobacco along with other socioeconomic and demographic characteristics were collected from respondents of 10 years of age and above. The data set has been used previously by Gurmu and Yunus (2008) in the context of binary response models. Here we focus on a 32 percent random sample consisting of 4800 individual respondents.

We consider joint modeling of two count variables, the daily number of smoking tobacco (*Smoking*) and number of chewing tobacco (*Chewing*). The average number of smoking tobacco used daily is about 3.2; the mean number of chewing tobacco is about 1.1 per day. The mean and standard deviations of both response variables show unconditional overdispersion. The range is 0 to 50 tobacco consumed per day for each of the counts. The count variables are unconditionally negatively correlated with a sample correlation of about -0.022, indicating that smoking and chewing tobacco may on average be substitutes. It is therefore desirable to consider a joint modeling strategy that allows for both negative and positive correlations between the counts. The average respondent is a Muslim (78%), married (57%), in his/her early thirties, lives in rural area, and has about 7 years of formal schooling. Although the country is mostly

agrarian, only around 11% of the respondents were related to agricultural occupation in either doing agricultural operations on their own farms or working as agricultural wage laborers. About 12% of the respondents belong to the service occupation. The benchmark occupational group consists of business and other occupations. More than one-half of the fathers and slightly less than two-thirds of the mothers of the respondents currently use or have used tobacco products in the past. Detailed summary statistics are given in a Supplemental appendix.

Table 2 presents coefficient estimates from the proposed bivariate count data model. Except for *income* and quadratics in *age* and *education*, the rest of the regressors are binary dummy variables. The semiparametric model dominates the bivariate Poisson and bivariate negative binomial models in terms of both the maximized value of the log-likelihood function and the AIC. SP2 with  $K = 2$  yields the smallest AIC of 20117 relative to AIC values of 48061 and 28776 for bivariate Poisson and bivariate negative binomial models, respectively. The two-factor SP2 model also dominates the one-factor generalized bivariate negative binomial model analyzed by Gurmu and Elder (2000) in terms of the AIC. Given the high proportion of reported zeros where about 65 percent of respondents are non-users of tobacco, we also estimate a zero-inflated (ZI) version of the proposed SP2 model with  $K = 1$  (AIC = 19501). The estimates of the correlation parameters ( $\rho$ 's) associated with SP2 models show that modeling higher order polynomials of unobserved heterogeneity components is important.

The bivariate Poisson model seems to be inadequate for joint estimation of overdispersed count data. There are some differences in the results from the semiparametric and the bivariate negative binomial models. In particular, there are differences in the statistical significance of some variables such as occupational and regional dummy variables in both equations. Tobacco consumption is concave in age; tobacco smoking for example reaches a maximum at about age 50. Male respondents consume significantly more smoking tobacco than women, while women tend to consume more chewing tobacco, a result which is in line with the custom of the country (Gurmu and Yunus

Table 2: Semiparametric coefficient estimates and t-ratios for smoking and chewing tobacco (SP2 with  $K = 2$ )

Variable	<i>Smoking</i>		<i>Chewing</i>	
	Est.	<i>t</i>	Est.	<i>t</i>
<i>Constant</i>	-3.926***	6.86	-4.845***	10.18
<i>Age</i>	0.149***	4.34	0.179***	4.85
<i>Age squared</i>	-0.002***	4.20	-0.001***	3.25
<i>Education</i>	-0.150**	2.43	0.052	0.343
<i>Education squared</i>	0.006	1.41	-0.019	1.10
<i>Income</i>	-0.001	0.71	0.009	0.48
<i>Male</i>	3.631***	16.33	-0.482	1.52
<i>Married</i>	0.180	0.47	0.285	0.63
<i>Muslim</i>	-0.334*	1.70	-0.027	1.12
<i>Region</i>	-0.164	0.99	1.040***	5.90
<i>Urban</i>	-0.246*	1.70	-0.044	0.48
<i>Father use</i>	-0.017	0.56	-0.184	0.36
<i>Mother use</i>	-0.143	0.51	-0.415**	2.01
$\log(\alpha_j)$	-1.601***	30.38	-2.475***	41.76
$\rho_{11}$	-0.059***	5.50		
$\rho_{12}$	-0.016	0.49		
$\rho_{21}$	0.030**	2.53		
$\rho_{22}$	0.022***	8.04		
Log-likel.			-10016.6	
AIC			20143	

NOTE: \*\*\* significant at 1%; \*\* significant at 5%; \* significant at 10%. Regression includes occupation effects.

2008).

We also computed the average marginal effects of changes in the regressors on the mean number of daily tobacco use. The marginal effects from bivariate SP2, zero-inflated bivariate SP2, bivariate Poisson and bivariate negative binomial (NB) are shown in Table 3. Generally the marginal effects from SP2 models are larger than those from the other bivariate models. The difference in the mean effects for various models is likely to be particularly important for significant regressors. The results from bivariate SP2 shows that on average two additional years of schooling would reduce smoking by about a half stick of cigarette per month. On average men smoke about 7 cigarettes per day more than women.

The average of the correlations between *Smoking* and *Chewing* is about 0.55 for the bivariate NB model, 0.16 for Bivariate SP2 and 0.27 for ZI-bivariate SP2. Given the characteristics of the individuals, the correlation between the two tobacco consumption variables estimated from ZI-bivariate SP2 model varies between a minimum of about -0.003 and maximum of 0.932 with 16.2% of the 4800 observations having negative correlations. These results suggest that, smoking and chewing tobacco may be substitutes for some individuals, rather than being complements for everyone.

Compared with univariate and bivariate standard models, there is preponderance of statistical evidence (e.g., using AIC and tests of independence) in favor of the SP class of models of tobacco use. Direct comparisons of the unconditional sample moments and estimated conditional and average (unconditional) moments are generally difficult because of the role of explanatory variables and the sampling distributions of the estimated moments in the latter. For example, introduction of covariates will generally dampen the amount of conditional overdispersion. Subject to these caveats, comparisons of models show that the estimated average means, degree of overdispersion, and correlations from the SP2 model are generally consistent with the corresponding unconditional sample moments. The SP2 model fits the empirical distribution better than bivariate Poisson and negative binomial models.

Table 3: Average marginal effects and average of moments of the number of daily tobacco use

Variable	Bivar Poisson		Bivar NB		Bivar SP2		ZI-Bivar SP2	
	<i>Smoke</i>	<i>Chew</i>	<i>Smoke</i>	<i>Chew</i>	<i>Smoke</i>	<i>Chew</i>	<i>Smoke</i>	<i>Chew</i>
<i>Age</i>	0.003	0.005	0.005	0.007	0.007	0.008	-0.001	-0.0002
<i>Education</i>	-0.002	-0.0001	-0.005	-0.001	-0.007	0.001	-0.003	0.0003
<i>Income</i>	-0.002	0.009	-0.012	0.016	-0.005	0.014	-0.007	0.014
<i>Male</i>	5.809	-0.635	6.725	-0.654	7.268	-0.724	6.010	-0.689
<i>Married</i>	1.225	0.052	0.561	-0.020	0.749	0.400	0.905	0.452
<i>Muslim</i>	-0.238	-0.142	-0.550	-0.291	-1.625	-0.042	-0.752	-0.357
<i>Region</i>	-0.471	0.614	-0.847	1.238	-0.729	1.533	-1.004	1.212
<i>Urban</i>	-0.368	-0.137	1.016	-0.161	-1.041	-0.067	-0.520	0.021
<i>Agri-labor</i>	0.260	-0.044	0.283	0.250	0.658	0.411	0.371	0.075
<i>Service</i>	-0.231	-0.109	-0.180	0.237	0.407	0.835	-0.200	0.107
<i>Business</i>	0.610	-0.068	0.862	0.114	1.769	0.602	0.937	0.166
<i>Self employ</i>	0.649	0.194	0.983	0.542	0.623	1.048	0.546	0.496
<i>Student</i>	-2.919	-1.081	-2.990	-1.162	-3.877	-1.301	-1.505	-0.724
<i>Father use</i>	0.039	-0.017	0.448	-0.107	-0.076	-0.298	-0.047	-0.021
<i>Mother use</i>	-0.163	-0.452	-0.942	-0.787	-0.650	-0.708	-0.379	-0.169
Mean( $y_j$ ) <sup>a</sup>	3.215	1.134	3.678	1.308	4.424	1.556	3.378	1.228
Var( $y_j$ ) <sup>a</sup>	3.215	1.134	186.202	26.941	523.520	184.542	133.137	58.445
Corr( $y_1, y_2$ ) <sup>b</sup>	0.00003		0.548		0.159		0.265	

<sup>a</sup> Sample averages of fitted marginal moments of smoking and chewing tobacco.

<sup>b</sup> Sample average of fitted correlations between smoking and chewing tobacco. The range of estimated correlation from ZI-Bivar SP2 model is -0.003 to 0.932.

In an earlier version of this paper, we provide an application to health care utilization measures; also given in the Supplemental appendix. Though not overwhelming, there are some differences between the SP2 model and existing realistic competitors in the significance of the variables and estimated marginal effects in both applications. Univariate models, such as negative binomial model, that appropriately handle overdispersion are better than bivariate Poisson model. The usefulness of the bivariate negative binomial model in empirical applications is limited since it imposes correlation between series to be non-negative. In general, the analysis in this paper suggests that bivariate mixture models with a common unobserved heterogeneity component are inadequate in empirical applications, in that a single unobserved variable may not be able to account both for the dependence between the counts and for the variation in the dependent variables due to changes in unobservables.

To obtain satisfactory starting values for the mean parameters in both applications, we sequentially estimate standard bivariate count models, such as the bivariate Poisson and bivariate negative binomial models. Using estimates from these models as our starting values, the optimization algorithm for our model is typically able to achieve the usual gradient-based convergence criteria, resulting in substantial improvements in the log likelihood function and AIC. As we increase the degree of polynomial, we use starting values from lower order polynomial model. At each step, we re-initiate the optimization procedure using different starting values to start a fresh search for a global maximum. In the Monte-Carlo analysis, for small proportion of the replications, the optimization routine was restarted with slightly altered starting values. Although optimization algorithms cannot guarantee that they achieve the global optimum, we feel confident that our parameter estimates are valid given the plausibility and the qualitative aspects of our results.

## 5 CONCLUSION

This paper has developed a flexible two-factor bivariate count data regression model based on series expansion for the joint density of the unobserved heterogeneity components. Multivariate generalizations as well as extensions to accommodate truncated, censored and zero-inflated correlated count data models are provided. We obtain a computationally tractable closed form of the bivariate model that allows for both negative and positive correlations. Our empirical application and simulation experiments show that the proposed model fits various features of the data well and compares favorably with existing bivariate count data models.

The gain in the proposed SP2 model comes from suitably modeling overdispersion as well as correlation between variables within a flexible likelihood framework. As compared to the Poisson log-normal mixture model, the SP2 model relies less on distributional assumptions and provides a closed form of the likelihood function. Note that the dispersion ( $\alpha$ 's) and correlation ( $\rho$ 's) parameters also appear in the marginal distributions implied by the correlated SP2 model. When a test of the null hypothesis of independence is rejected, the SP2 model is expected to fit both the bivariate distribution as well as the implied marginals better. If the correlation between  $y_1$  and  $y_2$  is not significantly different from zero, the SP2 model is simply fitting the marginals better. However, if the test of independence is not rejected, it is more sensible to estimate flexible univariate models such as those considered by Cameron and Johansson (1997), Gurmu (1997) and Gurmu *et al.* (1999).

As in univariate models, the semiparametric approach for multivariate models is particularly attractive for censored, truncated and related modified models where parametric maximum likelihood estimates are not robust to misspecification of the error distribution. Monte Carlo results reported in Gurmu *et al.* (1999) show that biases and the sampling variances resulting from censored and truncated parametric models, such as the censored negative binomial, can be substantial. Their Monte Carlo results

from the regular (uncensored) sample show that, although the SP procedure has an edge over standard parametric models, the biases resulting from parametric and SP models with heterogeneity are not substantial. By contrast, the Monte Carlo results in this paper show that the biases and sampling variances from the standard (uncensored/untruncated) bivariate models (such as bivariate negative binomial model) may be substantial.

While the sign of the correlation between counts is unrestricted, the range of possible values from the Poisson-unobserved factor mixture models is generally narrower than that of the corresponding mixing distribution. The range of possible correlations can be quite high in absolute terms for large values of the means of the counts, but this is not attainable in some common applications with small means. An ideal extension of the method proposed in this paper is to have a multivariate generalization that achieves the full range of correlations between the counts.

## APPENDIX: DERIVATIONS AND GENERALIZATIONS

### A Derivations and some Expressions of the SP2 Model

We outline the derivation of the MGF based Laguerre series expansion. The relevant  $k$ -th order generalized Laguerre polynomial associated with the random variable  $\nu_{ji}$  is

$$L_{kj}(\nu_{ji}) = \sum_{l=0}^k \binom{k}{l} \frac{\Gamma(k + \alpha_j)}{\Gamma(l + \alpha_j)\Gamma(k + 1)} \lambda_j^l (-\nu_{ji})^l, \text{ for } j = 1, 2, \quad (\text{A.1})$$

with baseline gamma weight

$$w(\nu_{ji}) = \frac{\nu_{ji}^{\alpha_j-1} \lambda_j^{\alpha_j}}{\Gamma(\alpha_j)} e^{-\lambda_j \nu_{ji}}. \quad (\text{A.2})$$

These polynomials are orthogonal,  $E_{\nu_j} [L_{kj}(\nu_{ji})L_{lj}(\nu_{ji})] = 0$  for  $k \neq l$ , and each with unit variance,  $E_{\nu_j} [h_{kj}^{-1}L_{kj}^2(\nu_{ji})] = 1$ . Since

$$\left[ \sum_{k=0}^K \sum_{r=0}^K \rho_{kr} P_k(\nu_1) P_r(\nu_2) \right]^2 = \sum_{k=0}^K \sum_{r=0}^K \sum_{l=0}^K \sum_{m=0}^K \rho_{kr} \rho_{lm} P_k(\nu_1) P_l(\nu_1) P_r(\nu_2) P_m(\nu_2),$$

the constant of proportionality in equation (9) takes the form

$$\begin{aligned} \varpi &= \sum_{k=0}^K \sum_{r=0}^K \sum_{l=0}^K \sum_{m=0}^K \rho_{kr} \rho_{lm} E [P_k(\nu_1) P_l(\nu_1)] E [P_r(\nu_2) P_m(\nu_2)] \\ &= \sum_{k=0}^K \sum_{r=0}^K \rho_{kr}^2 P_k^2(\nu_1) P_r^2(\nu_2) \\ &= \sum_{k=0}^K \sum_{r=0}^K \rho_{kr}^2; \end{aligned}$$

the last 2 lines follow from the properties that  $E [P_k(\nu_j) P_l(\nu_j)] = 0$  for  $k \neq l$  by orthogonality of polynomials and each with unit variance ( $E [P_k^2(\nu_j)] = 1$ ). The approximate density will then take the form

$$g_N(\nu_{1i}, \nu_{2i}) = \frac{w(\nu_{1i})w(\nu_{2i})}{\varpi} \left[ \sum_{k=0}^K \sum_{r=0}^K \rho_{kr} h_{k1}^{-1/2} h_{r2}^{-1/2} L_{k1}(\nu_{1i}) L_{r2}(\nu_{2i}) \right]^2, \quad (\text{A.3})$$

with its MGF defined by  $M_N(t_{1i}, t_{2i}) = E_{\nu} [\exp(t_1 \nu_{1i} + t_2 \nu_{2i})]$ .

The main result we need is the MGF

$$M_N(-\theta_{1i}, -\theta_{2i}) = \iint \exp(-\theta_{1i} \nu_{1i} - \theta_{2i} \nu_{2i}) g_N(\nu_{1i}, \nu_{2i}) d\nu_{1i} d\nu_{2i},$$

where  $g_N(\cdot)$  is given in (A.3). Using properties of orthonormal polynomials and since the integration and expectations are taken with respect to the baseline marginal distributions, this MGF can be reorganized and expressed readily as

$$\begin{aligned} M_N(-\theta_{1i}, -\theta_{2i}) &= \frac{1}{\varpi} \sum_{k=0}^K \sum_{r=0}^K \sum_{l=0}^K \sum_{m=0}^K \rho_{kr} \rho_{lr} (h_{k1} h_{r2} h_{l1} h_{m2})^{-1/2} \\ &\quad \times E_{\nu_{1i}} \left[ L_{k1}(\nu_{1i}) L_{l1}(\nu_{1i}) e^{-\theta_{1i} \nu_{1i}} \right] \\ &\quad \times E_{\nu_{2i}} \left[ L_{r2}(\nu_{2i}) L_{m2}(\nu_{2i}) e^{-\theta_{2i} \nu_{2i}} \right]. \end{aligned} \quad (\text{A.4})$$

We exploit the properties of the gamma function,  $\Gamma(\gamma) = \int z^{\gamma-1} e^{-z} dz$ , and a two-parameter gamma density,  $\int z^{\gamma-1} e^{-z/\delta} dz = \Gamma(\gamma)\gamma^\delta$ , to obtain a closed form solution for this MGF. The approximate moment generating function corresponding to the  $i$ -th observation is

$$M_N(-\theta_{1i}, -\theta_{2i}) = \frac{1}{\sum_{k=0}^K \sum_{r=0}^K \rho_{kr}^2} \left( \prod_{j=1}^2 \left( 1 + \frac{\theta_{ji}}{\lambda_j} \right)^{-\alpha_j} \Gamma(\alpha_j) \right) \times \sum_{k,r,l,m=0}^K \rho_{kr} \rho_{lm} (h_{k1} h_{r2} h_{l1} h_{m2})^{1/2} \Psi_{1kl}^0 \Psi_{2rm}^0, \quad (\text{A.5})$$

where

$$\Psi_{1kl}^0 = \sum_{k_1=0}^k \sum_{l_1=0}^l \binom{k}{k_1} \binom{l}{l_1} \frac{\Gamma(\alpha_1 + k_1 + l_1)}{\Gamma(\alpha_1 + k_1) \Gamma(\alpha_1 + l_1)} \left( -1 - \frac{\theta_{1i}}{\lambda_1} \right)^{-(k_1+l_1)}$$

and

$$\Psi_{2rm}^0 = \sum_{r_2=0}^r \sum_{m_2=0}^m \binom{r}{r_2} \binom{m}{m_2} \frac{\Gamma(\alpha_2 + r_2 + m_2)}{\Gamma(\alpha_2 + r_2) \Gamma(\alpha_2 + m_2)} \left( -1 - \frac{\theta_{2i}}{\lambda_2} \right)^{-(r_2+m_2)}.$$

Detailed derivations for (A.5) are available in a Supplemental Appendix.

Next, we show the implications of restricting the mean of each unobserved heterogeneity component to unity. Using (10), set  $M_N^{(1,0)}(0,0) = 1$  and  $M_N^{(0,1)}(0,0) = 1$ . These restrictions yield

$$\lambda_1 = \frac{1}{\sum_{k=0}^K \sum_{r=0}^K \rho_{kr}^2} \left[ \prod_{j=1}^2 \Gamma(\alpha_j) \right] \sum_{k=0}^K \sum_{r=0}^K \rho_{kr} (h_{k1} h_{r2})^{1/2} \sum_{l=0}^K \sum_{m=0}^K \rho_{lm} (h_{l1} h_{m2})^{1/2} \Psi_{1kl}(0,1) \Psi_{2rm}(0,0), \quad (\text{A.6})$$

and

$$\lambda_2 = \frac{1}{\sum_{k=0}^K \sum_{r=0}^K \rho_{kr}^2} \left[ \prod_{j=1}^2 \Gamma(\alpha_j) \right] \sum_{k=0}^K \sum_{r=0}^K \rho_{kr} (h_{k1} h_{r2})^{1/2} \sum_{l=0}^K \sum_{m=0}^K \rho_{lm} (h_{l1} h_{m2})^{1/2} \Psi_{1kl}(0,0) \Psi_{2rm}(0,1). \quad (\text{A.7})$$

Here, for example,  $\Psi_{1kl}(0,1) = \Psi_{1kl}(-\theta_{1i} = 0, y_{1i} = 1)$ . Finally, the correlation coefficient for the SP2 model is:

$$\text{Corr}(y_{1i}, y_{2i} \mid \mathbf{x}_i) = \frac{\theta_{1i} \theta_{2i} \left[ M_N^{(1,1)}(0,0) - 1 \right]}{\sqrt{\left[ \theta_{1i} + \theta_{1i}^2 \left( M_N^{(2,0)}(0,0) - 1 \right) \right] \left[ \theta_{2i} + \theta_{2i}^2 \left( M_N^{(0,2)}(0,0) - 1 \right) \right]}}. \quad (\text{A.8})$$

This is obtained readily using the general results in proposition 1 and equation (7) with the mean of each unobserved heterogeneity component set to unity.

Multivariate generalizations as well as extensions to accommodate truncated, censored and zero-inflated correlated count data models are provided in Supplementary Appendix B. Shaw (1988), Grogger and Carson (1991()), Silva (1997) and Gurmu et al. (1999) provide background and applications of truncated and censored univariate count data models. See, for example, Wang (2003) and Gurmu and Elder (2008) for background, applications and recent developments in univariate zero-inflated models.

### **SUPPLEMENTAL MATERIAL**

Supplemental appendices B through G contain generalizations, additional derivations, results, tables and graphs for the applications and Monte Carlo experiments. These appendices and the GAUSS program used in this paper can also be downloaded from the authors' homepages: [http://www2.gsu.edu/~ecosgg/research/pdf/ge\\_jbes.pdf](http://www2.gsu.edu/~ecosgg/research/pdf/ge_jbes.pdf).

### **ACKNOWLEDGEMENTS**

The authors thank the co-editor, Keisuke Hirano, the associate-editor, two anonymous referees, and numerous seminar and conference participants for helpful comments and suggestions. Mohammad Yunus graciously provided the data used in this article. Any errors that remain are solely our responsibility.

### **References**

- Aitchison, J., and Ho C.H. (1989), "The Multivariate Poisson-log Normal Distribution," *Biometrika*, 76, 643-653.
- Bierens, H. (2008), "Semi-nonparametric Interval-Censored Mixed Proportional Hazard Models: Identification and Consistency Results," *Econometric Theory*, 24, 749-794.
- Cameron, C., and Johansson P. (1997), "Count Data Regressions Using Series Expansions: With Applications," *Journal of Applied Econometrics*, 12, 203-223.

- Cameron, C., Trivedi P.K. (1993), "Tests of Independence in Parametric Models with Applications and Illustrations," *Journal of Business & Economic Statistics*, 11, 29-43.
- Cameron, C., and Trivedi P.K. (1998), *Regression Analysis of Count Data*, Cambridge University Press.
- Chib, S. and R. Winkelmann (2001), "Markov Chain Monte Carlo Analysis of Correlated Count Data," *Journal of Business & Economic Statistics*, 19, 428-135.
- Gallant, A. R., and Nychka, W. (1987), "Semi-nonparametric Maximum Likelihood Estimation," *Econometrica*, 55, 363-390.
- Gallant, A. R., and Tauchen, G. (1989), "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," *Econometrica*, 57, 1091-1120.
- Grogger, J. and J. Carson (1991), "Models for Truncated Counts," *Journal of Applied Econometrics*, 6, 225-238.
- Gurmu, S. (1997), "Semi-parametric Estimation of Hurdle Regression Models with an Application to Medicaid Utilization," *Journal of Applied Econometrics*, 12, 225-242.
- Gurmu, S., and Elder, J. (2000), "Generalized Bivariate Count Data Regression Models," *Economics Letters*, 68, 31-36.
- Gurmu, S., and Elder, J. (2007), "A Simple Bivariate Count Data Regression Model," *Economics Bulletin* 3 (11), 1-10.
- Gurmu, S., and Elder, J. (2008), "A bivariate zero-inflated count data regression model with unrestricted correlation" *Economics Letter* 100, 245-248.
- Gurmu, S., Rilstone, P., and Stern, S. (1999), "Semiparametric Estimation of Count Regression Models," *Journal of Econometrics*, 88, 123-150.

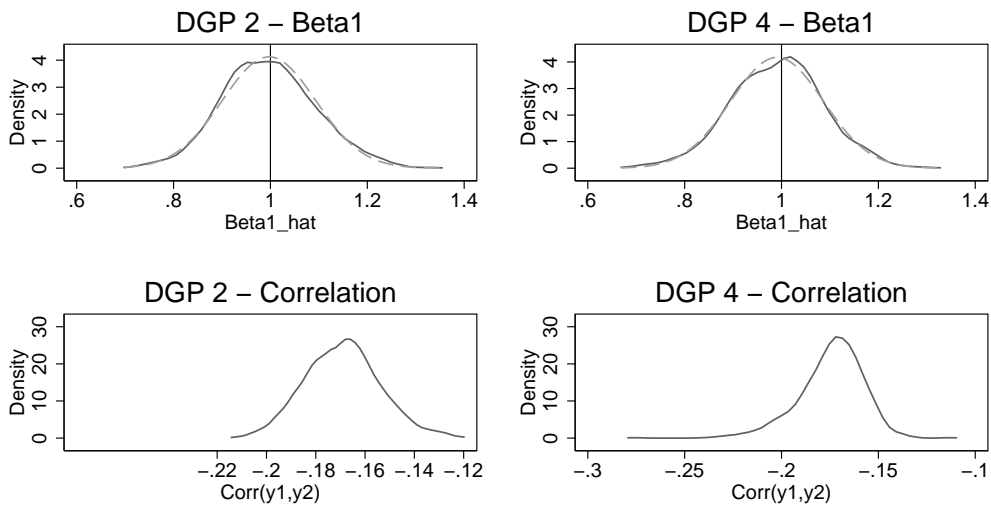
- Gurmu, S., and Trivedi, P. K. (1994), "Recent Developments in Event Count Models: A Survey," Thomas Jefferson Center Discussion Paper #261, Department of Economics, University of Virginia.
- Gurmu, S., and Yunus, M. (2008), "Tobacco Chewing, Smoking and Health Knowledge: Evidence from Bangladesh," *Economics Bulletin*, 9(10), 1-9.
- Hellstrom, J. (2006), "A Bivariate Count Data Model for Household Tourism Demand" *Journal of Applied Econometrics* 21, 213-226.
- Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997), *Discrete Multivariate Distribution*, New York: John Wiley.
- Jung, R. C., and Winkelmann, R. (1993), "Two Aspects of Labor Mobility: A Bivariate Poisson Regression Approach," *Empirical Economics*, 18, 543-556.
- Kocherlakota, S., and Kocherlakota, K. (1992), *Bivariate Discrete Distributions*, New York: Marcel Dekker.
- Lancaster, H.O. (1969), *The Chi-squared Distribution*, New York: Wiley.
- Mayer, W.J., and Chappell, W.F. (1992), "Determinants of Entry and Exit: An Application of the Compounded Poisson Distribution to US Industries, 1972-1977," *Southern Economic Journal*, 58, 770-778.
- Munkin, M., and Trivedi, P.K. (1999), "Simulated Maximum Likelihood Estimation of Multivariate Mixed-Poisson Regression Models, With Application," *Econometric Journal*, 1, 1-20.
- Ophem, H. van (1999), "A General Method to Estimate Correlated Discrete Random Variables," *Econometric Theory*, 15, 228-237.
- Shaw, D. (1988), "On-site Samples' Regression: Problems of Non-negative Integers, Truncation and Endogenous Stratification," *Journal of Econometrics*, 37, 211-233.

Silva, J.M.C. (1997), "Unobservables in Count Data Models for On-site Samples," *Economics Letters*, 54, 217-220.

Stephan, P. E., Gurmu, S., Sumell, A.J., and Black, G.C. (2007), "Who's Patenting in the University? Evidence from the Survey of Doctorate Recipients," *Economics of Innovation and New Technology*, 16, 71-99.

Wang, P. (2003), "A Bivariate Zero-inflated Negative Binomial Regression Model for Count Data with Excess Zeros," *Economics Letters*, 78, 373-378.

Figure 1. Kernel density estimates of distributions of coefficients and correlations  
DGPs 2 and 4 – SP2 K = 2, Sample size 1000



NOTE: The dashed lines in the top panel are the normal densities. For bottom panel, correlations were averaged over the sample observations prior to density estimation. The empirical average correlations between the count variables from DGPs 2 and 4 are  $-0.20$  and  $-0.27$ , respectively.

**Supplemental Materials Online to *Flexible***  
*Bivariate Count Data Regression Models*

Shiferaw Gurmu\* and John Elder†

October 29, 2011

---

\*Corresponding author: Department of Economics, Andrew Young School of Policy Studies,  
P.O. Box 3992, Georgia State University, Atlanta, GA 30302. Phone: (404) 413-0161; E-mail:  
sgurmu@gsu.edu

†Department of Finance & Real Estate, 1272 Campus Delivery, Colorado State University, Fort  
Collins, CO 80523

# Supplemental Appendices B through G: Derivations, Results, Tables and Graphs

*Contents:*

- **Appendix B:** Generalizations
- **Appendix C:** Derivation of the Moment Generating Function for the Mixing Distribution
- **Appendix D:** Tests of Independence in Bivariate Models
- **Appendix E:** Additional Monte Carlo and Related Results
- **Appendix F:** Additional Tables for Tobacco Data
- **Appendix G:** Application to Health-care Utilization

## **A Derivations and some Expressions of the SP2 Model**

Appendix A has been included in the main paper.

## **B Generalizations**

Apart from nesting some familiar baseline models, the approach based on series expansion also provides smooth estimation of the distribution of unobserved heterogeneity. In truncated and censored models, misspecification of the distribution of the unobserved heterogeneity leads to inconsistent parameter estimates. Accordingly the series expansion approach is particularly useful for truncated and censored models. Here, we extend the SP2 approach to the estimation of truncated and censored jointly dependent count regression models. We also provide multivariate generalizations and extension to accommodate excess zeros.

## B.1 Truncated and Censored Models

In these models, the ranges of the dependent variables may be constrained in much the same way as in the univariate count data models. The most common form of truncation is the omission of zeros. On-site surveys, such as interviewing people at the mall about the number of shopping trips, give rise to zero-truncated models. For discussion of related issues in models for on-site samples, see Shaw (1988), Grogger and Carson (1991) and Silva (1997). Censored samples may result when high counts are not observed, or may be imposed by survey design; see, for example, Gurmu *et al.* (1999). Thus, right censoring is the most common form in the analysis of univariate count models. Since the conditional mean in truncated and censored models depends on the density, the form of the density will affect the consistency of the regression parameters. Thus a misspecification of the distribution of the unobserved heterogeneity leads to inconsistency. It is therefore desirable to extend the semiparametric approach to the analysis of truncated and censored bivariate models.

Suppose the support of  $y = (y_1, y_2)$  is restricted to the region  $S^*$ . For example, for a zero truncated bivariate distribution,  $S^*$  is a set of positive integers in  $\mathfrak{R}^2$ . The density or probability mass function of the truncated distribution is

$$\frac{f(y_1, y_2; \delta)}{\phi}, \quad y \in S^*$$

where  $\delta$  is a parameter vector and  $\phi = \sum \sum_{y \in S^*} f(y_1, y_2; \delta)$ . Thus the log-likelihood function takes the general form

$$\mathcal{L}_T(\delta) = \sum_{i=1}^N [\ln f(y_{1i}, y_{2i}; \delta) - \ln \phi_i]. \quad (\text{B.1})$$

We focus on the empirically relevant case, models that are truncated at zero.

Consider the double truncation case, in which the zero class is missing for both dependent variables so that  $y_{ji} = 1, 2, 3, \dots$  for  $j = 1, 2$ . The approach of Section 2 can be extended to the estimation of zero-truncated bivariate regression models. To implement this for SP2, we need to obtain the normalization probability  $\phi_i$  associated

with the semiparametric mixture density (3) with  $M_N^{(y_1, y_2)}(-\theta_{1i}, -\theta_{2i})$  given in (10). The SP2 log-likelihood function for the double-truncated model is

$$\begin{aligned} \mathcal{L}_{sp2t}(\boldsymbol{\varphi}) &= \sum_{i=1}^N [y_{1i} \ln \theta_{1i} + y_{2i} \ln \theta_{2i} - \ln(\Gamma(y_{1i} + 1)) - \ln(\Gamma(y_{2i} + 1))] \\ &\quad + \ln M_N^{(y_1, y_2)}(-\theta_{1i}, -\theta_{2i}) - \ln \phi_i \Big], \end{aligned} \tag{B.2}$$

where  $\phi_i = 1 - M_N(-\theta_{1i}) - M_N(-\theta_{2i}) + M_N(-\theta_{1i}, -\theta_{2i})$ . Here  $M_N(-\theta_{ji})$  is the MGF of  $\nu_{ji}$  and  $M_N(-\theta_{1i}, -\theta_{2i})$  is given by (A.5) in Appendix A. The approach can easily be extended to the case where only a single-variable is truncated at zero. For example, if only  $y_j$  is truncated at zero, then  $\phi = 1 - f(y_j = 0)$ .

Since censoring is closely related to truncation, we provide only a sketch. Suppose that the bivariate counts are right censored at  $r = (r_1, r_2)$  so that  $y_{ji} = 1, 2, 3, \dots, r_j$  for  $j = 1, 2$ . If  $(y_{1i}, y_{2i}; \delta)$  denotes the complete bivariate density, the log-likelihood for the right-censored bivariate count model is

$$\mathcal{L}_c(\delta) = \sum_{i=1}^N d_i [\ln f(y_{1i}, y_{2i}; \delta)] + [1 - d_i] \ln \left[ 1 - \sum_{l=0}^{r_1-1} \sum_{m=0}^{r_2-1} f(y_{1i} = l, y_{2i} = m) \right],$$

where  $d_i = 1$  if  $y$  falls in the uncensored region, and  $d_i = 0$  otherwise. Thus, using the preceding equation, the log-likelihood for the right censored bivariate SP2 model can be constructed from (3) and (10).

## B.2 Zero-inflated Model and Other Extensions

The proposed bivariate count data model can be extended to allow for excess of zeros; see for example, Wang (2003) and Gurmu and Elder (2008) for background and recent developments. Consider the mixture density  $f(y_{1i}, y_{2i} | \mathbf{x}_i)$  of the bivariate SP2 model specified in (3) and (10). This can be accomplished by increasing the probability associated with the zero-zero state,  $(y_1 = 0, y_2 = 0)$ , and decreasing proportionately the values of the other joint probabilities. Define an indicator variable  $S_i = 1$  if  $(y_{1i} + y_{2i}) = 0$ , and  $S_i = 0$  otherwise. The ensuing zero-inflated bivariate SP2 density

is given as

$$h(y_{1i}, y_{2i} | x_i) = \begin{cases} \pi_i + (1 - \pi_i)f(y_{1i} = 0, y_{2i} = 0 | \mathbf{x}_i), & \text{if } S_i = 1 \\ (1 - \pi_i)f(y_{1i}, y_{2i} | \mathbf{x}_i) & \text{if } S_i = 0 \end{cases}, \quad (\text{B.3})$$

where the inflation parameter  $\pi_i$ ,  $0 \leq \pi_i < 1$ , can be parameterized in terms of vector  $z_i$  of observed factors using the logit form. In the empirical implementation of this model - reported in Section 4 above, the use of tobacco by the mother and father were included in  $\mathbf{z}$  of the inflation submodel, but were excluded from  $\mathbf{x}$  in the main parts of the smoking and chewing equations.

Multivariate extensions of the proposed semiparametric mixture approaches are feasible. Consider the case of  $J$  count variables where each  $y_{ji}$  is conditioned on the unobserved heterogeneity term  $\nu_{ji}$ . The unobserved heterogeneity components may be correlated. This gives rise to a multi-factor multivariate model. Without any loss of generality, let  $J = 3$ . Suppose  $(y_{ji} | \mathbf{x}_{ji}, \nu_{ji}) \sim \text{Poisson}(\theta_{ji}\nu_{ji})$  with  $(\nu_{1i}, \nu_{2i}, \nu_{3i})$  having a trivariate distribution  $g(\nu_{1i}, \nu_{2i}, \nu_{3i})$  in  $\mathfrak{R}_+^3$ . Techniques in multivariate density expansions may be used to approximate  $g(\cdot)$ . The ensuing mixture density takes a general form

$$f(y_{1i}, y_{2i}, y_{3i}; \boldsymbol{\varphi} | \mathbf{x}_i) = \left[ \prod_{j=1}^3 \frac{(\theta_{ji})^{y_{ji}}}{\Gamma(y_{ji} + 1)} \right] M_N^{(y_1, y_2, y_3)}(-\theta_{1i}, -\theta_{2i}, -\theta_{3i}),$$

where the MGF term associated with the trivariate distribution  $g(\nu_{1i}, \nu_{2i}, \nu_{3i})$  can be obtained in a manner analogous to the bivariate case. The approach can be extended to any arbitrary  $J$ . However, unlike one-factor multivariate models, if  $J$  is large, SPJ models based on multivariate mixing are computationally demanding. One solution is to assume that some of the unobserved heterogeneity components are common. These details depend largely on specific applications, and we do not pursue them further. In principle, multivariate generalizations of truncated and censored models are also possible.

## C Derivation of the Moment Generating Function for the Mixing Distribution

We provide additional details of the derivation of the MGF of the mixing density evaluated at  $(-\theta_{1i}, -\theta_{2i})$ ,  $M_N(-\theta_{1i}, -\theta_{2i})$ , given in A.5 of the paper. For convenience, equations in this web appendix are re-numbered as appropriate. Consider the approximate density in A.3

$$g_N(\nu_{1i}, \nu_{2i}) = \frac{w(\nu_{1i})w(\nu_{2i})}{\varpi} \left[ \sum_{k=0}^K \sum_{r=0}^K \rho_{kr} h_{k1}^{-1/2} h_{r2}^{-1/2} L_{k1}(\nu_{1i}) L_{r2}(\nu_{2i}) \right]^2, \quad (\text{C.1})$$

where for  $j = 1, 2$ :

$$w(\nu_{ji}) = \frac{\nu_{ji}^{\alpha_j-1} \lambda_j^{\alpha_j}}{\Gamma(\alpha_j)} e^{-\lambda_j \nu_{ji}}, \quad (\text{C.2})$$

$$\varpi = \sum_{k=0}^K \sum_{r=0}^K \rho_{kr}^2, \quad (\text{C.3})$$

$$L_{kj}(\nu_{ji}) = \sum_{l=0}^k \binom{k}{l} \frac{\Gamma(k + \alpha_j)}{\Gamma(l + \alpha_j) \Gamma(k + 1)} \lambda_j^l (-\nu_{ji})^l, \quad (\text{C.4})$$

$$h_{kj} = \frac{\Gamma(k + \alpha_j)}{\Gamma(\alpha_j) \Gamma(k + 1)}, \quad (\text{C.5})$$

$$P_k(\nu_{ji}) = h_{kj}^{-1/2} L_{kj}(\nu_{ji}). \quad (\text{C.6})$$

By properties of orthonormal polynomials, we have:

$$E_{\nu_j} [P_k(\nu_{ji})] = 0,$$

$$E_{\nu_j} [P_k(\nu_{ji}) P_l(\nu_{ji})] = 0, \quad \text{for } k \neq l \quad (\text{C.7})$$

$$E_{\nu_j} [P_k^2(\nu_{ji})] = 1. \quad (\text{C.8})$$

We also exploit the result that, for any random variable  $z \sim \text{Gamma}(\gamma, \delta)$  with mean  $E(z) = \gamma\delta$ ,

$$\int \frac{1}{\Gamma(\gamma)\gamma^\delta} z^{\gamma-1} e^{-z/\delta} dz = 1$$

so that

$$\int z^{\gamma-1} e^{-z/\delta} dz = \Gamma(\gamma)\gamma^\delta. \quad (\text{C.9})$$

The MGF of  $g_N(\nu_{1i}, \nu_{2i})$  at  $(t_{1i}, t_{2i}) = (-\theta_{1i}, -\theta_{2i})$  is

$$\begin{aligned} M_N(-\theta_{1i}, -\theta_{2i}) &= \iint \exp(-\theta_{1i}\nu_{1i} - \theta_{2i}\nu_{2i}) g_N(\nu_{1i}, \nu_{2i}) d\nu_{1i} d\nu_{2i} \\ &= \iint \exp(-\theta_{1i}\nu_{1i} - \theta_{2i}\nu_{2i}) \frac{w(\nu_{1i})w(\nu_{2i})}{\varpi} \\ &\quad \times \left[ \sum_{k=0}^K \sum_{r=0}^K \rho_{kr} h_{k1}^{-1/2} h_{r2}^{-1/2} L_{k1}(\nu_{1i}) L_{r2}(\nu_{2i}) \right]^2 d\nu_{1i} d\nu_{2i} \\ &= \iint \exp(-\theta_{1i}\nu_{1i} - \theta_{2i}\nu_{2i}) \frac{w(\nu_{1i})w(\nu_{2i})}{\varpi} \\ &\quad \times \sum_{k=0}^K \sum_{r=0}^K \sum_{l=0}^K \sum_{m=0}^K \left[ \rho_{kr} \rho_{lr} h_{k1}^{-1/2} h_{r2}^{-1/2} h_{l1}^{-1/2} h_{m2}^{-1/2} \right. \\ &\quad \left. \times L_{k1}(\nu_{1i}) L_{r2}(\nu_{2i}) L_{l1}(\nu_{1i}) L_{m2}(\nu_{2i}) \right] d\nu_{1i} d\nu_{2i}. \end{aligned}$$

The additional cross terms in moving from row 2 to row 3 above vanish due to the orthogonality property (C.7). Since the integration and expectations are taken with respect to the baseline marginal distributions, the above MGF can be reorganized and expressed readily as

$$\begin{aligned} M_N(-\theta_{1i}, -\theta_{2i}) &= \frac{1}{\varpi} \sum_{k=0}^K \sum_{r=0}^K \sum_{l=0}^K \sum_{m=0}^K \rho_{kr} \rho_{lr} (h_{k1} h_{r2} h_{l1} h_{m2})^{-1/2} \\ &\quad \times E_{\nu_{1i}} \left[ L_{k1}(\nu_{1i}) L_{l1}(\nu_{1i}) e^{-\theta_{1i}\nu_{1i}} \right] \\ &\quad \times E_{\nu_{2i}} \left[ L_{r2}(\nu_{2i}) L_{m2}(\nu_{2i}) e^{-\theta_{2i}\nu_{2i}} \right]. \end{aligned} \quad (\text{C.10})$$

Next, we will simplify the expectation associated with, say  $\nu_{1i}$ , using (C.2) and (C.4)

as follows:

$$\begin{aligned}
E_{\nu_{1i}} \left[ L_{k_1}(\nu_{1i}) L_{l_1}(\nu_{1i}) e^{-\theta_{1i} \nu_{1i}} \right] &= \int L_{k_1}(\nu_{1i}) L_{l_1}(\nu_{1i}) e^{-\theta_{1i} \nu_{1i}} w(\nu_{1i}) d\nu_{1i} \quad (\text{C.11}) \\
&= \int L_{k_1}(\nu_{1i}) L_{l_1}(\nu_{1i}) \frac{\nu_{1i}^{\alpha_1-1} \lambda_1^{\alpha_1}}{\Gamma(\alpha_1)} e^{-\lambda_1 \nu_{1i}} e^{-\theta_{1i} \nu_{1i}} d\nu_{1i} \\
&= \int \sum_{k_1=0}^k \sum_{l_1=0}^l \left[ \binom{k}{k_1} \frac{\Gamma(k+\alpha_1)}{\Gamma(k_1+\alpha_1)\Gamma(k+1)} \lambda_1^{k_1} (-\nu_{1i})^{k_1} \right. \\
&\quad \times \binom{l}{l_1} \frac{\Gamma(l+\alpha_1)}{\Gamma(l_1+\alpha_1)\Gamma(l+1)} \lambda_1^{l_1} (-\nu_{1i})^{l_1} \\
&\quad \times \left. \frac{\nu_{1i}^{\alpha_1-1} \lambda_1^{\alpha_1}}{\Gamma(\alpha_1)} e^{-\lambda_1 \nu_{1i}} e^{-\theta_{1i} \nu_{1i}} \right] d\nu_{1i} \\
&= \Gamma(\alpha_1) h_{k_1} h_{l_1} \int \sum_{k_1=0}^k \sum_{l_1=0}^l \left[ \binom{k}{k_1} \binom{l}{l_1} (-1)^{k_1+l_1} \right. \\
&\quad \times \frac{1}{\Gamma(k_1+\alpha_1)\Gamma(l_1+\alpha_1)} \lambda_1^{\alpha_1+k_1+l_1} \nu_{1i}^{\alpha_1+k_1+l_1-1} \\
&\quad \times \left. e^{-(\lambda_1+\theta_{1i})\nu_{1i}} \right] d\nu_{1i}.
\end{aligned}$$

By the property of gamma density (C.9),

$$\int \nu_{1i}^{\alpha_1+k_1+l_1-1} e^{-(\lambda_1+\theta_{1i})\nu_{1i}} d\nu_{1i} = \Gamma(\alpha_1+k_1+l_1) \left( \frac{1}{\lambda_1+\theta_{1i}} \right)^{\alpha_1+k_1+l_1} \quad (\text{C.12})$$

so that  $\lambda_1^{\alpha_1+k_1+l_1} (-1)^{k_1+l_1} \int \nu_{1i}^{\alpha_1+k_1+l_1-1} e^{-(\lambda_1+\theta_{1i})\nu_{1i}} d\nu_{1i}$  equals

$$\left( 1 + \frac{\theta_{1i}}{\lambda_1} \right)^{-\alpha_1} \left( -1 - \frac{\theta_{1i}}{\lambda_1} \right)^{-(k_1+l_1)} \Gamma(\alpha_1+k_1+l_1). \quad (\text{C.13})$$

Consequently C.11 simplifies as

$$\begin{aligned}
E_{\nu_{1i}} \left[ L_{k_1}(\nu_{1i}) L_{l_1}(\nu_{1i}) e^{-\theta_{1i} \nu_{1i}} \right] &= h_{k_1} h_{l_1} \left( 1 + \frac{\theta_{1i}}{\lambda_1} \right)^{-\alpha_1} \Gamma(\alpha_1) \quad (\text{C.14}) \\
&\quad \times \sum_{k_1=0}^k \sum_{l_1=0}^l \left[ \binom{k}{k_1} \binom{l}{l_1} \frac{\Gamma(\alpha_1+k_1+l_1)}{\Gamma(k_1+\alpha_1)\Gamma(l_1+\alpha_1)} \right. \\
&\quad \times \left. \left( -1 - \frac{\theta_{1i}}{\lambda_1} \right)^{-(k_1+l_1)} \right].
\end{aligned}$$

By analogy of the derivation of (C.14), we get:

$$\begin{aligned}
E_{\nu_{2i}} \left[ L_{r_2}(\nu_{2i}) L_{m_2}(\nu_{2i}) e^{-\theta_{2i}\nu_{2i}} \right] &= h_{r_2} h_{m_2} \left( 1 + \frac{\theta_{2i}}{\lambda_2} \right)^{-\alpha_2} \Gamma(\alpha_2) \\
&\times \sum_{r_2=0}^r \sum_{m_2=0}^m \left[ \binom{r}{r_2} \binom{m}{m_2} \frac{\Gamma(\alpha_2 + r_2 + m_2)}{\Gamma(\alpha_2 + r_2) \Gamma(\alpha_2 + m_2)} \right. \\
&\quad \left. \times \left( -1 - \frac{\theta_{2i}}{\lambda_2} \right)^{-(r_2+m_2)} \right].
\end{aligned} \tag{C.15}$$

Finally, plugging (C.14) and (C.15) into (C.10) and simplifying slightly gives

$$\begin{aligned}
M_N(-\theta_{1i}, -\theta_{2i}) &= \frac{1}{\sum_{k=0}^K \sum_{r=0}^K \rho_{kr}^2} \left( \prod_{j=1}^2 \left( 1 + \frac{\theta_{ji}}{\lambda_j} \right)^{-\alpha_j} \Gamma(\alpha_j) \right) \\
&\times \sum_{k=0}^K \sum_{r=0}^K \sum_{l=0}^K \sum_{m=0}^K \rho_{kr} \rho_{lr} (h_{k1} h_{r2} h_{l1} h_{m2})^{1/2} \Psi_{1kl}^0 \Psi_{2rm}^0,
\end{aligned} \tag{C.16}$$

where

$$\Psi_{1kl}^0 = \sum_{k_1=0}^k \sum_{l_1=0}^l \binom{k}{k_1} \binom{l}{l_1} \frac{\Gamma(\alpha_1 + k_1 + l_1)}{\Gamma(\alpha_1 + k_1) \Gamma(\alpha_1 + l_1)} \left( -1 - \frac{\theta_{1i}}{\lambda_1} \right)^{-(k_1+l_1)}$$

and

$$\Psi_{2rm}^0 = \sum_{r_2=0}^r \sum_{m_2=0}^m \binom{r}{r_2} \binom{m}{m_2} \frac{\Gamma(\alpha_2 + r_2 + m_2)}{\Gamma(\alpha_2 + r_2) \Gamma(\alpha_2 + m_2)} \left( -1 - \frac{\theta_{2i}}{\lambda_2} \right)^{-(r_2+m_2)}.$$

The MGF in (C.16) above is identical to equation (A.5) in the paper. This completes the proof of the main result for the MGF of the approximated bivariate density of the two-factor unobserved heterogeneity in equation (8) of the paper.

It is straightforward to find the cross-derivatives of  $M_N(-\theta_{1i}, -\theta_{2i})$ ,

$$\frac{\partial^{y_{1i}+y_{2i}} M_N(-\theta_{1i}, -\theta_{2i})}{\partial y_{1i}(-\theta_{1i}) \partial y_{2i}(-\theta_{2i})} \equiv M_N^{(y_{1i}+y_{2i})}(-\theta_{1i}, -\theta_{2i}),$$

by continuous differentiation involving a term of the type  $\left( 1 + \frac{\theta_{ji}}{\lambda_j} \right)^{-a_j}$  in equation (C.16), and making use of the definition  $\Gamma(a_j) = (a_j - 1) \Gamma(a_j - 1)$ . This process provides the result in equation (10) of the main paper.

## D Tests of Independence in Bivariate Models

Loosely speaking, the score test of independence between  $y_1$  and  $y_2$  is based on  $H_0 : E [P_k(y_{1i})P_m(y_{2i})] = 0$  for  $k, m = 1, 2, \dots, K$ , where  $P_q(y_{ji})$  denotes an orthogonal polynomial of degree  $q$  associated with variable  $y_j$ , and expectations are taken under the null hypothesis of independence. We focus on tests of zero correlations of order two ( $K = 2$ ) in the Poisson and negative binomial models.

The steps required to implement two variants of tests of independence in Poisson and negative binomial models are:

1. Assuming independence between  $y_1$  and  $y_2$ , obtain consistent estimates of  $\beta_j$  (and of  $\alpha_j$  in the case of the negative binomial model),  $j = 1, 2$ .
2. In Table D.1, let  $\theta_{ji} = \exp(x'_{ji}\beta_j)$ ,  $\varepsilon_{ji} = y_{ji} - \theta_{ji}$ ,  $b_{ji} = 1 + \frac{2\theta_{ji}}{\alpha_j}$ , and  $\sigma_{ji}^2 = \theta_{ji} + \frac{\theta_{ji}^2}{\alpha_j}$ , for  $j = 1, 2$ . To compute  $\tau_{21}^1$  in (D.1) based on Poisson pseudo-maximum likelihood, for example, we need  $P_2(y_{1i})P_1(y_{2i}) = ((y_{1i} - \theta_{1i})^2 - y_{1i})(y_{2i} - \theta_{2i})$ , evaluated at  $\hat{\beta}_1$  and  $\hat{\beta}_2$ . Using orthogonal polynomials based on Poisson and negative binomial models summarized in Table D.1, compute test statistic 1:

$$\tau_{km}^1 = \left( \sum_{i=1}^N P_k(y_{1i})P_m(y_{2i}) \right)^2 \left[ \sum_{i=1}^N (P_k(y_{1i})P_m(y_{2i}))^2 \right]^{-1} \quad \text{for } k, m = 1, 2. \quad (\text{D.1})$$

For example,  $\tau_{21}^1 = \left( \sum_{i=1}^N P_2(y_{1i})P_1(y_{2i}) \right)^2 \left[ \sum_{i=1}^N (P_2(y_{1i})P_1(y_{2i}))^2 \right]^{-1}$ .  $\tau_{km}^1$  is asymptotically distributed as  $\chi^2(1)$ . A large computed value of  $\tau_{km}^1$  results in the rejection of the null hypothesis. A sufficient condition for rejection of the null hypothesis of independence is that  $\tau_{11}^1$  be large.

3. Compute test statistic 2:

$$\tau_{km}^2 = \left( \sum_{i=1}^N Q_k(y_{1i})Q_m(y_{2i}) \right)^2 \left[ \left( \sum_{i=1}^N Q_k^2(y_{1i}) \right) \left( \sum_{i=1}^N Q_m^2(y_{2i}) \right) \right]^{-1} \quad \text{for } k, m = 1, 2, \quad (\text{D.2})$$

where, for example,  $Q_k(y_{2i})$  denotes the  $k$ -th order polynomial having a unit variance (see Table D.1).  $\tau_{km}^2$  also converges to a  $\chi^2(1)$  distribution.

Table D.1 Orthogonal Polynomials for Tests of Independence

Polynomial	Poisson	Negative Binomial
$P_1(y_{ji})$	$\varepsilon_{ji}$	$\varepsilon_{ji}$
$P_2(y_{ji})$	$\varepsilon_{ji}^2 - y_{ji}$	$\varepsilon_{ji}^2 - b_{ji}\varepsilon_{ji} - \sigma_{ji}^2$
$Q_1(y_{ji})$	$\frac{P_1(y_{ji})}{\sqrt{\theta_{ji}}}$	$\frac{P_1(y_{ji})}{\sqrt{\sigma_{ji}^2}}$
$Q_2(y_{ji})$	$\frac{P_2(y_{ji})}{(\sqrt{2})\theta_{ji}}$	$\frac{P_2(y_{ji})}{\sqrt{\sigma_{ji}^2 + 2\sigma_{ji}^4 + (6\sigma_{ji}^4/\alpha_j) - b_{ji}^2\sigma_{ji}^2}}$

The estimated variances of these tests are based on the outer-products of the gradients. The first test,  $\tau_{km}^1$ , requires the correct specification of the first  $k$  moments of  $y_{1i}$  and  $m$  moments of  $y_{2i}$ . Similarly,  $\tau_{km}^2$  requires the correct specification of the first  $2k$  moments of  $y_{1i}$  and  $2m$  moments of  $y_{2i}$ . In the empirical application involving smoking and chewing tobacco, the p-values for the first-order test  $\tau_{11}^1$  is 0.001 for Poisson and 0.104 for negative binomial. Moreover, most of the p-values corresponding to higher-order tests of independence are low. For example, the respective p-values corresponding to  $\tau_{12}^1$ ,  $\tau_{21}^1$ , and  $\tau_{22}^1$  are 0.002, 0.000, and 0.000 for negative binomial model, and 0.131, 0.080, and 0.000 for Poisson model.

## E Additional Monte Carlo and Related Results

This section presents additional tables and graphs from the Monte Carlo experiments from DGP 1 through 4 as well as results from two new simulation experiments described below. We also provide results from a numerical exercise showing the range of correlations the dependent variables could take, particularly as the means of the count

variables change and with focus on the empirically relevant cases of small means. The contents of tables and figures for this supplemental appendix are as follows.

- Table E.1: Monte Carlo results from DGPs 3 and 4 (Poisson - bivariate uniform mixture model) using bivariate negative binomial and bivariate SP2 estimators.
- Table E.2: Monte Carlo results from additional experiments with very low means (DGPs 5 and 6) using bivariate SP2 estimator.
- Table E.3: Monte Carlo results from DGPs 1 through 6 - summary statistics of correlations between the counts estimated by SP2.
- Table E.4: Correlation bounds for two-factor mixture models with varying heterogeneity parameters and means of the counts (Positive correlations).
- Table E.5: Correlation bounds for two-factor mixture models with varying heterogeneity parameters and means of the counts (Negative correlations).
- Figure E.1: Kernel density estimates of the distributions of  $\hat{\beta}_1$  and  $\hat{\beta}_2$  (estimates of the slope coefficients in MC designs) from DGPs 1 and 2 (Poisson-Bivariate lognormal mixture model) using SP2 estimates.
- Figure E.2: Kernel density estimates of the distributions of  $\hat{\beta}_1$  and  $\hat{\beta}_2$  (estimates of the slope coefficients in MC designs) from DGPs 3 and 4 (Poisson-Bivariate uniform mixture model) using SP2 estimates.
- Figure E.3: Kernel density estimates of the distributions of the correlations (at the means of covariates) between the response variables. Graphs obtained from results of MC experiments from DGPs 1 through 4 using SP2 estimators.
- Figure E.4: Kernel density estimates of the distributions of the inverse hyperbolic tangent of the correlations from Figure E3, with normal density added.

Table E.1 Monte Carlo results on parameter estimates by bivairate count models

DGP: Bivar	Est. Method	Parameter	Mean	RMSE	Mean
heterogeneity			bias		abs. bias
3. Uniform, $\text{Corr}(\nu_1, \nu_2) = 0.6$					
	Bivar NB	$\gamma_1$	0.0187	0.1185	0.0884
		$\beta_1$	0.6552	0.7021	0.6556
		$\gamma_2$	0.0201	0.1187	0.0879
		$\beta_2$	0.6696	0.7193	0.6697
	Bivar SP2 $K = 1$	$\gamma_1$	-0.0187	0.0328	0.0264
		$\beta_1$	-0.0184	0.0991	0.0783
		$\gamma_2$	-0.0177	0.0323	0.0260
		$\beta_2$	-0.0107	0.1012	0.0802
	Bivar SP2 $K = 2$	$\gamma_1$	-0.0086	0.0299	0.0238
		$\beta_1$	-0.0074	0.0984	0.0780
		$\gamma_2$	0.0077	0.0210	0.0241
		$\beta_2$	0.0006	0.1021	0.0813

NOTE: The true values for the parameters are  $\gamma_1 = \beta_1 = \gamma_2 = \beta_2 = 1$ .

Table E.1 Continued

DGP: Bivar	Est. Method	Parameter	Mean	RMSE	Mean
heterogeneity			bias		abs. bias
4. Uniform, $\text{Corr}(\nu_1, \nu_2) = -0.6$					
	Bivar NB	$\gamma_1$	0.0029	0.0443	0.0349
		$\beta_1$	0.3591	0.4038	0.3621
		$\gamma_2$	0.0037	0.0426	0.0333
		$\beta_2$	0.3654	0.4127	0.3673
	Bivar SP2 $K = 1$	$\gamma_1$	-0.0056	0.0276	0.0220
		$\beta_1$	-0.0145	0.0992	0.0792
		$\gamma_2$	-0.0051	0.0270	0.0213
		$\beta_2$	-0.0115	0.1035	0.0815
	Bivar SP2 $K = 2$	$\gamma_1$	-0.0024	0.0273	0.0217
		$\beta_1$	-0.0092	0.0961	0.0762
		$\gamma_2$	-0.0017	0.0264	0.0209
		$\beta_2$	-0.0055	0.0999	0.0790

NOTE: The true values for the parameters are  $\gamma_1 = \beta_1 = \gamma_2 = \beta_2 = 1$ .

Figures E.1 through E.4 are given at the end of this document. In contrast with estimated correlations from SP2 shown in figures E.3 and E.4, the distribution of estimated correlations from bivariate negative binomial model shifts to the right significantly in all cases and, as expected, the model predicts positive correlations for DGPs (2 and 4) with negative correlations. The estimated means (*standard deviations*) of  $\text{corr}(y_1, y_2)$  from bivariate NB are 0.50 (0.04), 0.34 (0.04), 0.57 (0.05), and 0.32 (0.04) based on DGPs 1, 2, 3, and 4, respectively. The respective empirical means of  $\text{corr}(y_1, y_2)$  for DGPs 1, 2, 3, and 4 are 0.24, -0.20, 0.27 and -0.27.

We use the means of the dependent count variables of between 2.8 and 3.1 in DGPs 1 through 4 to mimic reasonable applications of count data in the existing literature. We also run two additional simulations to (i) evaluate the performance of the SP2 approach and (ii) gauge the correlation bounds when the conditional means are relatively low, about 0.7. For both DGPs (called DGPs 5 and 6), the mean parameters of Poisson distributed count variables  $y_1$  and  $y_2$  are specified as in equation (14) of the paper. The values of explanatory variables  $x_1$  and  $x_2$  are generated independently from normal  $N(0, 1/16)$ , with true values for the regression parameters set to  $\gamma_1 = \gamma_2 = -0.5$  and  $\beta_1 = \beta_2 = 1$  in both experiments. The unobserved heterogeneity terms  $\nu_1$  and  $\nu_2$  are generated by bivariate lognormal distributions using the following specifications, along the implied average characteristics of the DGP.

5. Bivariate log-normal with positive correlation:  $\log(\nu_1) = \varepsilon_1$  and  $\log(\nu_2) = \varepsilon_2$  follow bivariate normal distribution with mean  $(0, 0)$ , variances  $\sigma_1^2 = \sigma_2^2 = 0.25$  and correlation parameter  $\rho_\varepsilon = 0.6$ . This gives  $\text{mean}(\nu_j) = 1.12$ ,  $\text{var}(\nu_j) = 0.37$  and  $\text{corr}(\nu_1, \nu_2) = 0.57$  and average moments:  $\text{mean}(y_j) = 0.71$ ,  $\text{var}(y_j) = 0.99$  and  $\text{corr}(y_1, y_2) = 0.11$ .
6. Bivariate log-normal with negative correlation:  $\log(\nu_1) = \varepsilon_1$  and  $\log(\nu_2) = \varepsilon_2$  follow bivariate normal distribution with mean  $(0, 0)$ , variances  $\sigma_1^2 = \sigma_2^2 = 0.25$  and correlation parameter  $\rho_\varepsilon = -0.6$ . This implies  $\text{mean}(\nu_j) = 1.12$ ,  $\text{var}(\nu_j) = 0.37$

Table E.2 Monte Carlo results on parameter estimates by SP2 with K=1

DGP: Bivar heterogeneity	Parameter	Mean bias	RMSE	Mean abs. bias
5. Lognormal, $\text{Corr}(\nu_1, \nu_2) = 0.57$				
	$\gamma_1$	-0.0027	0.0426	0.0338
	$\beta_1$	-0.0017	0.1719	0.1385
	$\gamma_2$	-0.0031	0.0413	0.0328
	$\beta_2$	0.0101	0.1711	0.1354
6. Lognormal, $\text{Corr}(\nu_1, \nu_2) = -0.49$				
	$\gamma_1$	-0.0026	0.0425	0.0338
	$\beta_1$	-0.0011	0.1729	0.1392
	$\gamma_2$	-0.0027	0.0414	0.0324
	$\beta_2$	0.0118	0.1685	0.1343

NOTE: The true values for the parameters are  $\gamma_j = -0.5$   
and  $\beta_j = 1$ , for  $j = 1, 2$ .

and  $\text{corr}(\nu_1, \nu_2) = -0.49$  and the average moments:  $\text{mean}(y_j) = 0.71$ ,  $\text{var}(y_j) = 0.99$  and  $\text{corr}(y_1, y_2) = -0.09$ .

Monte Carlo results from these new DGPs using our proposed SP2 with  $K = 1$  are shown in Table E.2. The biases of the coefficients are generally very low. As compared to results from DGPs with higher means, the variance of each slope estimate is higher (with an increase in RMSE of about 6%). Combining simulation experiments from DGPs 1 through 6, Table E.3 reports on the reliability of the proposed SP2 approach in correctly predicting both the signs and the magnitudes of the correlations allowed by the underlying DGPs.

On the other hand, using Proposition 1 and equations 4 to 6 in the main paper, tables E.4 and E.5 give correlation bounds from the two-factor bivariate mixture count models with varying heterogeneity parameters and means of the counts. The means of

Table E.3 Monte Carlo results for correlation estimates from SP2

DGP: Bivar heterogeneity	Mean( $y_j$ )	Expected corr		SP2 estimate of corr( $y_1, y_2$ )			
		( $\nu_1, \nu_2$ )	( $y_1, y_2$ )	Mean	SD	5%	95%
1. Lognormal	3.08	0.57	0.24	0.21	0.03	0.16	0.25
2. Lognormal	3.08	-0.49	-0.20	-0.17	0.02	-0.19	-0.14
3. Uniform	2.81	0.60	0.27	0.29	0.06	0.21	0.39
4. Uniform	2.81	-0.60	-0.27	-0.18	0.02	-0.20	-0.15
5. Lognormal	0.71	0.57	0.11	0.06	0.02	0.03	0.09
6. Lognormal	0.71	-0.49	-0.09	-0.05	0.01	-0.07	-0.03

NOTE: Results for DGPs 1 to 4 are from SP2 with  $K = 2$ ; those for DGPs 5 and 6 are from SP2 with  $K = 1$ . DGP moments in columns 2-4 are evaluated at the average of  $x_j$ . SD denotes standard deviation and the figures in the last two columns are the 5 and 95 percentiles.

the counts in this calibration exercise (not a Monte Carlo experiment) vary from 0.5 to about 40 such that  $E(y_j) = \theta_j E(\nu_j)$ , where  $\theta_j$  is a constant (no regressors). The range for mean counts captures most empirical applications using count data. For example, while the mean number of applications such as health care consultation trips, number of shopping trips and number of patents issued to individual scientists is quite low (Cameron et al., 1988; Gurmu et al. 1999; Stephan et al. 2007), applications such as the number of annual patents granted to firms have relatively large means of about 25 to 40 (e.g., Hausman et al., 1984). In specifications, denoted specs, (a) through (h) below, the moments of the bivariate unobserved heterogeneity (UH) used in generating the correlation bounds are consistent with parametrization from bivariate lognormal and uniform distributions as follows, with implied moments given in tables E.4 and E5.

- a. Bivariate log-normal such that  $\log(\nu_1) = \varepsilon_1$  and  $\log(\nu_2) = \varepsilon_2$  follow bivariate normal distribution with mean  $(0, 0)$ , variances  $\sigma_1^2 = \sigma_2^2 = 0.25$  and correlation

parameter  $\rho_\varepsilon = 0.6$  ( $\text{Corr}(\nu_1, \nu_2) = 0.57$ )

- b. Bivariate log-normal such that  $\log(\nu_1) = \varepsilon_1$  and  $\log(\nu_2) = \varepsilon_2$  follow bivariate normal distribution with mean  $(0, 0)$ , variances  $\sigma_1^2 = \sigma_2^2 = 0.36$  and correlation parameter  $\rho_\varepsilon = 0.85$  ( $\text{Corr}(\nu_1, \nu_2) = 0.82$ ).
- c. Bivariate uniform with correlation  $\text{corr}(\nu_1, \nu_2) = 0.6$ .
- d. Bivariate uniform with correlation  $\text{corr}(\nu_1, \nu_2) = 0.85$ .
- e. Bivariate log-normal such that  $\log(\nu_1) = \varepsilon_1$  and  $\log(\nu_2) = \varepsilon_2$  follow bivariate normal distribution with mean  $(0, 0)$ , variances  $\sigma_1^2 = \sigma_2^2 = 0.25$  and correlation parameter  $\rho_\varepsilon = -0.6$  ( $\text{Corr}(\nu_1, \nu_2) = -0.49$ )
- f. Bivariate log-normal such that  $\log(\nu_1) = \varepsilon_1$  and  $\log(\nu_2) = \varepsilon_2$  follow bivariate normal distribution with mean  $(0, 0)$ , variances  $\sigma_1^2 = \sigma_2^2 = 0.36$  and correlation parameter  $\rho_\varepsilon = -0.85$  ( $\text{Corr}(\nu_1, \nu_2) = -0.61$ )
- g. Bivariate uniform with correlation  $\text{corr}(\nu_1, \nu_2) = -0.6$ .
- h. Bivariate uniform with correlation  $\text{corr}(\nu_1, \nu_2) = -0.85$ .

The results reported in tables E.4 and E.5 provide numerical evidence of the correlation bounds given in Proposition 1 and equation 7 of the main paper. The magnitudes are narrower than the corresponding correlations from distributions of unobserved heterogeneity components. As the means of the response variables increase, the correlation between the counts approaches the correlation between the unobservables. On the other hand, as the means of the response variables decrease, the gap between the two correlations widens. This is potentially disconcerting since in many applications the means of the response variables may be low. However, as reported in Table E.3 and figures E.3 and E.4, the proposed series expansion approach does a good job of estimating the magnitude and sign of the correlations from the underlying data generating process

Table E.4 Correlation bounds for two-factor mixture models with varying heterogeneity parameters and means of the counts (Positive correlations)

Spec for Bivar heterogeneity	Mean( $y_j$ )	Var( $y_j$ )	Corr( $y_1.y_2$ )
a) Mean( $\nu_j$ ) = 1.13, Var( $\nu_j$ ) = 0.36, Corr( $\nu_1, \nu_2$ ) = <b>0.567</b>	0.57	0.66	0.079
	1.13	1.50	0.139
	2.27	3.73	0.223
	4.53	10.37	0.321
	11.33	47.80	0.435
	33.99	362.22	0.516
	45.33	628.84	0.529
b) Mean( $\nu_j$ ) = 1.20, Var( $\nu_j$ ) = 0.62, Corr( $\nu_1, \nu_2$ ) = <b>0.826</b>	0.60	0.76	0.170
	1.20	1.82	0.282
	2.39	4.88	0.421
	4.79	14.73	0.503
	11.97	74.08	0.693
	35.92	59.91	0.776
	47.89	1041.65	0.788

NOTE: Parameterization of moments of UH in specs (a) and (b) are based on bivariate lognormal; see text in Supplemental Appendix E.

Table E.4 Bounds for positive correlations (Continued)

Spec for Bivar heterogeneity	Mean( $y_j$ )	Var( $y_j$ )	Corr( $y_1, y_2$ )
c) Mean( $\nu_j$ ) = 1, Var( $\nu_j$ ) = 0.33, Corr( $\nu_1, \nu_2$ ) = <b>0.6</b>			
	0.5	0.58	0.086
	1.0	1.33	0.150
	2.0	3.33	0.240
	4.0	9.33	0.343
	10.0	43.33	0.462
	30.0	330.00	0.545
	40.0	573.33	0.558
d) Mean( $\nu_j$ ) = 1, Var( $\nu_j$ ) = 0.33, Corr( $\nu_1, \nu_2$ ) = <b>0.850</b>			
	0.5	0.58	0.121
	1.0	1.33	0.213
	2.0	3.33	0.340
	4.0	9.33	0.485
	10.0	43.33	0.654
	30.0	330.00	0.739
	40.0	573.33	0.791

NOTE: Parameterization of moments of UH

in specs (c) and (d) are based on bivariate uniform;

see text in Supplemental Appendix E.

Table E.5 Correlation bounds for two-factor mixture models with varying heterogeneity parameters and means of the counts (Negative correlations)

Spec for Bivar heterogeneity	Mean( $y_j$ )	Var( $y_j$ )	Corr( $y_1.y_2$ )
e) Mean( $\nu_j$ ) = 1.13, Var( $\nu_j$ ) = 0.36,			
Corr( $\nu_1, \nu_2$ ) = <b>-0.490</b>			
	0.57	0.66	-0.068
	1.13	1.50	-0.119
	2.27	3.73	-0.192
	4.53	10.37	-0.276
	11.33	47.80	-0.374
	33.99	362.22	-0.444
	45.33	628.84	-0.455
f) Mean( $\nu_j$ ) = 1.20, Var( $\nu_j$ ) = 0.62,			
Corr( $\nu_1, \nu_2$ ) = <b>-0.608</b>			
	0.60	0.76	-0.126
	1.20	1.82	-0.208
	2.39	4.88	-0.310
	4.79	14.73	-0.411
	11.97	74.08	-0.510
	35.92	59.91	-0.572
	47.89	1041.65	-0.580

NOTE: Parameterization of moments of UH

in specs (e) and (f) are based on bivariate uniform;

see text in Supplemental Appendix E.

Table E.5 Bounds for negative correlations (Continued)

Spec for Bivar heterogeneity	Mean( $y_j$ )	Var( $y_j$ )	Corr( $y_1.y_2$ )
g) Mean( $\nu_j$ ) = 1, Var( $\nu_j$ ) = 0.33,			
Corr( $\nu_1, \nu_2$ ) = <b>-0.6</b>			
	0.5	0.58	-0.086
	1.0	1.33	-0.150
	2.0	3.33	-0.240
	4.0	9.33	-0.343
	10.0	43.33	-0.462
	30.0	330.00	-0.545
	40.0	573.33	-0.558
h) Mean( $\nu_j$ ) = 1, Var( $\nu_j$ ) = 0.33,			
Corr( $\nu_1, \nu_2$ ) = <b>-0.850</b>			
	0.5	0.58	-0.121
	1.0	1.33	-0.213
	2.0	3.33	-0.340
	4.0	9.33	-0.485
	10.0	43.33	-0.654
	30.0	330.00	-0.739
	40.0	573.33	-0.791

NOTE: Parameterization of moments of UH

in specs (g) and (h) are based on bivariate lognormal;

see text in Supplemental Appendix E.

Table F.1 Definition, mean and standard deviation of variables for tobacco data

Name	Definition	Mean	St. Dev.
<i>Smoking</i>	Number of smoking tobacco used daily	3.215	6.867
<i>chewing</i>	Number of chewing tobacco used daily	1.134	3.640
<i>Age</i>	Age in years	30.350	14.875
<i>Education</i>	Years of formal schooling	6.896	4.679
<i>Income</i>	Monthly family income in '000 of Taka	7.572	10.131
<i>Male</i>	=1 if male	0.546	0.498
<i>Married</i>	=1 if married	0.574	0.495
<i>Muslim</i>	= 1 if religion is Islam	0.787	0.410
<i>Region</i>	=1 if region is Rangapur	0.499	0.500
<i>Urban</i>	=1 if urban	0.380	0.486
<i>Agri-labor</i>	=1 if agriculture labor occupation	0.110	0.312
<i>Service</i>	=1 if service occupation	0.125	0.331
<i>Business</i>	=1 if business occupation	0.130	0.336
<i>Self employ</i>	=1 if self-employed or household chores	0.305	0.460
<i>student</i>	=1 if student	0.268	0.443
<i>Wage labor</i>	=1 if wage labor occupation	0.063	0.243
<i>Father use</i>	=1 if father uses tobacco	0.544	0.498
<i>Mother use</i>	=1 if mother uses tobacco	0.649	0.477

## F Additional Tables for Tobacco Data

Table F.1 gives detailed definitions, means and standard deviations of variables for tobacco data. Table F.2 reports the empirical versus fitted frequencies (%) from standard bivariate models and 2 estimators using SP2 approach.

Table F.2 Empirical and Fitted Joint Percentage Frequencies from Bivariate Models

Counts	Empirical	Bivar	Bivar	SP2	ZI-SP2
<i>Smoke, Chew</i>		Poisson	NB	K=2	K=1
0, 0	65.9	29.7	62.0	65.6	64.9
0, 1-6	4.7	15.0	7.2	6.4	4.7
0, >6	5.6	1.4	0.9	2.4	2.0
1-6, 0	3.9	18.7	12.9	14.5	13.3
7-12, 0	9.1	5.7	1.7	2.8	3.5
13-24, 0	5.3	1.2	0.8	2.4	2.9
Others	5.5	28.3	14.6	6.0	8.7

## G Application to Health-care Utilization

Models of health-care utilization have been used to estimate relationships of economic interest, and to predict outcomes such as the probability of not utilizing doctor consultations or emergency room service. Here we analyze jointly two health utilization measures, the number of consultations with a doctor and the number of non-doctor consultations. The data were originally employed by Cameron *et al* (1988) in their analysis of various measures of health-care utilization using a sample of 5190 single-person households from the 1977-78 Australian Health Survey. Here we model two possibly correlated dependent variables: (1) the number of consultations with a doctor during the two-week period prior to the survey (*Doctor*), and (2) the number of consultations with non-doctor health professionals (chemist, optician, physiotherapist, social worker, district community nurse, chiropodist or chiropractor) during the past 4 weeks (*Nondoc*). The data are obtained from the *Journal of Applied Econometrics* June 1997 Data Archive. These health utilization measures have two interesting features - overdispersion and very high proportion of nonusers. The mean and the standard deviation of doctor visits are 0.302 and 0.798. The corresponding values for health professionals are 0.215 and 0.965. The frequencies of zero visits in the *Doctor* and *Nondoc*

samples are 80% and 91%, respectively. The sample correlation coefficient between the two series is 0.148.

The explanatory variables are:

- Socio-economic variables - a dummy variable for whether female (*Sex*), age in years divided by 100 (*Age*), age-squared (*Agesq*), and annual income in ten-thousands of dollars (*Income*).
- Insurance and health status variables - indicator variable for private insurance coverage (*Levyplus*), free government insurance cover due to low income (*Freeepoor*), free government coverage due to old age, disability or veteran status (*Freerepa*), default government Medibank insurance cover paid for by income levy (*Levy*), number of illness in the past two weeks (*Illness*), number of days reduced activity in past two weeks due to illness or injury (*Actdays*), general health questionnaire score using Goldberg's method with high score indicating bad health (*Hscore*), indicator variable for chronic condition not limiting activity (*Chond1*), and indicator variable for chronic condition limiting activity (*Chond2*).

See Cameron *et al.* (1988) for summary statistics of the covariates.

Given that the 2 dependent variables *Doctor* and *Nondoc* have been analyzed previously using univariate models, we shall focus on results from the bivariate regressions. An important consideration is whether the two health utilization variables are independent or not. We carried out conditional moment tests of independence in the Poisson and negative binomial models. The p-values of the first order tests of independence for *Doctor* and *Nondoc* are all below a significance level of 2%, which is a sufficient condition to reject the test of independence in each case. Moreover, most of the p-values corresponding to higher-order tests of independence are low. Overall, there is an overwhelming evidence that *Doctor* and *Nondoc* are dependent counts, and hence joint estimation is desirable.

Table G.1 presents coefficient estimates from the proposed bivariate count data model. The semiparametric model dominates the bivariate Poisson and bivariate negative binomial models in terms of both the maximized value of the log-likelihood function and the AIC. The two-factor SP2 model also dominates the one-factor generalized bivariate negative binomial model analyzed by Gurmu and Elder (2000) in terms of the AIC. Overall, SP2 with  $K = 2$  yields the minimum AIC. The estimates of the correlation parameters ( $\rho$ 's) show that modeling higher order polynomials of unobserved heterogeneity components is important. The bivariate Poisson model seems to be inadequate for joint estimation of overdispersed count data. There are some differences in the results from the semiparametric and the bivariate negative binomial models. In particular, there are differences in the statistical significance of some variables such as Sex, Levyplus, and Chond2 in the *Doctor* equation. The semiparametric estimates show that recent health status measures (Illness, Actdays) and one of the measures of long-term health status (Hscore) are important determinants of both doctor and non-doctor health professional visits. All of these are significant at 5%. The positive coefficient on the health status insurance variable Levyplus indicates that, relative to the default government Medibank insurance coverage, private insurance is associated with higher use of health services. Further, based on results from the SP2 regression, Non-doctor consultation is responsive to Freerepa, Chond1, and Chond2. The model predicts a convex relationship between Non-doctor health care utilization and age. Both health utilization measures are unresponsive to changes in income. This result is consistent with previous studies (*e.g.*, Cameron *et al.* 1988).

We also computed the predicted marginal effects of changes in the regressors on the mean number of visits. The marginal effects from bivariate SP2, bivariate Poisson and bivariate negative binomial (NB) are shown in Table G.2. Generally the marginal effects are greater in the bivariate models than in the independent Poisson and NB models originally analyzed by Cameron *et al.* (1988) using the same data set. The difference in the mean effects for various models is likely to be particularly important for

Table G.1 Semiparametric Coefficient Estimates and t-ratios

SP2 Estimates ( $K = 2$ )				
Variable	<i>Doctor</i>		<i>Nondoc</i>	
	Est	$t$	Est	$t$
<i>Constant</i>	-2.229	13.28	-2.820	6.04
<i>Sex</i>	.217	2.59	.288	1.41
<i>Age</i>	-.155	.94	-3.103	1.04
<i>Agesq</i>	.555	1.46	4.414	1.37
<i>Income</i>	-.143	1.02	-.036	0.62
<i>Levyplus</i>	.145	1.26	.348	2.28
<i>Freepoor</i>	-.551	2.11	-.169	0.46
<i>Freerepa</i>	.155	.96	.534	2.00
<i>Illness</i>	.217	8.78	.142	2.68
<i>Actdays</i>	.141	17.36	.134	7.66
<i>Hscore</i>	.040	2.60	.083	2.59
<i>Chond1</i>	.110	.89	.450	2.79
<i>Chcond2</i>	.205	1.72	1.241	6.50
$\log(\alpha_j)$	.073	.65	-2.059	23.89
$\rho_{11}$	.004	.11		
$\rho_{12}$	-.085	2.90		
$\rho_{21}$	.101	6.90		
$\rho_{22}$	.043	1.49		
Log-likel.	-5346.5			
AIC	10757.0			

Table G.2 Estimates of Marginal Effects and Moments of the Number of Visits

Variable	Bivar Poisson		Bivar NB		Bivar SP2	
	<i>Doctor</i>	<i>Nondoc</i>	<i>Doctor</i>	<i>Nondoc</i>	<i>Doctor</i>	<i>Nondoc</i>
<i>Sex</i>	.048	.072	.061	.087	.070	.081
<i>Age</i>	.075	.261	.164	.425	.126	.490
<i>Income</i>	-.062	-.007	-.048	-.001	-.047	-.010
<i>Levyplus</i>	.038	.072	.043	.084	.047	.098
<i>Freepoor</i>	-.129	.008	-.194	-.040	-.179	-.047
<i>Freerepa</i>	.023	.103	.075	.145	.050	.150
<i>Illness</i>	.056	.011	.082	.027	.071	.040
<i>Actdays</i>	.038	.021	.053	.034	.046	.038
<i>Hscore</i>	.009	.010	.015	.015	.013	.023
<i>Chond1</i>	.036	.113	.029	.115	.036	.127
<i>Chcond2</i>	.043	.232	.104	.278	.067	.349
Mean( $y_j$ )	.299	.212	.349	.254	.325	.281
Var( $y_j$ )	.302	.215	1.378	.931	.805	7.671
Corr( $y_j$ )	.019		.310		.058	

significant regressors. The sample average of the correlations between *Doctor* and *Non-doc* is about 0.3 in the bivariate NB. The average correlation from the SP2 regression is about 0.06.

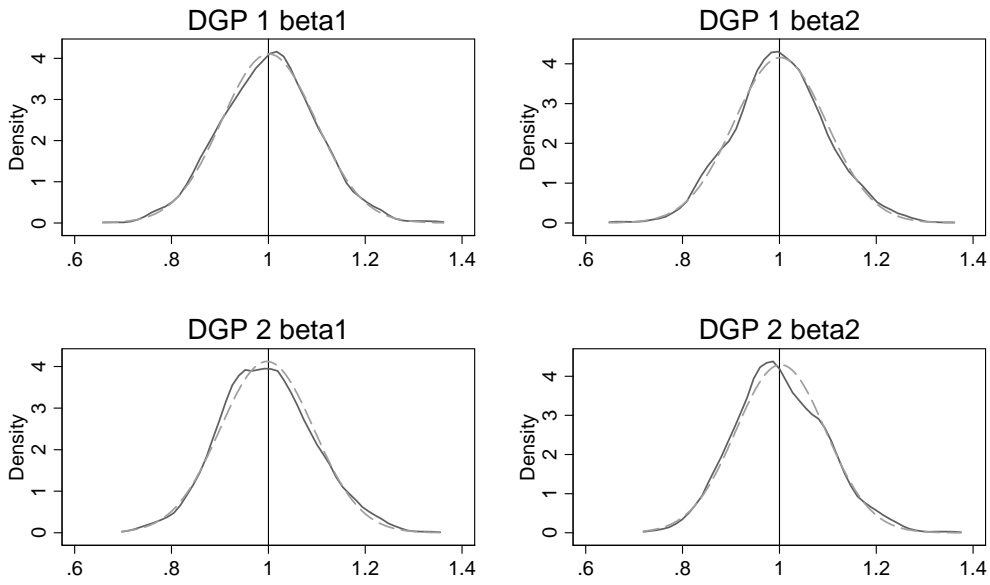
Compared with univariate and bivariate standard models, there is preponderance of statistical evidence (e.g., using AIC and tests of independence) in favor of the SP class of model of health care utilization. Direct comparisons of the unconditional sample moments and estimated conditional moments are generally difficult because of the role of explanatory variables and the sampling distributions of the estimated moments in the latter. For example, introduction of covariates will generally dampen the amount of overdispersion. Subject to this caveat, comparisons of models show that the estimated average means, variances, degree of overdispersion, and correlations from the SP2 model are generally consistent with the corresponding unconditional sample moments. Though not overwhelming, there are some differences between the results from the SP2 model and the results from realistic competitors in terms of the significance of the variables and the estimated marginal effects.

#### **Additional Reference**

Cameron, C., P. K. Trivedi, F. Milne and J. Piggott (1988), "A Microeconomic Model of the Demand for Health Care and Health Insurance in Australia," *Review of Economic Studies*, LV, 85-106.

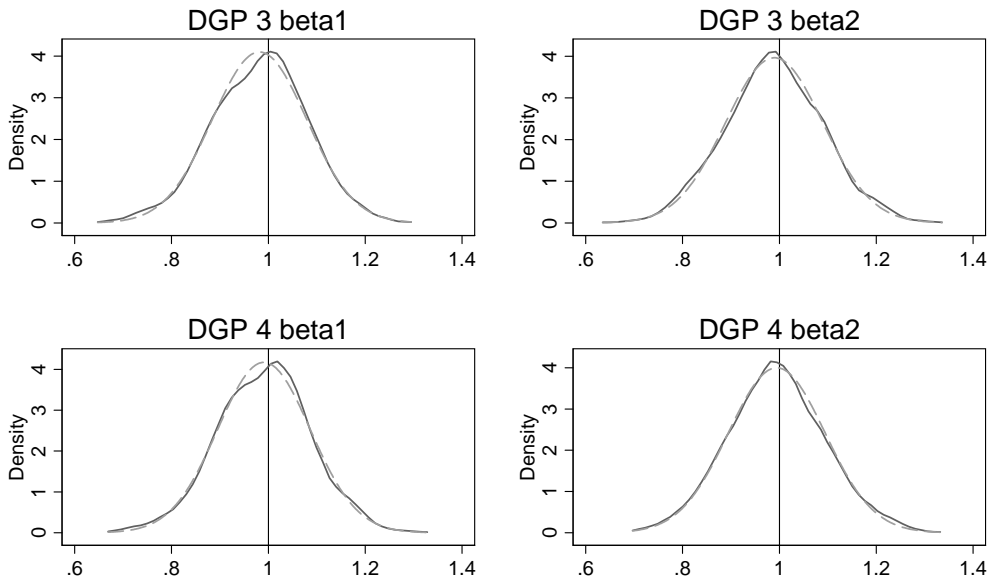
Hausman, J.A., B.H. Hall, Z. Griliches (1984). "Econometric Models for Count Data with Applications to the Patents R and D Relationship," *Econometrica*, 52, 909-938.

Figure E.1. Kernel density estimates of distributions of coefficients  
Bivariate Lognormal DGP - SP2, Sample size 1000



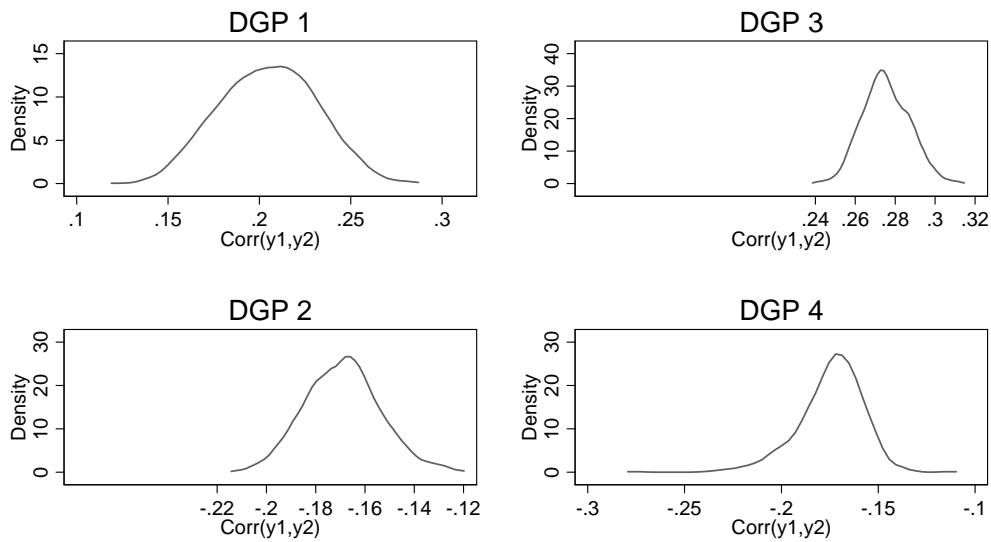
NOTE: The dashed line is the normal density

Figure E.2. Kernel density estimates of distributions of coefficients  
Bivariate Uniform DGP - SP2, Sample size 1000



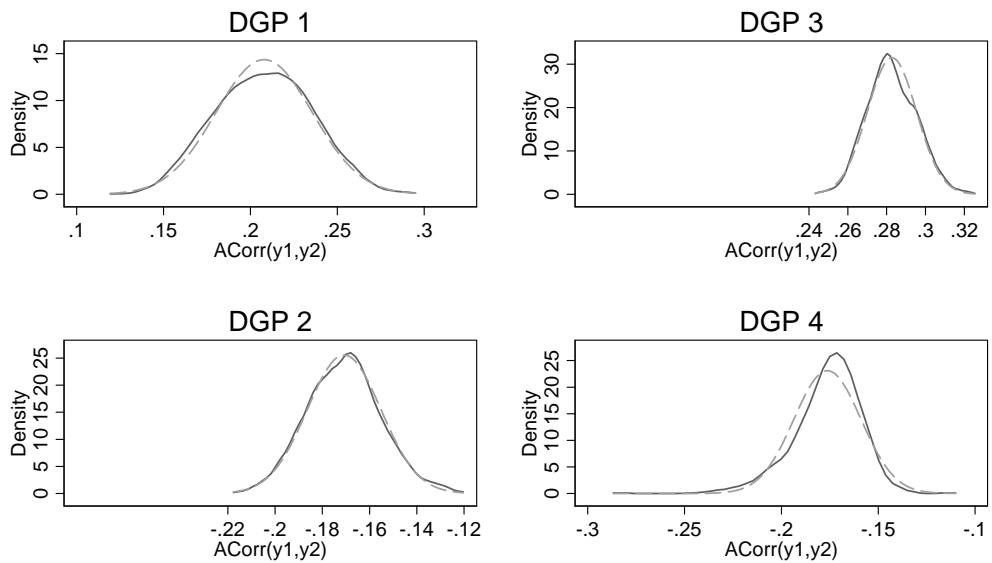
NOTE: The dashed line is the normal density

Figure E.3. Kernel density estimates of distributions of correlations  
SP2, Sample size 1000



NOTE: Correlations averaged over the sample observations prior to density estimation.  
The respective average correlations from DGPs 1, 2, 3 and 4 are 0.24, -0.20, 0.27 and -0.27.  
Bivariate heterogeneity distributions are lognormal for DGPs 1 and 2 and uniform for 3 and 4.

Figure E.4. Kernel density estimates of distributions of atanh of correlations SP2, Sample size 1000



NOTE: The dashed line is the normal density. atanh denotes the inverse hyperbolic tangent transformation. Bivariate heterogeneity distributions are lognormal for DGPs 1 and 2 and uniform for 3 and 4.