

A Multivariate Sarmanov Count Data Model^{*}

Eugenio J. Miravete[†]

August 5, 2010

Abstract

I present a multivariate, count data regression model that overcomes several difficulties to extend Poisson regressions beyond the univariate case. The estimator is computationally feasible and it accounts for both over and underdispersion. It also allows for correlations of any sign among counts independently of the dispersion parameters. I show that the proposed estimator has better small sample properties than the commonly used count data estimator based on a Gaussian copula. I apply the model to address whether the pricing strategies of competing duopolists in the early U.S. cellular telephone industry can be considered strategic complements or substitutes.

Keywords: Multivariate Count Data Models; Double Poisson; Sarmanov Distributions; Gaussian Copulas; Tariff Options.

JEL Codes: C16, C35, L11

^{*} Thanks are due to Jason Abrevaya, Stephen Donald, and Jerry Hausman for their comments to a previous version, as well to Thierry Magnac (editor) and three referees for their outstanding reviews. This paper was presented at the Texas Camp Econometrics XIV in New Braunfels.

[†]The University of Texas at Austin, Department of Economics, BRB 1.116, 1 University Station C3100, Austin, Texas 78712-0301; and CEPR, London, UK. Phone: 512-232-1718. Fax: 512-471-3510. E-mail: *miravete@eco.utexas.edu*; <http://www.eugeniomiravete.com>

1 Introduction

The absence of a sufficiently flexible multivariate distribution of counts has so far prevented the full information estimation of multivariate count data models. Simple analysis such as seemingly unrelated regressions, simultaneous equations models, or models with endogenous regressors remain mostly intractable whenever several endogenous counts are involved.¹ In the absence of such flexible joint distribution of counts, applied econometricians have turned their attention to computing intensive models such as orthogonal polynomial series expansions, *e.g.*, Cameron and Trivedi (1998, §8.5), moment-based estimation methods, *e.g.*, Gourieroux, Monfort, and Trognon (1984), and more recently copulas, in particular in the field of financial econometrics, *e.g.*, see Bien, Nolte, and Pohlmeier (2009), Heinen and Rengifo (2007), or Nolte (2008). However, most models available are difficult to extend beyond the bivariate case, and they fail to consider any effect of unobserved heterogeneity other than the well known overdispersion of the distribution of counts.² Furthermore, when correlated count events are considered, correlation is generally assumed to depend on the same parameter identifying overdispersion effects, therefore restricting it to be *necessarily positive* as the same parameter addresses

Dealing with over and underdispersion is the main goal of most developments in the single-equation count data regression models. Moving towards the multivariate case, the difficulty consists in building a framework that allows for the most flexible correlation pattern possible among counts. Cameron and Trivedi (1998, §8) best summarize the difficulties of estimating a

¹ Windmeijer and Santos-Silva (1997) successfully deal with one endogenous count and an endogenous continuous variable while Hausman, Leonard, and McFadden (1995) address the case of an endogenous count and an endogenous dichotomous variable.

² It has long been recognized that the Poisson model is generally too restrictive when estimating univariate count data regressions. Implicit to the Poisson model is the assumption of equidispersion of the distribution of counts, which is customarily rejected by the data. Many models, such as the Negative Binomial regression, have been suggested to address the existence of unobserved heterogeneity in the data that could explain the commonly observed overdispersion but not the less frequent underdispersion of the distribution of counts. Hausman, Hall, and Griliches (1984) even deal successfully with overdispersion in a univariate panel count data model.

multivariate count data regression model. Kocherlakota and Kocherlakota (1993) present what is perhaps the best known approach to deal with multiple and potentially correlated counts. They derive a bivariate Poisson distribution resulting from the addition of a common Poisson component to two independently distributed Poisson variables. The advantage of this approach —known as trivariate reduction— is that it allows for Poisson marginal distributions, but it also includes a *necessarily positive* correlation coefficient with a restricted range that characterizes the dependence structure of the count variables. Similarly, Marshall and Olkin (1990) generate a multivariate count data distribution from mixtures and convolutions of distributions of count events. The advantage of this second approach is that it allows for the simultaneous existence of unobserved heterogeneity that leads only to overdispersion and positive correlation of counts. Still, a serious limitation of all these models is that correlation among counts are *necessarily positive* because there is a single source of heterogeneity that explains *simultaneously* both the overdispersion of the marginal distributions of counts and their correlation.³ In this paper I make use of the Sarmanov family of distributions with double Poisson marginals to build a multivariate count data regression model that is flexible in the sense that it can accommodate both over and underdispersion independently of any correlation patterns among the counts.⁴

The Sarmanov count data regression model introduced in this paper has some remarkable features that overcome most difficulties in extending the existing single-dimensional count data regression models to multivariate environments. First, it can accommodate both over and un-

³ The multivariate Poisson-gamma mixture model of the random effects model of Hausman et al. (1984, §3) is a restricted version of the model of Marshall and Olkin (1990), while Gurmu and Elder (2000) and Winkelmann (2000) suggest multivariate negative binomial models. In all these works, only overdispersion is allowed but in the latter two cases, correlation is independent from dispersion, although still *necessarily positive*.

⁴ To the best of my knowledge, the double Poisson introduced by Efron (1986) is the only discrete univariate distribution that can accommodate both over and underdispersion. Winkelmann (1995) builds a similarly flexible unidimensional model but based on the continuous gamma distribution of latent waiting times that exploits the one-to-one relationship between the properties of the hazard rate of the distribution of waiting times and the over/underdispersion of the distribution of events that take place within an arbitrarily defined time interval. In a previous version of the present paper Miravete (2009) develops and estimate a multivariate version of the model of Winkelmann (1995) using the Sarmanov family of distributions.

derdispersion of the distribution of counts, therefore addressing a far larger pattern of behavior induced by the existence of unobserved individual heterogeneity. Second, any correlation sign is allowed, including the possibility of negative correlation among counts, thus reducing the possibility of misspecification.⁵ Third, dispersion and correlation of counts depend on different parameters of the model. This is an important property because it adds flexibility to the model by separating the effect of unobserved individual heterogeneity and correlation among counts (again reducing the likelihood of misspecification). Fourth, the model can be extended beyond the bivariate case. And finally, the estimation is not particularly time consuming: the likelihood function can always be written in closed form and there is no need to use simulation methods to obtain the parameter estimates.

The only drawback of the Sarmanov count data regression model is that in order to have a properly defined multivariate distribution of counts, the range of the estimates of correlations may be smaller than $[-1, 1]$. Indeed, the range of correlation is effectively bounded by the value of the rest of the parameters of the model. The estimation thus requires the use of constrained maximum likelihood methods. In order to deal with the possibility that some of the parameters are on the boundary of these constraints I make use of rescaled bootstrapping to obtain robust confidence intervals.

The use of the Gaussian copula with unrestricted correlation in many applied works (particularly in financial econometrics) may give the false impression that the proposed Sarmanov model is a step back. If that impression were correct, Gaussian copula would not only dominate the Sarmanov model but also any other multivariate model that makes use of families of distributions

⁵ There are other models that allow for correlation among counts of any sign based on the bivariate Poisson-lognormal distribution of Aitchison and Ho (1989), such as those of Hellström (2006), Munkin and Trivedi (1999), or Riphahn, Wambach, and Million (2003). These models can only address the case of overdispersed counts while estimation has to resort to simulation methods as the Poisson-lognormal mixture does not have a closed form expression. Gurmu and Elder (2008) obtain a closed form expression only after considering a first order Laguerre polynomial approximation to the bivariate distribution of unobservables. The Sarmanov regression model presented in this paper can address both over and underdispersion while being based on a well defined family of multivariate distributions.

such as the Farlie-Gumbel-Morgenstern or any copula function with a restricted correlation range. Correlations are generally restricted in order to ensure that the copula corresponds to a properly defined multivariate distribution function. The Gaussian copula does not compute the correlation between counts but rather the Pearson correlation between *continued* transformations of the corresponding cumulative marginal distributions. I argue in this paper that the use of the Gaussian copula to analyze truly discrete events such as the number of tariffs offered by competing firms may lead to seriously biased results. I show that both a Gaussian copula and my Sarmanov model produce similar estimates for most parameters of the model. However, the Gaussian copula model seriously overestimates the correlation coefficient and in many occasions produces significantly positive estimates while the data generating process only involves independent count distributions, *i.e.*, the Gaussian copula appears to be prone to type II errors when dealing with count events.

The need to estimate joint demands of countable products or services arise in many environments such as medical service, *e.g.*, Munkin and Trivedi (1999) or Riphahn et al. (2003); job changes, Jung and Winkelmann (1993); types of food, Meghir and Robin (1992); and recreational trips, Hausman et al. (1995), Hellström (2006), or Terza and Wilson (1990). In this paper I study the pricing strategies of competing duopolists in the early U.S. cellular telephone industry in order to evaluate whether the number of tariff options offered by competing firms are strategic complements and the pricing practices of firms corresponds to a supermodular game, *e.g.*, Topkis (1998, §4) and Vives (1990), or if alternatively, firms use their pricing offerings to differentiate themselves in attracting customers, *e.g.*, Yang and Ye (2008).

Back in the 1980s, as it is still today common in many industries, firms implemented nonlinear tariffs by means of a menu of self-selecting tariff options. In the absence of strategic considerations, firms will offer the number of tariff plans that is optimal to screen a customer base with some degree of heterogeneity while compensating for the costs of design and commercialization. In such a situation, estimating two independent count data regression models would be appropriate.

However, offering numerous rather than few tariff options might carry some strategic value and thus competitors may respond by offering a similar number of tariff plans in order to match the strategy of competitors. A significant positive estimates of the correlation among the count number of tariff plans offered supports the view that the number of tariff options in the early U.S. cellular industry are strategic complements. Alternatively, a negative correlation would arise if firms use the number of tariff plans as a device to segment markets and differentiate themselves from each other. However, results do not favor such interpretation.

The paper is organized as follows. Section 2 presents the bivariate count data regression model based on a bivariate and multivariate Sarmanov distributions with double Poisson marginals. This section also describes the properties of the Sarmanov family of distributions in reference to the proposed model with specific double Poisson marginal frequencies. Section 3 discusses the estimation of this model. Next, section 4 explores the small sample properties of the double Poisson-Sarmanov model and compares it to the Gaussian copula with double Poisson marginals. Section 5 estimates the two bivariate specification of the double Poisson-Sarmanov model to study the determinants of the number of tariff options offered by competing cellular telephone carriers in the U.S. during the mid-1980s. Section 6 concludes.

2 A Bivariate Double Poisson-Sarmanov Count Data Model

Let $y_k = 0, 1, 2, \dots$ be distributed as a double Poisson distribution with parameters μ_k and θ_k , conditional on a set of regressors \mathbf{x}_k in a sample with $i = 1, 2, \dots, n$ observations. After studying the properties of the double exponential family of distributions Efron (1986) shows that the probability frequency function of a double Poisson distribution is:

$$\tilde{f}_k(y_k|\mu_k, \theta_k) = c(\mu_k, \theta_k) f_k(y_k|\mu_k, \theta_k), \quad (1a)$$

$$f_k(y_k|\mu_k, \theta_k) = \sqrt{\theta_k} \exp(-\theta_k \mu_k) \exp(-y_k) \frac{y_k^{y_k}}{y_k!} \left(\frac{e\mu_k}{y_k} \right)^{\theta_k y_k}, \quad (1b)$$

$$\frac{1}{c(\mu_k, \theta_k)} = \sum_{y_k=0}^{\infty} f_k(y_k|\mu_k, \theta_k) \simeq 1 + \frac{1 - \theta_k}{12\theta_k \mu_k} \left(1 + \frac{1}{\theta_k \mu_k} \right), \quad (1c)$$

where $e = \exp(1)$, $\tilde{f}_k(y_k|\mu_k, \theta_k)$ denotes the exact double Poisson density, and $f_k(y_k|\mu_k, \theta_k)$ is the approximate probability mass function for the double Poisson family. The constant $c(\mu_k, \theta_k)$ makes $\tilde{f}_k(y_k|\mu_k, \theta_k)$ integrate to 1. Efron (1986) shows that $c(\mu_k, \theta_k)$ nearly equals 1 and thus he concludes that $f_k(y_k|\mu_k, \theta_k)$ is a good approximation for $\tilde{f}_k(y_k|\mu_k, \theta_k)$. The obvious advantage of the double Poisson over the standard Poisson distribution is that the mean and variance do not depend on the same single parameter. Thus, Efron (1986) also shows that conditional on a set of regressors \mathbf{x}_k , the expected count corresponding to observation i and its variance are:⁶

$$E[y_k|\mathbf{x}_k] \simeq \mu_k, \quad (2a)$$

$$\sigma_k^2 = \text{Var}[y_k|\mathbf{x}_k] \simeq \frac{\mu_k}{\theta_k}. \quad (2b)$$

Hence, the double Poisson includes the standard Poisson as a particular case when $\theta_k = 1$ but it allows for overdispersion if $\theta_k < 1$ as well as for underdispersion if $\theta_k > 1$. As it is commonly the case for count data regression models, I will specify an exponential mean function relating the observable characteristics to the expected number of counts:

$$\mu_k = \exp(\mathbf{x}'_k \beta_k). \quad (3)$$

⁶ Efron (1986, §3) show that in general, for the family of double exponential distributions, the difference between $\tilde{f}_k(y_k|\mu_k, \theta_k)$ and $f_k(y_k|\mu_k, \theta_k)$ is $O(n^{-1})$ while the differences between their means and variances are $O(n^{-2})$. Table 2 of the same paper explores the discrepancy of these two expressions for the double Poisson distribution as a function of parameters θ_k and μ_k . For the application of Section 5 the approximate probability is just 1.17% larger than the exact probability frequency function at the estimated value of the parameters. Notice however that all relations below make use of the exact double Poisson density.

Using Stirling's formula $z! \simeq \sqrt{2\pi z} \cdot z^z \cdot \exp(-z)$ repeatedly for $z = y_k$ and $z = \theta_k y_k$, the frequency function (1b) is accurately approximated by:⁷

$$f_k(y_k | \mu_k, \theta_k) \simeq \theta_k \exp(-\theta_k \mu_k) \frac{(\theta_k \mu_k)^{\theta_k y_k}}{\Gamma(\theta_k y_k + 1)}. \quad (4)$$

Sarmanov (1966) introduced a family of flexible bivariate distributions with given marginals. The double Poisson-Sarmanov count data regression models assumes that the marginal distributions are double Poisson. The bivariate probability frequency function takes the following form:

$$f_{12}(y_1, y_2) = f_1(y_1)f_2(y_2) \times [1 + \omega_{12}\psi_1(y_1)\psi_2(y_2)], \quad (5)$$

where $f_k(y_k)$ for $k = 1, 2$ corresponds to the double Poisson marginal frequency (4). For the case of positive counts, the mixing functions $\psi_k(y_k)$ are the bounded and nonconstant functions:⁸

$$\sum_{y_k=0}^{\infty} \psi_k(y_k) f_k(y_k) dy_k = 0. \quad (6)$$

Lee (1996, §4) explores a general approach for finding $\psi_k(y_k)$ and proves that for marginal distributions with support in \mathbb{R}_+ mixing functions are given by:

$$\psi_k(y_k) = \exp(-y_k) - L_k(1), \quad \forall y_k \geq 0, \quad (7)$$

where $L_k(1)$ is the value of the Laplace transform of the marginal distribution evaluated at $\zeta = 1$:

⁷ The use of Stirling's formula is useful for practical purposes to ensure the stability of estimation for large counts (not an issue in the application of this paper). Furthermore using it twice for $z = y_k$ and $z = \theta_k y_k$ simplifies $y_k^{y_k}/y_k!$ in (1b) and helps showing the convergence of the infinite sums of the double Poisson-Sarmanov model in Section 3.2.

⁸ In the literature on copulas mixing functions $\psi_k(y_k)$ are known as copula generators. See Fisher and Klein (2007, §2) and Nelsen (2006).

$$L_k(\zeta) = \sum_{y_k=0}^{\infty} \exp(-\zeta y_k) f_k(y_k) dy_k, \quad \text{at } \zeta = 1. \quad (8)$$

Substituting the marginal frequency function (4) we obtain an approximation to the Laplace transform of the double Poisson distribution, which evaluated at $\zeta = 1$ becomes:

$$L_k(1|\mu_k, \theta_k) \simeq c(\mu_k, \theta_k) \theta_k \exp(-\theta_k \mu_k) \sum_{y_k=0}^{\infty} \frac{(\theta_k \mu_k)^{\theta_k y_k} \exp(-y_k)}{\Gamma(\theta_k y_k + 1)}. \quad (9)$$

Thus, according to (7), the mixing function of the double Poisson-Sarmanov distribution is:

$$\psi_k(y_k|\mu_k, \theta_k) \simeq \exp(-y_k) - c(\mu_k, \theta_k) \theta_k \exp(-\theta_k \mu_k) \sum_{y_k=0}^{\infty} \frac{(\theta_k \mu_k)^{\theta_k y_k} \exp(-y_k)}{\Gamma(\theta_k y_k + 1)}. \quad (10)$$

A common theme in the literature on copula functions is the maximum range of variation of correlation. For expression (5) to properly define a bivariate density function the value of ω_{12} needs to fulfill the following constraint:

$$\omega_{12} \in \mathbb{R} : 1 + \omega_{12} \psi_1(y_1) \psi_2(y_2) \geq 0 \quad \forall y_1, y_2. \quad (11)$$

The Sarmanov family contains the Farlie-Gumbel-Morgenstern family of distributions as noticed by Johnson, Balakrishnan, and Kotz (2000, §44.13). The Sarmanov family shows not only a wider range for correlation coefficients, but also the possibility that correlation is negative, a rare feature of copula functions, *e.g.*, Joe (1997, 5.4). Indeed, Lee (1996) shows that restriction (11) holds when ω_{12} falls within the following bounds:

$$\underline{\omega}_{12} = \frac{-1}{\max\{L_1(1)L_2(1), [1 - L_1(1)][1 - L_2(1)]\}} \leq \omega_{12} \leq \frac{1}{\max\{L_1(1)[1 - L_2(1)], [1 - L_1(1)]L_2(1)\}} = \bar{\omega}_{12}, \quad (12)$$

which needs to be fulfilled for *every* observation in the sample.

It is useful to define the following mixing function weighted mean to compute the correlation coefficient:

$$\begin{aligned}\nu_k(\mu_k, \theta_k) &= \sum_{y_k=0}^{\infty} y_k \psi_k(y_k) f_k(y_k) dy_k = -L'_k(1|\mu_k, \theta_k) - L_k(1|\mu_k, \theta_k)\mu_k \\ &\simeq c(\mu_k, \theta_k)\theta_k \exp(-\theta_k\mu_k) \sum_{y_k=0}^{\infty} \frac{(\theta_k\mu_k)^{\theta_k y_k} \exp(-y_k)}{\Gamma(\theta_k y_k + 1)} (y_k - \mu_k),\end{aligned}\tag{13}$$

where $L'_k(1|\mu_k, \theta_k)$ denotes the value of the derivative of the Laplace transform (9) of the double Poisson frequency evaluated at $\zeta = 1$, and where $\Gamma(\theta_k y_k + 1)$ is the gamma function:

$$\Gamma(\theta_k y_k + 1) = \int_0^{\infty} \eta^{\theta_k y_k} \exp(-\eta) d\eta.\tag{14}$$

Next, notice that integrating the product $y_1 y_2$ with respect to (5) and making use of (13), the product moment can be written as:

$$E[y_1 y_2] = \mu_1 \mu_2 + \omega_{12} \nu_1 \nu_2,\tag{15}$$

so that the correlation coefficient of a well defined double Poisson-Sarmanov distribution can be:

$$\begin{aligned}\rho_{12} &= \frac{\omega_{12} \nu_1 \nu_2}{\sigma_1 \sigma_2} \\ &\simeq \omega_{12} \prod_{k=1}^2 \left\{ \frac{c(\mu_k, \theta_k)\theta_k \exp(-\theta_k\mu_k)}{\sqrt{\mu_k/\theta_k}} \sum_{y_k=0}^{\infty} \frac{(\theta_k\mu_k)^{\theta_k y_k} \exp(-y_k)}{\Gamma(\theta_k y_k + 1)} (y_k - \mu_k) \right\} \\ &= \omega_{12} \prod_{k=1}^2 Q(\mu_k, \theta_k).\end{aligned}\tag{16}$$

Thus, when $\omega_{12} = 0$ the correlation parameter is $\rho_{12} = 0$, and variables y_1 and y_2 are independent.⁹

⁹ Shubina and Lee (2004) study how condition (12) constrains the valid range of correlation (16) for different marginal distributions. These authors also show that the maximum ranges of other association measures such as Kendall's τ and Spearman's rank correlation coefficient are independent of the assumed marginal distributions.

Combining all these elements into (5) we obtain the probability of observing simultaneously a pair of counts $\{y_1, y_2\}$ generated by the double Poisson-Sarmanov distribution is:

$$f_{12}(y_1, y_2) \simeq \left(\prod_{k=1}^2 \left\{ c(\mu_k, \theta_k) \theta_k \exp(-\theta_k \mu_k) \frac{(\theta_k \mu_k)^{\theta_k y_k}}{\Gamma(\theta_k y_k + 1)} \right\} \right) \times \left(\frac{\prod_{m=1}^2 \left\{ \exp(-y_m) - c(\mu_m, \theta_m) \theta_m \exp(-\theta_m \mu_m) \sum_{y_m=0}^{\infty} \frac{(\theta_m \mu_m)^{\theta_m y_m} \exp(-y_m)}{\Gamma(\theta_m y_m + 1)} \right\}}{\prod_{m=1}^2 Q(\mu_m, \theta_m)} \right)^{1 + \rho_{12}}. \quad (17)$$

In addition, for this double Poisson-Sarmanov to be coherent and properly define a bivariate probability frequency function, the constraint corresponding to the general case (12) needs to hold. Making use of (9) and (16), the constraint can be written in terms of ρ_{12} as follows:

$$\underline{\omega}_{12} \prod_{k=1}^2 Q(\mu_k, \theta_k) \leq \rho_{12} \leq \bar{\omega}_{12} \prod_{k=1}^2 Q(\mu_k, \theta_k), \quad \forall i. \quad (18)$$

2.1 Multivariate Extensions

Multivariate extensions of this model are of clear interest for practical purposes. Lee (1996, §8) suggests a generalization of the joint density function of a multivariate Sarmanov distribution that accounts for higher order correlation among counts:

$$f_{1,2,\dots,K}(y_1, \dots, y_K) = \left[\prod_{k=1}^K f_k(y_k) \right] \times [1 + R_{\psi_1, \dots, \psi_K, \Omega_{\mathbf{K}}}(y_1, \dots, y_K)], \quad (19)$$

where correlations must fulfill the following condition:

$$\begin{aligned}
1 + R_{\psi_1, \dots, \psi_K, \Omega_K}(y_1, \dots, y_K) &= 1 + \sum_{1 \leq l_1 \leq l_2 \leq K} \sum \omega_{l_1 l_2} \prod_{m=1}^2 \psi_{l_m}(y_{l_m}) \\
&+ \sum_{1 \leq l_1 \leq l_2 \leq l_3 \leq K} \sum \omega_{l_1 l_2 l_3} \prod_{m=1}^3 \psi_{l_m}(y_{l_m}) + \dots + \omega_{12 \dots K} \prod_{m=1}^K \psi_m(y_m) \geq 0, \quad \forall y_1, \dots, y_K.
\end{aligned} \tag{20}$$

Extending the double Poisson-Sarmanov to more than two dimensions reduces to repeating the bivariate analysis of this section and substituting the probability frequency function (4) and mixing function (10) into (19) and (20):

$$\begin{aligned}
f_{1,2,\dots,K}(y_1, \dots, y_K) &\simeq \left[\prod_{k=1}^K \left\{ c(\mu_k, \theta_k) \theta_k \exp(-\theta_k \mu_k) \frac{(\theta_k \mu_k)^{\theta_k y_k}}{\Gamma(\theta_k y_k + 1)} \right\} \right] \times \\
&\left[1 + \sum_{1 \leq l_1 \leq l_2 \leq K} \rho_{l_1 l_2} \frac{\prod_{m=1}^2 \left\{ \exp(-y_{l_m}) - c(\mu_{l_m}, \theta_{l_m}) \theta_{l_m} \exp(-\theta_{l_m} \mu_{l_m}) \sum_{y_{l_m}=0}^{\infty} \frac{(\theta_{l_m} \mu_{l_m})^{\theta_{l_m} y_{l_m}} \exp(-y_{l_m})}{\Gamma(\theta_{l_m} y_{l_m} + 1)} \right\}}{\prod_{m=1}^2 Q(\mu_{l_m}, \theta_{l_m})} \right. \\
&+ \sum_{1 \leq l_1 \leq l_2 \leq l_3 \leq K} \rho_{l_1 l_2 l_3} \frac{\prod_{m=1}^3 \left\{ \exp(-y_{l_m}) - c(\mu_{l_m}, \theta_{l_m}) \theta_{l_m} \exp(-\theta_{l_m} \mu_{l_m}) \sum_{y_{l_m}=0}^{\infty} \frac{(\theta_{l_m} \mu_{l_m})^{\theta_{l_m} y_{l_m}} \exp(-y_{l_m})}{\Gamma(\theta_{l_m} y_{l_m} + 1)} \right\}}{\prod_{m=1}^3 Q(\mu_{l_m}, \theta_{l_m})} \\
&\left. + \dots + \rho_{12 \dots K} \frac{\prod_{m=1}^K \left\{ \exp(-y_m) - c(\mu_m, \theta_m) \theta_m \exp(-\theta_m \mu_m) \sum_{y_m=0}^{\infty} \frac{(\theta_m \mu_m)^{\theta_m y_m} \exp(-y_m)}{\Gamma(\theta_m y_m + 1)} \right\}}{\prod_{m=1}^K Q(\mu_m, \theta_m)} \right],
\end{aligned} \tag{21}$$

where, similarly to equation (16), higher order correlation coefficients are given by:

$$\rho_{12 \dots K} = \omega_{12 \dots K} \prod_{k=1}^K \left(\frac{\nu_k}{\sigma_k} \right), \tag{22}$$

which in turn makes use of the product moment:

$$E[y_1 y_2 \dots y_K] = \prod_{k=1}^K \mu_k + \omega_{12\dots K} \prod_{k=1}^K \nu_k. \quad (23)$$

3 Estimation

Despite the apparently cumbersome notation, estimation of the proposed model is relatively straightforward. Take for instance the scalar realizations y_{1i} and y_{2i} of two count random variables given two vectors of regressors \mathbf{x}_{1i} and \mathbf{x}_{2i} , parameter vectors γ_1 and γ_2 , as well as parameter scalar ω_{12} . Estimation by maximum likelihood maximizes the probability of jointly observing $\{y_{11}, y_{21}\}, \{y_{12}, y_{22}\}, \dots, \{y_{1n}, y_{2n}\}$ in an n -size sample. Using the general bivariate Sarmanov distribution (5), the log-likelihood function can be written as:

$$\mathcal{L}(\gamma_1, \gamma_2, \omega_{12}) = \sum_{i=1}^n \sum_{k=1}^2 \ln f_k(y_{ki} | \mathbf{x}_{ki}, \gamma_{ki}) + \sum_{i=1}^n \ln \left[1 + \omega_{12} \prod_{k=1}^2 \psi_k(y_{ki} | \mathbf{x}_{ki}, \gamma_{ki}) \right]. \quad (24)$$

Notice that ω_{12} only enters the term between brackets. The estimation thus proceeds iteratively, alternatively fixing the value of ω_{12} or γ_1 and γ_2 until we achieve convergence. Initial values $\hat{\gamma}_1^{(0)}$ and $\hat{\gamma}_2^{(0)}$ are obtained under the assumption of independence, *i.e.*, setting $\omega_{12} = 0$ and estimating two separate count data regression models. The initial estimate of ω_{12} is obtained by grid search, evaluating (24) over the interval defined by the constraint (12) while holding the estimated $\hat{\gamma}_1^{(0)}$ and $\hat{\gamma}_2^{(0)}$ constant. With this new value of $\hat{\omega}_{12}^{(0)}$, new estimates $\hat{\gamma}_1^{(1)}$ and $\hat{\gamma}_2^{(1)}$ are obtained by maximizing (24) while holding ω_{12} constant at the estimated value $\hat{\omega}_{12}^{(0)}$. The process is repeated until convergence is achieved.¹⁰

¹⁰ The Gauss code used for the estimation of the bivariate model of Section 5 is available upon request.

Estimating a trivariate or multivariate model is slightly more convoluted because in principle equation (20) would allow for multiple combinations of correlations coefficients that fulfill such constraint. However, the solution to this maximization problem is unique because Lee (1996, Theorem 5a) states that if $\{y_1, y_2, \dots, y_K\}$ are jointly distributed according to a K -variate Sarmanov distribution, then any subset of $\{y_1, y_2, \dots, y_K\}$ will also be distributed as a Sarmanov distribution. To see how this helps estimating the different correlation coefficients of a multivariate Sarmanov distribution, consider the trivariate case. The log-likelihood function of an n -size sample is:

$$\begin{aligned} \mathcal{L}(\gamma_1, \gamma_2, \gamma_3, \omega_{12}, \omega_{13}, \omega_{23}, \omega_{123}) &= \sum_{i=1}^n \sum_{k=1}^3 \ln f_k(y_{ki} | \mathbf{x}_{ki}, \gamma_{ki}) \\ &+ \sum_{i=1}^n \ln \left[1 + \sum_{1 \leq l_1 \leq l_2 \leq 3} \omega_{l_1 l_2} \prod_{m=1}^2 \psi_{l_m} (y_{l_m i} | \mathbf{x}_{l_m i}, \gamma_{l_m i}) + \omega_{123} \prod_{k=1}^3 \psi_k (y_{ki} | \mathbf{x}_{ki}, \gamma_{ki}) \right]. \end{aligned} \quad (25)$$

Under the assumption of independence, single dimensional count data regressions produce initial estimates for γ_1 , γ_2 , and γ_3 . Conditioning on $\hat{\gamma}_1^{(0)}$ and $\hat{\gamma}_2^{(0)}$ in (24) we obtain the estimate $\omega_{12}^{(0)}$ by the grid search procedure described above. The same approach can be used to obtain estimates $\omega_{13}^{(0)}$ and $\omega_{23}^{(0)}$ while conditioning the likelihood function (24) on $\{\hat{\gamma}_1^{(0)}, \hat{\gamma}_3^{(0)}\}$ and $\{\hat{\gamma}_2^{(0)}, \hat{\gamma}_3^{(0)}\}$, respectively. Then, maximizing (25) produces an estimate of ω_{123} while holding all the other parameters constant. Once $\hat{\omega}_{123}^{(0)}$ has been obtained, new estimates $\gamma_1^{(1)}$, $\gamma_2^{(1)}$, and $\gamma_3^{(1)}$ are estimated by maximizing (25) while holding $\{\hat{\omega}_{12}^{(0)}, \hat{\omega}_{13}^{(0)}, \hat{\omega}_{23}^{(0)}, \hat{\omega}_{123}^{(0)}\}$ constant. This procedure is then repeated until convergence is achieved.

3.1 Inference

We first need to evaluate whether we can consistently estimate parameters that may lie on the boundary generically defined by condition (11). Notice that the Sarmanov model only imposes a constraint on the correlation coefficient (although it must be fulfilled by *every* observation in

the sample). All other parameters, although they define the range of variation of the correlation coefficient, remain unrestricted. Furthermore, the range defined by (12) is a compact convex set so that any estimate of the correlation coefficient includes its neighborhood, thus fulfilling the requirements of Andrews (2000, §4.2) who studies the asymptotic distributions of estimators when the true parameter lies on the boundary of the parameter space.

Thus, in order to obtain consistent inference for these parameter estimates, we need to address the possibility that estimated parameters may lie on the boundary defined by the constraint (18). Andrews (1999) shows that standard bootstrapping does not produce consistent inference when a parameter is on the boundary of the parameter space defined by a nonlinear inequality such as general conditions (11) and (20) for the bivariate and multivariate case, respectively. Rather than computing common bootstrap standard errors Andrews (1999, §6.4) suggests the use of a rescaled bootstrap method in which bootstrap samples of size $b < n$ are employed.¹¹ Andrews (2000, §4) shows that this modified bootstrapping approach produces consistent standard errors estimates regardless of whether the true parameter is on a boundary of the parameter space or not. I thus employ rescaled bootstrapping to produce consistent inference for these constrained maximum-likelihood estimates. To speed up the process of obtaining robust inference I follow Andrews (2002) and compute a 10-step version of the rescaled bootstrap.

3.2 Econometric Implementation

An additional issue regarding the econometric implementation of this model remains to be addressed. We need to decide how many terms of the infinite sums in equations (9)-(17) to account for in the estimation. Equation (17) includes the following element:

¹¹ Horowitz (2001, §2.2) calls the rescaled bootstrapping *replacement subsampling* and discusses how it outperforms bootstrapping when the bootstrap is inconsistent. In the case of rescaled bootstrapping we use a given number of bootstrap samples of size b where some of these samples might be repeated. This differentiates rescaled bootstrapping from subsampling where some or all $n!/[(b!(n-b)!)]$ samples of size b (always without repetition) are employed in estimating each replication, *e.g.*, see Politis, Romano, and Wolf (1999, §2.1).

$$S_k(\mu_k, \theta_k) = \sum_{y_k=0}^{\infty} \frac{(\theta_k \mu_k)^{\theta_k y_k} \exp(-y_k)}{\Gamma(\theta_k y_k + 1)}. \quad (26)$$

Notice that for $\theta_k = 1$, the series $S_k(\mu_k, \theta_k)$ converges to:

$$S_k(\mu_k, 1) = \sum_{y_k=0}^{\infty} \frac{[\mu_k \exp(-1)]^{y_k}}{y!} = \exp\left(\frac{\mu_k}{e}\right), \quad (27)$$

because of the well known Taylor expansion of the exponential function. Using Stirling's formula:

$$\lim_{y_k \rightarrow \infty} \frac{(\theta_k \mu_k)^{\theta_k y_k} \exp(-y_k)}{\sqrt{2\pi} \sqrt{\theta_k y_k} (\theta_k y_k)^{\theta_k y_k} \exp(-\theta_k y_k)} = \lim_{y_k \rightarrow \infty} \frac{1}{\sqrt{2\pi \theta_k y_k}} \cdot \lim_{y_k \rightarrow \infty} \left(\frac{\mu_k}{y_k}\right)^{\theta_k y_k} \cdot \lim_{y_k \rightarrow \infty} [\exp(-y_k)]^{1-\theta_k} = 0, \quad (28)$$

so that the sum $S_k(\mu_k, \theta_k)$ converges for any value of θ_k . However, the length of the series needed to approximate $S_k(\mu_k, \theta_k)$ varies greatly with θ_k . We can rewrite equation (26) as:

$$S_k(\mu_k, \theta_k) = \sum_{y_k=0}^{\infty} \frac{[(\theta_k \mu_k)^{\theta_k} \exp(-1)]^{y_k}}{y_k!} \frac{y_k!}{(\theta_k y_k)!}, \quad (29)$$

so that the number of elements of the sum in (26) needed to approximate $S_k(\mu_k, \theta_k)$ decreases with θ_k for any given precision level. Thus, longer series are needed to approximate $S_k(\mu_k, \theta_k)$ the more overdispersed the distribution of y_k is.

The following expression can be used to determine the length of the series, *i.e.*, the value of y_k needed to increase estimate of $S_k(\mu_k, \theta_k)$ by at least δ :¹²

$$y_k^*(\delta) \in \operatorname{argmin}_{y_k} \frac{1}{\sqrt{2\pi \theta_k y_k}} \cdot \left(\frac{\mu_k}{y_k}\right)^{\theta_k y_k} \cdot [\exp(-y_k)]^{1-\theta_k} = \delta. \quad (30)$$

¹² The estimation of Section 5 below uses the first 40 elements of the infinite series to approximate (26), leading to an approximation error $\delta = 7 \times 10^{-107}$ for y_1 and $\delta = 1 \times 10^{-77}$ for y_2 at the current estimated values of the parameters. Consequently, varying the number of elements from 30 to 50 has no practical consequences for the estimation of the double Poisson-Sarmanov model.

4 Gaussian Copula vs. the Sarmanov Model

The Sarmanov family of distributions together with double Poisson marginal frequency functions combine into a properly defined multivariate frequency function as long as the correlations among the counts fulfill the general condition (20). A limited range for correlation (although always allowing for negative values) and the need to use scaled bootstrapping to obtain proper inference may deter practitioners from embracing the proposed model if simpler alternatives are available. One such alternative appears to be the Gaussian-Poisson copula recently used in financial econometrics, *e.g.*, Bien et al. (2009). Advocates of this copula vindicate its ability to accommodate any unrestricted correlation patterns among counts. Furthermore estimation employs straightforward maximum likelihood methods.¹³ The Monte Carlo analysis of this section shows that unfortunately the Gaussian-Poisson copula seriously overestimates the correlation between counts. Furthermore, the Gaussian-Poisson copula estimator of the correlation coefficient customarily leads to type II errors. Thus, estimates of the correlation coefficient are positive and significant despite the data generating process involving only two independent Poisson processes. This is a particularly acute problem for count processes with low means.

This section first reviews the main elements of a Gaussian-double-Poisson copula and then conducts a Monte Carlo analysis where the underlying data generating process consists of two independently distributed Poisson counts.

4.1 Gaussian Copula

To make the copula model directly comparable to the Sarmanov model presented in Section 2, I will assume that marginal frequency functions are given by the double Poisson of equations (1a)-(1c)

¹³ If this copula could provide with accurate estimates of the effect of regressors as well as correlations, it would turn obsolete for practical purposes most, if not all, research on multivariate discrete distributions summarized in Johnson, Kotz, and Balakrishnan (1997).

with cumulative frequency function $F_k(y_k|\mu_k, \theta_k)$.¹⁴ To ease the presentation I will focus on the bivariate case. As Heinen and Rengifo (2007, §2) discuss, copulas are uniquely defined under very general conditions if marginal distributions are continuous. For discrete distributions they are only uniquely defined on the lattice $\bigotimes_{k=1}^K \text{Range}(F_k(y_k|\mu_k, \theta_k))$. An easy way to solve this difficulty is the *continuisation* of counts suggested by Denuit and Lambert (2005) where a *continued* variable y_i^* is generated for each count, y_i so that the nonlinear measures of association among the *continued* variables are the same than among the original counts.¹⁵ The count data y_i is continued by standard uniformly distributed draws u as follows:

$$y_k^* = y_k + (u - 1). \quad (31)$$

Next, let denote by φ_k the probability integral transformation of y_k^* , the *continued* value of y_i with continuous marginal distribution $F_k^*(y_k^*|\mu_k, \theta_k)$:

$$\varphi_i = F_k^*(y_k^*|\mu_k, \theta_k) = F_k(y_k + (u - 1)|\mu_k, \theta_k) = F_k(y_k - 1|\mu_k, \theta_k) + f_k(y_k|\mu_k, \theta_k) \cdot u, \quad (32)$$

which is uniformly distributed, *e.g.*, Angus (1994). Furthermore, since $dy_k^* = du$:

$$f_k^*(y_k^*|\mu_k, \theta_k) = f_k(y_k|\mu_k, \theta_k). \quad (33)$$

Sklar (1959) showed that the following copula function $C^*(\cdot)$ describing the joint distribution of y_1 and y_2 exists and is unique:

¹⁴ This is exactly the copula function employed by Heinen and Rengifo (2007) in a time series environment. Unlike the Gaussian-Poisson copula of Bien et al. (2009), assuming double Poisson marginals has the advantage of reducing the risk of misspecification by wrongly imposing equidispersion of counts.

¹⁵ Heinen and Rengifo (2007, §2) make use of this *continuisation* approach to estimate their Gaussian copula model. Alternatively, Bien et al. (2009) and Cameron, Li, Trivedi, and Zimmer (2004) use first order differencing operators to link the multivariate frequency function with the cumulative multivariate copula distribution function.

$$F_{12}^*(y_1^*, y_2^* | \mu_1, \theta_1, \mu_2, \theta_2) = C^*(F_1^*(y_1^* | \mu_1, \theta_1), F_2^*(y_2^* | \mu_2, \theta_2)) . \quad (34)$$

Thus, a bivariate copula with standard uniform marginals can always be written:

$$C^*(\varphi_1, \varphi_2) = F^*(F_1^{*-1}(\varphi_1), F_2^{*-1}(\varphi_2)) , \quad (35)$$

The Gaussian copula makes use of a standard joint bivariate normal distribution with standard univariate normal marginal distribution functions in equation (35), thus allowing for the inclusion of parameter ρ to account for any correlation among monotone nonlinear transformations of the *continued* variables:

$$C_{\Phi}^*(\varphi_1, \varphi_2; \rho) = \Phi_{[2]} \left(\Phi_{[1]}^{-1}(\varphi_1), \Phi_{[1]}^{-1}(\varphi_2); \rho \right) , \quad (36)$$

where $\Phi_{[\lambda]}(\cdot)$ denotes the λ -dimensional standard normal probability distribution function. Notice however that ρ does not measure the correlation between the original counts y_1 and y_2 but rather the correlation between two normal scores $\Phi_{[1]}^{-1}(\varphi)$, so that indeed it measures the dependence between y_1^* and y_2^* through the monotone nonlinear transformation (32), *i.e.*, $\Phi_{[1]}^{-1}[F_k(y_k + (u - 1))]$. Combining Sklar's result (34), the Gaussian copula (36), and equations (32)–(33), the joint probability density function of the continued variables can be written as the product of the marginal frequency functions and the copula density:

$$f^*(y_1^*, y_2^*) = f_1^*(y_1^*)f_2^*(y_2^*) \frac{\partial^2 C^*(F_1^*(y_1^*), F_2^*(y_2^*))}{\partial F_1^*(y_1^*) \partial F_2^*(y_2^*)} = f_1(y_1)f_2(y_2)\phi_{[2]} \left(\Phi_{[1]}^{-1}(\varphi_1), \Phi_{[1]}^{-1}(\varphi_2); \rho \right) . \quad (37)$$

Therefore, the contribution of each observation to the log-likelihood is:

$$\sum_{i=1}^2 \log [f_i(y_{ij}, \mu_{ij}, \theta_i | \mathbf{x}_i)] + \log \left[\phi_{[2]} \left(\Phi_{[1]}^{-1}[F_1(y_{1j} + (u_j - 1))], \Phi_{[1]}^{-1}[F_2(y_{2j} + (u_j - 1))]; \rho \right) \right] . \quad (38)$$

4.2 Monte Carlo Analysis

Before the recent introduction of copula functions in econometrics Lee (1982) and Lee (1983) pioneered the use of the probability integral transformation and the Gaussian copula to allow for general multivariate distributions for the estimation of selection models. Interestingly, Goldberger (1983) concluded that the selectivity bias adjustment was quite sensitive to departures from normality. To the best of my knowledge there is no similar study available to evaluate the performance of copula functions in general and the Gaussian-Poisson copula in particular so widely used in the field of empirical econometrics. The Monte Carlo analysis of this section studies whether such copula approach produces reliable estimates of the correlation between counts when indeed they are independently distributed.

The data generating process consists of a couple of independently distributed Poisson variables with exponential mean functions:

$$\mu_k = \exp[\alpha_k + \beta_k x_j + \varepsilon_k], \quad j = 1, 2, \quad (39)$$

where x_1 and x_2 are independently drawn from $N(0, 1/16)$ and the unobserved heterogeneity components ε_1 and ε_2 are also independently distributed as $N(0, 1/4)$, *i.e.*, exactly as in Munkin and Trivedi (1999, §6). I generate 1,000 observations for these variables and heterogeneity components. Intercepts take nine different values $\alpha_1 = \alpha_2 = \{-1, -0.5, -0.25, -0.1, 0, 0.1, 0.25, 0.5, 1\}$ that define the nine different scenarios evaluated in Table 1. The true value of the other parameters are $\beta_1 = \beta_2 = 1$ and $\theta_1 = \theta_2 = 1$ (as the counts are equidispersed by construction). The true correlation between counts is always $\rho = 0$. For each set of parameter values I compute 500 replications and Table 1 reports the mean and standard deviations of the parameter estimates.

Results reported in Table 1 show that with the exception of ρ , other parameter estimates are almost identical independently of whether we use the Sarmanov model or the Gaussian copula

Table 1: Monte Carlo Simulations – Gaussian Copula vs. Sarmanov Model

	$\alpha = -1$	$\alpha = -0.5$	$\alpha = -0.25$	$\alpha = -0.1$	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 1$
Gaussian									
α_1	-0.8691 (0.0380)	-0.3646 (0.0305)	-0.1381 (0.0277)	0.0515 (0.0258)	0.1211 (0.0246)	0.2312 (0.0242)	0.3972 (0.0229)	0.6585 (0.0205)	1.1547 (0.0184)
β_1	0.9833 (0.1530)	0.9314 (0.1290)	1.1043 (0.1138)	1.0777 (0.1059)	0.9614 (0.0989)	1.1190 (0.0922)	1.0124 (0.0875)	1.0751 (0.0841)	1.0666 (0.0762)
θ_1	1.6746 (0.0427)	1.3203 (0.0318)	1.1448 (0.0253)	1.0415 (0.0272)	1.0077 (0.0229)	0.9714 (0.0246)	0.9371 (0.0238)	0.7891 (0.0208)	0.6174 (0.0177)
α_2	-0.8827 (0.0395)	-0.4024 (0.0328)	-0.1300 (0.0273)	-0.0170 (0.0265)	0.0894 (0.0243)	0.2138 (0.0244)	0.3724 (0.0214)	0.5938 (0.0202)	1.1143 (0.0175)
β_2	1.0261 (0.1591)	0.8912 (0.1185)	0.9559 (0.0168)	0.8975 (0.1148)	1.0328 (0.0956)	0.9305 (0.0952)	0.8442 (0.0942)	0.9622 (0.0878)	0.9999 (0.0706)
θ_2	1.6742 (0.0454)	1.3091 (0.0364)	1.1852 (0.0259)	1.0918 (0.0261)	1.0574 (0.0243)	0.9766 (0.0232)	0.9237 (0.0249)	0.8023 (0.0231)	0.6379 (0.0183)
ρ	0.0671 (0.0324)	0.1717 (0.0375)	0.0695 (0.0346)	0.0318 (0.0317)	0.0224 (0.0290)	0.0174 (0.0293)	0.0099 (0.0328)	0.0598 (0.0322)	0.0325 (0.0239)
Type II	55.80	100.00	57.60	17.60	9.60	6.00	7.60	47.60	11.00
Sarmanov									
α_1	-0.8691 (0.0380)	-0.3646 (0.0302)	-0.1382 (0.0277)	0.0514 (0.0258)	0.1210 (0.0246)	0.2313 (0.0242)	0.3972 (0.0229)	0.6585 (0.0205)	1.1547 (0.0184)
β_1	0.9833 (0.1530)	0.9312 (0.1292)	1.1042 (0.1140)	1.0777 (0.1059)	0.9614 (0.0990)	1.1190 (0.0922)	1.0124 (0.0875)	1.0751 (0.0841)	1.0666 (0.0762)
θ_1	1.6746 (0.0427)	1.3201 (0.0311)	1.1450 (0.0253)	1.0416 (0.0272)	1.0077 (0.0229)	0.9713 (0.0246)	0.9371 (0.0238)	0.7891 (0.0208)	0.6174 (0.0177)
α_2	-0.8827 (0.0395)	-0.4024 (0.0327)	-0.1303 (0.0273)	-0.0171 (0.0265)	0.0892 (0.0243)	0.2139 (0.0244)	0.3725 (0.0214)	0.5938 (0.0202)	1.1143 (0.0175)
β_2	1.0261 (0.1591)	0.8910 (0.1186)	0.9558 (0.1068)	0.8975 (0.1148)	1.0328 (0.0956)	0.9305 (0.0952)	0.8442 (0.0942)	0.9622 (0.0878)	0.9999 (0.0706)
θ_2	1.6742 (0.0454)	1.3093 (0.0358)	1.1853 (0.0259)	1.0918 (0.0261)	1.0574 (0.0243)	0.9766 (0.0232)	0.9237 (0.0249)	0.8023 (0.0231)	0.6379 (0.0183)
ρ	0.0008 (0.0003)	0.0076 (0.0260)	-0.0053 (0.0059)	-0.0027 (0.0030)	-0.0027 (0.0030)	0.0019 (0.0024)	0.0027 (0.0029)	0.0003 (0.0005)	0.0000 (0.0000)
Type II	0.00	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

500 replications. Mean estimate and standard deviation (between parentheses). Type II reports the percentage of replications leading to a significant estimate of ρ

approach. This is not surprising as both, the likelihood of the Sarmanov model (24) and the Gaussian copula (38) include the product of two double Poisson frequencies and an additional term that determines the value of the correlation coefficient. In the case of the Sarmanov model this additional term is required to fulfill some constraints so that the estimate is consistent with the Sarmanov family of distributions, but for the Gaussian copula case it is unrestricted although embedded on a bivariate normal distribution of some monotone transformation of the *continued* values of the counts. Thus, both models produce generally unbiased estimates of β_1 and β_2 and slightly overestimate α_1 and α_2 . Dispersion parameters θ_k are overestimated for low values of α_k (low expected mean of counts) and underestimated for high values of α_k (large expected mean of counts).

The most remarkable difference between these two models is the disparate estimates they provide for ρ . The Sarmanov model produces unbiased estimates while the average for those of the Gaussian copula are systematically larger than zero. Indeed, a detailed analysis of the estimates of each replication reveals that the Sarmanov model produces an estimate of ρ that is significantly different from zero only in ten occasions (out of 4,500 total replications). The estimate of ρ produced by the Gaussian copula is much more prone type II errors. This is particularly acute for low mean counts (most negative values of α_k), for some of which the Gaussian copula model always generates a significantly positive estimate of the correlation ρ . The Sarmanov model presented in this paper is far less likely to produce false significantly positive correlation between the counts. Thus, despite being numerically smaller than the estimates of the Gaussian copula, a significantly positive correlation estimate using the Sarmanov model provides a more powerful test to evaluate the existence of co-movements between counts. The small sample properties of these two estimators reassures our conviction that the results of the estimation of the Sarmanov model of the next section truly unveil the existence of strategic complementarities between the number of tariff options offered by competing cellular carriers.

5 Number of Tariff Options in Duopoly Competition

At the beginning of the 1980s, technology was a barrier for competition in cellular telephony, essentially because of the large size of bandwidth needed for transmission and the scarce radio spectrum available. To solve this problem the Federal Communications Commission (*FCC*) divided the U.S. into 305 non-overlapping markets corresponding to the Standard Metropolitan Statistical Areas (*SMSAs*). In 1981, the FCC set aside 50 MHz of spectrum in the 800 MHz band for cellular services. One of the two cellular channel blocks in each market —the B block or *wireline* license— was awarded to a local incumbent carrier, while the A block —the *nonwireline* license— was awarded by comparative hearing to an entrant carrier other than a local *wireline* incumbent. After awarding the first thirty *SMSA* licenses by means of this expensive and time consuming approach, rules were adopted in 1984 and 1986 to award the remaining *nonwireline* licenses through lotteries. Depending on the market, there were between 6 and 579 contenders for a single *nonwireline* license. The administrative decision to award the second license to one out of hundreds of applicants was customarily contested in court in a process that took several years. As the licenses were finally awarded, entrant firms had six month to be fully operative, something that was facilitated by the *FCC* requirement that the incumbent had to share its installed base of antennae with the entrant in this early stage of the market in order to promote competition.¹⁶

This section studies whether there were strategic consideration in designing the pricing strategies of local duopolists in local cellular telephone markets in the early U.S. cellular telephone industry. Tariffs in the early U.S. cellular industry were quite simple. A tariff option was normally a three-part tariff consisting of an allowance of “free” minutes per month, a fixed monthly fee, and a fixed rate per minute. Tariff options normally distinguished between peak (comprising on average

¹⁶ For an institutional and historical account of the poorly designed awarding process of licenses in the early U.S. cellular telephone industry see Hausman (2002), Parker and Röller (1997), or Murray (2002).

about 13 hours a day at that time) and off-peak marginal rates.¹⁷ I will thus focus on the total number of tariffs offered by the competing firms as strategic choice variable to sign up new customers with heterogeneous calling needs. Data available contain a complete description of the tariff options offered by any of the two firms present in the 100 largest markets of the U.S. between 1984 and 1988. I thus can compute the number of tariff options of each firm, PLANS. This information was collected by *Economic and Management Consultants International, Inc.* and reported in *Cellular Price and Marketing Letter*, Information Enterprises, various issues, 1984–1988. For year 1992, Marciano (2000) combined information for the same carriers from *Cellular Directions, Inc.*, the *Cellular Telephone Industry Association*, and direct interviews with managers.

Table 2 presents the market and firm specific characteristics used in the estimation.¹⁸ COMMUTING refers to the average daily commuting time in minutes in each city; POPULATION represents the number of inhabitants in each market measured in millions; EDUCATION is the median number of years of schooling; GROWTH is the average percent growth of the population in the 1980's; INCOME measures the median income in thousands of dollars; and BUSINESS accounts for the number of business in sectors with high demand for cellular services of each market and measured in thousands of firms.¹⁹ With the exception of GROWTH, all these variables are measured in logarithms. Two other interesting market indicators are MULTIMARKET and REGULATED. The former is the number of markets in which a particular couple of firms compete against each other.²⁰

¹⁷ Other value added services such as detailed billing, call waiting, no-answer transfer, call forwarding, three way calling, busy transfer, call restriction, and voice mail were priced independently and rarely bundled together with particular tariff options.

¹⁸ Other regressors are available but they are not significant in neither equation.

¹⁹ Businesses with potential high cellular demand include service firms, health care, professional, and legal services, contract construction, transportation, finance, insurance, and real estate. The source of all these demographics is the 1989 *Statistical Abstracts of the United States*; U.S. Department of Commerce, Bureau of the Census, using the Federal Communication Commission (*FCC*) Cellular Boundary Notices, 1982–1987, available in *The Cellular Market Data Book*, EMCI, Inc., as well as the 1990 U.S. Decennial Census.

²⁰ Busse (2000) addresses the relationship between multimarket contact on collusion so that the offering of certain tariff features allow firms to coordinate pricing.

Table 2: Descriptive Statistics

Variables	<i>Incumbent</i>		<i>Entrant</i>	
	Mean	Std.Dev.	Mean	Std.Dev.
PLANS	3.6402	1.2219	3.5541	1.3915
YEAR92	0.1199	0.3252	0.1199	0.3252
COMMUTING	3.1428	0.1512	3.1428	0.1512
POPULATION	0.0793	0.9583	0.0793	0.9583
EDUCATION	2.5752	0.0352	2.5752	0.0352
BUSINESS	3.2840	0.8876	3.2840	0.8876
GROWTH	0.9361	1.0274	0.9361	1.0274
INCOME	3.6406	0.1318	3.6406	0.1318
MULTIMARKET	3.1824	2.2808	3.1824	2.2808
REGULATED	0.5270	0.4997	0.5270	0.4997
AMERITECH	0.1554	0.3626	0.0942	0.2206
BELLATL	0.0574	0.2329	0.0671	0.1725
BELLSTH	0.0878	0.2833	0.0600	0.1652
CENTEL	0.0895	0.2857	0.0541	0.1623
CONTEL	0.0507	0.2195	0.0270	0.1204
GTE	0.1436	0.3510	0.0777	0.1970
MCCAW			0.2782	0.2473
NYNEX	0.0963	0.2952	0.0550	0.1734
PACTEL	0.0220	0.1467	0.0388	0.1354
SWBELL	0.1334	0.3403	0.0802	0.2174
USWEST	0.0895	0.2857	0.0566	0.1638

All variables are defined in the text. The number of observations is 592.

The latter is a dummy variable that indicates whether new tariffs need to be approved by the regulator.²¹ In order to control for potential firm effects, I also include firm dummies to identify the largest shareholder of each cellular carrier (available from the *FCC*). Only those carriers with at least 4% of licenses in this sample are identified.²² Lastly, YEAR92 identifies those observations from 1992, when arguably, the cellular market had matured.

²¹ Shew (1994) claims that the possibility of having request approval of new tariffs in the future prompts firms in this industry to offer an “excessive” number of options when they enter the market, the only time when they do not have to seek such approval as cost data is not yet available.

²² They are: *Ameritech Mobile* (AMERITECH), *Bell Atlantic Mobile* (BELLATL), *Bell South Mobile* (BELLSTH), *Century Cellular* (CENTEL), *Contel Cellular* (CONTEL), *GTE Mobilnet* (GTE), *McCaw Communications* (MCCAW), *Nynex Mobile* (NYNEX), *PacTel Mobile Access* (PACTEL), *SouthWest Bell* (SWBELL), and *US West Cellular* (USWEST).

Table 3: Frequency Distributions of Number of Tariff Options

Tariff Options	1984–1988				1992			
	<i>Incumbent</i>		<i>Entrant</i>		<i>Incumbent</i>		<i>Entrant</i>	
	Cases	Rel.Freq.	Cases	Rel.Freq.	Cases	Rel.Freq.	Cases	Rel.Freq.
1	14	0.0269	3	0.0423	51	0.0979	5	0.0704
2	71	0.1363	7	0.0986	76	0.1459	3	0.0423
3	198	0.3800	5	0.0704	122	0.2342	13	0.1831
4	128	0.2457	16	0.2254	162	0.3109	18	0.2535
5	63	0.1209	40	0.5634	55	0.1056	32	0.4507
6	47	0.0902	0	0.0000	55	0.1056	0	0.0000
Mean, (Var.)	3.5681	(1.4651)	4.1690	(1.3996)	3.4971	(1.9774)	3.9718	(1.4563)

Absolute and relative frequency distributions of the number of tariff options offered by each active firm.

Table 3 presents the marginal distribution of the total number of tariffs offered by incumbent and entrant carriers. Notice that incumbents only offer 3.5 and entrants 4 tariff options on average. When comparing pricing over time, it appears that there is a very slight reduction of options from 1984–1988 to 1992. This is consistent with the prediction of theoretical models of nonlinear pricing competition as markets mature and most potential customers have already signed up for one of the two carriers, *e.g.*, Armstrong and Vickers (2001) and Rochet and Stole (2002). Notice also that the unconditional distribution of tariff plans is always underdispersed, *i.e.*, the variance of the distribution of number of plans never exceeds the mean, which is the opposite of what most count data regression models address as the consequence of unobserved heterogeneity. These features, and in particular the low number of telephone options by two competing firms, make the Sarmanov model of this paper suitable to be used in the estimation of the determinants of the number of plans offered by different cellular carriers.

Table 4 presents the bivariate frequency of each combination of the number of tariff options offered by incumbent and entrant carriers. Simple unconditional association measures indicate that the number of tariff options appear to be strategic complements when we measure the association between these strategies regardless of any firm or market observed heterogeneity. It is clear from

Table 4: Correlation Among Number of Tariff Options

Plans	1984–1988							1992						
	1	2	3	4	5	6	All	1	2	3	4	5	6	All
1	9	0	1	4	0	0	14	0	0	1	1	1	0	3
2	20	35	11	4	0	1	71	2	1	1	2	1	0	7
3	9	15	55	68	26	25	198	1	0	2	1	1	0	5
4	8	19	42	36	9	14	128	0	0	4	3	9	0	16
5	5	7	9	34	7	1	63	2	2	5	11	20	0	40
6	0	0	4	16	13	14	47	0	0	0	0	0	0	0
All	15	76	122	162	55	55	521	5	3	14	18	32	0	72
Kendall’s τ	0.2928 (9.99)							0.1836 (2.26)						

Total cases for each combination of tariff options offered by the incumbent and entrant firm. Rows indicate the number of options of the entrant and columns those of the incumbent. Kendall’s τ measures the association among the number of tariff options. The corresponding absolute value t-statistics are shown in parentheses. There are 521 pairs of tariff strategies in the 1984–1988 sample and 72 pairs in the 1992 sample.

this table that in the 1984–1988 period, firms frequently offer either the same or very similar number of tariff options. Between 1984 and 1988, firms offered the same number of tariff options in 30% of cases while in 71% of cases, the difference between the number of tariff plans offered by the incumbent and the entrant does not exceed one. In the 1992 sample these percentages increase up to 71% and 75% of cases, respectively.

5.1 A Structural Justification

Unconditional measures of association between the number of tariff options offered by competing firms do not control for the effect of observable or unobservable heterogeneity to discount the effect of heterogeneity we need to postulate a model of firm behavior. In principle we could represent firms as selecting the number of tariff options to maximize their profits:

$$y_k \in \underset{y}{\operatorname{argmax}} = \pi_k(y, y_{-k}, \mathbf{x}_k, \epsilon; \vartheta_k, \beta_k), \quad (40)$$

where y_k is the number of tariff options that maximize expected profits $\pi_k(\cdot)$ of firm k with firm/market observable characteristics \mathbf{x}_k , unobservable environmental variables ϵ , and competitors offering y_{-k} tariff options. Variables in Table 2 are representative of firm/market characteristics \mathbf{x}_k while ϵ , a structural error component, may involve unobserved features such as the distribution of calling patterns of customers or their willingness to pay for cellular services in different markets. Parameter β_k identifies the effect of observable characteristics on the decision of firm k to offer fewer or more tariff options while ϑ_k would measure whether pricing practices of firms are strategic complements (positive sign) or strategic substitutes (negative sign). The particular functional form of the profit function and distribution of the structural error ϵ will account for the effect of unobserved heterogeneity.

An appropriate specification of the profit functions $\pi_k(\cdot)$ for $k = 1, 2, \dots, K$, will ensure that the corresponding nonlinear pricing game is supermodular. A game is supermodular if payoffs are monotone in the players' actions. In the present application it means that profits of firm k will *always* increase (or decrease) with the number of the options offered by any competitor. Supermodularity rules out the possibility that profits of firm k increase if a competitor increases her tariff offerings from 1 to 2 or from 3 to 4 but decreases if the competitor changes the number of tariff options from 2 to 3. Thus, a non-cooperative Nash equilibrium in the number of tariff options exists and it is the solution to the following system of reaction functions:²³

$$Y_k(y_k^*, y_{-k}^*, \mathbf{x}_k, \epsilon; \vartheta_k, \beta_k) = 0, \quad k = 1, 2, \dots, K. \quad (41)$$

A direct estimation of this system of equations involving count endogenous variables (even if reaction functions are linear equations) is far from straightforward. Coherence problems arise

²³ Observe that the strategy space is not a convex set and that the solution of the problem above is not defined on \mathbb{R} , but rather on \mathbb{N} . Topkis (1998, §4) and Vives (1990) make use of Tarski (1955) Fixed Point Theorem to show the existence of a non-cooperative solution of supermodular games defined on a lattice.

in the estimation of discrete choice simultaneous equations models when the specification involves actual dummies rather than an underlying continuous latent variables.²⁴ Firms decide whether to offer a countable number of tariff plans. The actual number of plans is not the result of maximizing over a continuous latent variable but rather the result of comparing the profitability of several combinations of tariff options for the firm and its competitors. Intuitively, coherence problems arise if a realization of the vector of unobservable structural errors ϵ can be associated with more than one profile of tariff offerings by the different competing firms.²⁵

Thus, rather than estimating the above system of reaction functions directly, we need to assume that its lattice-defined Nash equilibrium can be written as a system of reduced form, seemingly related count data regressions such as the one presented in Section 2. The dispersion parameters θ_k could capture the effect of individual unobservable heterogeneity, as it is commonly argued in the count data literature while ρ reflects the strategic complementarity between plan offerings. Evidently, in the absence of a particular specification for equation (40), the connection between the structural unobserved components ϵ and ϑ_k with θ_k and ρ have not been explicitly established. This discussion illustrates how the reduced form estimate of ρ in a Sarmanov model can be used to test the existence of complementarity between count strategies and eventually linked to a structural model of nonlinear pricing competition in tariff offerings.²⁶

5.2 Results

Table 5 presents the results of the estimation of the bivariate double-Poisson Sarmanov count data regressions model. Estimates capture the fact that the distribution of the number of tariffs are

²⁴ On coherence problems see Heckman (1978), Maddala (1983, §7.5), and particularly Schmidt (1981).

²⁵ See Miravete and Pernías (2010) for a detailed analysis of how coherence problems complicate testing for the existence of complementarities if firms' strategies are dichotomous.

²⁶ This is indeed the common practice in the empirical literature on management and strategy. See Arora (1996) and Arora and Gambardella (1990).

Table 5: Double Poisson – Sarmanov Regression

Variables	<i>Independent Regressions</i>				<i>Sarmanov Regression</i>					
	<i>Incumbent</i>		<i>Entrant</i>		<i>Incumbent</i>			<i>Entrant</i>		
CONSTANT	2.3986	(0.47)	-12.5831	(1.79)	2.3967	[25.87]	{52.93}	-12.6123	[27.21]	{61.25}
YEAR92	0.7212	(6.87)	0.6151	(4.18)	0.7115	[2.69]	{4.06}	0.6216	[2.98]	{3.68}
COMMUTING	-1.1548	(1.92)	1.5559	(2.06)	-1.1798	[11.88]	{14.25}	1.5674	[13.16]	{17.57}
POPULATION	-0.0489	(0.40)	0.0743	(0.59)	-0.0475	[0.41]	{0.54}	0.0652	[0.48]	{0.62}
EDUCATION	0.1227	(0.06)	2.8608	(0.97)	0.1284	[1.66]	{1.91}	2.8691	[20.19]	{36.52}
BUSINESS	0.0333	(0.26)	-0.2681	(2.01)	0.0237	[0.19]	{0.24}	-0.2694	[1.85]	{2.12}
GROWTH	0.0891	(1.51)	-0.4534	(6.76)	0.0874	[0.98]	{1.55}	-0.4500	[4.51]	{5.99}
INCOME	1.4633	(2.32)	1.3248	(1.64)	1.4941	[14.12]	{18.69}	1.3347	[13.57]	{18.36}
MULTIMARKET	0.0409	(1.80)	0.1082	(4.06)	0.0394	[0.80]	{1.60}	0.1037	[1.71]	{3.62}
REGULATED	0.0928	(0.87)	0.6520	(4.66)	0.0815	[0.54]	{0.76}	0.6342	[4.31]	{5.95}
AMERITECH	-0.2183	(0.87)	0.3169	(0.63)	-0.2299	[1.76]	{2.25}	0.2406	[1.80]	{2.07}
BELLATL	1.0770	(4.97)	0.1317	(0.31)	1.0957	[5.78]	{7.00}	0.0848	[0.54]	{0.62}
BELLSTH	-1.2825	(6.09)	-0.9200	(2.07)	-1.2362	[5.91]	{7.13}	-0.9414	[5.77]	{6.67}
CENTEL	-0.2719	(1.26)	1.3981	(2.94)	-0.2827	[1.58]	{2.02}	1.3036	[8.43]	{10.79}
CONTEL	-0.8500	(3.70)	-0.7116	(1.42)	-0.8524	[3.71]	{4.63}	-0.7340	[5.68]	{6.51}
GTE	-1.1022	(6.38)	-0.1997	(0.52)	-1.0929	[5.81]	{7.55}	-0.2429	[1.41]	{1.47}
MCCAW			0.8311	(2.76)				0.8508	[4.67]	{5.62}
NYNEX	0.9543	(5.30)	0.9591	(2.37)	0.9632	[5.53]	{6.32}	0.8880	[4.76]	{4.74}
PACTEL	-1.2295	(4.07)	-0.0734	(0.11)	-1.1967	[6.29]	{6.71}	-0.0748	[0.76]	{0.91}
SWBELL	-0.5886	(2.53)	0.0341	(0.06)	-0.5839	[3.49]	{4.55}	-0.0037	[0.03]	{0.04}
USWEST	-0.0150	(0.08)	0.7996	(1.69)	-0.0048	[0.03]	{0.03}	0.7491	[4.27]	{5.60}
θ	3.6324	(16.37)	2.3895	(15.24)	3.6585	[15.60]	{17.79}	2.4113	[13.82]	{18.33}
ρ						0.0396		[3.46]		{5.17}
$-\ln \mathcal{L}$	830.17		942.49		1,766.60					

Marginal effects evaluated at the sample mean of regressors. Endogenous variables are the number of tariff options of each competing firm. Absolute value t-statistics computed using a 2,000 replication, 10-step bootstrapping. Parentheses () report make use of rescaled bootstrapping while brackets [] refer to standard bootstrapping, respectively.

positively correlated and underdispersed. Table 5 also reports the estimates of the corresponding, restricted, independent, count data regression models. Table 5 shows that marginal effects are very similar. However the estimation of the correlation coefficient ρ is significant and the specification with independent count regression is rejected in favor of the bivariate double Poisson-Sarmanov model (likelihood ratio test of 12.12, 0.001 p-value). Since in addition of the sample considered, the effective range of the correlation coefficient is partially determined by the regressors included in the exponential mean function (3). Table 6 repeats the analysis without any regressors. Correlation is not significant but the specification is rejected in favor of that of Table 5.

Table 6: Double Poisson – Sarmanov Regression

Variables	<i>Independent Regressions</i>				<i>Sarmanov Regression</i>					
	<i>Incumbent</i>		<i>Entrant</i>		<i>Incumbent</i>			<i>Entrant</i>		
CONSTANT	4.6968	(40.10)	4.4959	(31.15)	4.6968	[16.13]	{39.36}	4.4959	[13.87]	{33.40}
θ	2.4171	(15.26)	1.7476	(14.88)	2.4171	[7.93]	{19.53}	1.7476	[7.96]	{19.42}
ρ					0.0000			[0.76]	{1.93}	
$-\ln \mathcal{L}$	955.11		1,040.50		1,995.61					

Marginal effects evaluated at the sample mean of regressors. Endogenous variables are the number of tariff options of each competing firm. Absolute value t-statistics computed using a 2,000 replication, 10-step bootstrapping. Parentheses () report make use of rescaled bootstrapping while brackets [] refer to standard bootstrapping, respectively.

The estimate of correlation between the number of tariff options offered by competing firms in Table 5 is small but positive and significant.²⁷ In view of the evidence from the Monte Carlo analysis of Section 4, we feel confident that such positive estimate of ρ is not the result of spurious correlation or misspecification of the model and thus, such result supports the view that the number of tariff options offered by competing cellular carriers are strategic complements.

Continuing with the effects of firm and market characteristics on the pricing decisions of cellular carriers, estimates show that ownership fixed effects are generally significant and indicate that the same firm group, in general, offer more tariff options in those markets where they act as new entrants relative to those markets where they are the incumbents. Many market characteristics have the same sign both as determinant of the number of tariffs of the incumbent and the entrant (although sometimes in one of the equations they fail to be significant). INCOME and YEAR92 have a positive effect on the number of tariffs offered by both competing carriers while MULTIMARKET only has a positive effect on the number of tariff options offered by the entrant. GROWTH and

²⁷ Notice that I have computed t-statistics based on two bootstrapping methods to evaluate whether the inference is sensible to dealing with the existence of bounds in the estimation given by the constraint (16). There are very few differences but in interpreting the results, I focus on the inference obtained making use of rescaled bootstrapping (t-statistics between brackets). This criteria is more conservative regarding the significance of parameters and also the appropriate one to deal with the existence of constraints involving the parameters estimate. Every estimation of the scaled bootstrap is run on a sample of size 1/3 of the full sample and where the same observation may be present multiple times.

BUSINESS have a negative effect on the number of tariff options offered by the entrant only, the latter being only marginally significant. These negative signs could be reconciled with situations where dynamic pricing considerations and switching cost are present. In the absence of a fast growing economy or if business customers are not numerous, the entrant has to offer several tariff options to segment the market of smaller users in order to induce them to subscribe. Offering just one or few tariffs that are less expensive than those offered by the incumbent will not secure the bulk of high valuation customers as the incumbent has previously targeted them and locked them in long term contracts. Finally, COMMUTING is the only variable that shows a significant opposite effect for each firm: in markets with longer commuting times entrants offer more tariff options than incumbents.

5.3 Interpreting the Sign of the Correlation Coefficient

The observed co-movements in the total number of plans offered by competing firms may respond not only to strategic interactions between firms who want to match competitors' practices (captured by ϑ_k above) but also by their attempt to extract as much surplus as possible from consumers' willingness to pay for cellular service (the effect of ϵ above) in a competitive environment.

Firms engaging in price discrimination commonly offer a few tariff options to screen a heterogeneous customer base. In principle the set of fully nonlinear tariffs offered by two competing firms are the best response to each other's tariffs given the distribution of consumer heterogeneity. The few existing theoretical results in this area show that equilibrium in nonlinear tariffs exists both in common agency or in exclusive agency environments (see Stole (2005) for an overview). However, such results refer to fully nonlinear tariffs rather than to how tariffs are commonly implemented, *i.e.*, through a menu of tariff options.

The use of few tariff options to screen consumer might be due to the existence of some commercialization costs or other marketing consideration. Thus, the foregone profits of an additional

tariff will eventually not compensate such cost, as foregone profits decline rapidly with the number of tariff options, *e.g.*, Wilson (1993, §8.3). Commercialization costs may refer not only to the cost of designing and selling this additional tariff option, but also the money value of the reputation effect that such strategy may have with customers who might value tariff complexity negatively. In any case, if this were the only reason determining the number of tariffs options offered by each carrier, we should expect that the number of tariffs offered by the first competitor (conditional on available firm and market characteristics) were uncorrelated with the number of tariff options offered by the second firm in the absence of synergies across commercialization costs of different firms.

Alternatively, with non-zero correlations between counts the number of tariff options offered becomes strategically relevant. If correlation among the conditional distribution of counts is positive, firms tend to offer a similar number of tariff options and their numbers are strategic complements. This environment responds to the equilibrium models of nonlinear pricing of Armstrong and Vickers (2001) and Rochet and Stole (2002) where firms end up offering similar, if not identical, two-part tariffs. This is the scenario supported by results of Table 5 and Table 6. On the contrary, if correlation among the number of tariff options offered is negative, firms might be attempting to differentiate their products through pricing and therefore use the number of tariff options as strategic substitute. Yang and Ye (2008) show that this situation could arise in mature markets where business stealing, rather than expanding the base of active customer, is the main effect of price discrimination. Results do not support this view, which is intuitive given the very early stage of development of the cellular industry during the 1980s.

6 Concluding Remarks

The Sarmanov count data model presented in this paper can accommodate both over and underdispersion and allows for the possibility that counts are not only positively but also negatively correlated. Furthermore, these two features of the joint distribution of counts are not driven by a common unobserved factor to all univariate marginal distributions and the parameterization of the likelihood function allows for all possible combinations of over or underdispersed marginals and correlation of any sign. The Sarmanov model thus overcomes most existing impediments to estimate a full multivariate count data regression model, thus reducing the risk of misspecification bias. Relative to alternative approaches commonly used in empirical finance such as the Gaussian-Poisson copula, the Sarmanov model presents better small sample properties in the sense that it rarely incurs in type II errors regarding the correlation coefficient estimate. Thus it provides a more powerful test of correlations between counts than copula based models.

Results from a particular application to study the pricing strategies of cellular carriers in the U.S. during the mid-1980s indicates that the number of tariff options offered by these firms can be considered strategic complements and, as postulated by Wilson (1993, §8.3), that competition in nonlinear tariffs is essentially driven by how heterogeneous the valuations of consumers are as well as by an attempt to match the competitor pricing practices to avoid losing early subscribers to the cellular service.

References

- AITCHISON, J. AND C. H. HO (1989): “The Multivariate Poisson-Lognormal Distribution.” *Biometrika*, 76, 643–653.
- ANDREWS, D. W. K. (1999): “Estimation When a Parameter Is on a Boundary.” *Econometrica*, 67, 1341–1383.
- ANDREWS, D. W. K. (2000): “Inconsistency of the Bootstrap When a Parameter Is on a Boundary of the Parameter Space.” *Econometrica*, 68, 399–405.
- ANDREWS, D. W. K. (2002): “Higher-Order Improvements of a Computationally Attractive k -Step Bootstrap for Extremum Estimators.” *Econometrica*, 70, 119–162.
- ANGUS, J. E. (1994): “The Probability Integral Transform and Related Results.” *SIAM Review*, 36, 652–654.
- ARMSTRONG, C. M. AND J. S. VICKERS (2001): “Competitive Price Discrimination.” *RAND Journal of Economics*, 32, 579–605.
- ARORA, A. (1996): “Testing for Complementarities in Reduced-Form Regressions: A Note.” *Economics Letters*, 50, 51–55.
- ARORA, A. AND A. GAMBARDELLA (1990): “Complementarity and External Linkages: The Strategies of the Large Firms in Biotechnology.” *The Journal of Industrial Economics*, 38, 361–379.
- BIEN, K., I. NOLTE, AND W. POHLMEIER (2009): “An Inflated Multivariate Integer Count Hurdle Model: An Application to Bid and Ask Quote Dynamics.” *Journal of Applied Econometrics*, 25, forthcoming.
- BUSSE, M. R. (2000): “Multimarket Contact and Price Coordination in the Cellular Telephone Industry.” *Journal of Economics & Management Strategy*, 9, 287–320.
- CAMERON, A. C., T. LI, P. K. TRIVEDI, AND D. M. ZIMMER (2004): “Modelling the Differences in Counted Outcomes Using Bivariate Copula Models with Application to Mismeasured Counts.” *Econometrics Journal*, 7, 566–584.
- CAMERON, A. C. AND P. K. TRIVEDI (1998): *Regression Analysis of Count Data*. New York, NY: Cambridge University Press.
- DENUIT, M. AND P. LAMBERT (2005): “Constraints on Concordance Measures in Bivariate Discrete Data.” *Journal of Multivariate Analysis*, 93, 40–57.
- EFRON, B. (1986): “Double Exponential Families and Their Use in Generalized Linear Regression.” *Journal of the American Statistical Association*, 81, 709–721.
- FISHER, M. AND I. KLEIN (2007): “Constructing Generalized FGM Copulas by Means of Certain Univariate Distributions.” *Metrika*, 65, 243–260.

- GOLDBERGER, A. S. (1983): “Abnormal Selection Bias.” In S. Karlin, T. Amemiya, and L. Goodman (eds.): *Studies in Econometrics, Time series, and Multivariate Statistics*. Academic Press.
- GOURIEROUX, C., A. MONFORT, AND A. TROGNON (1984): “Pseudo Maximum Likelihood Methods: Applications to Poisson Models.” *Econometrica*, 52, 701–720.
- GURMU, S. AND J. ELDER (2000): “Generalized Bivariate Count Data Regression Models.” *Economics Letters*, 68, 31–36.
- GURMU, S. AND J. ELDER (2008): “A Bivariate Zero-Inflated Count Data Regression Model with Unrestricted Correlation.” *Economics Letters*, 100, 245–248.
- HAUSMAN, J. A. (2002): “Mobile Telephone.” In M. Cave, S. K. Majumdar, and I. Vogelsang (eds.): *Handbook of Telecommunications Economics*, Vol. I. Amsterdam, The Netherlands: North-Holland.
- HAUSMAN, J. A., B. H. HALL, AND Z. GRILICHES (1984): “Econometric Models for Count Data with an Application to the Patents—R & D Relationship.” *Econometrica*, 52, 909–938.
- HAUSMAN, J. A., G. LEONARD, AND D. L. MCFADDEN (1995): “A Utility-Consistent, Combined Discrete Choice and Count Data Model: Assessing Recreational Use Losses due to Natural Resource Damage.” *Journal of Public Economics*, 56, 1–30.
- HECKMAN, J. J. (1978): “Dummy Endogenous Variables in a Simultaneous Equation System.” *Econometrica*, 46, 931–959.
- HEINEN, A. AND E. RENGIFO (2007): “Multivariate Autoregressive Modeling of Time Series Count Data Using Copulas.” *Journal of Empirical Finance*, 14, 564–583.
- HELLSTRÖM, J. (2006): “A Bivariate Count Data Model for Household Tourism Demand.” *Journal of Applied Econometrics*, 21, 213–226.
- HOROWITZ, J. L. (2001): “The Bootstrap.” In J. J. Heckman and E. Leamer (eds.): *Handbook of Econometrics*, Vol. 5. New York, NY: North-Holland.
- JOE, H. (1997): *Multivariate Models and Dependence Concepts*. New York, NY: Chapman & Hall/CRC.
- JOHNSON, N. L., N. BALAKRISHNAN, AND S. KOTZ (2000): *Continuous Multivariate Distributions*, Vol. 1, 2nd edn. New York, NY: John Wiley & Sons.
- JOHNSON, N. L., S. KOTZ, AND N. BALAKRISHNAN (1997): *Discrete Multivariate Distributions*. New York, NY: John Wiley & Sons.
- JUNG, C. J. AND R. WINKELMANN (1993): “Two Aspects of Labor Mobility: A Bivariate Poisson Regression Approach.” *Empirical Economics*, 18, 543–556.
- KOCHERLAKOTA, S. AND K. KOCHERLAKOTA (1993): *Bivariate Discrete Distributions*. New York, NY: Marcel Dekker.

- LEE, L. F. (1982): “Soma Approaches to the Correction of Selectivity Bias.” *Review of Economic Studies*, 49, 355–372.
- LEE, L. F. (1983): “Generalized Econometric Models with Selectivity.” *Econometrica*, 51, 507–513.
- LEE, M.-L. T. (1996): “Properties and Applications of the Sarmanov Family of Bivariate Distributions.” *Communications in Statistics, Part A — Theory and Methods*, 25, 1207–1222.
- MADDALA, G. S. (1983): *Limited Dependent and Qualitative Variables in Econometrics*. New York, NY: Cambridge University Press.
- MARCIANO, S. (2000): “Pricing Policies in Oligopoly with Product Differentiation: The Case of Cellular Telephony.” Ph.D. Thesis, University of Chicago Graduate School of Business.
- MARSHALL, A. W. AND I. OLKIN (1990): “Multivariate Distributions Generated from Mixtures of Convolution and Product Families.” In H. A. Block, A. R. Sampson, and T. H. Savits (eds.): *Topics in Statistical Dependence, IMS Lecture Notes Monograph*. Hayward, CA: Institute of Mathematical Statistics.
- MEGHIR, C. AND J.-M. ROBIN (1992): “Frequency of Purchase and Estimation of Demand Equations.” *Journal of Econometrics*, 53, 53–86.
- MIRAVETE, E. J. (2009): “Multivariate Sarmanov Count Data Models.” Working Paper 7463, CEPR.
- MIRAVETE, E. J. AND J. C. PERNÍAS (2010): “Testing for Complementarity when Strategies are Dichotomous.” *Economics Letters*, 106, 28–31.
- MUNKIN, M. AND P. K. TRIVEDI (1999): “Simulated Maximum Likelihood Estimation of Multivariate Mixed-Poisson Regression Models.” *Econometrics Journal*, 2, 29–48.
- MURRAY, J. B., JR (2002): *Wireless Nation. The Frenzied Launch of the Cellular Revolution in America*. Cambridge, MA: Perseus Publishing.
- NELSEN, R. B. (2006): *An Introduction to Copulas*, 2nd edn. New York, NY: Springer-Verlag.
- NOLTE, I. (2008): “Modeling a Multivariate Transaction Process.” *Journal of Financial Econometrics*, 6, 143–170.
- PARKER, P. M. AND L.-H. RÖLLER (1997): “Collusive Conduct in Duopolies: Multimarket Contact and Cross-Ownership in the Mobile Telephone Industry.” *RAND Journal of Economics*, 28, 304–322.
- POLITIS, D. M., J. P. ROMANO, AND M. WOLF (1999): *Subsampling*. New York, NY: Springer-Verlag.
- RIPHAHN, R. T., A. WAMBACH, AND A. MILLION (2003): “Incentive Effects in the Demand for Health Care: A Bivariate Panel Count Data Estimation.” *Journal of Applied Econometrics*, 18, 387–405.

- ROCHET, J.-C. AND L. A. STOLE (2002): “Nonlinear Pricing with Random Participation.” *The Review of Economic Studies*, 69, 277–311.
- SARMANOV, O. V. (1966): “Generalized Normal Correlation and Two-Dimensional Fréchet Classes.” *Soviet Mathematics — Doklady*, 168, 596–599.
- SCHMIDT, P. (1981): “Constraints on the Parameters in Simultaneous Tobit and Probit Models.” In C. F. Manski and D. L. McFadden (eds.): *Structural Analysis of Discrete Data with Applications*. Cambridge, MA: MIT Press.
- SHEW, W. B. (1994): “Regulation, Competition, and Prices in the U.S. Cellular Telephone Industry.” Mimeo, The American Enterprise Institute.
- SHUBINA, M. AND M.-L. T. LEE (2004): “On Maximum Attainable Correlation and Other Measures of Dependence for the Sarmanov Family of Bivariate Distributions.” *Communications in Statistics, Part A — Theory and Methods*, 33, 1031–1052.
- SKLAR, A. (1959): “Fonctions de répartition à n dimensions et leurs marges.” *Public Institute of Statistics of the University of Paris*, 8, 229–231.
- STOLE, L. A. (2005): “Price Discrimination and Imperfect Competition.” In C. M. Armstrong and R. H. Porter (eds.): *Handbook of Industrial Organization*, Vol. 3. New York, NY: North-Holland.
- TARSKI, A. (1955): “A Lattice-Theoretical Fixpoint Theorem and Its Applications.” *Pacific Journal of Mathematics*, 5, 285–308.
- TERZA, J. V. AND P. W. WILSON (1990): “Analyzing Frequencies of Several Types of Events: A Mixed Multinomial-Poisson Approach.” *Review of Economics and Statistics*, 72, 108–115.
- TOPKIS, D. M. (1998): *Supermodularity and Complementarity*. Princeton, NJ: Princeton University Press.
- VIVES, X. (1990): “Nash Equilibrium with Strategic Complementarities.” *Journal of Mathematical Economics*, 19, 305–321.
- WILSON, R. B. (1993): *Nonlinear Pricing*. New York, NY: Oxford University Press.
- WINDMEIJER, F. A. G. AND J. M. C. SANTOS-SILVA (1997): “Endogeneity in Count Data Models: An Application to Demand for Health Care.” *Journal of Applied Econometrics*, 12, 281–294.
- WINKELMANN, R. (1995): “Duration Dependence and Dispersion in Count-Data Models.” *Journal of Business & Economics Statistics*, 13, 467–474.
- WINKELMANN, R. (2000): “Seemingly Unrelated Negative Binomial Regression.” *Oxford Bulletin of Economics and Statistics*, 62, 553–560.
- YANG, H. AND L. YE (2008): “Nonlinear Pricing, Market Coverage, and Competition.” *Theoretical Economics*, 3, 123–153.