



Joint modelling of two count variables when one of them can be degenerate

Jacek Osiewalski¹ · Jerzy Marzec¹

Received: 22 February 2017 / Accepted: 24 July 2018 / Published online: 3 August 2018
© The Author(s) 2018

Abstract

We formulate a joint statistical model for two variables: one of them is either a count variable or just zero, and the other is a regular count variable. We consider a modelling framework based on switching between a bivariate Poisson regression model and a univariate one, where the switching depends on the observable outcome of the third, dichotomous variable. The ZIP–CP bivariate model (proposed quite recently) and the standard univariate Poisson regression model are used as basic elements of the switching (or mixture) model. Bayesian analysis is advocated; in two special cases of our Bayesian statistical model, consequences for inference are discussed. The empirical part is devoted to joint modelling of the numbers of cash payments and bank card payments in Poland, with the use of data for both cardholders and non-cardholders. Our Bayesian statistical test enables to examine whether it is appropriate to analyse each of two subsamples separately in order to infer on basic parameters. In the case of our data it is so, therefore inference on individual parameters is not affected by the sample selection error. However, inference on the correlation coefficient between two count variables is possible only within the proposed trivariate model.

Keywords Bivariate Poisson regression model · Zero inflated Poisson model · Bayesian inference · Count data models · Bank card payments · Cash payments

1 Introduction

Modelling univariate count data by means of Poisson type regression models is nowadays a routine approach, and some competing model specifications have been proposed

✉ Jerzy Marzec
marzecj@uek.krakow.pl
Jacek Osiewalski
eosiewa@cyf-kr.edu.pl

¹ Department of Econometrics and Operations Research, Cracow University of Economics, ul. Rakowicka 27, 31-510 Kraków, Poland

for bivariate count data as well (see, e.g. Lambert 1992; Kocherlakota and Kocherlakota 1992; Cameron and Trivedi 1998, 2005; Berkhout and Plug 2004; Famoye and Singh 2006; Lee et al. 2009; Winkelman 2008; Famoye 2010; Tsou 2016). It is worth mentioning that many empirical studies apply negative binomial regression because of the nature of count data. Multivariate count data occur in a wide range of applications, including accident analysis, sports statistics, economics, and many others (see e.g. Brijs et al. 2004, 2006; McHale and Scarf 2007; Ma et al. 2008; Baioa and Blangiardo 2010; Bermúdez and Karlis 2011; Shahtahmassebi and Moyeed 2016). Bayesian inference has sometimes been used in these applications. There are several approaches to the construction of the model for bivariate count data. The bivariate Poisson distribution based on the trivariate reduction method is restricted due to only positive correlation between two count variables (see Kocherlakota and Kocherlakota 1992; Brijs et al. 2006; Famoye 2010). Another approach is to model the joint distribution using copulas (see Van Ophem 1999; McHale and Scarf 2007); these models allow for more flexible specification of the dependence structure. Also, relatively flexible dependence structures appear in the models built upon the idea of the mixture of independent Poisson distributions (Aitchison and Ho 1989) or the conditional probabilities (Berkhout and Plug 2004). The models for multivariate count data are discussed in greater detail in, e.g. Cameron and Trivedi (1998) and Winkelman (2008).

In this paper we look at bivariate Poisson type regressions in the case when some observations follow a degenerate (i.e. univariate) distribution. Since a bivariate model for two non-degenerate count variables is the central part of our specification, we focus on some particular structure, following a simple and flexible modelling path, started by Berkhout and Plug (2004), which seems easy to generalize to dimensions higher than 2. However, our goal is to consider the possibility of degenerate bivariate observations, and not to propose one more new specification for related count variables.

In joint modelling of count variables we may face the situation when one of them is necessarily zero for many observed units. For example, if we analyze determinants of (and relation between) the numbers indicating how many times during a month people used public transport and how many times they used their own cars, then for any person without a car the number of using it is necessarily zero. It becomes crucial to examine the opportunities and consequences of inference (on determinants of using public transport and on the relationship between using public transport or private cars) on the basis of full data set in comparison to inference based on the data for car owners only. Using the latter means sample selection, which makes any generalization unjustified. In order to use all observations on two variables of interest, we propose a statistical model with switching between two specifications for count variables: a bivariate model and a univariate model. Switching is based on the third, zero–one variable (car ownership in the example above). Such approach enables to formulate testable hypotheses, e.g. that the mechanism generating values of the count variable which is always observed (never degenerate) is exactly the same in two groups of observed units.

The main part of the switching model, introduced in this paper, is a bivariate model of count variables representing the case where both variables are non-degenerate. We use the so-called ZIP–CP (*zero inflated Poisson–conditional Poisson*) specification, proposed by Marzec and Osiewalski (2012); it is a bivariate Poisson type regres-

sion, more general than P-CP (*Poisson-conditional Poisson*) model, introduced by Berkhout and Plug (2004). In the P-CP model, one of the variables is marginally Poisson and the other is conditionally Poisson. This model can be easily estimated and it allows correlation of any sign, but the sign of correlation between the two count variables depends on the sign of one parameter only and is independent of the explanatory variables of the model. In the ZIP-CP model of bivariate Poisson type regression, the marginal ZIP type distribution is used for the first variable (instead of the marginal Poisson distribution), which leads to the covariance sign dependent on the explanatory variables. The characteristics of the ZIP-CP model follow from the properties of the bivariate discrete ZIP-CP distribution, introduced and examined by Osiewalski (2012). The second part of the switching model proposed in this paper amounts to a univariate Poisson regression for the second variable—in case when the first variable is degenerate (with its full probability mass concentrated at zero). As it has been already mentioned, the third part is a dichotomous specification that describes switching between the bivariate (non-degenerate) and univariate (degenerate) case.

In the next section we present the probabilistic foundations of our switching model, i.e. the discrete distributions used to build the three parts of this model—in particular, the ZIP-CP distribution. Section 3 is devoted to our statistical model, the form of the likelihood function and the Bayesian analysis. Section 4 contains an empirical illustration, showing new results of joint modelling of the numbers of bank card and cash payments. In contrast to previous studies—see Polasik et al. (2012a) and Marzec and Osiewalski (2012)—we use all available data, for cardholders and non-cardholders. Our empirical example that refers to the research on non-cash transactions in Poland—see Polasik and Maciejewski (2009), Polasik (2015), Polasik et al. (2012b) and Goczek and Witkowski (2015, 2016)—serves as an illustration of modelling and inference problems with count variables, where one of them (the number of card payments) is degenerate for many observations (individuals without cards). In Sect. 5 concluding remarks are stated.

In the literature dealing with multivariate regression models we can find a variety of approaches to deal with dependent variables analyzed simultaneously of which at least one is partially observed. The practice of omitting data, also referred to as data selection, has been often used. For example, exploiting a dataset of over 11,000 payments, Bounie and Francois (2006) estimated the determinants of the probability of a transaction being paid by cash, check or bank card at the point of sale; a multinomial logit model was applied in their study. They excluded persons who did not hold bank cards or check accounts. The next simplification amounts to using variables which ignore the nature of the data. Kalckreuth et al. (2014) employed probit models for dependent variable. Although these practices seem useful, because they simplify the statistical modelling problem, they do not reflect the complexity of consumer behavior. From the viewpoint of statistical inference, such simplified approaches may be inappropriate, as we show in this paper. Stavins (2016) stated that “estimating consumers’ decisions to adopt and use payment instruments as independent events can lead to sample-selection problems”.

The aim of this paper is to present a new approach to payment behavior analysis by a tailor-made model, which is designed to cover all available observation units, e.g. consumers with or without cards. The advantage of our specification is that the

proposed trivariate model can capture salient features of the original data set. This modeling approach is free from the sample selection error and it allows to test for possible misspecification due to sample selection.

2 Probabilistic foundations of the new statistical model

We consider the joint distribution of three random variables (Y_1, Y_2, Y_3) , where the third one is a zero–one variable, the second variable can take any non-negative integer value, and the first one is concentrated at zero when $Y_3 = 0$ ($\Pr\{Y_1 = 0|Y_3 = 0\} = 1$), and can take any non-negative integer value if $Y_3 = 1$. Thus, when $Y_3 = 0$, the conditional distribution of (Y_1, Y_2) is the same as the distribution of $(0, Y_2)$ and corresponds to the univariate distribution of Y_2 . Only when $Y_3 = 1$ the distribution of the pair (Y_1, Y_2) is a bivariate distribution over the set of all pairs of non-negative integers; now we focus on its two special cases: P–CP (Berkhout and Plug 2004) and ZIP–CP (Osiewalski 2012). These distributions lead to particularly simple and useful bivariate Poisson type regression models. Other specifications impose restrictions on the correlation between two count variables or are more complicated from the statistical or numerical perspective.

If $Y_3 = 1$, the probability distribution of (Y_1, Y_2) is as follows:

$$\begin{aligned}\Pr\{Y_1 = i, Y_2 = j|Y_3 = 1\} &= \Pr\{Y_1 = i|Y_3 = 1\} \\ \Pr\{Y_2 = j|Y_3 = 1, Y_1 = i\} &= g(i) h(j, i),\end{aligned}\quad (1)$$

where $i, j \in N \cup \{0\}$. If the distribution of Y_1 is Poisson with mean (and variance) λ_1 and the conditional distribution of Y_2 given Y_1 is Poisson with mean (and variance) $\lambda_2 \exp(\alpha Y_1)$, i.e.

$$g(i) = \exp(-\lambda_1)(\lambda_1)^i / i!, \quad h(j, i) = \exp[-\lambda_2 \exp(\alpha \cdot i)](\lambda_2)^j \exp(\alpha \cdot i \cdot j) / j!, \quad (2)$$

then we have the bivariate P–CP distribution with the following moments (Berkhout and Plug 2004):

$$E(Y_2|Y_3 = 1) = \lambda_2 \exp[\lambda_1(e^\alpha - 1)], \quad (3)$$

$$E((Y_2)^2|Y_3 = 1) = E(Y_2|Y_3 = 1) + [E(Y_2|Y_3 = 1)]^2 \exp[\lambda_1(e^\alpha - 1)^2], \quad (4)$$

$$\text{Var}(Y_2|Y_3 = 1) = E(Y_2|Y_3 = 1) + [E(Y_2|Y_3 = 1)]^2 \{\exp[\lambda_1(e^\alpha - 1)^2] - 1\}, \quad (5)$$

$$E(Y_1 Y_2|Y_3 = 1) = \lambda_1 e^\alpha E(Y_2|Y_3 = 1). \quad (6)$$

If $\alpha \neq 0$, then the variance (5) of Y_2 is greater than its expectation (3). The dependence between the two variables leads to the inflated variance of Y_2 , which is usually observed in empirical count data. The Poisson distribution of Y_1 does not have this crucial property. This is the first reason to generalize the bivariate P–CP distribution through replacing the marginal Poisson distribution of Y_1 by a ZIP type distribution—in line of the approach presented by Lambert (1992), Cameron and Trivedi (1998, 2005)

and Winkelman (2008). The second reason to generalize the P-CP model lies in its restrictive approach to the dependence between two count variables, as the sign of covariance between Y_1 and Y_2 , i.e. of

$$\begin{aligned} Cov(Y_1, Y_2|Y_3 = 1) &= E(Y_1 Y_2|Y_3 = 1) \\ &- E(Y_1|Y_3 = 1)E(Y_2|Y_3 = 1) = \lambda_1(e^\alpha - 1)E(Y_2|Y_3 = 1), \end{aligned} \tag{7}$$

depends only on the sign of the real constant α , and not on λ_1 or λ_2 , which are parameterized through explanatory variables in statistical applications of this probabilistic model.

The generalization proposed by Osiewalski (2012) allows for dependence of the sign of covariance on λ_1 as well. This more general class of bivariate discrete distributions (denoted with a star) is characterized by the same conditional distribution of Y_2 given Y_1 :

$$Pr^*\{Y_2 = j|Y_3 = 1, Y_1 = i\} = h(j, i) = Pr\{Y_2 = j|Y_3 = 1, Y_1 = i\} \tag{8}$$

and by the ZIP-type distribution of Y_1 , with zero treated separately:

$$Pr^*\{Y_1 = i|Y_3 = 1\} = g^*(i) = \begin{cases} \gamma & \text{for } i = 0, \\ \frac{1-\gamma}{1-g(0)}g(i) & \text{for } i \in N, \end{cases} \tag{9}$$

where γ belongs to the $(0, 1)$ interval, and g and h are the same functions as in (2). If $\gamma = g(0)$, then $Pr^*\{Y_1 = i|Y_3 = 1\} = g^*(i) = g(i) = Pr\{Y_1 = i|Y_3 = 1\}$ and we are back in the P-CP case. If $\gamma > g(0)$, then the distribution of Y_1 is of the ZIP type, so we call the joint distribution ZIP-CP. However, note that the specification (9) is more general as it also allows $\gamma < g(0)$.

The moments of the ZIP-CP distribution are related to the moments of the P-CP case through the following general formula (assuming $0^m = 1$ for $m = 0$):

$$\begin{aligned} E^*(Y_1^m Y_2^n|Y_3 = 1) &= g_0[(1 - \gamma)E(Y_1^m Y_2^n|Y_3 = 1) + (\gamma - g(0))0^m] \\ E(Y_2^n|Y_3 = 1, Y_1 = 0), \end{aligned} \tag{10}$$

where $g_0 = (1 - g(0))^{-1}$.

In particular:

$$E^*(Y_1|Y_3 = 1) = g_0(1 - \gamma)E(Y_1|Y_3 = 1) = g_0(1 - \gamma)\lambda_1, \tag{11}$$

$$E^*(Y_1^2|Y_3 = 1) = g_0(1 - \gamma)E(Y_1^2|Y_3 = 1) = g_0(1 - \gamma)\lambda_1(1 + \lambda_1), \tag{12}$$

$$E^*(Y_2|Y_3 = 1) = g_0[(1 - \gamma)E(Y_2|Y_3 = 1) + (\gamma - g(0))\lambda_2], \tag{13}$$

$$E^*(Y_2^2|Y_3 = 1) = g_0[(1 - \gamma)E(Y_2^2|Y_3 = 1) + (\gamma - g(0))\lambda_2(1 + \lambda_2)], \tag{14}$$

$$\begin{aligned} E^*(Y_1 Y_2|Y_3 = 1) &= g_0(1 - \gamma)E(Y_1 Y_2|Y_3 = 1) = g_0(1 - \gamma)\lambda_1 e^\alpha E(Y_2|Y_3 = 1), \\ &\tag{15} \end{aligned}$$

$$Var^*(Y_1|Y_3 = 1) = \frac{1 - \gamma}{1 - g(0)}\lambda_1 \left(1 + \frac{\gamma - g(0)}{1 - g(0)}\lambda_1 \right), \tag{16}$$

$$\begin{aligned}
 \text{Var}^*(Y_2|Y_3 = 1) &= \frac{1 - \gamma}{1 - g(0)} \left\{ \text{Var}(Y_2|Y_3 = 1) + \frac{\gamma - g(0)}{1 - g(0)} [E(Y_2|Y_3 = 1) - \lambda_2]^2 \right. \\
 &\quad \left. + \frac{\gamma - g(0)}{1 - \gamma} \lambda_2 \right\}, \tag{17}
 \end{aligned}$$

$$\begin{aligned}
 \text{Cov}^*(Y_1, Y_2|Y_3 = 1) &= \frac{1 - \gamma}{1 - g(0)} \left\{ \text{Cov}(Y_1, Y_2|Y_3 = 1) + \frac{\gamma - g(0)}{1 - g(0)} \lambda_1 [E(Y_2) - \lambda_2] \right\}, \\
 &\tag{18}
 \end{aligned}$$

which leads to the correlation coefficient of the form

$$\begin{aligned}
 \text{Corr}^*(Y_1, Y_2|Y_3 = 1) &= \frac{\text{Cov}(Y_1, Y_2|Y_3 = 1) + \frac{\gamma - g(0)}{1 - g(0)} \lambda_1 (E(Y_2|Y_3 = 1) - \lambda_2)}{\sqrt{\lambda_1 \left(1 + \frac{\gamma - g(0)}{1 - g(0)} \lambda_1 \right) \left\{ \text{Var}(Y_2|Y_3 = 1) + \frac{\gamma - g(0)}{1 - g(0)} (E(Y_2|Y_3 = 1) - \lambda_2)^2 + \frac{\gamma - g(0)}{1 - \gamma} \lambda_2 \right\}}}, \\
 &\tag{19}
 \end{aligned}$$

where $E(Y_2|Y_3 = 1)$, $\text{Var}(Y_2|Y_3 = 1)$ and $\text{Cov}(Y_1, Y_2|Y_3 = 1)$ are the moments of the P-CP distribution in (3), (5) and (7). After simple manipulations we obtain

$$\begin{aligned}
 \text{Cov}^*(Y_1, Y_2|Y_3 = 1) &= (1 - g(0))^{-2} (1 - \gamma) \lambda_1 \left[(1 - g(0)) e^\alpha E(Y_2|Y_3 = 1) \right. \\
 &\quad \left. - (1 - \gamma) E(Y_2|Y_3 = 1) - (\gamma - g(0)) \lambda_2 \right] \\
 &= (1 - e^{-\lambda_1})^{-2} (1 - \gamma) \lambda_1 \lambda_2 \left\{ [(1 - e^{-\lambda_1}) e^\alpha - (1 - \gamma)] \right. \\
 &\quad \left. \exp(\lambda_1 (e^\alpha - 1)) - \gamma + e^{-\lambda_1} \right\}. \tag{20}
 \end{aligned}$$

Now it is clear that the variables (Y_1, Y_2) that follow the ZIP-CP distribution

1. are negatively correlated, if $[(1 - e^{-\lambda_1}) e^\alpha - (1 - \gamma)] \exp(\lambda_1 (e^\alpha - 1)) < \gamma - e^{-\lambda_1}$,
2. are positively correlated, if $[(1 - e^{-\lambda_1}) e^\alpha - (1 - \gamma)] \exp(\lambda_1 (e^\alpha - 1)) > \gamma - e^{-\lambda_1}$,
3. are uncorrelated, if $[(1 - e^{-\lambda_1}) e^\alpha - (1 - \gamma)] \exp(\lambda_1 (e^\alpha - 1)) = \gamma - e^{-\lambda_1}$.

When $\gamma = g(0) = \exp(-\lambda_1)$, i.e. if Y_1 is Poisson (under $Y_3 = 1$), the complicated formulas (18) and (20) reduce to the much simpler form (7), where the sign of covariance depends only on the sign of α . In other cases, i.e. when Y_1 is of ZIP type, the sign of covariance in (20) depends on the values of λ_1 and α (not only on the sign of the latter constant). Obviously, the value of covariance in the ZIP-CP distribution (and not only its sign) as well as the value of the correlation coefficient (19) depend on all the constants appearing in the ZIP-CP probability function, i.e. on $\gamma, \lambda_1, \lambda_2$ and α .

Remind that increasing the probability of the zero value of Y_1 (in comparison to the Poisson distribution with mean and variance λ_1), that is assuming the ZIP type distribution with $\gamma > g(0)$, leads to variance (16) greater than expectation (11). The ZIP-CP distribution class enables inflating variances of both count variables, although they are not symmetrically treated.

As yet our considerations has been focused on the conditional distribution of the pair (Y_1, Y_2) given $Y_3 = 1$, that is on the complicated part of our trivariate structure.

The distribution of Y_2 given $Y_3 = 0$ (and the only possible zero value of Y_1) is specified in such a way as to make it easy to test in our statistical model whether the conditional distribution of Y_2 given $Y_1 = 0$ is the same in both situations: $Y_3 = 0$ and $Y_3 = 1$. Therefore we assume the Poisson distribution with the probability function

$$\Pr\{Y_2 = j | Y_3 = 0\} = \Pr\{Y_2 = j | Y_3 = 0, Y_1 = 0\} = h_0(j) = \exp(-\lambda_{2,0}) (\lambda_{2,0})^j / j!, \tag{21}$$

with $\lambda_{2,0}$ possibly different from λ_2 .

Summing up all the assumptions we have already introduced, we propose the following joint distribution of three discrete variables:

$$\Pr\{Y_1 = i, Y_2 = j, Y_3 = l\} = \begin{cases} p g^*(i) h(j, i), & i, j \in N \cup \{0\}, l = 1, \\ (1 - p) h_0(j), & i = 0, j \in N \cup \{0\}, l = 0, \\ 0, & i \in N, j \in N \cup \{0\}, l = 0, \end{cases} \tag{22}$$

where $p = \Pr\{Y_3 = 1\}$. The marginal distribution of the pair (Y_1, Y_2) is a particular mixture of the bivariate ZIP-CP distribution and the univariate Poisson distribution:

$$\Pr\{Y_1 = i, Y_2 = j\} = p g^*(i) h(j, i) + (1 - p) I_{\{0\}}(i) h_0(j), \quad i, j \in N \cup \{0\}, \tag{23}$$

where $I_A(\cdot)$ denotes the characteristic function of the set A ; the moments can be written as:

$$E(Y_1^m Y_2^n) = p E^*(Y_1^m Y_2^n | Y_3 = 1) + (1 - p) 0^m E(Y_2^n | Y_3 = 0, Y_1 = 0), \tag{24}$$

with $E^*(Y_1^m Y_2^n | Y_3 = 1)$ denoting the appropriate moment of the ZIP-CP distribution, see (10), and $E(Y_2^n | Y_3 = 0, Y_1 = 0)$ coming from the Poisson distribution with parameter $\lambda_{2,0}$.

3 The Bayesian statistical model

Consider T trivariate observations $(Y_{1t}, Y_{2t}, Y_{3t}; t = 1, 2, \dots, T)$, where Y_{3t} are dichotomous. For $Y_{3t} = 1$, pairs (Y_{1t}, Y_{2t}) have different ZIP-CP distributions, i.e.

$$\Pr^*\{Y_{1t} = i, Y_{2t} = j | Y_{3t} = 1\} = g_t^*(i) h_t(j, i) \quad (i, j \in N \cup \{0\}), \tag{25}$$

where

$$\Pr^*\{Y_{1t} = i | Y_{3t} = 1\} = g_t^*(i) = \begin{cases} \gamma_t & \text{for } i = 0, \\ \frac{1 - \gamma_t}{1 - g_t(0)} g_t(i) & \text{for } i \in N; \end{cases} \quad g_t(i) = e^{-\lambda_{1t}} (\lambda_{1t})^i / i!, \tag{26}$$

$$\Pr^*\{Y_{2t} = j | Y_{3t} = 1, Y_{1t} = i\} = h_t(j, i) = \exp[-\lambda_{2t} e^{\alpha \cdot i}] (\lambda_{2t})^j e^{\alpha \cdot i \cdot j} / j!, \tag{27}$$

$$\lambda_{1t} = \exp(\mathbf{x}_t \boldsymbol{\beta}_1), \quad \lambda_{2t} = \exp(\mathbf{w}_t \boldsymbol{\beta}_2), \quad \gamma_t = \exp(-e^\delta \lambda_{1t}) = \exp(-\exp(\delta + \mathbf{x}_t \boldsymbol{\beta}_1)), \tag{28}$$

\mathbf{x}_t and \mathbf{w}_t are row vectors consisting of values of explanatory variables that determine the probabilities of particular pairs of values of Y_{1t} and Y_{2t} . The role of the explanatory variables depends on the column vectors of parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, but also on the dependence parameter α and the ZIP parameter δ , which governs the deviation of $\Pr^*\{Y_{1t} = 0 | Y_{3t} = 1\}$ from the value corresponding to the Poisson distribution. Now the moments of the distribution of (Y_{1t}, Y_{2t}) given Y_{3t} , presented in the previous section, depend on the explanatory variables.

The specification based on (26) is known in the literature as the *hurdle model* (Cameron and Trivedi 2005, p. 680); Winkelman (2008) compares it to the original ZIP model. The *hurdle model* form of our ZIP type specification is very simple, thus making estimation and testing quite easy.

When $Y_{3t} = 0$, the pairs $(Y_{1t}, Y_{2t}) = (0, Y_{2t})$ have degenerate distributions, where for Y_{2t} we assume Poisson distributions—as in (24); that is

$$\Pr\{Y_{2t} = j | Y_{3t} = 0, Y_{1t} = 0\} = h_{0,t}(j) = \exp[-\lambda_{2t,0}] (\lambda_{2t,0})^j / j!, \tag{29}$$

$$\lambda_{2t,0} = \exp(\mathbf{w}_t \boldsymbol{\beta}_{2,0}).$$

If $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_{2,0}$, then $\Pr^*\{Y_{2t} = j | Y_{3t} = 1, Y_{1t} = 0\} = \Pr\{Y_{2t} = j | Y_{3t} = 0, Y_{1t} = 0\}$ and the mechanism that generates values of Y_{2t} given $Y_{1t} = 0$ is exactly the same, no matter what the value of Y_{3t} is. In order to test the hypothesis $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_{2,0}$ we need a trivariate statistical model. Under our assumptions, this model amounts to the following parametric class of distributions:

$$\Pr\{Y_{1t} = i, Y_{2t} = j, Y_{3t} = l; \theta\} = \begin{cases} p_t g_t^*(i) h_t(j, i), & i, j \in N \cup \{0\}, l = 1, \\ (1 - p_t) h_{0,t}(j), & i = 0, j \in N \cup \{0\}, l = 0, \\ 0, & i \in N, j \in N \cup \{0\}, l = 0, \end{cases} \tag{30}$$

where $p_t = \Pr\{Y_{3t} = 1\} = 1 - F(-\mathbf{z}_t \boldsymbol{\beta}_3)$, \mathbf{z}_t is the row vector of explanatory variables and F is the distribution function representing the particular dichotomous model for Y_{3t} . In the empirical section we use the logit model, i.e. we assume that F is the distribution function of the logistic distribution. Other models of the dichotomous variable Y_{3t} are worth considering, especially the one based on the skewed Student t distribution, which Osiewalski and Marzec (2004a, b) introduced as a relatively general alternative for the logit and probit specifications. In our statistical model for the triple (Y_{1t}, Y_{2t}, Y_{3t}) the parameter vector $\boldsymbol{\theta}$ is a column grouping $\delta, \alpha, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3$ and $\boldsymbol{\beta}_{2,0}$. We assume that, for any $\boldsymbol{\theta}$, trivariate observations are stochastically independent.

When $Y_{1t} = y_{1t}, Y_{2t} = y_{2t}$ and $Y_{3t} = y_{3t}$ ($t = 1, 2, \dots, T$) have been observed, the likelihood function takes the form

$$\begin{aligned}
 L(\theta; y) &= \left[\prod_{t: y_{3t}=1, y_{1t}=0} \gamma_t h_t(y_{2t}, 0) \right] \left[\prod_{t: y_{3t}=1, y_{1t}>0} \frac{1 - \gamma_t}{1 - g_t(0)} g_t(y_{1t}) h_t(y_{2t}, y_{1t}) \right] \\
 &\quad \left[\prod_{t: y_{3t}=0, y_{1t}=0} h_{0,t}(y_{2t}) \right] \left[\prod_{t: y_{3t}=1} p_t \right] \left[\prod_{t: y_{3t}=0} (1 - p_t) \right] \\
 &= L_1(\beta_1, \beta_2, \alpha, \delta) L_2(\beta_{2,0}) L_3(\beta_3), \tag{31}
 \end{aligned}$$

where y denotes the $(3 \times T)$ matrix of the observed values of Y_{1t} , Y_{2t} and Y_{3t} . The first two products in (31) correspond to the bivariate component of the mixture model and form the function L_1 of $\delta, \alpha, \beta_1, \beta_2$; the third product in (31) corresponds to the univariate Poisson component and is the function L_2 of $\beta_{2,0}$; the fourth and fifth products in (31) correspond to the dichotomous switching variable and constitute the function L_3 of β_3 . If there is no relation among these three groups of parameters, then inference on each of them can be conducted separately, using only the appropriate function L_r ($r = 1, 2, 3$) instead the full likelihood function. The situations of “no relations” or their presence can be precisely formalized within the Bayesian statistics, where a probability measure (prior distribution) on the parameter space is defined, prior independence between parameters can be formally stated and posterior independence can be considered. Here we focus on two situations: the case of prior independence among the three groups of parameters and the case of $\beta_2 = \beta_{2,0}$.

Under the separability of the likelihood function, obvious from (31), prior independence among $(\delta, \alpha, \beta_1, \beta_2), \beta_{2,0}$ and β_3 leads to their posterior independence, which means complete separability of inference on each group of parameters. In this case, using only observations with $y_{3t} = 1$ for estimating $(\delta, \alpha, \beta_1, \beta_2)$ is fully justified as well as using only observations with $y_{3t} = 0$ for estimating $\beta_{2,0}$ alone. Obviously, inference on such functions of θ that involve parameters from different groups, e.g. on $Corr(Y_{1t}, Y_{2t}|\theta)$ —the unconditional correlation coefficient between the first two elements of the triple (Y_{1t}, Y_{2t}, Y_{3t}) , must be based on the joint posterior density of $\theta, p(\theta|y)$, which uses the full likelihood function and complete data. The joint posterior is needed if one wants to compare the unconditional correlation coefficient $Corr(Y_{1t}, Y_{2t}|\theta)$ and the conditional one, $Corr(Y_{1t}, Y_{2t}|Y_{3t} = 1, \theta) = Corr^*(Y_{1t}, Y_{2t}|Y_{3t} = 1, \delta, \alpha, \beta_1, \beta_2)$, derived in the ZIP-CP model using formula (19).

In the case of $\beta_2 = \beta_{2,0}$, when (given $Y_{1t} = 0$) Y_{2t} is explained in exactly the same way no matter what Y_{3t} is, L_1 and L_2 in the likelihood function cannot be considered separately as both depend on β_2 . In this case inference has to be based on all data, the full likelihood and the joint posterior. Making inferences with the use of the data with $y_{3t} = 1$ only would mean sample selection error. Of course, testing $\beta_2 = \beta_{2,0}$ requires the general model, without this restriction.

Complete specification of our Bayesian statistical model [with the sampling distribution (30) that leads to the likelihood function (31)] requires the prior distribution of θ . Obviously, our prior choice is related to the model structure, not to the data that are analysed in the empirical part. We assume prior independence and the standard normal prior $N(0, 1)$ for each parameter. Zero prior expectations mean that the simplest model (with no ZIP effect, no dependence and no explanatory variables) gets the highest prior

chance, but unitary standard deviations ensure significant prior chances for specifications being far from the simplest one. It seems that such simple joint prior distribution introduces little initial information and guarantees easy Monte Carlo simulations of the posterior distribution. Obviously, sensitivity of inferences with respect to the form of the prior distribution is an empirical question, to be answered for the data at hand, but it is of greater importance mainly in small data-sets. According to basic Bayesian asymptotic results, under any regular prior, the posterior based on a sufficiently large number of observations can be approximated by an appropriate multivariate normal distribution centred at the maximum likelihood estimate. Thus, in empirical studies based on large data-sets, sensitivity with respect to the prior distribution becomes much less important.

In this study we implement the random-walk Metropolis–Hastings MCMC algorithm to simulate samples from the posterior distribution of θ (Gamerman 1998). This algorithm was started either at zero values of the parameters or at maximum likelihood estimates obtained by estimating each sub-model separately (due to separability of the likelihood function). It turned out that the selection of starting values was not important for convergence. We generated a candidate random variable from a multivariate Student distribution; preliminary runs were used to calibrate its precision matrix. The algorithm involved 1,000,000 cycles, and the acceptance rate was about 10%. Convergence of single chains from the MCMC sampler was confirmed by the graphical procedure proposed by Yu and Mykland (1998).

4 Joint modelling of the numbers of card and cash payments

In order to illustrate the empirical usefulness of the proposed statistical model, we use the data collected for the research that was financed by the National Bank of Poland and described by Polasik et al. (2012a) and Marzec et al. (2013). The data consist of the information whether person t is a cardholder (y_{3t}) as well as the number of his/her cash payments (y_{2t}) and card payments (y_{1t}) within a month. $T = 2518$ persons were questioned in October or November 2010, or in January 2011. The fraction of cardholders was 47.3%.

Frequency distributions of the numbers of cash payments for $y_{3t} = 1$ and $y_{3t} = 0$ are presented in Table 1. For non-cardholders, the average number of cash payments during a month was 22.5 (with the empirical standard deviation 19.8); for cardholders, the average number of cash payments during a month was lower: 20.5 (with the empirical standard deviation 17.3). The value $W^2 = 10.9$ of the modified test statistic of Anderson and Darling (1954) indicates dissimilarity of these two discrete distributions. For cardholders, the average number of card payments during a month is 5 (with standard deviation 6.7); the empirical correlation coefficient between y_{1t} and y_{2t} (given $y_{3t} = 1$) is 0.008, which indicates no linear dependence.

The results obtained by Polasik et al. (2012a)—within the P–CP model on the basis of the data for 1190 cardholders—showed very small positive correlation between the numbers of cash and card payments. Marzec and Osiewalski (2012) confirmed this using the ZIP–CP model, indicating at the same time that the P–CP model is not a valid reduction of the more general ZIP–CP case, as both parameters α and

Table 1 Frequency distributions of the numbers of cash payments y_{2t} for cardholders ($y_{3t} = 1$) and non-cardholders ($y_{3t} = 0$). *Source:* Own calculations

Number of cash payments y_{2t} (interval)	Number of cardholders (t such that $y_{3t} = 1$)	Frequency	Number of non-cardholders (t such that $y_{3t} = 0$)	Frequency
0	24	2%	0	0%
(0; 5]	126	11%	60	5%
(5; 10]	248	21%	275	21%
(10; 15]	196	16%	224	17%
(15; 20]	148	12%	208	16%
(20; 25]	108	9%	151	11%
(25; 30]	85	7%	123	9%
(30; 35]	66	6%	73	5%
(35; 40]	55	5%	57	4%
(40; 45]	32	3%	55	4%
(45; 50]	32	3%	26	2%
>50	70	6%	76	6%
Sum	1190	100%	1328	100%
Mean	20.5	–	22.5	–
Median	16	–	18	–

δ are significantly different from zero. Note that univariate empirical distributions for cardholders only (i.e. with $y_{3t} = 1$) require a bivariate model with inflated zeros for both count variables; the ZIP-CP specification meets this requirement, while the P-CP model does not. Moreover, for cardholders, formal Bayesian model comparison led to the conclusion that, in the ZIP-CP, model Y_{1t} must represent the number of card payments and Y_{2t} —cash payments (not vice versa); see Marzec and Osiewalski (2012). The necessity of establishing which count variable is the first one comes from the asymmetric structure of the bivariate model under consideration.

Now we present the results obtained for the full dataset, which includes non-cardholders. Similarly as Marzec and Osiewalski (2012), we have modelled raw data, without weights indicating the degree of representativeness of individual observations; such weights were used by Polasik et al. (2012a) and Marzec et al. (2013). The motivation to use weighted (adjusted) data amounts to adequately represent the population from which the sample has been drawn. Information about demographic characteristics, such as gender, age, marital status, and place of residence, are used to develop the weights. In this paper we model raw data, as we do not focus on representativeness issues.

The structure of our complete trivariate model—shown in Fig. 1—consists of two separate count variables models: for $T_1 = 1190$ pairs (Y_{1t}, Y_{2t}) with $Y_{3t} = 1$ and for $T_2 = 1328$ variables Y_{2t} if $Y_{3t} = 0$, as well as of the specification for the dichotomous variable Y_{3t} that links all $T = T_1 + T_2 = 2518$ observations. The same main characteristics of the questioned individuals are used as explanatory variables in all three parts of our joint model, that is $x_t = w_t = z_t$.

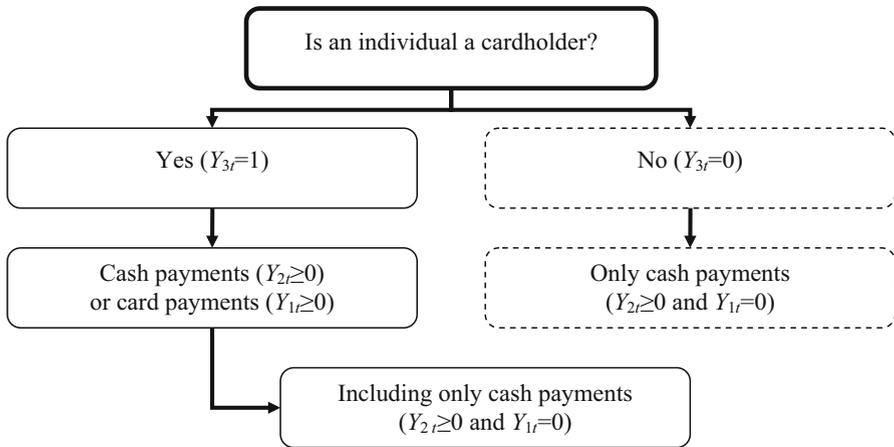


Fig. 1 The structure of the full trivariate statistical model. Source: authors' elaboration

Table 2 Typical (average or most frequent) values of explanatory variables. Source: authors' elaboration

Explanatory variable	$T = 2518$ All observations	$T_1 = 1190$ Cardholders	$T_2 = 1328$ Non-cardholders
Gender (1—man, 0—woman)	0	0	0
Age (years)	41.2	40.1	42.2
Marital status (1—married, 0—not married)	1	1	0
Residence (1—city, 0—countryside)	1	1	1
Monthly income (thousand PLN)	2.9	3.3	2.5
Education (years of schooling)	12.3	13.2	11.5
Access to Internet at home (1—yes, 0—no)	1	1	0

In Table 2 we present the typical values of our explanatory variables, i.e. the most frequent values for zero–one variables and the arithmetic means for other variables. It seems that the main determinants of having a bank card are: income, education, marital status and the access to Internet at home. The role of the access to Internet and marital status, as well as of the place of residence, is also suggested by the information presented in Table 3 (i.e. the fraction of ones in the case of dichotomous explanatory variables). We assume that the access to Internet is a proxy variable for consumer openness to technology adaptation.

In our empirical research we have used the statistical model presented in (30), together with the joint prior distribution, proposed in the previous section and assuming independence among all parameters of the trivariate model. Taking advantage of posterior independence, which results from the separability of the likelihood in (31) and prior independence, we have used three independent Metropolis–Hastings chains in order to simulate from the posterior distribution in each part of our model. That is, we have separately estimated $(\beta_1 \beta_2, \alpha, \delta)$ in the ZIP–CP model (M_1) , $\beta_{2,0}$ in the

Table 3 Fraction (%) of ones in the case of dichotomous explanatory variables. *Source:* authors' elaboration

Explanatory variable	$T = 2518$	$T_1 = 1190$	$T_2 = 1328$
	All observations	Cardholders	Non-cardholders
Gender (1—man, 0—woman)	44	45	42
Marital status (1—married, 0—not married)	56	65	48
Residence (1—city, 0—countryside)	63	71	56
Access to Internet at home (1—yes, 0—no)	61	76	49

Poisson model for the number of cash payments for non-cardholders (M_2) and β_3 in the logit model M_3 . The total number of parameters is 34.

In Table 4 we present the posterior means and standard deviation of all individual parameters; the results are printed in bold if the absolute value of the posterior mean is greater than two posterior standard deviations.

Referring to the assumed prior distribution of parameters, we see that our $N(0, 1)$ priors appear relatively vague in this application, because the posterior standard deviations are much lower than the prior standard deviations and almost all posterior means are in the interval $(-2, 2)$, and most of them in $[-1, 1]$. We checked that our results are robust to changes in the prior distribution. On the other hand, in studies (like ours) where the number of observations is large, prior sensitivity becomes much less important.

For cardholders we see (in M_1) that all seven explanatory variables that we have used are obviously important to explain the number of cash payments. But only the access to Internet, education and income significantly (and positively) affect the number of card payments. In the pure Poisson model for non-cardholders (M_2), not all seven variables are important to explain the number of cash payments—gender, income and age are not. Our results show that a cardholder's education, being in a marriage and the access to Internet have a negative effect on cash payments. Note, however, that the impact of these three variables on the number of cash payments is positive for non-cardholders. Living in a city will lead to more frequent use of cash as the payment method for both consumer types. Additionally, there is significant positive influence of age only on cash payments in the case of cardholders. In the logit model (M_3), five variables (except for gender and age) are the determinants of possessing a bank card. We confirmed that being in a marriage, living in a city, having higher income, staying in education for a longer period and having the access to Internet increase the probability of having a bank card.

Let us stress the differences between posterior distributions of the parameters describing the number of cash payments in M_1 and M_2 . In the case of four explanatory variables (marital status, income, education and the access to Internet), the signs of the posterior means are different. As the standard deviations of most of the parameters are small, we suspect that the equality $\beta_2 = \beta_{2,0}$ does not hold.

In order to verify the hypothesis $\beta_2 = \beta_{2,0}$ we use a Lindley-type Bayesian test (similar to the highest posterior density interval test, see Lindley 1965 p. 58; Zellner 1971, pp. 298–302). Let $\kappa = \beta_2 - \beta_{2,0}$; building upon the classical F or Chi squared

Table 4 Posterior means and standard deviations of the parameters of each part of the trivariate model. *Source:* authors' calculations

Variable/parameter	M_1 (ZIP-CP): card payments	$E(\cdot y)$	$D(\cdot y)$	M_3 : logit model for having a bank card	$E(\cdot y)$	$D(\cdot y)$
"1"		0.911	0.098		-5.455	0.330
Gender		-0.044	0.025		0.181	0.092
Age		-0.002	0.001		0.001	0.003
Marital status		-0.048	0.029		0.596	0.102
Residence		-0.007	0.028		0.483	0.096
Income		0.051	0.010		0.185	0.039
Education		0.056	0.006		0.297	0.024
Internet		0.361	0.039		0.622	0.106
"1"		2.825	0.050	M_2 : Poisson model – cash payments (for non-cardholders)	2.133	0.051
Gender		-0.101	0.013		-0.014	0.013
Age		0.008	0.001		0.002	0.001
Marital status		-0.158	0.015		0.082	0.015
Residence		0.145	0.015		0.126	0.015
Income		0.016	0.006		-0.009	0.006
Education		-0.008	0.003		0.062	0.003
Internet		-0.085	0.016		0.152	0.016
α	-	0.004	0.001		-	-
δ	-	-1.876	0.041		-	-

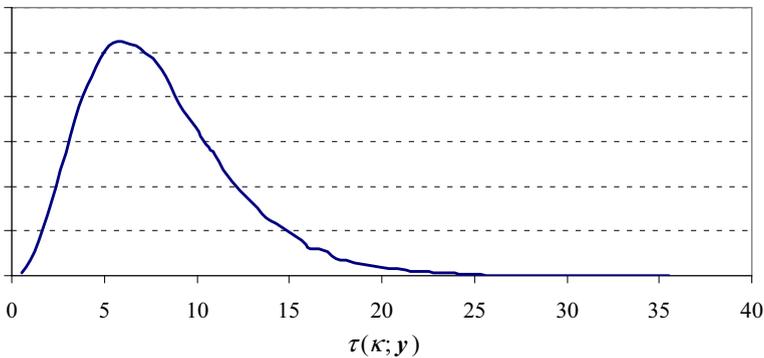


Fig. 2 Posterior distribution of $\tau(\kappa; y)$. *Source:* authors' calculations

tests, we consider the following quadratic form (see also Osiewalski and Steel 1993; Marzec and Osiewalski 2008):

$$\tau = \tau(\kappa; y) = (\kappa - E(\kappa|y))'(V(\kappa|y))^{-1}(\kappa - E(\kappa|y)), \tag{32}$$

where $E(\kappa|y) = E(\beta_2|y) - E(\beta_{2,0}|y)$ and $V(\kappa|y) = V(\beta_2|y) + V(\beta_{2,0}|y)$; the sum of covariance matrices is a result of posterior independence between β_2 and $\beta_{2,0}$ in the general model (without any restrictions). Univariate variable τ is random as a function of both the observations and parameters of our Bayesian model. Inferences on τ are based on its posterior distribution with the density function $p(\tau|y)$. In our Lindley-type approach, testing the restriction $\kappa = 0$ amounts to checking whether the value $\tau(0; y)$ belongs to the region of the highest posterior density $p(\tau|y)$ and the close to one posterior probability mass. If so, we do not reject the hypothesis $\kappa = 0$ and go to the model based on this restriction, which enables the separate analysis of two subsamples (of cardholders and non-cardholders). If $\tau(0; y)$ is outside the highest posterior density region, then the equality $\kappa = 0$ is not supported by the data—so we reject it and stay with the results obtained in the general, unrestricted model.

In Fig. 2 we present the posterior density $p(\tau|y)$. The value $\tau(0; y) = 973.85$ lies far in the tail of the posterior distribution of τ , so the equality $\beta_2 = \beta_{2,0}$ is strongly rejected. Thus, for our dataset, inferences (on individual parameters) based on the separability of the likelihood in (31) are free from any sample selection error.

Finally we present the posterior results for the conditional and unconditional correlation coefficients between the numbers of card and cash payments (Y_{1t}, Y_{2t}). Remind that the unconditional correlation coefficient $Corr(Y_{1t}, Y_{2t}|\theta)$ is a function of all parameters of the three sub-models, so its posterior distribution can be obtained only in the joint trivariate model, irrespectively of the outcome of the test we have considered above. In Table 5 we present main results. For all data ($T = 2518$) we have obtained the posterior distributions of the unconditional correlation coefficient $Corr(Y_{1t}, Y_{2t}|\theta)$ that were concentrated close to zero—but only on the positive side. The minimum posterior mean was 0.031, the maximum was 0.16 and the average posterior mean was 0.072 (always with a relatively small posterior standard deviation). It means that the unconditional correlation between the numbers of card and cash payments is very small,

Table 5 Posterior means of correlation coefficients between (Y_{1t}, Y_{2t}) , averaged over observations. *Source:* authors' calculations

Correlation coefficient	Average posterior mean	
$Corr(Y_{1t}, Y_{2t} \theta)$	0.072 (for $T = 2518$)	
–	Cardholders ($Y_{3t} = 1$)	Non-cardholders ($Y_{3t} = 0$)
$Corr(Y_{1t}, Y_{2t} \theta)$	0.065 (for $T_1 = 1190$)	0.079 (for $T_2 = 1328$)
$Corr(Y_{1t}, Y_{2t} Y_{3t} = 1, \theta)$	0.073 (for $T_1 = 1190$)	–

but positive. In the ZIP–CP model—for the cardholders only—the average posterior mean of the conditional correlation coefficient $Corr(Y_{1t}, Y_{2t}|Y_{3t} = 1, \theta)$ was 0.073. The overall average estimate of the unconditional correlation coefficient is practically the same as the average estimate of the conditional correlation coefficient given $y_{3t} = 1$, although the specific average estimates of $Corr(Y_{1t}, Y_{2t}|\theta)$ obtained for cardholders and non-cardholders are quite different.

5 Concluding remarks

The trivariate discrete distribution and Bayesian statistical model have been proposed in order to jointly model two count variables in the case where one of them can be degenerate. Our statistical model amounts to using a zero–one variable to switch between two separate models for count variables. The first model is bivariate and the second one is only univariate—but from the same class as the conditional part of the bivariate model. While the proposed modelling scheme is quite general, the choice of the sub-models (the building blocks of the trivariate structure) is rather specific and can be changed. Simplicity is the main criterion in choosing the ZIP–CP model (as the bivariate specification for count variables) and the logistic model (for the zero–one switching variable); both lead to a tractable trivariate model. Replacing the logistic part by a different dichotomous specification—e.g. based on a skewed Student t distribution and allowing for interactions of explanatory variables (see Osiewalski and Marzec 2004b)—is not difficult and may improve the data fit. However, replacing the ZIP–CP specification, which is the main part of our trivariate model, would be much more difficult. Using alternative structures for two related count variables is left for future research.

As far as the prior specification is concerned, our particular form of the prior distribution can easily be changed, but two crucial properties should be kept in mind. The separability of the likelihood function can be fully exploited only under prior independence of parameters describing sub-models, so their independence is a natural element of each prior specification. Also remind that particular, standard normal prior distribution (that we have assumed for each individual parameter) is not important if the number of observations is large—like in our empirical example. Obviously, small samples would require sensitivity analysis within a larger class of prior distributions (e.g. Student t with unknown degrees of freedom).

In the proposed Bayesian model one can easily use our Lindley-type test (the Bayesian counterpart of the F or Chi squared tests) in order to verify the fundamental restriction, which makes the parameters describing the non-degenerate count variable identical for both values of the switching zero–one variable. It would be interesting to use formal Bayesian model comparison (through Bayes factors and posterior model probabilities) for testing different specifications that could appear in future research. This would require an efficient estimator of the marginal data density value in each model. It seems that, in the case of the Markov Chain Monte Carlo simulations of the posterior distribution, the corrected arithmetic mean estimator proposed by Pajor (2017) is an appropriate tool.

Our trivariate model is constructed in such a way that separability of the likelihood function is preserved. Thus it is a useful tool to examine the consequences of sample selection caused by deleting all observations with only one non-degenerate count variable (i.e. deleting the whole subsample of non-cardholders). In our empirical example we have shown that inference on individual parameters is not affected by the sample selection error, since the restriction linking parameters of two sub-models is not supported by the data. We have also shown that any deeper inference on correlation between two count variables—that is, inference on the unconditional correlation coefficient as opposed to the conditional one—is possible only within our full trivariate specification.

Let us stress that the proposed trivariate model always enables making inference for all available data, without applying any preliminary tests. Instead, our model itself constitutes a useful testing framework, in particular for testing particular conditions that lead to sample selection errors. This is the main contribution of the paper.

Acknowledgements The authors acknowledge support from research funds granted to the Faculty of Management at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Aitchison J, Ho CH (1989) The multivariate Poisson-log normal distribution. *Biometrika* 76:643–653
- Anderson TW, Darling DA (1954) A test of goodness of fit. *J Am Stat Assoc* 49:765–769
- Baioa G, Blangiardo M (2010) Bayesian hierarchical model for the prediction of football results. *J Appl Stat* 37(2):253–264
- Berkhout P, Plug E (2004) A bivariate Poisson count data model using conditional probabilities. *Stat Neerl* 58:349–364
- Bermúdez L, Karlis D (2011) Bayesian multivariate Poisson models for insurance ratemaking. *Insur Math Econ* 48:226–236
- Bounie D, Francois A (2006) Cash, check or bank card? The effects of transaction characteristics on the use of payment instruments. <http://ssrn.com/paper=891791>—SSRN eLibrary. Accessed 26 July 2017
- Brijs T, Karlis D, Swinnen G, Vanhoof K, Wets G, Marchanda P (2004) A multivariate Poisson mixture model for marketing applications. *Stat Neerl* 58:322–348

- Brijs T, Van de Bossche F, Wets G, Karlis D (2006) A model for identifying and ranking dangerous accident locations: a case study in Flanders. *Stat Neerl* 60:457–476
- Cameron AC, Trivedi PK (1998) *Regression analysis of count data*. Cambridge University Press, New York
- Cameron AC, Trivedi PK (2005) *Microeconometrics: methods and application*. Cambridge University Press, New York
- Famoye F (2010) On the bivariate negative binomial regression model. *J Appl Stat* 37:969–981
- Famoye F, Singh KP (2006) Zero-inflated generalized Poisson regression model with an application to domestic violence data. *J Data Sci* 4:117–130
- Gamerman D (1998) *Markov chain Monte Carlo. Stochastic simulation for Bayesian inference*. Chapman and Hall, London
- Goczek Ł, Witkowski B (2015) The determinants of cash-free transactions. The National Bank of Poland Working Paper Series no. 146
- Goczek Ł, Witkowski B (2016) Determinants of card payments. *Appl Econ* 48:1530–1543
- Kalckreuth U, Schmidt T, Stix H (2014) Choosing and using payment instruments: evidence from German microdata. *Empir Econ* 46:1019–1055
- Kocherlakota S, Kocherlakota K (1992) *Bivariate discrete distributions*. Marcel Dekker, New York
- Lambert D (1992) Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 34:1–14
- Lee J, Jung BC, Jin SH (2009) Tests for zero inflation in a bivariate zero-inflated Poisson model. *Stat Neerl* 63:400–417
- Lindley DV (1965) *Introduction to probability and statistics from a Bayesian view point. Part 2: inference*. Cambridge University Press, Cambridge
- Ma J, Kockelman K, Damien P (2008) A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accid Anal Prev* 40:964–975
- Marzec J, Osiewalski J (2008) Bayesian inference on technology and cost efficiency of bank branches. *Bank i Kredyt* 39:29–43
- Marzec J, Osiewalski J (2012) Dwuwymiarowy model typu ZIP-CP w łącznej analizie zmiennych licznikowych. *Folia Oeconomica Cracoviensia* 53:5–20
- Marzec J, Polasik M, Fiszeder P (2013) Wykorzystanie gotówki i karty płatniczej w punktach handlowo-usługowych w Polsce: zastosowanie dwuwymiarowego modelu Poissona. *Bank i Kredyt* 44:375–402
- McHale I, Scarf P (2007) Modelling soccer matches using bivariate discrete distributions with general dependence structure. *Stat Neerl* 61:432–445
- Ophem Van H (1999) A general method to estimate correlated discrete random variables. *Econ Theory* 15:228–237
- Osiewalski J (2012) Dwuwymiarowy rozkład ZIP-CP i jego momenty w analizie zależności między zmiennymi licznikowymi, [in:] *Spotkania z królową nauk (Księga jubileuszowa dedykowana Profesorowi Edwardowi Smadze)*. Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie, Kraków, pp 147–154
- Osiewalski J, Marzec J (2004a) Uogólnienie dychotomicznego modelu probitowego z wykorzystaniem skośnego rozkładu Studenta. *Przełąd Statystyczny* 51:13–24
- Osiewalski J, Marzec J (2004b) Model dwumianowy II rzędu i skośny rozkład Studenta w analizie ryzyka kredytowego. *Folia Oeconomica Cracoviensia* 45:63–83
- Osiewalski J, Steel MF (1993) Una perspectiva bayesiana en selección de modelos. *Cuadernos Economicos* 55/3, pp 327–351 (A Bayesian perspective on model selection, original English version available at: <http://www.cyfronet.krakow.pl/~eosiewa/pubo.htm>)
- Pajor A (2017) Estimating the marginal likelihood using the arithmetic mean identity. *Bayesian Anal* 12:261–287
- Polasik M (2015) Stan i potencjał rozwoju sieci akceptacji kart płatniczych w Polsce. *Acta Universitatis Nicolai Copernici, Oeconomia* 46:23–58
- Polasik M, Maciejewski K (2009) Innowacyjne usługi płatnicze w Polsce i na świecie. *Materiały i Studia NBP* no. 241, NBP, Warszawa
- Polasik M, Marzec J, Fiszeder P, Górka J (2012a) Modelowanie wykorzystania metod płatności detalicznych na rynku polskim. *Materiały i Studia NBP* no. 265, NBP, Warszawa
- Polasik M, Wiśniewski TP, Lightfoot G (2012b) Modelling customers' intentions to use contactless cards. *Int J Bank Acc Finance* 4:203–231
- Shahtahmassebi G, Moyeed R (2016) An application of the generalized Poisson difference distribution to the Bayesian modelling of football scores. *Stat Neerl* 70(3):260–273

- Stavins J (2016) The effect of demographics on payment behavior: panel data with sample selection. Federal Reserve Bank of Boston Working Paper No. 16-5
- Tsou TS (2016) Robust likelihood inference for multivariate correlated count data. *Comput Stat* 31:845–857
- Winkelmann R (2008) *Econometric analysis of count data*. Springer, Berlin
- Yu B, Mykland P (1998) Looking at Markov samplers through cusum path plots: a simple diagnostic idea. *Stat Comput* 8:275–286
- Zellner A (1971) *An introduction to Bayesian inference in econometrics*. Wiley, New York