O P E R A T I O N S   R E S E A R C H   A N D   D E C I S I O N S
No. 3–4 2010

Robert KAPŁON*

HETEROGENEITY IN MODELS OF PURCHASE FREQUENCY. A COMPARISON OF POISSON-GAMMA MIXTURES WITH FINITE POISSON MIXTURES

Poisson models are fundamental in the modelling of purchase frequencies. However, very often they are statistically incompatible with the data. This stems from the fact that the mean is assumed to be equal to the variance and, in consequence, this fails to capture heterogeneity. Thus Poisson mixture models are often considered instead. The most commonly used of these models is the Poisson-gamma mixture model, which is very often applied to problems in marketing. Hence, it would be advisable to discover its limitations. Using real marketing data sets, we point out the limitations of this approach. Furthermore, we compare it with finite Poisson mixtures.

Keywords: *Poisson mixture models, heterogeneity, purchase frequency*

1. Introduction

A count refers to the number of times an event occurs and is treated as the realization of a nonnegative integer-valued random variable. Examples of such variables are: visits to the doctor and other types of health care utilization, occupational injuries and illnesses, absenteeism in the workplace, recreational or shopping trips, claims on automobile insurance, labour mobility, entry and exits from an industry, takeover activity in business, defaults on mortgage repayments and loans, bank failures, patent registration in connection with industrial research and development [2]. Dividing counts by the total sample size yields frequencies, e.g. the frequency of purchases by customers.

As one may note, count data is common in many disciplines, therefore many statistical models have been proposed [4], [6], [7], [16]. The most commonly used count

*Institute of Organization and Management, Wrocław University of Technology, Smoluchowskiego 25, 50-372 Wrocław, Poland, e-mail: robert.kaplon@pwr.wroc.pl

models are based on the Poisson distribution. However, this leads to *overdispersion,* because the variability of the data, as measured by the variance, exceeds the mean. Violation of this mean-variance property may be due to unobserved heterogeneity. This implies that a more complex and reliable model is needed.

As a remedy, mixtures of distributions have been widely used. Two categories of mixture distributions play a prominent role. First, when a certain distribution of the mean and the variance is imposed in the Poisson model. As a result one can obtain distributions such as the negative binomial distribution, Sichel's distribution, the Neyman type A distribution, the Poisson–Pascal distribution or the Polya–Aeppli distribution [8]. Another category of model concerns the specification of a finite mixture in which the underlying distribution of counts is approximated by a finite number of Poisson distributions with different parameters. This stems from the assumption that each observed count can be viewed as arising from a subpopulation.

BROCKETT et al. [1] considered a wide class of Poisson mixture models and contrasted them with the well known negative binomial distribution (NBD). In their empirical study, they showed that the generalized compound Poisson–Pascal distribution (GCPP) strongly outperforms Poisson-gamma mixture distributions (NBD). The Pearson goodness-of-fit statistic indicated that the NBD never fitted their marketing data, in contrast to the GCPP. These findings are in contradiction with the commonly held belief about the robustness of NBD. Unfortunately, they did not take finite mixtures of Poisson distributions into account. It might be argued that this category of mixture distribution is very flexible and allows a researcher to estimate the parameters of such a model in a very simple way. Thus the aim of this paper is to compare these two categories of models – Poisson-gamma mixtures and finite Poisson mixtures – in the context of an application to data on purchase frequency.

2. Poisson-gamma mixture model

Let N be the number of purchases of a product during the observed period. The conditional probability mass function of N given that the purchase rate $\Lambda = \lambda$ is the Poisson distribution defined by

$$Pr(N = n \mid \Lambda = \lambda) = \frac{\lambda^n \exp(-\lambda)}{n!}, \qquad n = 0, 1, \ldots$$

Heterogeneity can be captured by setting a distribution for Λ. Therefore, different Poisson mixture models can be defined depending on the assumptions made regarding the distribution of Λ One can assume: the gamma distribution, the inverse Gaussian distribution, and the log-normal distribution [17]. The first of these distributions is most commonly used, because the Poisson-gamma mixture model has a closed form that leads to the well-known negative binomial distribution. To see this, let g be the gamma density function with parameters (α, β) [13]:

$$g(\lambda \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} \mathrm{e}^{-\beta\lambda}, \qquad \alpha, \beta > 0,$$

The unconditional probability distribution can be found by integrating with respect to λ:

$$Pr(N = n \mid \alpha, \beta) = \int_0^{\infty} Pr(N = n \mid \Lambda = \lambda; \alpha, \beta) g(\lambda \mid \alpha, \beta) d\lambda$$

$$= \int_0^{\infty} \frac{\lambda^n \exp(-\lambda)}{\Gamma(n-1)} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} d\lambda = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)\Gamma(n+1)} \left(\frac{\beta}{1+\beta}\right)^{\alpha} \left(\frac{1}{1+\beta}\right)^{n}.$$

This is Ehrenberg's NBD model. Ehrenberg confirmed the usefulness of this model many times over 50 years of research: the NBD was found to fit the data, regardless of whether big, medium or small brands of very varied grocery-type products from soap to soup were considered. This was examined for various countries, means of analysis, points of time, shorter and longer periods of analysis, younger and older consumers [5]. The NBD gave a very close fit to the observed purchasing data. What is clear, then, is that the NBD serves as a reference point or benchmark for comparison to other models.

The parameters can be estimated using the method of maximum likelihood. In order to derive estimators, we formulate the likelihood:

$$L(\alpha, \beta \mid \mathbf{n}, \mathbf{k}) = \prod_{i=0}^{I} \left[\frac{\Gamma(\alpha+n_i)}{\Gamma(\alpha)\Gamma(n_i+1)} \left(\frac{\beta}{1+\beta}\right)^{\alpha} \left(\frac{1}{1+\beta}\right)^{n_i} \right]^{k_i} \tag{1}$$

where n_i is the number of purchases of the product and k_i is a count. Maximizing expression (1) with respect to α, β is equivalent to maximizing the log likelihood. After some manipulation, we obtain:

$$\log L(\alpha, \beta \mid \mathbf{n}, \mathbf{k}) \propto \alpha K \log \frac{\beta}{1+\beta} + \sum_{i=1}^{I} k_i \sum_{j=1}^{n_i} \log(\alpha + n_i - j) + K\overline{n} \log \frac{\beta}{1+\beta},$$

$$K = \sum_{i=0}^{I} k_i, \qquad \overline{n} = K^{-1} \sum_{i=0}^{I} n_i k_i.$$

Equating the first order partial derivatives to zero, we obtain:

$$\hat{\beta} = \frac{\alpha}{\bar{n}},$$
$$\sum_{i=1}^{I} k_i \sum_{j=1}^{n_i} \frac{1}{\alpha + n_i - j} + K \log \frac{\alpha}{\bar{n} + \alpha} = 0, \tag{2}$$

and the second equation needs to be solved numerically.

Inference concerning the estimated parameters requires computing their variances. For this purpose, we use the observed Fisher information matrix (see [11]). Thus the variance-covariance matrix takes the form:

$$\mathrm{Var}(\hat{\alpha}, \hat{\beta}) = -\begin{bmatrix} h_{\hat{\beta}\hat{\beta}} & h_{\hat{\alpha}\hat{\beta}} \\ h_{\hat{\alpha}\hat{\beta}} & h_{\hat{\alpha}\hat{\alpha}} \end{bmatrix}^{-1},$$

where

$$h_{\hat{\beta}\hat{\beta}} = \left.\frac{\partial^2 \log L(\alpha, \beta)}{\partial \beta^2}\right|_{\beta=\hat{\beta}, \alpha=\hat{\alpha}} = K(\hat{\alpha} + \bar{n})\left(\frac{1}{(1+\hat{\beta})} - \frac{1}{\hat{\beta}^2}\right),$$
$$h_{\hat{\alpha}\hat{\alpha}} = \left.\frac{\partial^2 \log L(\alpha, \beta)}{\partial \alpha^2}\right|_{\alpha=\hat{\alpha}} = -\sum_{i=0}^{I} k_i \sum_{j=1}^{n_i} \frac{1}{(\hat{\alpha} + n_i - j)^2},$$
$$h_{\hat{\alpha}\hat{\beta}} = \left.\frac{\partial^2 \log L(\alpha, \beta)}{\partial \beta \partial \alpha}\right|_{\beta=\hat{\beta}} = K\left(\frac{1}{\hat{\beta}} - \frac{1}{1+\hat{\beta}}\right).$$

3. Finite mixture Poisson model

An alternative approach to modelling heterogeneity is to assume that each individual belongs to one class or subpopulation C_s (s = 1, …, S). Unfortunately, we do not know in advance which class each individual belongs to. Thus we presume that a randomly selected individual in the sample belongs to class C_s with the probability π_s, satisfying the following constraints [10]:

$$\sum_{s=1}^{S} \pi_s = 1, \qquad \pi_s > 0.$$

The conditional probability mass function of N, given that N comes from class C_s, is:

$$\Pr(N = n \mid i \in C_s; \lambda_s) = \frac{\lambda_s^n \exp(-\lambda_s)}{n!}, \qquad n = 0,1,\ldots$$

Hence, the unconditional probability mass function of N has the following mixture form:

$$\Pr(N = n; \boldsymbol{\pi}, \boldsymbol{\lambda}) = \sum_{s=1}^{S} \pi_s \frac{\lambda_s^n \exp(-\lambda_s)}{n!}, \qquad n = 0,1,\ldots$$

A variety of algorithms can be used to estimate the parameters via the method of maximum likelihood such as gradient methods like Fisher's scoring, Newton–Raphson or quasi Newton–Raphson. In the context of finite mixture models, it is advisable to use the EM algorithm [3], [9]. In accordance with the EM algorithm, we introduce non-observed data z_{is} that are independent and identically distributed with a multinomial distribution. It is further assumed that the N_i are conditionally independent given z_{is}. Hence, the complete log-likelihood function can be written as:

$$\log L_c(\boldsymbol{\Psi} \mid \mathbf{N}, \mathbf{Z}) = \sum_{i=0}^{I} \sum_{s=1}^{S} k_i \left[z_{is} \log P(N = n_i \mid i \in C_s; \lambda_s) + z_{is} \log \pi_s \right], \tag{3}$$

where $\boldsymbol{\Psi} = (\pi_1, \ldots, \pi_S, \lambda_1, \ldots, \lambda_S)$. To obtain ML estimates from the EM algorithm, two steps are required. In the first, the E-step, in the $(r + 1)$-th iteration the conditional expectation of the complete log likelihood function (3) is calculated using some known set of values $\boldsymbol{\Psi}^{(r)}$:

$$Q(\boldsymbol{\Psi} \mid \boldsymbol{\Psi}^{(r)}) = \mathbb{E}_{\boldsymbol{\Psi}^{(k)}} \left[\log L_c(\boldsymbol{\Psi} \mid \mathbf{N}, \mathbf{Z}) \middle| \mathbf{N}, \boldsymbol{\Psi}^{(r)} \right]. \tag{4}$$

In the M-step, the maximum value of the function (4) is found with respect to $\boldsymbol{\Psi}$:

$$\boldsymbol{\Psi}^{(r+1)} = \arg\max_{\boldsymbol{\Psi}} Q\left(\boldsymbol{\Psi} \mid \boldsymbol{\Psi}^{(r)} \right). \tag{5}$$

Following this procedure, from (3)–(5) we obtain the estimates:

$$\hat{\pi}_s^{(r+1)} = K^{-1} \sum_{i=0}^{I} \hat{\tau}_{is}^{(r)} k_i,$$

$$\hat{\lambda}_c^{(r+1)} = \left(K \hat{\pi}_s^{(r+1)} \right)^{-1} \sum_{i=0}^{I} \hat{\tau}_{is}^{(r)} k_i n_i,$$

where the posterior probability that observation i belongs to class C_s is given by:

$$\hat{\tau}_{is}^{(r)} = \frac{\hat{\pi}_s^{(r)} P(N = n_i \mid i \in C_s; \hat{\lambda}_c^{(r)})}{\sum_{c=1}^{S} \hat{\pi}_c^{(r)} P(N = n_i \mid i \in C_s; \hat{\lambda}_c^{(r)})}.$$

The E and M steps are alternated repeatedly until a convergence criterion is met, e.g. no further major improvements in the value of the log likelihood function can be obtained. One important fundamental property of this algorithm is that the sequence of EM iterates will converge to a local maximizer of the log likelihood function. On the other hand, the EM algorithm can be very slow to converge and several extensions have been proposed to speed up convergence [3], [9], [18]. Fortunately, we did not observe this problem for the finite Poisson mixture models considered.

In contrast to the Poisson-gamma mixture model, calculation of the incomplete data information matrix would be algebraically tedious. Hence, numerical approximation will be used, in order to compute the standard errors of ML estimates.

4. Empirical comparison of PFM and PGM

In this section we compare the Poisson-gamma mixture (PGM) model to the finite Poisson mixture (PFM) model using four data sets adopted from the paper [1]. The data was supplied to the authors by the MRCA Information Services from their National Panel of Household Customers. The data set includes purchase frequencies for four products: salty snacks, potato chip brand 1, potato chip brand 2 and potato chip brand 3 (all other potato chip brands). For the last three products, many very small (including zero) purchase counts were observed. Hence, we truncated them at counts of 48, 24 and 25 getting 4313, 4069 and 4182 purchases, respectively. In this way, 1% of purchases were discarded but such an approach is justifiable in the context of a comparative study. As a result, estimation of the parameters is stable. In the instance of salty snacks, the data consist of purchase counts up to 65 and a total of 3852 purchases.

Prior to utilizing the estimation procedure for the finite Poisson mixture, described in the previous section, one needs to define the number of classes*. Unfortunately, the actual number of classes is not known and must be inferred from the data. In addition, the likelihood ratio statistic, allowing the comparison of two models with s and $s+1$ classes, does not asymptotically have a chi-squared distribution (under the null hypothesis, some parameters are on the boundary of the parameter space and hence the regularity conditions fail) and therefore cannot be used [10], [14]. To resolve this problem, several approaches have been proposed, among which the use of information criteria is the most common. Thus we use Akaike's information criterion (AIC), Akaike's consistent information criterion (CAIC) and the Bayesian information criterion (BIC). The EM algorithm is repeated 1000 times (for each data set and for each

*The computations were done using R [12].

number of components) with random initialization and the constraints $\lambda_1 < \lambda_2 < \ldots < \lambda_S$, in order to find the global optimum* and verify identifiability. It is worth noting that mixtures of Poisson distributions are generically identifiable [15], [19]. Thus these constraints force a unique labelling and unequal parameters.

Comparing several models (from 1 to 8 classes) with respect to the minimum values for these criteria, we conclude that salty snacks and the potato chip brands from 1 to 3 should have 7, 6, 4 and 4 classes, respectively. The estimated parameters are statistically significant given the error estimate (see the appendix). Additionally, the goodness-of-fit of the models to the data based on the Pearson chi-squared test are given in Table 1**. Although for the first two products, the p values are too small to accept the hypothesis that the models fit the data well, we decided not to increase the number of classes. The motivation behind such an approach came from additional analysis showing that there was little improvement when we increased the number of classes. It is worth also noting that taking the significance level to be equal to 0.001, one should not reject the hypothesis that the models fit the data well. Evaluating the models of the purchase patterns for the other two products, we find them closely related to the observed purchase frequencies in the data. This is confirmed by the Pearson chi-squared test.

Table 1. Estimates for the finite Poisson mixture and the Poisson-gamma mixture models

Parameter	Salty snacks		Potato chip brand 1		Potato chip brand 2		Potato chip brand 3	
	PFM	PGM	PFM	PGM	PFM	PGM	PFM	PGM
log$L$	–15227	–15540	–12167	–12203	–6921	–6939	–7224	–7244
chi-square	80.0	441.9	66.0	102.4	23.2	53.2	18.9	59.2
d.f.	52	63	37	46	17	22	18	23
#parameters	13	2	11	2	7	2	7	2
p value	0.008	0	0.002	0	0.14	0.0002	0.4	0.0001
#classes	7	–	6	–	4	–	4	–
AIC	30 480	31 085	24 356	24 410	13 855	13 881	14 461	14 492
CAIC	30 575	31 099	24 437	24 425	13 907	13 896	14 512	14 507
BIC	30 562	31 097	24 426	24 423	13 900	13 894	14 505	14 505

In turn, the Poisson-gamma mixture models do not reproduce the data well, as can be seen from the Pearson statistic, although the estimated parameters are statistically significant (see appendix). Based on this, we can safely reject the PGM models for each product in favour of the PFM. In this instance, making an indirect comparison is conclusive.

*Local identification was established from the matrix of the second derivatives of the log-likelihood, which was negative definite.

**The sources for all tables are calculations of the author.

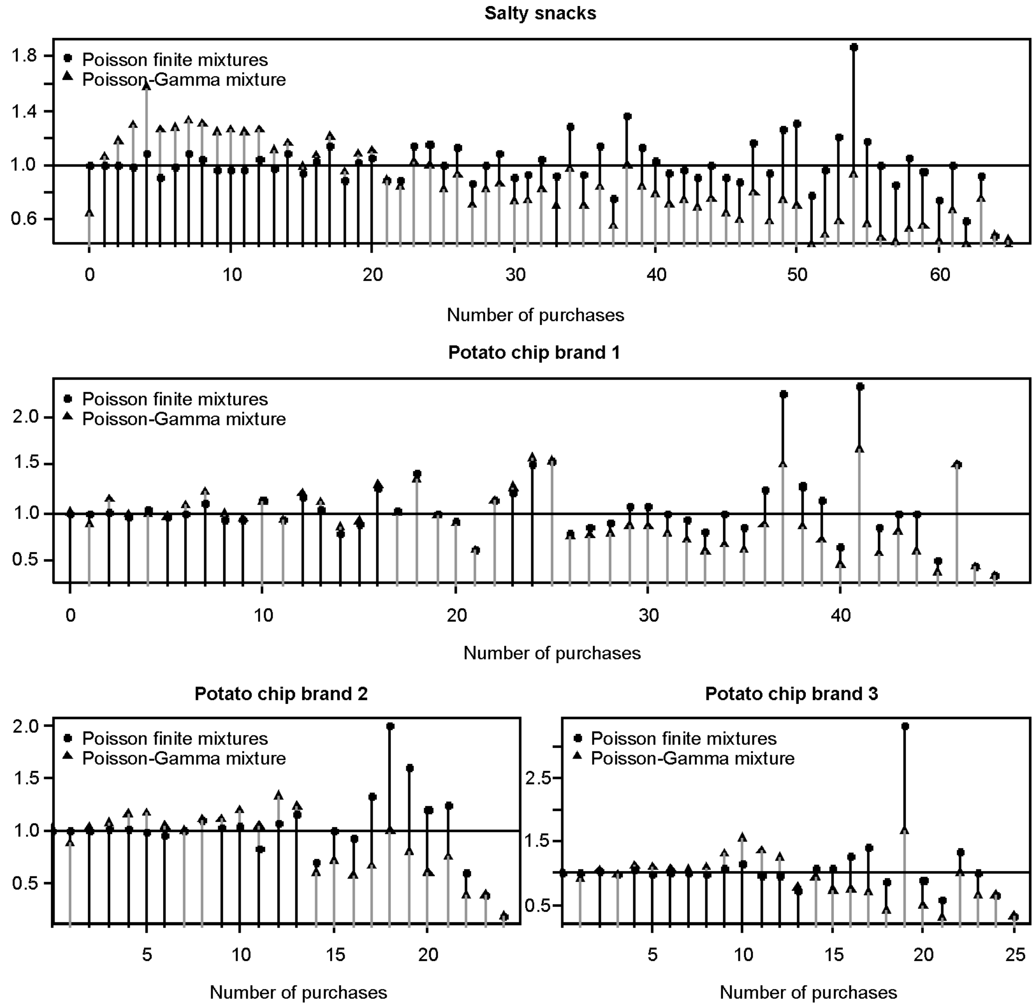

Fig. 1. Expected counts divided by the observed counts with the reference line at 1 indicating a perfect fit

But it is not rare that, according to the Pearson statistic, two different models fit the data well. Therefore, the question remains as to which of them to choose. Since PGM and PFM are not nested, a direct comparison based on a likelihood ratio test is not appropriate. Hence, one can use information criteria. To strengthen our conclusion based on the Pearson statistic, information criteria were estimated (see Table 1). The conclusions given above were confirmed only for the salty snacks data. For the remaining data, AIC favors the PFM models, whereas CAIC and BIC favor the PGM models. This suggests ambiguity but, as a rule, BIC and CAIC tend to select less complex models (with fewer parameters) than Akaike's information criterion does. Since the differences between the values of the information criteria are often minor, it is credible that although one model gives a smaller value of the information criteria, the other model fits the data better (this is confirmed by Table 1). It is often stated that in the first place a researcher should rely on statistical tests and only then on information criteria.

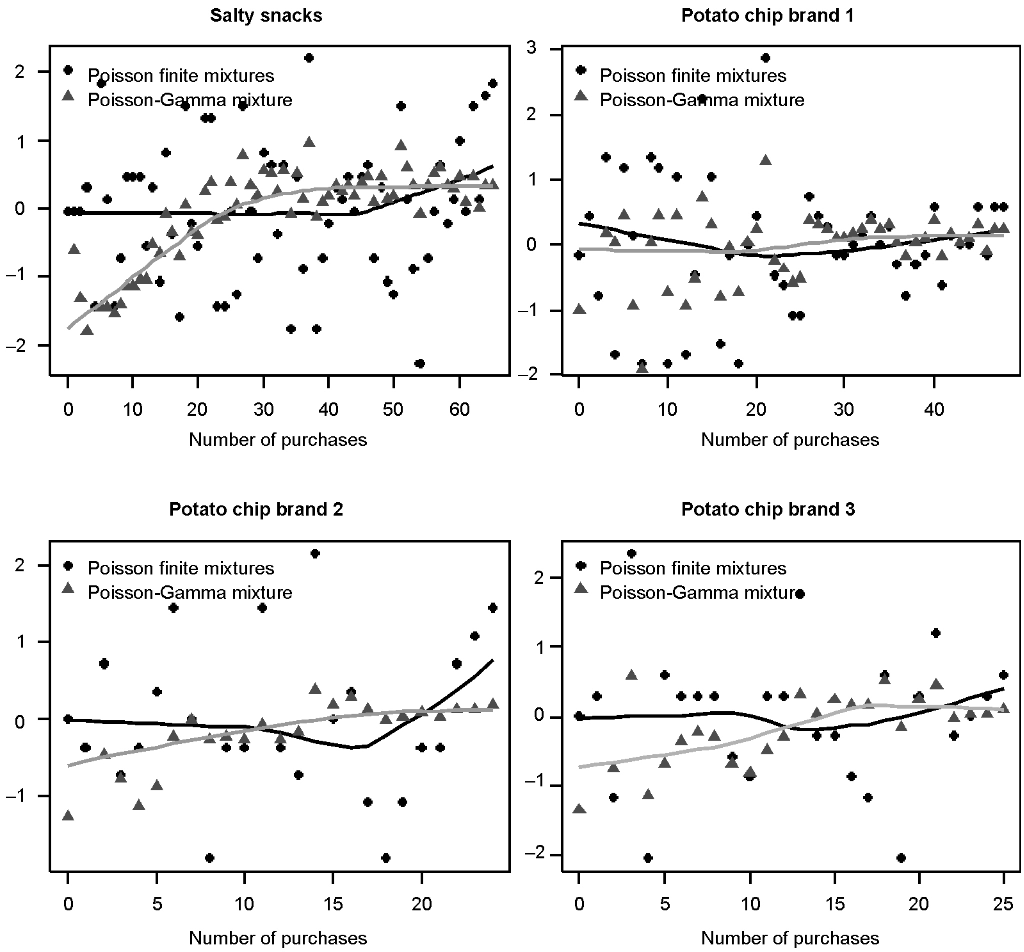

Fig. 2. Residuals and smoothed curves for the PGM and PFM models

To facilitate further comparison, we divided the fitted value by the observed counts for each product. As can be seen in Figure 1, the previous comparison based on the Pearson test confirms the superiority of the finite Poisson mixture models. The largest difference between models concerns the salty snacks data, which exhibit a relatively long right tail. Some patterns and behaviour of the ratio show why the PGM fails to capture heterogeneity. For the lowest number of purchases, apart from zero, the

PGM overestimates these counts, while for above 20 purchases underestimates them. Thus the PGM appears to lack flexibility and, in consequence, an unexplained pattern exists in the residuals. To substantiate this observation, standardized residuals were plotted (see Figure 2). We also fitted a non-parametric smoothed curve through the residuals. The graph of salty snacks for the PGM model clearly shows the relationship between the number of purchases and the residuals, which means that the model is deficient. In contrast, the residuals for the PFM model do not display any particular pattern.

The differences between the models for the second data set (chip brand 1) are not so large in contrast to the salty snacks data. The counts for small numbers of purchases are predicted with almost the same level of accuracy. A disparity becomes noticeable for the numbers of customers who purchased potato chips brand 1 at least 29 times (see Figure 1). The PFM model outperforms its rival and fits worse for only a few numbers of purchases. The residuals for both models do not exhibit any particular pattern (Figure 2). This confirms the previous finding that a relatively long tail might not be captured by the PGM model.

Although the potato chip brand 2 and 3 data do not have such a long tail, Figure 1 highlights the behaviour of the PGM models that was highlighted above. The predicted counts for low numbers of purchases are overestimated, while for high numbers of purchases they are underestimated. This tendency is not as strong as for the salty snacks data, but is distinct, as can be seen in Figure 2. This evidence indicates that besides a limited possibility for capturing a long tail, the PGM models lacks flexibility. Examining the data, one notices that there is a large discrepancy between the frequencies. The first few counts constitute the vast majority of the observations. Almost 85% of customers bought potato chips of brand 2 or 3 no more than 3 times. In contrast, for salty snacks and potato chip brand 1, these figures were 21% and 55%, respectively. This appears to be the main reason for why the PGM models fail.

The PFM models based on the potato chip brand 2 and 3 data outperform their rivals at least in predicting counts of small numbers of purchases (see Figure 1). For larger numbers of purchases, the differences between these two kinds of models become smaller. Moreover, it sometimes happens that the predictive accuracy of the PGM models is better. Unfortunately, this is not very important, because the relative frequencies are very small and thus they do not greatly influence performance. Taking into account the undesirable property mentioned above (over- and underestimated predicted values), PFM models should be preferred.

5. Conclusions

In the paper, a Poisson-gamma mixture model with a finite Poisson mixture model have been compared. Since both models offer alternative ways of capturing heterogeneity in the data and the former, known as the negative binomial model, is very popular in modelling marketing data, we highlight some differences between them. To achieve this goal, we have used four real sets of data on the frequency of purchases.

Despite our empirical findings being conditioned on the data sets used, some understanding of the merits of both models emerges. Based on the first data set, we find that the Poisson-gamma model fails to capture a long tail, as opposed to the finite Poisson mixture. Analyzing the residuals, an undesirable pattern is observed for the Poisson-gamma model: counts of low numbers of purchases are overestimated, whereas those of large numbers are underestimated. Moreover, the smoothed curve clearly shows a trend in the residuals, which indicates that the model does not fit the data well.

The residuals behave similarly, although the trend is not so distinct, when the vast majority of counts are linked to very small numbers of purchases (e.g. the data sets for potato chips brand 2 and 3). This evidence indicates that a Poisson-gamma mixture model lacks flexibility as well.

What emerges from this empirical investigation is that finite Poisson mixture models do not have these shortcomings and under such circumstances should be used instead.

Appendix

Table A1. Maximum-likelihood estimates for the parameters of the finite Poisson mixture models (data: salty snacks and potato chip brand 1)

Class	Salty snacks				Potato chip brand 1			
	$\hat{\lambda}_s$	$\hat{\pi}_s$	$sd(\hat{\lambda}_s)$	$sd(\hat{\pi}_s)$	$\hat{\lambda}_s$	$\hat{\pi}_s$	$sd(\hat{\lambda}_s)$	$sd(\hat{\pi}_s)$
1	0.2	0.09	0.14	0.024	0.4	0.39	0.03	0.015
2	2.0	0.12	0.50	0.015	3.4	0.28	0.21	0.014
3	6.8	0.16	0.65	0.013	9.0	0.17	0.55	0.012
4	14.4	0.18	0.90	0.014	16.8	0.09	0.96	0.011
5	24.4	0.17	1.24	0.015	27.4	0.04	1.63	0.007
6	35.9	0.14	1.22	0.017	38.7	0.03	1.35	0.006
7	52.9	0.14	0.55	0.009	–	–	–	–

Table A2. Maximum-likelihood estimates for the parameters of the finite Poisson mixture models (data: potato chips brand 2 and 3)

Class	Potato chip brand 2				Potato chip brand 3			
	$\hat{\lambda}_s$	$\hat{\pi}_s$	$\mathrm{sd}(\hat{\lambda}_s)$	$\mathrm{sd}(\hat{\pi}_s)$	$\hat{\lambda}_s$	$\hat{\pi}_s$	$\mathrm{sd}(\hat{\lambda}_s)$	$\mathrm{sd}(\hat{\pi}_s)$
1	0.2	0.60	0.03	0.04	0.2	0.59	0.03	0.032
2	1.9	0.26	0.26	0.03	2.0	0.26	0.25	0.022
3	6.2	0.11	0.41	0.01	6.0	0.11	0.43	0.016
4	15.4	0.03	0.55	0.00	15.6	0.04	0.49	0.004

Table A3. Maximum-likelihood estimates for the parameters of the Poisson-gamma mixture models (all data sets)

Snack	$\hat{\alpha}$	$\hat{\beta}$	$sd(\hat{\alpha})$	$sd(\hat{\beta})$
Salty snacks	0.90	0.04	0.013	0.0002
Potato chip brand 1	0.51	0.08	0.009	0.0005
Potato chip brand 2	0.35	0.19	0.008	0.0022
Potato chip brand 3	0.35	0.19	0.008	0.0021

References

[1] BROCKETT P.L., GOLDEN L.L., PANJER H.H., *Flexible purchase frequency modeling*, Journal of Marketing Research, 1996, 33, 94–107.

[2] CAMERON A.C., TRIVEDI P.K., *Regression analysis of count data.* Cambridge University Press, 1998.

[3] DEMPSTER A.P., LAIRD N.M., RUBIN D.B., *Maximum likelihood from incomplete data via EM algorithm*, Journal of the Royal Statistical Society Series B-Methodological, 1977, 39, 1–38.

[4] DILLON W., KUMAR A. (1994), *Advanced methods of marketing research*, [In:] R.P. Bagozzi Ed.), *Latent structure and other mixture models in marketing: An integrative survey and overview*, Oxford, Blackwell, 1994, 295–351.

[5] EHRENBERG A, *My research in marketing: how it happened,* Marketing Research, 2004, 16, 36–41.

[6] GREENE W., *Functional forms for the negative binomial model for count data*, Economics Letters, 2008, 99, 585–590.

[7] GUO J.Q., TRIVEDI P., *Flexible parametric models for long-tailed patent count distributions*, Oxford Bulletin of Economics and Statistics, 2002, 64, 63–82.

[8] JOHNSON N.L., KEMP A.W., KOTZ S., *Univariate discrete distributions*, Hoboken, Wiley, 2005.

[9] MCLACHLAN G.J., KRISHNAN T., *The EM algorithm and extensions*, Wiley, New York, 2008.

[10] MCLACHLAN G.J., PEEL D., *Finite mixture models*, Wiley, New York, 2000.

[11] PAWITAN Y., I*n all likelihood: statistical modelling and inference using likelihood.* Oxford University Press, Oxford, 2001.

[12] *R Development Core Team. R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, 2009. URL http://www.R-project.org.

[13] ROSS S.M., *Introduction to probability models*, Academic Press, Amsterdam, 2007.

[14] SELF S.G., LIANG K.Y., *Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions*, Journal of the American Statistical Association, 1987, 82, 604–610.

[15] TEICHER H., *On the mixture of distributions*, The Annals of Mathematical Statistics, 31, 1960, 55–73.

[16] WEDEL M., DESARBO W.S., BULT J.R., RAMASWAMY V., *Latent class Poisson regression model for heterogeneous count data*, Journal of Applied Econometrics, 8, 1993, 397–411.

[17] WINKELMANN R., *Econometric analysis of count data*, Springer, Berlin, 2008.

[18] WU C.F.J., *On the Convergence properties of the EM Algorithm*, The Annals of Statistics, 1983, 11, 95–103.

[19] YAKOWITZ S.J., SPRAGINS J.D., *On the identifiability of finite mixtures*, The Annals of Mathematical Statistics, 1968, 39, 209–214.