Folia Oeconomica
Acta Universitatis Lodziensis
ISSN 0208-6018 e-ISSN 2353-7663

https://www.czasopisma.uni.lodz.pl/foe/
1(366) 2024
DOI: https://doi.org/10.18778/0208-6018.366.02

Piotr Sulewski https://orcid.org/0000-0002-0788-6567

Institute of Exact and Technical Sciences, Pomeranian University Słupsk, Poland
piotr.sulewski@upsl.edu.pl

Probability Distribution Modelling of Scanner Prices and Relative Prices Using Theoretical Distributions with Two, Three, Four, and Five Parameters

Abstract: This article addresses the problem of proper adjustment of the theoretical probability distribution to the empirical distribution of scanner prices. In the empirical study, we use scanner data from one retail chain in Poland, i.e., monthly data on natural yogurt, yogurt drinks, long grain rice and coffee powder sold in 212 outlets in January and February 2022. Prices and relative prices are modelled using fifty two-, three-, four-, and five-parameter probability distributions with non-negative support. Some of them consist of somewhat known distributions which are called their special cases. The study indirectly involves over a hundred of these distributions. Information criteria such as AIC, BIC, HQIC and p-values of goodness-of-fit tests are used for comparative analysis. This article shows that models such as Frechet, Pareto IV and Log-Logistic could be distinguished as very accurate, which provides a good background for simulation research on price indices or for the construction of the so-called population price indices. The Appendix presents the cumulative distribution function formulas of the models used and the necessary R codes for conducting the research.

Keywords: data modelling, scanner data, price distributions

JEL: C13, C43, E31

1. Introduction

Scanner data are a relatively new and cheap alternative data source in inflation measurement. The volume of scanner data is enormous compared to the datasets obtained as part of the traditional data collection, and they provide detailed information about the products sold at the barcode level. This kind of data are usually obtained with high frequency (monthly, weekly, and in some countries even daily), which enables effective modelling of scanner prices. As a result, having well-matched theoretical probability distributions to empirical price distributions provides a good background for simulation research on price indices or for the construction of the so-called population price indices.

This article addresses the problem of proper adjustment of the theoretical probability distribution to the empirical distribution of scanner prices. A natural question arises about the desirability of this type of consideration.

In order to justify undertaking the research problem, let us note that knowledge of the distribution of prices and the distribution of relative prices permits the construction of the so-called population price indices. It is possible then to generalise the so-called sample elementary indices (the Dutot index (Dutot, 1738), the Carli index (Carli, 1804) or the Jevons index (Jevons, 1865)) to the entire population of products from a given segment by determining the so-called population elementary price indices (Silver, Heravi, 2007; Białek, 2022). With certain technical assumptions about consumption levels (quantity distributions), it is also possible to infer the population Laspeyres price index (Białek, 2015).

Another argument may be the fact that by having accurate probabilistic price models, we are able to effectively construct simulation experiments to study the nature of price indices. For example, knowing the expected values of such distributions and using theorems about the distribution of sums and quotients of random variables, we can formulate expectations for price indices understood as random variables, and then check whether the indices determined on the basis of empirical data are close to these expectations. The above-presented approach was used, for instance, in the papers by Białek and Bobel (2019) or Białek and Beręsewicz (2021) to optimise the choice of a multilateral price index.

The article attempts to model the prices of food products, among others, due to the multitude of their representatives. In the empirical study, we use scanner data from one retail chain in Poland, i.e., monthly data on natural yogurt, yogurt drinks, long grain rice and coffee powder sold in 212 outlets in January and February 2022. More details will be presented in the further parts of the paper. Prices and relative prices are modelled using fifty two-, three-, four, and five-parameter probability distributions with non-negative support. Some of them consist of somewhat known distributions which are called their special cases. The total number of these distributions indirectly participating in the study is over a hundred. Information criteria such as AIC, BIC, HQIC and p-values of Kolmogorov-Smirnov (KS), Anderson-Darling (AD) and Cramer-von Mises (CVM) goodness-of-fit tests (GoFTs) are used for comparative analysis.

The article of Sulewski and Białek (2022) is also devoted to modelling scanner prices and relative scanner prices for the same food products using ten models covering seven, two and one distributions with two, three and four parameters, respectively. The authors selected the GB2 and lognormal models as the most appropriate for modelling scanner prices and relative scanner prices. By clearly expanding the family of distributions for data modelling in this article, we obtain conclusions that are significantly different from those obtained in the shortened version of this article (Sulewski, Białek, 2022). The authors of the aforementioned work described the importance of scanner data, along with their main advantages, but also methodological challenges related to the implementation of these data in inflation measurement.

This article consists of five sections. Section 2 presents fifty probability distributions with two (13), three (20), four (10) and five (7) parameters supported to semi-infinite intervals $\left[u,\infty\right)$ or $\left(u,\infty\right)\left(u\geq 0\right)$. Distribution families are also presented in detail. Section 3 presents numerical measures for comparisons of the quality of data modelling, i.e., the information criteria such as AIC, BIC, HQIC and p-values of the KS, AD and CVM GoFTs. Section 4 describes the main stages of the implemented scanner data processing, it also presents and describes results obtained for the sets of selected models and applied GoFTs. Section 5 is devoted to summary and conclusions. Cumulative distribution function (CDF) formulas of models used and R codes necessary to perform the Monte Carlo simulation are presented in the Appendix.

2. Theoretical probability distribution considered

In our study, theoretical probability distributions used for modelling of scanner prices and relative prices are divided into four groups according to the number of distribution parameters. The II symbol denotes, for example, distributions with two parameters.

Group II includes thirteen distributions: Log-Weibull (LW) (Gumbel, 1958), Nakagami (NA) (Nakagami, 1960), Inverse Normal (IN) (Chhikara, Folks, 1989), Inverse Weibull (IW) (Drapella, 1993), Shifted Gompertz (SGO) (Bemmaor, 1994), Gompertz (GO) (Johnson, Kotz, Balakrishnan, 1995), Inverse Gamma (IG) (Witkovsky, 2001), Log-Logistic (LOGL) (Kleiber, Kotz, 2003) Log-Laplace (LL) (Lindsey, 2004), Flexible Weibull Extension (FEW) (Bebbington, Lai, Zitikis, 2007), Power Lindley (PL) (Ghitany et al., 2013), Inverse Extended Weibull (IEW) (El-Gohary, El-Bassiouny, El-Morshedy, 2015), and Chen (CH) (Nadarajah, Rocha, 2016).

Group III includes twenty distributions: Lognormal (LOG) (Gaddum, 1945), Birnbaum-Saunders (BS) (Birnbaum, Saunders, 1969), Beta Prime (BP) (Johnson, Kotz, Balakrishnan, 1995), Gamma-Gompertz (GGO) (Johnson, Kotz, Balakrishnan, 1995), Log-Gamma (LOGG) (Kotz, Nadarajah, 2000), Marshall-Olkin Extended Weibull (MOEW) (Ghitany, Al-Hussaini, Al-Jarallah, 2005), Frechet (FR) (Castillo et al., 2005), Odd Weibull (OW) (Cooray, 2006), Extended Weibull (EW) (Tieling, Min, 2007), Generalised Inverse Weibull (GIW) (Felipe, Edwin, Gauss, 2009), Sarhan and Zaindin's Modified Weibull (SZMW) (Sarhan, Zaindin, 2009), Weibull Geometric (WG) (Barreto-Souza, de Morais, Cordeiro, 2011), Weibull Poisson (WP) (Lu, Shi, 2012), Kumaraswamy Inverse Weibull (KIW) (Shahbaz, Shahbaz, Butt, 2012), Generalised Gompertz (GeGO) (El-Gohary, Alshamrani, Al-Otaibi, 2013), Gamma Nadarajah–Haghighi (GNH) (Bourguignon et al., 2015), Weighted Generalised Exponential-Exponential (WGEE) (Mahdavi, 2015), Marshall-Olkin Extended Inverse Weibull (MOEIW) (Okasha et al., 2017), Cosine Sine Exponential (CSE) (Chesneau, Bakouch, Hussain, 2018), and Quasi XGamma Poisson (QXP) (Subhradev, Mustafa, Haitham, 2018).

Group IV includes ten distributions: Pareto IV (PIV) (Brazauskas, 2003), Transformed Beta (TB) (Kleiber, Kotz, 2003), Generalised Beta Type 2 (GB2) (McDonald, 1984), Generalised Modified Weibull (GMW) (Carrasco, Ortega, Cordeiro, 2008), Beta Generalised Exponential (BGE) (Barreto-Souza, Santos, Cordeiro, 2010), Exponentiated Generalised Gamma (EGG) (Cordeiro, Ortega, Silva, 2011), Exponentiated Modified Weibull Extension (EMWE) (Sarhan, Apaloo, 2013), Marshall-Olkin Kappa (MOK) (Javed, Nawaz, Irfan, 2019), Generalisation of Generalised Gamma (GGG) (Shanker, Shukla, 2019), and Chen-Pareto (CHP) (Awodutire, 2020).

Group V includes seven distributions: Beta Generalised Gamma (BGG) (Cordeiro et al., 2013), Transmuted Exponentiated Modified Weibull (TEMW) (Eltehiwy, Ashour, 2013), Five Parameter Lindley (FPL) (Al-Babtain et al., 2015), Exponentiated Transmuted Modified Weibull (ETMW) (Pal, Tiensuwan, 2015), Kumuraswamy Exponentiated Linear Exponential (KELE) (Yusuf, Qureshi, 2019), Exponentiated Uniform-Pareto (EUP) (Abdollahi Nanvapisheh, 2019), and Modified Beta Linear Exponential (MBLE) (Bakouch, Saboor, Khan, 2021). In total, we have fifty distributions with at least two and at most five parameters. These distributions are supported to semi-infinite intervals $\left[u,\infty\right)$ or $\left(u,\infty\right)\left(u\geq 0\right)$. The value of u is positive in six cases, i.e., for BS, LOG, FR, PIV, CHP, and EUP.

The PIV, TB, GB2, GMW, BGE, EGG, EMWE, MOK, GGG (see group IV) and BGG, TEMW, FPL, ETMW, KELE, MBLE (see group V) distributions are actually distribution families. These families consist of other, somewhat known, distributions which are referred to as their special cases (see Tables 1–15).

CDFs of distributions from groups II – V are presented in the Appendix (see Tables A1–A4). Substituting the parameter values from Tables 1–15 into CDF formulas in the Apendix, we obtain a very large family of models. Summarising, we use 50 models directly in the empirical study, and over a hundred models indirectly.

Table 1. Sub-models of the PIV distribution with parameters a, b, c, d

No	a	b	c	d	Name
1	a	b	1	d	Pareto type I
2	a	b	1	d	Pareto type II
3	a	b	c	1	Pareto type III
4	0	b	1	d	Lomax

Source: own elaboration based on Jędrzejczak, Pekasiewicz, 2020

Table 2. Sub-models of the TB distribution with parameters a, b, c, d

No	a	b	c	d	Sub-model
1	a	b	1	d	Burr
2	a	1	c	d	Generalised Pareto
3	1	b	c	d	Inverse Burr
4	1	1	c	d	Inverse Pareto
5	1	b	b	d	Inverse Paralogistic
6	1	b	1	d	Loglogistic
7	a	a	1	d	Paralogistic
8	a	1	1	d	Pareto

Source: own elaboration based on R documentation

Table 3. Sub-models of the GB2 distribution with parameters a, b, c, d

No	a	b	c	d	Sub-model
1	a	b	1	d	Singh–Maddala (Burr XII)
2	a	b	c	1	Dagum (Burr III)
3	1	b	c	d	Beta type II
4	1	b	1	1	Fisk (log-logistic)
5	1	b	1	d	Lomax (Pareto type II)
6	a	b	1	1	Fisk
7	1	b	c	1	Inverse Lomax

Source: own elaboration based on Jędrzejczak, Pekasiewicz, 2020

Table 4. Sub-models of the GMW distribution with parameters a, b, c, d

No	a	b	c	d	Sub-model
1	a	1	c	0	Weibull
2	a	1	0	d	Extreme-value
3	a	b	c	0	Exponentiated Weibull
4	a	1	c	d	Modified Weibull

Source: own elaboration based on Carrasco, Ortega, Cordeiro, 2008

Table 5. Sub-models of the BGE distribution with parameters a, b, c, d

No	a	b	c	d	Sub-model
1	1	1	c	d	Generalised Exponential
2	1	1	c	1	Exponential
3	a	b	c	1	Beta Exponential

Source: own elaboration based on Barreto-Souz, Santos, Cordeiro, 2010

Table 6. Sub-models of the EGG distribution ($n \in N$) with parameters a, b, c, d

No	a	b	c	d	Sub-model
1	a	1	1	1	Exponential
2	a	1	1	d	Gamma
3	a	b	1	1	Weibull
4	2	1	1	$n/2$	Chi-square
5	$\sqrt{2}$	2	1	$n/2$	Chi
6	a	2	1	1	Rayleigh
7	$\sigma\sqrt{2}$	2	1	1.5	Maxwell
8	a	b	1	d	Generalised Gamma

Source: own elaboration based on Cordeiro, Ortega, Silva, 2011

Table 7. Sub-models of the EMWE distribution with parameters a, b, c, d

No	a	b	c	d	Sub-model
1	a	1	c	d	Generalised Gompertz
2	∞	1	c	d	Generalised Exponential
3	∞	b	c	d	Exponentiated Weibull
4	1	b	c	1	Chen (Chen, 2000)

Source: own elaboration based on Sarhan, Apaloo, 2013

Table 8. Sub-models of the MOK distribution with parameters a, b, c, d

No	a	b	c	d	Sub-model
1	a	b	c	1	Kappa-III
2	a	b	1	1	Kappa-II

Source: own elaboration based on Javed, Nawaz, Irfan, 2019

Table 9. Sub-models of the GGG distribution ($n \in N$) with parameters a, b, c, d

No	a	b	c	d	Sub-model
1	1	d	1	1	Exponential
2	1	d	b	1	Gamma
3	1	d	1	c	Weibull
4	1	2	$n/2$	2	Chi-square
5	1	$\sqrt{2}$	$n/2$	2	Chi
6	1	$\sigma\sqrt{2}$	1	2	Rayleigh
7	1	$\sigma\sqrt{2}$	1.5	2	Maxwell
8	1	d	b	c	Generalised Gamma

Source: own elaboration based on Stacy, Mihram, 1965

Table 10. Sub-models of the BGG distribution with parameters a, b, c, d, e

No	a	b	c	d	e	Sub-model
1	1	1	c	λ	β	Generalised Gamma
2	1	1	1	λ	β	Gamma
3	1	1	1	0.5	$n/2$	Chi-square
4	1	1	1	λ	1	Exponential
5	1	1	c	λ	1	Weibull
6	1	1	2	λ	1	Rayleigh
7	1	1	2	λ	3/2	Maxwell
8	1	1	2	$1/\sqrt{2}$	1/2	Folded Normal
9	1	1	c	λ	∞	Lognormal
10	a	1	1	λ	β	Exponentiated Gamma
11	a	1	1	0.5	$n/2$	Exponentiated Chi-square
12	a	1	1	λ	1	Exponentiated Exponential
13	a	1	c	λ	1	Exponentiated Weibull
14	a	1	2	λ	1	Exponentiated Rayleigh
15	a	1	2	λ	3/2	Exponentiated Maxwell
16	a	1	2	$1/\sqrt{2}$	1/2	Exponentiated Folded Normal

No	a	b	c	d	e	Sub-model
17	a	1	1c	λ	∞	Exponentiated Lognormal
18	a	b	1	λ	β	Beta Gamma
19	a	b	1	0.5	n/2	Beta Chi-square
20	a	b	1	λ	1	Beta Exponential
21	a	b	c	λ	1	Beta Weibull
22	a	b	2	λ	1	Beta Rayleigh
23	a	b	2	λ	3/2	Beta Maxwell
24	a	b	2	$1/\sqrt{2}$	1/2	Beta Folded Normal
25	a	b	c	λ	∞	Beta Lognormal

Source: own elaboration based on Cordeiro et al., 2013

Table 11. Sub-models of the TEMW distribution with parameters a, b, c, d, e

No	a	b	c	d	e	Sub-model
1	a	b	c	d	0	Exponentiated Modified Weibull
2	1	b	c	d	0	Modified Weibull
3	1	b	c	d	e	Transmuted Modified Weibull
4	a	b	c	0	0	Exponentiated Weibull
5	a	b	c	0	e	Transmuted Exponentiated Weibull
6	1	b	c	0	0	Weibull
7	1	b	c	0	e	Transmuted Weibull
8	a	1	c	0	0	Exponentiated Exponential
9	a	1	c	0	e	Transmuted Exponentiated Exponential
10	1	1	c	0	0	Exponential
11	1	1	c	0	e	Transmuted Exponential
12	a	2	c	d	0	Generalised Linear Failure Rate
13	a	2	c	d	e	Transmuted Generalised Linear Failure Rate
14	1	2	c	d	0	Linear Failure Rate
15	1	2	c	d	e	Transmuted Linear Failure Rate
16	a	2	c	0	0	Generalised Rayleigh
17	a	2	c	0	e	Transmuted Generalised Rayleigh
18	1	2	c	0	0	Rayleigh
19	1	2	c	0	e	Transmuted Rayleigh

Source: own elaboration based on Eltehiwy, Ashour, 2013

Table 12. Sub-models of the FPL distribution $(v \in N)$ with parameters a, b, c, d, e

No	a	b	c	d	e	Sub-model
1	a	b	1	1	0	Gamma
2	a	1	1	1	0	Exponential
3	a	1	2	1	1	Lindley
4	a	v	1	1	0	Erlang
5	a	1	2	d	e	Quasi Lindley
6	a	b	$\alpha+1$	1	e	Generalised Lindley
7	a/e	1	2	1	e	Janardan
8	a	b	c	1	1	New Generalised Lindley
9	0.5	$0.5v$	1	1	0	Chi-square
10	a	1	2	d	1	Two parameter Lindley
11	a	b	c	1	e	Four parameter Lindley type I
12	a	b	c	d	1	Four parameter Lindley type II

Source: own elaboration based on Al-Babtain et al., 2015

Table 13. Sub-models of the ETMW distribution with parameters a, b, c, d, e

No	a	b	c	d	e	Sub-model
1	a	b	c	1	e	Transmuted Modified Weibull
2	0	b	c	1	e	Transmuted Weibull
3	a	1	c	1	e	Transmuted Exponential
4	0	b	c	d	e	Exponentiated Transmuted Weibull
5	a	1	c	d	e	Exponentiated Transmuted Exponential
6	a	b	c	d	0	Exponentiated Modified Weibull
7	0	b	c	d	0	Exponentiated Weibull
8	a	1	c	d	0	Exponentiated Exponential
9	a	b	c	1	0	Modified Weibull
10	a	2	c	1	0	Linear Failure Rate
11	0	b	c	1	0	Weibull
12	0	2	c	1	0	Rayleigh
13	a	1	0	0	e	Exponential

Source: own elaboration based on Pal, Tiensuwan, 2015

Table 14. Sub-models of the KELE distribution with parameters a, b, c, d, e

No	a	b	c	d	e	Sub-model
1	1	1	1	λ	0	Exponential
2	1	1	1	0	θ	Rayleigh
3	1	1	1	λ	θ	Linear Failure Rate
4	a	1	1	0	θ	Exponentiated Rayleigh
5	a	1	1	0	λ	Exponentiated Exponential
6	a	1	1	θ	λ	Exponentiated Linear Failure Rate
7	1	1	α	θ	λ	Exponentiated Linear Exponential
8	a	1	α	θ	λ	Exponentiated Exponentiated Linear Exponential
9	a	b	1	θ	0	Kumuraswamy Rayleigh
10	a	b	1	0	λ	Kumuraswamy Exponential
11	a	b	1	θ	λ	Kumuraswamy Linear Failure Rate

Source: own elaboration based on Yusuf, Qureshi, 2019

Table 15. Sub-models of the MBLE distribution with parameters a, b, c, d, e

No	a	b	c	d	e	Sub-model
1	1	1	1	d	e	Linear Exponential
2	a	1	1	0	e	Generalised Exponential
3	a	1	1	d	0	Generalised Rayleigh
4	a	1	1	d	e	Generalized Linear Failure Rate
5	a	b	1	0	e	Beta Exponential
6	a	b	1	d	0	Beta Rayleigh
7	a	b	1	d	e	Beta Linear Failure Rate

Source: own elaboration based on Bakouch, Saboor, Khan, 2021

3. The goodness-of-fit tests and information criteria used

Let $M(\Theta)$ be the model with the vector of parameters Θ and $f_M(x;\Theta)$ be the probability density function (PDF) of this model. Let $x_1^*, x_2^*, ..., x_n^*$ be a random sample of size n from the $M(\Theta)$. Our target is to estimate the unknown parameters Θ by using the maximum likelihood estimation (MLE) method. The likelihood function is given by (Brandt, 2014).

$$L(\Theta) = \prod_{i=1}^{n} f_M(x_i^*;\Theta), \quad (1)$$

then the log-likelihood function is defined as

$$l(\Theta) = \ln L(\Theta) = \sum_{i=1}^{n} \ln\left[f_M\left(x_i^*;\Theta\right)\right]. \tag{2}$$

Formulas $dl/d\Theta$ have complex forms. In practice, the calculation of these derivatives is not necessary.

The maximum likelihood estimates of parameters Θ were calculated in R software (R Core Team, 2021) using the *fitdistr()* function (package *MASS*). To avoid local maxima of the log-likelihood function, the optimisation routine was run repeatedly each time from different starting values that are widely scattered in the parameter space. The ranges of initial values are different for the distribution parameters and are selected intuitively in such a way that the optimisation procedure works optimally.

The KS, AD and CVM GoFTs were used for model fitting, while the information criteria (IC) such as AIC (Akaike IC), BIC (Bayesian IC) and HQIC (Hannan-Quinn IC) were used for model comparisons. Let us note that (Akaike, 1974; Schwarz, 1978; Hannan, Quinn, 1979):

$$AIC = -2l + 2p, BIC = -2l + pln(n), HQIC = -2l + 2pln(\ln(n)),$$

where l is the log-likelihood function (2), n is the sample size, and p is the number of model parameters.

4. Empirical study

4.1. Description of scanner datasets used

In the following empirical study, we use scanner data from one retail chain in Poland, i.e., monthly data on natural yogurt (subgroup of COICOP 5 group: 011441), yogurt drinks (subgroup of COICOP 5 group: 011441), long grain rice (subgroup of COICOP 5 group: 011111) and coffee powder (subgroup of COICOP 5 group: 012111) sold in 212 outlets in January and February 2022 (52,618 records, which means 42 MB of data). These groups will be designated in our study as Cases 1–4, respectively. We defined a homogeneous product at the most detailed level, i.e., at the EAN bar code level. We detected the following number of different EANs with respect to analysed product groups: 59 (natural yogurt), 106 (yogurt drinks), 28 (long grain rice), and 98 (coffee powder). For each EAN, the monthly price was calculated as the so-called *unit value*, i.e., the monthly product price was determined as the quotient of the total value of sales of a given product divided by the number of units of the product sold. For each analysed Case, the following variants for the price samples were considered: prices from the beginning of the research period (denoted by "B"), prices from the end of the research period (denoted by "E") and the variant with partial price indices (the "I" variant with relative prices, i.e., ratios of February prices to January prices).

Before fitting the probability distributions, the above-mentioned datasets were carefully prepared. A full description of scanner data processing can be found in (Sulewski, Białek, 2022).

The data were from the lowest level of aggregation (GTIN code level) and it can be assumed that the prices of products understood in this way were independent of each other (the prices of a product are determined by the demand for it and not the price of products from the same homogeneous product group).

4.2. Main results

All calculations are performed in R software. PDF (d prefix), CDF (p prefix) and pseudo-random generator (r prefix) are calculated using built-in functions from R packages for 35 models. Details can be found in Table 16, in which models are sorted by parameter number and publication date. User (not built-in) functions with prefixes d, p, and r are used for 15 models of the R codes which are presented in the Appendix.

Table 16. Built-in functions from R packages for models used in the empirical study

No	Model	Package	Function	No	Model	Package	Function
1	LW	RelDists	LW	26	WP	RelDists	WP
2	NA	VGAM	naka	27	KIW	RelDists	KumIW
3	IN	SuppDists	invGauss	28	GeGO	x	GeGO
4	IW	RelDists	IW	29	GNH	x	GNH
5	SGO	extraDistr	sgomp	30	WGEE	RelDists	WGEE
6	GO	x	GO	31	MOEIW	RelDists	MOEIW
7	IG	actuar	invgamma	32	CSE	RelDists	CS2e
8	LOGL	actuar	llogis	33	QXP	RelDists	QXGP
9	LL	LaplacesDemon	llaplace	34	PIV	VGAM	paretoIV
10	FEW	RelDists	FWE	35	TB	actuar	trbeta
11	PL	RelDists	PL	36	GB2	GB2	gb2
12	IEW	x	IEW	37	GMW	RelDists	GMW
13	CH	x	CH	38	BGE	RelDists	BGE
14	BS	extraDistr	fatigue	39	EGG	RelDists	EGG
15	LOG	Fadist	lnorm3	40	EMWE	RelDists	EMWEx
16	BP	extraDistr	betapr	41	MOK	RelDists	MOK

No	Model	Package	Function	No	Model	Package	Function
17	GGO	x	GGO	42	GGG	x	GGG
18	LOGG	VGAM	lgamma	43	CHP	x	CHP
19	MOEW	RelDists	MOEW	44	BGG	x	BGG
20	FR	VGAM	frechet	45	TEMW	x	TEMW
21	OW	RelDists	OW	46	FPL	x	FPL
22	EW	RelDists	ExW	47	ETMW	x	ETMW
23	GIW	RelDists	GIW	48	KELE	x	KELE
24	SZMW	RelDists	SZMW	49	EUP	x	EUP
25	WG	RelDists	WG	50	MBLE	x	MBLE

Source: own elaboration

The empirical study consists of two steps. In the first, the AIC and KS statistic values are calculated for all analysed models. For cases I–IV and samples B, E and I, the best models are selected according to the measures used and marked in bold (see Table 17, MLEs and IC values are presented and p-values of the KS, AD, CVM GoFTs are calculated for the best models from the first step (see Tables 21–24). The best models in Tables 21–24 are in bold.

The p-values for a given model were calculated as follows. Let Θ be the vector of model parameters. Having estimated parameters vector $\hat{\Theta}$ for a given sample of size n, we calculated test statistics $T\left(\hat{\Theta},n\right)$. Next, we generated 10^5 samples of size n for a given model with the estimated parameters vector $\hat{\Theta}$. For each obtained sample s, we calculated the value of $T_i^s\left(\hat{\Theta},n\right)$. Finally, the p-value can be approximated as follows:

$$p \approx \#\left\{i : T_i^s\left(\hat{\Theta},n\right) > T\left(\hat{\Theta},n\right)\right\}10^{-5}.$$

Table 17. AIC and KS statistic values for models (M) and data (D). Case I (natural yogurt)

D	B		E		I		D	B		E		I	
M	AIC	KS	AIC	KS	AIC	KS	M	AIC	KS	AIC	KS	AIC	KS
LW	327.38	0.21	347.66	0.24	– 109.01	0.37	WP	376.02	0.40	374.73	0.42	– 163.33	0.31
NA	294.08	0.10	302.42	0.12	– 183.64	0.20	KIW	296.70	0.10	302.58	0.12	– 209.66	0.15
IN	305.89	0.13	308.19	0.15	– 190.66	0.19	GeGO	296.70	0.10	302.68	0.12	– 199.42	0.18
IW	337.36	0.21	336.86	0.22	– 200.56	0.19	GNH	295.84	0.09	302.07	0.12	– 182.10	0.20
SGO	296.39	0.10	299.06	0.11	– 119.20	0.41	WGEE	301.82	0.11	304.26	0.13	– 58.57	0.43
GO	316.06	0.21	329.73	0.24	– 83.69	0.52	MOEIW	295.37	0.09	296.61	0.08	**– 224.29**	0.12
IG	316.11	0.14	316.88	0.16	– 194.02	0.18	CSE	**291.76**	0.09	297.68	0.09	168.48	0.47
LOGL	292.87	0.09	**294.21**	**0.08**		**0.11**	QXP	374.07	0.42	373.61	0.41	123.50	0.59
LL	332.49	0.25	331.77	0.25	310.27	0.97	PIV	292.53	**0.08**	294.98	**0.08**	**– 233.71**	**0.08**
FEW	301.96	0.13	312.06	0.16	– 124.07	0.35	TB	298.14	0.10	304.45	0.12	– 181.87	0.20
PL	294.06	0.10	302.12	0.13	– 123.48	0.36	GB2	292.78	0.09	**294.89**	0.08	**– 234.89**	**0.09**
IEW	341.35	0.32	341.00	0.33	– 200.32	0.19	GMW	316.05	0.19	330.07	0.23	– 176.31	0.23
CH	310.92	0.19	324.82	0.21	– 80.96	0.44	BGE	297.25	0.10	302.94	0.11	– 204.25	0.15
BS	294.07	0.09	300.35	0.11	– 202.10	0.16	EGG	297.42	0.09	303.34	0.11	– 149.27	0.31
LOG	293.90	0.09	299.80	0.10	– 184.57	0.19	EMWE	297.62	0.09	303.52	0.11	– 200.75	0.17
BP	297.31	0.10	302.13	0.12	– 91.66	0.13	MOK	**291.79**	0.08	**293.28**	0.08	**– 225.29**	**0.10**
GGO	**290.56**	**0.08**	**293.47**	**0.07**	– 95.87	0.50	GGG	297.95	0.10	304.12	0.12	– 180.67	0.20
LOGG	298.62	0.10	311.75	0.12	– 173.05	0.24	CHP	1930.57	0.10	1935.26	0.12	1408.34	**0.12**
MOEW	**290.74**	0.08	294.99	0.09	– 187.69	0.24	BGG	299.35	0.09	305.27	0.10	– 203.33	0.17
FR	298.48	0.10	301.06	0.11	– 201.37	0.18	TEMW	296.76	**0.08**	302.12	0.10	– 211.39	0.14
OW	292.02	**0.08**	**294.88**	0.08	– 218.66	0.13	FPL	293.59	0.09	301.13	**0.08**	– 150.27	0.31
EW	**290.94**	**0.08**	295.55	0.10	– 191.00	0.23	ETMW	296.75	0.11	302.40	0.10	– 206.75	0.14
GIW	339.36	0.21	338.86	0.22	– 198.56	0.19	KELE	298.89	0.09	305.27	0.12	– 179.20	0.23

D	B		E		I		D	B		E		I	
M	**AIC**	**KS**	**AIC**	**KS**	**AIC**	**KS**	**M**	**AIC**	**KS**	**AIC**	**KS**	**AIC**	**KS**
SZMW	375.26	0.36	374.87	0.35	– 120.04	0.41	EUP	343.46	0.21	342.96	0.22	– 194.43	0.19
WG	294.33	0.08	296.07	0.08	124.42	0.64	MBLE	294.40	0.09	297.71	**0.08**	– 203.12	0.21

Source: own elaboration

As shown in Table 17, the models that deserve more in-depth analysis for Case 1 regarding natural yogurt are the following: LOGL, PIV, MOK, GB2, OW, and EW (see Table 21).

Table 18. AIC and KS statistic values for models (M) and data (D). Case II (yogurt drinks)

D	B		E		I		D	B		E		I	
M	**AIC**	**KS**	**AIC**	**KS**	**AIC**	**KS**	**M**	**AIC**	**KS**	**AIC**	**KS**	**AIC**	**KS**
LW	887.995	0.424	927.321	0.432	– 269.057	0.239	WP	627.955	0.228	627.575	0.231	– 301.774	0.229
NA	685.149	0.313	706.516	0.293	– 326.949	0.234	KIW	576.856	0.221	590.119	0.165	– 353.978	0.159
IN	592.104	0.254	607.133	0.205	– 331.328	0.237	GeGO	604.323	0.253	623.778	0.211	– 340.029	0.209
IW	533.524	0.142	**553.812**	0.097	**– 355.914**	0.158	GNH	601.295	0.244	615.749	0.186	– 324.049	0.239
SGO	608.673	0.251	629.533	0.209	– 212.854	0.315	WGEE	603.290	0.250	622.351	0.217	– 233.006	0.302
GO	691.730	0.367	699.970	0.360	– 156.059	0.479	MOEIW	571.773	0.147	593.906	0.130	**– 355.380**	**0.141**
IG	560.936	0.206	576.315	0.157	– 257.330	0.248	CSE	663.361	0.221	683.785	0.200	– 299.654	0.251
LOGL	565.659	0.152	582.962	0.126	– 339.686	0.157	QXP	601.536	0.195	615.071	0.187	2.004	0.535
LL	609.235	0.218	620.057	0.212	90.667	0.440	PIV	**524.559**	**0.096**	**548.625**	**0.052**	**– 364.229**	**0.124**
FEW	670.257	0.383	685.930	0.373	– 280.722	0.258	TB	**527.093**	**0.118**	**549.680**	**0.062**	**– 356.442**	**0.155**
PL	646.503	0.251	662.611	0.267	202.610	0.529	GB2	**526.894**	**0.112**	**549.891**	**0.069**	**– 366.414**	**0.122**
IEW	555.094	0.213	575.982	0.174	**– 355.868**	0.158	GMW	586.816	0.231	603.280	0.182	– 335.447	0.214
CH	721.476	0.348	737.335	0.349	– 232.354	0.275	BGE	607.020	0.237	611.512	0.227	– 332.364	0.217

D	B		E		I		D	B		E		I	
M	AIC	KS	AIC	KS	AIC	KS	M	AIC	KS	AIC	KS	AIC	KS
BS	539.678	0.178	567.246	0.145	– 349.637	0.177	EGG	571.753	0.220	586.570	0.164	– 89.824	0.370
LOG	**528.919**	0.133	555.597	0.102	– 325.622	0.233	EMWE	578.049	0.228	594.458	0.179	– 342.121	0.200
BP	564.088	0.207	579.596	0.159	– 329.470	0.227	MOK	**529.399**	**0.097**	**551.455**	**0.070**	– 336.115	0.158
GGO	693.466	0.369	591.929	0.189	– 175.089	0.430	GGG	598.133	0.222	612.840	0.231	– 322.756	0.241
LOGG	758.599	0.298	788.579	0.287	– 320.226	0.24	CHP	**526.347**	**0.111**	**551.393**	**0.076**	– 353.724	**0.140**
MOEW	621.333	0.278	621.502	0.200	– 307.103	0.187	BGG	544.085	0.188	564.543	0.139	– 349.871	0.167
FR	**524.183**	**0.107**	**548.798**	**0.073**	– 353.944	0.159	TEMW	592.100	0.216	608.241	0.174	– 334.427	0.207
OW	564.912	0.148	582.308	0.122	**– 354.968**	**0.151**	FPL	577.538	0.224	607.744	0.169	– 252.072	**0.135**
EW	620.808	0.224	622.432	0.228	– 311.236	0.185	ETMW	569.040	0.200	590.448	0.162	– 245.530	0.156
GIW	535.521	0.143	555.748	0.101	– 353.914	0.158	KELE	636.450	0.234	650.934	0.242	– 341.303	0.200
SZMW	694.903	0.388	702.086	0.366	– 278.736	0.256	EUP	530.688	0.130	555.398	**0.092**	– 349.566	0.158
WG	568.093	0.167	585.033	0.128	– 279.565	0.256	MBLE	613.702	0.265	585.948	0.154	– 335.840	0.211

Source: own elaboration

As shown in Table 18, the models that deserve more in-depth analysis regarding yogurt drinks are the following: PIV, GB2, CHP, TB, FR, and MOK (see Table 22).

Table 19. AIC and KS statistic values for models (M) and data (D). Case III (long grain rice)

D	B		E		I		D	B		E		I	
M	AIC	KS	AIC	KS	AIC	KS	M	AIC	KS	AIC	KS	AIC	KS
LW	190.063	0.193	182.282	0.179	– 90.462	0.127	WP	181.821	0.183	178.011	0.166	– 90.467	0.090
NA	174.781	0.161	**170.679**	0.128	**– 92.445**	**0.085**	KIW	173.288	0.105	170.913	0.107	**– 90.787**	**0.089**
IN	**171.124**	0.103	**168.443**	0.101	**– 92.191**	0.092	GeGO	173.908	0.113	170.974	0.110	– 88.861	0.107

D	B		E		I		D	B		E		I	
M	AIC	KS	AIC	KS	AIC	KS	M	AIC	KS	AIC	KS	AIC	KS
IW	**172.153**	**0.091**	**170.591**	0.105	– 85.018	0.138	GNH	173.956	0.109	170.996	0.113	**– 90.709**	**0.085**
SGO	**172.431**	0.116	**169.444**	0.113	– 58.633	0.332	WGEE	173.856	0.121	170.983	0.104	– 59.082	0.303
GO	182.754	0.177	177.189	0.152	– 41.394	0.441	MOEIW	173.285	**0.097**	171.220	**0.090**	– 89.856	0.093
IG	**171.184**	0.105	**168.988**	**0.091**	– 66.3445	0.233	CSE	180.759	0.143	176.818	0.113	– 81.498	0.176
LOGL	173.320	0.112	170.836	0.107	**– 91.999**	**0.076**	QXP	176.472	0.139	172.954	0.107	16.535	0.439
LL	186.735	0.176	186.735	0.184	25.160	0.443	PIV	**171.991**	**0.093**	**169.959**	**0.087**	– 88.841	0.093
FEW	173.600	0.163	**169.430**	0.121	**– 91.186**	0.120	TB	175.274	0.105	172.854	0.101	– 88.782	**0.089**
PL	174.348	0.147	173.824	0.216	**– 90.675**	0.122	GB2	175.281	0.105	172.853	0.101	– 88.779	**0.089**
IEW	**170.352**	**0.094**	**168.597**	0.103	– 85.032	0.138	GMW	175.411	0.109	172.835	0.106	– 88.796	**0.089**
CH	181.47	0.165	176.054	0.144	– 86.1139	0.16	BGE	174.056	**0.095**	172.318	0.100	– 88.213	**0.085**
BS	**171.846**	**0.097**	**169.829**	**0.088**	– 90.522	0.086	EGG	175.204	0.105	172.913	**0.093**	– 88.792	0.090
LOG	**172.493**	0.099	**170.422**	**0.090**	**– 90.524**	0.086	EMWE	175.237	0.107	172.736	0.101	– 86.637	0.110
BP	**173.217**	0.105	170.937	**0.096**	90.194	0.086	MOK	176.780	0.110	173.392	0.096	– 88.002	**0.077**
GGO	**173.120**	**0.091**	179.081	0.148	– 44.856	0.403	GGG	175.754	0.108	173.023	**0.094**	– 88.673	0.090
LOGG	181.628	0.186	176.239	0.139	**– 90.615**	0.086	CHP	174.287	0.099	172.213	0.890	– 88.781	0.088
MOEW	178.282	0.118	175.005	0.106	– 88.845	0.084	BGG	174.553	0.571	171.982	0.571	– 86.783	0.089
FR	173.522	0.106	171.254	0.098	– 84.361	0.132	TEMW	177.122	0.106	174.872	0.101	– 86.773	0.084
OW	175.235	0.112	172.321	**0.085**	– 89.975	**0.076**	FPL	175.329	**0.072**	172.206	**0.087**	– 62.620	0.246
EW	177.729	0.115	174.305	0.105	– 89.745	0.080	ETMW	176.927	0.106	174.558	0.096	– 86.739	0.086
GIW	174.153	**0.091**	172.587	0.107	– 83.018	0.138	KELE	175.919	0.116	173.636	0.101	– 86.240	0.082
SZMW	202.716	0.352	201.316	0.356	– 89.168	0.119	EUP	178.170	**0.090**	176.626	0.108	– 78.834	0.137
WG	174.631	0.113	171.597	0.106	**– 91.052**	0.080	MBLE	187.204	0.264	184.333	0.256	– 86.775	0.088

Source: own elaboration

As shown in Table 19, the models that deserve more in-depth analysis regarding long grain rice are the following: PIV, BS, LOG, IN, IEW, and NA (see Table 23).

Table 20. AIC and KS statistic values for models (M) and data (D). Case IV (coffee powder)

D	B		E		I		D	B		E		I	
M	**AIC**	**KS**	**AIC**	**KS**	**AIC**	**KS**	**M**	**AIC**	**KS**	**AIC**	**KS**	**AIC**	**KS**
LW	1150.447	0.337	1215.83	0.371	– 201.615	0.268	WP	1004.671	0.804	1022.726	0.808	– 263.732	0.739
NA	1016.879	0.232	1081.54	0.253	– 252.433	0.164	KIW	964.861	0.098	982.029	0.095	– 246.752	0.193
IN	971.116	0.131	991.299	0.130	– 241.859	0.179	GeGO	1123.797	0.399	1166.953	0.434	– 258.129	0.173
IW	999.655	0.156	1019.23	0.176	– 127.830	0.339	GNH	965.131	0.097	982.669	0.092	– 215.692	0.268
SGO	965.175	**0.070**	984.287	**0.071**	– 160.456	0.347	WGEE	966.026	**0.075**	985.262	**0.068**	– 164.503	0.309
GO	1008.888	0.200	1028.07	0.201	– 132.921	0.474	MOEIW	**962.777**	0.083	**976.604**	**0.070**	– 234.574	0.261
IG	982.690	0.159	1003.12	0.158	– 216.442	0.263	CSE	983.083	0.086	1003.344	0.087	– 265.375	0.233
LOGL	**963.385**	0.083	**974.588**	0.071	**– 302.195**	**0.079**	QXP	1017.229	0.239	1030.828	0.229	12.243	0.502
LL	979.515	0.130	990.352	0.139	91.525	0.429	PIV	**963.308**	**0.075**	**977.072**	**0.063**	**– 301.275**	**0.080**
FEW	1001.764	0.190	1035.69	0.262	– 218.624	0.246	TB	**963.290**	**0.073**	**976.721**	**0.059**	**– 307.250**	**0.076**
PL	970.475	0.082	991.238	0.084	– 223.368	0.236	GB2	**963.290**	**0.073**	**976.721**	**0.059**	**– 307.226**	**0.076**
IEW	1013.389	0.261	1032.88	0.271	– 123.473	0.331	GMW	967.719	0.095	985.675	0.088	– 256.079	0.165
CH	1014.457	0.161	1044.18	0.197	– 153.063	0.357	BGE	967.798	0.086	1018.895	0.166	– 254.870	0.156
BS	965.983	0.085	984.990	0.078	– 253.367	0.162	EGG	966.874	0.095	984.313	0.092	– 256.141	0.167
LOG	**960.781**	0.087	979.875	0.080	– 253.916	0.163	EMWE	967.488	**0.077**	985.416	0.098	– 252.307	0.177
BP	**963.505**	0.093	979.816	0.089	– 244.378	0.174	MOK	964.402	**0.079**	**977.485**	**0.062**	**– 315.653**	**0.062**
GGO	976.138	0.155	997.966	0.136	– 123.116	0.457	GGG	967.009	0.101	984.537	0.099	– 253.064	0.171
LOGG	1044.360	0.148	1102.62	0.193	– 255.956	0.161	CHP	967.492	0.104	984.731	0.100	– 232.811	0.232
MOEW	1072.337	0.327	978.725	**0.064**	**– 291.046**	**0.126**	BGG	968.606	0.093	985.902	0.090	– 254.384	0.157
FR	**962.2882**	0.086	**977.205**	0.076	– 163.634	0.310	TEMW	966.062	0.086	981.911	0.080	**– 282.119**	**0.117**

D	B		E		I		D	B		E		I	
M	**AIC**	**KS**	**AIC**	**KS**	**AIC**	**KS**	**M**	**AIC**	**KS**	**AIC**	**KS**	**AIC**	**KS**
OW	**962.992**	0.084	**976.839**	0.072	**– 299.916**	**0.079**	FPL	**959.396**	**0.078**	**971.702**	0.071	– 242.660	0.184
EW	965.113	0.081	979.633	**0.068**	**– 298.091**	**0.104**	ETMW	965.982	**0.078**	981.806	**0.067**	– 281.631	**0.129**
GIW	1001.655	0.156	1021.22	0.175	– 125.830	0.339	KELE	973.317	0.113	990.468	0.096	– 249.565	0.163
SZMW	1017.112	0.240	1030.78	0.229	– 217.865	0.247	EUP	1006.759	0.158	1026.273	0.177	– 120.456	0.340
WG	1017.730	0.227	**976.585**	0.074	– 215.301	0.288	MBLE	975.907	**0.077**	981.683	0.083	– 270.899	0.131

Source: own elaboration

As shown in Table 20, the models that deserve more in-depth analysis regarding coffee powder are the following: TB, GB2, PIV, MOK, LOGL, and OW (see Table 24).

Table 21. MLEs, information criteria and p-values of GoFTs for models (M) and data (D). Case I (natural yogurt)

M	D	MLEs	Information criteria			p-value		
			AIC	BIC	HQIC	KS	AD	CVM
LOGL	B	$\hat{a}=5.052,\ \hat{b}=7.949$	292.87	**297.03**	294.49	0.76	0.63	0.77
	E	$\hat{a}=4.999,\ \hat{b}=7.812$	294.21	**298.37**	**295.84**	0.84	0.61	0.79
	I	$\hat{a}=57.007\ \hat{b}=0.985$	– 228.32	– 224.17	– 226.70	0.49	0.29	0.45
PIV	B	$\hat{a}=-10.917,\hat{b}=18.986$ $\hat{c}=0.079,\hat{d}=1.002$	292.53	300.84	295.77	0.80	**0.84**	0.80
	E	$\hat{a}=-12.183,\hat{b}=19.361$ $\hat{c}=0.065,\ \hat{d}=0.693$	294.98	303.29	298.22	**0.87**	0.78	0.83
	I	$\hat{a}=0.631,\ \hat{b}=0.340$ $\hat{c}=0.034,\ \hat{d}=0.499$	– 233.71	– 225.40	– 230.46	**0.80**	**0.59**	**0.78**
MOK	B	$\hat{a}=4.305,\hat{b}=4.595$ $\hat{c}=1.382,\hat{d}=29.155$	291.79	300.10	295.03	0.78	0.79	0.71
	E	$\hat{a}=4.433,\hat{b}=4.348$ $\hat{c}=1.322,\hat{d}=34.069$	**293.28**	301.59	296.53	0.81	**0.83**	0.82
	I	$\hat{a}=6.101,\hat{b}=0.896$ $\hat{c}=9.786,\hat{d}=290.598$	– 225.29	– 216.98	– 222.04	0.58	0.35	0.52
GB2	B	$\hat{a}=9.743,\ \hat{b}=9.520,$ $\hat{c}=0.321,\ \hat{d}=0.696$	292.78	301.09	296.02	0.70	0.81	0.80
	E	$\hat{a}=11.463,\ \hat{b}=8.717,$ $\hat{c}=0.284,\ \hat{d}=0.455$	294.89	303.20	298.13	0.79	0.75	0.79
	I	$\hat{a}=120.365,\ \hat{b}=0.975$ $\hat{c}=0.553,\ \hat{d}=0.294$	**– 234.89**	**– 226.58**	**– 231.65**	0.68	0.42	0.50

M	D	MLEs	Information criteria			p-value		
			AIC	BIC	HQIC	KS	AD	CVM
OW	B	$\hat{a}=0.096, \hat{b}=1.396$ $\hat{c}=2.615$	292.02	298.26	294.46	**0.85**	0.76	**0.83**
	E	$\hat{a}=0.081, \hat{b}=0.800$ $\hat{c}=4.517$	294.88	301.11	297.31	0.82	0.65	0.80
	I	$\hat{a}=0.920, \hat{b}=3.737$ $\hat{c}=9.994$	– 218.66	– 212.43	– 216.23	0.29	0.14	0.24
EW	B	$\hat{a}=1.071, \hat{b}=0.840$ $\hat{c}=495.414$	**290.94**	297.17	**293.37**	0.81	**0.83**	**0.83**
	E	$\hat{a}=1.206, \hat{b}=0.792$ $\hat{c}=492.169$	295.55	301.78	297.98	0.59	0.62	0.76
	I	$\hat{a}=8.178, \hat{b}=4.640$ $\hat{c}=1994.359$	– 191.00	– 184.77	– 188.57	0.00	0.00	0.00

Source: own elaboration

As shown in Table 21, the best models for modelling natural yogurt distributions are one, two and three models with two, three and four parameters, respectively. The best model for data B in terms of the AIC, HQIC and p-values of the AD and CVM tests is EW. The best model in terms of p-values of the KS and CVM tests is OW. The best model for data E in terms of the BIC and HQIC is LOGL, while in terms of the AIC and p-value of the AD test is MOK. The best model for data I in terms of the AIC, BIC, HQIC and p-values of KS, AD, and CVM tests are GB2 and PIV, respectively.

Table 22. MLEs, information criteria and p-values of GoFTs for models (M) and data (D). Case II (yogurt drinks)

M	D	MLEs	Information criteria			p-value		
			AIC	BIC	HQIC	KS	AD	CVM
PIV	B	$\hat{a}=4.322,\ \hat{b}=2.133,$ $\hat{c}=0.407,\ \hat{d}=0.614$	524.56	535.21	528.88	**0.26**	**0.61**	**0.51**
	E	$\hat{a}=3.477,\ \hat{b}=2.542,$ $\hat{c}=0.267,\ \hat{d}=0.408$	**548.62**	559.28	552.94	**0.92**	**0.90**	**0.92**
	I	$\hat{a}=0.25,\ \hat{b}=0.726,$ $\hat{c}=0.015,\ \hat{d}=0.26$	– 364.22	– 353.57	– 359.90	0.06	0.24	0.23

M	D	MLEs	Information criteria			p-value		
			AIC	BIC	HQIC	KS	AD	CVM
GB2	B	$\hat{a}=8.49$, $\hat{b}=2.843$, $\hat{c}=188.76$, $\hat{d}=0.285$	527.09	537.75	531.41	0.10	0.45	0.36
	E	$\hat{a}=12.589$, $\hat{b}=5.423$, $\hat{c}=1.01$, $\hat{d}=0.18$	549.78	560.43	554.10	0.64	0.82	0.79
	I	$\hat{a}=184.262$, $\hat{b}=0.985$, $\hat{c}=0.354$, $\hat{d}=0.122$	**– 367.48**	**– 356.83**	**– 363.17**	0.06	**0.71**	0.24
CHP	B	$\hat{a}=4.47$, $\hat{b}=1.656$, $\hat{c}=1.941$, $\hat{d}=2.277$	526.35	537.00	530.67	0.14	0.45	0.38
	E	$\hat{a}=3.926$, $\hat{b}=1.569$, $\hat{c}=2.004$, $\hat{d}=3.312$	551.39	562.05	555.71	0.55	0.64	0.69
	I	$\hat{a}=0.895$, $\hat{b}=-4.541$, $\hat{c}=30.47$, $\hat{d}=2.156$	– 354.72	– 344.06	– 350.40	0.01	0.06	0.08
TB	B	$\hat{a}=0.284,$, $\hat{b}=8.503$, $\hat{c}=109.768$, $\hat{d}=3.034$	527.09	537.75	531.41	0.10	0.44	0.35
	E	$\hat{a}=0.082$, $\hat{b}=26.327$, $\hat{c}=0.418$, $\hat{d}=5.605$	549.60	560.26	553.92	0.83	0.86	0.84
	I	$\hat{a}=0.641$, $\hat{b}=40.654$, $\hat{c}=3.312$, $\hat{d}=0.956$	– 357.13	– 346.48	– 352.81	**0.24**	0.36	**0.35**
FR	B	$\hat{a}=1.847$, $\hat{b}=3.328$, $\hat{c}=3.201$	**524.18**	**532.17**	**527.42**	0.16	0.51	0.41
	E	$\hat{a}=1.971$, $\hat{b}=2.669$, $\hat{c}=3.904$	548.80	**556.79**	**552.04**	0.59	0.70	0.72
	I	$\hat{a}=35.327$, $\hat{b}=-0.321$, $\hat{c}=1.314$	– 353.94	– 345.95	– 350.71	0.01	0.07	0.07
MOK	B	$\hat{a}=0.023$, $\hat{b}=10.45$, $\hat{c}=98.23$, $\hat{d}=0.126$	529.32	539.97	533.64	0.22	0.41	0.45
	E	$\hat{a}=0.02$, $\hat{b}=11.009$, $\hat{c}=99.854$, $\hat{d}=0.121$	551.38	562.03	555.70	0.65	0.67	0.82
	I	$\hat{a}=1.005$, $\hat{b}=0.877$, $\hat{c}=38.544$, $\hat{d}=220.978$	– 335.67	– 325.02	– 331.35	0.01	0.01	0.02

Source: own elaboration

As shown in Table 22, the best models for modelling yogurt drinks distributions are one and five models with three and four parameters, respectively. The best model for data B in terms of the AIC, BIC, HQIC and p-values of KS, AD, and CVM tests are FR and PIV, respectively. The best model for data E in terms of the AIC and p-values of the KS, AD, and CVM tests is PIV, while in terms of the BIC and HQIC is FR. The best model for data I in terms of AIC, BIC, HQIC and p-values of the AD test is GB2, while in terms of the KS and CVM tests is TB.

Table 23. MLEs, information criteria and p-values of GoFTs for models (M) and data (D). Case III (long grain rice)

M	D	MLEs	Information criteria			p-value		
			AIC	BIC	HQIC	KS	AD	CVM
PIV	B	$\hat{a}=5.345,\ \hat{b}=197.298,\ \hat{c}=0.835,\ \hat{d}=51.865$	171.92	177.25	173.55	0.95	0.98	**0.98**
	E	$\hat{a}=5.159,\ \hat{b}=275.144,\ \hat{c}=0.776,\ \hat{d}=106.497$	169.75	175.08	171.38	**0.98**	**0.97**	**0.98**
	I	$\hat{a}=0.807,\ \hat{b}=0.515,\ \hat{c}=0.214,\ \hat{d}=99.658$	– 88.85	– 83.53	– 87.22	0.96	**0.99**	**0.99**
BS	B	$\hat{a}=0.687,\ \hat{b}=7.101,\ \hat{c}=3.549$	171.85	175.84	173.07	0.93	0.98	0.97
	E	$\hat{a}=0.6,\ \hat{b}=7.773,\ \hat{c}=2.853$	169.83	173.83	171.05	0.97	0.97	0.97
	I	$\hat{a}=0.002,\ \hat{b}=17.976,\ \hat{c}=-16.993$	– 90.53	– 86.53	– 89.30	0.92	0.75	0.91
LOG	B	$\hat{a}=0.653,\ \hat{b}=1.981,\ \hat{c}=3.484$	172.49	176.49	173.71	0.92	0.97	0.97
	E	$\hat{a}=0.561,\ \hat{b}=2.096,\ \hat{c}=2.596$	170.42	174.42	171.64	0.96	0.96	0.97
	I	$\hat{a}=0.017,\ \hat{b}=0.954,\ \hat{c}=-1.613$	– 90.43	– 86.43	– 89.21	0.97	0.99	0.98
IN	B	$\hat{a}=12.338,\ \hat{b}=62.029$	171.12	173.79	171.94	0.90	0.96	0.92
	E	$\hat{a}=12.034,\ \hat{b}=64.754$	**168.44**	**171.11**	**169.26**	0.91	0.96	0.96
	I	$\hat{a}=0.983,\ \hat{b}=502.182$	– 92.19	– 89.53	– 91.38	0.97	0.99	0.98

M	D	MLEs	Information criteria			p-value		
			AIC	BIC	HQIC	KS	AD	CVM
IEW	B	$\hat{a}=11.33,\ \hat{b}=0.136$	**170.35**	**173.02**	**171.17**	**0.95**	**0.98**	0.96
	E	$\hat{a}=11.222,\ \hat{b}=0.14$	168.60	171.26	169.41	0.90	0.94	0.92
	I	$\hat{a}=10.207,\ \hat{b}=11.068$	– 85.03	– 82.37	– 84.22	0.74	0.55	0.58
NA	B	$\hat{a}=1.503,\ \hat{b}=182.574$	174.78	177.45	175.60	0.42	0.68	0.58
	E	$\hat{a}=0.649,\ \hat{b}=170.138$	170.68	173.34	171.49	0.70	0.76	0.71
	I	$\hat{a}=129.464,\ \hat{b}=0.969$	**– 92.44**	**– 89.78**	**– 91.63**	**0.98**	**0.99**	**0.99**

Source: own elaboration

As shown in Table 23, the best models for modelling long grain rice distributions are three, two and one models with two, three and four parameters, respectively. The best model for data B in terms of the AIC, BIC, HQIC and p-values of the KS and AD tests is EIW. The best model for data E in terms of the AIC, BIC, and HQIC is IN, while in terms of p-values of the KS, AD, and CVM tests is PIV. The best model for data I in terms of all measures used is NA.

Table 24. MLEs, information criteria and p-values of GoFTs for models (M) and data (D). Case IV (coffee powder)

M	D	MLEs	Information criteria			p-value		
			AIC	BIC	HQIC	KS	AD	CVM
TB	B	$\hat{a}=1.487,\ \hat{b}=2.794,$ $\hat{c}=0.902,\ \hat{d}=70.618$	963.29	973.63	967.47	0.64	**0.78**	0.76
	E	$\hat{a}=0.907,\ \hat{b}=3.798,$ $\hat{c}=0.592,\ \hat{d}=69.914$	976.72	987.06	980.90	0.86	0.83	0.86
	I	$\hat{a}=0.263,\ \hat{b}=130.989,$ $\hat{c}=0.203,\ \hat{d}=1.061$	– 307.25	– 296.91	– 303.06	**0.98**	**0.69**	**0.87**
GB2	B	$\hat{a}=2.795,\ \hat{b}=70.602,$ $\hat{c}=0.901,\ \hat{d}=1.486$	963.29	973.63	967.47	0.63	**0.78**	0.76
	E	$\hat{a}=3.801,\ \hat{b}=69.871,$ $\hat{c}=0.592,\ \hat{d}=0.906$	976.72	987.06	980.90	0.86	0.83	0.86
	I	$\hat{a}=144.308,\ \hat{b}=1.062,$ $\hat{c}=0.181,\ \hat{d}=0.24$	– 307.81	– 297.47	– 303.63	0.61	0.35	0.52

M	D	MLEs	Information criteria			p-value		
			AIC	BIC	HQIC	KS	AD	CVM
PIV	B	$\hat{a}=0.003,\ \hat{b}=71.159,\ \hat{c}=0.383,\ \hat{d}=1.646$	963.31	973.65	967.49	0.62	**0.78**	**0.77**
	E	$\hat{a}=-4.357,\ \hat{b}=70.057,\ \hat{c}=0.333,\ \hat{d}=1.28$	977.07	987.41	981.25	0.81	**0.84**	0.87
	I	$\hat{a}=-1.587,\ \hat{b}=2.634,\ \hat{c}=0.009,\ \hat{d}=0.837$	– 301.28	– 290.94	– 297.10	0.64	0.39	0.59
MOK	B	$\hat{a}=1.61,\ \hat{b}=25.656,\ \hat{c}=1.962,\ \hat{d}=11.433$	964.17	974.51	968.35	0.61	0.73	0.76
	E	$\hat{a}=2.076,\ \hat{b}=26.879,\ \hat{c}=1.56,\ \hat{d}=12.766$	976.93	987.27	981.11	**0.87**	0.83	**0.88**
	I	$\hat{a}=6.937,\ \hat{b}=0.953,\ \hat{c}=6.634,\ \hat{d}=98.59$	**– 316.56**	**– 306.22**	**– 312.37**	0.88	0.54	0.69
LOGL	B	$\hat{a}=2.993,\ \hat{b}=0.018$	**960.78**	**965.95**	**962.87**	0.68	0.66	0.73
	E	$\hat{a}=2.959,\ \hat{b}=0.017,$	**974.59**	**979.76**	**976.68**	0.66	0.68	0.81
	I	$\hat{a}=39.701,\ \hat{b}=0.95$	– 302.20	– 297.03	– 300.10	0.56	0.28	0.46
OW	B	$\hat{a}=0.008,\ \hat{b}=0.459,\ \hat{c}=4.698$	961.81	969.56	964.95	**0.79**	0.72	0.72
	E	$\hat{a}=0.006,\ \hat{b}=0.33,\ \hat{c}=6.477$	975.72	983.48	978.86	0.21	0.41	0.39
	I	$\hat{a}=0.797,\ \hat{b}=2.094,\ \hat{c}=13.743$	– 301.74	– 293.98	– 298.60	0.54	0.28	0.47

Source: own elaboration

As shown in Table 24, the best models for modelling coffee powder distributions are one, one and four models with two, three and four parameters, respectively. The best model for data B in terms of the AIC, BIC, and HQIC is LOGL, while in terms of p-values of the AD and CVM tests is PIV. The best model for data E in terms of the AIC, BIC, and HQIC is LOGL, while in terms of p-values of the KS and CVM tests is MOK. The best model for data I in terms of the AIC, BIC, and HQIC is MOK, while in terms of p-values of the KS, AD, and CVM tests is TB.

5. Summary and conclusions

We used fifty distributions with at least two and at most five parameters supported to semi-infinite intervals to model scanner prices and relative scanner prices of natural yogurt, yogurt drinks, long grain rice, and coffee powder. Some of these distributions form a family consisting of other more or less known distributions which are referred to as their special cases (see Tables 1–15).

For the ranking of selected models, we used the information criteria and p-values of the GoFTs. Interestingly, the ranking of models according to the AIC criterion is not always the same as according to the BIC and HQIC criteria. Moreover, the ranking of models based on the p-values of the KS test does not always align with the rankings based on the p-values of the AD and CVM tests. Please note that the ranking in terms of the information criteria differs from the analogical ranking based on p-values.

This article shows that the greater the number of model parameters, the more special cases a given model has (see Tables 1–15). One might expect that as the number of model parameters increases, the model will fit the data better. Among best models for modelling of distributions of the analysed food products are five, six and thirteen models with two, three and four parameters, respectively. This is probably due to the specificity of the data, and estimation problems in the five-parameter model. Undoubtedly, the dominant group are four-parameter models.

The best models for price modelling from the beginning of the research period (B) are:

– natural yogurt – Extended Weibull (AIC, HQIC, AD, and CVM p-values) and Odd Weibull (KS, CVM),
– yogurt drinks – Frechet (AIC, BIC, and HQIC) and Pareto IV (KS, AD, and CVM),
– long grain rice – Inverse Extended Weibull (all analysed measures),
– coffee powder – Log-Logistic (AIC, BIC, and HQIC) and Pareto IV (AD, CVM).

The best models for price modelling at the end of the research period (E) are:

– natural yogurt – Log-Logistic (BIC, HQIC) and Marshall-Olkin Kappa (AIC, AD),
– yogurt drinks – Pareto IV (AIC, KS, AD, and CVM) and Frechet (BIC, HQIC),
– long grain rice – Inverse Normal (AIC, BIC, and HQIC) and Pareto IV (KS, AD, and CVM),
– coffee powder – Log-Logistic (AIC, BIC, and HQIC) and Marshall-Olkin Kappa (KS, CVM).

The best models for the variant with partial price indices (I) (i.e., ratios of E to B) are:

– natural yogurt – GB2 (AIC, BIC, and HQIC) and Pareto IV (KS, AD, and CVM),
– yogurt drinks – GB2 (AIC, BIC, HQIC, and AD) and Transformed Beta (KS, CVM),
– long grain rice – Nakagami (all analysed measures),
– coffee powder – Marshall-Olkin Kappa (AIC, BIC, and HQIC) and Transformed Beta (KS, AD, and CVM).

In summary, for the data modelling of analysed food products from the beginning and at the end of research periods, the best models such as Frechet, Pareto IV and Log-Logistic are repeated. Pareto IV , also popular in income modelling, is the only model that stands out in data B, E, and I.

Several of the remaining models also seem to be of good quality in price modelling, with the final selection of the model probably depending on the product segment and the definition of a homogeneous product (the lower the aggregation level, the greater the price fluctuations we observe).

The aim of the article was achieved because thanks to a significant extension of the family of distributions for data modelling, we obtain conclusions that are significantly different from those obtained in the shortened version of this article (Sulewski, Białek, 2022).

Appendix

Abbreviations used:

– $Ö[x,a,b]$ – the CDF of $N(a,b)$; $Ã(x), B(a,b), sgn(x)$ – the gamma, beta and sign functions, respectively;
– $Ã_L[a,x]$, $Ã_U[a,x]$ – the lower and upper incomplete gamma function, respectively;
– $B_L(x,a,b), B_R(x,a,b), B_U(x,a,b)$ – the lower, regularised and upper incomplete beta, function, respectively;
– a,b,c,d,e – model parameters.

Table A1. CDFs of models used for modelling scanner prices and relative prices. Group II

$F_{LW}(x;a,b) = 1 - exp\left[-\exp\left(\frac{x-a}{b}\right)\right](x>0; a\in R, b>0)$
$F_{NA}(x;a,b) = \frac{Ã_L[b, bx^2/a]}{Ã(b)}(x \geq 0; b \geq 0.5, a>0)$
$F_{IN}(x;Öa,b) =/ \left[\sqrt{b/\ a,\theta,1\ x}\exp\right]2+ /\ (Ö b\ \ a) / \left[-\sqrt{b/\ a,\theta,1\ x}\ \ 0;\right](x>0\ a\ b>\ )$
$F_{IW}(x;a,b) = \exp(-ax^{-b})(x>0; a,b>0)$
$F_{SGO}(x;a,b) = \left[1-e^{-ax}\right] * \exp\left[-be^{-ax}\right](x \geq 0; a,b \geq 0)$

$F_{GO}(x;a,b)=1-\exp\left\{-a\left[e^{bx}-1\right]\right\}(x\geq 0;a,b>0)$
$F_{IG}(x;a,b)=\frac{\tilde{A}_U(b/x,a)}{\tilde{A}(a)}(x>0;a,b\geq 0)$
$F_{LOGL}(x;a,b)=\frac{x^a}{b^a+x^a}(x\geq 0;a,b\geq 0)$
$F_{LL}(x;a,b)=\frac{1}{2}\left\{1+sgn\left[\ln(x)-a\right]\left[1-\exp\left(\frac{-\lvert\ln(x)-a\rvert}{b}\right)\right]\right\}(x>0;a,b>0)$
$F_{FWE}(x;a,b)=1-\exp\left[-\exp(ax-b/x)\right](x>0;a,b>0)$
$F_{PL}(x;a,b)=1-\left(1+\frac{bx^a}{b+1}\right)\exp\left(-bx^a\right)(x>0;a,b>0)$
$F_{IEW}(x;a,b)=\exp\left[-\exp\left(\frac{a}{x}-bx\right)\right](x>0;a,b>0)$
$F_{CH}(x;a,b)=1-\exp\left[b\left(1-\exp\left(x^a\right)\right)\right](x>0;a,b>0)$

Source: own elaboration

Table A2. CDFs of models used for modelling scanner prices and relative prices. Group III

$F_{BS}(x;a,b,c)=\Phi\left[\frac{1}{a}\left(\sqrt{\frac{x-c}{b}}-\sqrt{\frac{b}{x-c}}\right),0,1\right](x>c;c\in R;a,b>0)$
$F_{LOG}(x;a,b,c)=\Phi\left[\ln(x-c),b,a\right](x>c;a>0,b\geq 0)$
$F_{BP}(x;a,b,c)=B_R\left(\frac{x}{x+c},a,b\right)(x>0;a,b,c>0)$
$F_{GGO}(x;a,b,c)=1-\frac{c^b}{\left(c-1+e^{ax}\right)^b}(x\geq 0;a,b,c>0)$
$F_{LOGG}(x;a,b,c)=\frac{\Gamma_L\left[c,\exp((x-a)/b)\right]}{\Gamma(c)}(x>0;a\in R;b,c>0)$

$F_{MOEW}(x;a,b,c)=1-\frac{a\exp\left[-(a/x)^{b}\right]}{1-(1-a)\exp\left[-(a/x)^{b}\right]}(x>0;a,b,c>0)$
$F_{FR}(x;a,b,c)=\exp\left[-\left(\frac{x-b}{c}\right)^{-a}\right](x>b;a,c>0;b\in R)$
$F_{OW}(x;a,b,c)=1-\left\{1+\left[\exp\left((ax)^{b}\right)-1\right]^{c}\right\}^{-1}(x>0;a,bc>0)$
$F_{EW}(x;a,b,c)=1-\frac{1-\exp\left(-ax^{b}\right)}{1-(1-c)\exp\left(-ax^{b}\right)}(x>0;a,b,c>0)$
$F_{GIW}(x;a,b,c)=\exp\left[-c\left(\frac{a}{x}\right)^{b}\right](x>0;a,b,c>0)$
$F_{SZMW}(x;a,b,c)=1-\exp\left(-ax-bx^{c}\right)(x>0;a,b,c>0)$
$F_{WG}(x;a,b,c)=\frac{1-\exp\left[-(ax)^{b}\right]}{1-c\exp\left[-(ax)^{b}\right]}(x>0;a,b>0,0<c<1)$
$F_{WP}(x;a,b,c)=\frac{\exp\left[c\exp\left(-ba^{x}\right)\right]-\exp(c)}{1-\exp(c)}(x>0;a,b,c>0)$
$F_{KIW}(x;a,b,c)=1-\left[1-\exp\left(-bx^{-a}\right)\right]^{c}(x>0;a,b,c>0)$
$F_{GeGO}(x;a,b,c)=\left[1-\exp\left(-\frac{a}{b}\left(e^{bx}-1\right)\right)\right]^{c}(x>0;a,b,c>0)$
$F_{GNH}(x;a,b,c)=\frac{\Gamma_{U}\left[a,(1+cx)^{b}-1\right]}{\Gamma(a)}(x>0;a,b,c>0)$
$F_{WGEE}(x;a,b,c)=\frac{\exp\left(-ax^{-b}\right)}{c-(c-1)\exp\left(-ax^{-b}\right)}(x>0;a,b,c>0)$
$F_{CSE}(x;a,b,c)=\frac{b\cos\left(0.5\pi e^{-x/c}\right)}{a\sin\left(0.5\pi e^{-x/c}\right)+b\cos\left(0.5\pi e^{-x/c}\right)}(x>0;a,b,c>0)$
$F_{QXP}(x;a,b,c)=\frac{e^{c}-\exp\left[ce^{-bx}\frac{1+a+bx+0.5b^{2}x^{2}}{a+1}\right]}{e^{c}-1}(x>0;a,b,c>0)$

Source: own elaboration

Table A3. CDFs of models used for modelling scanner prices and relative prices. Group IV

$F_{PIV}(x;a,b,c,d)=1-\left[1+\left(\frac{x-a}{b}\right)^{\frac{1}{c}}\right]^{-d}\quad(x\geq a;a\in R;b,c,d>0)$
$F_{TB}(x;a,b,c,d)=\int_0^x\frac{b(y/d)^{bc}}{B(a,c)y\left[1+(y/d)^b\right]^{a+c}}dy(x>0;a,b,c,d>0)$
$F_{GB2}(x;a,b,c,d)=\int_0^x\frac{a}{bB(c,d)}\left(\frac{y}{b}\right)^{ac-1}\left[1+\left(\frac{y}{b}\right)^a\right]^{-(c+d)}dy(x\geq 0;a,b,c,d>0)$
$F_{GMW}(x;a,b,c,d)=\left\{1-\exp\left[-ax^c*\exp(d*x)\right]\right\}^b\quad(x>0;a>0,b,c,d\geq 0)$
$F_{BGE}(x;a,b,c,d)=\int_0^x\frac{cd\exp(-cx)(1-\exp(-cx))^{ad-1}}{B(a,b)\left[1-(1-\exp(-cx))^d\right]^{1-b}}dy\quad(x>0;a,b,c,d>0)$
$F_{EGG}(x;a,b,c,d)=\left\{\frac{\Gamma_L\left[d,(x/a)^b\right]}{\Gamma(d)}\right\}^c\quad(x>0;a,b,c,d>0)$
$F_{EMWE}(x;a,b,c,d)=\left\{1-\exp\left(ac\left(1-\exp\left(x^b/a^b\right)\right)\right)\right\}^d\quad(x\geq 0;a,b,c,d>0)$.
$F_{MOK}(x;a,b,c,d)=\left\{d*\left[\frac{a}{x^{ac}/b^{ac}}+1\right]^{\frac{1}{a}}+1-d\right\}^{-1}\quad(x>0;a,b,c,d>0)$
$F_{GGG}(x;a,b,c,d)=1-\frac{\Gamma_u\left[b,a(dx)^c\right]}{\Gamma(b)}(x\geq 0;a,b,c,d>0)$
$F_{CHP}(x;a,b,c,d)=\frac{1-\exp\left\{b\left[1-\exp\left(1-a^c/x^c\right)^d\right]\right\}}{1-\exp(b-b\exp(1))}(x\geq a;a,b,c,d>0)$

Source: own elaboration

Table A4. CDFs of models used for modelling scanner prices and relative prices. Group V

$F_{BGG}(x;a,b,c,d,e)=\frac{1}{B(a,b)}\int_0^{\frac{\Gamma(e,(xd)^c)}{\Gamma(e)}}\omega^{a-1}(1-\omega)^{b-1}d\omega(x>0;a,b,c,d,e>0)$
$F_{TEMW}(x;a,b,c,d,e)=(1-U)^a+\left[1+e-e(1-U)^a\right]$ $U=\exp\left[-\left(dx+cx^b\right)\right],\left(x>0;a,b,c,d>0,\lvert e\rvert\le 1\right)$
$F_{FPL}(x;a,b,c,d,e)=\frac{\mathrm{ad}\Gamma_L(b,ax)+e\Gamma_L(c,ax)}{ad+e}(x>0;d,e\ge 0;a,b,c>0)$
$F_{ETMW}(x;a,b,c,d,e)=\left[\left(1-\exp\left(-ax-cx^b\right)\right)\left(1+e\exp\left(-ax-cx^b\right)\right)\right]^d$ $\left(x>0;a,b,c,d>0,\lvert e\rvert\le 1\right)$
$F_{KELE}(x;a,b,c,d,e)=1-\left\{1-\left[1-\exp\left(-\left(\frac{e}{2}x^2+dx\right)^c\right)\right]^a\right\}^b(x\ge 0;a,b,c>0;e,d\ge 0)$
$F_{EUP}(x;a,b,c,d,e)=\left[1-\frac{bc^d}{x^d(b-a)}\right]^e\left(x>c\left(\frac{b}{b-a}\right)^{\frac{1}{d}};a,b\in R(a\le b);c,d,e>0\right)$
$F_{MBLE}(x;a,b,c,d,e)=\frac{B_L\left[1+\left(-c\exp\left(0.5dx^2+ex\right)\right)^{-1};a,b\right]}{B(a,b)}(x\ge 0;a,b,d,e\ge 0;c>0)$

Source: own elaboration

R codes of user functions for the PDF of the models used

```
dGO = function(x,a,b) return(a*b*exp(a+b*x-a*exp(b*x)))
dIEW = function(x,a,b) return(exp(-exp(a/x-b*x))*(exp(a/x-b*x)*(a/x^2+b)))
dCH = function(x,a,b) return(exp(a*(1-exp(x^b)))*(a*(exp(x^b)*(x^(b-1)*b))))
dGGO = function(x,a,b,c) return(a*b*c^b*exp(a*x)/(c-1+exp(a*x))^(b+1))
dGeGO = function(x,a,b,c) return(a*c*exp(b*x)*exp(-a*(exp(b*x)-1)/b)*(1-exp(-
  a*(exp(b*x)-1)/b))^(c-1))
dGNH = function(x,a,b,c) return(b*c*(1+c*x)^(b-1)*((1+c*x)^(b)-1)^(a-1)*exp(-
  ((1+c*x)^(b)-1))/gamma(a))
dGGG = function(x,a,b,c,d) return(c*(a*d^c)^b*x^(b*c-1)*exp(-a*(x*d)^c)/
  gamma(b))
dCHP = function(x,a,b,c,d) {return(1/(1-exp(b*(1-exp(1))))*(exp(b*(1-exp((1-(a/
  x)^c)^d)))*
  (b*(exp((1-(a/x)^c)^d)*((1-(a/x)^c)^(d-1)*(d*((a/x)^(c-1)*(c*(a/x^2))))))))    }
dBGG = function(x,a,b,c,d,e){
ly=max(length(x),length(a),length(b),length(e),length(d));
  x=rep(x,length=ly);b=rep(b,length=ly)
a=rep(a,length=ly);e=rep(e,length=ly); d=rep(d,length=ly);dg=c*d^(c*e)*x^(c*e-
  1)*exp(-(d*x)^c)/gamma(e)
vF=pgamma((d*x)^c,shape=e,log=TRUE);
  vS=pgamma((d*x)^c,shape=e,log=TRUE,lower.tail=FALSE)#1-F
logfy=(a-1)*vF+(b-1)*vS+log(dg)-lbeta(a,b); return(logfy)  }
dTEMW = function(x,a,b,c,d,e) { return((1-exp(-(d*x+c*x^b)))^(a-1)*(a*(exp(-
  (d*x+c*x^b))*
  (d+c*(x^(b-1)*b))))*(1+e-e*(1-exp(-(d*x+c*x^b)))^a)-((1-exp(-(d*x+c*x^b)))^a)*
  (e*((1-exp(-(d*x+c*x^b)))^(a-1)*(a*(exp(-(d*x+c*x^b))*(d+c*(x^(b-1)*b)))))))    }
dFPL = function (x a,b,c,d,e) {
return((d*(a*x)^(b-1)/gamma(b)+e*(a*x)^(c-1)/a/gamma(c))*exp(-a*x)*a^2/(e+a*d))   }
dETMW = function(x,a,b,c,d,e) { return(((1-exp(-a*x-c*x^b))*(1+e*exp(-a*x-
  c*x^b)))^(d-1)*
(d*(exp(-a*x-c*x^b)*(a+c*(x^(b-1)*b))*(1+e*exp(-a*x-c*x^b))-(1-exp(-a*x-c*x^b))*
(e*(exp(-a*x-c*x^b)*(a+c*(x^(b-1)*b)))))))    }
dKELE = function(x,a,b,c,d,e) { return((1-(1-exp(-(0.5*e*x^2+d*x))^c)^a)^(b-
  1)*(b*((1-exp(-(0.5*e*x^2+ d*x))^c)^(a-1)*(a*(exp(-(0.5*e*x^2+d*x))^(c-1)*(c*(exp(
  (0.5*e*x^2+d*x))*(0.5*e*(2*x)+d))))))) }
dEUP = function(x,a,b,c,d,e)  return((1-b*(c/x)^d/(b-a))^(e-1)*(e*(b*((c/x)^(d-
  1)*(d*(c/x^2)))/(b-a))))
```

```
dMBLE = function(x,a,b,c,d,e) { return(c^a*(d*x+e)*exp(-b*d/2*x^2-b*e*x)*(1-
  exp(-0.5*d*x^2-x*e))^(a-1)
*(1-(1-c)*(1-exp(-0.5*d*x^2-x*e)))^(-a-b)/beta(a,b))   }
```

R codes of user functions for the CDF of the models used

```
pGO = function(x,a,b) return(1-exp(-a*(exp(b*x)-1)))
pIEW = function (x,a,b) return(exp(-exp(a/x-b*x)))
pCH = function (x,a,b)  return(1-exp(a*(1-exp(x^b))))
pGGO = function(x,a,b,c) return(1-c^b/(c-1+exp(a*x))^b)
pGeGO = function(x,a,b,c) return((1-exp(-a/b*(exp(b*x)-1)))^c)
pGNH = function(x,a,b,c) return(1-incgam((1+c*x)^b-1,a)/gamma(a))
pGGG = function(x,a,b,c,d) return(1-incgam(a*(d*x)^c,b)/gamma(b))
pCHP = function(x,a,b,c,d) return(1/(1-exp(b*(1-exp(1))))*(1-exp(b*(1-exp((1-(a/
  x)^c)^d)))))
pBGG = function(q,a,b,c,d,e) {
  lq=max(length(q),length(a),length(b),length(e),length(d))
q=rep(q,length=lq); b=rep(b,length=lq); a=rep(a,length=lq);
  e=rep(e,length=lq); d=rep(d,length=lq)
vF=pgamma((d*q)^c,shape=e); cdf=pbeta(vF,a,b); return(cdf)   }
pTEMW = function (x,a,b,c,d,e) return(((1-exp(-(d*x+c*x^b)))^a)*(1+e-e*(1-exp(-
  (d*x+c*x^b)))^a))
pFPL = function (x,a,b,c,d,e) return(1-(a*d*incgam(a*x,b)/
  gamma(b)+e*incgam(a*x,c)/gamma(c))/(e+a*d))
pETMW= function (x,a,b,c,d,e) return(((1-exp(-a*x-c*x^b))*(1+e*exp(-a*x-
  c*x^b)))^d)
pKELE = function (x,a,b,c,d,e) return(1-(1-(1-exp(-(0.5*e*x^2+d*x))^c)^a)^b)
pEUP = function (x,a,b,c,d,e) return((1-b*(c/x)^d/(b-a))^e)
pMBLE=function (x,a,b,c,d,e) return(Ibeta(1+(-c*exp(0.5*d*x^2+e*x)+c-1)^(-1),
  a, b, lower=TRUE)/beta(a,b))
```

R codes of user functions for the pseudo-random generator of the models used

```
rGO = function(n,a,b) {x=numeric(n);  for (i in 1:n)  x[i]=(1/b)*log(-(1/
  a)*log(1-runif(1,0,1))+1);  return(x) }
rIEW = function(n,a,b) {x=numeric(n);  for (i in 1:n) { aa=log(-
  log(runif(1,0,1)));
x[i]=0.5/b*(-aa+sqrt(aa^2+4*a*b))};  return(x)  }
rCH = function(n,a,b) {x=numeric(n);  for (i in 1:n) x[i]=(log(1-1/a*log(1-
  runif(1,0,1))))^(1/b);  return(x) }
```

```
rGGO = function(n,a,b,c) {x=numeric(n);  for (i in 1:n) x[i]=log(c*(1-
  runif(1,0,1))^(-1/b)+1-c)/a;  return(x) }
rGeGO = function(n,a,b,c) {x=numeric(n); for (i in 1:n) x[i]=1/b*log(1-b/
  a*log(1-runif(1,0,1)^(1/c))); return(x)}
rGNH = function(n,a,b,c) {x=numeric(n);  for (i in 1:n)
  x[i]=((1+rgamma(1,a))^(1/b)-1)/c;  return(x) }
INV _ GGG = function (p,a,b,c,d) {x0=0
  while (pGGG(x0,a,b,c,d)<p) {x0=x0+1}; x0=x0-1
  while (pGGG(x0,a,b,c,d)<p) {x0=x0+0.1}; x0=x0-0.1
  while (pGGG(x0,a,b,c,d)<p) {x0=x0+0.001}; return (x0)  }
rGGG=function (n, a, b, c, d) {rows _ =runif(n,0,1); num=c()
 for (i in 1:n) num=c(num, INV _ GGG(rows _ [i], a, b, c, d)); return (num) }
rCHP=function(n,a,b,c,d) { W=1/(1-exp(b*(1-exp(1))));  x=numeric(n)
for (i in 1:n) x[i]=(1-(log(1-1/b*log(1-runif(1,0,1)/W)))^(1/d))^(1/c);
  return(a/x)  }
INV _ BGG = function (p,a,b,c,d,e) { x0=0
  while (pBGG(x0,a,b,c,d,e)<p) {x0=x0+1}; x0=x0-1
  while (pBGG(x0,a,b,c,d,e)<p) {x0=x0+0.1}; x0=x0-0.1
  while (pBGG(x0,a,b,c,d,e)<p) {x0=x0+0.001};  return (x0) }
rBGG = function (n, a, b, c, d, e) { rows _ =runif(n,0,1);  num=c()
  for (i in 1:n) num=c(num, INV _ BGG(rows _ [i], a, b, c, d, e));  return
  (num) }
INV _ TEMW = function (p,a,b,c,d,e) {x0=0
  while (pTEMW(x0,a,b,c,d,e)<p) {x0=x0+1};  x0=x0-1
  while (pTEMW(x0,a,b,c,d,e)<p) {x0=x0+0.1};  x0=x0-0.1
  while (pTEMW(x0,a,b,c,d,e)<p) {x0=x0+0.001};  return (x0) }
rTEMW = function (n, a, b, c, d, e) { rows _ =runif(n,0,1);  num=c()
  for (i in 1:n) num=c(num, INV _ TEMW(rows _ [i], a, b, c, d, e));  return
  (num)  }
rFPL = function(n,a,b,c,d,e) {x=numeric(n);
for (i in 1:n) { U=runif(1,0,1); Y1=rgamma(1,b,1/a); Y2=rgamma(1,c,1/a)
    x[i]=ifelse(U>a*d/(e+a*d),Y2,Y1) };  return(x) }
INV _ ETMW = function (p,a,b,c,d,e) {x0=0
  while (pETMW(x0,a,b,c,d,e)<p) {x0=x0+1}; x0=x0-1
  while (pETMW(x0,a,b,c,d,e)<p) {x0=x0+0.1}; x0=x0-0.1
  while (pETMW(x0,a,b,c,d,e)<p) {x0=x0+0.001};  return (x0)  }
rETMW = function (n,a,b,c,d,e) { rows _ =runif(n,0,1);  num=c()
  for (i in 1:n) num=c(num, INV _ ETMW(rows _ [i], a, b, c, d, e));  return
  (num) }
```

```
INV _ KELE = function (p,a,b,c,d,e) {x0=0
  while (pKELE(x0,a,b,c,d,e)<p) {x0=x0+1}; x0=x0-1
  while (pKELE(x0,a,b,c,d,e)<p) {x0=x0+0.1}; x0=x0-0.1
  while (pKELE(x0,a,b,c,d,e)<p) {x0=x0+0.001};  return (x0)  }
rKELE = function (n,a,b,c,d,e) { rows _ =runif(n,0,1);  num=c()
  for (i in 1:n) num=c(num, INV _ KELE(rows _ [i], a, b, c, d, e)); return
   (num) }
INV _ EUP = function (p,a,b,c,d,e) {x0=c*(b/(b-a))^(1/d)
  while (pEUP(x0,a,b,c,d,e)<p) {x0=x0+1}; x0=x0-1
  while (pEUP(x0,a,b,c,d,e)<p) {x0=x0+0.1};  x0=x0-0.1
  while (pEUP(x0,a,b,c,d,e)<p) {x0=x0+0.001};  return (x0)  }
rEUP = function (n,a,b,c,d,e) {rows _ =runif(n,0,1);  num=c()
  for (i in 1:n) num=c(num, INV _ EUP(rows _ [i], a, b, c, d, e));  return
   (num) }
INV _ MBLE = function (p,a,b,c,d,e) { x0=0
  while (pMBLE(x0,a,b,c,d,e)<p) {x0=x0+1}; x0=x0-1
  while (pMBLE(x0,a,b,c,d,e)<p) {x0=x0+0.1};  x0=x0-0.1
  while (pMBLE(x0,a,b,c,d,e)<p) {x0=x0+0.001};  return (x0)  }
rMBLE = function (n, a, b, c, d, e) {rows _ =runif(n,0,1);  num=c()
  for (i in 1:n) num=c(num, INV _ MBLE(rows _ [i], a, b, c, d, e));  return
   (num)  }
```

References

Abdollahi Nanvapisheh A. (2019), *A New Five-Parameter Distribution: Properties and Applications*, "International Journal of Mathematical Modelling and Computations", vol. 9(3), pp. 201–212.

Akaike H. (1974), *A new look at the statistical model identification*, "IEEE Transactions on Automatic Control", vol. 19(6), pp. 716–723.

Al-Babtain A., Eid A.M., Ahmed A.H.N., Merovci F. (2015), *The five parameter Lindley distribution*, "Pakistan Journal of Statistics", vol. 31(4).

Awodutire P. (2020), *Chen Pareto Distribution: Properties and Application*, "Pakistan Journal of Statistics and Operation Research", vol. 16(4), pp. 812–826.

Bakouch H.S., Saboor A., Khan M.N. (2021), *Modified beta linear exponential distribution with hydrologic applications*, "Annals of Data Science", no. 8, pp. 131–157.

Barreto-Souza W., Morais A.L. de, Cordeiro G.M. (2011), *The Weibull-geometric distribution*, "Journal of Statistical Computation and Simulation", vol. 81(5), pp. 645–657.

Barreto-Souza W., Santos A.H., Cordeiro G.M. (2010), *The beta generalized exponential distribution*, "Journal of Statistical Computation and Simulation", vol. 80(2), pp. 159–172.

Bebbington M., Lai C.D., Zitikis R. (2007), *A flexible Weibull extension*, "Reliability Engineering and System Safety", vol. 92(6), pp. 719–726.

Bemmaor A.C. (1994), *Modeling the diffusion of new durable goods: Word-of-mouth effect versus consumer heterogeneity*, [in:] G. Laurent, G.L. Lilien, B. Pras (eds.), *Research Traditions in Marketing*, Kluwer, Boston, pp. 201–229.

Białek J. (2015), *Construction of confidence intervals for the Laspeyres price index*, "Journal of Statistical Computation and Simulation", vol. 85(14), pp. 2962–2973.

Białek J. (2022), *Elementary price indices under the GBM price model*, "Communications in Statistics – Theory and Methods", vol. 51(5), pp. 1232–1251.

Białek J., Beręsewicz M. (2021), *Scanner data in inflation measurement: from raw data to price indices*, "The Statistical Journal of the IAOS", no. 37, pp. 1315–1336.

Białek J., Bobel A. (2019), *Comparison of price index methods for CPI measurement using scanner data*, 16th Meeting of the Ottawa Group on Price Indices, Rio de Janeiro.

Birnbaum Z.W., Saunders S.C. (1969), *A new family of life distributions*, "Journal of Applied Probability", vol. 6(2), pp. 637–652.

Bourguignon M., Lima M.D.C.S., Leão J., Nascimento A.D., Pinho L.G.B., Cordeiro G.M. (2015), *A new generalized gamma distribution with applications*, "American Journal of Mathematical and Management Sciences", vol. 34(4), pp. 309–342.

Brandt S. (2014), *Data analysis*, Springer International Publishing, Switzerland.

Brazauskas V . (2003), *Information matrix for Pareto (IV), Burr, and related distributions*, "Communications in Statistics-Theory and Methods", vol. 32(2), pp. 315–325.

Carli G. (1804), *Del valore e della proporzione de'metalli monetati*, "Scrittori Classici Italiani di Economia Politica", no. 13, pp. 297–336.

Carrasco J.M., Ortega E.M., Cordeiro G.M. (2008), *A generalized modified Weibull distribution for lifetime modelling*, "Computational Statistics and Data Analysis", vol. 53(2), pp. 450–462.

Castillo E., Hadi A.S., Balakrishnan N., Sarabia J.S. (2005), *Extreme Value and Related Models with Applications in Engineering and Science*, Wiley-Interscience, Hoboken.

Chen Z. (2000), *A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function*, "Statistics and Probability Letters", no. 49, pp. 155–161.

Chesneau C., Bakouch H.S., Hussain T. (2018), *A new class of probability distributions via cosine and sine functions with applications*, "Communications in Statistics-Simulation and Computation", vol. 48(8), pp. 2287–2300.

Chhikara R.S., Folks J.L. (1989), *The Inverse Gaussian Distribution: Theory, Methodology and Applications*, Marcel Dekker, New York.

Cooray K. (2006), *Generalization of the Weibull distribution: The odd Weibull family*, "Statistical Modelling", vol. 6(3), pp. 265–277.

Cordeiro G.M., Ortega E.M., Silva G.O. (2011), *The exponentiated generalized gamma distribution with application to lifetime data*, "Journal of Statistical Computation and Simulation", vol. 81(7), pp. 827–842.

Cordeiro G.M., Castellares F., Montenegro L.C., Castro M. de (2013), *The beta generalized gamma distribution*, "Statistics", vol. 47(4), pp. 888–900.

Drapella A. (1993), *The complementary Weibull distribution: unknown or just forgotten?*, "Quality and Reliability Engineering International", vol. 9(4), pp. 383–385.

Dutot C.F. (1738), *Reflexions Politiques sur les Finances et le Commerce*, Les Freres, The Hague.

El-Gohary A., Alshamrani A., Al-Otaibi A.N. (2013), *The generalized Gompertz distribution*, "Applied Mathematical Modelling", vol. 37(1–2), pp. 13–24.

El-Gohary A., El-Bassiouny A.H., El-Morshedy M. (2015), *Inverse flexible Weibull extension distribution*, "International Journal of Computer Applications", vol. 115(2), pp. 46–51.

Eltehiwy M., Ashour S. (2013), *Transmuted Exponentiated Modified Weibull Distribution*, "International Journal of Basic and Applied Sciences", vol. 2(3), pp. 258–269.

Felipe R.Sd.G., Edwin M.M.O, Gauss M.C. (2009), *The generalized inverse Weibull distribution*, "Statistical Papers", vol. 52(3), pp. 591–619.

Gaddum J.H. (1945), *Lognormal distributions*, "Nature", vol. 156(3964), pp. 463–466.

Ghitany M.E., Al-Hussaini E.K., Al-Jarallah R.A. (2005), *Marshall–Olkin extended Weibull distribution and its application to censored data*, "Journal of Applied Statistics", vol. 32(10), pp. 1025–1034.

Ghitany M.E., Al-Mutairi D.K., Balakrishnan N., Al-Enezi L.J. (2013), *Power Lindley distribution and associated inference*, "Computational Statistics and Data Analysis", no. 64, pp. 20–33.

Gumbel E.J. (1958), *Statistics of Extremes*, Columbia University Press, New York.

Hannan E.J., Quinn B.G. (1979), *The determination of the order of an autoregression*, "Journal of the Royal Statistical Society: Series B (Methodological)", vol. 41(2), pp. 190–195.

Javed M., Nawaz T., Irfan M. (2019), *The Marshall-Olkin kappa distribution: properties and applications*, "Journal of King Saud University-Science", vol. 31(4), pp. 684–691.

Jevons W.S. (1865), *The variation of prices and the value of the currency since 1782*, "Journal of the Statistical Society of London", no. 28, pp. 294–320.

Jędrzejczak A., Pekasiewicz D. (2020), *Teoretyczne rozkłady dochodów gospodarstw domowych i ich estymacja*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.

Johnson N.L., Kotz S., Balakrishnan N. (1995), *Continuous univariate distributions*, vol. 2, John Wiley & Sons, New York.

Kleiber C., Kotz S. (2003), *Statistical Size Distributions in Economics and Actuarial Sciences*, Wiley-Interscience, Hoboken.

Kotz S., Nadarajah S. (2000), *Extreme Value Distributions: Theory and Applications*, Imperial College Press, London.

Lindsey J.K. (2004), *Statistical analysis of stochastic processes in time*, vol. 14, Cambridge University Press, Cambridge.

Lu W., Shi D. (2012), *A new compounding life distribution: the Weibull–Poisson distribution*, "Journal of Applied Statistics", vol. 39(1), pp. 21–38.

Mahdavi A (2015), *Two Weighted Distributions Generated by Exponential Distribution*, "Journal of Mathematical Extension", vol. 9(1), pp. 1–12.

McDonald J.B. (1984), *Some generalized functions for the size distribution of income*, "Econometrica", vol. 52(3), pp. 647–663.

Nadarajah S., Rocha R. (2016), *Newdistns: An R package for new families of distributions*, "Journal of Statistical Software", no. 69, pp. 1–32.

Nakagami M. (1960), *The m-Distribution – A General Formula of Intensity Distribution of Rapid Fading*, [in:] W.C. Hoffman (ed.), *Statistical Methods in Radio Wave Propagation*, Pergamon, Oxford, pp. 3–36.

Okasha H.M., El-Baz A.H., Tarabia A.M.K., Basheer A.M. (2017), *Extended inverse Weibull distribution with reliability application*, "Journal of the Egyptian Mathematical Society", vol. 25(3), pp. 343–349.

Pal M., Tiensuwan M. (2015), *Exponentiated transmuted modified Weibull distribution*, "European Journal of Pure and Applied Mathematics", vol. 8(1), pp. 1–14.

R Core Team (2021), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, https://www.R-project.org/ [accessed: 24.11.2023].

Sarhan A.M., Apaloo J. (2013), *Exponentiated modified Weibull extension distribution*, "Reliability Engineering and System Safety", no. 112, pp. 137–144.

Sarhan A.M., Zaindin M. (2009), *Modified Weibull distribution*, "APPS. Applied Sciences", no. 11, pp. 123–136.

Schwarz G. (1978), *Estimating the dimension of a model*, "The Annals of Statistics", vol. 6(2), pp. 461–464.

Shahbaz M.Q., Shahbaz S., Butt N.S. (2012), *The Kumaraswamy–Inverse Weibull Distribution*, "Pakistan Journal of Statistics and Operation Research", vol. 8(3), pp. 479–489.

Shanker S., Shukla K.K. (2019), *A generalization of Generalized Gamma distribution*, "International Journal of Computational and Theoretical Statistics", vol. 6(1), pp. 33–42.

Silver H., Heravi S. (2007), *Why elementary price index number formulas differ: Evidence on price dispersion*, "Journal of Econometrics", no. 140, pp. 874–883.

Stacy E.W., Mihram G.A. (1965), *Parameter estimation for a generalized gamma distribution*, "Technometrics", vol. 7(3), pp. 349–358.

Subhradev S., Mustafa C.K., Haitham M.Y. (2018), *The Quasi XGamma-Poisson distribution: Properties and Application*, "Istatistik: Journal of the Turkish Statistical Assocation", vol. 11(3), pp. 65–76.

Sulewski P., Białek J. (2022), *Probability Distribution Modelling of Scanner Prices and Relative Prices*, "Statistika: Statistics & Economy Journal", vol. 102(3).

Tieling Z., Min X . (2007), *Failure Data Analysis with Extended Weibull Distribution*, "Communications in Statistics – Simulation and Computation", no. 36, pp. 579–592.

Witkovsky V . (2001), *Computing the distribution of a linear combination of inverted gamma variables*, "Kybernetika", vol. 37(1), pp. 79–90.

Yusuf A., Qureshi S. (2019), *A five parameter statistical distribution with application to real data*, "Journal of Statistics Applications and Probability Letters", no. 8, pp. 11–26.

Modelowanie rozkładu cen skanowanych i indeksów cen za pomocą rozkładów teoretycznych z dwoma, trzema, czterema i pięcioma parametrami

Streszczenie: W artykule podjęto problematykę odpowiedniego dopasowania teoretycznego rozkładu prawdopodobieństwa do empirycznego rozkładu cen skanerów. W badaniu empirycznym wykorzystano dane skanerowe z jednej sieci handlowej w Polsce, tj. miesięczne dane dotyczące jogurtów naturalnych, napojów jogurtowych, ryżu długoziarnistego i kawy w proszku, sprzedanych w 212 placówkach w styczniu i lutym 2022 roku. Ceny i ceny względne modelowano za pomocą pięćdziesięciu dwu-, trzy-, cztero- i pięcioparametrowych rozkładów prawdopodobieństwa z nieujemną dziedziną. Niektóre z nich składały się z dość znanych rozkładów, które nazywane są ich specjalnymi przypadkami. Łączna liczba tych rozkładów, które pośrednio wzięły udział w badaniu, to ponad sto. Do analizy porównawczej wykorzystywano kryteria informacyjne, takie jak AIC, BIC, HQIC i wartości *p* testów dobroci dopasowania. W artykule wykazano, że modele takie jak Frechet, Pareto IV i Log-Logistic można uznać za bardzo dokładne, co stanowi dobrą podstawę do badań symulacyjnych wskaźników cen czy też konstrukcji tzw. wskaźników cen ludności. Wzory na dystrybuantę wykorzystanych modeli oraz kody R niezbędne do przeprowadzenia badań przedstawiono w załączniku.

Słowa kluczowe: modelowanie danych, dane skanowane, rozkłady cen

JEL: C13, C43, E31

Received: 2023-11-18; revised: 2024-02-12. Accepted: 2024-06-17