TJMM
8 (2016), No. 1, 77-87

STATISTICAL ANALYSIS OF CONSUMER PRICE INDICES

ANAMARIA POPESCU

Abstract. This paper attempts to characterize through statistical indicators-quantitative measures-chronological series that we have available. The purpose of this paper is to present the calculation of statistical indicators that are used frequently for statistical characterization and adjustment of statistical series time series by identifying the most appropriate statistical model then evolves over time this causes. We can thus analyze central tendency, and data variability, form and concentration distributions package data using analytical tools in Microsoft Excel that enables automatic calculation of descriptive statistics using Data Analysis option from the Tools menu.

The analysis of consumer price indices in the period 1990 - 2014 and information website provided by the National Statistics Institute (INS) seems to be a certain tendency towards eccentricity and asymmetry of financial data series.

1. Introduction

Starting from concepts and data describe trends in the indices of consumer prices in Romania, we have made some statistical conclusions on them. After identifying a heterogeneous set of variables selected were evaluated by descriptive statistics homogeneity, asymmetry (skewness), vaulting (kurtosis) and normal distribution of data for a period of 25 years.

In this study used four sets of data on consumer price indices, ie annual index of consumer prices total annual index of consumer prices of food products, the annual index of consumer prices of non-food goods and the annual index of prices consumer services.

Annual consumer price index (CPI) measures the overall evolution of prices for purchased goods and tariffs for services used by the population in the current year over the previous year (or another year chosen as the reference period).

This index is calculated as a ratio, expressed as a percentage, between the average index of prices in the current year and the average index of the previous year (or another year chosen as the reference period).

The annual inflation rate is calculated by subtracting 100 from the annual index of consumer prices.

CPI inflation is often associated with that refers to the overall price increase in an economy. Known as inflation, CPI indicate the change in the price of consumer goods and services. Although, there are many different methodologies to measure inflation, CPI uses a constant basket of goods and services, and these are: food, non-food and services.

Most often, the information we are presented as a numeric enumeration value whose meaning is at first glance cipher. For example, we find on the Internet, the NSI website, database CPI Consumer Price Index for a period of 25 years.

Thus, sequencing numbers in the first phase gives us information about the unit observed (point in time). If you want information but overall, the period of time or people

2010 *Mathematics Subject Classification.* 62P20, 91B38, 62J02.

Key words and phrases. descriptive statistics, asymmetry, eccentricity, normal distribution, correlation, regression equation, Data Analysis.

observed the sequencing data remains less conclusive. In this case, the question of extracting information from the data stream, which can be done through different ways of processing the data.

Numerical data processing comes with some additional information, discovering certain features of the phenomenon (eg, media) that were not available directly from the data stream. Processing consists of calculating parameters that define summarizes overall data set, such as mean, median, modal, standard deviation, asymmetry, excess etc.

A more complete picture is obtained when the series calculate various parameters that describe a whole series of data.

We mention three groups of parameters i.e. parameters that describe the phenomenon of central tendency (mean, median, modal), or variation of parameters describing the values of the population.

Arithmetic mean. If the variable is normally distributed, the average indicates the middle of the range between minimum and maximum (the distribution range of data). Also in the case of normal distribution around the mean (ie within-average standard deviation, mean ± standard deviation are most of the data).

The standard error. The standard error is involved in the calculation of 95% confidence interval around the mean (only one variable with normal distribution), it is also involved in the statistical inference.

The median is a value of the series so that half of the observations have values less than (or equal) and the other half have values greater than (or equal). If normal distribution mean and median are equal. So the median and arithmetic mean are indicators for normal distribution, with values closer as they have the more likely that the variable is normally distributed.

The module is the value which has the highest frequency in the series. If a situation arises module is where the series is not how, that all values appear only once, when displayed value #N/A. Another possible situation is that the series to be bimodal or tri-modal. Then only the first value will be displayed in order of their appearance in the series. In this case all values for determining the unit can make a frequency table.

The module is useful in the case of qualitative variables ordered, but also in other types of variables, for example in the case of continuous variable with normal distribution module is likely to have a value close to the average.

Standard Deviation or standard deviation can be calculated and the standard deviation STDEV or STDEV.P population.

Standard deviation shows that the root mean square deviation is the arithmetic variable. If you have a low value, then the data varies slightly around the average. If the distribution is represented by Gauss (distribution is normal) distributions occur following data (according Cebâșev's theorem):

The standard deviation shows us the deviation mean square from the arithmetic average of the variable values. If it has a low value when the data vary slightly around the average. In the case where the distribution is represented by the curve of Gauss (distribution is normal) take place following distributions of data (according Cebâșev's theorem):

– Measure the height vaulting excess or flattening or vaulting a distribution compared to a normal distribution. The excess is zero for a series of data with a normal distribution is positive for a series of data with train higher than a normal distribution and negative for a number of data train is lower than that of a normal distribution . Because distribution is close to normal, vaulting should be close to 0.

– Measured asymmetry deviation from the symmetrical layout and direction of asymmetry (positive or negative) over the normal curve. Asymmetry is 0 for a series of data with a normal distribution is negative for a number of data asymmetrical left (series has several smaller values) is positive for a data series asymmetrical right (series has several values big). Because distribution is close to normal asymmetry should be close to 0.

Conditions which must be met for a series to have normal distribution:

– Arithmetic average ≈ median ≈ mode
– Asymmetry and vaulting to be as close to 0 or 0.

Frequently, in scientific research, occurring phenomena in which dominant partition law is normal distribution. Support for this law is central limit theorem of probability theory. By virtue of this theorem amount a sufficiently large number of independent random variables

$$S_n = X_1 + X_2 + ... + X_n \tag{1}$$

tends to the limit by normal distribution, with media and finished variations $V(X_k) = \sigma_k^2$, $k \in \mathbb{N}^*$.

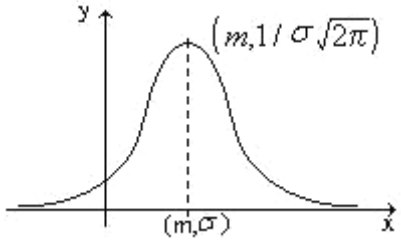


FIGURE 1. The density of the normal distribution curve

A continuous random variable follows a normal distribution if it has the following probability density function (p.d.f.):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}},\ x \in \mathbb{R},\ m \in \mathbb{R},\ \sigma > 0. \tag{2}$$

The probability density function (the cumulative distribution function) is given by:

$$P(\mathrm{x}, \mathrm{m}, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \int_{x_1}^{x_2} f(x)dx. \tag{3}$$

The properties of normal distribution curves may be formulated:

– allow a maximum point $x = m$:

$$f(m) = \frac{1}{\sigma\sqrt{2\pi}} \tag{4}$$

and decreases asymptotically toward the x-axis to the left and to the right.
– are symmetrical to the corresponding ordered mean ($x = f(m)$- the axis of symmetry).
– the bell-shaped with the convexity facing upward, in the area of the maximum point; curve has two inflection points located at $x = \pm\sigma$ a distance from the mean.
– ordered function is all the greater as the mean square deviation σ is less; if σ increase curves "flatten it" increasingly more to flattening (curves leptycurtiks).

The probability that the random variable continues to take values within the range $(-\infty, +\infty)$ is:

$$P(\mathrm{x} < +\infty) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = 1 \tag{5}$$

i.e. $\forall\, x \in \mathbb{R}$, $\exists\, P(\mathrm{x}) \leq 1$.

The probability density $f(x)$ is continuous, non-negative and satisfies the condition:

$$\int_{-\infty}^{+\infty} f(x)dx = 1 = P(x). \tag{6}$$

To simplify the calculation function $P(x, m, \sigma)$ operation is performed to change the variable

$$z = \frac{x - m}{\sigma} \tag{7}$$

a normal random variable, of course.

With this notation, the normal distribution $N(m, \sigma)$ becomes $N(0, 1)$. Normal probability density variable rules, in this case it can write:

$$f(z) = f(z, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \tag{8}$$

and distribution function takes the form:

$$P\left(z < z_p\right) = \frac{1}{\sqrt{2\pi}} \int_{-¥}^{z_p} e^{-\frac{z^2}{2}} dz. \tag{9}$$

The links which exist between the two statistical variables can be studied using two techniques: correlation and regression.

The correlation shows how strong is the link, the dependency of the variables, and regression helps in explaining and forecasting a factor based on the value of another (others).

The coefficient of correlation shows the extent to which the variations in a variable are correlated with variations in other variables and can take values between -1 and $+1$.

Between -1 and 0, the connection between the two variables is of the opposite direction and is even more intense, as it draws to a close. Between 0 and 1, the connection between the two variables is direct and is more intense, the more it approaches 1.

The coefficient of determination R^2 measures the rate of change in one variable that can be assigned (or explained) by the variation of other variables.

Presentation of the link between two linear variables, when it exists, is called the method of linear regression (linear regression).

For this purpose it is considered one of the independent variable or variables as a predictor variable and the other variable as the dependent variable or variable response (outcome).

Linear link between the two variables is described by an equation linear regression equation (regression equation) which corresponds geometrically right regression (the regression line): $y = a + bx$ where a is called interceptor and b regression coefficient.

2. Results

As an example of the specificity of financial data series was selected a range of data from the National Institute of Statistics (INS) and the specialized website of NIS, between 1990 and 2014.

The four series of data on consumer price indices are annual index of consumer prices overall ($Y1$) annual index of consumer prices of food products ($X1$), the annual index of consumer prices of non-food goods ($X2$) and index annual consumer price services ($X3$).

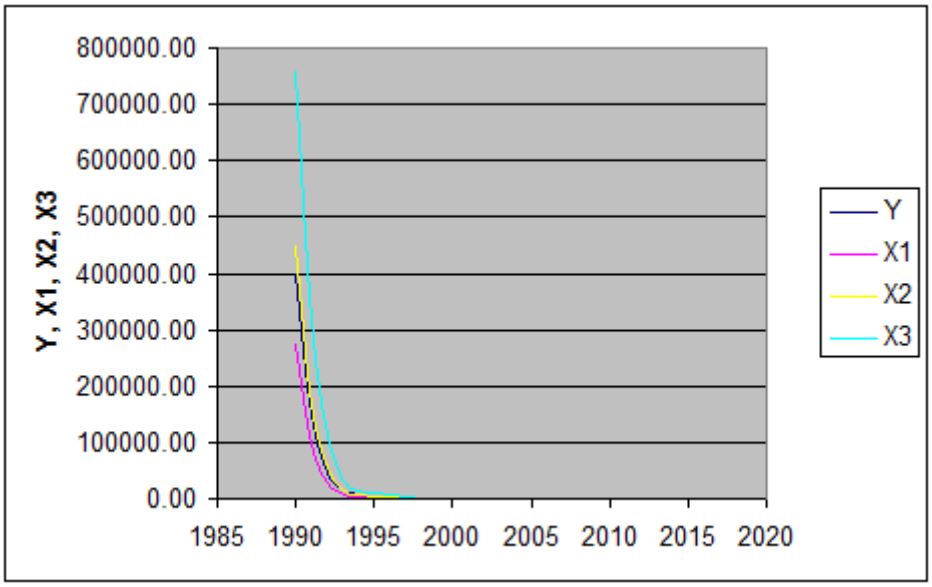
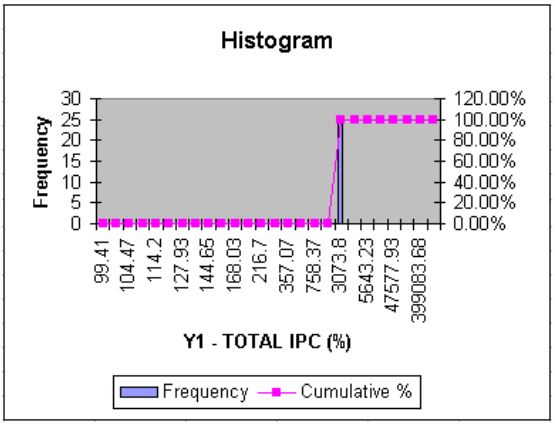


FIGURE 2. Evolution of price indices in Romania between 1990-2014
Source: INS, `http://statistici.insse.ro/shop/`

Microsoft Excel enables automatic plotting of the histogram using the Data Analysis option from the Tools menu.

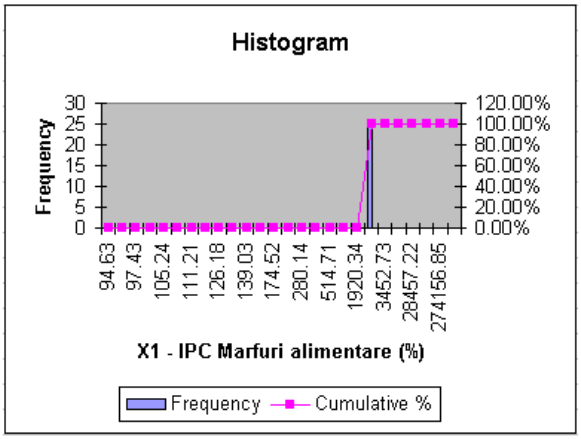
FIGURE 3. Histograms of consumer price indices
Annual consumer price index total

TOTAL IPC (%)	*Frequency*	*Cumulative %*
99.41	0	0.00%
100.47	0	0.00%
104.47	0	0.00%
107.96	0	0.00%
114.2	0	0.00%
121.16	0	0.00%
127.93	0	0.00%
137.97	0	0.00%
144.65	0	0.00%
154.14	0	0.00%
168.03	0	0.00%
187.99	0	0.00%
216.7	0	0.00%
265.54	0	0.00%
357.07	0	0.00%
520.13	0	0.00%
758.37	0	0.00%
1206.53	0	0.00%
3073.8	25	100.00%
4266.72	0	100.00%
5643.23	0	100.00%
13360.69	0	100.00%
47577.93	0	100.00%
147699.36	0	100.00%
399083.68	0	100.00%
More	0	100.00%



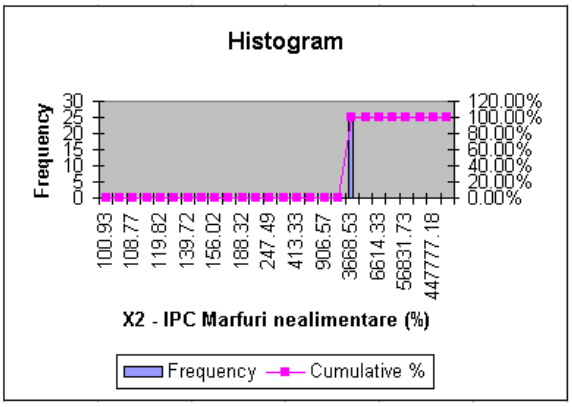
Annual index of consumer prices of food products

IPC Marfuri alimentare (%)	Frequency	Cumulative %
94.63	0	0.00%
96.2	0	0.00%
97.43	0	0.00%
99.27	0	0.00%
105.24	0	0.00%
107.7	0	0.00%
111.21	0	0.00%
121.46	0	0.00%
126.18	0	0.00%
131.03	0	0.00%
139.03	0	0.00%
152.2	0	0.00%
174.52	0	0.00%
206.45	0	0.00%
280.14	0	0.00%
402.59	0	0.00%
514.71	0	0.00%
763.97	0	0.00%
1920.34	0	0.00%
2618.67	25	100.00%
3452.73	0	100.00%
8156.52	0	100.00%
28457.22	0	100.00%
95792.05	0	100.00%
274156.85	0	100.00%
More	0	100.00%



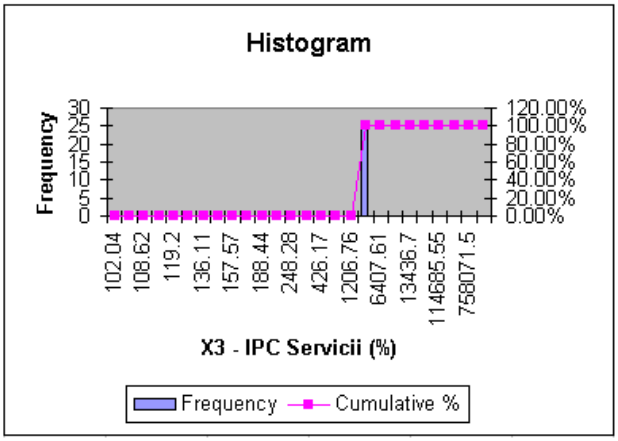
Annual index of consumer prices of non-food goods

IPC Marfuri nealimentare (%)	Frequency	Cumulative %
100.93	0	0.00%
103.4	0	0.00%
108.77	0	0.00%
112.87	0	0.00%
119.82	0	0.00%
131.54	0	0.00%
139.72	0	0.00%
148.61	0	0.00%
156.02	0	0.00%
169.24	0	0.00%
188.32	0	0.00%
213.17	0	0.00%
247.49	0	0.00%
310.62	0	0.00%
413.33	0	0.00%
595.19	0	0.00%
906.57	0	0.00%
1452.76	0	0.00%
3668.53	25	100.00%
5102.58	0	100.00%
6614.33	0	100.00%
15401.29	0	100.00%
56831.73	0	100.00%
167205.82	0	100.00%
447777.18	0	100.00%
More	0	100.00%



Annual consumer price index services

IPC Servicii (%)	Frequency	Cumulative %
102.04	0	0.00%
105.26	0	0.00%
108.62	0	0.00%
114.12	0	0.00%
119.2	0	0.00%
124.9	0	0.00%
136.11	0	0.00%
147.77	0	0.00%
157.57	0	0.00%
170.49	0	0.00%
188.44	0	0.00%
216.19	0	0.00%
248.28	0	0.00%
314.73	0	0.00%
426.17	0	0.00%
655.72	0	0.00%
1206.76	0	0.00%
2317.64	25	100.00%
6407.61	0	100.00%
9415.63	0	100.00%
13436.7	0	100.00%
33699.56	0	100.00%
114685.55	0	100.00%
321625.58	0	100.00%
758071.5	0	100.00%
More	0	100.00%



From histograms can be seen that the four annual indices of consumer prices did not have a normal distribution.

Microsoft Excel allows the automatic calculation of descriptive statistics using Data Analysis option from the Tools menu.

Obtained automatically arithmetic mean, mode, median, variance etc. for the four data series submitted for analysis, the annual index of consumer prices overall ($Y1$) annual index of consumer prices of food products ($X1$), the annual index of consumer prices of non-food goods ($X2$) and the index's annual consumer price services ($X3$).

From the table we conclude that the coefficient of vaulting / flattening for the four annual indices of consumer prices is positive (series platikurtic).

Asymmetry coefficient is positive and greater than 1 for the four annual indices of consumer prices, situation that shows that the series are significantly asymmetrical right and a distribution data appears to the left (mean > median).

The financial series generates greater uncertainty because of econometric modeling or depreciating significantly intensifies and correlation of data series that characterize the financial markets, as it becomes apparent from the correlation matrix applied to the series of consumer price indices.

Matrix of correlation applied to independent variables (1990 - 2014)

	$X1$	$X2$	$X3$
$X1$	1		
$X2$	0.999544767	1	
$X3$	0.996887245	0.998756	1

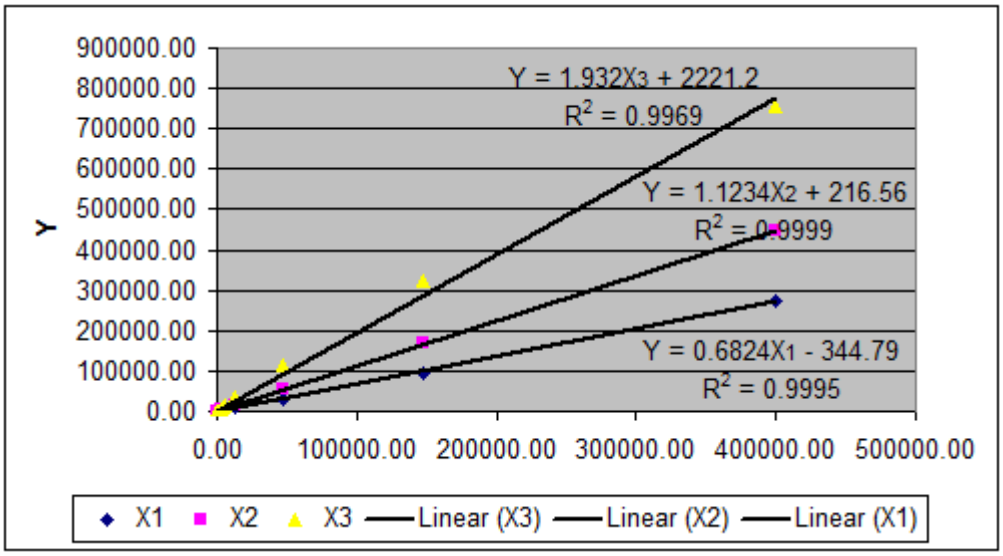
The correlation coefficients resulting from the comparison of independent variables $X1$, $X2$ and $X3$ is close to 1, it is clear that there is a correlation between the significant variables, i.e. they are more linear dependent from each other.

FIGURE 4. Descriptive statistics indices of consumer prices (1990-2014)

	Y	X1	X2	X3
Mean	25023.93	16731.13	28328.79	50568.09
Standard Error	16729.83	11419.28	18795.09	32372.67
Median	216.7	174.52	247.49	248.28
Mode	#N/A	#N/A	#N/A	#N/A
Standard Deviation	83649.17	57096.39	93975.47	161863.3
Sample Variance	7E+09	3.26E+09	8.83E+09	2.62E+10
Kurtosis	18.2436	18.90615	18.09156	16.55444
Skewness	4.170265	4.249459	4.150746	3.971102
Range	398984.3	274062.2	447676.3	757969.5
Minimum	99.41	94.63	100.93	102.04
Maximum	399083.7	274156.9	447777.2	758071.5
Sum	625598.1	418278.3	708219.8	1264202
Count	25	25	25	25
Confidence Level(95.0%)	34528.68	23568.23	38791.17	66813.9

Software used MS Excel
Sursa: INS, `http://statistici.insse.ro/shop/`

FIGURE 5. The correlation between Y and $X1$, $X2$, $X3$



Dependence of independent variables and Y is positive: an increase in the variables $X1$, $X2$ and $X3$ involves an increase in the total consumer price index, Y. regression Lines have an uptrend, and dispersion diagrams indicate an increasing trend.

Interpretation of R^2: 99% of the variance in Y can be explained by the linear relationship of $X1$, $X2$ and $X3$.

The coefficient of the variable $X1$ from the equation for the regression Y shall be construed as follows: for each increase in the $X1$ with a unit of measure (%p.a.), Y increases with 0.68 percentage points, the coefficient of the variable $X2$ from the equation

FIGURE 6. Sequence data sheet of the program with input and automated statistical analysis results

TOTAL IPC (%)			f(x)	P(x)
399083.68	TOTAL IPC - Y			*1.00*
147699.36			*0.14*	*0.93*
47577.93	Mean	25023.9252	*0.38*	*0.61*
13360.69	Standard Ei	16729.8332	*0.40*	*0.44*
5643.23	Median	216.7	*0.39*	*0.41*
4266.72	Mode	#N/A	*0.39*	*0.40*
3073.80	Standard D	83649.1658	*0.39*	*0.40*
1206.53	Sample Var	6997182946	*0.38*	*0.39*
758.37	Kurtosis	18.2436031	*0.38*	*0.39*
520.13	Skewness	4.1702655	*0.38*	*0.38*
357.07	Range	398984.27	*0.38*	*0.38*
265.54	Minimum	99.41	*0.38*	*0.38*
216.70	Maximum	399083.68	*0.38*	*0.38*
187.99	Sum	625598.13	*0.38*	*0.38*
168.03	Count	25	*0.38*	*0.38*
154.14	Confidence	34528.6784	*0.38*	*0.38*

for the regression Y shall be construed as follows: for each increase in the $X2$ with a unit of measure (%p.a.), Y increases with 1.12 percentage points and the coefficient of the variable $X3$ from the equation for the regression Y shall be construed as follows: for each increase in the $X3$ with a unit of measure (%p.a.), Y increases with 1.93.

Assuming normal distributions of the four strings, each string individually, averages and dispersions in Figure 4 can be calculated for each element in each row, the value density function of normal distribution with relation (8) applied for the determination of point values:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{10}$$

Figure 6 also shows the use of expression (10) and values fixed m and σ^2 and the function of the distribution of probability, calculated as a function of cumulative frequencies, $P(x)$.

The density of the distribution is calculated by writing function =1/SQRT(2 * PI()) * EXP(- ((x-mean) $\hat{}$2/2/variance)), and the function of the distribution by applying instruction = NORMDIST(x, mean, variance, cumulative).

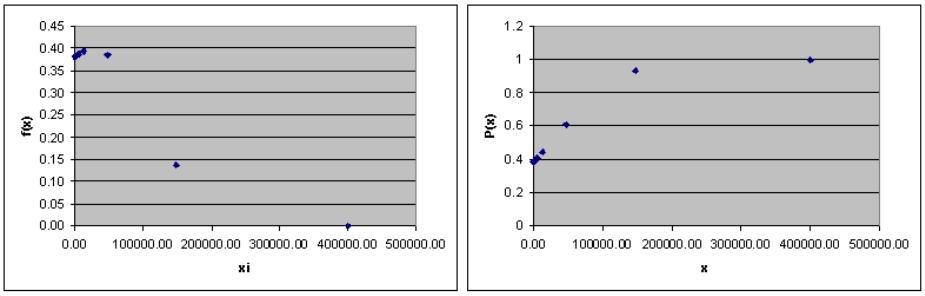
Using the data in columns $f(x)$ and $P(x)$ plot graphs of density distribution and the probability distribution function.

It can be seen in Figure 7, the four indices of consumer prices follow a distribution leptokurtic (mean > median), which indicates that the data are grouped and close to median batch of data with a high degree of heterogeneity of values.

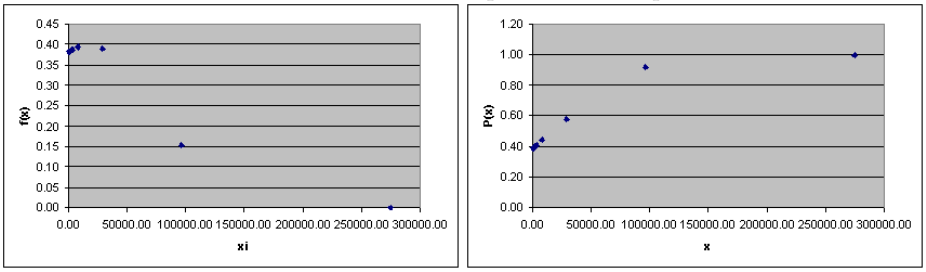
The data series have module (#N/A), which shows that all values appear only once.

FIGURE 7. Graphs density function of the distribution and of the function of the distribution of the probability

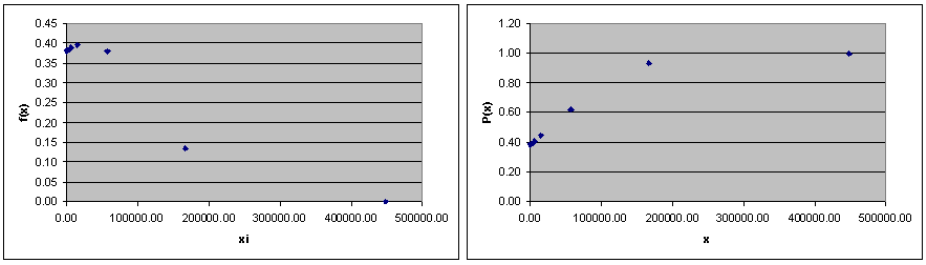
Annual consumer price index total



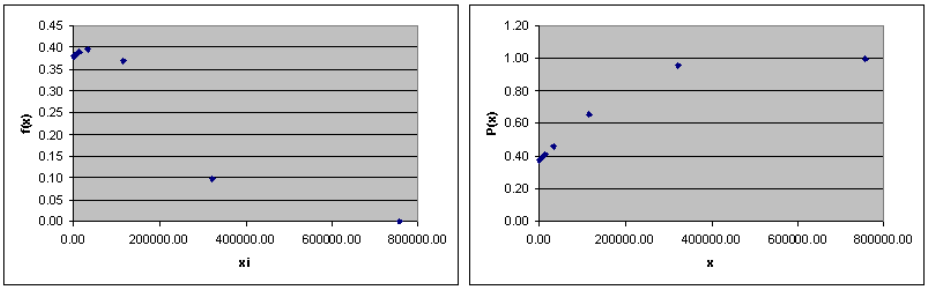
Annual index of consumer prices of food products



Annual index of consumer prices of non-food goods



Annual consumer price index services



3. CONCLUSIONS

Descriptive statistics on the investigation of the four sets of financial data points out that the coefficients ai asymmetry indices of consumer prices are nearly symmetrical, analyzed 1990-2014.

The investigation descriptive statistics on the four sets of financial stresses that asymmetry coefficients of annual consumer price indices are almost symmetrical analyzed period 1990-2014.

In leptokurtic distribution, the probability of an event is extremely high probability of occurrence of a normal distribution (and vice versa, in a leptokurtic distribution, the probability of an event is extremely lower probability of occurrence of a normal distribution). As a result, valuation models stock prices and yields can cause errors if we assume that their distribution is normal.

In general empirical data series are slightly or moderately asymmetrical as proof that on certain days in the financial market has been no quotations for example higher or even very large or smaller, and very low in comparison with previous day (defining moderate or pronounced asymmetries positive and negative).

References

[1] Anghelache, C., Anghel, M.C., *Metode de analiză a PIB prin utilizarea modelelor statistico-econometrice*, Revista / Journal ,,Economica" nr. 1 (91), 2015.
[2] Anghelache, C., Pagliacci, G.R.M., Prodan, L., *Model for macroeconomic – Analyse based on the regression function*, Romanian Statistical Review 1/2013.
[3] Florens, J.P., Malavorti, L., *Instrumental regression with discrete endogeneous variables*, Working paper, GREMAQ, Universite de Sciences Sociales, Toulouse, 2003.
[4] Oroian, M., Radnoti-Szakacs, A., Simonfi, A., *Studiu empiric asupra relațiilor de cauzalitate dintre rata inflației și unele variabile macroeconomice*, Academica Science Journal Studia Series, No. (2) 1 - 2013, ISSN: 2285-9314, Dimitrie Cantemir University.
[5] Popescu, A., Stoicuța, N., *Statistical analysis of monetary policy indicators variability*, The 23rd Conference on Applied and Industrial Mathematics CAIM 2015, September 17-20, 2015, Suceava, Romania.
[6] Săvoiu, G., *Statistical characteristics of financial data sets*, Romanian Statistical Review nr. 4/2013.
[7] `http://statistici.insse.ro/shop/`

Department of Mathematics & Computer Science
Faculty of Science
University of Petroșani
20 University Street, Petroșani, România
E-mail address: `am.popescu@yahoo.com`