A C T A U N I V E R S I T A T I S L O D Z I E N S I S
FOLIA OECONOMICA 3(302), 2014

Aleksandra Baszczyńska[*], *Dorota Pekasiewicz*[**]

BIAVERAGE AND MULTIMODALITY IN INVESTIGATING DISTRIBUTION OF ELECTRICITY PRICES

Abstract. In the paper chosen statistical methods concerning analysis of random variable distributions are presented. Investigating modality of distribution is one of the most interesting and important stages in random variable analysis. Among others, the following methods can be used: kernel density estimation, the Hartigan test of unimodality and the biavarage estimation. An example showing application of these methods for data from the one-day-ahead market of electricity is presented.
Keywords: kernel estimation, Hartigan test, dip statistic, biaverage, one-day-ahead market

I. INTRODUCTION

Discovering the modality of random variable distribution is one of the most interesting and important stages of statistical analysis. The number of modes can be regarded as the fundamental information in choosing the proper statistical procedure. The analysis of unimodality of distribution of a random variable may result, for example, in the possibility (or impossibility) of using classical methods. The problem of multimodality is also closely connected with distinguishing populations, which are represented in a sample. The question of number of populations arises in economic, social and medical phenomena.

The issue of modality is widely discussed in statistical literature. There are methods of estimation of the whole density function that allow for visual evaluation of, among others, the number of modes, for example: Parzen (1962) and Silverman (1996). There are also methods of estimation mode and biaverage values, for example Sokołowski (2013) and Antoniewicz (2005). Verification of the null hypothesis concerning modality, including the special case of the hypothesis that the distribution is unimodal, can be found, for example, in

[*] Ph.D., Department of Statistical Methods, University of Łódź.
[**] Ph.D., Department of Statistical Methods, University of Łódź.

Hartigan, Hartigan (1985), where the dip test is based on histogram and in Silverman (1981), where the test is based on kernel density estimator. It is claimed that methods mentioned above should be used to a great extend in statistical analysis of random variable distribution. The choice of methods assessing multimodality, that are presented in the paper, results from their availability and computational complexity.

Presentation and indicating the possibility of application of chosen methods of assessing multimodality are the main aims of this paper.

The paper is organized as follows. Section II presents methods usually used in the first stages of analysis of distribution – kernel estimation of density function and the statistical test verifying the hypothesis of unimodality of the distribution. Section III describes useful method of mode estimation in bimodal distribution by the biaverage. Section IV presents an attempt to apply the considered methods in analysis of distribution of electricity prices in Poland. The results of empirical study and conclusions are presented in section V.

II. KERNEL DENSITY ESTIMATION AND HARTIGAN DIP TEST FOR UNIMODALITY

Kernel density estimation is a nonparametric procedure which provides a way of approximating the structure of data.

Let $X_1, X_2, ..., X_n$ be a sample of size *n* from a continuous random variable X with density f.

The kernel density estimator is defined in the following way (Silverman (1996)):

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1}$$

where: n is a sample size, h is a smoothing parameter and $K(\cdot)$ is a kernel function.

Kernel function and smoothing parameter control, respectively, the shape and the spread of kernel used in estimation. In many practical implementations of estimator (1) the smoothing parameter is specified as $h = 1{,}06\hat{\sigma} n^{-1/5}$, where $\hat{\sigma}$ is the sample standard deviation. The procedure requires also the specification of kernel function, that is, in many cases, the density function of the standardized normal distribution is used $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$. The discussion concerning the choices of other parameters of kernel methods and their influence on the kernel density estimator is presented, for example, in Baszczyńska (2012).

The problem of the verification of hypotheses about the number of modes can be formulated as :

$$H_0: \ j \le k,$$

$$H_1: \ j > k,$$

where j is the true number of modes and k is a fixed positive integer number.

For the case $k = 1$, we can apply the dip test of unimodality presented in Hartigan, Hardigan (1985). This test is one of the widely used test in assessing the unimodality. The "dip" put forward by Hartigan, measures departures from unimodality, so the dip statistic is defined as the maximum difference between the observed distribution (empirical distribution function) and the best fitting unimodal distribution function that minimizes that maximum difference. The probability of the rejection of the unimodal distribution is calculated empirically and tabulated as a function of sample size.

The Hartigan dip test is available in R or Matlab.

III. BIAVERAGE FOR BIMODAL DISTRIBUTION

The term biaverage for bimodal distribution and, as generalization, k-average for k-modal distribution was proposed by Antoniewicz in 2005. Biaverage is defined as a pair of parameters given by the concentration of probabilistic masses. The following condition allows, under the assumption that random variable has four first moments, to calculate the values of biaverage (m_1, m_2):

$$\min_{a,b} E\left((X-a)(X-b)\right)^2 = E\left((X-m_1)(X-m_2)\right)^2 . \tag{2}$$

When the random variable considered has four moments and its variance is not equal to zero, then the solution of equation (2) is the following (cf. Antoniewicz (2005)):

$$m_1 = \frac{1}{2}\left(P - \sqrt{P^2 + 4Q}\right), \tag{3}$$

$$m_2 = \frac{1}{2}\left(P + \sqrt{P^2 + 4Q}\right), \tag{4}$$

where

$$P = \frac{E\left(X^3\right) - E\left(X^2\right)E(X)}{E\left(X^2\right) - E^2(X)}, \tag{5}$$

$$Q = \frac{E\left(X^2\right)E\left(X^2\right) - E\left(X^3\right)E(X)}{E\left(X^2\right) - E^2(X)}. \tag{6}$$

The dispersion of the random variable around the biaverage can be calculated as

$$V_0 = E\left((X - m_1)(X - m_2)\right)^2. \tag{7}$$

Then, the standard deviation of the biaverage has the following form:

$$\sigma_0 = \sqrt[4]{V_0}. \tag{8}$$

The value of biaverage can be evaluated using the random sample $X_1, X_2, ..., X_n$ chosen from a bimodal population. First, the following values are calculated:

$$p = \frac{n\sum_{i=1}^{n} x_i^3 - \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}, \tag{9}$$

$$q = \frac{\left(\sum_{i=1}^{n} x_i^2\right)^2 - \sum_{i=1}^{n} x_i^3 \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{10}$$

and $c = \frac{1}{2}\sqrt{p^2 + 4q}$.

Then

$$c=\frac{\sqrt{\left(n\sum_{i=1}^{n}x_i^3-\sum_{i=1}^{n}x_i^2\sum_{i=1}^{n}x_i\right)^2+4\left(\left(\sum_{i=1}^{n}x_i^2\right)^2-\sum_{i=1}^{n}x_i^3\sum_{i=1}^{n}x_i\right)\left(n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2\right)}}{2\left(n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2\right)}. \quad (11)$$

Next the estimator of the biaverage $(\hat{m}_1,\ \hat{m}_2)$ is found in such a way that:

$$\hat{m}_1=\frac{1}{2}p-c\,, \quad (12)$$

$$\hat{m}_2=\frac{1}{2}p+c\,. \quad (13)$$

It is shown (Antoniewicz (2006)) that if sample moments are good estimators of population moments, the biaverage is a good estimator of modes and specifies concentration of two probability masses.

IV. EMPIRICAL ANALYSIS OF THE DISTRIBUTION OF ELECTRICITY PRICES

The empirical analysis was carried out to show clearly some aspects of multimodality in the distribution of electricity prices in Poland. The data was taken from the database of Market Data on Polish Power Exchange (http://www.tge.pl). Weighted average prices (in PLN/MWh) of transactions on the trading session during the continuous trading for particular hours on day-ahead market were taken into account. In the empirical study two periods of time were chosen:

period A: 26.08.2013 – 21.09.2013 (592 observations),

period B: 9.12.2013 (24 observations).

The choice of these periods was made on the basis of the analysis of the database and has the advantage that small and big samples were considered in the study and, in addition, different features of modality can be presented.

In the initial stage of the study, the kernel density estimation was made for the periods considered. The results are presented in Figures 1–2.

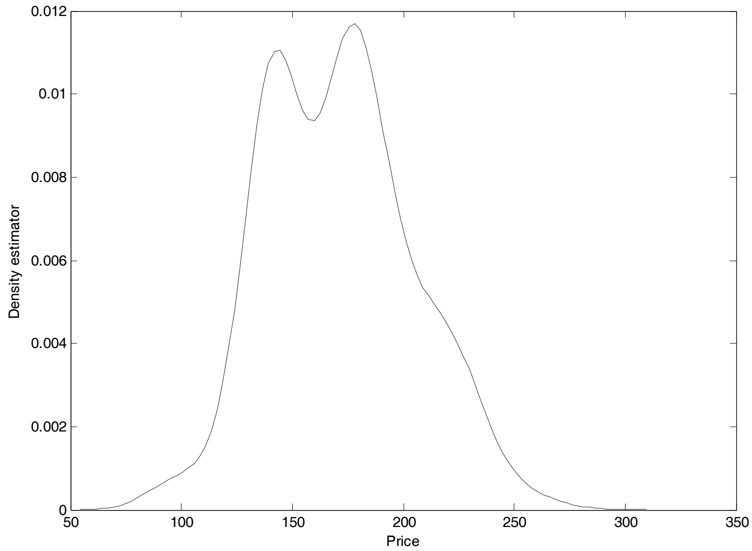

Figure 1. Kernel density estimation for data from period A
Source: own work.

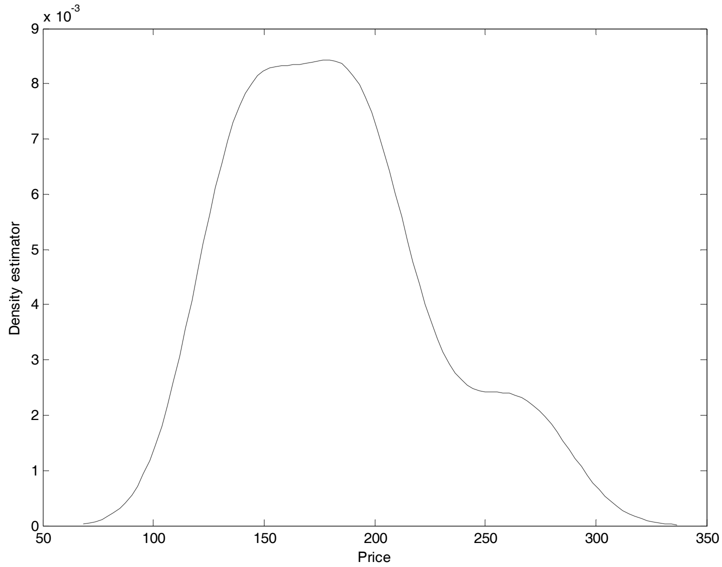

Figure 2. Kernel density estimation for data from period B
Source: own work.

At the same initial stage, the Hartigan dip test was used to assess the modality of distribution of prices in two periods. When 592 observations were taken into account (period A) the value of the dip statistic was 0.0602. For sample of 24 observations (period B) this value was 0.0875.

In the case of period A the results of kernel density estimation indicated the bimodality of the distribution of electricity prices. For period B it was rather difficult to assess modality visually on the grounds of the kernel estimation only. The *p*-value for the Hartigan dip test was equal to 0.000 for period A, so we rejected the hypothesis about unimodality of the distribution of the random variable considered. For period B, *p*-value was equal to 0.127 indicating that we could not reject this hypothesis. The results of the Hartigan dip test are consistent with Figures 1–2.

In the next stage, we studied biaverages for two periods A and B. The results were as follows:

period A: ${}_A\hat{m}_1 = 110.655$, ${}_A\hat{m}_2 = 240.135$, ${}_A\hat{\sigma}_0 = 58.460$;

period B: ${}_B\hat{m}_1 = 108.476$, ${}_B\hat{m}_2 = 283.314$, ${}_B\hat{\sigma}_0 = 78.468$.

The values $\hat{m}_1$, $\hat{m}_2$ can be treated as the mode estimates of electricity prices distribution.

V. CONCLUSIONS

The results of the considered empirical study show that modality should be assessed using different statistical methods and procedures. Kernel density estimation allows to evaluate the basic characteristic features of distribution including modality. But in many cases the big number of modes (or large level of smoothing the estimator) can be caused by wrong-selected smoothing parameter. The Hartigan dip test did not confirm bimodality in period B, but the application of biaverage to mode estimation proved to be useful.

The Hartigan dip test is a useful method, but the form of the null hypothesis is a drawback. If we decide that the null hypothesis should be rejected, we do not know exactly what kind of multimodality is in the population. In this situation using *k*-average provides a useful statistical method in the analysis of multimodal populations.

The generalization of the results of empirical study needs further research using, for example, simulation methods.

ACKNOWLEDGEMENTS

This work was supported by the project number DEC-2011/01/B/HS4/02746 from the National Science Centre.

REFERENCES

Antoniewicz R. (2005), *O średnich i przeciętnych*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław.

Antoniewicz R. (2006), *Pięć lat pojęcia dwuprzeciętnej*, Zeszyty Naukowe Uniwersytetu Szczecińskiego, 450, 11–15.

Baszczyńska A. (2012), *Estymacja funkcji gęstości z pakietem MATLAB*, Rola informatyki w naukach ekonomicznych i społecznych. Innowacje i implikacje interdyscyplinarne, Wydawnictwo Wyższej Szkoły Handlowej w Kielcach, 7–16.

Hartigan J. A., Hartigan P. M. (1985), *The Dip Test of Unimodality*, The Annals of Statistics, Vol. 13, No. 1, 70–84.

Parzen E. (1962), *On Estimation of a Probability Density Function and Mode*, Annals of Mathematical Statistics, 33, 1065–1076.

Silverman B. W. (1981), *Using Kernel Density Estimates to Investigate Multimodality*, Journal of the Royal Statistical Society, Series B (Methodological), Vol. 43, No. 1, 97–99.

Silverman B. W. (1996), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London.

Sokołowski A. (2013), *Bezpośrednie estymatory modalnej*, Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie, Kraków.

http://www.tge.pl [11.12.2013]

Aleksandra Baszczyńska, *Dorota Pekasiewicz*

BIŚREDNIA I WIELOMODALNOŚĆ W BADANIACH ROZKŁADU CEN ENERGII

W pracy przedstawiono wybrane metody statystyczne dotyczące analizy rozkładu zmiennej losowej. Do badania modalności, która jest jednym z ważniejszych etapów analizy zmiennej losowej mogą być zastosowane, między innymi, następujące metody: estymacja jądrowa gęstości, test Hartigana jednomodalności oraz biśrednia. Metody te wykorzystane zostały do badania rozkładu cen energii na rynku dnia następnego w Polsce.