DOI: https://doi.org/10.53486/cike2023.47

UDC: 330.322.54(560)

ESTIMATION OF THE TURKISH STOCK INVESTOR NUMBERS BASED ON KERNEL METHOD

ERCAN ÖZEN
Faculty of Applied Sciences,
University of Uşak, 64200, Uşak, Türkiye
Email: ercan.ozen@usak.edu.tr

MEHMET NIYAZI ÇANKAYA
Faculty of Applied Sciences,
University of Uşak, 64200, Uşak, Türkiye
Email: mehmet.cankaya@usak.edu.tr

Abstract: The investigation of investments in stock markets is an important and a challenging topic. Knowing the number of Turkish stock investor makes a benefit to conduct a smooth economical circulation. In this study, it is aimed that the Turkish stock investor numbers are estimated by using the kernel estimation method. The built functions in the Mathematica software are used. The built functions are responsible to model the data sets smoothly. After necessary components of the statistical results in the built function, the random number generation is performed for the estimated function, $\hat{f}$, which is the sampling form of the assumed function f for the population representing the investor numbers. The numbers are estimated to be around 5 million for the real data sets if the maximum order statistics are used. The confidence interval as a covering probability being %95 in the asymptotic theory shows that the estimated value can be around 7 million if the new population occurs.

Key words: Investor, Stock market, Estimation, Investment, Statistics.

JEL Clasification: Q1, R1, H254

Introduction

The investigation of Turkish stock market investments is an important and a challenging topic in the Turkish economy. Nowadays, there are many factors affecting the behaviors of potential investors. Estimation of the stock market investor numbers help the economy in various aspects. The companies can use it as one of the decision tools. To get credit, borrowing, issue bonds or going public are some options for companies to get working capital or fixed asset capital. In general to get short term capital or bond issuing can be considered cheaper capital for the companies as long as business profit is higher than their cost. However, in other cases, providing cash by equity can be considered a more appropriate, reliable and stable source. Thus, it is important for companies to know how many investors will be in the capital markets and how much capital support they will provide to companies.

Since the economy is an accelerator at many directions in the working field of the business, etc, the estimation of the Turkish stock investor numbers will help to set the budget or many indictors in the working of government or many fields business. The balanced economical movement can trust the number of investors who play role and support the process of economical evaluation. In this principle,

the conditions set by investigation and circulation in the umbrella of economy will also affect and support the number of investors. For these reasons, the economy depends many factors which can already be populations. Since these factors play role, the estimated number can change and the different statistical methods should be performed as well.

Equipment based on mathematical and statistical perspectives are required for us to model actual life created randomly by people who try to live and conduct random occurrences that are not predicted in most circumstances. For these reasons, the probabilities appear in the findings of the experiment performed on the universe's life. Because the random events we encounter happen whenever users touch and play with the experiment, the observations should be described by some functions. According to the statistical perspective, several mathematical tools exist in the theoretical component of the discipline. We can select one of them to conduct the analysis. We can build a prediction and forecasting for the analyzing principle.

All investors are decentralized, which means that the market is influenced/determined as much by the government or bank groups in exchange and the currency can be inflated, and so the currency can be changed, rotated, and cycled in various ways. Economic marketplaces are defined by currency buyers and sellers; they allow for various payments that affect the economy. Because of the decentralized market structure, the currency market is extremely volatile; traders can either make a lot of money or face significant loses based on the assessed valuations of important items in the cyclic economy. Investors can be the most valuable indicators in the world. In addition to, there can be a merging among firms as well, which can lead to increase the numbers of investors (Komáromi, 2006; Lhabitant, & Gregoriou, 2008).

Since the mixing of many factors play role in creating the artificial data sets, many summarizing statistics should also be given to depict or try to discover the role of factors as many as possible to do so. In other words, the maximum value of a data set can be regarded as another population which is a moved one from a population. The same approach can be considered for the case in which the minimum value of a data set is chosen. Other summarizing statistics such as mean, median, etc. as a central tendency of population can also be used to summarize the whole data completely. Other statistics such as standard deviation, mean absolute deviation, etc. should be provided to see where and how the data sets behave or occur.

In this study, we estimate the number of Turkish stock investors. The true data set for Turkish investor numbers will be utilized to estimate how many Turkish investors are computed in order to give a suggestion for economic actors. The economy, in particular, is a critical role in establishing favorable economic conditions. Section 1 gives a quick summary of the preliminary information provided by the literature. Section 2's research technique provides an introduction to the kernel estimation method. Section 3 presents the key findings. The final section of the article is reserved for the paper's discussions and conclusions.

Literature review.

Kernel density smoothing, like the histogram, is a technique for identifying structure in the data without the use of a parametric model. The formula for the kernel density estimator is

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

where K is called as the kernel function and guarantees

$$\int_{-\infty}^{\infty} K(x)dx = 1,$$

and h is called the smoothing parameter or bandwidth at histogram ($h > 0$). By downscaling, the aforementioned equation can be expressed more concisely $K_h(u) = \frac{1}{h}K(\frac{u}{h})$ such that:

$$\hat{f}(x;h) = \frac{1}{n}\sum_{i=1}^{n} K_h(x-X_i).$$

By centering a scaled kernel at X and taking the average of the n kernel ordinates there, one can derive the kernel estimate. In comparison to regions with fewer observations, the estimate is larger the more samples that fall inside a bandwidth. This method is comparable to the histogram method in that we are now adding the kernel ordinates over the window width h rather than adding the number of observations in a bin.

The Gaussian distribution or standard normal is one frequent option for K:

$$\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{x^2}{2\sigma^2}\right).$$

The random variable X is assumed to be normal and $X \sim N(0,\sigma^2)$.

The mixture model, which can be thought of as a type of kernel approach, is a helpful tool for density estimation. The form of the Gaussian mixture model is

$$\hat{f}(x) = \sum_{m=1}^{M}\alpha_m\phi(x;\mu_m,\Sigma_m),$$

with mixing propositions $\alpha_m, \Sigma_m\alpha_m = 1$ as a summation of weightings. Note that $0 \le \alpha_m \le 1$. μ_m and Σ_m are location and scale parameters of the normal distribution.[7] [3,4].

[7] https://alain-vandormael.netlify.app/post/ksmooth/, (Accessed on 12.08.2023); https://mathematica.stackexchange.com/questions/61827/structure-of-kernelmixturedistribution-and-smoothkerneldistribution, (Accessed on 12.08.2023)

Research methodology

The role of software is inevitable if we need to apply the highly computational process. In order to conduct the process, the Mathematica software is used. The statistical viewpoint is about the kernel estimation method and its mixing form which provides an attractive and comprehensive competence in modelling the data set if the non-identically occurs at high degree or the assumption of being identical is violated importantly. The built functions in Mathematica software are capable to perform the kernel smoothing techniques precisely. On the other hand, after modelling the data set is performed, we can have the sampling form of function or the estimate form of the function f, i.e., $\hat{f}$. Since $\hat{f}$ is determined, the built functions in Mathematica provides a compatible platform to add this $\hat{f}$ into random generator function in Mathematica, that is,

$$\text{RandomVariate}[\text{SK}[w], n];$$

where SK represents the smooth kernel function:

$$\text{SK}[w] = \text{SmoothKernelDistribution}[x, \{\text{Adaptive}, \text{Automatic},.1\}, \text{PerformanceGoal} \rightarrow \text{Quality}];$$

and

$$\text{SK}[w] = \text{KernelMixtureDistribution}[x, \text{Automatic}, \text{"SemiCircle"}];$$

Comprehending KernelMixtureDistribution, whose PDF SmoothKernelDistribution interpolates, is essential for comprehending SmoothKernelDistribution. Simple coordinates for that interpolation make up the SmoothKernelDistribution internals (Lhabitant, & Gregoriou, (2008)

Note that even if the real data set is discrete, it is logical to expect that a discrete data set can convergence to continuous data set. For example, binomial distribution can convergence to normal distribution (Hastie, et. al., 2009; Casella & Berger, 2021).

The computational procedure and summarizing

There are two main buildings which are the modelling on real data sets and the generating data sets. The following computational schema is used to perform the computation.

1. Import the real data set;
2. Perform the kernel estimation method via KernelMixtureDistribution and SmoothKernelDistribution functions in Mathematica;
3. Set a random number generation based on the kernel method;
4. Set the number of sample size in the computationally generated data set;
5. Get the statistics such as mean, median, standard deviation, median absolute deviation, minimum, maximum;
6. Set these values;

7. Summarize and make the graphical illustrations for quickly summarizing the results.

The following section is divided for the results of the functions in Mathematica.

Results

There are two main figures which are based on the sample sizes chosen as $n = 246$ and $n = 1000$. The number of first sample size represents the number of real data sets. The second number of that is chosen arbitrarily; because, if the number of sample size increased, we aim to observe the behavior of artificial data set based on the determined function $\hat{f}$ based on the kernel estimation method.

The kernel estimation method uses the continuous kernel function such as normal distribution. In the Mathematica software, there is an adaptative method based on the behavior of data set.

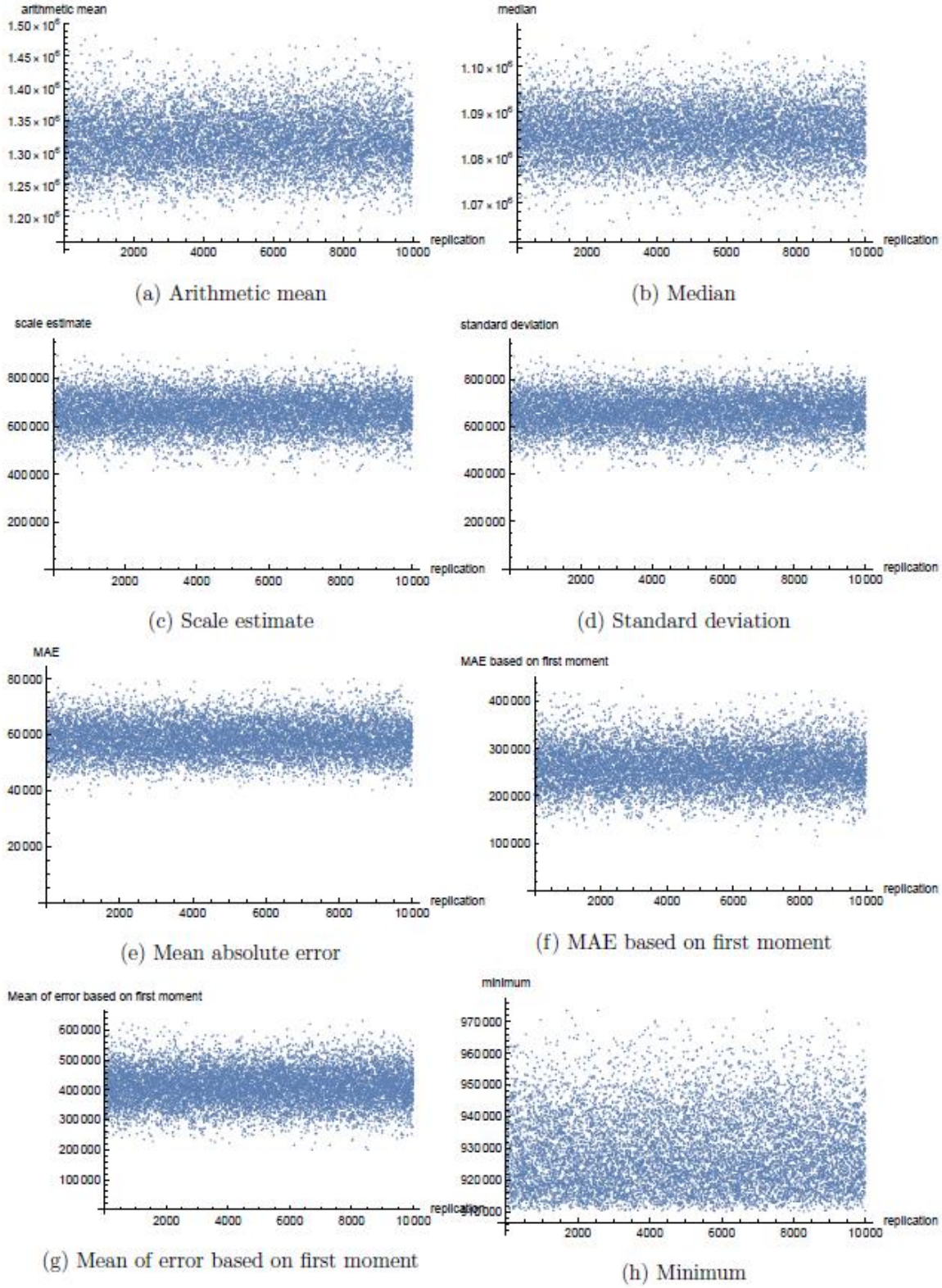
The following figures represent the statistics about arithmetic mean, median, scale estimate based on kernel method, standard deviation, mean absolute error, its variants such as moments based on the kernel method, minimum and maximum statistics as well as percents for the quantile statistics where and how the data set takes their values in order to observe which value can be generated for the probabilities such as 0.01, 0.25, 0.50, 0.75 and 0.99. These probabilities can be alternative statistics for the minimum and maximum values as well.

It is importantly noted that when the maximum values are taken into account, it can be observed that there is a potential estimation for the values being lower 5 million. However, since we perform an estimation procedure, it is logical to expect that there can be over 5 million. That is, it is confidently said that we can estimate values around 5 million for the numbers of Turkish stock investors.

On the other hand, the estimated value can be supported by means of the confidence interval as given below:

$$\hat{\theta} + z_{\frac{\alpha}{2}}\sqrt{Var(\hat{\theta})},$$

where $\hat{\theta}$ represents the maximum values of the statistic occurred due to the replicated form conducted by simulation. In this case, $[5 - 2 * 1{,}5 + 2 * 1] = [3{,}7]$ as a confidence interval of the maximum values.



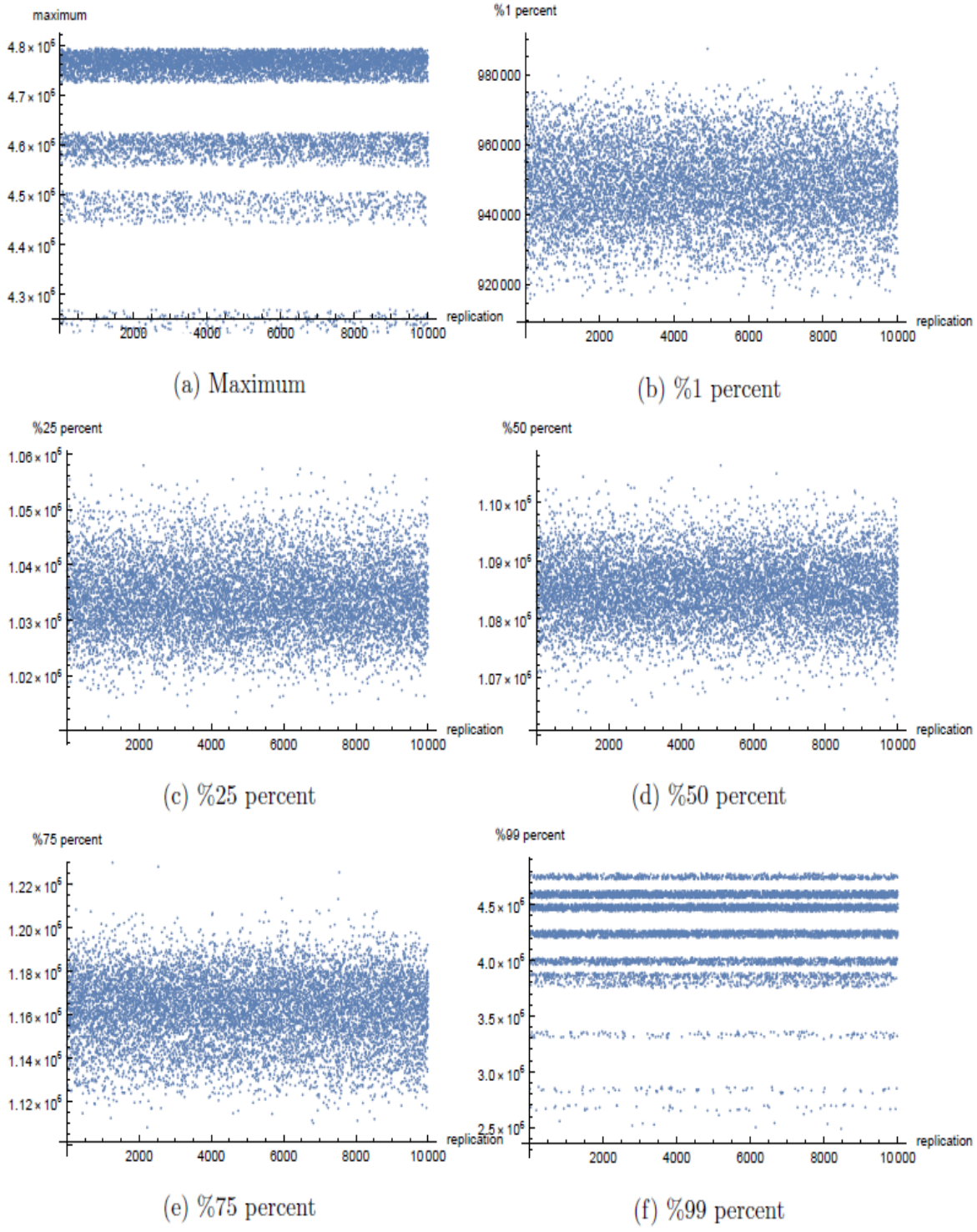
(a) Arithmetic mean (b) Median

(c) Scale estimate (d) Standard deviation

(e) Mean absolute error (f) MAE based on first moment

(g) Mean of error based on first moment (h) Minimum

Figure 1: The statistics about the artificial data set when $n = 246$ (continue)

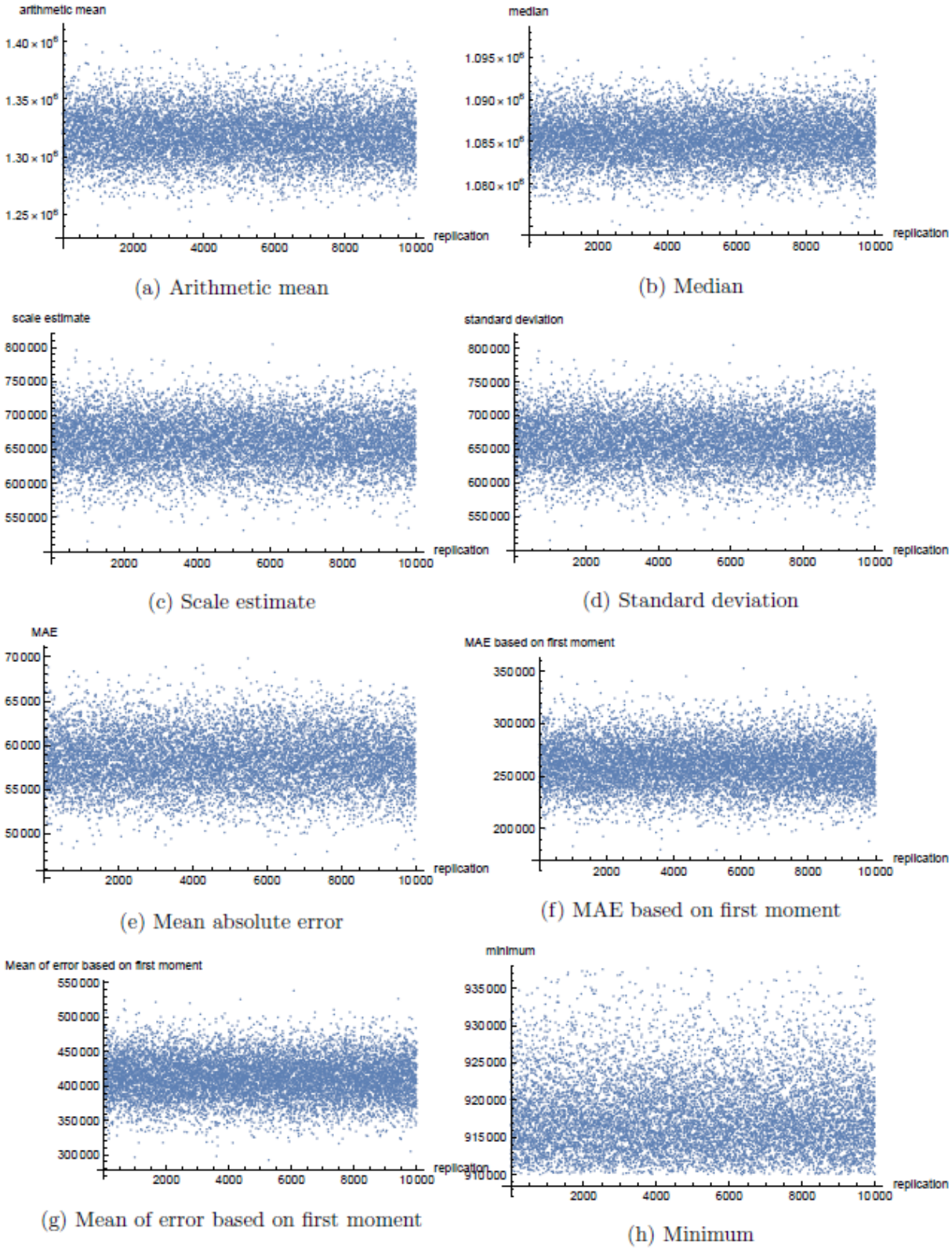


(a) Maximum (b) %1 percent

(c) %25 percent (d) %50 percent

(e) %75 percent (f) %99 percent

Figure 1: The statistics about the artificial data set when $n = 246$ (continue)



(a) Arithmetic mean (b) Median

(c) Scale estimate (d) Standard deviation

(e) Mean absolute error (f) MAE based on first moment

(g) Mean of error based on first moment (h) Minimum

Figure 2: The statistics about the artificial data set when $n = 1000$

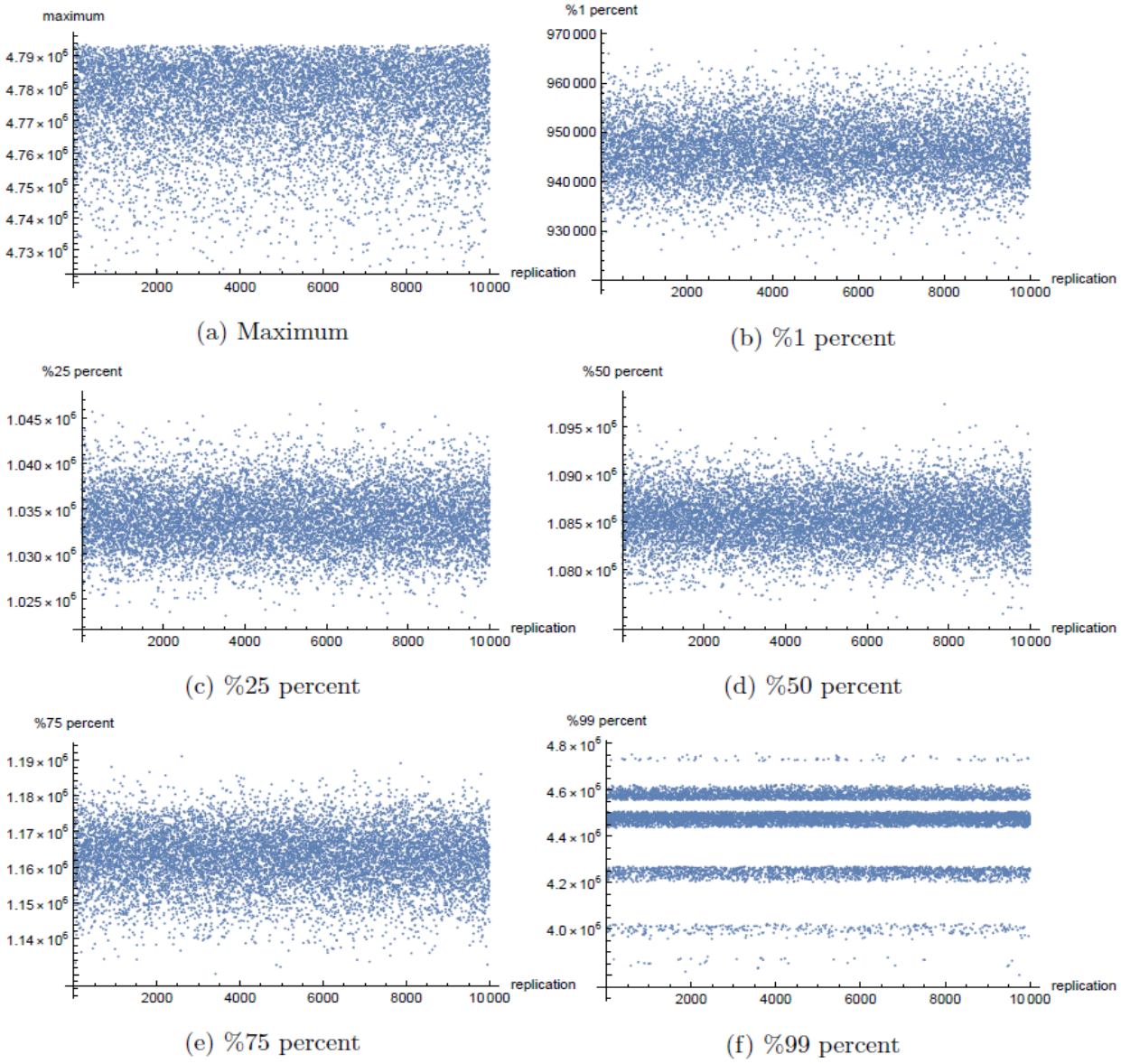


Figure 2: The statistics about the artificial data set when $n = 1000$ (continue)

Figure 1 and Figure 2 are mainly same statistics for the sample sizes 246 and 1000, respectively. The figures 1 and 2 cont. (a) represent the numbers for the statistics based on the maximum value. In addition, the scale estimate used for the confidence interval is given by the figures 1 and 2 cont. (c). The values from these figures labelled as (a) and (c) are responsible to construct a confidence interval given above. It should be noted that since these estimates for the scale parameter can be expected to be near or over 1 million due to the fact that we perform an estimation procedure.

Discussion and conclusions

Since the kernel method is used to estimate the numbers of Turkish stock investor, the results of computational procedure have been based on the data-adaptive approach; because, the empirical

cumulative distribution function is mainly used in the kernel estimation method. In other words, the main idea in the kernel is the empirical probabilistic behavior and so the estimated values from the generated data based on this empirical probabilistic behavior have been observed around the values of real data. The estimated values for the maximum value of statistics are around 5 million for Turkish stock investor. The upper bound from the confidence interval is around 7 million. It is generally commented that we can accomplish to perform a good estimation for the numbers of Turkish stock investor and the kernel estimation method is performable to conduct a good estimation for such real data sets.

The forthcoming studies will be based on the parametric approaches and the mixing of the parametric models.

REFERENCES:

[1] Komáromi, G. (2006). Anatomy of stock market bubbles. ICFAI UP.
[2] Lhabitant, F. S., & Gregoriou, G. N. (2008). Stock market liquidity: implications for market microstructure and asset pricing (Vol. 420). John Wiley & Sons.
[3] https://alain-vandormael.netlify.app/post/ksmooth/, (Accessed on 12.08.2023)
[4]https://mathematica.stackexchange.com/questions/61827/structure-of-kernelmixturedistribution-and-smoothkerneldistribution, (Accessed on 12.08.2023)
[5] Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.
[6] Casella, G., & Berger, R. L. (2021). Statistical inference. Cengage Learning.

Appendix: Turkish Stock Market Investor Number