NATIONAL BANK OF KAZAKHSTAN

DETERMINATION OF THE OPTIMAL INTERVAL SIZE FOR CALCULATING THE MEDIAN OF INFLATION EXPECTATIONS.

Monetary Policy Department
Working Paper №2024-8

Yerzhan I.
Akpanov A.

Working papers and analytical notes of the National Bank of the Republic of Kazakhstan (“the NBK”) are meant to disseminate the outcomes of the NBK’s studies as well as of other academic research works of the NBK staff. Economic studies are disseminated in order to encourage discussions. The opinions expressed in the paper reflect the personal standing of the authors and may not coincide with the official standpoint of the NBK.

Determination of The Optimal Interval Size for Calculating the Median of Inflation Expectations

November 2024

NBRK – WP – 2024 – 8

Abstract

Yerzhan I.[1]
Akpanov A.[2]

Inflation expectations play a pivotal role for central banks, particularly in countries with inflation targeting regimes. They exert a significant influence on the actual level of inflation by shaping the decisions of both consumers and businesses. This influence manifests through demand-pull and cost-push inflation mechanisms. On the one hand, consumers adjust their purchasing decisions for goods and services based on anticipated price trends. On the other hand, businesses set prices according to expected costs, including wage dynamics, which contribute to cost-push inflation.

The precise measurement of inflation expectations is critically important for conducting effective monetary policy. This process is influenced by various factors, including the sample size of respondents and the calculation methodology employed. Since January 2023, the National Bank of the Republic of Kazakhstan has been employing median estimates of inflation expectations, making the determination of optimal interval widths a key issue. This study analyzes existing methods for determining the optimal interval width for the median of grouped data on public inflation expectations and identifies the most suitable interval width for Kazakhstan.

Key Words: inflation expectations, interval width, median estimates, Berk method

JEL-Classification: E31, E52, C83, C43

[1] Yerzhan I.S. – Deputy Head, Macroeconomic Research and Forecasting Division, Monetary Policy Department, National Bank of the Republic of Kazakhstan. E-mail: Islam.Yerzhan@nationalbank.kz
[2] Akpanov A.K. – Leading analyst, Macroeconomic Research and Forecasting Division, Monetary Policy Department, National Bank of the Republic of Kazakhstan. E-mail: Arman.Akpanov@nationalbank.kz

CONTENT

1. INTRODUCTION

The measurement of inflation expectations is critically important for central banks in making monetary policy decisions, as expectations directly influence the economic behavior of consumers and investors (Vellekoop and Wiederholt, 2019). These expectations shape the dynamics of actual inflation, as decisions regarding consumption, investment, pricing, and wage-setting reflect households' and businesses' anticipations of future price changes (IMF, 2023). Anchored inflation expectations help reduce uncertainty and improve the forecasting of economic conditions.

Accurate measurement of inflation expectations enables central banks to effectively assess and manage monetary policy, achieving long-term inflation targets. The precision of these estimates depends on numerous factors, among which the selection of an appropriate calculation methodology plays a particularly significant role.

In Kazakhstan, the National Bank of the Republic of Kazakhstan (NBRK) has been conducting monthly surveys of the population since January 2016 to assess inflation expectations. The survey covers the adult population (18 years and older), with the sample distributed by gender, age, ethnicity, and region in accordance with official demographic statistics. The representativeness of the sample is ensured through random selection of respondents, allowing for the inclusion of various regions and socio-demographic groups.

Initially, the calculation of inflation expectations in Kazakhstan was carried out using the Berk quantification method. However, this method has its methodological limitations, which are associated with the quantification of inflation expectations based on a probabilistic approach. It assumes that inflation expectations depend on the current level of inflation, indicating the adaptive nature of inflation expectations. As a result, the quantitative estimates of the population's inflation expectations tend to be close to the actual annual inflation, which, in turn, may create the illusion of anchored inflation expectations.

Since January 2023, the NBRK has ceased publishing quantified values of expected inflation. Currently, in its communications, the NBRK uses median estimates of perceived and expected inflation intervals, which provide a more accurate representation of the distribution of public expectations. At present, this approach is one of the most widely used methods for assessing inflation expectations. Most central banks in developed countries adopt this method, as it is independent of current inflation and is considered more objective. Estimates obtained through this method reflect the numerical range of expected price changes, making them more informative and useful for analyzing the population's inflation expectations.

The median formula for grouped data accounts for interval sizes, frequencies, and cumulative frequencies, with the assumption that the interval width influences the accuracy of the median estimate. The main hypothesis of this study is that dynamic interval sizes are preferable to fixed ones for grouping survey data and subsequently calculating the population's inflation expectations. The aim of this paper is to determine the optimal interval width for calculating the median of grouped data on inflation expectations used in the NBRK's calculations.

Within the scope of this research, median estimates are analyzed using various interval widths for grouped data. Given that there is currently no unified method for determining the optimal interval width, the study tests multiple tools. For instance, the Freedman-Diaconis equation represents a statistical method for determining the optimal bin width when constructing a histogram of the data. This rule helps create a histogram that most accurately reflects the true data distribution, avoiding excessive smoothing or over-detailing. Additionally, the methodology proposed by Tao et al. (2023) is applied for analysis, where a novel approach to data grouping – the BSI method – is introduced to determine the optimal interval width required for constructing histograms of multimodal datasets. Each of the different interval widths is used to construct corresponding histograms, followed by statistical deconvolution[3]. Additionally, this study analyzes the adequacy of the sample size used in the NBRK's inflation expectation surveys by applying the Cochran and Slovin formulas – two widely used methods for sample size calculation in statistical research. The application of these formulas allows for optimizing the data collection process, ensuring the representativeness and adequacy of the sample for conducting accurate analyses.

This paper consists of several sections. The second section provides a review of international literature on the subject of the study. The third section examines international practices regarding the number of respondents and the intervals used in various countries. Following this, the methodology for determining the optimal number of respondents is described, and various methods for identifying the optimal interval width for grouping data are discussed. The concluding sections of the paper present the calculation results and conclusions.

2. LITERATURE REVIEW

Inflation expectations are one of the key elements in macroeconomic analysis and decision-making regarding monetary policy. They influence the behavior of consumers and businesses, shaping the dynamics of actual inflation and economic growth. The concept of inflation expectations was introduced into economic theory within the frameworks of adaptive and rational expectations. Muth (1961) introduced the concept of rational expectations, which became fundamental in modern macroeconomics, while Friedman (1968) emphasized the importance of economic agents' expectations in his work on the role of monetary policy.

Anchored inflation expectations reduce economic uncertainty and contribute to lower actual inflation, despite various shocks (Mishkin, 2007). In this context,

[3] In statistics and data analysis, deconvolution is a method used to "unfold" or refine data, providing a more accurate representation of its true distribution. This approach is particularly valuable when the data has been grouped into intervals or distorted due to measurement constraints. By applying deconvolution, hidden information can be uncovered, significantly enhancing the accuracy and reliability of the analysis.

managing inflation expectations is a key tool for achieving price stability and ensuring the effectiveness of monetary policy (Woodford, 2003).

Orphanides and Williams (2005) demonstrated in their work that accurately measuring inflation expectations allows central banks to respond more effectively to inflation shocks and adjust policies in line with changing economic conditions. Furthermore, central bank transparency and communication regarding inflation targets strengthen public trust and help align expectations with policy objectives (Svensson, 1997).

Thus, the development and application of reliable methods for measuring inflation expectations are critically important for central banks within the framework of inflation targeting. This not only enhances the effectiveness of monetary policy but also strengthens the trust of economic agents, ultimately contributing to price stability (Bernanke et al., 1999).

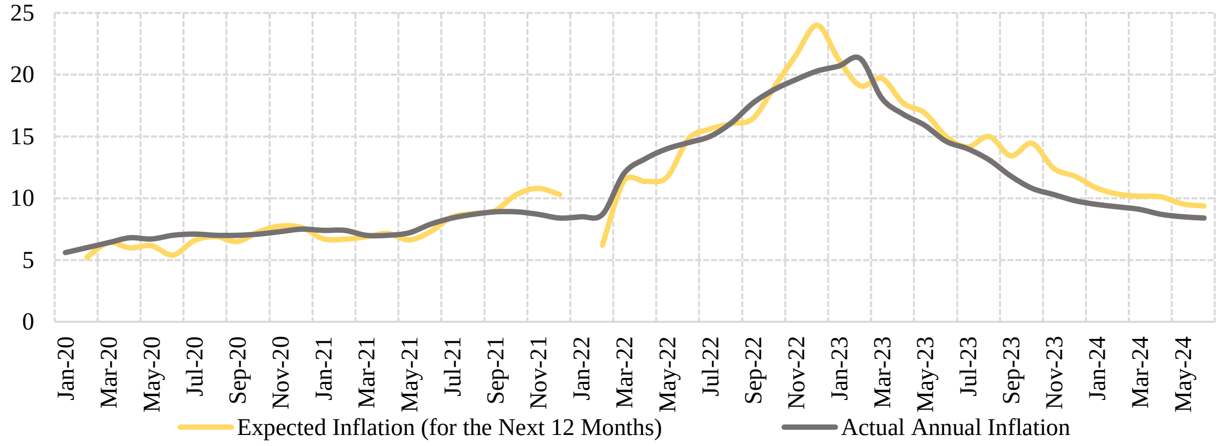
One of the early methods for quantitatively assessing inflation expectations is the method proposed by Berk (1999). This method is based on converting respondents' qualitative assessments into quantitative indicators. The essence of quantification lies in aligning qualitative responses with current inflation indicators. Mathematical expectations of the distribution are then calculated based on this data, serving as quantified estimates of expected inflation.

$$\pi_m^e = \frac{\pi(C+D)}{C+D-(A+B)}, \quad (1)$$

where $A=F^{-1}(1-a^e)$; $B= F^{-1}(1-a^e-b^e)$; $C= F^{-1}(1-a^e-b^e-c^e)$; $D=F^{-1}(1-e^e)$, F^{-1} *– Inverse of the normal distribution function,* a^e *– share of respondents expecting prices to increase more rapidly,* b^e *– share of respondents expecting prices to increase at the same rate,* c^e *– share of respondents expecting prices to increase more slowly,* e^e *– share of respondents expecting prices to decrease.*

Initially, the NBRK calculated and published estimates of the population's inflation expectations using the probabilistic quantification method based on Berk's methodology. However, this method has several weaknesses. Quantification using this approach implies a dependence of inflation expectations on its current value, inherently assuming the adaptive nature of expectations (Nardo, 2003). An additional drawback of Berk's quantification method is the smoothing of inflation expectations, as survey results are multiplied by the actual annual inflation rate of the previous month. This results in quantitative estimates of the population's inflation expectations being close to the actual annual inflation rate, which, in turn, may create the "appearance" of anchored inflation expectations (Graph 1). Moreover, converting qualitative responses into quantitative estimates can introduce errors and distortions, affecting the accuracy of measurements (Smith and McAleer, 1995). This issue arises because quantification methods often rely on assumptions of linearity and uniformity in the distribution of respondents' answers, which may not reflect reality.

Graph 1. Dynamics of Inflation and Quantified Inflation Expectations in Kazakhstan, %



Source: NB RK, BNS ASPR RK

Another method for assessing inflation expectations involves the use of median estimates based on direct quantitative questions in population surveys. The University of Michigan in the United States introduced the method of directly surveying consumers about expected inflation in the mid-1970s (Curtin, 2005).

This method differs in that respondents are asked direct quantitative questions about expected inflation, eliminating the need to convert qualitative responses into quantitative estimates. Such an approach improves the accuracy and reliability of the collected data, reducing potential distortions associated with the subjective interpretation of questions. Furthermore, the use of median estimates makes results less sensitive to outliers and extreme values, contributing to a more stable analysis of inflation expectations (Bryan and Palmqvist, 2005). This also mitigates the influence of current inflation rates on respondents' expectations, reflecting more objective forecasts for the future.

The methodology for calculating median inflation expectations involves collecting direct quantitative responses from respondents regarding their expected level of inflation over a specific period. Respondents are asked to provide a specific numerical value, expressed as a percentage, which reflects their expectations regarding future price changes. The responses are then grouped into intervals, and the median estimate is calculated by accounting for the frequency of responses within each interval and the width of those intervals. A specific formula, which considers cumulative frequencies, is applied for a more precise determination of the median within a given interval.

$$Me = x_0 + i * \frac{\frac{1}{2} * \sum f_i - S_{Me-1}}{f_{Me}}, \quad (2)$$

where x_0 – the lower bound of the median interval, for which the cumulative frequency of responses exceeds half of the total sum of frequencies, i – the width of the median interval x_0 (the difference between the upper and lower bounds of the interval), f_i – the sum of all frequencies of the intervals, S_{Me-1} – the cumulative frequency of the interval preceding the median interval x_0, f_{Me} – the frequency of the median interval x_0 (the number of observations falling within this interval).

The median value represents the inflation threshold at which exactly half of the respondents report higher price growth rates, while the other half believe inflation was lower than this level. Thus, the median reflects a form of public consensus and divides all respondents into two equal groups by size (Bank of Russia).

Since January 2023, the NBRK has incorporated median estimates of expected inflation intervals in its communications.

At the same time, in addition to median estimates, several central banks also use mean estimates of inflation expectations. Comparing median and mean values reveals that mean estimates are generally more volatile and higher than median values, as means are more susceptible to outliers in the data (D'Acunto et al., 2019). This raises a debate about whether outliers should be excluded from the data, considering them unreliable, or retained to preserve information about respondents who expected significant increases in inflation. In this context, the median may serve as a more accurate measure of central tendency (Curtin, 2005; Armantier et al., 2016).

Another method for assessing inflation expectations is the probabilistic approach, where respondents are asked to distribute probabilities across various inflation ranges. Instead of providing a single number or selecting one interval, respondents express their expectations as a probability distribution over different inflation intervals. For example, a respondent might indicate that there is a 20% probability inflation will be below 2%, a 50% probability it will be between 2% and 4%, and a 30% probability it will exceed 4%. This method allows measurement of not only the mean expected inflation but also the level of uncertainty or confidence respondents have in their forecasts (Manski, 2004). Additionally, it enables the study of the distribution of expectations within the population, which is important for assessing risks and potential economic scenarios (Engelberg et al., 2009).

However, this method has its drawbacks: it requires respondents to operate with probabilities, which can reduce response accuracy or increase refusal rates (Bruine de Bruin et al., 2010). Furthermore, analyzing such data requires more complex statistical tools, complicating its practical implementation (Manski, 2004).

In this study, we focus in greater detail on the calculation of median inflation expectations and the determination of intervals. Different interval widths can reflect data measurement characteristics in different ways. Wider intervals can accommodate more data, leading to fewer overall intervals and thus reducing noise caused by sampling randomness. Conversely, narrower intervals increase the frequency of observed intervals, making histograms more sensitive to sampling noise or insufficient smoothing. As a result, varying the interval width for the same dataset can lead to different data distributions (Tao et al., 2023).

Within the scope of this study, we will analyze how the choice of interval width affects the accuracy of inflation expectation estimates. Previous studies demonstrate that selecting the correct methodology and calculation parameters can significantly improve the quality of estimates, ultimately contributing to more effective analysis of inflation expectations and their incorporation into economic policy development.

3. INTERNATIONAL EXPERIENCE

Currently, the most common method for assessing inflation expectations is the use of median values derived from respondents' answers to quantitative questions. This approach is widely adopted by central banks, as these measures are not tied to actual inflation and are considered more objective. The resulting estimates reflect a numerical range of expected price changes, making them more informative for analysis and decision-making in monetary policy.

Table 1 presents information on methods used to assess inflation expectations in various countries, including sample sizes, published indicators, and the starting dates of time series. Most inflation expectation surveys are conducted on a monthly basis; however, in some countries, they are performed quarterly.

In global practice, the number of respondents varies depending on the socio-economic characteristics of the population, acceptable error margins, methodology, and available resources. In certain countries, the proportion of surveyed individuals relative to the total population may be small, but the sample is carefully designed to ensure representativeness across various socio-demographic parameters. The average sample size is approximately 2,000 respondents, which ensures a high level of representativeness relative to the general population. Median estimates of inflation expectations, supported by an adequate sample size, have become the modern standard in central bank practices worldwide.

Table 1. Methods for Assessing Inflation Expectations by Country

Country		Sample	Population of the Country	Published Indicators	Start Date of the Time Series
ECB	Spain	3000	48 373 340	Median, mean	April 2020
	Italy	3000	58 761 150		
	France	3000	68 170 230		
	Germany	3000	84 482 270		
	Belgium	1000	11 822 590		
	Netherlands	1000	17 879 490		
	Austria	1000	9 132 380		April 2022
	Portugal	1000	10 525 350		
	Greece	1000	10 361 300		
	Ireland	1000	5 262 380		
	Finland	1000	5 584 260		
Canada		2000	40 097 760	Median	Q4 2014
United Kingdom		2000	68 350 000	Median	November 1999
Russia		2000	143 826 130	Median	April 2010
Kazakhstan		1500	19 900 180	Median	January 2016
Turkey		4884	85 326 000	Mean	January 2015
New Zealand		1000	5 223 100	Median, mean	March 1995
Norway		1000	5 519 590	Median, mean	Q3 2002
Japan		4000	124 516 650	Median, mean	Q2 2006
Federal Reserve Bank of New York		1300	334 914 900	Median	June 2013

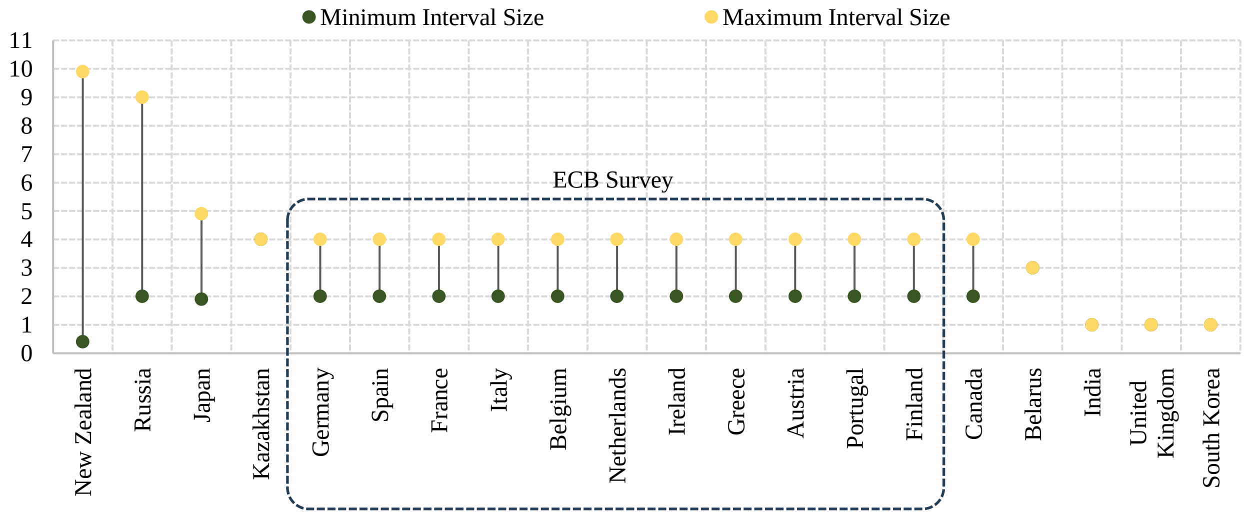
India	5300	1 428 627 660	Median, mean	September 2008
South Korea	2500	51 712 620	Median	February 2002

Source: Official websites of central banks, World Bank
** Population data as of the end of 2023*

The accuracy of median inflation expectation estimates depends on the selected intervals. While some countries use static intervals, the majority adopt dynamic ones (Graph 2). The range of these intervals can vary depending on the volume and precision of the data, and there is no unified global approach to the number and width of intervals.

When dynamic intervals are used, the interval sizes depend on the magnitude of inflation expectations themselves. As the level of inflation expectations increases, the width of the intervals typically expands. This is because higher inflation expectations are accompanied by greater uncertainty and a wider dispersion of opinions among respondents regarding future price changes. Expanding the interval width allows for a more accurate capture of a broader range of expected inflation values, thereby improving the precision of median estimates. Additionally, applying wider intervals when inflation expectations are high helps avoid bias in estimates that may arise from using overly narrow or fixed intervals.

Graph 2. Range of Intervals for Calculating Inflation Expectations by Country, p.p.



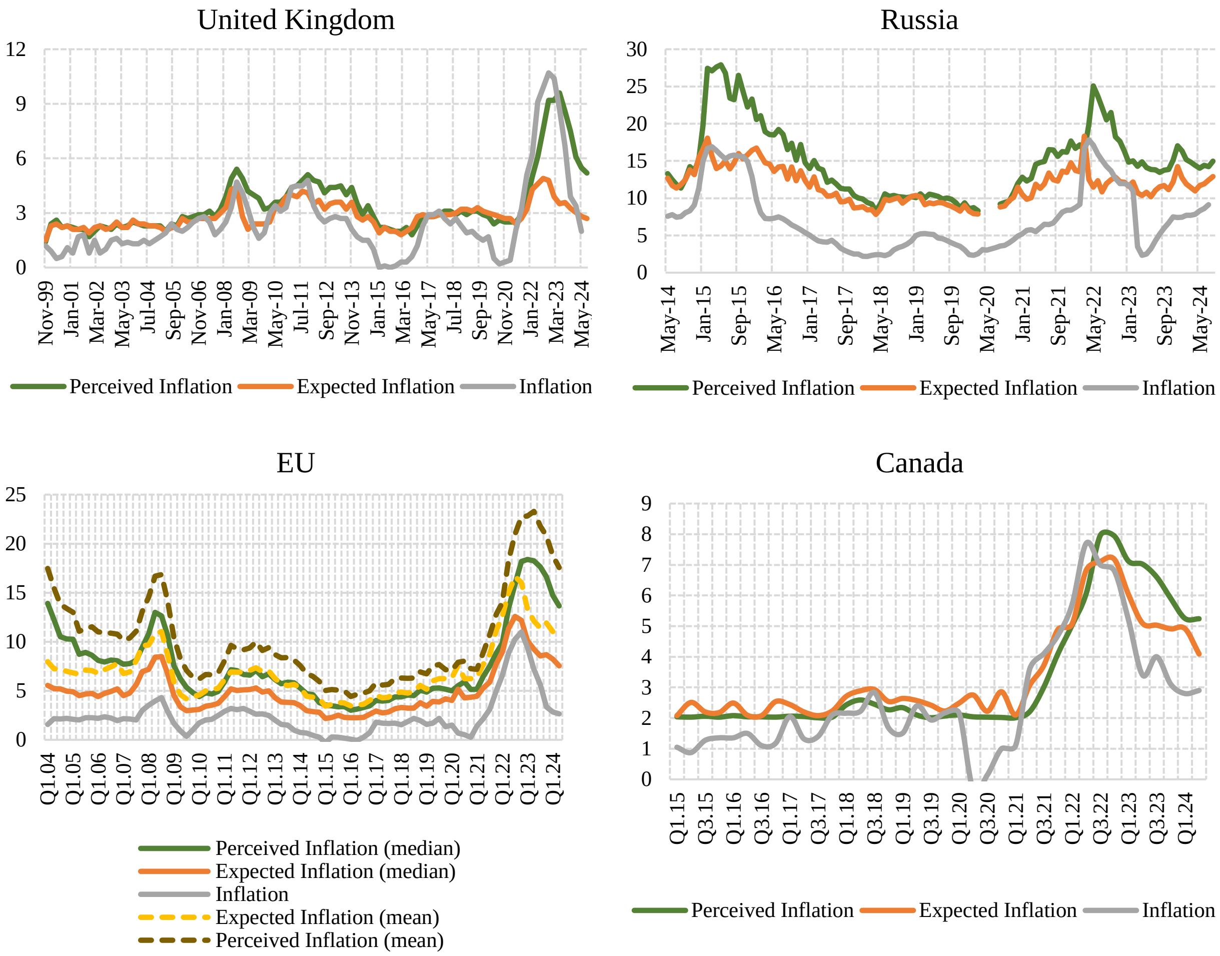
Source: Official websites of central banks

Despite the varying interval widths used by countries to assess inflation expectations, a common trend is that the median and mean numerical values of these expectations often exceed the actual level of inflation (Weber et al., 2022). This phenomenon can be attributed to the fact that perceived and expected inflation indicators are influenced by a multitude of factors, including psychological, economic, and political aspects, price changes in key goods, past experiences of high inflation, the news environment, and other factors (Graph 3).

One explanation for the elevated level of inflation expectations is that households perceive current inflation to be higher than the rate measured by the Consumer Price Index (CPI) and extrapolate this perception into their future expectations (Jonung, 1981). Specifically, this discrepancy can arise because households tend to overestimate their daily expenses (D'Acunto et al., 2018; Cavallo et al., 2017), pay more attention to price increases than to decreases (D'Acunto et al., 2015), and rely on information from external sources—such as mass media, rumors, and social networks—that extend beyond their own consumer experience (Ehrmann et al., 2017).

Thus, understanding the reasons behind elevated inflation expectations is crucial for central banks when designing effective monetary policy. By taking into account the factors influencing the public's perception of inflation, policymakers can enhance communication strategies and implement measures to anchor inflation expectations more effectively.

Graph 3. Inflation Expectations in Selected Countries Worldwide, %



Source: Official websites of central banks

4. METHODOLOGY

When conducting statistical research, including the assessment of inflation expectations among the population, it is important to determine the optimal sample size to ensure representativeness and accuracy of the results. In this case, Cochran's formula (Cochran, 1979) can be used. It is widely applied in statistics to determine the required sample size for estimating population parameters with a specified level of precision and confidence.

$$n = \frac{Z^2 * p * (1-p)}{E^2}, (3)$$

where n – required sample size, Z – estimation for the selected confidence level (e.g., for 95% confidence, z=1.96), p – assumed proportion (commonly set to 0.5 if the exact distribution is unknown), E – margin of error (e.g., E=0.05 for 5%).

In addition to Cochran's formula for determining sample size, this study also employed Slovin's formula (Slovin, 1960) to calculate the sample size, taking into account the acceptable margin of error.

$$n = \frac{N}{1+N*e^2}, (4)$$

where n – the required sample size, N – size of the general population, e – margin of error.

The application of both formulas in this study is driven by the goal of ensuring high accuracy and reliability of the obtained data. Cochran's formula helps determine the necessary sample size for large populations with a specified confidence level and precision, while Slovin's formula adjusts this size by accounting for the finiteness of the population and the acceptable margin of error.

It is important to note that the correct choice of sample size directly affects the quality and reliability of research results. An insufficient sample size can increase statistical error and reduce data representativeness, while an excessively large sample size may be inefficient in terms of cost and time required for data collection and processing. Additionally, using both methods allows for comparison of results to ensure their robustness. This is particularly critical when studying inflation expectations among the population, where data accuracy and reliability are of paramount importance.

Thus, determining the correct sample size is only the first step toward obtaining reliable data. However, for a more in-depth and precise evaluation of inflation expectations, it is equally important to select optimal intervals in which respondents' answers will be grouped.

To date, the method for determining the optimal number and size of intervals has not yet become universally accepted. In the past, many attempts have been made to propose various criteria for assessing the optimal interval sizes necessary for grouping data. One of the earliest works in the literature dedicated to histogram construction was the study by Sturges (1926). In this work, Sturges proposed an empirical rule for selecting the optimal number of intervals (or bins) for a histogram.

$$k = 1 + \log_2 n, (5)$$

where k – the number of intervals for constructing a histogram, n – the number of observations in the sample.

Later, Scott (1979), in his work, proposed a method for determining the optimal bin width for histograms, which takes into account the standard deviation of the data. This approach allows for a more precise selection of interval widths, particularly for large samples or datasets with high variability.

$$h = \frac{3.5 * \sigma}{\sqrt[3]{n}}, (6)$$

where h – the interval width, σ – standard deviation of the data, n – the number of observations in the sample.

Subsequently, this equation was improved by the Freedman-Diaconis rule (Freedman & Diaconis, 1981), which provides a more accurate estimate of interval width by utilizing the interquartile range (IQR). The use of the IQR makes the calculated value less sensitive to outliers in the dataset.

$$h = \frac{2 * IQR}{\sqrt[3]{n}}, (7)$$

where h – the interval width, IQR – the interquartile range (the difference between the first and third quartiles), used to estimate the spread of the data, n – the number of observations in the dataset.

In this study, the Freedman-Diaconis rule is applied to determine interval widths, ensuring an optimal balance between detail and smoothing of the data, which plays a key role in analyzing the distribution of population inflation expectations.

Furthermore, the methodology proposed by Tao et al. (2023) is employed to determine the optimal interval size for constructing histograms of multimodal data. The primary objective of this approach is to minimize errors in histogram construction, ensuring the most accurate representation of the true probability density function for a variable with multiple modes.

The first step of the method involves selecting the initial interval width using the Freedman-Diaconis formula (7). This formula determines the optimal initial interval size, taking into account data variability and sample size. The obtained value h serves as the starting point for testing alternative interval widths.

The intervals must not be too wide, as this leads to excessive generalization of the histogram and may obscure important details in the data. The method involves testing several interval width options to select the one that most accurately reflects the data distribution.

To assess the accuracy of the histogram compared to the true distribution, the sum of squared errors (SSE) is used. This measure determines how well the histogram reproduces the structure of the original data.

$$SSE = \sum_{i=1}^{N_b} w_i (\hat{y}_i - y_i)^2, (8)$$

where N_b – the number of intervals in the histogram, y_i – the frequency or probability for each interval in the actual histogram, $\hat{y}_i$ – frequency or probability for each interval in the analytical histogram, w_i – the weight of the interval (typically equal to 1).

This formula demonstrates how significantly the values in the histogram deviate from what is predicted by the mathematical model. The goal of the method is to minimize this error so that the histogram best represents the actual distribution.

As the number of peaks (modes) increases, the model becomes more complex. In such cases, the degrees of freedom (DOF) are considered, representing the number of independent parameters that influence the final estimate. This helps prevent overcomplicating the model and overfitting, while maintaining a balance between the model's accuracy and its generalization capability.

$$DOF = N_b - 3K - 1, \quad (9)$$

where N_b – the number of intervals in the histogram, K – the number of modes (peaks) in the distribution.

This formula accounts for the fact that each mode (peak) requires three parameters: A_j (proportion), μ_j (mean), and σ_j (standard deviation). The more modes there are, the fewer degrees of freedom remain for constructing the model, as each additional peak increases the model's complexity.

For each interval, the standard error (SE) is calculated as:

$$S_E = \sqrt{\frac{SSE}{DOF}}, \quad (10)$$

where SSE – sum of squared errors, DOF - degrees of freedom.

Since errors can vary significantly between different intervals, they need to be normalized. This normalization makes all errors comparable, allowing for their correct comparison.

$$S_{EN} = \frac{|S_E - \mu_S|}{\sigma_S}, \quad (11)$$

where S_{EN} – normalized standard error, S_E – standard error for a specific interval, μ_s – mean standard error across all intervals, σ_s – standard deviation of the standard error across all intervals

To select the optimal interval, the Bin Size Index (BSI) is used:

$$BSI = \frac{|2 * \ln(S_{EN})|}{K}, \quad (12)$$

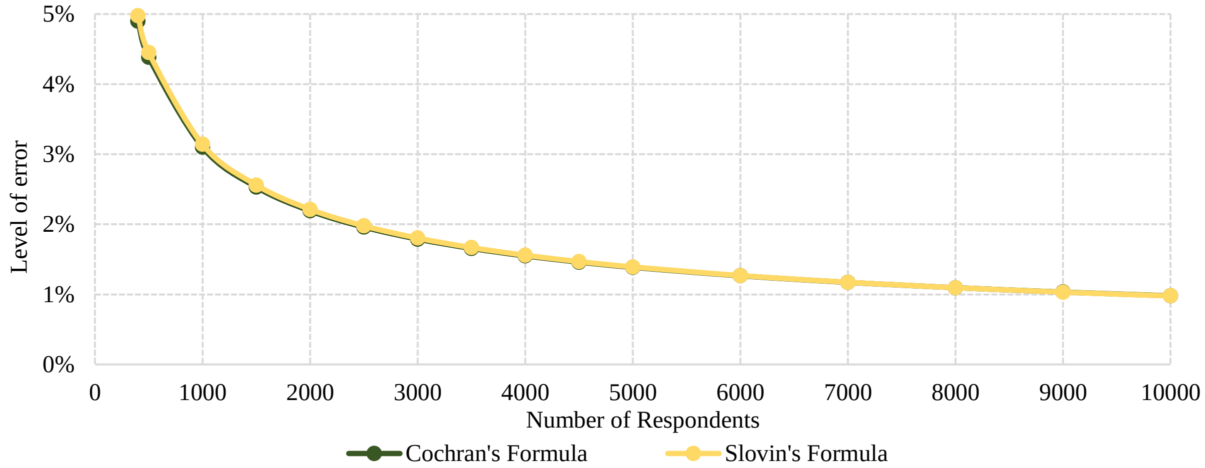
where $\ln(S_{EN})$ – logarithm of the normalized standard error, K – number of modes (peaks) in the distribution

The determination of the optimal interval width is based on selecting the maximum value of the BSI index.

5. RESULTS

According to the results of calculations using Cochran's formula (3) and Slovin's formula (4), a sample size of 1,500 people ensures a margin of error of approximately 2.5%. In statistical research, a margin of error below 5% is considered acceptable and provides sufficient accuracy of results. This margin of error often corresponds to a significance level of $\alpha = 0.05$, which indicates a 5% probability that the observed difference from the population is due to random chance (Graph 4).

Graph 4. Error Level as a Function of the Number of Respondents



Source: Authors' calculations

The minimum sample size required to ensure a 5% margin of error is approximately 400 respondents. Increasing the sample size to 5,000 respondents reduces the margin of error to 1.4%, and further increasing it to 10,000 respondents lowers the error to 1%. In this context, a sample size of 1,500 respondents appears optimal, as it provides an acceptable accuracy level with a margin of error of 2.5%, without requiring significant expenses for conducting surveys. This sample size strikes a balance between accuracy and efficiency while covering various socio-demographic groups of the population.

An increase in the sample size beyond this level does reduce the margin of error further; however, the rate of this reduction becomes insignificant, making it hard to justify additional costs. Therefore, a sample size of 1,500 respondents is a rational choice, supported by both theoretical calculations and practical considerations.

Continuing with the analysis, it is important to note that to calculate the intervals for median inflation expectations, the dataset was divided into three periods. The first period covers the time from February 2022 to December 2023, when inflation expectations were relatively high. The second period spans from

January 2020 to December 2021, which is characterized by moderate inflation expectations.

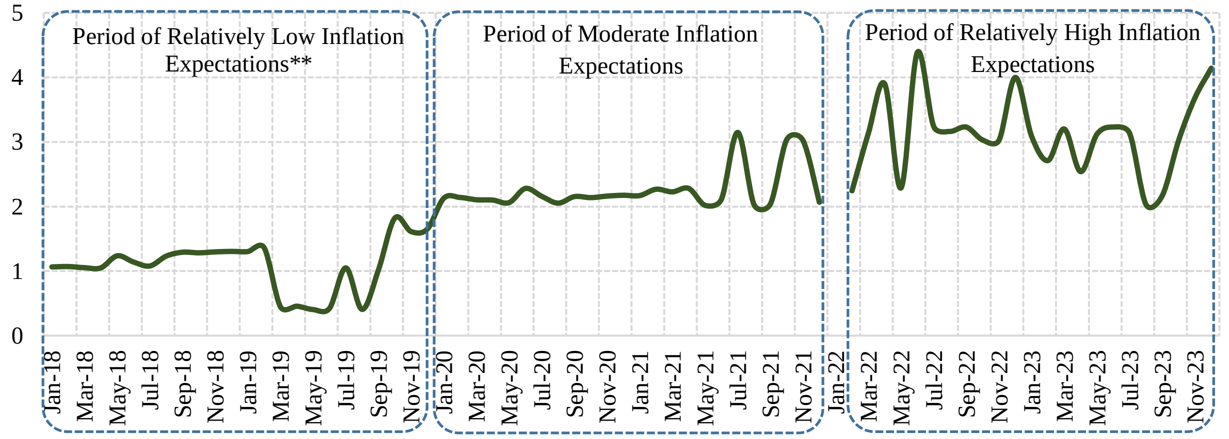
Considering that inflation expectations in Kazakhstan did not drop below 8.8%, an additional time period from January 2018 to December 2019 was generated, where the dynamics of inflation expectations were modeled at a level close to the 5% target inflation rate. This generated period is necessary to verify the accuracy of the calculations based on the Freedman-Diaconis formula (7) and the BSI index (12).

It is assumed that as the magnitude of inflation expectations changes, the width of intervals calculated by both methods will adjust accordingly, reflecting the necessity to adapt intervals to changing economic conditions.

This adaptive approach aligns with statistical principles, which state that the width of intervals should reflect the characteristics of the data distribution. For example, when the variance and variability of the data increase, the use of wider intervals helps to avoid excessive detail and "noise" while still providing an accurate representation of the overall trend.

The results of calculations using the Freedman-Diaconis formula (7) are presented in graph 5. The analysis confirms that as the magnitude of inflation expectations increases, the optimal interval width expands. During periods of relatively low inflation expectations, the interval width is approximately 1 percentage point (p.p.), which allows for the precise reflection of small fluctuations in the data and ensures high accuracy of median estimates. During periods of moderate inflation expectations, the range expands to 2–3 p.p., while during periods of relatively high inflation expectations, the interval width increases to 3–4 p.p. The increase in the width of optimal intervals as inflation expectations rise is associated with greater variability in respondents' answers and the need to capture a broader range of values.

Graph 5. Calculated Interval Widths, p.p.



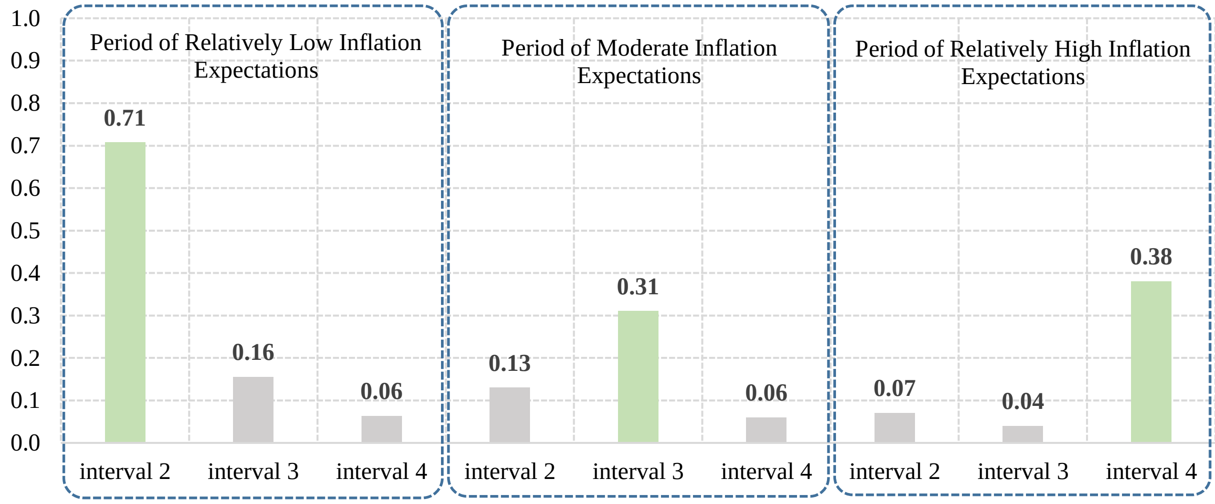
Source: Authors' calculations

** The estimates for January 2022 were not published as the survey results were influenced by the events in the country during January, which led to a smaller sample size and a lack of comparability with previous data.*

*** The data from January 2018 to December 2019 were generated for the analysis of low inflation expectations.*

Next, to obtain an alternative estimate of the optimal interval width, the BSI index calculation methodology was applied. Interval sizes of 2, 3, and 4 were selected for analysis (Graph 6).

Graph 6. Calculated BSI Index for Different Intervals



Source: Authors' calculations

The calculations, conducted similarly to the work of Tao et al. (2023) using specialized software Origin, indicate that during the period of relatively low inflation expectations, the maximum BSI index value (0.71) is observed at an interval of 2 percentage points (p.p.), confirming its optimality for accurately representing the data. In the second period, characterized by moderate inflation expectations, the maximum BSI value (0.31) is achieved at an interval of 3 p.p.. In the third period, when inflation expectations were relatively high, the maximum BSI value (0.38) corresponds to an interval of 4 p.p., indicating the preference for wider intervals at higher levels of inflation expectations.

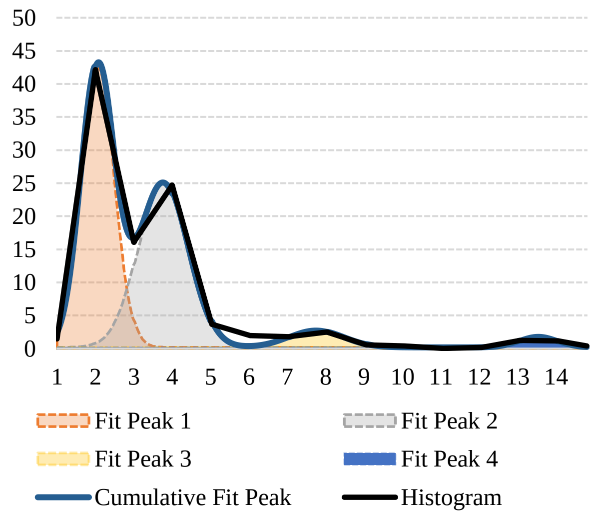
Similar to the results obtained using the Freedman-Diaconis formula (7), the BSI index confirms that as inflation expectations increase, the interval width should also increase. The results highlight the importance of adapting interval widths based on the level of inflation expectations to ensure high analytical accuracy.

For periods of moderate and relatively low inflation expectations, the Freedman-Diaconis formula (7) and the BSI method (12) produce similar results. However, under conditions of relatively low inflation expectations, the BSI index determines a wider interval size. In the study by Tao et al. (2023), it is noted that values calculated using the Freedman-Diaconis formula typically align with the maximum value for the BSI index. At the same time, the use of wider intervals in the BSI method reflects a preference for smoothing the distribution and highlighting key trends, which reduces the impact of random fluctuations.

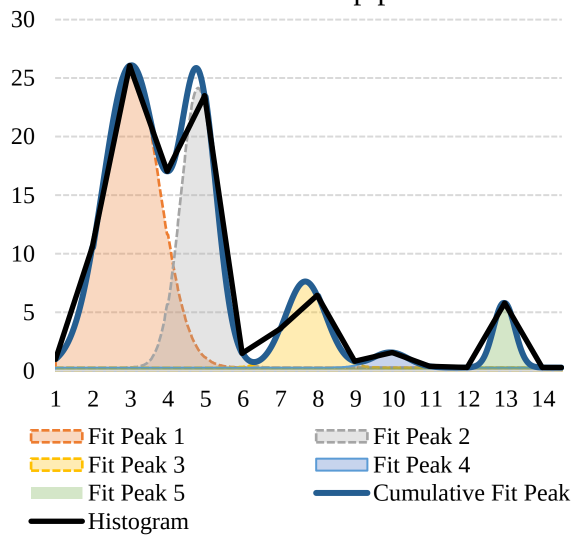
This approach is justified in situations where the data are characterized by high volatility or significant dispersion. In such cases, excessive detail may amplify noise levels, complicating the interpretation of results. In the context of inflation expectations, identifying the main trends while ignoring minor fluctuations appears more appropriate. Additionally, it should be noted that mathematical models, such as the Freedman-Diaconis method, often tend to overestimate the number of intervals, leading to the formation of narrower ranges (Sahann et al., 2023).

Next, graphs 7-9 present histograms of grouped data for the three periods using the optimal intervals. Each of these histograms was fitted with a multimodal distribution function (curve – Cumulative Fit Peak) with the corresponding number of peaks (Fit Peak). This, in turn, made it possible to determine the error between the analytical and experimental histograms.

Graph 7. Deconvolution Distribution of Data with an Interval Width of 2 p.p.



Graph 8. Deconvolution Distribution of Data with an Interval Width of 3 p.p.



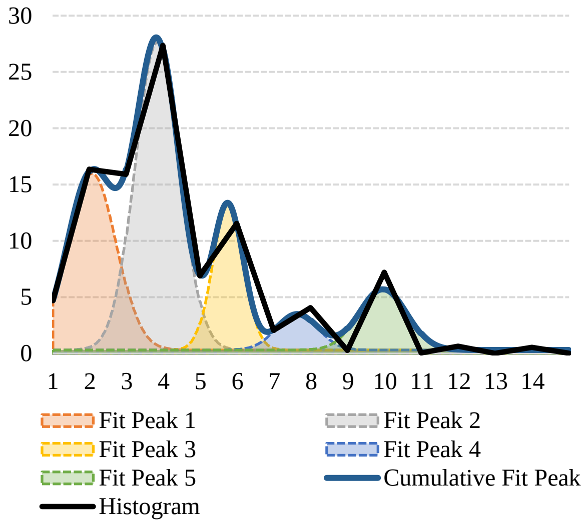
Source: Authors' calculations

The X-axis represents data intervals grouped into specific ranges, such as values of inflation expectations expressed in percentages. The Y-axis shows the number of observations or the frequency of data falling into each interval, allowing for a visual assessment of the data distribution and identification of key peaks.

Graph 7 presents the analysis for low inflation expectations, where narrow intervals of 2 percentage points (p.p.) were used. This approach reveals three distinct peaks (Fit Peak 1–3) and ensures high sensitivity to small fluctuations. The cumulative peak (Cumulative Fit Peak) combines all components, providing a comprehensive representation of the data distribution. Narrow intervals are particularly useful for analyzing low-volatility data, where precision and detail are essential.

Graph 8 displays the distribution using intervals with a width of 3 p.p.. This approach identifies five distinct peaks (Fit Peak 1–5), enabling a more detailed analysis of data characterized by moderate inflation expectations. These intervals maintain a balance between detail and smoothing, capturing small deviations while minimizing the impact of random fluctuations. This level of detail is especially useful for analyzing moderate inflation expectations.

Graph 9. Deconvolution Distribution of Data with an Interval Width of 4 p.p.



Source: Authors' calculations

Graph 9 demonstrates the use of intervals with a width of 4 percentage points (p.p.), which allows for smoothing the distribution and highlighting key trends. Despite the presence of the same five peaks as in graph 8, their contours become broader and smoother. This approach reduces the impact of random fluctuations and focuses on the main trends, which is particularly important under conditions of high inflation expectations and uncertainty.

As a result, the conducted analysis demonstrated that the optimal interval width for the data used is dynamic. During periods of low inflation expectations, it is preferable to use narrower intervals, while for high expectations, wider intervals are more appropriate.

Currently, a static interval width of 4 percentage points (p.p.) is used to calculate inflation expectations, specifically the following ranges: 1–5%, 6–10%, 11–15%, 16–20%, and over 20%. However, based on the conducted analysis, the most optimal approach appears to be the use of dynamic intervals with widths ranging from 2 to 4 p.p. Accordingly, the optimal intervals can be structured as follows: 1–3%, 4–6%, 7–10%, 11–15%, 16–20%, and over 20%. This approach will provide a more accurate representation of the distribution of inflation expectations, adapting the intervals to changing economic conditions and reducing the impact of statistical noise.

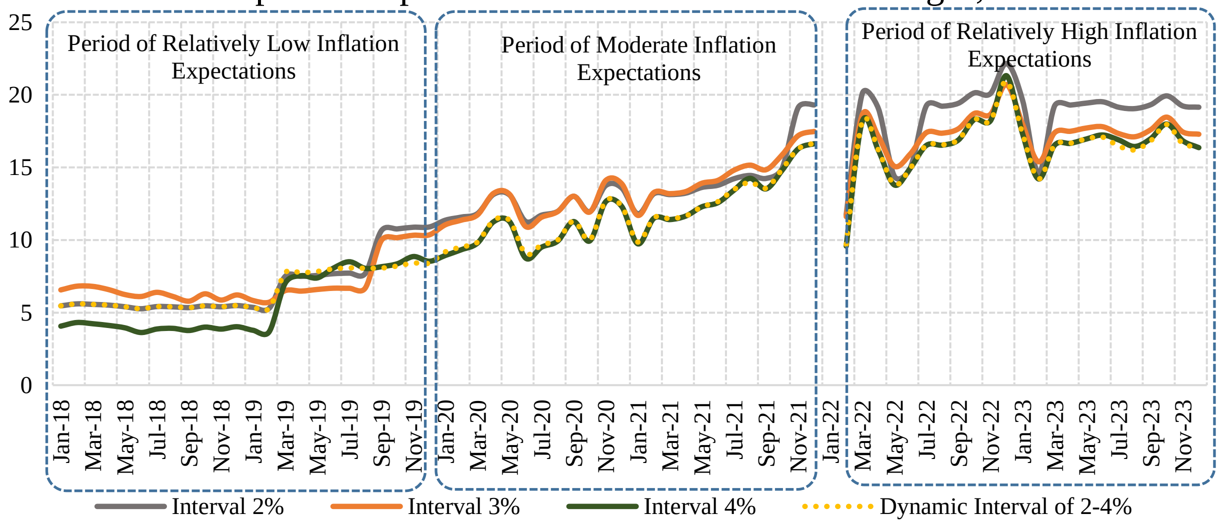
During periods of low inflation expectations, it is advisable to use intervals with a width of 2 percentage points for the ranges 1–3% and 4–6%. As the magnitude of inflation expectations increases, the interval width expands to 3 percentage points, with the interval becoming 7–10%. This division appears logical because the new intervals, unlike the previous ones, conditionally separate inflation expectations into expectations below the target inflation level (1–3%), expectations around the target level (4–6%), and expectations above the target level (7–10%). The interval 7–10% also captures expectations that are multiples of 5% and 10%, which is important since respondents often round their answers, and this interval effectively takes such values into account.

For subsequent intervals of inflation expectations (above 10%), considering the increase in inflation expectations themselves, the optimal width is 4 percentage points. Thus, the intervals are as follows: 11–15% and 16–20%. These intervals are similar to the existing ones currently used for calculating the median inflation expectations by the NBRK. Overall, the use of dynamic intervals should not lead to a significant revision of inflation expectations, as since 2016, most of them have remained above 10%. In these ranges, the new dynamic intervals coincide with the static ones and amount to 4 percentage points.

However, as our calculations have shown, during periods of low inflation expectations, it is necessary to use narrower intervals. In the future, as inflation expectations decrease, the use of dynamic narrow intervals for low inflation expectation values will provide more accurate and stable results with less volatility compared to a static interval width of 4 percentage points.

Graph 10 illustrates a comparison of inflation expectations calculated using intervals of various widths: 2 p.p., 3 p.p., 4 p.p., as well as a dynamic interval of 2–4 p.p.

Graph 10. Expected Inflation with Different Ranges, %



Source: Authors' calculations

The use of intervals of 2 p.p., 3 p.p., and 4 p.p. shows that, while changing the width, the results remain relatively similar but with slight differences in smoothness and sensitivity to sharp changes. The dynamic interval of 2–4 p.p. adapts to volatility, demonstrating high flexibility under rising inflation expectations, while during periods of low inflation expectations, the dynamic interval of 2–4 p.p. exhibits lower volatility. This happens because, similar to wide intervals, it is less susceptible to the influence of minor deviations. However, unlike the interval width of 4 p.p., the dynamic interval does not underestimate inflation expectation estimates, maintaining a more balanced picture. Such adaptability is particularly important for accurate analysis in conditions of stable and low inflation expectations, where it is necessary to account for small fluctuations without introducing significant distortions.

6. CONCLUSION

In this study, the optimal sample size and interval width for calculating the median inflation expectations based on survey results of Kazakhstan's population were determined. According to Cochran's formula (3) and Slovin's formula (4), a sample size of 1,500 respondents ensures a margin of error of approximately 2.5%. In statistical research, a margin of error below 5% is considered acceptable and provides sufficient accuracy of results. Such a margin of error means there is a 5% probability that the observed difference from the population is random.

In this context, a sample size of 1,500 respondents appears optimal, as it ensures acceptable accuracy with a 2.5% margin of error without requiring significant costs for conducting surveys. While increasing the sample size beyond this level does reduce the margin of error further, the rate of reduction becomes insignificant, making it hard to justify the additional expenses.

Regarding the determination of the optimal interval width, according to the Freedman-Diaconis formula (7) and the BSI methodology (12), the optimal interval width varies depending on the magnitude of inflation expectations. The analysis demonstrated that the optimal interval width for the data used is dynamic. Thus, during periods of low inflation expectations, it is preferable to use narrower intervals, whereas during periods of high expectations, wider intervals are more appropriate.

When inflation expectations are low, the distribution of respondents' answers tends to be more concentrated, and data variability is lower. Using narrower intervals allows for capturing small differences in expectations more accurately and provides a detailed representation of the data distribution. Wide intervals under conditions of low variability may lead to excessive smoothing, which can obscure important features and trends within the data. Narrow intervals help maintain detail and increase the sensitivity of the analysis to small changes in inflation expectations.

When inflation expectations are high, respondents' answers become more dispersed, and data variability increases. Using wider intervals helps group the data into larger categories, which reduces the impact of random fluctuations and statistical noise. Wide intervals under high variability provide more reliable statistical estimates because each interval contains more observations. This improves the stability of median and mean values and reduces the volatility of estimates.

Currently, the National Bank uses a static interval width of 4 percentage points (p.p.), specifically the ranges 1–5%, 6–10%, 11–15%, 16–20%, and over 20%. However, based on the conducted analysis, the most optimal approach appears to be the use of dynamic intervals with widths ranging from 2 to 4 p.p. Thus, the results of the study indicate that the following dynamic ranges are more preferable: 1–3%, 4–6%, 7–10%, 11–15%, 16–20%, and over 20%.

The findings of this study provide practical recommendations for improving the methodology for calculating median inflation expectations in Kazakhstan. The application of the proposed dynamic ranges will contribute to a more accurate assessment of inflation expectations.

REFERENCES

1. Armantier, O., Topa, G., Van der Klaauw, W. and Zafar, B., 2017. The Survey of Consumer Expectations. *Federal Reserve Bank of New York Economic Policy Review*, 23(2), pp.51–72.
2. Berk, J.M., 1999. Measuring inflation expectations: A survey data approach. *Applied Economics*, 31(11), pp.1467–1480.
3. Bernanke, B.S., Laubach, T., Mishkin, F.S. and Posen, A.S., 1999. *Inflation Targeting: Lessons from the International Experience*. Princeton: Princeton University Press.
4. Bryan, M.F. and Palmqvist, S., 2005. Testing near-rationality using detailed survey data. *European Economic Review*, 49(2), pp.577–598.
5. Cavallo, A., Cruces, G. and Perez-Truglia, R., 2017. Inflation Expectations, Learning, and Supermarket Prices: Evidence from Survey Experiments. *American Economic Journal: Macroeconomics*, 9(3), pp.1–35.
6. Cochran, W.G., 1977. *Sampling Techniques*. 3rd ed. New York: John Wiley & Sons.
7. Curtin, R.T., 1996. *Procedure to Estimate Price Expectations*. Ann Arbor, MI: University of Michigan Survey Research Center.
8. Curtin, R.T., 2005. Inflation Expectations: Theoretical Models and Empirical Tests. Paper presented at the National Bank of Poland, Warsaw, December 2005.
9. D'Acunto, F., Hoang, D. and Weber, M., 2015. Inflation Expectations and Consumption Expenditure. *University of Chicago Booth School of Business Working Paper*.
10. D'Acunto, F., Malmendier, U., Ospina, J. and Weber, M., 2019. Exposure to Daily Price Changes and Inflation Expectations. *NBER Working Paper No. 26237*.
11. D'Acunto, F., Malmendier, U. and Weber, M., 2018. What Do the Data Tell Us About Inflation Expectations? *AEA Papers and Proceedings*, 108, pp.562–566.
12. Ehrmann, M., Pfajfar, D. and Santoro, E., 2017. Consumers' Attitudes and Their Inflation Expectations. *International Journal of Central Banking*, 13(1), pp.225–259.
13. Engelberg, J., Manski, C.F. and Williams, J., 2009. Comparing the Point Predictions and Subjective Probability Distributions of Professional Forecasters. *Journal of Business & Economic Statistics*, 27(1), pp.30–41.
14. Freedman, D. and Diaconis, P., 1981. On the histogram as a density estimator: L2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57(4), pp.453–476.

15. Friedman, M., 1968. The role of monetary policy. *American Economic Review*, 58(1), pp.1–17.
16. IMF, 2023. *World Economic Outlook: Navigating Global Divergences*.
17. Jonung, L., 1981. Perceived and Expected Rates of Inflation in Sweden. *American Economic Review*, 71(5), pp.961–968.
18. Manski, C.F., 2004. Measuring Expectations. *Econometrica*, 72(5), pp.1329–1376.
19. Mishkin, F.S., 2007. Inflation Dynamics. *International Finance*, 10(3), pp.317–334.
20. Muth, J.F., 1961. Rational expectations and the theory of price movements. *Econometrica*, 29(3), pp.315–335.
21. Nardo, M., 2003. The Quantification of Qualitative Survey Data: A Critical Assessment. *Journal of Economic Surveys*, 17(5), pp.645–668.
22. Nathanael, V. and Wiederholt, M., 2019. Inflation Expectations and Choices of Households. *SAFE Working Paper Series 250*. Leibniz Institute for Financial Research SAFE.
23. Orphanides, A. and Williams, J.C., 2005. Inflation scares and forecast-based monetary policy. *Review of Economic Dynamics*, 8(2), pp.498–527.
24. Sahann, R., Möller, T., and Schmidt, J., 2023. Histogram binning revisited with a focus on human perception. *Scientific Reports.*
25. Scott, D.W., 1979. On optimal and data-based histograms. *Biometrika*, 66(3), pp.605-610.
26. Slovin, E., 1960. Slovin’s formula for sampling technique.
27. Smith, J. and McAleer, M., 1995. Alternative procedures for converting qualitative response data to quantitative expectations: An application to Australian manufacturing. *Journal of Applied Econometrics*, 10(2), pp.165–185.
28. Sturges, H.A., 1926. The choice of a class interval. *Journal of the American Statistical Association*, 21(153), pp.65-66.
29. Svensson, L.E.O., 1997. Inflation Forecast Targeting: Implementing and Monitoring Inflation Targets. *European Economic Review*, 41(6), pp.1111–1146.
30. Tao, J., Yucheng, L., Yongkang W., 2023. A new bin size index method for statistical analysis of multimodal datasets from materials characterization. *Scientific Reports*.
31. Weber, M., D'Acunto, F., Gorodnichenko, Y. and Coibion, O., 2022. The Subjective Inflation Expectations of Households and Firms: Measurement, Determinants, and Implications. *NBER Working Paper No. 30034*.

32. Woodford, M., 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton: Princeton University Press.
33. Bank of Russia, *The methodology of the study of inflation expectations and consumer sentiment of the population.*