

# Project 1

## Iran's Household Income Expenditure Survey

Fatemeh Abbasian-Abyaneh

March 4, 2024

### Introduction

The Household Income Expenditure Survey (HIES) is a comprehensive survey conducted to collect information on the living standards, income distribution, and consumption patterns of households. The survey encompasses a wide array of socio-economic indicators that provide insights into the well-being of the population and the economic conditions of the country.

In the analysis of the HIES data for Iran, we focus on the period spanning from the year 1398 to 1401. This period, characterized by socio-economic fluctuations and policy changes, offers a fertile ground for examining trends in various facets of household economics, including workforce participation, consumption, housing conditions, and inequality measures.

### Data Preparation

Data preparation is a critical step in ensuring the accuracy and reliability of our analysis. The process began with the extraction of the HIES data from a comprehensive database using the Open Database Connectivity (ODBC) interface. Following the retrieval of the data, the datasets were meticulously organized and cleaned to facilitate the analysis.

Demographic information, including urban (UR=1) and rural (UR=0) household designation and the year of the survey, was generated for each dataset to distinguish between different segments of the population. Separate datasets for rural and urban areas across the four years of interest were created and saved, ensuring that the urban-rural divide was adequately represented in the analysis.

A consistent naming convention was applied to variables across datasets to maintain uniformity. For example, household addresses were labeled as *Address*, and weights were designated as *W*. Descriptive labels were also added to variables to enhance the interpretability of the data. For instance, *Weight* was labeled "Household Weight," indicating the number of households represented by the survey data.

Key economic indicators such as total expenditure, clothing expenditure, and rent expenditure were extracted and adjusted for inflation where necessary. This was done by generating new variables that encapsulate the total expenditure on various commodities per household size, thus normalizing the data to allow for equitable comparisons.

Subsequently, the datasets were collapsed to summarize the total expenditures across all households, delineated by year and urban-rural classification. This provided a foundational dataset upon which further analysis, such as quartile and decile expenditure trends, could be built.

In closing, the data preparation phase was conducted with rigorous attention to detail and methodical organization to ensure that subsequent analyses could be performed with the highest level of precision and confidence in the integrity of the data.

# 1 Women's Work Force participation

## 1.1 Question

Identify the trend of changes in women's participation rate by separating the whole country, urban, and rural areas.

### 1.1.1 Data

The analysis utilizes data from the Combined Family Characteristics (CombinedP1FamChar.dta) dataset, which encompasses demographic details, educational levels, and employment status across provinces for the period from 1398 to 1401. The dataset includes both urban and rural populations, allowing for a comprehensive analysis of women's participation in the workforce.

### 1.1.2 Methodology

The methodology involves creating binary variables to determine the gender and employment status of individuals. The data is filtered to focus on females aged 15 and above. Employment status is categorized into employed, unemployed, and active (either employed or actively seeking work). The analysis calculates the participation rate, defined as the proportion of active women to the total weighted number of women, both overall and segmented by urban and rural residency.

### 1.1.3 Visualization

The trends in women's work force participation rates are visualized using line graphs that distinguish between the total country, urban, and rural areas. The graphs depict changes in participation rates over the four-year period, with separate lines for each category. Check Figure 1.

### 1.1.4 Interpretation

The provided graph, "Female Participation Rate Trends 1398-1401 by Area," illustrates the shifts in the female workforce participation rate across different regions of Iran within a span of four years. The delineation by area — Total Country, Urban, and Rural — allows for a comparative analysis of the varying dynamics in women's workforce engagement.

- **Total Country (Black Line):** Nationally, the participation rate exhibits minor fluctuations. It begins with a marginal decline from 1398 to 1399, stabilizes in the following year, and experiences a slight increase by 1401. This pattern suggests a relatively stable national trend in female workforce participation.
- **Urban (Green Line):** Urban areas follow the national trend with a decline from 1398 to 1399. Nonetheless, there is a notable rebound from 1399 to 1400, with rates surpassing those of 1398, although there is a slight drop observed in 1401. This recovery hints at either the transient nature of the factors affecting urban female participation or the effectiveness of targeted interventions.

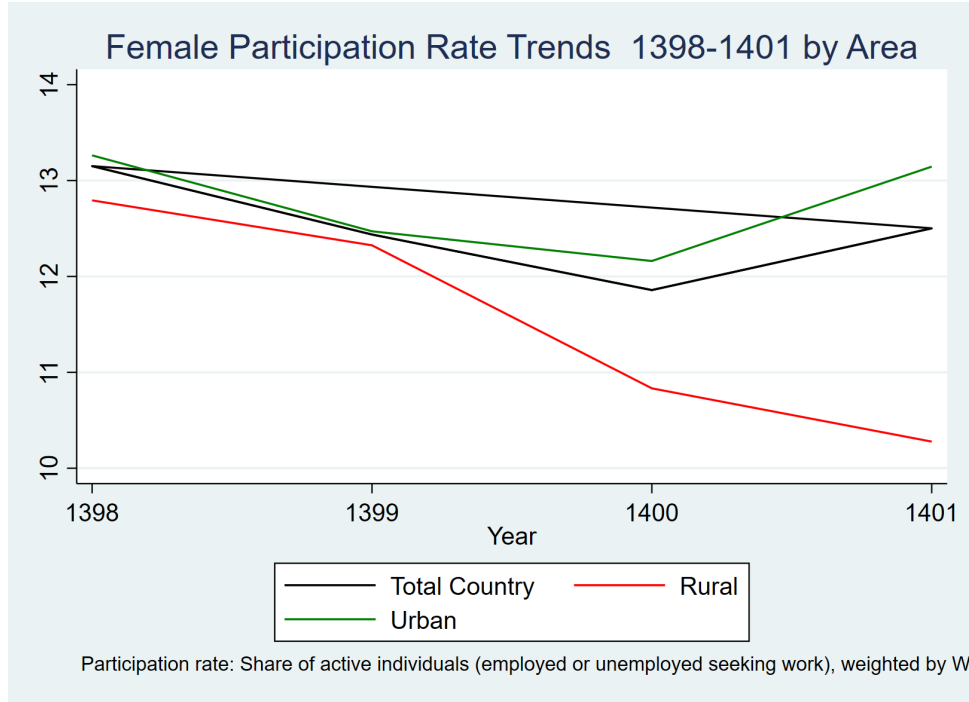


Figure 1: Female Participation Rate Trends 1398-1401 by Area

- **Rural (Red Line):** Contrary to urban trends, the rural areas show a consistent and alarming decline over the four years. This persistent downtrend points to ongoing challenges that are disproportionately impacting rural women's ability to enter or sustain employment.

## 1.2 Question

Identify the trend of changes in married women's participation rate across education levels.

### 1.2.1 Data

The dataset titled "Q1.dta" contains individual characteristics including marital status, gender, education level, and workforce activity status. Analysis in this section is confined to married women to scrutinize their participation in the workforce in relation to their education levels.

### 1.2.2 Methodology

The methodology includes filtering the dataset to retain only records of married females. The educational attainment of these women is categorized from primary education to postgraduate degrees. Workforce participation rates are then computed for each education level, weighted by the individual household weights to ensure representativeness. These rates are calculated annually to observe temporal trends.

### 1.2.3 Visualization

The workforce participation rates by education level are visualized through line and bar charts, displaying the temporal trends from 1398 to 1401. The line chart delineates the trends for each education level, while the bar chart provides a comparative static snapshot, allowing for easy comparison between years and education levels.( Figure 2 & Figure 3)

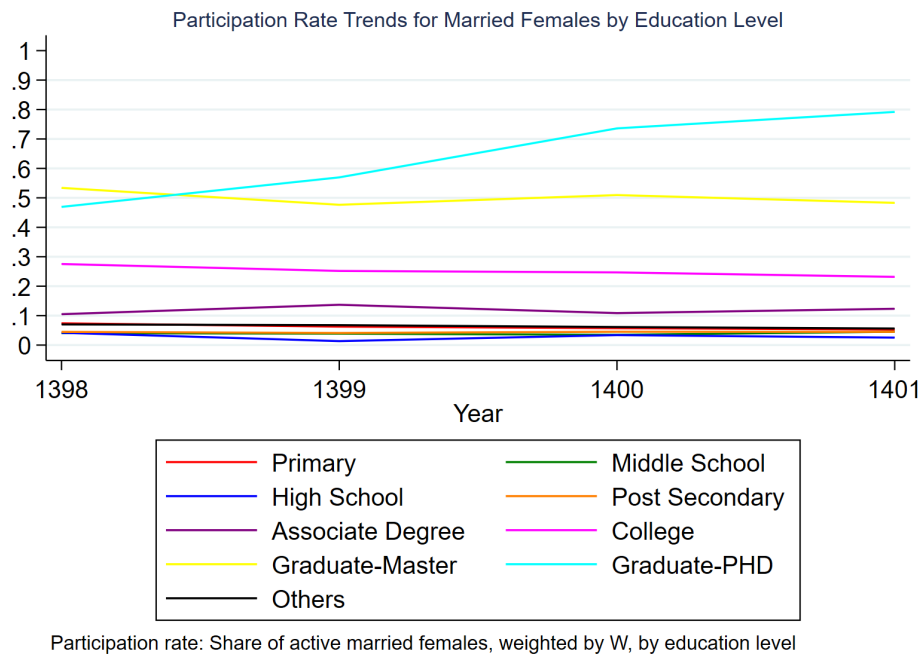


Figure 2: Line chart of Participation Rate Trends for Married Females by Education Level.

### 1.2.4 Interpretation

The bar chart titled "Participation Rate Trends for Married Females by Education Level" displays the participation rates of married women in the workforce across different education levels over a span from 1398 to 1401. Each color represents a distinct education level, ranging from Primary to Graduate-PHD.

- In 1398, the participation rates across all education levels were relatively low, with the highest rates observed in the College (Post Secondary) and Graduate-Master categories.
- A significant increase is noticeable in 1399 for those with a College education, indicating a possible shift in the economic or social factors influencing the participation of higher-educated married women.
- The year 1400 shows a further increase in the participation rate, particularly for those with an Associate Degree, College education, and Graduate-Master level, suggesting a trend where higher education is potentially correlating with higher workforce engagement.

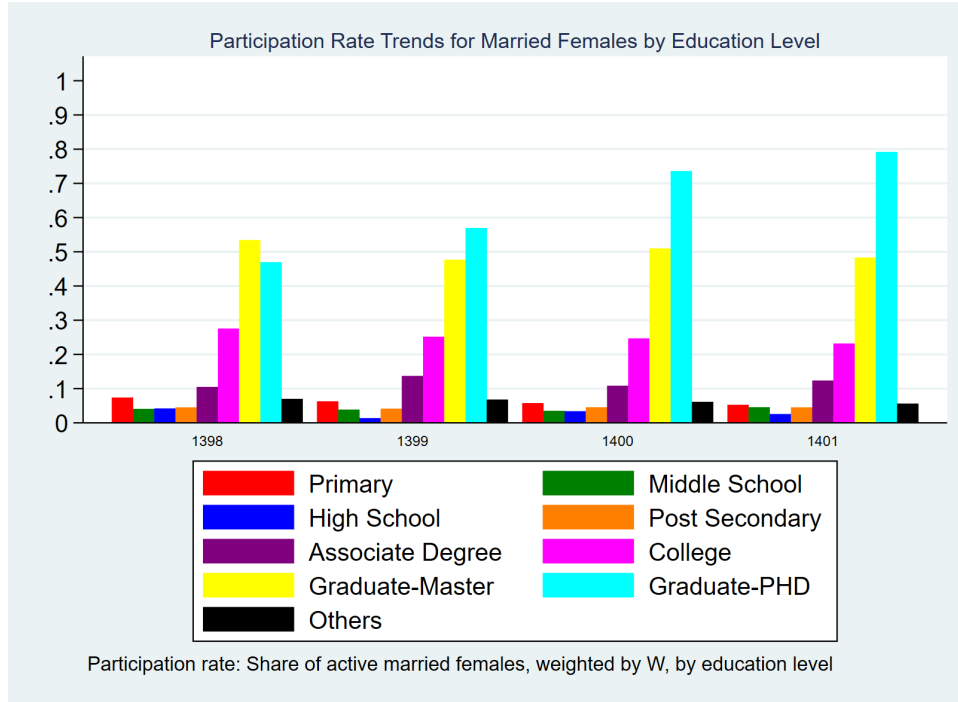


Figure 3: Bar chart of Participation Rate Trends for Married Females by Education Level.

- By 1401, there is a prominent rise in participation rates for women with College and Graduate-Master education, while the rates for other education levels remain stable or experience only minor increases.

The overall trend suggests that married females with higher education levels, particularly those with College and Graduate-Master degrees, are increasingly participating in the workforce. This trend could be indicative of a growing recognition of the value of higher education in improving employment prospects for married women or a reflection of broader social changes that support the active economic engagement of highly educated married women.

### 1.3 Question

Identify the trend of changes in women's participation thread across provinces

#### 1.3.1 Data

Data for this analysis is derived from "Q1.dta," which comprises individual characteristics, including gender, activity status, and location by province. This section concentrates on the female population's engagement in the workforce, with a specific lens on geographic distribution.

#### 1.3.2 Methodology

We compute the weighted participation rate for each province annually. The participation rate is defined as the proportion of active women in the workforce relative to the total female population, adjusted by the household weight to accurately reflect the provincial

representation. This rate is then collated to form a provincial perspective of women’s participation trends over the years.

### 1.3.3 Visualization

The analysis culminates in a table that juxtaposes the participation rates across different provinces from 1398 to 1401. This tabular representation provides a clear visual comparison of the participation rates, highlighting trends and disparities among provinces. (Table 1)

Province	Year (%)			
	1398	1399	1400	1401
Markazi	8%	6%	7%	8%
Gilan	13%	12%	10%	10%
Mazandaran	13%	14%	15%	12%
Azarbajjan-E-Sharghi	13%	12%	12%	10%
Azarbajjan-E-Gharbi	10%	11%	8%	9%
Kermanshah	15%	15%	17%	15%
Khouzestan	11%	10%	11%	10%
Fars	11%	8%	8%	13%
Kerman	11%	12%	11%	10%
Khorasan-E-Razavi	13%	12%	10%	10%
Isfahan	12%	11%	10%	12%
Sistan-va-Balouchestan	10%	11%	11%	12%
Kordestan	13%	12%	10%	10%
Hamedan	16%	15%	12%	15%
Bakhtyari	15%	14%	14%	14%
Lorestan	16%	17%	13%	12%
Ilam	16%	15%	16%	12%
Kohkilouye	16%	18%	19%	17%
Boushehr	15%	13%	12%	14%
Zanjan	21%	17%	17%	21%
Semnan	12%	13%	11%	13%
Yazd	15%	11%	11%	13%
Hormozgan	10%	11%	10%	10%
Tehran	15%	15%	14%	16%
Ardabil	18%	14%	14%	14%
Qom	11%	7%	9%	11%
Qazvin	13%	11%	11%	12%
Golestan	17%	16%	14%	14%
Khorasan-E-Shomali	21%	16%	15%	13%
Khorasan-E-Jonoubi	12%	10%	11%	11%
Alborz	12%	13%	12%	14%

Table 1: Percent of female participation in the workforce in Iran by provinces for the years 1398 to 1401.

### 1.3.4 Interpretation

The graph titled "Female Participation Rate Trends 1398-1401 by Area" demonstrates a comparative analysis of female workforce participation rates across urban and rural areas, alongside the total country figures, from the year 1398 to 1401. Urban areas consistently exhibit higher rates of participation, possibly reflecting the abundance of employment opportunities or cultural acceptance of women in the workforce. Contrastingly, rural areas record the lowest rates, potentially due to limited employment options or prevailing cultural norms. A noticeable decline in rural participation rates over the years indicates an increasing gap between urban and rural women's workforce engagement. Interestingly, a subtle rise in urban participation in 1401 hints at possibly positive shifts in the urban employment landscape. The data underscores the necessity for region-specific policy measures to bolster female workforce inclusion, with a particular focus on addressing the stark rural-urban divide.

## 2 Consumption In Iran

### 2.1 Question

Obtain the trend of the per capita fruit consumption rate change for each household, separated by total, urban, and rural.

#### 2.1.1 Data

The study on consumption utilizes the "*CombinedP3S01Fruits.dta*" dataset, which has been merged with clothing expenditure data from "*CombinedP3S03Clothes.dta*" and family size data from "*Familysize.dta*" to form a comprehensive view of household expenditures. This data integration facilitates the examination of per capita fruit consumption across various household types.

#### 2.1.2 Methodology

The methodology involves normalizing fruit expenditures by family size to obtain per capita values, followed by adjusting for inflation using the Consumer Price Index (CPI) provided by the Central Bank of Iran. The base year for CPI adjustments is set to 1395, ensuring the calculation of real values. The expenditure is then weighted by household weight (W) to accurately represent the sample population. This allows for the assessment of real per capita fruit consumption rates for urban, rural, and total household categories.

#### 2.1.3 Visualization

Data visualization is accomplished through both bar and line charts, which portray the trends in real per capita fruit consumption across the specified years for the total, urban, and rural household categories.

#### 2.1.4 Interpretation

The bar graph shows that urban areas have a higher consumption rate in comparison to rural areas for each year. The total consumption rate seems to be closer to the urban rate,

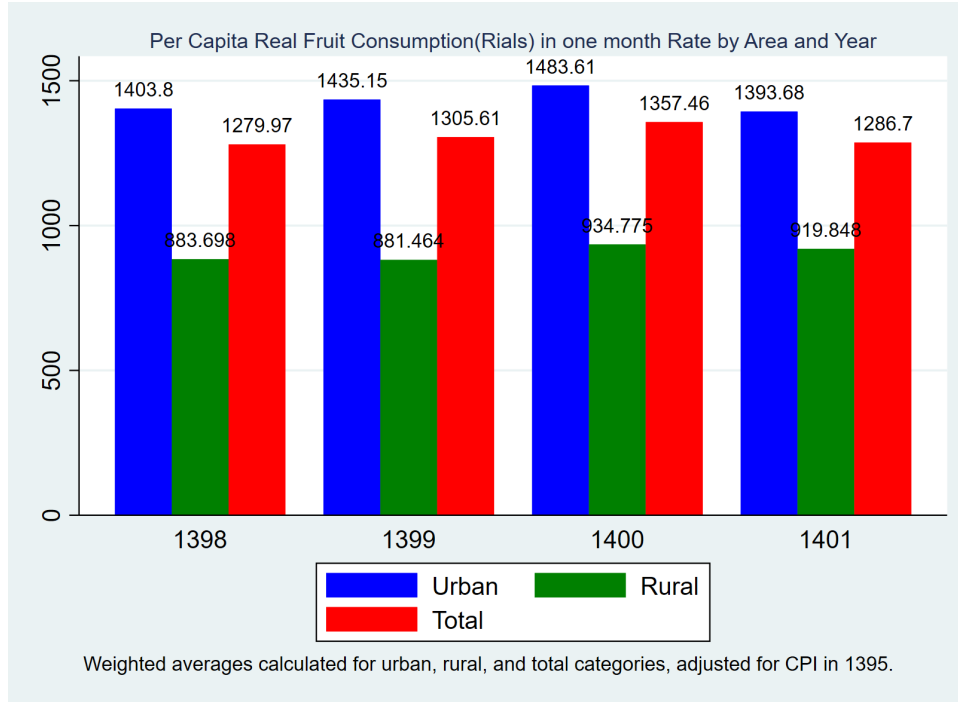


Figure 4: Bar chart depicting Per Capita Real Fruit Consumption (Rials) per month Rate by Area and Year.

indicating that urban consumption might be influencing the overall rate more significantly than rural consumption. There is a slight decrease in consumption from year 1398 to 1401 in urban areas, while rural areas show a slight increase in the same period. The total consumption rate fluctuates but ends slightly lower in 1401 than it started in 1398.

Both graphs indicate that urban residents consume more fruit per capita than rural residents, and the total consumption per capita is decreasing over the years. These trends are adjusted for the Consumer Price Index (CPI) as of the year 1395, which suggests that the rates take into account inflation and are meant to show real changes in consumption rather than changes due to price fluctuations.

## 2.2 Question

On average, how much has each Iranian household (rural, urban, and total) spent on clothing in these years?

### 2.2.1 Data

This analysis is predicated on data extracted from "Q2.dta," which amalgamates information on clothing expenditures with household demographics and sizes. The focus is on discerning the real value of clothing expenditures, adjusted for inflation, to gauge the financial commitment of Iranian households to clothing across different years and regions.

### 2.2.2 Methodology

The methodology encompasses adjusting the nominal values of clothing expenditures for inflation to derive real expenditure figures per household. This adjustment uses the



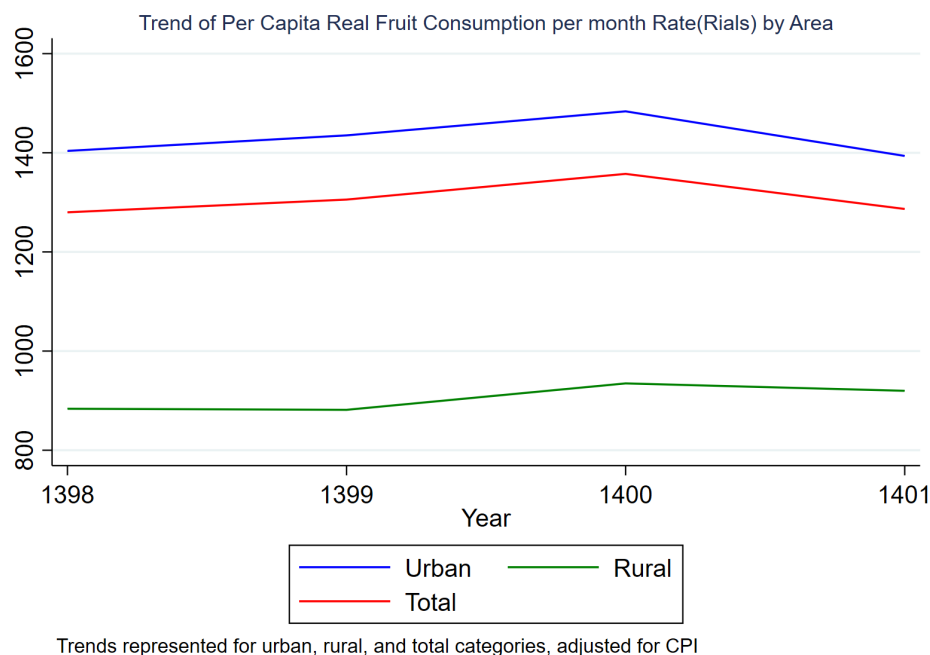


Figure 5: Line chart showing the Trend of Per Capita Real Fruit Consumption per month Rate (Rials) by Area.

Consumer Price Index (CPI) as provided by the Central Bank of Iran, with 1395 as the base year. Subsequently, the real expenditures are weighted by the household weight ( $W$ ) to ensure a representative analysis across different household types. This process allows for the calculation of the average real clothing expenditure for urban, rural, and total categories of households for each year within the study period.

### 2.2.3 Visualization

Visual representation of the data includes both bar and line charts. The bar chart delineates the average real clothing expenditure for urban, rural, and total household categories across the specified years. The line chart, conversely, plots the trend over time, providing a dynamic view of how real clothing expenditure has evolved.

**Note:** These figures highlight the variances in clothing expenditure among urban and rural households, providing insights into consumption patterns.

### 2.2.4 Interpretation

The first set of images shows the per capita real fruit consumption measured in Rials over one month. The bar chart indicates that urban areas consistently have higher fruit consumption rates compared to rural areas each year. The line graph corroborates this by showing an overall higher trend line for urban areas. There's a slight decrease in consumption in urban areas from year 1399 to 1401, while rural areas show a moderate increase. The total consumption (presumably a weighted average of urban and rural) slightly declines over the years.

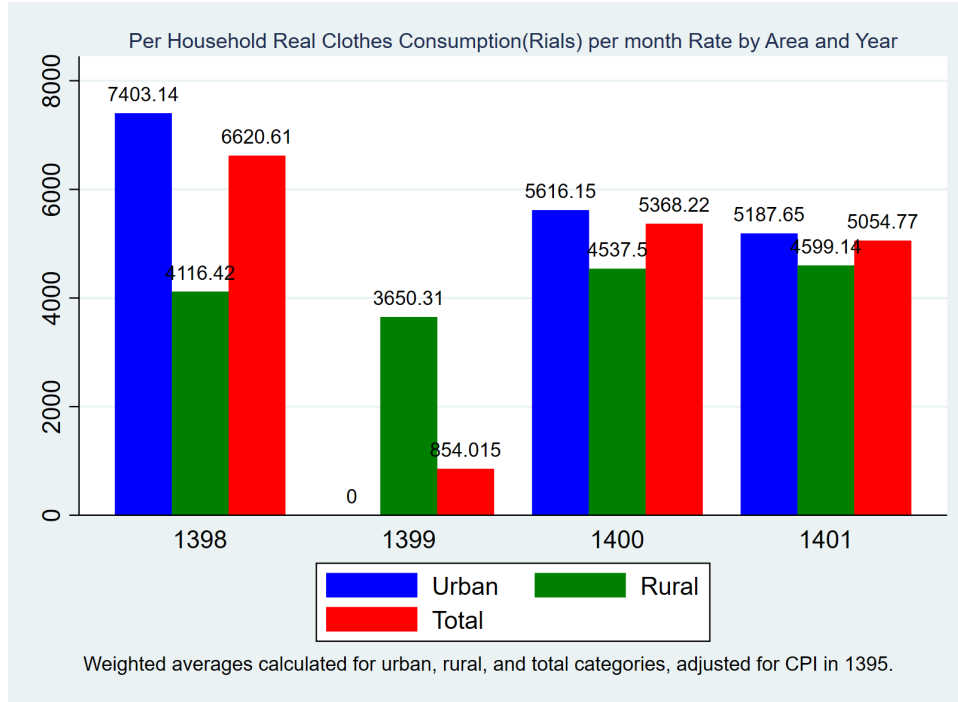


Figure 6: Average Real Clothing Expenditure (Rials) per Household by Area and Year.

The second set of images presents data on per household real clothes consumption per month. The bar chart reflects a significant decrease in clothes consumption from 1398 to 1399 in both urban and rural areas, with a subsequent recovery in 1400 and a slight decline in 1401. The line graph mirrors this with a sharp decline and recovery between 1398 and 1400. Urban areas have a higher rate of clothes consumption than rural areas overall, but both follow a similar trend over the years. Total consumption dips in 1399 but then stabilizes closer to the rates seen in 1398 by 1401.

## 3 Housing Condition In Iran

### 3.1 Question

What percentage of Iranian households, over these years, categorized by total, urban, and rural areas, have had rental, mortgaged, or owned housing?

#### 3.1.1 Data

Data for this section is drawn from the "CombinedP2Housing.dta" dataset, which includes detailed information on the types of housing (owned, rented, mortgaged, others) occupied by households. This dataset enables an analysis of housing conditions across different demographic segments, specifically urban and rural areas.

#### 3.1.2 Methodology

The analysis commences with categorizing households into different housing types based on available data. Weighted counts for each housing type are calculated for urban, rural, and total categories, ensuring that each observation is proportionately represented. These

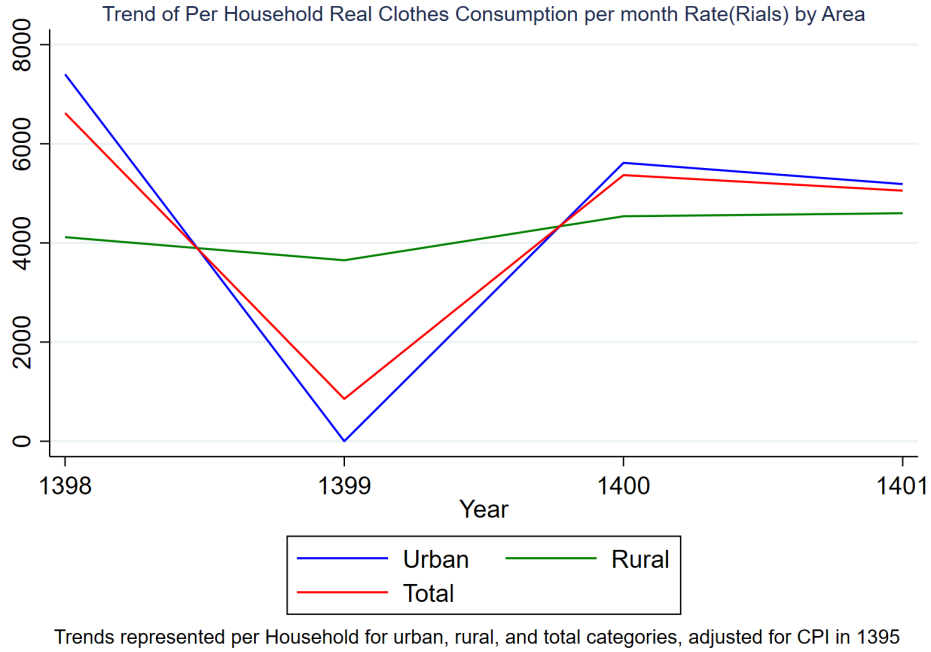


Figure 7: Trend of Per Household Real Clothing Expenditure (Rials) by Area.

counts are then used to compute the percentage of households in each housing category by area (urban, rural) and year, allowing for a temporal and spatial analysis of housing conditions.

### 3.1.3 Visualization

Visualizations include bar charts that depict the distribution of housing types by year for urban and rural households. These charts provide a clear, comparative view of housing trends, highlighting any shifts in housing conditions over time.(Figure 8 & 9 &10)

**Note:** The figures above delineate the percentages of owned, rented, and mortgaged housing among urban and rural households.

### 3.1.4 Interpretation

The charts display the distribution of housing types—owned, mortgaged, and rented—in urban and rural areas, as well as the overall total, from the years 1398 to 1401. The data is adjusted for household weight, implying it may be representative of the relative proportions of different housing types within the population.

In urban areas, owned housing is predominant, occupying the largest percentage by a significant margin, followed by mortgaged and rented properties. There appears to be a consistent pattern across the years with only slight fluctuations. Rented housing remains the smallest category, indicating that urban dwellers are more likely to own or be paying a mortgage on their homes than to rent.

Rural areas show a similar trend with owned housing being the most common. However, the percentage of owned housing is even higher in rural areas compared to urban areas, and the percentage of rented housing is lower. Mortgaged housing in rural areas

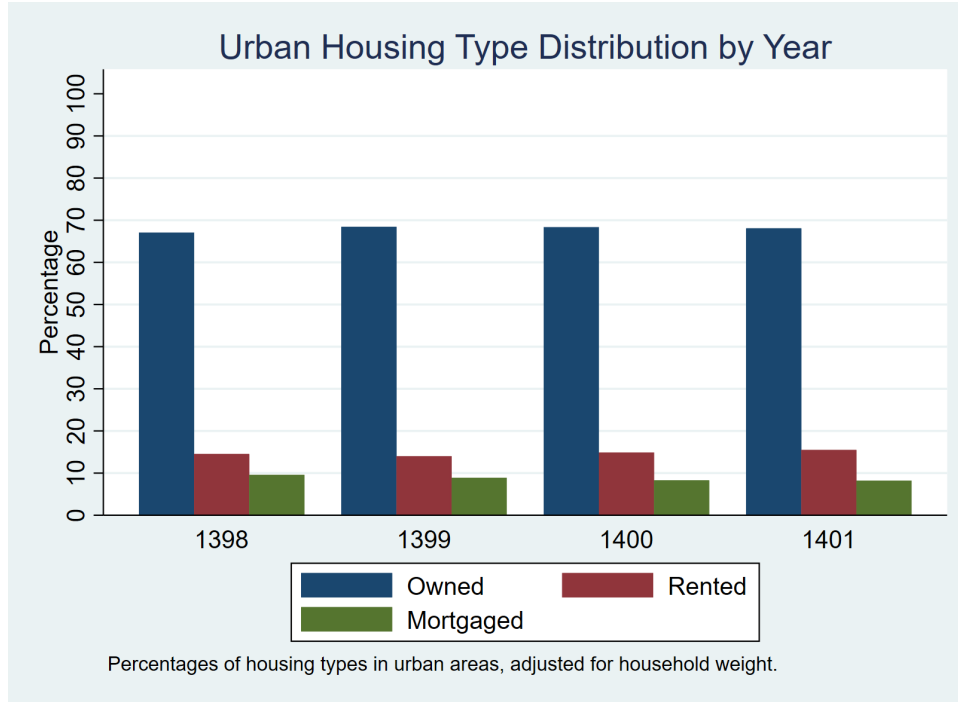


Figure 8: Urban Housing Type Distribution by Year.

is the least common of the three categories and remains relatively low and stable across the four years.

The total housing type distribution chart combines the data from both urban and rural areas. The dominance of owned houses is evident, with a very small percentage of houses being mortgaged or rented. This overall trend suggests that regardless of the area, the majority of dwellings are owned outright, with a minority being mortgaged or rented. This information could be crucial for understanding housing market dynamics, informing policy on homeownership, and planning for housing assistance programs.

## 3.2 Question

On average, how much has each Iranian household spent on assumed and real rents in these years? (calculate for total, urban, and rural households)

### 3.2.1 Data

This subsection utilizes data from "CombinedP3S04Housings.dta," focusing on rent expenditures of households across urban and rural settings.

### 3.2.2 Methodology

The analysis adjusts rent expenditures for inflation to obtain real values, ensuring the comparability across years. It then calculates the average real rent expenditure per household, weighted by household size, for urban, rural, and the total population.

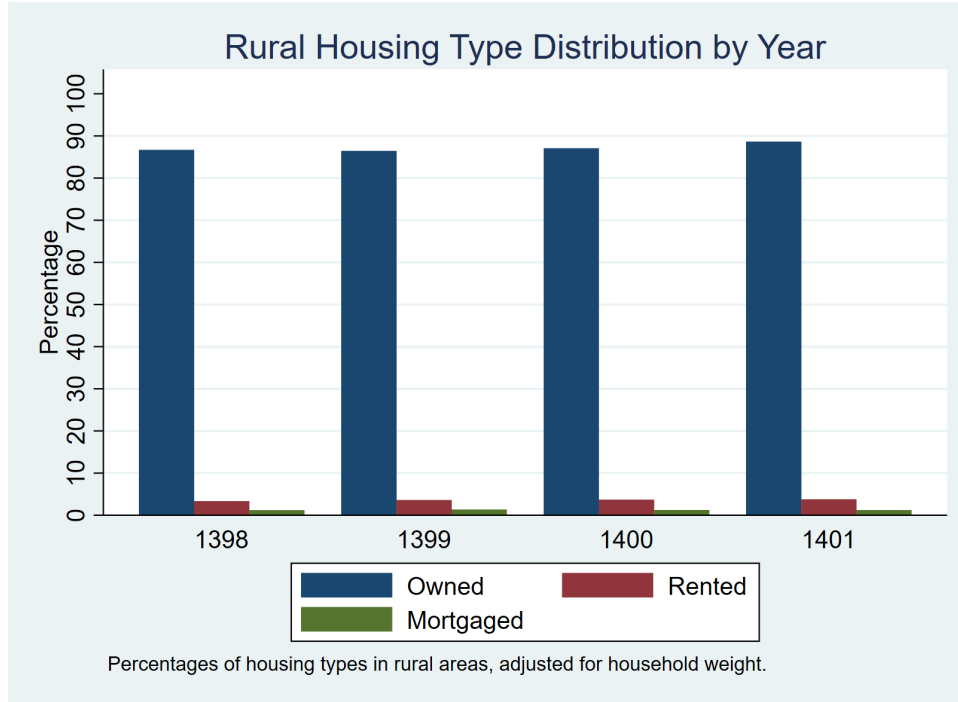


Figure 9: Rural Housing Type Distribution by Year.

### 3.2.3 Visualization

The data visualization includes both bar and line charts, depicting the trend of real and assumed rent per household by urban, rural, and total categories, adjusted for inflation. (Figure 13 & Figure 14)

### 3.2.4 Interpretation

In the urban areas, there is a noticeable increase in the rent rate from 1398 to 1401, with a significant jump in 1401. This suggests a growing demand or cost associated with urban housing. The rural areas, on the other hand, show a relatively stable rent rate across the four years, with a slight increase in 1401. The stability might indicate a more balanced housing market or less dynamic changes in rural housing demand or valuation.

The overall trend, represented by the total category, shows a steady rise in rent rates from 1398 to 1400 and a sharp increase in 1401. This rise in the total category is predominantly driven by the urban rates, which are substantially higher than the rural rates each year.

## 4 Inequality In Iran

### 4.1 Question

One of the indicators for measuring inequality is the ratio of expenditures of the tenth decile to the first decile. Using this indicator, identify the changes in inequality during the years 1380 to 1400 for the entire country, and separately for urban and rural areas. Have the subsidy programs implemented by governments made any changes in this trend?

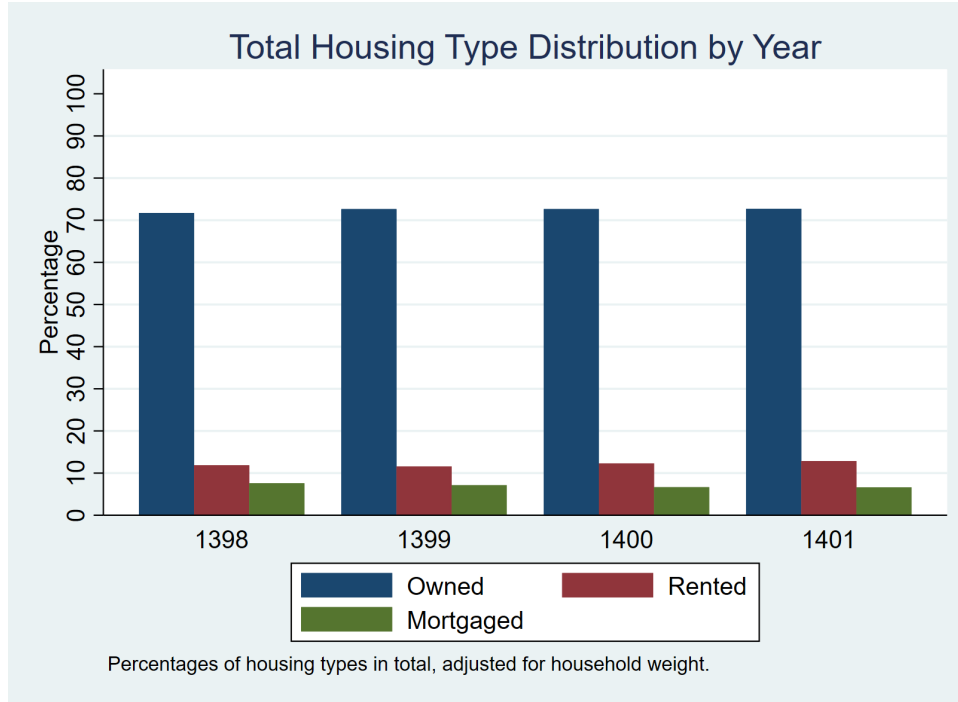


Figure 10: Total Housing Type Distribution by Year.

#### 4.1.1 Data Preparation

The analysis began with a thorough data preparation process, involving the importation of household expenditure data from Excel spreadsheets for both rural and urban households across two decades. Data cleaning steps included renaming variables for clarity, destringing numeric fields, and calculating total household expenditures by combining various expenditure categories and adjusting for household size.

#### 4.1.2 Methodology

To assess inequality, we created decile groups based on total household expenditure, allowing for a nuanced analysis of spending patterns across the economic spectrum. The critical metric for our analysis, the ratio of expenditures between the highest (tenth) and lowest (first) deciles, was then computed for each year, providing a direct measure of inequality.

The analysis was stratified to separately evaluate urban and rural households, recognizing the distinct socio-economic environments that might influence expenditure patterns differently. This approach enables a comparative analysis of inequality within and between these two segments over time.

#### 4.1.3 Analysis

The computation of decile ratios required aggregating household weights within each decile group, ensuring that our inequality measures accurately reflected the broader population's distribution. This process involved detailed scripting to automate the calculation of weighted expenditures and their ratios for each decile group across years and by urban and rural classifications.

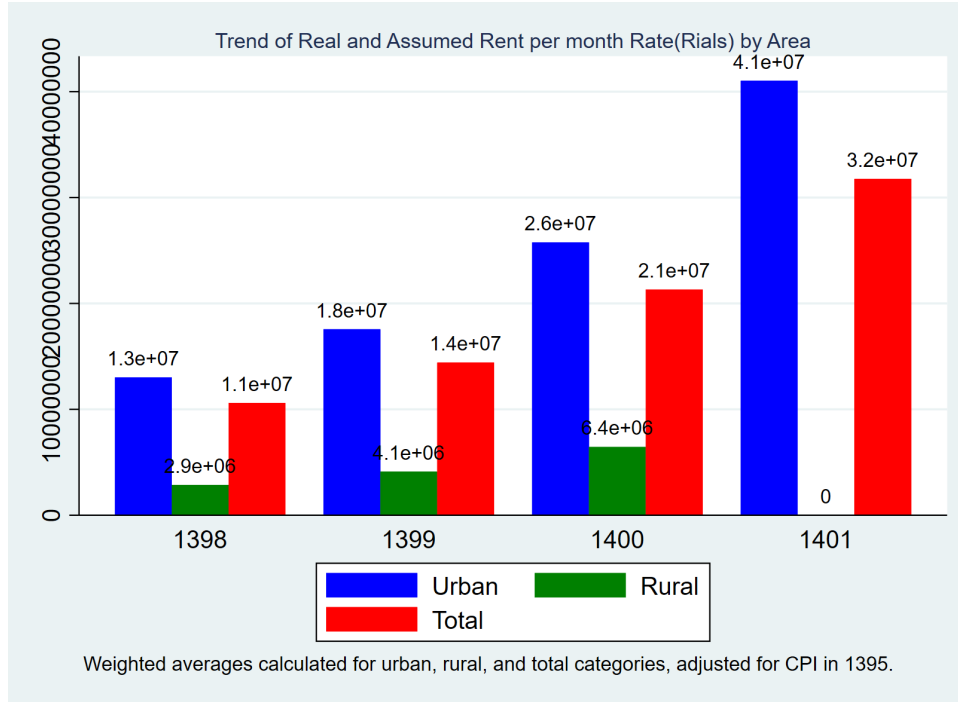


Figure 11: Trend of Real and Assumed Rent per Month Rate (Rials) by Area.

#### 4.1.4 Visualization

Graphical representations were created to depict the trend of the 10th/1st expenditure ratio over the period of study. Separate lines were plotted for the total, urban, and rural categories to visually compare inequality trends across different demographics.(figure 15)

**Note:** This figure illustrates the dynamics of economic inequality across urban and rural areas.

#### 4.1.5 Interpretation

A higher ratio indicates greater inequality, as it implies that the expenditure of the wealthiest decile (tenth) is many times that of the poorest (first). From 1380 to 1384, inequality was rising, as seen by an upward trend in the expenditure ratio for both urban and rural areas. In 1385, the implementation of subsidy programs appears to have made an impact, as the inequality ratio decreases sharply in the subsequent years, most notably in urban areas. This suggests that subsidy programs have been more effective in urban settings.

During the years following the start of the subsidy programs, up to around 1392, the inequality ratio continued to decline or remain stable for urban areas, while rural areas experienced a more modest decrease. After 1392, the inequality ratio for urban areas started to increase again, while rural areas showed a slight fluctuation but generally maintained a level trend.

By 1400, the trend lines for urban and rural areas converge slightly, indicating that the gap between urban and rural inequality has reduced, yet inequality in urban areas remains higher than in rural areas. The total ratio for the country reflects the combined trend of both areas, with a notable decrease in inequality after the subsidy program's

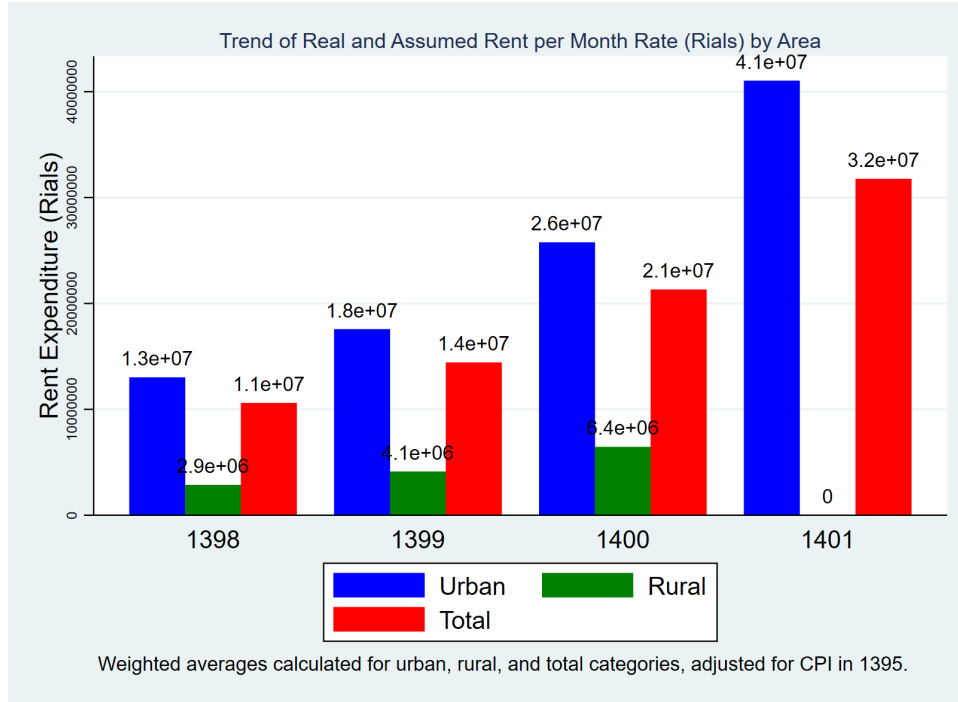


Figure 12: Trend of Real and Assumed Rent per Month Rate (Rials) by Area - Line Chart.

initiation, and then a gradual increase post-1392, although not returning to pre-subsidy levels by 1400.

This analysis suggests that the subsidy programs initially reduced inequality, particularly in urban areas. However, the effects seem to have diminished over time, with a gradual return to rising inequality, indicating that the impact of such programs may be temporary or influenced by other economic factors over the longer term.

## 4.2 Question

Many economists believe that inequality manifests itself at the highest and lowest percentiles. To observe this phenomenon, using the summary data of the year 1400, calculate the ratio of expenditures for each percentile to the previous percentile, and plot the trend of these changes.

### 4.2.1 Data Preparation

Data for the year 1400 was meticulously prepared by importing expenditure information from separate Excel files for rural and urban households. After importing, the data was cleaned and standardized, variables were renamed for clarity, and total household expenditures were calculated by summing various expenditure categories and adjusting for household size. This unified dataset, "Q4-2.dta," served as the basis for further analysis.



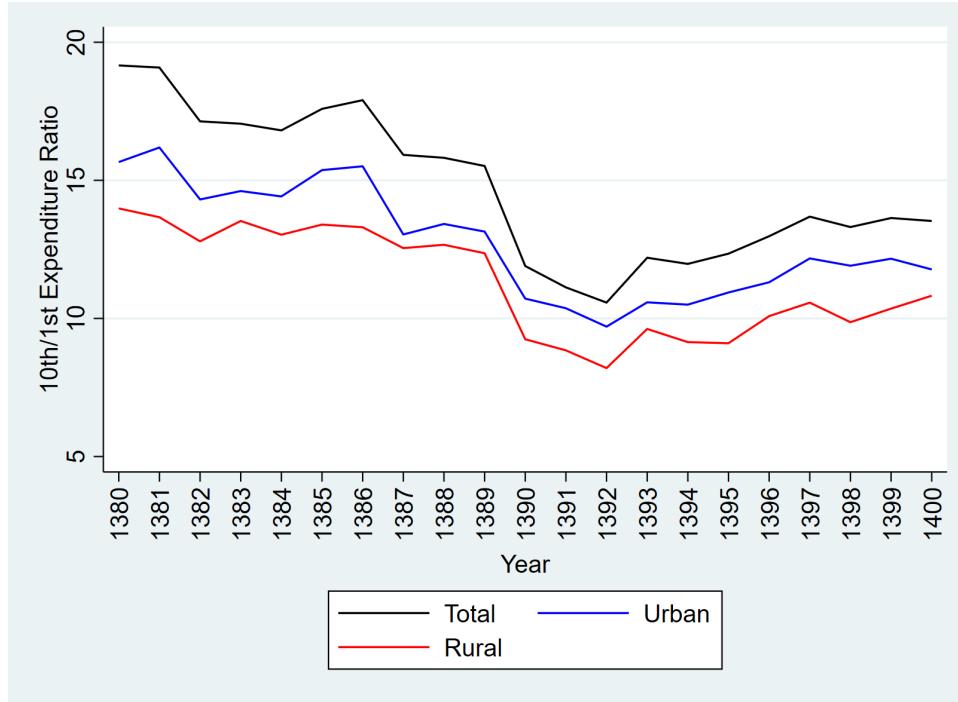


Figure 13: Trend of the 10th/1st Expenditure Ratio by Area, 1380-1400.

#### 4.2.2 Methodology

The methodology involved segmenting the households into percentiles based on their total expenditure, allowing for a granular analysis of spending patterns. The focus was on calculating the ratio of expenditure for each percentile compared to the previous percentile. This approach was applied separately to urban, rural, and the combined dataset to observe differential patterns of inequality.

#### 4.2.3 Analysis

For the analysis, households were divided into 100 percentiles, and the average expenditure for each percentile was calculated, weighted by the household weight to ensure representativeness. The ratios of consecutive percentiles' expenditures were then computed to identify any significant jumps in expenditure, which could indicate points of inequality.

#### 4.2.4 Visualization

An enhanced graph was generated to visualize the expenditure ratio trends across percentiles, distinguishing between urban, rural, and total categories. This visualization helps to pinpoint where the most significant disparities in expenditure occur, shedding light on the nature and extent of inequality within the population.(Figure 16)

#### 4.2.5 Interpretation

The graph shows that for the vast majority of the percentiles, particularly from around the 5th to the 80th percentile, the ratio is relatively close to 1 for both urban and rural areas as well as the total, which indicates a small incremental increase in expenditures

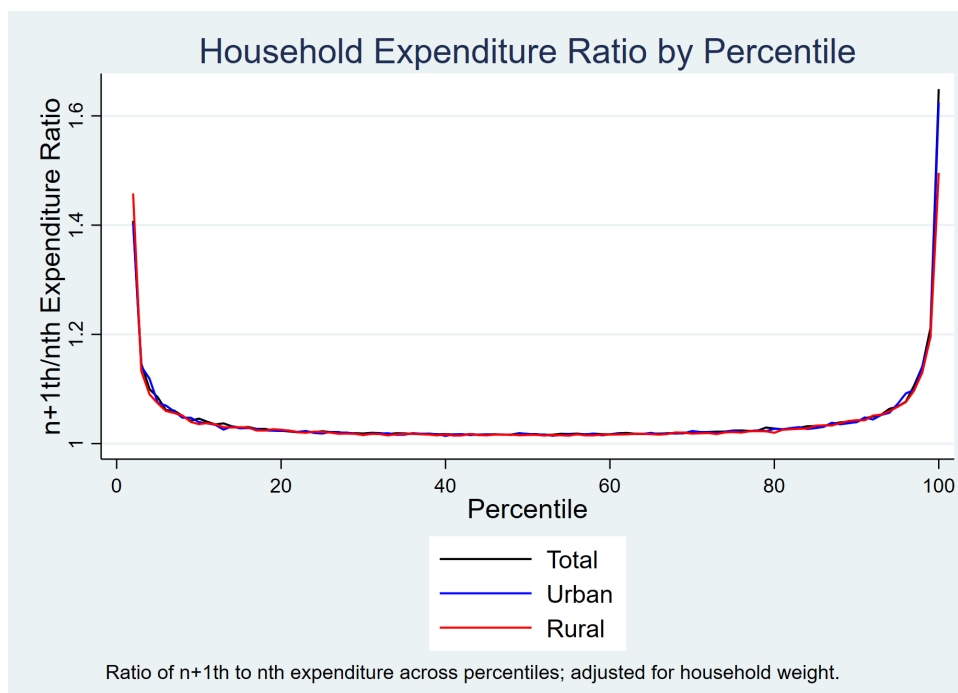


Figure 14: Household Expenditure Ratio by Percentile for the Year 1400.

from one percentile to the next. This suggests a level of homogeneity in expenditure within the middle class and working population. However, at the very lowest and highest percentiles, there are significant spikes. The sharp rise at the beginning of the graph, particularly for rural areas, indicates that the expenditure of the second percentile is much higher than that of the first percentile. This could be indicative of extreme poverty or a significant disparity within the lowest earners. Similarly, the sharp upturn at the end of the graph demonstrates that the expenditure of the top percentile (100th) is disproportionately higher than that of the 99th percentile. This is true for both urban and rural areas, although it's most pronounced in urban areas. This reflects significant inequality at the top of the income distribution, with the wealthiest households spending much more than those just below them in the distribution.