NHH

TERM PAPER
TECH2

Fall, 2025

Start: 2025-11-03 09:00

End: 2025-11-05 16:00

THE TERM PAPER SHOULD BE SUBMITTED IN WISEFLOW

You can find information on how to submit your paper here:
https://www.nhh.no/en/for-students/examinations/home-exams-and-assignments/

Your candidate number will be announced on StudentWeb. The candidate number should be noted on all pages (not your name or student number). In case of group examinations, the candidate numbers of all group members should be noted.

Collaboration between individuals or groups on submission preparation, as well as exchange of self-produced materials between individuals or groups is prohibited. The answer paper must consist of individual's or the group's own assessments and analysis. All communication during the home exam is considered cheating. All submitted assignments are processed in Ouriginal, a plagiarism control system used by NHH. Use of artificial intelligence (AI) in this exam is determined by the course coordinator. Information on AI usage and guidelines can be found on the course's Canvas page.

SUPPLEMENTARY REGULATIONS FOR EXAMINATIONS

You can find supplementary regulations under the headline “Regulations”

https://www.nhh.no/en/for-students/regulations/

Find more information under chapter 4.0 in the Supplementary provisions to the regulations for fulltime study programmes

Number of pages, including front page: 7

Number of attachments: Zip file with CSV files

About the group term paper

- Your solution needs to be submitted on Wiseflow.
- Submissions are done in groups of 2-3 students. Cooperation across groups is not allowed.
- Deadline: Wednesday, November 5 at 16:00. Late submissions will not be accepted.

Requirements

- You are allowed to use all online resources for help, including generative AI. You must include a statement on how you used these resources to solve the tasks.
- You should submit a *single* Jupyter notebook which contains your solution.
- Make sure your notebook runs without errors (`Restart` and `Run all`).
- Your implementation *must* work with the Anaconda environment TECH2 created from the environment file `environment.yml` in the GitHub repository (the same environment we used in part 2 of the course).
- Your notebook should be well structured and visually appealing, e.g.,:
  - Use markdown cells with headings and other formatting elements.
  - Each part should be in a clearly distinguishable sub-section.
  - Each function should be defined in its own code cell.
  - Each graph should be in its own cell.
  - Graphs should be visually appealing. They should include legends and labels so they are easy to comprehend.
  - Don't print hundreds of lines of output in your final notebook.
- Your code should satisfy the following criteria:
  - Your code is reusable and avoids duplication. An excellent submission should contain only one implementation of each function that is called repeatedly to solve different tasks (this is particularly relevant for parts 3-5).
  - Your code should be commented to help the reader understand what it is doing.
  - Your code should be properly formatted. You can use the "Black Formatter" VS Code extension for best results.
- Be sure to provide an explanation for your results where applicable.

Assessment

Your term paper will be assessed according to the following criteria:

Component	Points
Part 1	5
Part 2	15
Part 3	15
Part 4	30
Part 5	30
Notebook runs without errors	5

Note that a correct result is not sufficient to obtain full points: code efficiency, code reuse, elegance of the solution, and the presentation of the results are also factored in.

Expectations about inflation, house prices, and the stock market

You have just completed your NHH studies and started as an analyst at Norges Bank, the central bank of Norway. The bank governor, Ida Wolden Bache, is interested in how individuals form expectations about inflation and asset prices, since managing inflation expectations is an important part of running a central bank.

Your team is looking at household expectations in several countries, and you have been assigned to analyze the data for the United States. The data for this task comes from the Survey of Consumer Expectations (SCE), a monthly online survey of approximately 1,200 respondents that has been running since 2014 and is administered by the Federal Reserve Bank of New York, a branch of the US central bank. A colleague has already started processing the survey data and converted a subset of the variables to CSV format. These files are located in the `data/` folder. The appendix to this document contains a description of the variables present in these files. In this project, you will focus on the following subset of variables:

1. Demographic variables (gender, education, age)
2. Questions about a respondent's expectations over the next 12 months regarding:
    1. Inflation (changes in prices of a broadly defined consumption basket)
    2. Changes in the average house price.
    3. The probability that the stock market is going to increase in nominal terms.
3. Seven questions aimed at assessing a respondent's numeracy (numerical literacy). Specifically, you will be working with the seven indicator variables (with the suffix `_correct`) which are 1 if the question was answered correctly and 0 otherwise.

The governor is particularly interested in three sets of questions:

1. Do average expectations of inflation, house prices, and stock markets differ across gender, education, and numeracy (or numerical literacy, the ability to perform simple numerical tasks)?
2. Do expectation dynamics over time differ across these groups, and how do important geopolitical events affect these expectations?
3. Are expectations correlated with the actual outcomes, or are individuals more likely to extrapolate from past experiences?

The following sections guide you through answering these questions.

Part 1 — Reading in the data

The individual-level data for this report are stored in several files in the folder `data/`, which have filenames following the pattern `SCE-MMM-YYYY.csv`, where `MMM` is the month and `YYYY` is the year in which a particular wave of the survey was run. Each line in these files corresponds to an individual respondent in that month.

The SCE is a rotating panel: individuals are recruited into the survey and usually remain for 12 months (yielding 12 observations per individual) and are then replaced with new respondents. A handful of individuals are observed more than 12 times, while others stop responding before the 12th month.

Each individual is assigned a unique identifier (variable `userid`), which remains constant across all survey waves in which they are observed. Each survey wave is also assigned a unique identifier (variable `wid`).

1. Write code to read all SCE files and merge them into a single `DataFrame`.
2. Report the following statistics to get an idea of the sample size:
    1. Number of unique individuals in the dataset.

2. Number of observations (rows) in the dataset.
3. Number of unique survey waves.
4. The first and last dates observed in the dataset.

Part 2 — Data pre-processing

Before performing the actual analysis, you need to drop or fill in missing observations and create some additional variables:

1. The numeracy questions are only asked the first time a respondent enters the survey and are missing in subsequent waves. For each individual, fill in the missing numeracy variables (`num_lit_X_correct` where `X` is a number) using the values from the first observation.
2. Drop all observations (rows) with missing values for the following variables:
    1. Demographic information (gender, age, education)
    2. The three expectations questions about inflation, house price changes, and the stock market
    3. The seven numeracy questions (after you have forward-filled nonmissing values in step 1!)

    Report the number of observations dropped at each step.
3. Drop outliers (implausibly small or large values). For each expectations response:
    1. Compute the 0.1th percentile (0.001 quantile) and drop observations below this value.
    2. Compute the 99.9th percentile (0.999 quantile) and drop observations above this value.

    Report the number of observations dropped at each step.
4. Create a new column `college` equal to 1 if an individual has at least a bachelor's degree, and 0 otherwise.
5. For each individual, compute the total number of correct numeracy responses and report the fraction of individuals with 0, 1, . . . , 7 correct responses (e.g., 36.2% of individuals got all 7 right).

    Create a new column `num_lit_high` ("high numerical literacy") equal to 1 if an individual had *more* correct responses than the median, and 0 otherwise.
6. Report the same sample statistics as in Part 1 for the final data set.

Part 3 — Average expectations by group

With the processed data set in hand, you now turn to answering the governor's first question: how do expectations differ by gender, education, and numeracy?

For this part, perform the same analysis three times, each time splitting the sample by a different variable:

1. Males vs. females
2. Non-college vs. college
3. Low vs. high numeracy

For each part,

1. Compute the average for each expectations variable (inflation, house prices, stock market) separately for each group (males and females, non-college and college, etc.)
2. Create a figure with three panels (one per expectations variable) which depicts these group averages as bar charts.

The complete analysis should produce three figures, each with three panels, each panel containing two bars.

Part 4 — Expectation dynamics by group

To answer the governor's second set of questions, you need to investigate how average expectations evolved over time for the period of 2015-2024 covered by the sample.

For each grouping variable (`female`, `college`, `num_lit_high`):

1. Collapse the data to monthly averages for each of the expectation variables (inflation, house prices, stock market) for each group (e.g., for males vs. females).
2. Create a figure with three vertically stacked panels (one panel per expectation variable). Each panel should show the group time series (two series per panel, e.g., males vs. females) with time on the x-axis.
3. The governor wants to know how expectations reacted to important geopolitical events. Add vertical lines and annotations to each panel indicating the following events:
   - Trump elected US president for the first time (November 8, 2016)
   - COVID-19 pandemic goes global (February 1, 2020)
   - Biden elected US president (November 3, 2020)
   - Russia's full-scale invasion of Ukraine (February 24, 2022)
   - Nobel Prize in Literature awarded to Jon Fosse (October 3, 2023)
   - Trump elected US president for the second time (November 5, 2024)

   Which events had sizeable effects on expectations? Can you detect differences in how different groups adjusted their expectations?

The complete analysis should produce three figures, each with three panels, each panel containing two lines.

Part 5 — Correlations with past & future realizations

Finally, the governor is interested in whether people's expectations indicate realized future inflation or simply reflect past inflation experienced when answering the survey.

To answer this question, you first need to obtain data on realized inflation. A colleague has already downloaded data on the level of the Consumer Price Index (CPI) from FRED and stored it as a CSV file in the `data/` folder.

1. First, compare expectations to realized *future* inflation:
   1. Using this monthly CPI data, compute the realized inflation over the *next* 12 months; i.e., for each month t compute the forward-looking annual inflation as

$$Inflation_t = \frac{CPI_{t+12} - CPI_t}{CPI_t} \times 100$$

   2. Merge this inflation measure with the monthly averages by gender from Part 4. Specifically, match the average expected inflation by gender i in month t from the SCE, $ExpInflation_{it}$, with the forward-looking inflation measure $Inflation_t$ from the CPI data.
   3. Create a figure with two panels (one per gender), each showing a scatter plot of realized future inflation (y-axis) versus average expected inflation by gender.

      Compute the correlation between expected and realized inflation for each gender and add the correlation coefficient as text to the corresponding panel.

2. Repeat steps 1–3, but instead of forward-looking inflation, compute realized inflation over the *past* 12 months:

$$Inflation_t = \frac{CPI_t - CPI_{t-12}}{CPI_{t-12}} \times 100$$

Do you find differences in these correlation coefficients? What do these results say about how individuals form beliefs about inflation? Are there notable gender differences?

Note: In this part, use monthly data only; individual-level SCE data is not needed.

Data description

Variables

Legend for variable type:

`C` – continuous; `I` – indicator (0 or 1); `O` – ordinal categorical

Variable	Type	Description
userid	—	Unique respondent identifier
wid	—	Unique wave identifier
date	C	Date of survey interview
weight	C	Survey weight
age	C	Respondent's age at first interview
female	I	Respondent is female
couple	I	Married or living with partner
educ	O	Education of reference person (1 = no high school/GED, 2 = high school or GED, 3 = some college or associate degree, 4 = Bachelor's degree or higher)
num_kids	C	Number of children
black	I	Respondent is Black/AA non-Hispanic
hispanic	I	Respondent is Hispanic
owner	I	Respondent is a home owner
inflation	C	Question: What do you expect the rate of inflation to be over the next 12 months? Please give your best guess. (in percent)
house_price_change	C	Expected house price change over the next 12 months from the interview (in percent)
prob_stocks_up	C	Question: What do you think is the percent chance that 12 months from now, on average, stock prices in the U.S. stock market will be higher than they are now? (in percent)
num_lit_1	C	Question: In a sale, a shop is selling all items at half price. Before the sale, a sofa costs $300. How much will it cost in the sale?
num_lit_2	C	Question: Let's say you have $200 in a savings account. The account earns ten per cent interest per year, and interest accrues at each anniversary. If you never withdraw, how much will you have in the account at the end of two years?
num_lit_3	C	Question: In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1,000 people each buy a single ticket?
num_lit_5	C	Question: If the chance of getting a disease is 10 percent, how many people out of 1,000 would be expected to get the disease?
num_lit_6	C	Question: The chance of getting a viral infection is 0.0005. Out of 10,000 people, about how many of them are expected to get infected?
num_lit_8	O	Question: Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After one year, how much would you be able to buy with the money in this account? Answer Options: • More than today (1) • Exactly the same (2) • Less than today (3)

Variable	Type	Description
num_lit_9	O	Question: Please tell me whether this statement is true or false: Buying a single company's stock usually provides a safer return than a stock mutual fund. Answer Options: • True (1) • False (2)
num_lit_1_correct	I	Indicator whether respondent correctly answered question num_lit_1
num_lit_2_correct	I	Indicator whether respondent correctly answered question num_lit_2
num_lit_3_correct	I	Indicator whether respondent correctly answered question num_lit_3
num_lit_5_correct	I	Indicator whether respondent correctly answered question num_lit_5
num_lit_6_correct	I	Indicator whether respondent correctly answered question num_lit_6
num_lit_8_correct	I	Indicator whether respondent correctly answered question num_lit_8
num_lit_9_correct	I	Indicator whether respondent correctly answered question num_lit_9

Reference

- The Survey of Consumer Expectations (SCE) is administered by the Federal Reserve Bank of New York
- URL: https://www.newyorkfed.org/microeconomics/sce#/
- See here for the detailed questionnaire.