

Advanced Empirical Research Methods

MT Examined Problem Set - MT2024

You must answer all questions in Part 1 and Part 2. You must use Julia and R for the formal analysis (see exam rules in each Part). Please upload separate files/folders containing your answers (one per question) via Inspira by noon on Monday of Week 0 HT. Please do not include your name or other identifying information in your submission.

Part 1

Exam rules:

1. Use the Julia language, and write the code in Jupyter notebook or VS Code.
2. Provide a cleanly formatted PDF that contains all answers and results, as well as a separate set of files that contains all code to replicate your results. Export all figures and estimates to a separate folder called **results**.
3. All results must be replicable using your code.
4. Adding plots to the PDF that you are not able to replicate using your own solution (e.g., because you know the consumption policy is concave in wealth, but your algorithm gives you something convex) is considered cheating and leads to failure.
5. Do this exam on your own. Do not copy other student's code. If this happens, both students will fail the exam. So please do not ask a fellow student for this kind of favor.
6. You can use code from your homework, build on ChatGPT, etc., and use built-in functions, but your code must demonstrate that you understand what is happening.
7. If you find that the exam is missing assumptions to solve a problem (e.g., I forgot to specify the discount factor), make assumptions that will allow you to proceed, and state clearly what these assumptions are.

Setup: Non-Homothetic Household Savings

Time is discrete. Consider an economy with heterogeneous households subject to labor income shocks. Households can save in a risk-free bond b that pays gross return $1 + r$, and there is an ad-hoc borrowing constraint $b \geq 0$. The interest rate is set exogenously at 10%. The wage per efficiency unit of labor equals one, $w = 1$. This is the standard incomplete market model from class where there are no insurance markets, but households have access to a risk-free bond to insure themselves.

Households live forever and solve the following dynamic problem:

$$V = \max_{\{c_t\}_{t \geq 0}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t, b_t),$$

subject to the budget constraint:

$$\begin{aligned} b_{t+1} &= (1 + r)b_t + h_t - c_t, \\ b_{t+1} &\geq 0, \end{aligned}$$

where h_t are the efficiency units a household is endowed with, which are stochastic (I already used the fact that the wage is one), and a standard transversality condition holds.

The non-standard part of the problem is the flow utility $u(c, b)$ that depends not only on consumption but potentially on assets as well. We assume the following functional form:

$$u(c, b) = \frac{c^{1-\gamma}}{1-\gamma} + \phi \frac{(\kappa + b)^{1-\gamma}}{1-\gamma}.$$

Two observations are noteworthy: First, if $\phi = 0$, we get our standard CRRA utility back. If $\phi > 0$, there will be an additional incentive to hold assets – think of this as the joy of feeling wealthy. Second, for $\kappa > 0$, assets in the utility function are a “luxury” good.

We will explore what this means as well as solve the model. In short, there is an additional savings motive, which is relatively stronger for high-income households, on top of the standard consumption smoothing and precautionary savings motive.

Use the following parameter values:

Parameter	Value
β	0.85
γ	1.5
r	0.10
ρ	0.88
κ	1.5
tol_{HJB}	10^{-5}
tol_{KFE}	10^{-7}

Exercise 1: Standard Consumption-Saving Problem

Suppose $\phi = 0$, so we have the standard CRRA flow utility:

$$u(c, b) = \frac{c^{1-\gamma}}{1-\gamma}.$$

1. Construct a discrete grid of log labor income with 3 grid points, a lower bound of -1 , an upper bound of 1 , and centered around zero, using Rouwenhorst’s method. The coefficient of autocorrelation is $\rho = 0.88$.
2. After you derive the stochastic matrix that embodies this income process, iterate on this matrix until you have found the stationary income distribution $\mu(h)$ using the tolerance tol_KFE . Stop iterating when:

$$\max_h |\mu^{n+1}(h) - \mu^n(h)| < \text{tol}_{\text{KFE}}.$$

Plot the stationary probability mass function of log labor income.

3. Use the stationary distribution to compute the unconditional variance of the log of labor income. What is the variance?
4. Construct a linearly spaced asset grid with 1000 grid points where the highest value of assets is 20 times the highest labor income in levels. The lower bound is zero.
5. Solve the household consumption-saving problem using value function iteration. Plot the consumption policy function for the lowest and highest income types against assets. Label your plot appropriately.
6. Let household savings be measured as:

$$s(b, h) = \frac{h + rb - c}{h + rb},$$

i.e., savings are defined as household income minus household consumption relative to household income. Compute average household savings by income group using the stationary distribution $\mu(b, h)$, then compute:

$$\tilde{s}(h) = \sum_{b \in B} s(b, h) \cdot \mu(b, h),$$

where B includes all possible asset positions. The result will be the weighted average saving rate $\tilde{s}(h)$ in the economy for households of a particular labor income type. Make a bar diagram where each bar plots the average saving rate for low, middle, and high-income households, respectively. Is the saving rate increasing or decreasing in income? Why is that? Explain in max 2 sentences.

Exercise 2: Non-homothetic demand for savings

Suppose that the utility function now features an additional motive to hold assets by setting $\phi = 0.13$:

$$u(c, b) = \frac{c^{1-\gamma}}{1-\gamma} + 0.13 \cdot \frac{(1.5 + b)^{1-\gamma}}{1-\gamma}.$$

This specification reflects the additional utility benefit through the second term, which could represent status gains associated with being wealthy. We assumed that $\kappa = 1.5$.

1. Maintain the same setup as before (same income and asset grid) but adjust the flow payoff to include the non-homothetic asset holding motive.
2. Solve the value function and plot the consumption policy function for low and high-income households. Is there any difference to the previous consumption policy function?
3. Compute the average saving rate within each income group for this new solution and compare them to the old results from exercise 1.6. Create a bar plot with all six saving rates next to each other, cleanly labeled.
4. Are you able to create saving rates that are higher than the ones from the previous exercise? For which types does the saving rate increase or decrease, and why do you think that is? (max 4 sentences)

Part 2

Exam rules

Use R and produce your answers by generating a PDF from an R **Markdown** file which includes your code, its output, and your answers to questions in a single document. Each time you are asked to comment, it should be no more than two short paragraphs, unless specified otherwise. Rules 5 - 7 above apply.

Returns to a STEM degree

Examining the returns to education (how much one more year of education increases your lifetime income) is a well-studied problem in applied econometrics. Returns to different degrees is less studied but seems quantitatively just as important. One usual distinction is between STEM (Science, Technology, Engineering and Mathematics) and other degrees.

The data we will use is from a graduate survey. The population pools several cohorts of graduates in Germany who have completed tertiary education. They are observed five years after graduation.

The data contains the following variables:

- **age**: age at the time of the survey.
- **fem**: equals 1 if the respondent is a woman.
- **stemdeg**: equals 1 if the respondent has a STEM degree.
- **grade_abitur**: the grade at the secondary school exam (Abitur). It has been normalised to be between 0 and 100 percent.
- **binfathered** (resp. **binmothered**): equals 1 if the father (resp. the mother) has tertiary education.
- **income_5yrs**: the hourly wage after five years of graduation.
- **no_regret**: the respondent's answer to the question: 'if you could choose again, would you do the same degree?' It equals 1 when the answer is 'yes'.
- **related_job**: equals 1 if the respondent holds a job related to his/her degree.
- **bin_prop_faculty_stem**: equals 1 if proportion of female faculty in the year and region of Abitur is above the national average.

A researcher is interested in the distribution of graduates outcomes after graduation in a STEM degree.

- (1) Present a data summary of pertinent variables and compare the STEM degree holders to the non-degree holders. Tipp: Some data cleaning: Drop the extreme values of wages (below 5% and above 95% quantile). It is standard to use log-income rather than income level.
- (2) The researcher assumes first that all selection into STEM degree is based on observable characteristics. Denote (Y_0, Y_1) the potential outcomes, where the subscript 1 refers to STEM degrees. How would you estimate the marginal distributions $F_{Y_0}(y)$ and $F_{Y_1}(y)$?
- (3) Implement your choice for Y being successively the variables **income_5yrs**, **no_regret**, and **related_job**. If necessary, show the results on a graph. Explain your choice of control variables. Finally comment on the difference between both distributions.
- (4) The researcher focuses on the returns to STEM education for women and would like to use a binary instrumental variable Z . What are the conditions to estimate the marginal distributions of the potential outcomes $F_{Y_0}(y)$ and $F_{Y_1}(y)$ in the LATE model? How would you estimate them?
- (5) The researcher argues that one can use either of the following variables as an excluded variable:
 - the mother's education **binmothered**, which equals 1 when the mother has obtained tertiary education.
 - the proportion of female STEM faculty **bin_prop_female_faculty_stem**, which equals one, if in the year and in the province where one has completed secondary school, the proportion of faculty teaching STEM subjects is above the average proportion in the country. Could you argue in favour or against each instrument? Which one would you prefer? Please explain.
- (6) Compare the population of compliers (among women) between each of the two instruments. Please summarise your statistics in a table.
- (7) Estimate the marginal distributions for the compliers (among women) by adapting estimator (2) of Frohlich (2007). The formula in the lecture notes was:

$$\hat{E}(Y_1 - Y_0 | T = c) = \frac{\sum_i (Y_i Z_i / \hat{\pi}(X_i) - Y_i (1 - Z_i) / (1 - \hat{\pi}(X_i)))}{\sum_i (D_i Z_i / \hat{\pi}(X_i) - D_i (1 - Z_i) / (1 - \hat{\pi}(X_i)))}$$

where

$$\hat{\pi}(x) = \hat{E}(D | X = x), \text{ a nonparametric estimator of the propensity score.}$$

Comment on the difference of distributions for Y being successively the variables **income_5yrs**, **no_regret**, and **related_job**. Tipp: Assume that **grade_abitur** is the only regressor that you need to include.

- (8) Test for the hypothesis $H_0 : F_{Y_0} = F_{Y_1}$ among the compliers. Explain carefully your testing procedure. Tipp: ignore any regressor.

- (9) Given all your findings above, write a short abstract (max. 150 words) of a paper that would present your research about the returns to STEM degrees in Germany.