Course Project:

Structural Estimation in Household Finance

Deadline: December 21, 2023

Household Finance (PhD426)

Fall 2023

In this project, we are going to estimate three parameters, the time discount factor, risk aversion, and cost of the stock market participation, in a highly stylized model of consumption and portfolio choice over the lifecycle. Data moments regarding stock market participation, portfolio choice, and wealth are provided. A Matlab code that simulates participation, portfolio choice, and wealth accumulation is provided as well ("*simul.m*"); but of course, you may write your own code via any programming language if you wish. Please upload a PDF report of max 3 pages (excluding figures and tables) on Canvas. You may do this projects in groups of two students. Please attach all your codes as well. Below I sketch out the setup, estimation guidance, and what you are asked to do and report.

1 Model

An individual i lives in an environment with a risk-free asset, which delivers a time-invariant gross return R_f, and a risky asset, which delivers $R_t = \exp(z_t)$, where z_t is *iid* over time and is normally distributed (truncated at $+/-2.5$ STDs). The individual earns a *deterministic* labor income $Y_t = A_i \exp(gt)$, where $A_i = 1$, representing the national wage average at the base year, is set for normalization. The individual solves the following optimization program

$$U = \max_{C_{iT}(\cdot),\pi_{iT}(\cdot),W_{it}} \quad \mathbf{E}\{\sum_{t=0}^{T-1} \beta^t \frac{C_{it}^{1-\gamma}}{1-\gamma} + \beta^T b \frac{W_{iT}^{1-\gamma}}{1-\gamma}\}$$

such that

$$W_{it+1} = [W_{it} + Y_t - C_{it} - \phi_{it}] \cdot [\pi_{it} R_t + (1-\pi_{it}) R_f]$$

where C is the momentary consumption and π is the risky share in portfolio. W_{it} is wealth over time, the initial wealth W_{i0} is taken as given, and W_{iT} is the endogenous terminal wealth. $\phi_{it} = \Phi Y_t \mathbf{1}\{\pi_{it} > 0\}$ is the cost of the stock market participation, which is expressed in units of earnings. ϕ_{it} is paid only if the risky share is positive, i.e., if the individual participates in the stock market ($\pi_{it} > 0$). The individual earns labor income until age T, after which she is retired with the endogenous terminal wealth W_{iT}. $t \in [0, T-1]$ correspond to working periods. The parameter b determines the desire to save for retirement, which reflects both life expectancy and the desire to leave a bequest. Lastly, note that the individual may not borrow!

$$W_{it} \geq 0$$

at each point in time, in all states of the world.

2 Estimation

We calibrate g, R_f, mean and variance of R_t, which determines the stochastic process for z, b, and T as follows. Note that these are all fake numbers; do not cite them!

Parameter	Value
g	.01
R_f	1.01
$\mathbf{E}\{R_t\} - R_f$	.05
$\frac{\mathbf{E}\{r_t\}-r_f}{\mathbf{std}\{r_t\}}$	1/3
b	20
T	45

We then try to estimate β, γ, and Φ. To do so, we first solve the optimization program above, and then simulate the observable outcomes—wealth, participation, and risky share in the portfolio, for a cohort of 100,000 individuals. All such individuals have been exposed to a specific realization of stock returns R_t, recorded in the history of the data, which I have attached (*"realized_R.txt"*). We approximate the distribution of initial wealth across this sample of individuals by

$$W_{i0} = \exp(\mu_w - \sigma_w^2/2 + \sigma_w \nu_i)$$

where ν_i is normally distributed, with the fit-implied values $\mu_w = -1,\ \sigma_w = 1$. Recall that we normalize the national wage at year 0, $A = 1$, which sets the scale for all parameters and outputs with dollar units, including W_{i0} and its distribution across individuals.

The simulation code attached delivers the optimal policy for consumption-wealth ratio, risky share in the portfolio (which indicates participation as well: risky share is zero for those who do not participate), and wealth over the lifecycle pre retirement (45 years) for this cohort of individuals. Use this code and estimate β, γ and Φ. To do so, first, build informative simulated moments via aggregating simulated outcomes; then, match the simulated moments with data counterparts. We know the following moments from data (also attached, "*data_moments.txt*"), which can be useful for identification. Note that these are all fake numbers; do not cite them!

Moment	Description	Value in Data
$\mathbf{E}[\pi_{it} > 0]$	participation rate	0.8399
$\mathbf{E}[\pi_{it}]$	mean risky share	0.5941
$\mathbf{E}[\pi_{it} \vert \pi_{it} > 0]$	conditional mean risky share	0.7073
$\mathbf{E}[W_{it}]$	mean wealth	1.2151
$\mathbf{STD}[W_{it}]$	std wealth	1.9791
$\mathbf{E}[W_{iT}]$	mean wealth at retirement	4.9608
$\mathbf{STD}[W_{iT}]$	std wealth at retirement	0.4073

Notes: Except the last two rows, all rows report statistics across all individuals over all years, observations being weighted equally. The last two rows report the mean and std of wealth in the cross-section at the retirement age, respectively.

Note: Some of these moments might be irrelevant and not useful for the estimation. You do not need to target all of these moments for estimating the parameters, and you may even get "wrong" results by "polluting" your estimation with irrelevant moments. Simulation noise might be large for some of these moments: they might be sensitive/not robust to grid size/range, convergence criteria, etc. It is your job to figure out the relevant and useful, and robust set of targeted moments for estimating β, γ, and Φ. Please defend your choice.

Unfortunately, standard errors for these reported moments are not available. As a result, we may not get standard error of estimated parameters, and we may not do any statistical inference. Please only report your point estimates. But please compare your numbers with estimates in the literature for validity. Lastly, it is also your job to find out a relevant SMM weight matrix. Please describe how you set it.

Note: There might be bugs in the code that I am giving to you; go over it to understand how it works and make sure there is no issue, before using it for your estimation.

Note: The range to which these unknown parameters of estimates belong most likely is $\beta \in [.8, .99]$, $\gamma \in [1, 5]$, and $\Phi \in [0, .01]$ (also commented in the attached code).

Hint: It may take a long time, especially if you do it on a CPU, to estimate this model using traditional optimization algorithms (at least via the simulation code that I am giving to you). I suggest you use the 'approximate' SMM technique that we covered in the class. Based on my knowledge a training set of 1000 observations is more than enough in our case. And you can divide the job of building this set, or share/trade your set with each other :) Don't duplicate a computation task! By doing the approximate SMM you may also answer some of the question asked below (about subsample analysis and robustness to the measurement error) easily and more precisely ;) In any case, please briefly explain the algorithm you use, basically what you do to estimate the model.

3 Counterfactual Analyses

Having estimated the model parameters, try to adjust the code/write a new piece of code to answer the following counterfactual question. What is the impact of introducing a 30% capital income tax on the average wealth of this cohort of individuals at retirement? Do a partial equilibrium analysis, and assume the government spends this tax revenue outside of the model. How much this tax revenue would be btw? I mean, in present-value terms... Did you need a model to answer this question? If you just take the average wealth and participation rate and risky share from data, and do some naive accounting (back-of-envelope calculations), with some further assumptions of course, how far would be the end result? What if you repeat this policy experiment with a restricted model with participation cost $\Phi = 0$? What if you repeat this policy experiment in an environment with more risk averse individuals: $\gamma \to 2\gamma$?

Note: You have to adjust both parts of the code regarding value/policy function derivation, and then the simulation part for observable outcomes to answer this question.

4 Robustness

What is the robustness of your results to a potential noise in the reported mean wealth moment with an std of .1 (if you use that moment in your estimation, of course)? For example, instead of the moment mean wealth in the data be $\mathbf{E}[W_{it}] = 1.2151$, it can be any number $\mathbf{E}[W_{it}] = 1.2151 + 0.1 * z$ for a noise term z drawn from a standard normal distribution $z \sim \mathcal{N}(0,1)$.

What about the robustness of estimates to the set of targeted moments? I.e., How would the parameter estimates change, if you target various subset of moments in your SMM?

5 Heterogeneity

Take your estimate of the model above as given, noted by β, γ, Φ. Suppose there is indeed heterogeneity in the true economy, in the following sense. Half of the individuals have discount factor $\beta + 0.01$, risk aversion $\gamma - 0.5$ and participation cost 0. The other half have $\beta - 0.01$, $\gamma + 0.5$ and 2Φ. Simulate the behavior of the two subpopulation, each represented with 50,000 individuals. And then compute the same set of aggregated moments (represented in the table of data moments above), across *all* 100,000 individuals, and name it m^*. Now suppose a naive economist ignores heterogeneity and targets m^* in estimating the underlying parameters of a homogeneous economy version (just like section 1 above). Name this set of parameters by $\beta^*, \gamma^*, \Phi^*$. How large is the difference between the estimated parameters based on the average individual outcomes $\beta^*, \gamma^*, \Phi^*$, and the true average value of the fundamental parameters in the population β, γ, Φ?