Problem set 2 - Dynamic Macroeconomics

Nguyen Quoc Hieu

May 2025

Question 1

Data Processing

To create a new dataset for analysis, the VHLSS 2008 dataset was combined to enable the analysis of household income, consumption, and wealth.

Household Identification and Aggregation

First, for the data to be aggregated by household, the data for each individual must be classified so that the data is correctly arranged into each household. The process was done by selecting key features to identify household members: `tinh`, `huyen`, `xa`, `diaban`, and `hoso`. The indexes `tinh`, `huyen`, `xa`, and `diaban` were used to ensure that all members of the same household live in the same place, and `hoso` represents the household number. For example, if individuals in the dataset share the same value for `hoso`, it indicates they belong to the same household. Additionally, if the values for `tinh`, `huyen`, `xa`, `diaban`, and `hoso` are the same, the individuals are considered to live in the same household. Therefore, to aggregate individuals into households, the values of these variables must match.

Variable Selection for Income, Consumption, and Wealth

The next step is to select variables to sort by income, consumption, and wealth. The method was to compare the questionnaire answer codes with the dataset to locate the relevant variables. The data from VHLSS 2008 was divided into eight main sections. Sections 1 to 3 were used for demographic and basic household information. Section 4 was used to extract household income data, while Sections 5 and 7 were used to collect information on household consumption. Finally, Section 6 of the dataset was used to extract information related to household wealth.

Data Filtering Based on Household Head Characteristics

As mentioned, we group all household indicators according to the previous structure. We also continue to filter for households with male heads of households whose ages are also representative of that household. Columns for data filtering are also added to our dataset, `m1ac2` for gender and `m1ac5` for age. The goal of selecting data like this is to ensure that the household data is as uniform as possible for easy processing in the next steps. This also minimizes the number of variance effects of gender, age, and legal age for labor participation. Removing these data noise will minimize data noise, creating stability to be able to create in-depth interpretation for the study.

Data Filtering for Household Income, Consumption and Wealth

Category	Variable(s)	Description
Income	m4ac11, m4ac12f	Income from full-time jobs
	m4ac21, m4ac22, m4ac25	Income from other jobs that individuals may have besides their full-time job
Consumption	m5a1c2b, m5a1c3b, m5a2c6, m5a2c10	Food and drink consumption
	m5b1c4, m5b1c5	Non-food and drink consumption
	m5b2c2, m5b2c3	Annual consumption expenditure
	m5b3c2	Other annual household expenses
	m7c32, m7c36, m7c39	Utility bills: water, electricity, and garbage collection
Wealth	m6ac3, m6ac6, m6ac7	Fixed assets including land, houses, and infrastructure — recorded by quantity, market price, and condition
	(Durable goods not used)	Household appliances like TVs and refrigerators are excluded as they are not considered national assets

Table 1: Summary of variables used for measuring Income, Consumption, and Wealth

Age-Specific Productivity Adjustment Factor G_t

The geometric mean of total household income, G_t, is used to represent the average value of income of each household according to the age of the household. First, to avoid the skewness of the average income value, the logarithm function is used, and then the average value is calculated for household heads of the same age. Next, we exponentiate the results to convert the average value into average income value. In this model, G_t is used as an age-specific productivity adjustment factor, changing the income of households based on their age.

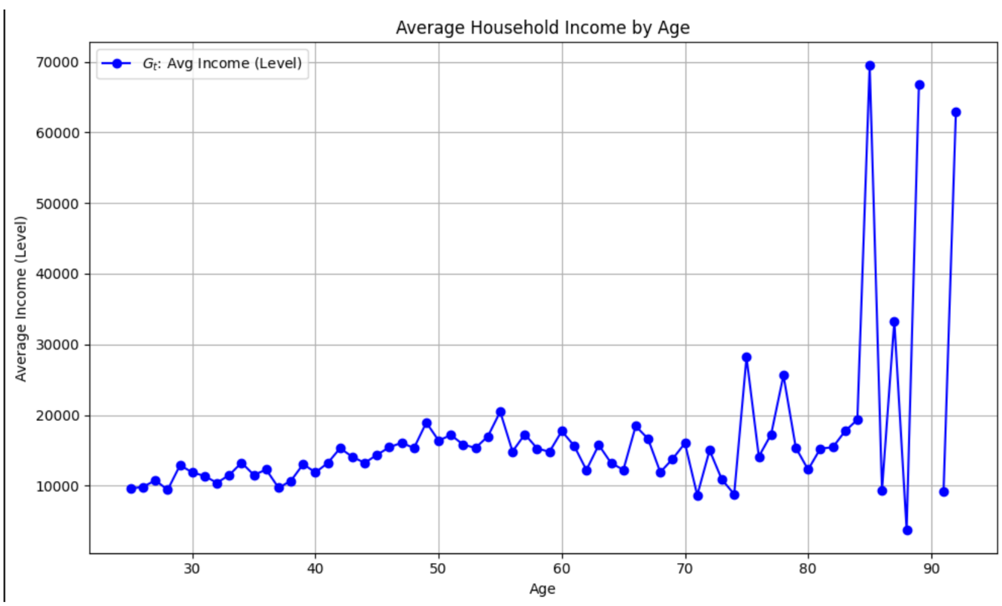


Figure 1: Average Household Income by Age (G_t)

Looking at Graph 1, we can see that the G_t chart increases gradually with age, and at age 50, G_t is measured at the highest income point and then gradually decreases. However, at age 70, the model data is no longer accurate as less data is collected, leading to fluctuations that make the results less robust. The pattern observed from G_t is plausible: as we enter the workforce, our income increases over time, peaks at age 50, when we have the most success in our job, and then declines over time as we age out of the workforce.

Simulation Model of Household Consumption Policy Function

Simulation Model of Consumption Policy for Households with Low Levels of Wealth

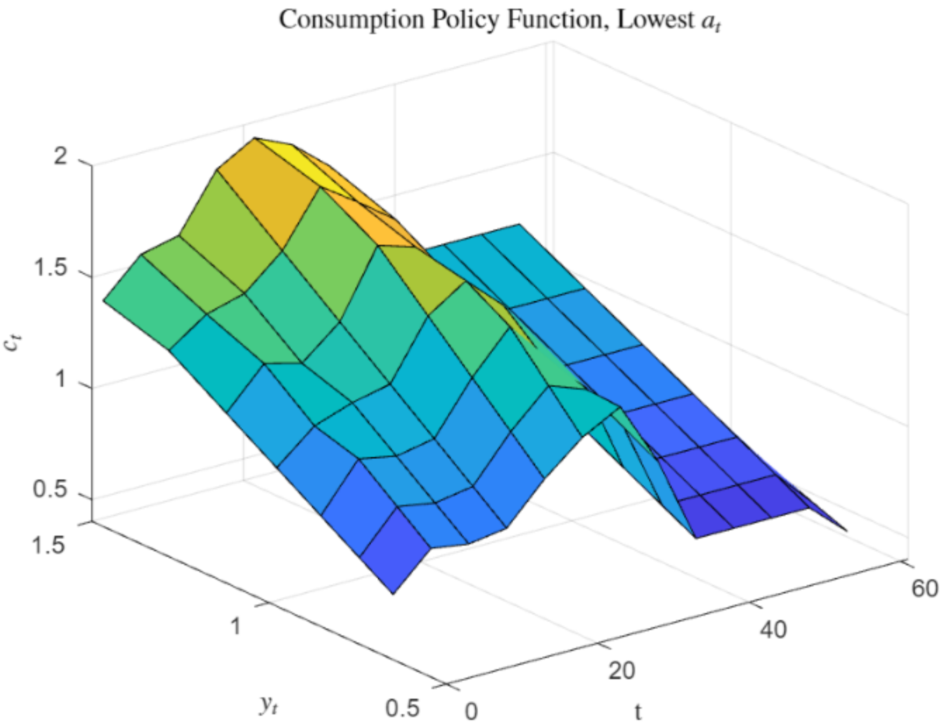


Figure 2: Simulation Model of Consumption Policy for Households with Low Levels of Wealth

This graph illustrates the simulation model for household consumption policy function based on three main factors: age, income, and consumption. This simulation will help us see how optimal consumption changes when there are factors of age and income levels, especially in households with minimal wealth. This simulation model reveals that consumption levels increase with age, peaking just before retirement and then gradually decreasing as they get older. This is also reasonable because, based on the G_t model, households' income will gradually increase as they grow up and join the labor force, which means their consumption will also increase with age during this period and peak just before retirement. Households' consumption will then gradually decrease as their income also decreases, and they no longer have the same consumption needs as before.

Simulation Model of Consumption Policy for Households with High Levels of Wealth

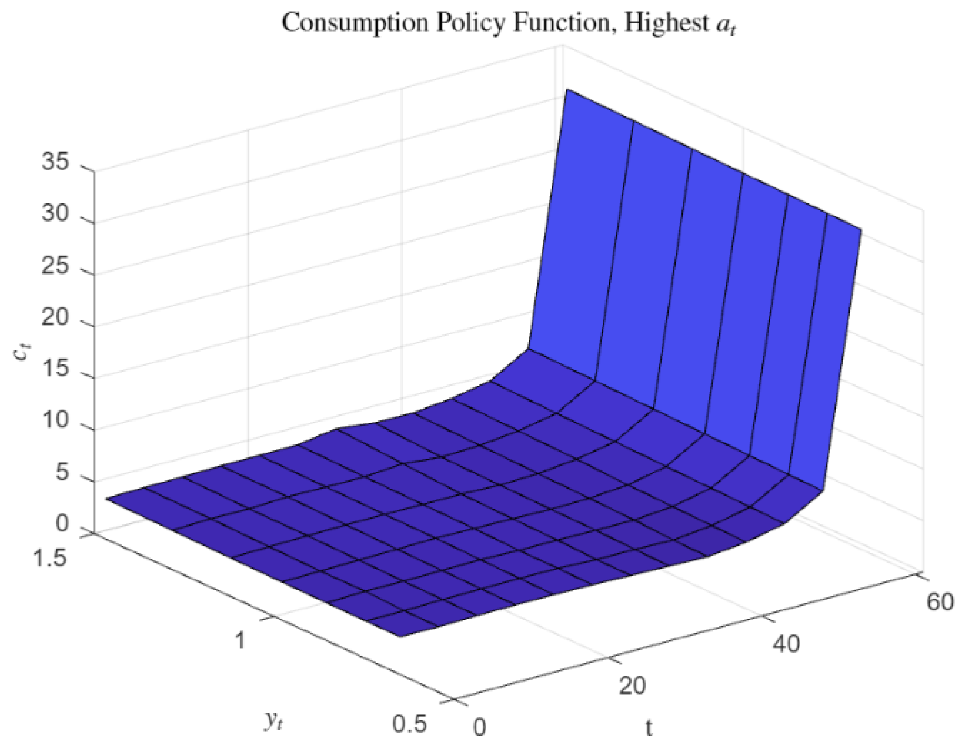


Figure 3: Simulation Model of Consumption Policy for Households with High Levels of Wealth

The second simulation model depicts the consumption policy function for the households with the highest wealth at the start of each period. This model allows us to analyze how consumption policy functions at low and high levels of wealth may differ. The optimal consumption model for the maximum level of wealth is similar in structure to the minimal level of wealth; it follows optimal consumption patterns when age and income level are combined. Unlike the results in Graph 1, where consumption is tightly related to and influenced by income level throughout time, Graph 2 demonstrates that consumption is much higher at any age, particularly after retirement.

This finding suggests that wealthy households consume more and can sustain that level regardless of income or age. We may conclude that these households can maintain high consumption levels for many years, and when they reach adulthood and enter the labor force, their consumption is higher than that of those with low levels of wealth, even though income may vary. This implies they may rely on their financial reserves and household assets to sustain their level of life, which is reinforced by the fact that after retirement, their household consumption can continue even if they no longer have a source of income. This is in stark contrast to their counterparts with limited to no assets, who see a dramatic drop in consumption after retirement.

Simulation Model of Saving Policy for Households with Low Levels of Wealth

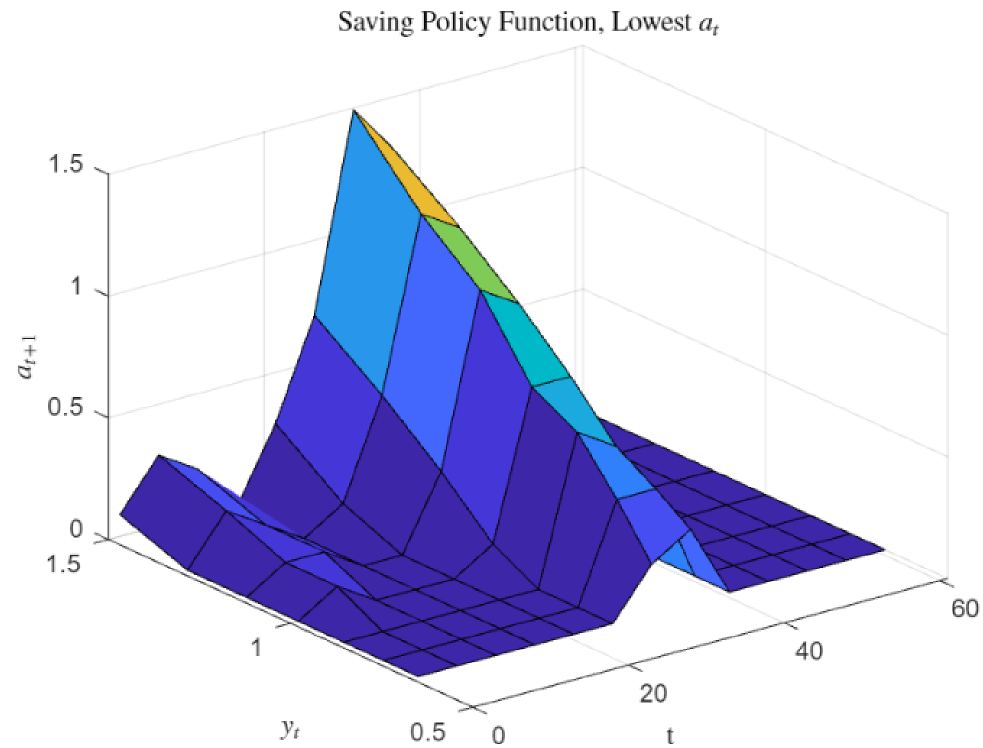


Figure 4: Simulation Model of Saving Policy for Households with Low Levels of Wealth

The third simulation model illustrates the saving policy function, the optimal savings for households in the next period with a low level of assets, and the same as previous simulation models. This model also incorporates the variables of age and income level. This function is based on a life cycle model, simulating the savings decisions of households when they have little or no savings. The result of this simulation depicts that in the early stages of life, these households have very low or zero savings. It will start to increase gradually and peak at the age of 30-50 for households with increasing income in midlife. Finally, in the near-retirement stage, saving continues to decline and is almost zero.

The reason for this result can be explained by the fact that in the early stages, incomes are usually low, and households with low levels of wealth spend almost all of their income. In the mid-life stage, when households have improved income sources, households with higher income levels will have more surpluses to save. This is also the time when these households have to save for retirement. Lastly, saving will decrease in the late stage of the life cycle, household incomes gradually decrease, while some expenses may increase related to health. It can be seen that families with minimal wealth levels try to smooth consumption throughout their lives by saving when incomes are high for retirement spending, but because they are households with few or no assets, they reduce their ability to earn and save effectively.

Simulation Model of Saving Policy for Households with High Levels of Wealth

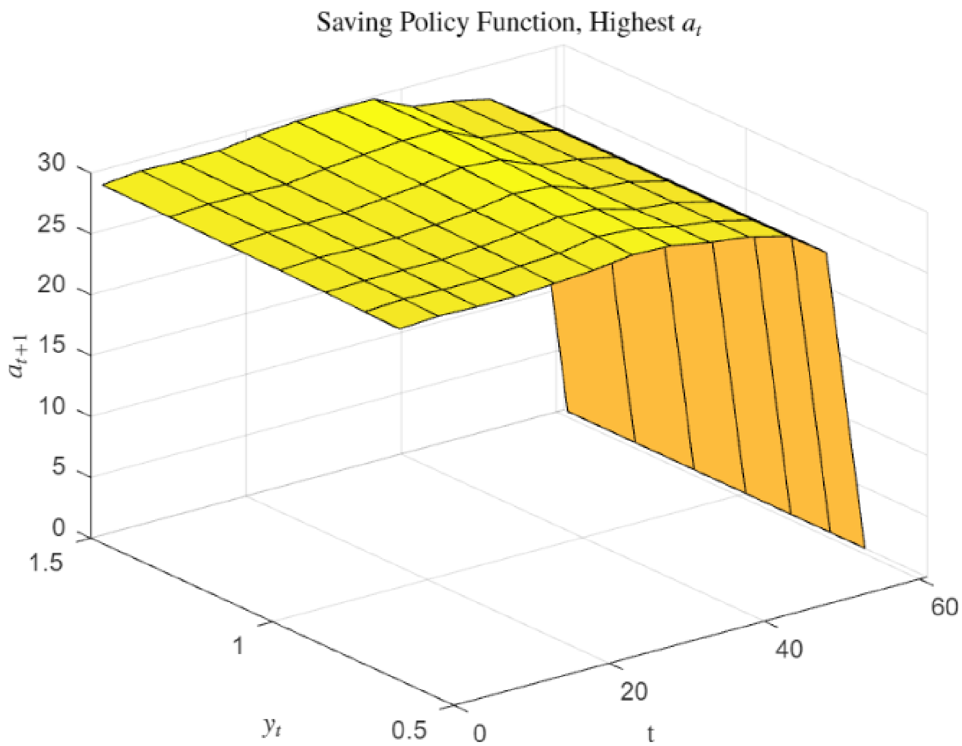


Figure 5: Simulation Model of Saving Policy for Households with High Levels of Wealth

The saving policy function for households with a high level of assets shows the savings for the next period of households with many financial reserves throughtout their life cycle. The simulation result shows that the savings of these households are much higher than those of minimal asset households. Until retirement, savings are very high and are maintained at this level until midlife. However, in post-retirement, the same pattern can be observed with minimal asset-level households: savings decrease sharply with increasing age.

To explain this observation, wealthy households have a lot of financial reserves, together with the income they have earned during their working years, which helps them maintain these savings levels during this period. However, after retirement, they no longer have these sources of income but still maintain their consumption levels when they are still working, which leads to a sharp decrease in savings in this period as they use their financial reserves to meet their living needs when they no longer need to save for later.

Simulation Model of Value Function for Households with Low Levels of Wealth

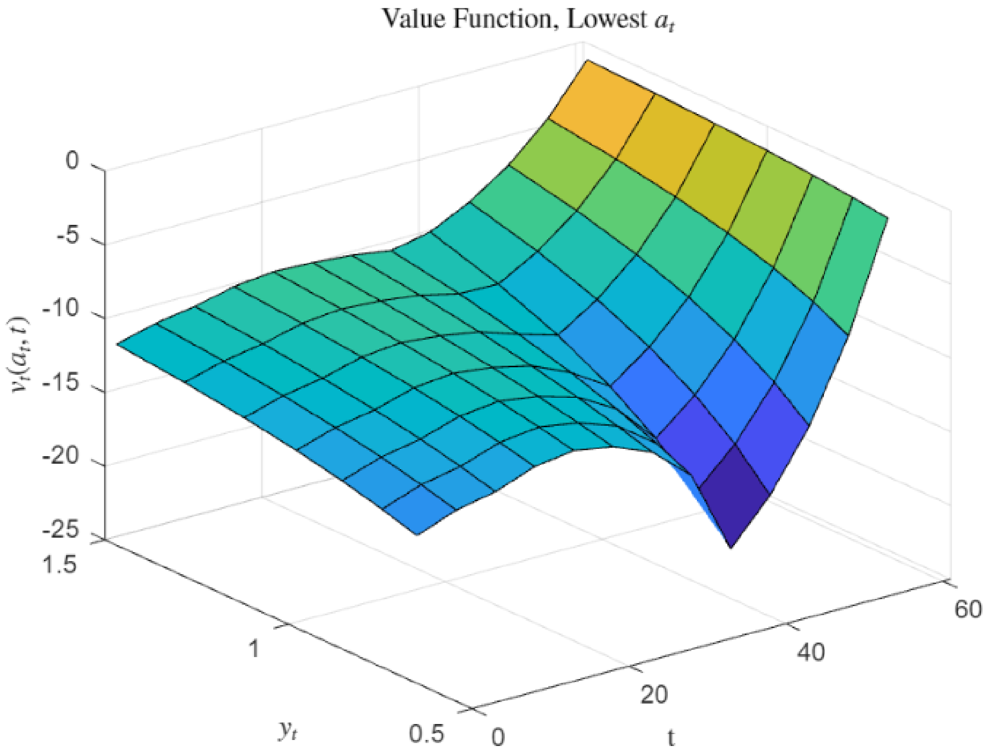


Figure 6: Simulation Model of Value Function for Households with Low Levels of Wealth

This simulation models the value function, which is the expected lifetime utility of a household with minimal levels of wealth. The model illustrates that when we are at the early stage in life, the value function is higher, showing that at this stage, we have a higher expected lifetime utility. In the next stage of life, it shows that expectation for lifetime utility decreases as age increases. This model also shows that for low level of wealth households, their expected utility will be higher when their income level is higher than for households with lower income levels at any age.

In the life cycle perspective, the value function reflects that young people, tend to have a more positive perspective for the future; perhaps they can make more money and plan for their future. As they age, people feel diminished in the actual outcomes compared to what they expect at the early stage, which leads to a decrease in expectation about utility. This can be more significant in low-income and low-level-wealth households, as the negative impacts can strike them more extremely.

Simulation Model of Value Function for Households with High Levels of Wealth

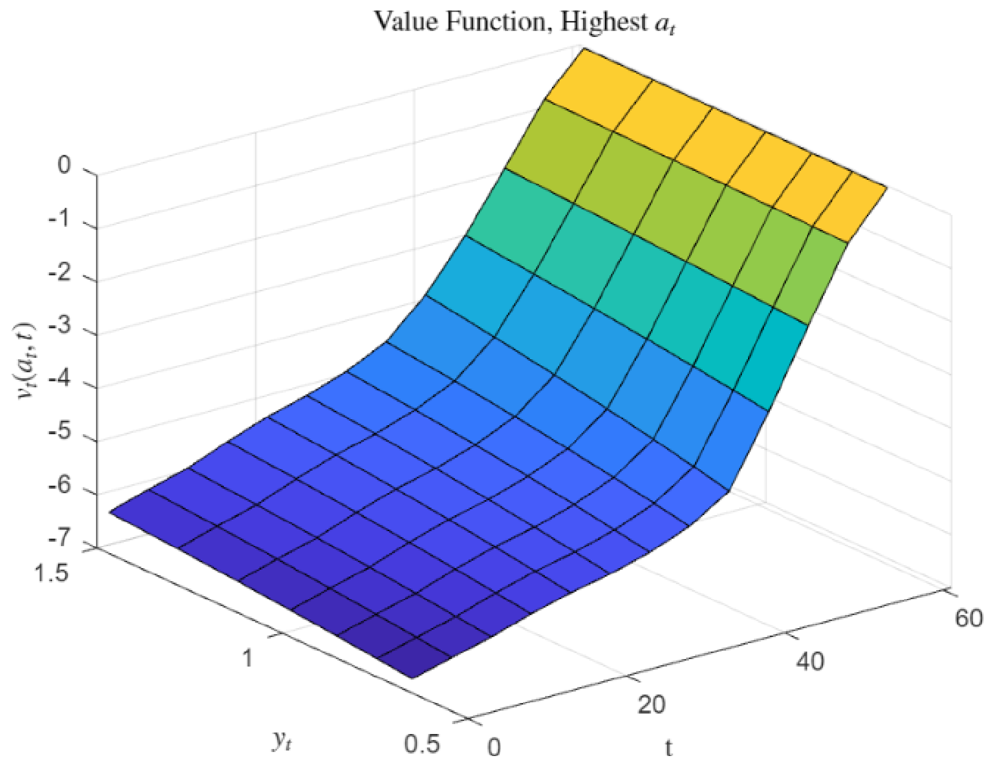


Figure 7: Simulation Model of Value Function for Households with High Levels of Wealth

This simulation model illustrates the value function for households with high levels of assets at the beginning of their life cycle. Compared to low to non assets households, the result for the value function shows that the level was higher than that of low asset households regardless of all ages and income levels. As age increases, the value function will also increase, especially, the value function will increase sharply when households are in the age range near, during, and after retirement. We can also observe that the difference in value function between households with higher income compared to the lower-income group in the high level asset household is not as significant as in the low level asset household.

To explain this, people in this group have stronger financial reserves, which makes them have more freedom and higher expectations in viewing their utility in the upcoming periods. Combined with their income, households in this group even have higher utility in later life, as they can rely on their savings and assets.

Simulation Model of Life Cycle Profile for Consumption

For this simulation model, we only extract the data for the household heads who are age 25 or above. This means that for the x-axis (age), the age needs to be increased by 25 to represent the actual age recorded in the simulation.

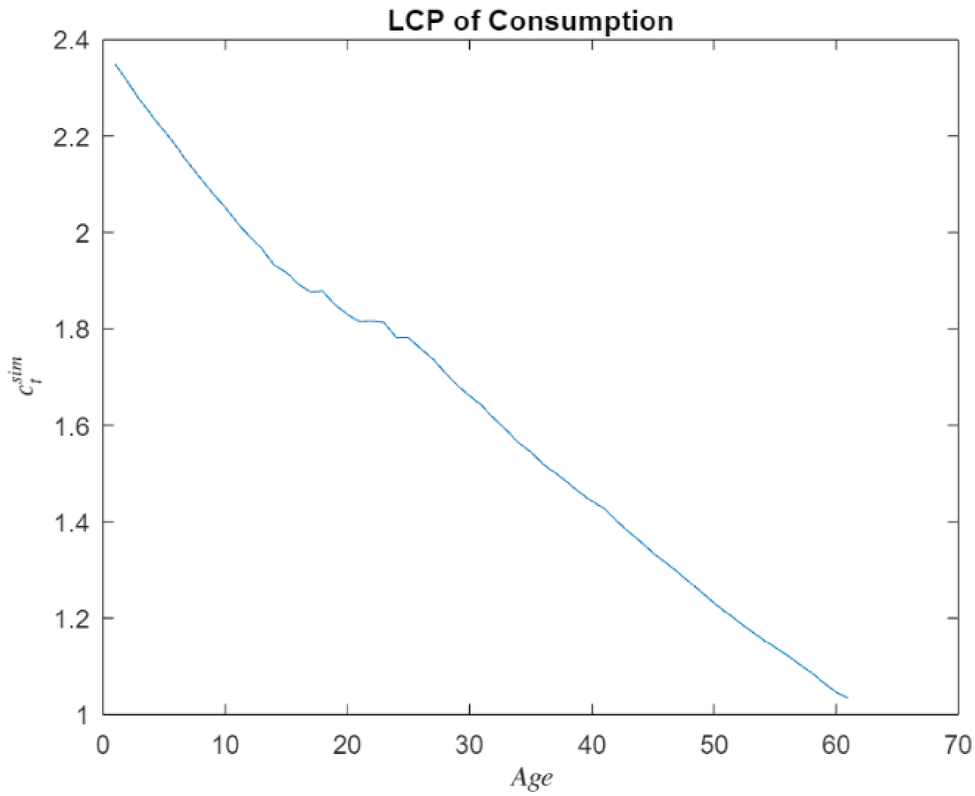


Figure 8: Simulation Model of Life Cycle Profile for Consumption

The graph shows the Life Cycle Profile (LCP) of consumption for each age of the representative household heads. The graph exhibits a decreasing trend, with consumption starting high at the beginning of the life cycle stage and then decreasing steadily as age increases. Throughout the period, the pattern of decrease shows little fluctuation, and no point on the graph indicates any sudden change in consumption habits.

A possible interpretation of this pattern can be explained by the early stage when individuals begin to enter the labor force (around age 25). Young people tend to consume more as they start living independently, facing costs such as rent and bills. Additionally, they may anticipate future income growth, leading them to maintain a higher level of consumption early on, expecting that increased earnings over time will offset their initial spending.

From the midlife stage onward, consumption levels gradually decline. This may be due to insufficient savings or a natural reduction in household consumption needs as individuals grow older and their lifestyles become more stable and less demand-intensive.

Simulation Model of Life Cycle Profile for Savings

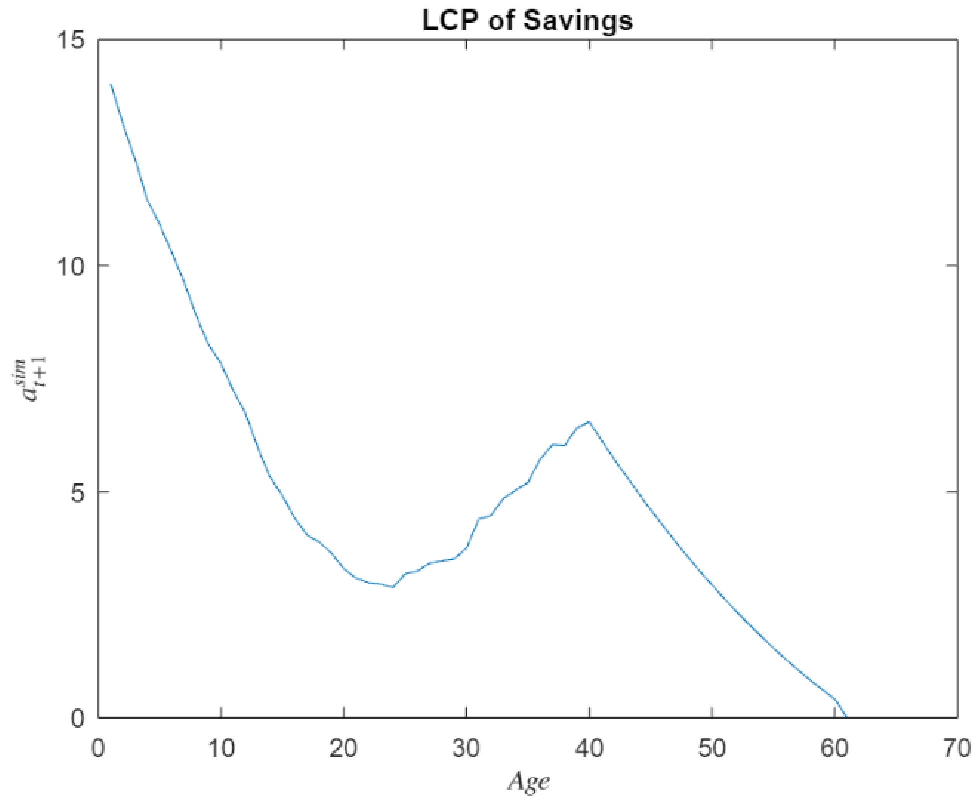


Figure 9: Simulation Model of Life Cycle Profile for Savings

For the simulation of LCP Savings, the trend line declines sharply from age 25 to around midlife, then the model shows that savings will increase steadily until retirement age. Finally, savings will continue to decline steadily and reach 0 when the person is over 80.

Explanation for this: In the early stage when entering the labor market, workers tend to save less when life is unstable and they have to pay for many things. In midlife, perhaps when their career is at its peak, they proactively save more to prepare for retirement, and finally, after retirement until death, they consume based on their savinsg, so the simulation shows that this data decreases until the end of the cycle.

Simulation Model of Life Cycle Profile for Utility

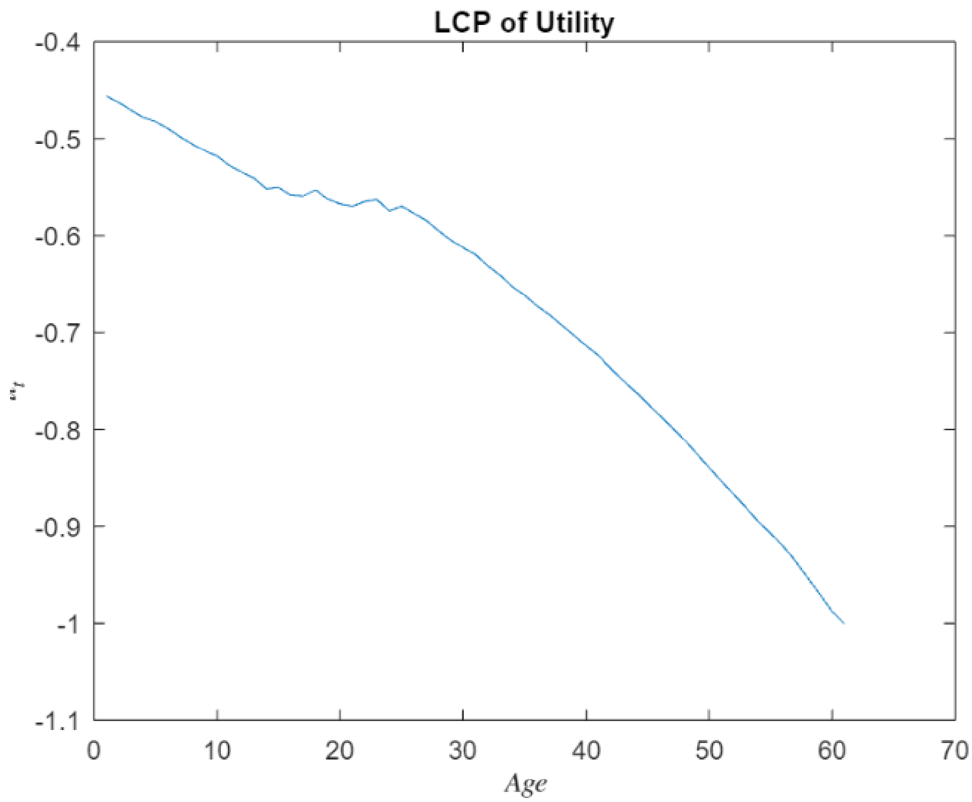
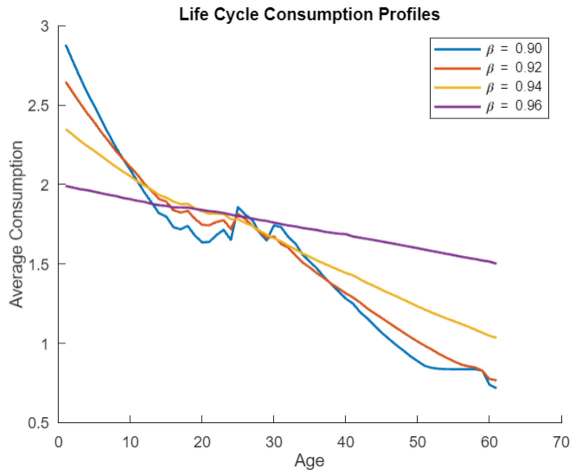


Figure 10: Simulation Model of Life Cycle Profile for Utility

Since both saving and consumption have a common pattern of decreasing with age, their utility also decreases over time. Living standards and future security worsen, leading to lower expectations for the future.

Simulation Model for different values of betas for Consumption



(a) Line plot for different values of betas for Consumption

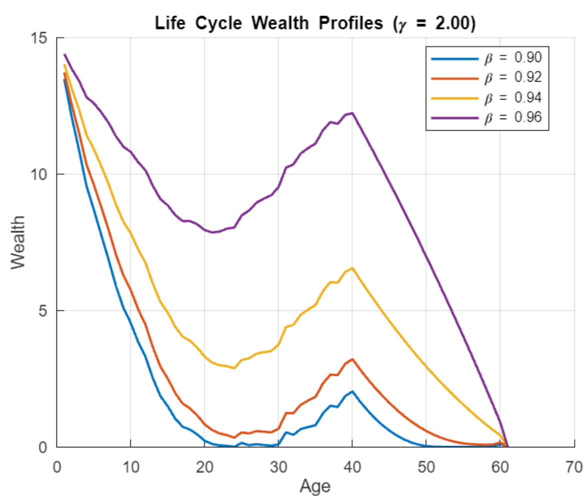
Beta	Mean	Std	Min	Max
0.9	1.5459	0.55875	0.71627	2.8804
0.92	1.5669	0.50794	0.76649	2.6465
0.94	1.6334	0.37153	1.0341	2.3483
0.96	1.7482	0.1403	1.5007	1.9908

(b) Descriptive table for different values of betas for Consumption

Figure 11: Simulation Model for different values of betas for Consumption

This simulation model shows savings at different levels of betas and a value of gamma. The graph and descriptive statistic table show that households with higher betas - that is, they are more patient in their consumption - show smoother consumption paths over time. When the descriptive statistic is conducted, the standard deviation decreases as the beta increases, and the mean increases. A Beta of 0.96 results in a mean of 1.7482, a standard deviation of 0.1403, and a minimum consumption level of 01.5007. On the other hand, the lowest beta results in the lowest mean, highest standard deviation, and lowest minimum consumption level. This suggests that households with higher betas have smoother consumption patterns than households with lower betas, which helps them spend more evenly and allocate their spending resources more efficiently.

Simulation Model for different values of betas for Wealth



(a) Line plot for different values of betas for Wealth

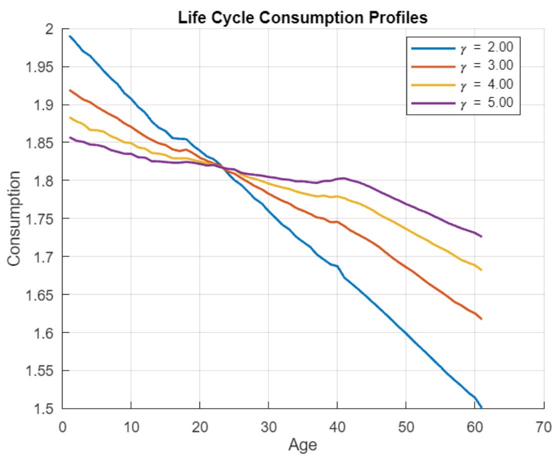
Beta	Mean	Std	Min	Max
0.9	1.9598	3.2394	0	13.476
0.92	2.6787	3.341	0	13.717
0.94	4.9629	3.12	0	14.024
0.96	8.9061	3.2099	0	14.393

(b) Descriptive table for different values of betas for Wealth

Figure 12: Simulation Model for different values of betas for Wealth

This simulation model shows wealth at different levels of betas and a value of gamma. The pattern of this simulation is similar to the pattern simulated previously for consumption. As beta increases, households become more patient and save more, which also helps them accumulate more wealth over than households with lower betas. The descriptive statistic helps us see this pattern more clearly, with households with higher betas having significantly higher average wealth than households with lower betas (8.9061 compared to 1.9598), and higher betas also helping households have higher maximum levels of wealth than other households. In particular, regardless of the beta level, the simulated minimum wealth can be zero, suggesting that despite higher patience, households can lose all their wealth at some point.

Simulation Model for different values of gamma for Consumption



(a) Line plot for different values of gammas for Consumption

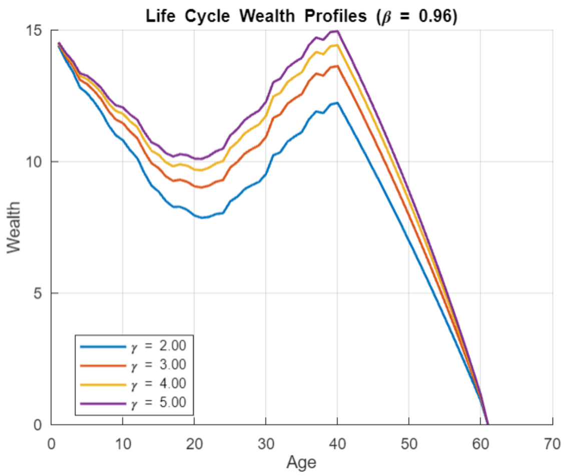
Gamma	Mean	Std	Min	Max
2	1.7482	0.1403	1.5007	1.9908
3	1.7757	0.084927	1.6173	1.9192
4	1.7912	0.053608	1.6817	1.8832
5	1.8016	0.033163	1.7258	1.8568

(b) Descriptive table for different values of gammas for Consumption

Figure 13: Simulation Model for different values of gamma for Consumption

This simulation model shows the consumption behavior of a household with different levels of risk aversion (represented by gamma equals 2, 3, 4, 5), with the beta index held constant at 0.96 over a life cycle. The pattern of the graph and the data in the stats table show that households with higher risk aversion (higher gamma) have a smoother consumption behavior pattern than those with lower risk aversion. Additionally, as beta increases, the average consumption of the household also improves, while the volatility of the data decreases. This shows that increasing the level of risk aversion of households will help them have higher and more stable consumption levels over a life cycle. Moreover, the highest and lowest consumption levels of these households are also higher, which, according to this simulation, brings more consumption benefits to households when they increase their level of risk aversion.

Simulation Model for different values of gamma for Wealth



(a) Line plot for different values of gammas for Wealth

Gamma	Mean	Std	Min	Max
2	8.9061	3.2099	0	14.393
3	9.8474	3.3647	0	14.466
4	10.381	3.4725	0	14.503
5	10.737	3.5526	0	14.946

(b) Descriptive table for different values of gammas for Wealth

Figure 14: Simulation Model for different values of gamma for Wealth

This simulation model shows how the total accumulated wealth of a household is related to different levels of risk aversion, represented by gamma, given a value of beta. As can be seen from the graph, total accumulated wealth increases as gamma increases from 2 to 5, and this pattern is maintained throughout the life cycle. The simulation also shows that the trend line reaches its highest peak before the retirement age and declines after that. Risk aversion is most evident in midlife, when households with high gamma increase their wealth accumulation. The stats table also supports and reinforces what the graph shows: increasing gamma also helps the average wealth of these households improve compared to other households. However, standard deviation increases with gamma, indicating a wider distribution of total wealth among these households. Although maximum wealth increases slightly as households become more risk averse, the minimum wealth of all risk averse levels is equal to zero. This also explains why standard deviation increases as gamma increases, since households, no matter how risk-averse, can still lose their entire wealth.

Heatmap of Average Wealth by Patience (β) and Risk Aversion (γ)

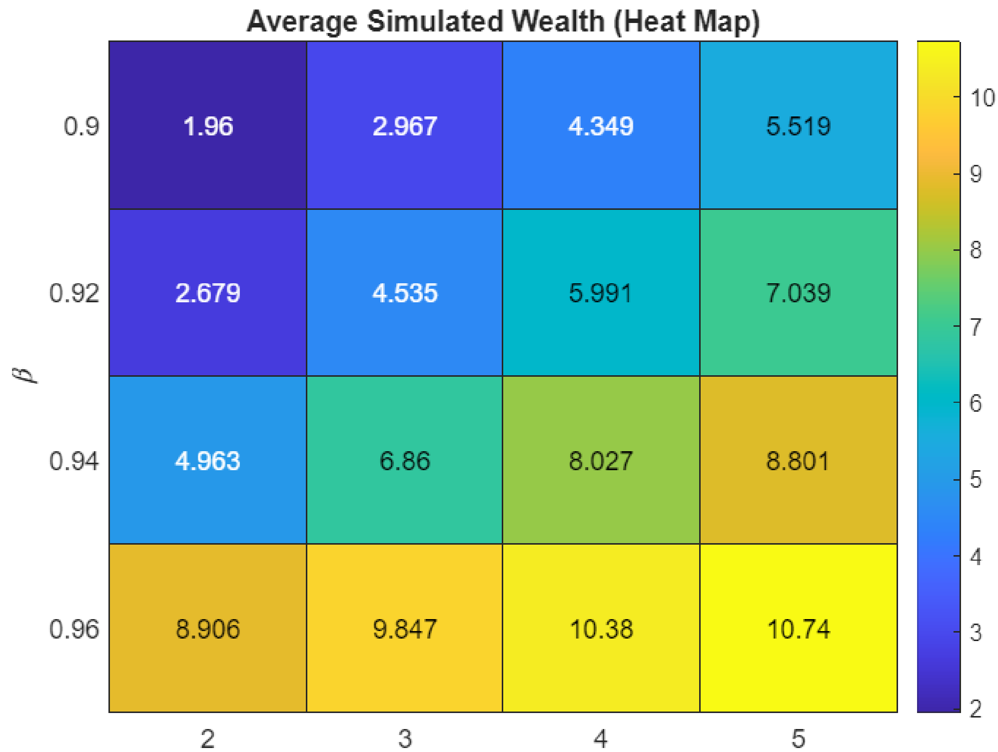


Figure 15: Heatmap of Average Wealth by Patience (β) and Risk Aversion (γ)

Here is a heat map showing the average simulated wealth with different values of β and γ:

β (beta) on the vertical axis: household patience

γ (gamma) on the horizontal axis: aversion to risk of the household.

Each square in the figure shows the simulated average wealth under the influence of different combinations of β and γ. Different colors have different meanings; dark blue represents lower wealth, and yellow represents higher wealth.

Key patterns: Wealth increases with β: On the vertical axis, as β increases from the top to the bottom ($\beta = 0.90 \rightarrow 0.96$), wealth increases steadily with each value of γ. This shows that more patient households accumulate more wealth.

Wealth increases with γ: On the horizontal axis, as γ moves from left to right ($\gamma = 2 \rightarrow 5$), wealth increases steadily with each value of β, although the increase is not as significant as with β. This shows that households with higher risk aversion accumulate more wealth.

The highest simulated wealth level was recorded to be 10.74, with $\beta = 0.96$ and $\gamma = 5$, while the lowest simulated wealth was recorded to be 1.96, with $\beta = 0.90$ and $\gamma = 2$.

This heatmap confirms that households with a combination of patience and risk aversion accumulate the most wealth.

Using Working Hours to Understand Household Consumption in Vietnam

When building a model to better understand how households in Vietnam spend money, it is extremely important to choose variables that influence daily spending. In many parts of Vietnam, especially rural areas, some people do not have stable jobs that pay them a fixed monthly salary. They have to do jobs such as selling goods on the street or working for hire. Therefore, measuring their income systematically and accurately is almost impossible.

So, instead of just looking at the income of one person, in this model total working hours is suggested. The index is easier to measure, clearly showing the effort of households in earning money. This also shows that the working and leisure time of each household affects the level of household's needs to spend on consumption.

For example, a household with more total working hours is more likely to have a higher income expectation, leading to more consumption. Conversely, households with lower total working hours will have less income, consume less, or value leisure more than extra income.

To test the relationship of the new variable, working hours, on household consumption, I used a scatter plot to observe the relationship between the two variables. The scatter plot clearly shows that household consumption increases with working hours, strengthening the idea of incorporating this variable into the optimization problem to depict labor choice.

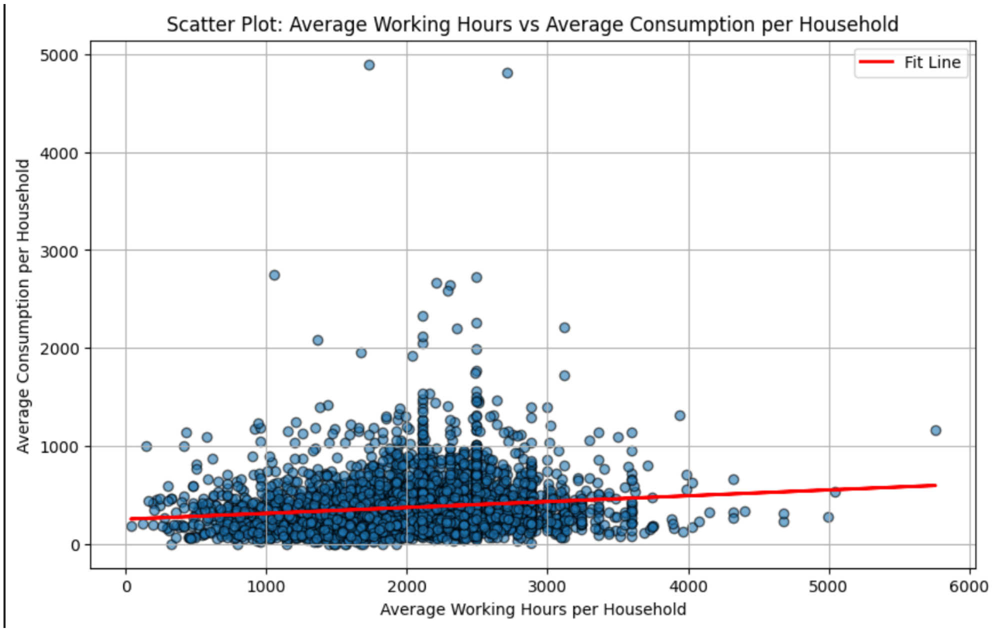


Figure 16: Scatter Plot of Household Consumption vs. Working Hours

To compute the total number of working hours, we use the following formula:

$$\text{Individuals Total Working Hours} = (\texttt{m4ac6} \times \texttt{m4ac7} \times \texttt{m4ac8}) + (\texttt{m4ac16} \times \texttt{m4ac17} \times \texttt{m4ac18})$$

where:

- `m4ac6` and `m4ac16` represent the number of working months in a year (for the first and second jobs, respectively),
- `m4ac7` and `m4ac17` represent the number of working days per month (for the first and second jobs, respectively),
- `m4ac8` and `m4ac18` represent the number of working hours per day (for the first and second jobs, respectively).

The next step is to filter out individuals with zero working hours. After computing the *Individual Total Working Hours* as described above, we retain only those individuals whose working hours are greater than zero.

We apply the same method as before to aggregate total working hours at the household level. Individuals sharing the same values for `tinh`, `huyen`, `xa`, `diaban`, and `hoso` are considered to belong to the same household. The sum of their individual working hours constitutes the *Household Total Working Hours*.

Finally, we compute the *Average Working Hours per Individual in the Household* by dividing the household's total working hours by the number of individuals in that household whose individual working hours are greater than zero.

To test whether total working hours actually affect the consumption behavior of each household, a scatter plot is used to compare the hours worked by households and the amount consumed by that household. The scatter plot shows that households that work more will consume more. Strengthen the relationship of workings hours and consumption

Because working hours can be measured and tested more easily, as well as have a clear relationship with consumption, I believe this is a useful factor for the model, helping to explain the consumption behavior of households, as well as improving the model in areas that income does not explain or that the data is insufficient.

Incorporating Leisure into the Optimization Model

$$V_t(a_t, t, y_t) = \max_{c_t, n_t, a_{t+1}} \left\{ \frac{c_t^{1-\sigma}}{1-\sigma} + \gamma \frac{(1-n_t)^{1+\frac{1}{\nu}}}{1+\frac{1}{\nu}} + \beta V_{t+1}(a_{t+1}, t+1, y_{t+1}) \right\}$$

subject to:

$$a_{t+1} = (1+r)(a_t + y_t - c_t),$$

$$y_t = \begin{cases} G_t n_t e^{(\rho \log y_{t-1} + \epsilon_t)} & \text{if } t < t_r, \\ \kappa y_{t_r - 1} & \text{if } t \geq t_r, \end{cases}$$

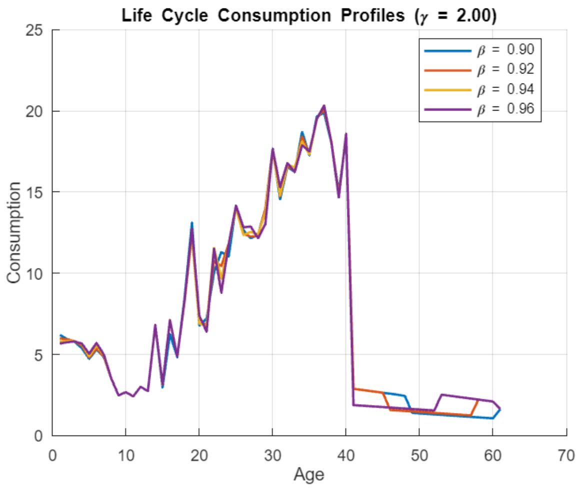
$$a_t \geq 0, \quad a_0 \geq 0 \text{ given}, \quad a_T = 0, \quad c_t > 0,$$

$$1 = n_t + l_t.$$

In this new model, the addition of working hours is an important change, completely changing the way the households' lifetime utility function was previously defined. Unlike the previous model, where the utility of households was based only on consumption, leisure time is now added to consider the utility function in addition to consumption. In which, nt is working time, and 1-nt is leisure time. With the addition of leisure time, the utility function of each household (yt) is not only affected by exogenous factors, which only depend on fixed factors such as income and consumption, but also by endogenous factors such as leisure time or working hours. Thus, households will have to face a trade-off: working more will help them have higher income and increase consumption, but this will simultaneously reduce the household's leisure time. On the contrary, if the household prioritizes having more leisure time, this will lead to a decrease in income, and less consumption.

With this new model, it will help simulate the value function of a household better and more realistically. Thanks to the ability to control the number of working hours, households now control both their income through their working hours and their consumption. An improvement over the old model, where only income affects the household's consumption.

Simulation Model for different values of betas for Consumption in the new Optimization problem



(a) Line plot for different values of betas for Consumption

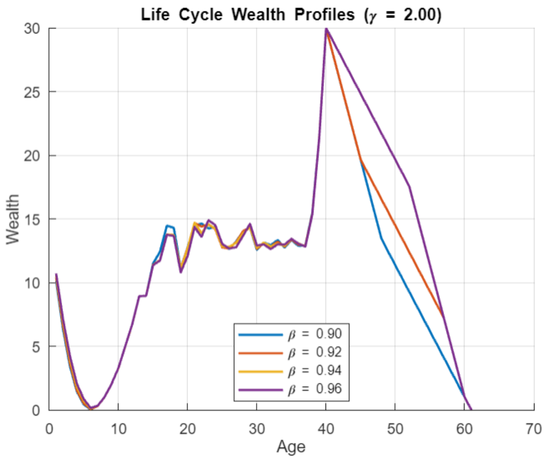
Gamma	Mean	Std	Min	Max
2	8.9061	3.2099	0	14.393
3	9.8474	3.3647	0	14.466
4	10.381	3.4725	0	14.503
5	10.737	3.5526	0	14.946

(b) Descriptive table for different values of betas for Consumption

Figure 17: Simulation Model for different values of betas for Consumption in the new Optimization problem

The simulated model shows the change in household consumption over the life cycle for different values of the discount factor beta (0.90, 0.92, 0.94, and 0.96), which reflects the degree of patience, while gamma is kept constant. In general, the pattern of household consumption tends to be similar to the simulated model without working hours. However, with the new model, households with higher betas show more patience while also maintaining higher consumption levels, especially in peak working years. The descriptive table supports this finding, indicating that the average consumption level increases with increasing beta, from 7.1927 at beta 0.90 to 7.2546 at beta 0.96. The measured consumption value for households also increases with beta, from 19.871 to 20.337.

Simulation Model for different values of betas for Wealth in the new Optimization problem



(a) Line plot for different values of betas for Wealth

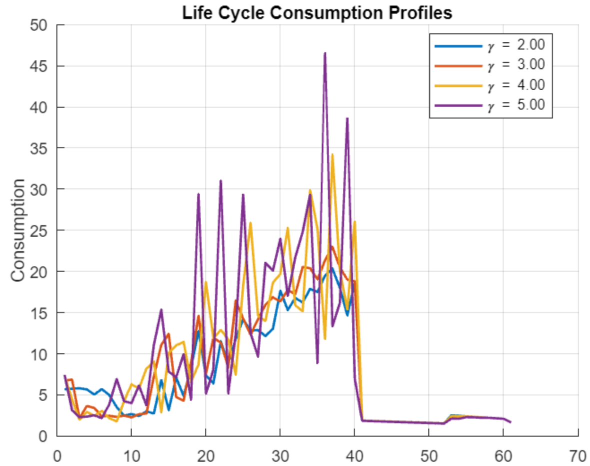
Beta	Mean	Std	Min	Max
0.9	11.094	6.912	0	29.984
0.92	11.687	6.7967	0	30
0.94	12.718	7.6878	0	30
0.96	12.713	7.6695	0	30

(b) Descriptive table for different values of betas for Wealth

Figure 18: Simulation Model for different values of betas for Wealth in the new Optimization problem

The simulated model shows the change in wealth over the life cycle of a household with different values of the discount factor beta (0.90, 0.92, 0.94, and 0.96), the value reflecting the level of patience, while gamma is kept the same. Based on the graph, the trend line of the new model shows more fluctuations than the original model without working hours. When the old model reflects that wealth tends to decrease from the beginning of the cycle to the midlife stage and then gradually increases. While the trend line of the new model decreases sharply in the first 5 years of the cycle, it increases in the midlife stage, accumulating and leveling off before peaking in the pre-retirement stage. However, the graph in the new model does not clearly show that more patience will help households accumulate more assets. On the other hand, looking at the descriptive table, the average wealth is 11,094 when beta is 0.90, and increases to 12,713 when beta is 0.96. At the same time, from beta 0.92 and above, the maximum wealth measured reaches the highest value of the simulation. This confirms the descriptive table that higher consumption will give households more wealth over their lifetime.

Simulation Model for different values of gammas for Consumption in the new Optimization problem



(a) Line plot for different values of gammas for Consumption

Gamma	Mean	Std	Min	Max
2	7.2546	6.0315	1.5467	20.337
3	7.9835	7.0503	1.5401	22.981
4	8.9127	8.5055	1.5418	34.167
5	9.4251	10.456	1.519	46.495

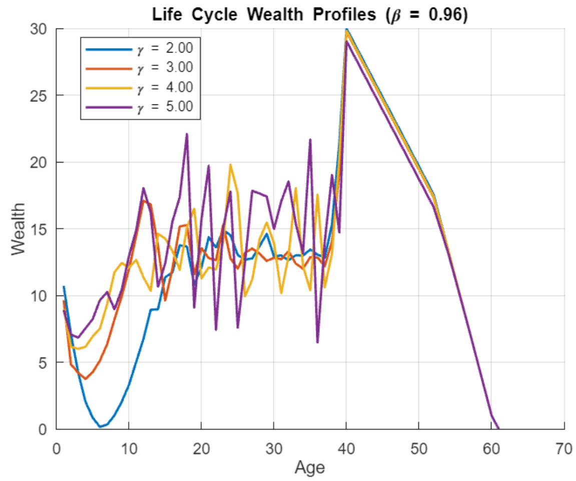
(b) Descriptive table for different values of gammas for Consumption

Figure 19: Simulation Model for different values of gammas for Consumption in the new Optimization problem

The new graph shows the change of the average consumption level of the new optimization problem with gamma values (2, 3, 4, 5), which reflects the level of risk aversion. From the simulated graph using the new model, the consumption level is higher when the gamma value is higher and vice versa. This trend is reinforced by the data from the descriptive model: the average consumption level increases from 7.2546 with gamma equal to 2 to 9.4251 when gamma equals 5, while the simulated maximum consumption value also has a sharp increase from 20.337 to 46.495. This result shows that households with higher risk aversion will have higher consumption when they are risk-averse. Moreover, compared with the original model, the pattern line of the new model, although witnessing many fluctuations, still follows an overall upward trend from the beginning of the period to before retirement, in contrast to the decreasing trend seen in the original model.

Important note: In this new optimization model, gamma is used not only to denote the risk aversion of a household but also for leisure hours. So in these simulation results, gamma in here is used to explain the risk aversion factor of a household.

Simulation Model for different values of gammas for Wealth in the new Optimization problem



(a) Line plot for different values of gammas for Wealth

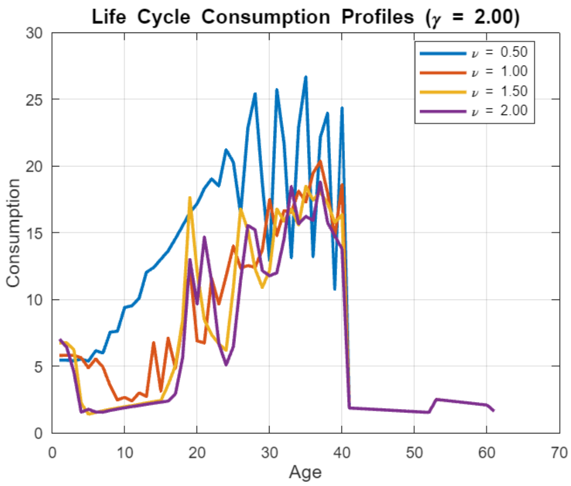
Gamma	Mean	Std	Min	Max
2	12.713	7.6695	0	30
3	13.652	6.72	0	29.772
4	14.064	6.593	0	29.833
5	14.645	6.5656	0	29.049

(b) Descriptive table for different values of gammas for Wealth

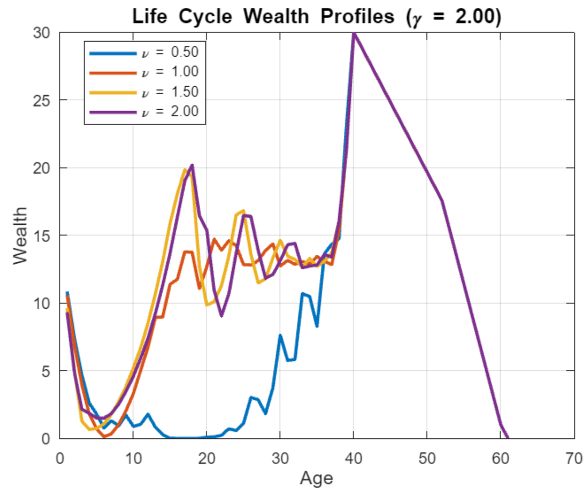
Figure 20: Simulation Model for different values of gammas for Wealth in the new Optimization problem

In line with the measured trend for consumption, the accumulated wealth of the model also increases as the value of gamma increases. Overall, both the model with and without working hours show an increasing trend and reach their highest value in retirement. The statistics also confirm this, showing that the average wealth increases from 12,713 at gamma 2 to 14,645 at gamma 5. This can be explained by the fact that higher risk aversion or caution will help families prioritize saving money, leading to more accumulated wealth compared to less risk-averse households.

Sensitivity Analysis of Consumption and Wealth Response to Income Changes Using ν



(a) Line plot for different values of nu for Consumption



(b) Line plot for different values of nu for Wealth

nu	mean_consumption	std_consumption	min_consumption	max_consumption	mean_wealth	std_wealth	min_wealth	max_wealth
0.5	10.449	8.2238	1.5467	26.672	8.8749	9.3895	0	30
1	7.2504	6.0237	1.5467	20.345	12.718	7.6878	0	30
1.5	6.6607	6.0849	1.4104	18.481	13.187	7.7316	0	29.984
2	6.2511	5.7415	1.5465	18.809	13.169	7.7081	0	29.993

(c) Descriptive table for different values of nu for Consumption and Wealth

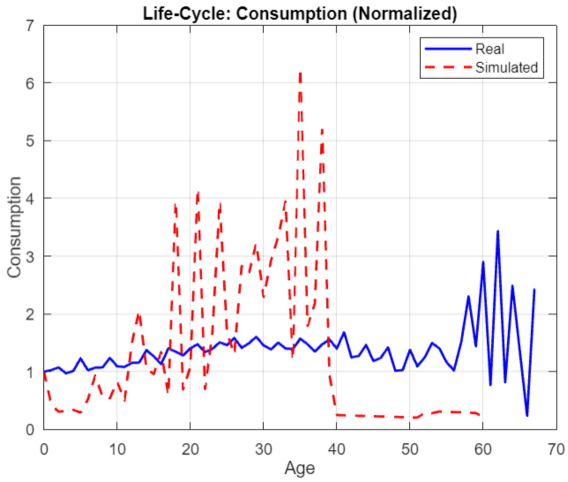
Figure 21: Sensitivity Analysis of Consumption and Wealth Response to Income Changes Using ν

To test the sensitivity of adding working hours to the model, I adjusted two variables to test this new simulation model: γ, the level of leisure time preference of each household, and ν, the inverse of the Frisch elasticity of labor supply, which reflects the effect of income on changes in hours worked. Here, I introduce a new variable, ν, which captures the extent to which households are willing to trade off working hours for changes in income. When ν is high, households are reluctant to change their working hours regardless of whether income increases or decreases. For example, when income increases, households may prefer to work more hours, but this depends on the preferences of the household.

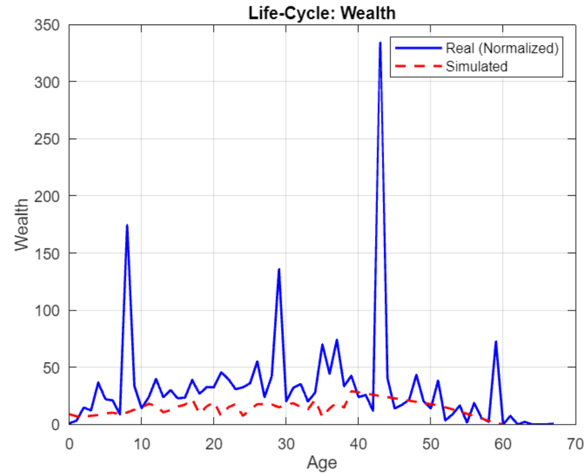
The first graph shows the average consumption of households over the life cycle with different values of ν (0.5, 1.0, 1.5, 2.0). The average consumption of households increases gradually from age 45 to before retirement. Households with lower ν values show higher average consumption, and vice versa. This is supported by the descriptive table, where the average household consumption value decreases as ν increases. This might be because as ν decreases, households are more flexible with their working hours as income changes, making them more willing to work more as income increases

The second figure shows average wealth with ν of 0.5, 1.0, 1.5, and 2.0. Accumulated wealth increases gradually and peaks at the retirement age throughout the cycle. Higher ν values represent higher total wealth for households. In the descriptive model, mean wealth increases gradually, from 8.8749 when $\nu = 0.5$ to 13.169 when $\nu = 2.0$. This may be because as households become less flexible with their working hours, they tend to save more and accumulate more assets. When they are comfortable with changing their hours, they save less, because they can now increase their income from working more hours instead of saving more.

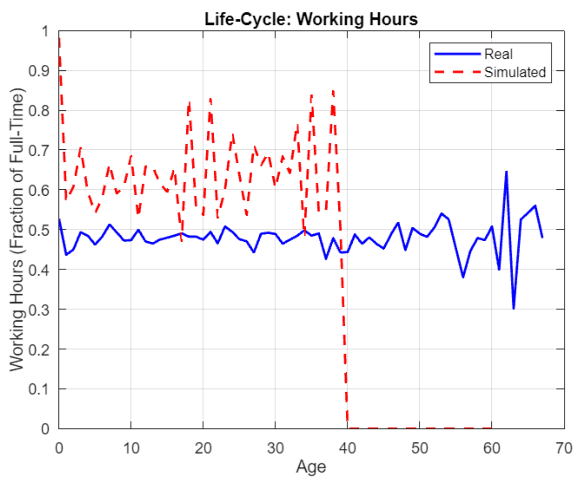
Model Validation Against Empirical Data



(a) Simulated vs Real data: Consumption



(b) Simulated vs Real data: Wealth



(c) Simulated vs Real data: Labor choice

Figure 22: Simulated model compared to the real data from VHLSS data set

	Mean_Real	Std_Dev_Real	Mean_Sim	Std_Dev_Sim
Consumption	1.3777	0.46192	1.2679	1.4066
Wealth	33.177	46.943	14.645	6.5656
Working Hours	0.47912	0.041758	0.42012	0.31991

Figure 23: Descriptive table for Simulated and Real data

In the last section, I compared the simulated data with the real data collected in the VHLSS model, noting that in this comparison, the age on the x-axis must be increased by 25.

First graph: This graph compares real (red-dashed line) and simulated (blue solid line) consumption for households throughout the life cycle. For the real data, consumption increases sharply from midlife to pre-retirement, then begins to decline sharply in the following years. The simulated line in general follows the same trend; however, it stays consistently lower than the real data and shows very small fluctuations compared to the real data throughout the life cycle. Overall, the simulated line—or the model—underestimates both the level and the volatility of consumption compared to the real data.

Second graph: This graph compares real (red-dashed line) and simulated (blue solid line) wealth for households throughout the life cycle. For the real data, wealth increases gradually from the beginning of the period, peaks at pre-retirement age, and then declines gradually in the following years. The simulated line for wealth behaves similarly to that of consumption: it remains consistently below the real data line and shows less fluctuation. Based on this, we can conclude the same for wealth—that the simulated line from the model underestimates both the accumulation and the volatility of wealth compared to the real data.

Third graph: This graph compares real (red-dashed line) and simulated (blue solid line) working hours for households throughout the life cycle. For the real data, working hours remain relatively flat throughout the period, with small fluctuations. However, the simulated line is above the real data line, indicating higher simulated working hours, along with greater fluctuations. The simulated line also stops exactly after retirement. Overall, the model overestimates working hours and their volatility, likely due to the parameters ν or γ.

Question 2

Data Processing

To create a new dataset for analysis, the Enterprise Survey data was combined for the years 2015 and 2005 to analyze variable inputs, costs, and to categorize the size of the enterprise (large and small).

Firm size categorization

First, to distinguish firms in different years, we take the identifier of that company for that year: `idstd` is for the global unique identifier, and `id` is for the country unique identifier (for example, `idstd2015` and `id2015` are for the global unique identifier and country unique identifier, respectively). Since the dataset I want to create consists only of firms for 2 years, we will only take firms that have `idstd2015` and where the `id2015` index is not empty; the same method is applied for `idstd2005` and `id2005`.

The next stage is to determine whether the company is a large company or a small company. This process involves 2 steps. I first take the median of all the firm sizes (`l1`) collected in the Enterprise Survey for all years. Then, following the questionnaire in the survey, it states that if the firm has more than or equal to 20 employees, it is considered medium-sized, and if it has more than or equal to 100 employees, it is considered large. Finally, if it has fewer than 20 employees, it's a small firm. The answer is recorded in the index `a6a`, where 1, 2, and 3 represent small, medium, and large, respectively. However, to simplify this, I only consider large and small firms. The conditions for small firms are the same, while medium and large are now combined to be classified as large only. Additionally, the last step of categorizing is that the final dataset will only include large firms with more than 20 employees and larger than the computed median; if the conditions are not met, it will be classified as a small firm. This results in the total number of large firms being 936, and small firms being 1106.

Variable Selection for the data in 2015 and 2005

For 2015 and 2005, the following table represents the variable costs, production inputs, and financing activities of firms recorded in the enterprise survey.

Table 2: Variable Selection by Year and Category

Year	Index	Description
2015: Costs		
2015	c9b	Annual losses due to power outages
2015	d10	Losses due to theft as % of the value of the products
2015	d11	Losses due to breakage or spoilage as % of the value of the products
2015	i2b	Total annual cost of security
2015	i4b	Total annual value of losses due to theft, robbery, vandalism
2015	n2a	Total annual cost of labor
2015	n2b	Total annual cost of electricity
2015	n2e	Total annual cost of raw materials and intermediate goods used in production
2015	n2f	Total annual cost of fuel
2015	n2i	Total annual cost of sales (for retails)
2015: Production Inputs		
2015	n5a	Purchase of new or used machinery, vehicles, and equipment
2015	n5b	Lands & buildings
2015	_2015_h8	Cost of formal research and development activities (innovation activities)
2015	l1	Permanent, full-time workers at the end of last fiscal year
2015: Financing		
2015	k1c	% Purchased on credit (loans)
2015	k2c	% Sold on credit (receivables)
2005: Costs		
2005	_2005_q86a1	Total sales
2005	_2005_q86a3	Total purchases of raw materials and intermediate goods
2005	_2005_q86a4	Total cost of labor
2005	_2005_q86a5	Depreciation
2005	_2005_q86a6	Rent on land and buildings
2005	_2005_q86a8	Rent on machinery, equipment, and vehicles
2005	_2005_q86a9	Interest charges
2005	_2005_q86a10	Energy cost
2005	_2005_q86a11	Taxes
2005: Production Inputs		
2005	q91a1	Total fixed assets
2005	q91a7	Total current assets (excluding receivables)
2005	_2005_q87a	Net profits (after tax) in 2004
2005	_2005_q87d	% of net profits reinvested in the establishment
2005: Financing		
2005	q91a12	Receivables

Derived Variables and Calculations

Once the above variables have been selected, some further calculations are required, as noted in the questionnaire from the data. Below is the structure that applies to the dataframe of large firms; the same calculations are also applied to small firms.

For 2015

- `main_activity_sales`: Total sales derived from the main activity or product.

  `df_large['main_activity_sales']` = `df_large['d2']` × `df_large['d1a3']`

- `theft_loss_value`: The monetary value of theft losses.

  `df_large['theft_loss_value']` = `df_large['d2']` × `df_large['d10']`

- `spoilage_loss_value`: The monetary value of spoilage losses.

  `df_large['spoilage_loss_value']` = `df_large['d2']` × `df_large['d11']`

- `total_purchases_2015`: Total capital investment in 2015.

  `df_large['total_purchases_2015']` = `df_large['n5a']` + `df_large['n5b']` + `df_large['`$_2 015\_h8'$]

- `purchased_on_credit`: Investment amount made via credit financing.

  `df_large['purchased_on_credit']` = `df_large['total_purchases_2015']` × `(df_large['k1c'] / 100)`

For 2005

- `reinvestment_2005`: Percentage of the profits reinvested into the firm.

  `df['reinvestment_2005']` = `df['q87a']` × `(df['q87d'] / 100)`

- `total_assets_2005`: Excluding receivable assets from total assets.

  `df['total_assets_2005']` = `df['q91a1']` − `df['q91a12']`

- `total_cost_2005`: Total cost calculated by summing specific cost components:

```
df['total_cost_2005'] = df['_2005_q86a3'] + df['_2005_q86a4'] + df['_2005_q86a5'] +
                        df['_2005_q86a6'] + df['_2005_q86a8'] + df['_2005_q86a9'] +
                        df['_2005_q86a10'] + df['_2005_q86a11']
```

Regression Analysis: Impact of Production Inputs on Firm Debt in 2015

Large firm

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     purchased_on_credit   R-squared:                       0.406
Model:                             OLS   Adj. R-squared:                  0.400
Method:                  Least Squares   F-statistic:                     65.70
Date:                 Thu, 15 May 2025   Prob (F-statistic):           2.64e-42
Time:                         16:38:24   Log-Likelihood:                -11963.
No. Observations:                  389   AIC:                         2.394e+04
Df Residuals:                      384   BIC:                         2.396e+04
Df Model:                            4
Covariance Type:             nonrobust
=============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------
const       1.839e+11   2.99e+11      0.616      0.538   -4.03e+11    7.71e+11
l1         -6.165e+07   1.81e+08     -0.341      0.733   -4.17e+08    2.94e+08
n5a           10.0130      1.390      7.203      0.000       7.280      12.746
n5b           -8.3120     31.065     -0.268      0.789     -69.391      52.767
_2015_h8       0.1611      0.012     13.707      0.000       0.138       0.184
=============================================================================
Omnibus:                     861.113   Durbin-Watson:                   2.001
Prob(Omnibus):                 0.000   Jarque-Bera (JB):          1694583.369
Skew:                         16.582   Prob(JB):                         0.00
...
l1                 -0.007859
n5a                 0.100602
n5b                -0.012454
_2015_h8            1.000000
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

Figure 24: OLS regression between production inputs and debt for Large Firms

To show the relationship between production inputs and debt for firms, in 2015, I take the OLS regression test to model `purchased_on_credit` (debt) as the dependent variable, and the independent variables are `l1`, `n5a`, `n5b`, and `_2015_h8` (production inputs).

The result of this regression has shown that 0.406 of the variance in debt is explained by the selected variables for production inputs. Adjusted R-squared is slightly lower than R-squared, at 0.4. F-statistic: 65.70 (Prob(F-statistic): 2.64e-42) has a very small corresponding p-value, almost equal to 0, showing that the regression is statistically significant, meaning that the selected independent variables have a significant relationship with the dependent variable.

- For `l1` (permanent, full-time workers at the end of last fiscal year), the coefficient is `-6.165e+07`, and the p-value is `0.733`, meaning that as the number of workers increases, this might reduce debt for the firm; however, this number is not statistically significant.
- For `n5a` (purchase of new or used machinery, vehicles, and equipment), the coefficient is `10.0130`, and the p-value is < 0.001, meaning that as `n5a` increases, this might increase debt for the firm, and this value is critically significant.
- For `n5b` (lands & buildings), the coefficient is `-8.3120`, and the p-value is `0.789`, meaning that as `n5b` increases, this might reduce debt for the firm; however, this number is not statistically significant.
- For `_2015_h8` (R&D Costs), the coefficient is `0.1611`, and the p-value is < 0.001, meaning that as `_2015_h8` increases, this might increase debt for the firm, and this value is intensely significant.

Small firm

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     main_activity_sales   R-squared:                       0.095
Model:                             OLS   Adj. R-squared:                  0.089
Method:                  Least Squares   F-statistic:                     15.64
Date:                 Thu, 15 May 2025   Prob (F-statistic):           3.55e-12
Time:                         18:09:00   Log-Likelihood:                -15633.
No. Observations:                  600   AIC:                         3.128e+04
Df Residuals:                      595   BIC:                         3.130e+04
Df Model:                            4
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       1.867e+10   2.18e+09      8.558      0.000    1.44e+10     2.3e+10
l1          1.868e+08   2.42e+07      7.734      0.000    1.39e+08    2.34e+08
n5a            0.5765      0.571      1.010      0.313      -0.544       1.697
n5b            0.1737      1.768      0.098      0.922      -3.298       3.645
_2015_h8   -1.428e-05   3.26e-05     -0.438      0.662   -7.84e-05    4.98e-05
==============================================================================
Omnibus:                      741.375   Durbin-Watson:                   1.854
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            78509.986
Skew:                           6.091   Prob(JB):                         0.00
Kurtosis:                      57.699   Cond. No.                     6.69e+13
==============================================================================
```

Figure 25: OLS regression between production inputs and debt for Small Firms

The OLS regression result for production inputs and debt is not as significant as that for large firms, which means that the impact of production inputs on the debt of small firms is less than that of large firms.

The result of this regression has shown that only 0.095 of the variance in debt is explained by the selected variables for production inputs, which is significantly smaller than the OLS regression for large firms. The adjusted R-squared is slightly lower than the R-squared, at 0.089. The F-statistic is 15.64 (Prob(F-statistic): 3.55×10^{-12}), which has a very small corresponding p-value, almost equal to 0. This shows that the regression is intensely significant, meaning that the selected independent variables have a significant relationship with the dependent variable.

- For `l1` (permanent, full-time workers at the end of the last fiscal year), the coefficient is 1.868×10^8, and the p-value is < 0, meaning that as the number of workers increases, this might increase debt for the firm. This result is statistically significant.
- For `n5a` (purchase of new or used machinery, vehicles, and equipment), the coefficient is 0.5765, and the p-value is 0.313, meaning that as `n5a` increases, this might increase debt for the firm; however, this value is not statistically significant.
- For `n5b` (lands & buildings), the coefficient is 0.1737, and the p-value is 0.922, meaning that as `n5b` increases, this might increase debt for the firm; however, this result is not statistically significant.
- For `2015_h8` (R&D Costs), the coefficient is -1.428×10^{-5}, and the p-value is 0.662, meaning that as `2015_h8` increases, this might reduce debt for the firm; however, this value is not statistically significant.

Why there is such a gap between the OLS results for small and large firms:
For small firms, it may be because they have fewer resources, and for a small firm, their most important resource is their employees. So we see that they are less dependent on other items such as infrastructure, machinery, and R&D, especially on credit.

Calculate Key Inputs for Dynamic Firm Model

So far, we have collected all the variable inputs and categorized those inputs into small and large firms for 2005 and 2015. In the next step, to find the average input quantities of production and investment for firms in 2015 and 2005, we go through a series of steps: first, finding data from other sources for the prices of production to calculate the average price of production; next, calculating average variable costs; then taking the average variable cost divided by the average price for the average input quantities of production.

Below table is the information on the prices for inputs production.

Table 3: Price Inputs for Calculating Average Annual Cost (Vietnam, 2015)

Variable	**Value (VND)**	**Source / Notes**
Average Electricity Price (2015)	25,008	2084 VND/month × 12 months. Source: *CEIC Data* — `https://www.ceicdata.com/en/vietnam/average-retail-price-ho-chi-minh-city/average-retail-price-hcmc-electricity`
Average Annual Wage (2015)	56,400,000	4,700,000 VND/month × 12 months. Source: *Vietnam Labour Force Survey 2015*, Ministry of Planning and Investment – General Statistics Office
Producer Price Index (2015)	132.350	Index value (no unit). Source: *CEIC Data* — `https://www.ceicdata.com/en/vietnam/producer-price-index-2010100/producer-price-index-ppi-industrial`
Average Annual Price (Used as input price)		Calculated as the average of the three inputs: $\frac{\text{Wage+Electricity+PPI}}{3}$

Table 4: Economic Metrics for Small and Large Firms in Vietnam, 2015

Metric	**Small Firms**	**Large Firms**
Average Annual Price (VND)	18 808 380.116 666 667	18 808 380.116 666 667
Average Variable Cost per Firm (VND)	4 562 044 648.141 666	9 573 890 636.190 231
xt (Quantity)	242.553 830 784 135 56	509.022 604 647 729 4
Average Investment (pt) (VND)	3 100 864 084 999.7	1 661 478 195 372.4275