Master Thesis: Climate change economic effect on health through temperature

Paulo Gugelmo Cavalheiro Dias

May 14, 2025

Contents

1 Introduction

1.1 Introduction to the subject

Climate change is projected to strongly impact temperature distribution in the upcoming century. Therefore, it is important to try to identify the different effects of temperature on the economy. One of the particular effect channels is health. How will temperature change affect the health status of the population, and therefore the economic production?

1.2 Related literature

In order to study this question,

The relationship between temperature, health, and economics is deeply intricated in several literature fields.

First, since the 2000s, the relationship on how climate and economics are intertwined has been studied in a more climate-related literature.

Second, health economics has identified important findings in the last years.

- Health economics: - Health and productivity. - Health and life expectancy. - Climate and economics: - Hotter temperature and conflicts.

1.3 Research question and strategy

The goal of this Master Thesis is to propose a simple approach to model the effect of temperature on the health of individuals, and subsequently on the economy.

In order to achieve this, two main parts are identified.

First, an empirical part studying the link between temperature and health will be done.

Second, a macroeconomic model is presented to explain the economic mechanisms that will be affected by the identified empirical relationships.

2 Setting

This section is dedicated to the presentation of the general setting in which individuals live in this model. First, the relationships between the different elements in a general framework will be presented. Then, the data used will be presented and discussed. The methods used to estimate the functional forms of the relationships will then be presented, and will be refered as the specific case of this work. Finally, results of the estimation process of these relationships will be presented and discussed.

2.1 Goal

2.2 Formal Description

At each period, an exogenous weather realization occurs, and individuals draw a health and living status.

2.2.1 History vectors

In a general framework, we can think of the health of an individual at time t as a vector $\mathcal{H}_t \in \mathbb{R}^t$ containing all the health status of the individual throughout their life. Similarily, we can think of the weather experienced by an individual at time t as a vector $\mathcal{W}_t \in \mathbb{R}^t$ containing all the weather conditions experienced by the same individual throughout their life.

2.2.2 Health Status

Let $H_t \in \Omega(H)$ be a random variable denoting the health status of an individual at time t. The functional form of its distribution f_h will depend on its sample space $\Omega(H)$. The past health history and the temperature also affect the probability distribution of health status. Generically, we can therefore write:

$$H_t \sim f_h(\mathcal{H}_{t-1}, \mathcal{W}_t) \tag{1}$$

Also, it follows that:

$$\mathcal{H}_t = (H_1, ..., H_t) \tag{2}$$

2.2.3 Living Status

Let $L_t \in \{0, 1\}$ be a random binary variable denoting the living status of an individual at time t. It is determined by a Bernoulli distribution with parameter p_t such that:

$$L_t \sim \mathcal{B}(p_t) \tag{3}$$

The probability parameter p_t depends on their health, age, and temperature. In a general approach, we can rewrite the first equation such as:

$$L_t \sim \mathcal{B}(p_t(\mathcal{H}_t, \mathcal{W}_t)) \tag{4}$$

As such, the probability of an individual to be alive at period t is:

$$Pr(L_t = 1 | \mathcal{H}_t, \mathcal{W}_t) = \prod_{j=1}^{t} p_j(\mathcal{H}_j, \mathcal{W}_j) \tag{5}$$

2.3 Data

Three main datasets were used to estimate these relationships: the Health and Retirement Study (HRS) dataset for health status and survival, the Berkeley Earth dataset, and finally the Federal Reserve Bank of Saint Louis (FRED) dataset for other economic variables.

2.3.1 HRS Data

The HRS has a main survey performed every two years on a panel of individuals in the United States of America (USA). An exit survey occurs in parallel, that targets individuals identified as dead, in which questions are asked to relatives. The exit survey was used to identify dead individual, for which the living status was noted as 0 at the year of the survey. Dead individuals were kept in the final analyzed dataset if they had been observed in the immediate previous survey. To avoid mortality effects of the COVID-19, the latest year selected for this study was 2018. Due to different encoding, the 2000 and prior surveys were not taken into account.

Four variables were used in this dataset: Year of the survey, age of individuals, living status, and health status. The age of individuals was determined based on the year of birth question, that was substracted to the year of the survey. Living status was encoded as 1 if an individuals was present in the main survey, and 0 if they were in the exit survey. The health status of dead individuals was encoded as the same as in the previous survey. The health status was identified through the self-reported health. This measure is less precise than alternative composites, based on weighted average of specific questions of health, but perform well enough and is more tractable.

In the HRS data, health status can take eight possible values:
- 1: Excellent
- 2: Very Good
- 3: Good
- 4: Fair
- 5: Poor
- 8, -8, or 9: Non Available, Not answered, or refused to answer[1]

[1] This last category was removed from the final analyzed dataset.

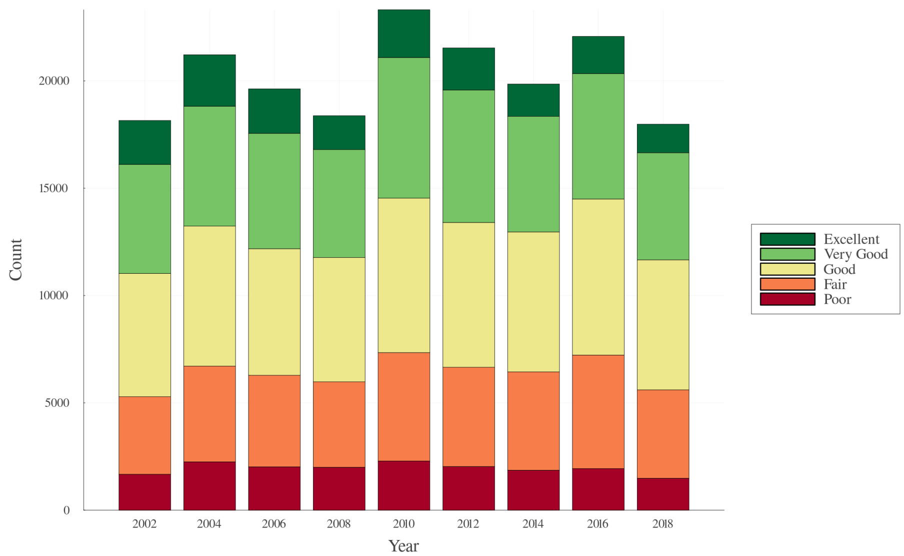

Figure 1: Health Status distribution per Year, from the HRS data

2.3.2 Climate Data

The climate data of Berkeley Earth was used to collect information on average annual or pluriannual global temperature on land. The name of the chosen dataset was the "Land Monthly Average Temperature". For each month, from 1880 to 2022, two anomaly extrema values are given. They correspond to the 95% confidence interval for the average temperature. The average temperature is computed as a deviation from the average annual temperature computed between Janurary 1951 and December 1980.

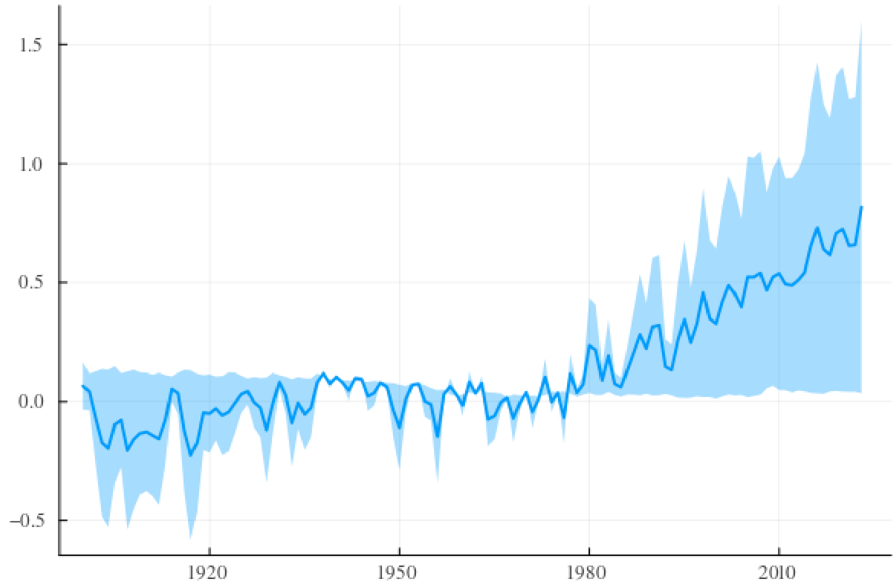

Figure 2: Average annual temperature from the Berkeley Dataset
The light blue area is delimited by the anomalies extrema of each year, and the dark blue line represents the average of the anomalies.

2.3.3 Economic Data

Finally, the FRED data was used to get two economic variables. First, the annual Gross Domestic Product (GDP) of the USA. Second, the interest rate, whose average is then used to calibrate the economic model.

3 Estimation Methods

3.1 Health Transition

The HRS dataset provides rich information regarding the health status of individuals. To make use of the five different values of the self reported health, it was therefore decided not to recode the variables into a binary health status variable $H \in \{Bad, Good\}$.

Several methods were considered to estimate the relationship between current health and past health and other covariates. Given that health is here an discrete, ordinal variable, ordered response models were chosen. More specifically, the ordered logit and ordered probit models were selected, due to their simplicity and broad usage (**Wooldridge˙2010**).

For tractability and computational reasons, it was chosen not to overload the function with history parameters, and just to focus on recent health, and on current temperature. Therefore, $f_h(\mathcal{H}_{t-1}, \mathcal{W}_t)$ is considered as $f_h(H_{t-1}, T_t)$ from now onwards.

Since $\Omega(H_t) = [\![1, 5]\!]$, we can consider f_h as a categorical distribution function[2]. It can therefore be written as:

$$f_h(H_{t-1} = j, T_t) = f_{h,j}(T_t),\ \forall j \in [\![1, 5]\!] \tag{6}$$

To estimate $f_{h,j}(T_t)$, i.e. the probabilities to go to another health state given that $H_{t-1} = j$, a first naive approach would consist in running an ordinal logistic regression with H_t as the dependent variable, and the with covariates including the age, temperature, and control variables.

There are two main issues with this regression. First, there is a large amount of omitted variables affecting the survival probability in this formulation. Second, when we try to include economic variables reflecting the progress in medicine, economic development, or other kinds of control to take into account possible omitted variables, we are faced with a colinearity issues with temperature.

Another important element to take into account to estimate $f_h(\cdot)$ are the interaction effects between the different covariates. For example, it seems plausible that age and previous health status interact: Being in "Fair" health should not have the same effect on the health transition probability for a 20 years old individual compared to a 80 years old individual.

To tackle these problems, it is possible to use a IV-based approach to try to isolate the effect of temperature on health. To do so, let us define a Health Proxy ($HP_{i,t}$) as the sum of binary variables at time t for individual i indicating if they have health issues. I retained four possible health accident, that can all be related to exposure to high levels of ambient temperature, such that:

$$X^h_{i,t} = \begin{bmatrix} \text{High Blood Pressure}_{i,t} \\ \text{Lung Disease}_{i,t} \\ \text{Hearth Condition}_{i,t} \\ \text{Stroke}_{i,t} \end{bmatrix}$$

Formally, the Health Proxy can therefore be written as:

$$HP_{i,t} = \sum_{j \in X^h_{i,t}} j \tag{7}$$

We can thus first run the following linear regression, to estimate the marginal effect of temperature on health through the above mentioned health accidents:

$$\widehat{HP}^I_{i,t} = \widehat{\beta_0} + \widehat{\beta_A} \cdot Age_{i,t} + \widehat{\beta_T} \cdot Temperature_t \tag{8}$$

[2]Also called generalized Bernoulli distribution, that can be represented as a Markov transition matrix.

Once we have the estimate $\widehat{HP}^I_{i,t}$, we can then run a second regression, with the transition probability as the dependent variables, and with $H_{i,t-1}$ and $\widehat{HP}^I_{i,t}$ as covariates.

3.2 Living Status

The living status variable, being binary, a simple logistic regression was possible.

To estimate $p_t(\mathcal{H}_t, \mathcal{W}_t)$, the first naive approach would have led to the same problems as previously seen. For example, running the non-IV regression with X containing only GDP, we find a negative coefficient on temperature, but also with GDP. This is due to the recent tendency in the USA in which the life expectancy stays stable or decrease slightly, while the GDP continues to raise importantly.

With $\Lambda(\cdot)$ being the logistic function, we could therefore run the following logistic regression to estimate the probability parameter $p_{i,t}$, i.e. the survival probability of i at t, with the previous estimate $\widehat{HP}^I_{i,t}$:

$$\widehat{p_{i,t}} = \Lambda\left(\widehat{\beta_0} + \widehat{\beta_H} \cdot Health_{i,t} + \widehat{\beta_{HP}} \cdot \widehat{HP}^I_{i,t}\right) \tag{9}$$

3.3 Estimation Results

The first regression to explain the Health Proxy yields:

	HP
(Intercept)	-0.487***
	(0.068)
Age	0.020***
	(0.001)
Temperature	-0.032
	(0.123)
Age × Temperature	0.005**
	(0.002)
N	182,947
R^2	0.086

Table 1: Regression of Health Proxy on Age, Temperature, and their interaction.

3.3.1 Health transition

The regression to estimate the health transition probabilities yields.

PLOT

From these estimates, it is possible to simulate the collective health trajectory of a population.

PLOT

3.3.2 Survival

When taking into account the health transition estimates and the survival probability estimates, it is then possible to visualize the demographic evolution of a population trhough time.

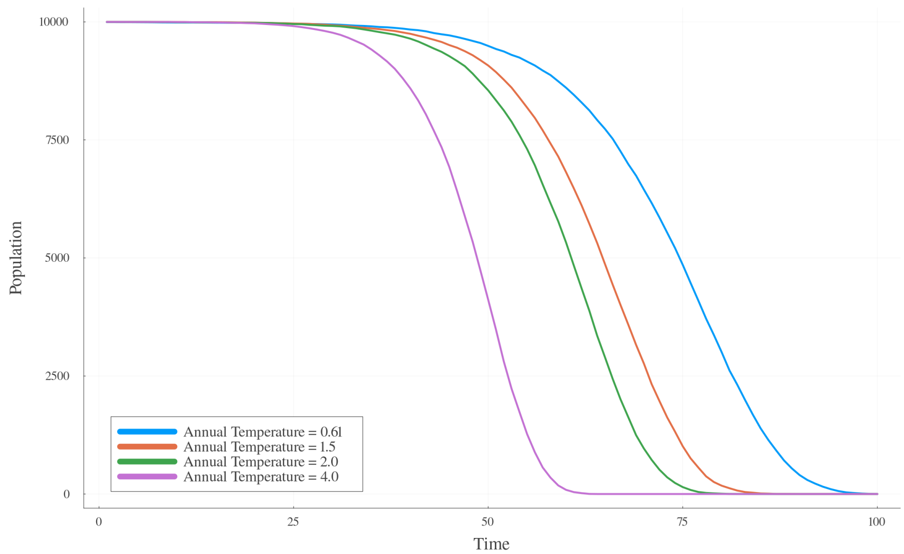

Figure 3: Annual probability of survival as a function of age and health status, obtained with N_0 =10 000

Temperature	Life expectancy
4.0	48.0
2.0	59.86
1.5	63.95
0.61	73.26
$0.5 - 4$	57.33
$0.5 - 2$	64.83
$0.0 - 1.5$	69.29

Table 2: Life expectancy simulations with fixed and variable temperatures.

We see that the effect of temperature on life expectancy is negative. This is due to the double effect of temperature on health transition and on survival probability.

DISCUSSION

These estimates are now going to be used in the economic model.

4 Model

This section is dedicated to the formal description of the model, as well as its analytical analysis. First, its main mechanisms will be explained, and then, the inexistence of analytical solution in most cases will be shown.

4.1 Baseline specification

The agents maximizes:

$$\max_{\{c_t, l_t, s_{t+1}\}_{t=1}^{100}} \mathbb{E}\left[\sum_{t=1}^{100} \beta^t \cdot u(c_t, l_t)\right]$$

Their utility function is:

$$u(c_t, l_t) = \frac{c_t^{1-\rho}}{1-\rho} - \xi_t \cdot \frac{l_t^{1+\varphi}}{1+\varphi}$$

With :
- c_t the consumption
- l_t the quantity of labor supply provided by the agent
- h_t the health status
- w_t the weather variable, which is here temperature
- ξ_t the labor disutility coefficient
- ρ the risk aversion coefficient
- φ the Frisch elasticity

The agent is subject to the following budget constraint:

$$c_t + s_{t+1} \leq l_t \cdot z_t + s_t \cdot (1 + r_t)$$

With:
- c_t the consumption at period t
- s_{t+1} the savings for period $t+1$
- l_t the labor supply provided by the agent at period t
- z_t the productivity at time t
- s_t the savings available at the beginning of period t
- r_t the interest rate at period t

Also, agents are subject to the following borrowing constraint, defined as:

$$s_{t+1} \geq \underline{s}, \forall t \in [\![1, T]\!]$$

We can note the First Order Conditions, such that:

$$c_t^{-\rho} \cdot z_t = \xi_t \cdot l_t^{\varphi} \iff \begin{cases} c_t = \left[\frac{\xi_t \cdot l_t^{\varphi}}{z_t}\right]^{-\frac{1}{\rho}} \\ l_t = \left[\frac{c_t^{-\rho} z_t \cdot}{\xi_t}\right]^{\frac{1}{\rho}} \end{cases} \tag{10}$$

And

$$c_t^{-\rho} = \beta \cdot \mathbb{E}\left[c_{t+1}^{-\rho} \cdot (1 + r_{t+1})\right] + \gamma_t \tag{11}$$

The first corresponds to the equalization of marginal benefit and cost of labor, and the second corresponds to the Euler equation.

The first equilbrium condition implies an within decision, driven by the labor disutility coefficient ξ and the productivity z. There is a unique mapping between consumption and labor at any period, to equalize the benefits and the costs of labor.

The second equilibrium condition implies an intertemporal decision. The marginal utility of consumption at one period must be equal to the expected marginal utility of conusmption next period, discounted by the discounting factor β and the interest rate next period $(1 + r_{t+1})$, plus the marginal benefit of violating the borrowing constraint at the current period.

It is now important to describe what the Expectation operator $\mathbb{E}$ entails here. In a generic formulation, one could expect the uncertainty to affect the interest rate at the next period, which is the reason $(1 + r_{t+1})$ is within the operator.

Another specification could exclude any uncertainty from the interest rate. The uncertainty could then come from the health and survival draw. If the uncertainty only comes from these two draws, the expectation operator can be formalized such as:

$$\mathbb{E}\left[c_{t+1}\right] \equiv p_{t+1}(\mathcal{H}_t, \mathcal{W}_t) \cdot c_{t+1}$$

In this specification, the only uncertainty is whether the agent will be alive or not in next period. The baseline model will first focus on this simplified assumption. Variations will be introduced later.

4.2 Analytical Solution Inexistence

Proposition This maximization program is impossible to solve analytically in most cases[3].

If we consider the model altogether, it is impossible to describe analytically the optimal policy functions of the three choice variables. While the entire proof is available in the appendix, a quick explanation is possible here. First, the objective functions is linear with the savings at next period s_{t+1}, making it disappear from the F.O.C.s. This term requires therefore the labor and consumption policies to be solved, and then plugged into the budget constraint, to

[3] The proof of this proposition is in the Appendix.

have a solution. However, if we try to solve the two other policy functions, we end up with transcendantal equations of the form $a \cdot x^{\alpha} + b \cdot x + c = 0$, with $\alpha \notin \mathbb{N}$. This transcendantal equations can be overcome with specific combinations of parameters, among which $\varphi = -\rho$, in which case we obtain:

$$EQUATION$$

To allow for more flexibility in the resolution, the model was mainly solved numerically. The next section discusses the different methods used in order to do so.

5 Numerical methods

Several ways have been considered to solve this model numerically. This section is dedicated to the presentation of the different methods used in order to do so. First, the auxiliary functions are presented. Second, the different main algorithms specifications and their performance are presented. Finally, the aggregation methods and different numerical results are discussed.

5.1 Functions

To proceed to a numerical solving of the model, some fundamental functions were used.

- **Budget clearing function:** The budget clearing function computes the amount of non used income for a set of state and choice variables. Since at optimal, the budget constraint is binding, the underlying theoreical result indicates that the budget clearing function should be zero. Given the imprecision of numerical methods, the average budget clearing function was used as a measure of the precision performance of each algorithm.

  It is equal to:

  $$B(s_t, l_t, c_t, s_{t+1}) = l_t \cdot z_t + s_t \cdot (1 + r_t) - s_{t+1} - c_t$$

- **Bellman function:** The Bellman function takes as an argument the value function next period, and maximizes the current utility plus the discounted value function next period.

  It is equal to:

  $$V(s_t) = \max_{\{c_t, l_t, s_{t+1}\} \in \Gamma(s_t)} \{u(c_t, l_t) + \beta \cdot V(s_{t+1})\}$$

  With $\Gamma(s_t)$ the feasibility set given by the state variable s_t.

- **Backwards function:** The backwards function iterates the Bellman function from the last period to the first one. In the case of policy iteration, it iterates over the policy function, and not the value function.

It aggregates the optimal decisions and returns a grid of optimal choices associated to each period and state variable value.

For the pure numerical value function iteration, the backwards function is as following:

```
for t from 100 to 1

    if t is equal to 100
        Bellman next period = Vector of zeros
    end

    for s in the possible set of s
        for c,l,s' in the feasible set of the current s value

            # We compute the budget clearing:

            bc = budget_clearing(c,l,s')

            # If it does not,
            # we set the value function to a very low number.

            if bc < 0

                V[c,l,s',s] = -Inf

            # If it does, we compute the value function
            # for this combination of choice variables.

            else if bc >= 0

                V[c,l,s',s] =
                    utility(c,l,s') +
                    beta * probability of survival *
                    Bellman next period[s']
            end
        end

        # We set the value function for s and t to
        # the maximum value found.

    end

# We set the value function at next period to current one
Bellman_next_period = Value_function[index_s,t]

end
```

5.2 Algorithms

This section details the different algorithms developed to perform a numerical solution of the model.

5.2.1 Pure numerical value function iteration

The pure numerical value function iteration algorithm consists in verifying all possible combination of choice variables for each level of state variable to determine what is the best possible response given a certain amount of state variable [4].

Here, the algorithm goes through all the possible values of c, l, and s', without using any approximation obtained through the FOC mentioned above. This is quite computational-intensive, but has the advantage of not using analytical results, which can lead to approximation depending on the resolution of the ranges used.

5.2.2 F.O.C. approximated value function iteration

The FOC value function iteration algorithms make use of the two FOC derived in the previous section. They are faster by order of magnitudes, but contain more errors.

Note that it is impossible to use the second FOC, i.e. the Euler equation, containing the Lagrangien multiplier γ_t that is not solvable analytically.

5.2.3 Interpolated algorithms

The interpolated algorithms use interpolation techniques to approximate the value of the next period Bellman equation. This interpolation can be implemented in the pure numerical algorithm, and in the FOC-approximated ones.

They allow for a smoother shape of policy function, and are more easily graphically interpretable. However, their results on speed performance is slightly worse, and their results on precision performance is ambiguous.

5.2.4 Policy iteration algorithms

The policy iteration algorithms use a backward function iterating on the policy function, and not on the value function.

[4] The different steps of the algorithms and their source code are available online. The steps and comments of the present work are available here: `https://www.paulogcd.com/Master_Thesis/`, and the documented replication package, coded in Julia, is available here: `https://www.paulogcd.com/Master_Thesis_Paulogcd_2025/`.

6 Results

7 Conclusion

8 References

9 Appendix

9.1 Proof of Impossibility

The associated Lagrangien is:

$$\begin{aligned}\mathcal{L}(c_t, l_t, s_{t+1}; \lambda_t, \gamma_t) = \mathbb{E}\Big[\sum_{t=1}^{100} \beta^t \cdot ((\frac{c_t^{1-\rho}}{1-\rho} - \xi_t \cdot \frac{l_t^{1+\varphi}}{1+\varphi}) \\ + \lambda_t \cdot (l_t \cdot z_t + s_t \cdot (1+r_t) - c_t - s_{t+1}) \\ + \gamma_t \cdot (s_{t+1} - \underline{s}))\Big]\end{aligned} \tag{12}$$

The First Order Conditions are the following:

$$\frac{\partial \mathcal{L}}{\partial c_t} = 0 \iff {c_t}^{-\rho} = \lambda_t$$

$$\frac{\partial \mathcal{L}}{\partial l_t} = 0 \iff \lambda_t \cdot z_t = \xi_t \cdot l_t^{\varphi}$$

$$\frac{\partial \mathcal{L}}{\partial s_{t+1}} = 0 \iff \lambda_t = \beta \cdot \mathbb{E}\left[\lambda_{t+1} \cdot (1+r_{t+1})\right] + \gamma_t$$

We first note that we must obtain a closed-form solution for c_t and l_t to obtain the optimal value of s_{t+1}.

Replacing the expression of λ_t in the two other equation yields:

$${c_t}^{-\rho} \cdot z_t = \xi_t \cdot l_t^{\varphi} \iff \begin{cases} c_t = \left[\frac{\xi_t \cdot l_t^{\varphi}}{z_t}\right]^{-\frac{1}{\rho}} \\ l_t = \left[\frac{{c_t}^{-\rho} z_t \cdot}{\xi_t}\right]^{\frac{1}{\rho}} \end{cases} \tag{13}$$

And

$${c_t}^{-\rho} = \beta \cdot \mathbb{E}\left[c_{t+1}^{-\rho} \cdot (1+r_{t+1})\right] + \gamma_t \tag{14}$$

Assuming that the budget constraint binds, it becomes:

$$c_t + s_{t+1} = l_t \cdot z_t + s_t \cdot (1+r_t) \iff c_t = l_t \cdot z_t + s_t \cdot (1+r_t) - s_{t+1}$$

Plugging this expression into the maximization program, it can therefore be rewritten such as:

$$\max_{\{l_t, s_{t+1}\}} u(l_t) = \frac{[l_t \cdot z_t + s_t \cdot (1 + r_t) - s_{t+1}]^{1-\rho}}{1-\rho} - \xi_t \cdot \frac{l_t^{1+\varphi}}{1+\varphi} \tag{15}$$

The F.O.C implies:

$$l_t^{\varphi} \cdot \xi_t = [l_t \cdot z_t + s_t \cdot (1 + r_t) - s_{t+1}]^{-\rho} \cdot z_t \tag{16}$$

We can develop the decomposition of consumption if and only if $\rho \in \mathbb{N}$. Indeed, this equation is of form $x = (x - \alpha)^{\beta} \cdot z$. With $\beta \notin \mathbb{N}$, is a transcendantal equation.

If we try to solve it backwards, we now go to the last period. At the last period, $s_{t+1} = \underline{s}$ for sure: Since there is no future, the agent will borrow as much as they can, or will at least not save anything more than what is imposed by the constraint.

For simplification, let $\underline{s}$ be fixed such that: $\underline{s} = 0$. The new optimality condition is:

$$l_t^{\varphi} \cdot \xi_t = [l_t \cdot z_t + s_t \cdot (1 + r_t)]^{-\rho} \cdot z_t \tag{17}$$

This is still a transcendantl equation, and the problem remain the same.

Special cases:

Now, if we plug the F.O.C. 2, (equation 7), in the budget constraint, we obtain:

$$\begin{cases} c_t = \left[\frac{\xi_t \cdot l_t^{\varphi}}{z_t}\right]^{-\frac{1}{\rho}} \\ c_t = l_t \cdot z_t + s_t \cdot (1 + r_t) - s_{t+1} \end{cases}$$

$$\iff$$

$$\left[\frac{\xi_t \cdot l_t^{\varphi}}{z_t}\right]^{-\frac{1}{\rho}} = l_t \cdot z_t + s_t \cdot (1 + r_t) - s_{t+1}$$

$$\iff$$

$$l_t^{-\frac{\varphi}{\rho}} \cdot \left(\frac{\xi_t}{z_t}\right)^{-\frac{1}{\rho}} = l_t \cdot z_t + s_t \cdot (1 + r_t) - s_{t+1}$$

$$\iff$$

$$l_t^{-\frac{\varphi}{\rho}} \cdot \left(\frac{\xi_t}{z_t}\right)^{-\frac{1}{\rho}} - l_t \cdot z_t - s_t \cdot (1 + r_t) - s_{t+1} = 0$$

This is a transcendantal equation of form $x^{\alpha} \cdot b - x \cdot y - a = 0$

Which admits a solution if and only if: $\frac{\varphi}{\rho} \in \mathbb{N}$