Economic Studies 213

Ao Fei
 Essays on Taxation, Externalities, and Poverty Traps
Department of Economics, Uppsala University

Visiting address: Kyrkogårdsgatan 10, Uppsala, Sweden
Postal address: Box 513, SE-751 20 Uppsala, Sweden
Telephone: +46 18 471 00 00
Telefax: +46 18 471 14 78
Internet: http://www.nek.uu.se/

ECONOMICS AT UPPSALA UNIVERSITY

The Department of Economics at Uppsala University has a long history. The first chair in Economics in the Nordic countries was instituted at Uppsala University in 1741.

The main focus of research at the department has varied over the years but has typically been oriented towards policy-relevant applied economics, including both theoretical and empirical studies. The currently most active areas of research can be grouped into six categories:

* Labour economics
* Public economics
* Macroeconomics
* Microeconometrics
* Environmental economics
* Housing and urban economics
Ao Fei

Essays on Taxation, Externalities, and Poverty Traps
Abstract

This thesis consists of three self-contained essays.

Essay I characterizes the optimal mix of linear commodity taxation and non-linear income taxation in a dynamic economy where consumption gives rise to positional and environmental externalities. Both externalities are modeled as stock-externalities by reflecting people’s present and past consumption. With non-atmospheric positional externalities, the principle of targeting does not apply, which means that the policy rules for all tax instruments are adjusted in response to the positional externality. The results of the model imply simultaneous motives for corrective taxation, which interacts in an important way.

Essay II presents a poverty trap model at the firm level, which is driven by time capacity constraints, and involves the optimal allocation of time between a basic and a more-productive business. To operate the more-productive business requires additional time input but enhances the ability to process information in the future, which in turn, determines the firm’s time capacity. As a result, a firm with a low ability may encounter challenges in expanding time capacity for further growth. We highlight the importance of passing proficiency thresholds in operating the more-productive business to achieve sustainable growth. The model can explain why aid that increases the ability to process information generates heterogeneous effects on firms, in terms of both short-run and long-run growth.

Essay III analyzes the optimal aid distribution in the presence of poverty traps and funds deficiency. The aid programs include a top-down and a bottom-up policy. The former aims to lift poverty permanently by prioritizing those who are better off in terms of economic prospects; the latter is for a momentary improvement with the opposite prioritization. The trade-off between the long-run and short-run effects shapes the optimal aid distribution, where the current value of lifting poverty from productivity improvement and available funds jointly matter. I characterize the conditions when funding the top-down policy exclusively or two policies simultaneously is optimal. In the latter situation, certain aid receivers have to be left in poverty traps.

Keywords: Optimal Taxation, Externalities, Poverty traps, Aid
To my mother and all who were part of this journey.
Acknowledgements

As I penned the preface of my dissertation, a flurry of emotions engulfed me. Earning a Ph.D. in Economics from Uppsala University marked a pivotal point in my academic journey, a dream nurtured by both me and my family since high school. In retrospective, I realized that doing a Ph.D. was quite different from excelling as a bachelor and master student, as signaled by good academic scores. In my Ph.D. stage, experiencing frequent research setbacks proved to be more common than I initially anticipated. In order to overcome stagnation, dedicating a significant amount of time was essential, but it alone was not sufficient. Moreover, as soon as one challenge was resolved, a new one swiftly emerged. Even when I made progress in the results I had been seeking, it was quite common to face skepticism or doubt. Consequently, it was easy to become frustrated and lose sight of oneself, leading to numerous moments of self-doubt. This self-doubt hampered both creativity and productivity. Such a negative cycle of emotions became one of the primary adversaries to conquer. However, I was lucky because I met many kind people who lifted me up on my Ph.D. journey.

I start with my supervisors, who are my bosses and also friends. I learned from you how to generate research ideas, construct models, articulate those ideas precisely, and collaborate to enhance the quality of research papers. This journey also cultivated my ability to remain calm, perseverant, and accurate. These skills have proven invaluable in my daily life. Daniel Spiro, I am grateful for your supervision and the opportunities you have provided—such as introductions to your co-authors, and facilitating my visit to University of Oslo. Thomas Aronsson, working under your knowledgeable and supportive supervision has been an immense privilege. Your contributions have significantly shaped my academic path, granting me clarity on my research complexities and bolstering my confidence. The fortune of meeting you and benefiting from your expertise, encouragement, and patience has been instrumental in my personal and professional development. In addition, I would like to thank Gaute Torsvik for hosting me at UiO, which was an excellent experience.

The leader of our co-authored project, Jonas Hjort, meeting you at the University of Oslo (UiO) was fortuitous. From the outset, your warm smile and optimism, combined with your diligent work ethic, left an indelible impression. It is an honor to be part of our forward-thinking team. Your guidance, insightful feedback, and support have been instrumental in my personal and professional development. In addition, I would like to thank Gaute Torsvik for hosting me at UiO, which was an excellent experience.
The journey would have been more difficult without families’ support. My deepest gratitude goes to my dear mother. I would not be who I am without you. Since 2006, our journey has been fraught with challenges: from my father’s illness and subsequent passing to family debts, academic pressures, external criticisms of my studies abroad, my elder brother’s relationship struggles, and my own health setbacks. At times, it felt as though we were suffocating under the weight of it all. Yet, these hardships also solidified our bond, fortifying our mutual trust and love. To my dear elder brother, Xiang, I am truly grateful that our relationship has blossomed into something beautiful after overcoming several years of difficulties. Your maturity and kindness are deeply appreciated. Trusting you with our mother’s care comes naturally. Your devotion to her in my absence is invaluable. Because of your support, studying in Sweden has been peaceful. Uncle Chang, your kindness stands out distinctly. Despite the differences in our age and profession, our bond is effortless. Fate has been kind in bringing us together. I wish both you and Aunt Song good health and happiness.

My cherished friends are also unforgettable, Xiaowei Wang, Jian Wang, Jing Guo, Haiyan Chen, Qian Zhang, Ester Hetergård, Dominik Elsner, Georgios Miaris, Anders Lindström, Johan Holmberg, Dennis Hastad and the wonderful cohorts Anton, Elin, Edvin, Simon, Maxi, and Zeynep. The first special shoutout goes to my office roommate, Zeynep, for not only tolerating my noon naps and my preference for keeping the windows shut but also for the countless fun conversations and fika sessions we have shared. The second one goes to Edvin for his generous help with my studies and life guidance in Sweden, including searching for an apartment, moving in and out, and the numerous fun moments we have had together. My dear apartment roommates Tsun-Tsun and Xiaolong, you have witnessed my moments of joy and sorrow. It has been great to have your company.

I extend immense gratitude to my outstanding academic colleagues: Adam, Alice, Caio, Mikael, Fredrik, Hiep, Ola, Per, Spencer, Torben, Tsz Chun, Xue-meng, Yaroslav, and Zunyuan as well as the members of the Uppsala Center for Fiscal Studies and Theory Club. I am equally thankful for the support from my administrative colleagues Ulrika, Johanna, and Nina.

A special nod goes to the encouraging Chinese researchers, including L. Zhou, F. Wang, Y. Zhou, Y.T. Han, M.X. Wang, R. Zhe, D.X. Tian, among others. Additionally, I offer my heartfelt thanks to my high school English teacher, G.Y. Wang, and my middle school chemistry teacher, Y.Q. Sun. Both of you have guided me through various stages of my life.

Living in Sweden introduced me to a healthy work-life balance. I am delighted to have discovered a hobby-volleyball. This newfound routine has significantly enhanced my energy, efficiency, and overall well-being. Despite my beginner skills, comparable to learning the first-order condition for an economist, I was grateful to have been welcomed into the volleyball community. Thank you to my volleyball friends Bryan, David, Flora, Francesca,
Kristina, Maria, Milan, Pedro, Romee, Sanja, Sergei, Tomas, Tereza, Vijay, Yaser and others from GHIF, NNIF and Campus 1477 in Sweden and Norway. Playing alongside all of you has been truly memorable, and the moments we have shared are irreplaceable.

Thank you to all who have supported me along the way. Because of you, I have been able to persevere and silently await the dawn’s light during the long, dark nights. I am excited to see where our paths may take us in the future, and confident that our bond will remain strong.

Uppsala, November 2023
Fei Ao
Contents

Introduction .................................................................................................................. 13

I Optimal Taxation and Multiple Externalities in a Dynamic Economy .................. 19
  1 Introduction .......................................................................................................... 20
  2 Related literature ................................................................................................. 23
  3 The Model ............................................................................................................ 24
  4 The shadow prices of the externalities ............................................................... 32
  5 Optimal commodity taxation ............................................................................ 34
  6 Optimal marginal labor income and capital income taxation ....................... 37
  7 Results and Discussion ....................................................................................... 40
  A Appendix ............................................................................................................. 46

II Time Capacity and Firms’ Economic Stagnation .............................................. 57
  1 Introduction .......................................................................................................... 58
  2 Related Literature ............................................................................................... 60
  3 The model ............................................................................................................ 67
  4 Results .................................................................................................................. 69
  5 Aid ......................................................................................................................... 71
  6 Conclusions ......................................................................................................... 73
  A Appendix ............................................................................................................. 78

III Optimal Aid Distribution and Poverty Traps ................................................... 85
  1 Introduction .......................................................................................................... 86
  2 Model Setup ........................................................................................................ 90
  3 Policy Design ........................................................................................................ 92
  4 Robustness .......................................................................................................... 103
  5 Conclusions ......................................................................................................... 104
  A Appendix ............................................................................................................. 108
  B Supplementary materials to understand top-down prioritization ...................... 125
Introduction

This dissertation explores applied theory within the realms of public and development economics, examining two central themes: optimal taxation, addressed in Essay I, and poverty traps, which are the focus of Essays II and III. It acknowledges individual decision-making while noting the potential misalignment between personal choices and societal welfare. The study concentrates on formulating optimal policy frameworks, as investigated in Essays I and III, and provides a thorough evaluation of interventions, as discussed in Essay II.

The first essay—Optimal Taxation and Multiple Externalities in a Dynamic Economy—examines the implications of tax policies in a complex economic environment where consumption externalities, influenced both by peers’ behavior and environmentally detrimental consumption, intermingle and accumulate. This essay presents an analytical framework featuring a two-tier tax model that accommodates proportional taxes on commodities alongside non-linear taxes on labor and capital income. The model addresses varying income-earning capacities while considering the challenges a social planner faces due to the inability to discern individual ability types, yet maintains a focus on achieving redistributive objectives.

The dynamic framework draws on a growing body of literature dealing with redistribution under asymmetric information in dynamic economies, including works by Ordover and Phelps (1979), Pirttilä and Tuomala (2001), and Brett (1997), among others. This study extends the works of Howarth (1996) and Eckerstorfer (2014) to an intertemporal economic context. It also contributes to optimal redistributive taxation in dynamic economies with multiple externalities, building on Aronsson and Johansson-Stenman (2010), (2014), and Aronsson and Sjögren (2018), who primarily studied single externality scenarios.

1 The concept of taxation in the presence of externalities has its origins in the work of Warming (1911) and Pigou (1920). It has evolved in recent literature, as evidenced by Sandmo (1975), Boskin and Sheshinski (1978), Layard (1980), and Cremer and Gahvari (2001), among others.

2 For dynamic framework, Howarth (2006) is partially related to my research. He focuses on the interaction between income taxes and emission taxes in a dynamic model. The main conclusion is that the appearance of relative consumption concerns leads to higher taxation of carbon dioxide emissions given the pre-existing income tax, which is consistent with relatively high environmental taxes when relative consumption concerns are present. But Howarth (2006) did not look into the interaction between different consumption externalities, nor did he characterize commodity taxes and marginal capital income taxes.
The second and third essays look into the multifaceted issue of poverty. To provide a deeper understanding of chronic poverty, the discussion pivots towards the poverty trap theory and how interventions might fare in the presence of such self-reinforcing systems. In the second essay *Time Capacity and Firms’ Economic Stagnation*, I start the literature review in poverty traps with macro socioeconomics factors where micro agents can do nothing but suffer chronic poverty if in the poverty trap. Then following Ghatak (2015), poverty traps can be further remarked by two sources: external frictions and scarcity-driven mechanisms. While the former refers to elements beyond an individual’s control, the latter is more aligned with behaviors ingrained by scarcity.

Essay II focuses on poverty at a firm level by narrowing its lens to scarcity-driven poverty traps. Here, the focus is predominantly on small firms in developing countries facing growth impediments. Various research studies, including those by Bloom et al. (2010), Hsieh and Klenow (2014), and E. Verhoogen (2021), have highlighted that these firms often encounter challenges that stymie further growth. The barriers they face are manifold, stemming from both production-side restrictions (as noted by McKenzie and Woodruff, 2014; Quinn and Woodruff, 2019) and demand-side factors (as observed by E. A. Verhoogen, 2008; Atkin and Donaldson, 2012; Donaldson and Hornbeck, 2016).

From the demand side literature in explaining why small firms suffer further economic growth, firms in developing countries have the potential to offer goods and services of comparable quality to those in developed countries. It highlights that limited opportunities are what keep firms in low-income countries stuck in economic growth constraints. Building on this strand of thought, Hjort et al. (2020) aim to explain the challenges businesses face when attempting to penetrate new markets due to limited access to pertinent information for Liberian firms, examining how their limited knowledge of market prospects can hinder their expansion.

I developed a firm-level poverty trap model to provide a theoretical basis for the empirical findings of Hjort et al. (2020). My conceptual framework focuses on the limitations of time capacity that prevents firms’ entrance into more challenging yet profitable and productive business-bidding for contracts-by making the acquisition of information costly. The model pinpoints two vital thresholds that firms must surpass to realize lasting performance improvements and achieve sustainable enhancements in operational efficiency. Interventions, such as training programs aimed at enhancing information processing skills and time capacity, are particularly beneficial for firms with experience in bidding for contracts and robust internet access. These benefits can be observed in both the short and long term.

Poverty at an individual level matters as well. United Nations Development Programme (UNDP) (2023) reports that 1.1 billion out of 6.6 billion people in 110 developing countries are grappling with multidimensional poverty. How-
ever, it is worth noting that the actual numbers might be higher, impacted by external factors like the COVID-19 pandemic, the Russia-Ukraine conflict, and the ongoing global climate crisis. The third essay *Optimal Aid Distribution and Poverty Traps* differs from the second essay by not focusing on the mechanisms of firm stagnation and how they influence the varying impacts of aid but welfare analysis for aid receivers in a simplified poverty trap model, where it applies specifically to the external friction aspect of the poverty trap theory at an individual level. This type of poverty trap is empirically supported by Balboni et al. (2022), where individuals lack access to better opportunities, leading to enduring poverty. The concept of “big pushes”, which entails allocating substantial funds to break the cycle of poverty, was introduced to help ultra-poor individuals, as demonstrated effectively by Murphy et al. (1989), Banerjee et al. (2021), and Balboni et al. (2022).

Essay III underscores the importance of extending interventions not just to ultra-poor individuals but also to all poor individuals, including the implicitly poor who are relatively richer than the ultra-poor and are thus approaching the threshold for escaping the poverty trap. This research zeroes in on devising optimal assistance strategies within a constrained funding framework to maximize overall social welfare, where the trade-off between top-down policies for permanent poverty alleviation and bottom-up policies offering temporary relief is decisive. Such an approach is pertinent in real-world situations where policymakers grapple with limited resources while aiming for a substantial impact on poverty reduction. It reveals that optimal aid distribution should consistently give priority to those nearing the poverty line, recognizing the substantial increase in productivity that occurs when individuals overcome the poverty threshold. The implementation of a combined policy strategy depends on the incremental value derived from productivity enhancements in alleviating poverty and the total funds accessible.
References


Essay I. Optimal Taxation and Multiple Externalities in a Dynamic Economy

Acknowledgements: Special thanks to Thomas Aronsson for valuable comments and suggestions. The comments from Chuanzhong Li, Spencer Bastani, Dominik Elsner, Anders Lindström, Giwrgos Miaris, Adam Gill, Francesco Alosa, Jeanne Poulain and UCFS participants are helpful. Additional gratitude extends to Yikai Wang, Per Engström and all others who offered feedback, enhancing the quality of this paper. Writing the manuscript was possible thanks to funding from the project “The Economics of Planetary Boundaries”, FORMAS grant 2020-00371, hosted by Daniel Spiro.
1 Introduction

It is well known that externalities coexist and accumulate over time, but their relation to the theory of optimal taxation remains to be developed. This paper investigates how multiple consumption externalities affect the optimal tax structure in a dynamic economy, with a special focus on the coexistence of environmental and positional externalities.

Following Sandmo (1975), Pirittlä and Tuomala (1997), Cremer et al. (1998), Micheletto (2008) and Wendner (2014), consumption externalities related to environmental degradation and social comparison will be discussed, which is motivated by the fact that different types of consumption can substitute/complement each other and therefore transmit such relation to the interaction between different externalities. The individual consumption of dirty goods lowers environmental quality for everyone in the form of an environmental externality, e.g., global warming caused by excessive consumption of fossil fuels. A continuous rise of pollutants can also be seen in the environment, e.g., the accumulation of greenhouse gas in the atmosphere,\(^1\) indicating that the impact of environmental externalities may persist over time once generated. A growing strand of literature on preferences indicates that doing better than a reference group matters for individual well-being (Solnick and Hemenway, 1998; Johansson-Stenman et al., 2002; Alpizar et al., 2005). Positional preferences are typically modeled by assuming that individuals derive well-being from their consumption relative to that of referent others, where the consumption of a positional good is not merely for satisfying one’s own real needs. A positional consumption externality can be summarized as a behavior in which increasing one’s consumption gives a higher utility for himself or herself but diminishes others’ utility by raising the consumption reference levels. This behavior is identified as Keeping-up-with-the-Joneses when individuals compare their current consumption with that of their peers. When this comparison includes others’ past consumption, it leads to a Catching-up-with-the-Joneses dynamic. Two factors can drive this. Firstly, the lingering impact of status effects over time means past consumption continues to influence current social comparisons. Secondly, the inevitable delay in receiving information about social references, even with media access, implies that comparisons often occur between an individual’s present consumption and others’ past consumption.

\(^1\)In 2022, Statista reported that the average annual level of carbon dioxide in the atmosphere was 418.56 parts per million (ppm), marking a 0.5 percent increase from the previous year. The global carbon dioxide emissions from cars and vans increased by approximately 1.4 percent compared to the previous year, reaching a total of 3.53 billion metric tons. 
My work develops and examines a model of optimal mixed taxation in a dynamic economy. It also examines whether the additivity property described by Sandmo (1975) applies. The additivity property is a canonical principle for designing taxes when there are externalities to correct. The additivity property is a consequence of the principle of targeting and implies that the shadow price of the externality enters additively in the policy rule for the tax on the externality-generating activity, while it does not directly affect the policy rules for the other tax instruments. With a dynamic model, the additivity property has a slightly different meaning by involving an additional tax instrument on capital income. The analysis takes place in a two-group tax model where earning abilities are heterogeneous, which connects to Stiglitz (1982). The set of tax instruments includes proportional commodity taxes and non-linear labor and capital income taxes. Such a mixed tax system is interesting for at least two reasons. First, the joint use of linear commodity and non-linear income taxes is common in industrialized countries. Second, the mixture of these instruments roots in a social planner’s informational ground. This echoes several assumptions: (i) Private income is observable but specific earning abilities are not. The government can, therefore, use a non-linear income tax targeting individual income but cannot apply ability-type specific lump-sum taxes. (ii) Following Eckerstorfer (2014), the government can observe positional consumption at the aggregate level but not necessarily at an individual level, even if it can be observed by peers, such as certain types of clothes, art, and jewelry. I make the same assumption for the consumption of dirty goods. The anonymous transaction makes non-linear commodity taxes infeasible, which leads the designer to resort to linear commodity taxes.

To make the analysis concise without loss of generality, I consider a three-good economy in a dynamic world, where one of the goods is a numeraire that does not generate any consumption externality, and the two remaining goods give rise to negative consumption externalities which are stock-featured, meaning that both the past and current consumption matters for the externalities. One is a positional externality generated by the consumption of a positional good, and the other is an environmental externality caused by the consumption of a dirty good. In a general case with non-atmospheric externalities, meaning that the marginal contribution to an overall externality differs among ability types, the additivity property usually fails. This nature of a

---

2 An alternative assumption is made in Jacobs and De Mooij (2015), where the individual consumption of polluting goods is observable. This assumption enables the social planner to implement a nonlinear commodity tax on the dirty good, in contrast to other models where the consumption of polluting goods may not be observable or easily measured.

3 The research realm in optimal mixed taxation was discussed by Atkinson and Stiglitz (1976), Mirrlees (1976) as well as Edwards et al. (1994). They represent seminal contributions to the literature on optimal mixed taxation, albeit do not address resource allocation problems with externalities.

4 The relevant term “atmospheric” externality was introduced by Meade (1952).
consumption externality indeed matters for the optimal tax structure, which has been verified by Micheletto (2008), Eckerstorfer and Wendner (2013) and Eckerstorfer (2014). I, therefore, allow the positional externality to be non-atmospheric by differentiating the marginal contribution to the social reference level made by low and high-ability consumers. This captures the phenomenon that some consumers are more influential in forming reference levels.

The model results show that a linear tax on the positional good can no longer perfectly target the positional externality, meaning that the other tax instruments will play a supplemental role in correcting the positional externality. Thus, the dirty good tax depends directly on the shadow price of the positional externality in addition to that of the environmental externality, where the shadow prices of the two externalities directly interact over time when determining the optimal tax mix. It indicates that the extreme condition formulated by Micheletto (2008) to maintain the additive property in a commodity tax system is invalid. The condition requires that cross-substitution effects, how a change in the price of one commodity impacts the demand for another commodity, vanish. It may eliminate the term associated with the shadow price of positional externality, but not affect the direct interaction between environmental and positional externalities in the dirty good tax formula. The marginal income tax instruments are also used for correcting the non-atmospheric externality, indicating a direct interaction with other tax instruments in correcting the positional externality.

The present paper contributes to Howarth (1996) and Eckerstorfer (2014) by extending to an intertemporal economy, where Howarth focuses on the correlation between consumption externalities and production externalities in the first-best, while Eckerstorfer examines the interaction between different positional consumption externalities in the second-best framework. While my research is related to Eckerstorfer (2014), it extends the analysis by focusing on a dynamic economy with stock-externalities; and extends the analysis of Aronsson and Johansson-Stenman (2010), (2014) and Aronsson and Sjögren (2018) who focused on optimal redistributive taxation in dynamic economies with one single externality. Building on previous literature, the coexistence of various externalities, which can extend beyond two externalities, over time, is found to be important in my work. These can interact with each other dynamically, influencing the determination of optimal tax. Policy design for the dirty good tax should consistently aim to correct social comparison, even if the cross-substitution effect diminishes to zero. It suggests that the current tax on dirty goods might not be optimal because it overlooks the need to adjust for status-driven consumption in each period. As a general rule, labor and capital income taxation should incorporate the goal of correcting social comparisons.

The remainder of this paper has six sections. Section 2 covers and summarizes the related literature. The model is laid out in section 3. This is followed by discussions of the shadow prices of the externalities in section 4, optimal
commodity taxation in section 5, and optimal non-linear income taxation in section 6. Section 7 provides a general discussion and concluding remarks.

2 Related literature

There is relatively large literature on optimal taxation under externalities. It dates back at least to Warming (1911) and Pigou (1920) who proposed taxes/subsidies to fill the gap between the social marginal benefits and the private marginal benefits of various activities. More recent literature can be seen as Sandmo (1975), Boskin and Sheshinski (1978), Layard (1980), Ng (1987), Bovenberg and De Mooij (1994), Pirttilä and Tuomala (1997), Cremer et al. (1998), Ljungqvist and Uhlig (2000), Cremer and Gahvari (2001), Kopczuk (2003), Micheletto (2008), Aronsson and Johansson-Stenman (2008), (2010).

Sandmo (1975) contributes by integrating externality correction and redistribution in a model of optimal linear commodity taxation. If the externality is atmospheric, the principle of targeting applies. Following Sandmo’s assumption on an atmospheric consumption externality, Pirttilä and Tuomala (1997) found that the additivity property extends to a mixed taxation system where a non-linear income tax and linear commodity taxes are accessible to the government. If the externality is non-atmospheric, Micheletto (2008) shows that the additivity property does not apply except under very specific conditions.

The first strand of studies closest to mine refers to research on optimal redistributive taxation in dynamic economies, such as Aronsson and Johansson-Stenman (2010), (2014) and Aronsson and Sjögren (2018). Compared to a static model, Aronsson and Johansson-Stenman (2010) derive tax implications by paying attention to relative consumption comparisons in an OLG model where relative consumption concerns are modeled such that households are *Keeping-up-with-the-Joneses*. They provide an insightful idea on dynamic consumption externalities, where they analyze how positional consumption alters non-linear income tax formulas. Their main analysis is based on mean-value comparisons, which is the conventional assumption in the literature. Furthermore, two extensions are made by considering within-generation comparisons and upward comparisons. The latter only modifies the tax policy rules for high-ability individuals, because the consumption of low-ability individuals does not generate any social harm. Aronsson and Johansson-Stenman (2014) use an OLG model to analyze *Keeping-up-with-the-Joneses* and *Catching-up-with-the-Joneses* mechanisms simultaneously for designing optimal non-linear income taxes. One of their main results is that relative consumption concerns over own past consumption do not alter the policy rule for marginal income taxation but the comparison with others’ past and current consumption does. Their study shows that the two mechanisms of social comparison have similar policy implications. Aronsson and Sjögren (2018) analyze optimal mixed taxation- a non-linear income tax and a linear com-
modity tax - in an economy where dirty good consumption causes a stock externality with a persisting effect, making the shadow price of the environmental externality forward-looking. Their research examines both dynamic and non-atmospheric externalities, but they do not address multiple externalities. My study extends their research in optimal taxation in a dynamic economy by including multiple externalities.

The other critical reference strand refers to the policy implications of multiple externalities, such as Eckerstorfer (2014) and Howarth (1996). Eckerstorfer (2014) explores an interdependence between multiple positional externalities when deriving an optimal mixed taxation policy formula. His model separates two positional goods from one non-positional good. Each specific positional commodity can only generate one type of positional externality. The measure of reference consumption used by Eckerstorfer is flexible enough to encompass within-group comparisons, meaning that the reference differs by group, as well as a non-atmospheric positional externality. As a result, the shadow prices of different externalities interact with each other and appear in the tax rules for the non-positional good and labor income. However, his contribution of extending the tax policy framework to multi-externality settings takes place in a static economy. My research extends the analysis to a dynamic economy, which yields the possibility of capital income taxation for realizing the social planner’s objectives.

This paper’s main contribution is combining both of these strands of the literature. I, therefore, characterize the optimal commodity tax and income tax structure in a dynamic economy with environmental and positional externalities. This is also reactive to the fact that while there are a large number of studies on optimal taxation under environmental externalities and a reasonably large number of studies on optimal taxation under positional externalities, very few studies combine them to characterize optimal mixed taxation in an intertemporal economy.

3 The Model
Consider two agent-types that live for two periods; each individual works during the first-period of life (young) and retires in the second period (old). They distinguish themselves from one another by their different innate abilities $j$ which could either be a low-ability type $j = L$ or high-ability type $j = H$. Differences in ability are reflected in the pre-tax wages $w^H_t > w^L_t$ in any period. The total population at time $t$, $\sum_j \left(n^j_t + n^j_{t-1}\right)$, which is normalized to one in

---

5Hart (2022) is closer to Howarth (1996) by including pollution and consumption externalities in a static economy. He looks into how labor supply reacts as productivity progresses when there exist status effects by the consumption of a positional good and environmental damage by the production of the positional good. In conclusion, the tax on positional goods can be more essential to reach optimal environmental quality than the tax on pollution emissions.
each period for notational convenience, consists of the sum of young and old individuals across abilities, where \( n_j^t \) is the population share of ability type \( j \) who was born at time \( t \) and assumed to be constant.

### 3.1 Individual behavior

The life-time utility function, as a sum of discounted utility over two periods for an individual of ability type \( j \) born in period \( t \), reads as

\[
U_j^t = U \left( c_{1,t}^j, x_{1,t}^j, y_{1,t}^j, z_{1,t}^j, E_t, R_t^j \right) + \beta U \left( c_{2,t+1}^j, x_{2,t+1}^j, y_{2,t+1}^j, 1, E_{t+1}, R_{t+1}^j \right) = u \left( c_{1,t}^j, x_{1,t}^j, y_{1,t}^j, z_{1,t}^j, E_t, \bar{y}_t \right) + \beta u \left( c_{2,t+1}^j, x_{2,t+1}^j, y_{2,t+1}^j, E_{t+1}, \bar{y}_{t+1} \right)
\]

(1)

where the time separability may not yield a dramatic deviation from economic intuition, since it is only across two periods instead of an infinite time horizon. The instantaneous utility function \( U(\cdot) \) and \( u(\cdot) \) for each type is strictly concave and twice continuously differentiable.

In the first line of (1), \( U(\cdot) \) is increasing in each argument except the environmental externality \( E \), which is the disutility due to the concern over environmental quality. Individuals derive utilities from private consumption (i) of the numeraire good when young, \( c_{1,t}^j \), and old, \( c_{2,t+1}^j \), (ii) of a positional good when young, \( y_{1,t}^j \), and old, \( y_{2,t+1}^j \), and (iii) of a dirty good when young, \( x_{1,t}^j \), and old, \( x_{2,t+1}^j \). The incorporation of these multiple commodities considers the coexistence of both positional and non-positional consumption (Alpizar et al., 2005; Solnick and Hemenway, 1998, 2005) and the consumption of dirty and clean goods (Aronsson and Sjögren, 2018). The dirty good and the positional good are assumed to be normal goods. The numeraire \( c \) represents all other commodities in this analysis. Young individuals enjoy leisure \( z_{1,t}^j \), which is equal to the individual’s time endowment less the time spent working, i.e.,

\[
z_{1,t}^j = 1 - I_{1,t}^j
\]

where the time endowment is normalized to one. Consumers will not work when old such that \( z_{2,t+1}^j = 1 \). The relative consumption can be defined as \( R_{t}^j = y_{1,t}^j - \bar{y}_t \) and \( R_{t+1}^j = y_{2,t+1}^j - \bar{y}_{t+1} \) in young and old life stages.\(^7\)

It could either benefit consumers if their positional consumption outweighs

\(^6\)The assumption that the life-time utility is time separable is helpful in simplifying the process of deriving taxes. The separability indicates that the utility today at time \( t \) is independent of that in the future at time \( t + 1 \). The time separability may not yield a dramatic deviation from economic intuition, since it is only across two periods instead of an infinite time horizon.

\(^7\)Following a bunch of literature to model the mechanism of a social comparison Akerlof (1997), Corneo and Jeanne (1997), Ljungqvist and Uhlig (2000), Bowles and Park (2005) Carlsson et al. (2007) and Aronsson and Johansson-Stenman (2008), etc., I assume that the relative consumption can be described by the difference between the individual’s consumption of the positional good and the economy-wide social reference consumption of this good. Alternatively, ratio comparisons—the individual’s own consumption divided by the referenced consumption-(Boskin
a social reference level \( y^j > \bar{y} \) \((R > 0)\) or harm the agent if their positional consumption falls short of it \( y^j < \bar{y} \) \((R < 0)\).

The second line in equation (1) is a reduced-form utility function explicitly referring to the environmental and positional externalities in this analysis. \( u(\cdot) \) is directly decreasing in reference consumption \( \bar{y} \) and the environmental externality \( E \), and increasing in the other arguments. The reduced form simplifies some of the calculations by modeling the reference of positional consumption as a direct detriment to society, without losing the intuition that if an individual increases his/her consumption, this will increase the reference which, in turn, reduces the relative consumption for all others.

The intertemporal budget constraint for a consumer of ability-type \( j \) at time \( t \) is expressed as

\[
c^j_{1,t} + q_{x,t}x^j_{1,t} + q_{y,t}y^j_{1,t} + s^j_{1,t} = w^j_{1,t}I^j_{1,t} - T_t \left( w^j_{1,t}I^j_{1,t} \right) \]  

\[
c^j_{2,t+1} + q_{x,t+1}x^j_{2,t+1} + q_{y,t+1}y^j_{2,t+1} = (1 + r_{t+1}) s^j_{1,t} - \Phi_{t+1} \left( r_{t+1} s^j_{1,t} \right) . \]  

Equation (2) is the budget constraint for a young individual with ability \( j \). He/she considers the allocation of net labor income, where \( T_t \left( w^j_{1,t}I^j_{1,t} \right) \) is a non-linear labor income tax instrument, between private consumption on \( c^j_{1,t}, x^j_{1,t}, y^j_{1,t} \) and savings \( s^j_{1,t} \). \( q_{x,t} = p_{x,t} + \tau_{x,t} \) and \( q_{y,t} = p_{y,t} + \tau_{y,t} \) are the consumption prices of goods \( x \) and \( y \), respectively. \( q_{x,t} \) and \( q_{y,t} \) are defined relative to the numeraire \( c \). Compared to \( c \), \( x \) and \( y \) are externality-generating commodities and they are taxed. The price is the sum of the production price \( p \) and the commodity tax rate \( \tau \). The same consumer has a similar situation to consider when old at period \( t + 1 \) and is shown in equation (3) where \( \Phi_{t+1} \left( r_{t+1} s^j_{1,t} \right) \) is a non-linear capital income tax. Equations (2) and (3) differ in the source of disposable income: the income for an old consumer comes from previous savings and additional interest income \((1 + r_{t+1}) s^j_{1,t}\) reduced by capital income taxes. This is because labor income taxation takes place when an individual is young, i.e., in his/her first period of life, while capital income taxes are paid when an individual is old, i.e., the period of retirement without any labor supply.

In this optimization problem, the consumer prices, the tax structure (such as the form of the tax functions), and the two externalities are all exogenous to consumers, but consumers can choose how much to consume, to save and how many hours to use for labor supply. The optimal tax problem is formulated in terms of conditional demand functions and conditional indirect utility functions. These functions are defined conditional on work hours and savings.

and Sheshinski, 1978; Layard, 1980) and ordinal ranks (Frank, 1985; Hopkins and Kornienko, 2004) can be used for modeling consumption comparisons. Typically, difference comparisons and ratio comparisons give the same qualitative results.
The two-stage methods from Christiansen (1984) are helpful in deriving them. Specifically, in the first stage, a utility-maximizing consumer with ability \( j \) at time \( t \) makes a direct choice over \( c^j_{1,t}, x^j_{1,t}, y^j_{1,t} \) and \( c^j_{2,t+1}, x^j_{2,t+1}, y^j_{2,t+1} \) subject to the budget constraints

\[
\begin{align*}
b^j_{1,t} &= w^j_{1,t} l^j_{1,t} - T_t \left( w^j_{1,t} l^j_{1,t} \right) - s^j_{1,t} \quad \text{(4)} \\
b^j_{2,t+1} &= (1 + r_{t+1}) s^j_{1,t} - \Phi_{t+1} \left( r_{t+1} s^j_{1,t} \right) \quad \text{(5)}
\end{align*}
\]

where \( b^j_{1,t} \) and \( b^j_{2,t+1} \) denote the income available for consumption in each period. With (4) and (5), I transform private budget constraints at each life stage into

\[
b^j_{1,t} = c^j_{1,t} + q^j_{x,t} x^j_{1,t} + q^j_{y,t} y^j_{1,t} \quad \text{(6)}
\]

and

\[
b^j_{2,t+1} = c^j_{2,t+1} + q^j_{x,t+1} x^j_{2,t+1} + q^j_{y,t+1} y^j_{2,t+1} \quad \text{(7)}
\]

The conditional demand functions for three different types of commodities in two periods are therefore given by

\[
\begin{align*}
\zeta^j_{1,t} &= \zeta \left( b^j_{1,t}, q^j_{x,t}, q^j_{y,t}, z^j_{1,t}, E_t, \bar{y}_t \right) \quad \text{for} \quad \zeta = c, x, y \quad \text{(8)} \\
\xi^j_{2,t+1} &= \xi \left( b^j_{2,t+1}, q^j_{x,t+1}, q^j_{y,t+1}, E_{t+1}, \bar{y}_{t+1} \right) \quad \text{for} \quad \xi = c, x, y \quad \text{(9)}
\end{align*}
\]

In what follows, the conditional indirect utility function for ability type \( j \) individual at time \( t \) over a lifespan reads

\[
V^j_t = \underbrace{v \left( b^j_{1,t}, q^j_{x,t}, q^j_{y,t}, z^j_{1,t}, E_t, \bar{y}_t \right)}_{v^j_{1,t}} + \underbrace{\beta v \left( b^j_{2,t+1}, q^j_{x,t+1}, q^j_{y,t+1}, E_{t+1}, \bar{y}_{t+1} \right)}_{v^j_{2,t+1}} \quad \text{(10)}
\]

where I introduce the notation \( v^j_{1,t} \) and \( v^j_{2,t+1} \) for a concise expression of conditional indirect utility function at different life stages.

In the second stage, hours of work \( l^j_{1,t} \) and savings \( s^j_{1,t} \) are chosen to maximize the conditional indirect utility function under the budget constraints. That is

\[
\begin{align*}
\text{Max}_{l^j_{1,t}, s^j_{1,t}} & \quad v^j_{1,t} + \beta v^j_{2,t+1} \\
\text{s.t.} & \quad b^j_{1,t} = w^j_{1,t} l^j_{1,t} - T_t \left( w^j_{1,t} l^j_{1,t} \right) - s^j_{1,t} \quad \text{(11)} \\
b^j_{2,t+1} &= (1 + r_{t+1}) s^j_{1,t} - \Phi_{t+1} \left( r_{t+1} s^j_{1,t} \right).
\end{align*}
\]

The first-order conditions with respect to labor supply and savings can be written as

\[
MRS^j_{z,b} = \frac{\partial v / \partial z^j_{1,t}}{\partial v / \partial b^j_{1,t}} = w^j_{t} \left[ 1 - T'_t \left( w^j_{1,t} \right) \right] \quad \text{(12)}
\]
\[ MRS_{j_1,j_2}^{i,t} = \frac{\partial v/\partial b_{1,t}^j}{\beta \partial v/\partial b_{2,t+1}^j} = 1 + r_{t+1} \left[ 1 - \Phi_{t+1}^j \left( r_{t+1} s_{1,t}^j \right) \right]. \quad (13) \]

Equation (12) means that the marginal rate of substitution between leisure \( z_{1,t}^j \) and disposable income \( b_{1,t}^j \) equals the marginal wage rate. Equation (13) states that the marginal rate of substitution between intertemporal disposable income \( b_{1,t}^j \) and \( b_{2,t+1}^j \) fits the marginal interest factor, i.e., one plus the interest rate net of marginal capital income taxation.

3.2 Social planner’s behavior

The social planner aims at maximizing a general social welfare function reflecting the preference over both old and young individuals with different abilities over periods

\[ W \left( \sum_j n_{i}^j V_{i,j}^j, \sum_j n_{i+1}^j V_{i+1,j}^j, \ldots \right), \quad (14) \]

which is in line with the literature on optimal taxation with redistribution purposes in an OLG model (Pirttilä and Tuomala, 2001; Aronsson and Johansson-Stenman, 2010; Aronsson and Sjögren, 2018). To this end, he needs to design a tax scheme that efficiently redistributes from the high-ability to the low-ability type and corrects the environmental and positional externalities. The availability of a mix of non-linear income taxes and linear commodity taxes is rooted in the information structure. Specifically, the government must base the taxation on observable earnings \( w_{i,t}^j l_{1,t}^j \) and capital income \( r_{t+1} s_{1,t}^j \) due to asymmetric information concerning consumers’ specific abilities. Therefore, a non-linear income tax on individual income is feasible while ability-type specific lump-sum taxes are not. In addition, aggregate consumption of goods can be observed instead of individual consumption. As a result, the government

---

8 Individuals of the same generation (who enter the economy at the same time) are given the same welfare weight. Note that this formulation is assumed to be additive for a given generation as in Pirttilä and Tuomala (2001) and Aronsson and Sjögren (2018). It can also be written as non-additive within one generation as in Aronsson and Johansson-Stenman (2010) with comparable results in tax implications, because the tax implications are independent of a social welfare function form. Meanwhile, the results remain the same if the social planner aims to maximize the utility of one specific type of consumers in one period when the minimum utility requirement for the rest of consumers holds. Note that the model in question can be solved without specifying the exact welfare function, making it applicable to various functional forms. Details are provided in the government’s first-order conditions (see Equation A.2). Here, the derivative of \( W \) regarding the indirect utility function for any individual can be replaced with other terms when determining shadow prices in A.1, and subsequently canceled out during the derivation of non-linear taxes in A.5.
can only levy linear commodity taxes but not non-linear ones. The planner needs to consider feasibility sets before setting a tax schedule.\(^9\)

### 3.2.1 The public budget in each period \( t \)

Given two generations of both low and high-type consumers active in each period, the public budget could accordingly be written as

\[
\sum_j \left[ n_j^T w_{1,t}^l + n_j^T \Phi_t \left( r_j s_{1,t}^l - 1 \right) + \tau_{x,t} \left( n_j^x x_{1,t}^l + n_j^{x-1} x_{2,t}^l \right) + \tau_{y,t} \left( n_j^y y_{1,t}^l + n_j^{y-1} y_{2,t}^l \right) \right] = 0. \tag{15}
\]

The first two terms represent the sum of payments of labor income and capital income taxes in period \( t \). The remaining two terms refer to commodity tax revenues. To connect private savings to the public budget constraints, I write down physical capital accumulation

\[
K_{t+1} = \sum_j n_j^l s_{1,t}^l. \tag{16}
\]

The capital stock at time \( t + 1 \) comes from the sum of previous savings at \( t \) made by the young generation. The initial capital stock is exogenous by assumption. The combination of private budget constraints (4), (5) and social capital accumulation (16) updates the public budget constraint as

\[
K_t \left( 1 + r_t \right) - K_{t+1} + \sum_j \left[ n_j^l w_{1,t}^l - \left( n_j^l b_{1,t}^l + n_j^{l-1} b_{2,t}^l \right) \right] + \sum_j \tau_{x,t} \left( n_j^x x_{1,t}^l + n_j^{x-1} x_{2,t}^l \right) + \sum_j \tau_{y,t} \left( n_j^y y_{1,t}^l + n_j^{y-1} y_{2,t}^l \right) = 0. \tag{17}
\]

The revenue from non-linear income taxes \( \sum_j \left[ n_j^T w_{1,t}^l + n_j^T \Phi_t \left( r_j s_{1,t}^l - 1 \right) \right] \) is now expressed by a net physical capital surplus, \( K_t \left( 1 + r_t \right) - K_{t+1} \), and total labor income reduced by disposable income, \( \sum_j \left[ n_j^l w_{1,t}^l - \left( n_j^l b_{1,t}^l + n_j^{l-1} b_{2,t}^l \right) \right] \).

### 3.2.2 The self-selection constraint

Given the government’s redistributive purpose and no access to private ability information, a high-ability individual could potentially mimic the low-ability type consumer in order to take advantage of the redistribution policy. In order to make the potential mimicker’s behavior unattractive for high-ability type

---

\(^9\)The social planner selects non-linear taxes as well as commodity taxes. However, opting for non-linear income taxes is equivalent to choosing the labor supply and disposable income for each type of agent, in accordance with the given constraints in equation (4) and (5).
individuals, the self-selection constraint reads as follows

$$V_t^H = v\left(b_{1,t}^H, q_{x,t}, q_{y,t}, z_{1,t}^H, E_t, \bar{y}_t\right) + \beta v\left(b_{2,t+1}^H, q_{x,t+1}, q_{y,t+1}, E_{t+1}, \bar{y}_{t+1}\right)$$

$$\geq \hat{V}_t^H = v\left(b_{1,t}^H, q_{x,t}, q_{y,t}, z_{1,t}^H, E_t, \bar{y}_t\right) + \beta v\left(b_{2,t+1}^H, q_{x,t+1}, q_{y,t+1}, E_{t+1}, \bar{y}_{t+1}\right),$$

(18)

where $z_{1,t}^H = 1 - \phi_t^1L_{1,t}$ refers to the potential mimicker’s leisure time and is higher than that of low-ability consumers. The mimicker’s efficiency is expressed by the wage ratio $\phi_t = w_t^f/w_t^H < 1$. The conditional indirect utility function for a retired low-ability individual and a retired mimicker will converge as $\hat{V}_{2,t+1}^H = \sqrt{V}_{2,t+1}$. Because the budget constraint for a potential mimicker will be exactly the same as for a low-ability type in the second period when old individuals do not work.

3.2.3 Externality-generating mechanisms

Stock-externalities mean that if an externality does not entirely dissolve in the next period but accumulates over time, the consumers will suffer from its aggregation. $E_t$ the stock of pollution in period $t$ comes from the previous consumption of a dirty good i.e., the pollution stock level $E_{t-1}$ and current consumption $x_t$:

$$E_t = (1 - \rho)E_{t-1} + \sum_j \left(n^j_1x^j_{1,t} + n^j_{t-1}x^j_{2,t}\right).$$

(19)

Despite the abilities of self-restoration by the environment itself, where $\rho \in (0, 1)$ is the pollutant decay rate, the continuous rise of pollutant contents is seen as emissions into the environment currently lay above this self-purification threshold.. The dirty commodities with a tight connection to environmental pollution as a stock externality are ubiquitous. For example, the effects by consuming gasoline, plastic-made products, over-decorated garments, papers, medicine, cleansers, etc., remain a while. The second term in (19) shows that the mechanism of pollution flow is atmospheric, meaning that the marginal contribution is the same across individuals.

In line with Ljungqvist and Uhlig (2000), where people’s utility depends on the average consumption of both the present and the past, I develop the stock of the positional externality at time $t$ measuring the reference level by which people compare their own consumption in period $t$ as

$$\bar{y}_t = (1 - \eta)\bar{y}_{t-1} + \eta \left(\sum_j n^j_1\bar{y}^j_{1,t} + \sum_j n^j_{2,t}\bar{y}^j_{2,t}\right).$$

(20)

Both the reference level in the last period $\bar{y}_{t-1}$ and the weighted average of people’s current consumption of the positional good contribute to the reference measure at time $t$. The dynamics are featured by the first term measuring
the contribution by the previous period’s positional externality level. It reflects a *Catching-up-with-the-Joneses* mechanism where individuals compare their own current consumption with other people’s previous consumption as in Aronsson and Johansson-Stenman (2014).\(^\text{10}\)

The second term implies that the two ability types differ in their marginal harm by the positional consumption, i.e. \(\alpha^L \neq \alpha^H\), which means that the positional externality is non-atmospheric. This nature captures the generalization that some individuals may be considered more influential in forming reference levels than others. This could be an upward comparison, where only high-ability individuals contribute to the reference measure, a downward comparison, or situations between the two extremes.\(^\text{11}\) The special case with an atmospheric positional externality arises when \(\alpha^L = \alpha^H\). The second term also embeds a *Keeping-up-with-the-Joneses* effect, as the reference level \(\bar{y}_t\) depends on the current consumption of a positional good. The parameter \(\eta \in (0, 1)\) partially translates people’s current positional consumption into the reference measure in the same period.

### 3.2.4 Lagrange Function

Given the social objective and the constraints facing the social planner, the optimization thereafter translates into selecting \(l_{1,t}^j, b_{1,t}^j, b_{2,t}^j, \tau_{x,t}, \tau_{y,t}, K_t, E_t, \bar{y}_t\) for \(j = 1, 2\), and all \(t\) to maximize the following Lagrangean

\[
L_0 = W \left( \sum_j n_0^j V_0^j + \sum_j n_1^j V_1^j + \ldots + \sum_j n_t^j V_{t+1}^j + \ldots \right) + \sum_t \lambda_t \left( V_t^H - \tilde{V}_t^H \right)
+ \sum_t \mu_t \left[ E_t - (1 - \rho) E_{t-1} - \sum_j \left( n_j^1 x_{1,t}^j + n_j^1 x_{2,t}^j \right) \right]

+ \sum_t \delta_t \left[ \bar{y}_t - (1 - \eta) \bar{y}_{t-1} - \eta \sum_j \alpha_j \left( n_j^1 y_{1,t}^j + n_j^1 y_{2,t}^j \right) \right]

+ \sum_t \gamma_t \left[ K_t (1 + r_t) - K_{t+1} + \sum_j \left( n_t^j \left( x_{1,t}^j \tau_{x,t}^j + y_{1,t}^j \tau_{y,t}^j + w_{1,t}^j l_{1,t}^j - b_{1,t}^j \right) \right.ight.

\left. \left. + n_{t-1}^j \left( x_{2,t}^j \tau_{x,t}^j + y_{2,t}^j \tau_{y,t}^j - b_{2,t}^j \right) \right) \right].
\]

The Lagrange multipliers \(\mu_t\) and \(\delta_t\) are associated with the environmental and positional externality, respectively. \(\lambda_t\) is the multiplier for self-selection constraint and \(\gamma_t\) denotes the shadow price of the resource constraint. The quotients \(\mu_t / \gamma_t\) and \(\delta_t / \gamma_t\) can be interpreted as the social marginal value of avoiding the externalities, measured in terms of tax revenue. I next turn to the first

---

\(^{10}\)This mechanism was initially introduced to explain why the average value of the equity premium is high in Abel (1990) and Campbell and Cochrane (1999), where the lagged level of aggregate consumption per capita matters in consumers’ utility functions.

\(^{11}\)It would be possible to relax the assumption that the environmental externality is atmospheric, but one non-atmospheric externality will be adequate to present the intuition, though.
proposition for the formulas of shadow prices of externalities, because they are of high relevance for tax policy expressions.

4 The shadow prices of the externalities

Before formulating the expressions of the shadow prices, it will be helpful to put forward related definitions at time \( t \) across ages \( k \) (for young \( k = 1 \) and old \( k = 2 \) ) and abilities \( j \) (for low \( j = L \) and high abilities \( j = H \)). Let

\[
MW P^j_{Mb,k,t} = -\frac{\partial v^j_{k,t}}{\partial v^j_{k,t}} / M_t
\]

(22)

denote the marginal willingness to pay to avoid the environmental externality \((M = E)\) or the positional externality \((M = \bar{y})\) for type \( j \), and let

\[
\overline{MW P}^H_{Mb,k,t} = -\frac{\partial \hat{v}^H_{k,t}}{\partial \hat{v}^H_{k,t}} / E_t
\]

(23)

be the analogous marginal willingness to pay for a potential mimicker. Considering that the environmental (positional) externality is treated as pollution (psychological suffering) and therefore impairs consumers’ utility, the marginal willingness to pay for externality correction is positive.

By using \( \tilde{x}^j_t \left( \tilde{y}^j_t \right) \) to denote the compensated conditional demand for the dirty (positional) good, I can define the following compensated derivatives with the Slutsky equation. The externality effect is

\[
\frac{\partial \tilde{x}^j_{k,t}}{\partial M_t} = \frac{\partial \tilde{x}^j_{k,t}}{\partial M_t} MW P^j_{Mb,k,t} \quad \text{and} \quad \frac{\partial \tilde{y}^j_{k,t}}{\partial M_t} = \frac{\partial \tilde{y}^j_{k,t}}{\partial M_t} MW P^j_{Mb,k,t}.
\]

(24)

Relevantly, there are two externality feedback factors related to compensated conditional demand. \( \frac{1}{\sigma^E_t} = 1 - \sum_j \left( n^j_t \frac{\partial \tilde{x}^j_{k,t}}{\partial E_t} + n^j_{t-1} \frac{\partial \tilde{x}^j_{k,t}}{\partial E_t} \right) \) is the environment feedback to compensated conditional demand for the dirty good. In a different context, Sandmo (1980) shows that this type of feedback effect is positive because \( |\partial \tilde{x}^j_{k,t}/\partial E_t| \) is assumed to be moderate to ensure that the aggregate demand function is stable. This is specified by \( \left| \sum_j \left( n^j_t \frac{\partial \tilde{x}^j_{k,t}}{\partial E_t} + n^j_{t-1} \frac{\partial \tilde{x}^j_{k,t}}{\partial E_t} \right) \right| < 1 \).

The other is the positional externality feedback factor

\[
\frac{1}{\sigma^F_t} = 1 - \eta \sum_j \alpha^j \left( n^j_t \frac{\partial \tilde{y}^j_{k,t}}{\partial \bar{y}_t} + n^j_{t-1} \frac{\partial \tilde{y}^j_{k,t}}{\partial \bar{y}_t} \right). \]

It is also positive based on a similar argument as the environment feedback factor.
Proposition 1 The shadow prices of an environmental and a positional externality measured in monetary units at $t$ are given by

$$\mu_t = \sum_{m=0}^{\infty} \left(1 - \rho\right)^{m} \Pi_{t=0}^{m} \sigma_{i+m}^{E} \sum_{j} \left(n_{i+m}^{j} MW P_{E_b,1,t+m}^{j} + n_{i+m-1}^{j} MW P_{E_b,2,t+m}^{j}\right)$$

$$+ \sum_{m=0}^{\infty} \left(1 - \rho\right)^{m} \Pi_{t=0}^{m} \sigma_{i+m}^{E} \left[\lambda_{i+m,1}^{1} \left(MWP_{E_b,1,t+m}^{L} - \overline{MWP_{E_b,1,t+m}^{H}}\right)\right]$$

$$- \sum_{m=0}^{\infty} \left(1 - \rho\right)^{m} \Pi_{t=0}^{m} \sigma_{i+m}^{E} \left[\gamma_{i+m,1}^{1} \sum_{j} \left(n_{i+m}^{j} \frac{\partial \hat{x}_{1,t+m}^{j}}{\partial \bar{E}_{t+m}} + n_{i+m-1}^{j} \frac{\partial \hat{x}_{2,t+m}^{j}}{\partial \bar{E}_{t+m}}\right)\right]$$

$$- \sum_{m=0}^{\infty} \left(1 - \rho\right)^{m} \Pi_{t=0}^{m} \sigma_{i+m}^{E} \left[\gamma_{i+m,1}^{1} \sum_{j} \left(n_{i+m}^{j} \frac{\partial \hat{y}_{1,t+m}^{j}}{\partial \bar{E}_{t+m}} + n_{i+m-1}^{j} \frac{\partial \hat{y}_{2,t+m}^{j}}{\partial \bar{E}_{t+m}}\right)\right]$$

(25)

and

$$\delta_{t} = \sum_{m=0}^{\infty} \left(1 - \eta\right)^{m} \Pi_{t=0}^{m} \sigma_{i+m}^{F} \sum_{j} \left(n_{i+m}^{j} MW P_{y_b,1,t+m}^{j} + n_{i+m-1}^{j} MW P_{y_b,2,t+m}^{j}\right)$$

$$+ \sum_{m=0}^{\infty} \left(1 - \eta\right)^{m} \Pi_{t=0}^{m} \sigma_{i+m}^{F} \left[\lambda_{i+m,1}^{1} \left(MWP_{y_b,1,t+m}^{L} - \overline{MWP_{y_b,1,t+m}^{H}}\right)\right]$$

$$- \sum_{m=0}^{\infty} \left(1 - \eta\right)^{m} \Pi_{t=0}^{m} \sigma_{i+m}^{F} \left[\tau_{i+m,1}^{1} \sum_{j} \left(n_{i+m}^{j} \frac{\partial \bar{y}_{1,t+m}^{j}}{\partial \bar{E}_{t+m}} + n_{i+m-1}^{j} \frac{\partial \bar{y}_{2,t+m}^{j}}{\partial \bar{E}_{t+m}}\right)\right]$$

$$- \sum_{m=0}^{\infty} \left(1 - \eta\right)^{m} \Pi_{t=0}^{m} \sigma_{i+m}^{F} \left[\tau_{i+m,1}^{1} \sum_{j} \left(n_{i+m}^{j} \frac{\partial \bar{y}_{1,t+m}^{j}}{\partial \bar{E}_{t+m}} + n_{i+m-1}^{j} \frac{\partial \bar{y}_{2,t+m}^{j}}{\partial \bar{E}_{t+m}}\right)\right]$$

(26)

where $\lambda_{i+m,1}^{1} = \frac{\lambda_{i+m}^{1} \frac{\partial \hat{y}_{1,t+m}^{H}}{\partial \bar{E}_{t+m}}}{\bar{E}_{t+m} \frac{\partial \hat{y}_{1,t+m}^{H}}{\partial \bar{E}_{t+m}}}$. Proof: see Appendix A.2.\(^{12}\)

I will explain the meanings of the formulas in parallel since expressions (25) and (26) are symmetric in several aspects. One of the standard features of the shadow prices is the forward-looking discounting term, i.e., $\sum_{m=0}^{\infty} \left(1 - \rho\right)^{m} \Pi_{t=0}^{m}$ in (25) and $\sum_{m=0}^{\infty} \left(1 - \eta\right)^{m} \Pi_{t=0}^{m}$ in (26). The discounting term consists of the interest rate and the decay rate of pollution (social comparisons).

The first line represents the present value of the consumers’ marginal willingness to avoid pollution in (25) (a positional externality in (26)) across generations and earning abilities. The second line illustrates the sum of discounted disparity in the marginal willingness to pay to avoid the environmental externality (the positional externality) between a low-ability type and a potential mimicker.\(^{13}\) The third line is an own tax-base effect. It arises as long as

---

\(^{12}\)The structure of the results is similar to previous literature such as Aronsson and Sjögren (2018). However, my results characterize the possibility of direct interaction between the externalities over time.

\(^{13}\)The sign of this component depends on whether the marginal willingness to pay to avoid the environmental externality increases or decreases in leisure time. The sign will be positive if this
the compensated conditional demand for a dirty (positional) good reacts to a change in the environmental quality (consumption references).

The last line is a cross-tax-base effect, which is relevant to the corrective purposes. In (25), the cross-tax-base effect depends on \( \tau_{y,t+m} - \eta \alpha^i \delta_{t+m}/\gamma_{t+m} \) the distortionary part of the tax on the positional good \( y \), which is difference between the positional commodity tax rate and the marginal damage in terms of monetary units generated by individuals’ positional consumption, and \( \partial \tilde{y}_{t+m}/\partial E_{t+m} \) how the compensated conditional demand for \( y \) reacts to the environmental quality. If \( \partial \tilde{y}_{t+m}/\partial E_{t+m} \neq 0 \), meaning that a marginal change in the environmental pollution affects the compensated conditional demand for the positional good, there will be a direct dynamic interaction between environmental and positional externalities.

Specifically, \( \mu_t/\gamma \) directly interacts with \( \delta_{t+m}/\gamma_{t+m} \) for \( m = 0,1,\ldots,\infty \) since \( \tau_{y,t+m} - \eta \alpha^i \delta_{t+m}/\gamma_{t+m} \) will contain \( \delta_{t+m}/\gamma_{t+m} \) by the expression of commodity taxation (30) in the coming section. The intuition is that the linear commodity tax on the positional good is not sufficiently flexible for internalizing the positional externality, where the marginal contribution differs across individuals. Therefore, the uncorrected positional externalities in both present and future periods become part of the current shadow price of the environmental externality. The direct interaction between \( \mu_t/\gamma \) and \( \delta_{t+m}/\gamma_{t+m} \) has important implications for tax policy. The social value of decreasing an environmental externality at time \( t \) embeds that of correcting for the positional externality from \( t \) onwards. This interaction vanishes in the special case where the positional externality is atmospheric, because the linear tax on the positional good can perfectly target uniform marginal social harm of positional consumption, yielding that \( \tau_{y,t+m} - \eta \alpha^i \delta_{t+m}/\gamma_{t+m} \) does not depend on \( \delta_{t+m}/\gamma_{t+m} \). Without this special situation, the other tax instrument will supplement the tax on the positional good to rectify the positional externality. To see this more clearly, I proceed to the commodity taxes.

5 Optimal commodity taxation

For simplicity, suppose that the government commits to the tax policy decided on in period 0.\(^{14}\) To understand the potential time inconsistency in this case, note that when each individual has revealed his/her type at the end of the first

---

\(^{14}\)This is the conventional assumption in earlier studies on optimal redistributive taxation in dynamic economies. Earlier references can be seen in Pirttilä and Tuomala (1997), Aronsson and
period in life, the government may want to use this new information for revising the structure of income and commodity taxation. I assume away this option because, by assumption, the government commits to the second-best policy.

To explain the commodity tax expressions, I follow the concept in equation (24) to write down the own-price effects for the compensated conditional demands

\[
\frac{\partial \tilde{x}_{i,t}}{\partial q_{x,t}} = \frac{\partial x_{i,t}^j}{\partial q_{y,t}} + x_{i,t}^j \frac{\partial x_{i,t}^j}{\partial b_{k,t}^j} \quad \text{and} \quad \frac{\partial \tilde{y}_{k,t}}{\partial q_{y,t}} = \frac{\partial y_{k,t}^j}{\partial q_{x,t}} + y_{k,t}^j \frac{\partial y_{k,t}^j}{\partial b_{k,t}^j} \tag{27}
\]

and the cross-price effects

\[
\frac{\partial \tilde{x}_{i,t}}{\partial q_{y,t}} = \frac{\partial x_{i,t}^j}{\partial q_{y,t}} + y_{k,t}^j \frac{\partial x_{i,t}^j}{\partial b_{k,t}^j} \quad \text{and} \quad \frac{\partial \tilde{y}_{k,t}}{\partial q_{x,t}} = \frac{\partial y_{k,t}^j}{\partial q_{x,t}} + x_{i,t}^j \frac{\partial y_{k,t}^j}{\partial b_{k,t}^j}. \tag{28}
\]

**Proposition 2** The second-best optimal commodity taxation system follows

\[
\tau_{x,t} = \frac{\lambda_{1,t}^* (x_{1,t}^L - x_{1,t}^H)}{\Omega_t} \sum_j \left( n_{t,j}^i \frac{\partial \tilde{x}_{1,t}^j}{\partial q_{y,t}} + n_{t-1,j}^i \frac{\partial \tilde{x}_{2,t}^j}{\partial q_{y,t}} \right)
- \frac{\lambda_{1,t}^* (y_{1,t}^L - y_{1,t}^H)}{\Omega_t} \sum_j \left( n_{t,j}^i \frac{\partial \tilde{x}_{1,t}^j}{\partial q_{x,t}} + n_{t-1,j}^i \frac{\partial \tilde{x}_{2,t}^j}{\partial q_{x,t}} \right)
+ \frac{\mu_0}{\gamma_t} + \frac{\delta_t}{\gamma_t} \left( \alpha^L - \alpha^H \right) \left( n_{t,j}^L \frac{\partial \tilde{x}_{1,t}^j}{\partial q_{x,t}} + n_{t-1,j}^L \frac{\partial \tilde{x}_{2,t}^j}{\partial q_{x,t}} \right) \left( n_{t,j}^H \frac{\partial \tilde{x}_{1,t}^H}{\partial q_{y,t}} + n_{t-1,j}^H \frac{\partial \tilde{x}_{2,t}^H}{\partial q_{y,t}} \right)
+ \frac{\delta_t}{\gamma_t} \left( \alpha^H - \alpha^L \right) \left( n_{t,j}^H \frac{\partial \tilde{x}_{1,t}^H}{\partial q_{x,t}} + n_{t-1,j}^H \frac{\partial \tilde{x}_{2,t}^H}{\partial q_{x,t}} \right) \left( n_{t,j}^L \frac{\partial \tilde{x}_{1,t}^L}{\partial q_{y,t}} + n_{t-1,j}^L \frac{\partial \tilde{x}_{2,t}^L}{\partial q_{y,t}} \right) \tag{29}
\]

\[
\tau_{y,t} = \frac{\lambda_{1,t}^* (y_{1,t}^L - y_{1,t}^H)}{\Omega_t} \sum_j \left( n_{t,j}^i \frac{\partial \tilde{x}_{1,t}^j}{\partial q_{x,t}} + n_{t-1,j}^i \frac{\partial \tilde{x}_{2,t}^j}{\partial q_{x,t}} \right)
- \frac{\lambda_{1,t}^* (y_{1,t}^L - y_{1,t}^H)}{\Omega_t} \sum_j \left( n_{t,j}^i \frac{\partial \tilde{x}_{1,t}^j}{\partial q_{x,t}} + n_{t-1,j}^i \frac{\partial \tilde{x}_{2,t}^j}{\partial q_{x,t}} \right)
+ \frac{\delta_t}{\gamma_t} \left( \alpha^L \theta^L + \alpha^H \theta^H \right). \tag{30}
\]

Johansson-­Stenman (2010), and Aronsson and Sjögren (2018). There is also a strand of research on time-consistent optimal taxation without commitment. Aronsson and Sjögren (2016) and Brett and Weymark (2019) analyze time-consistent optimal taxation in different contexts under asymmetric information.
where $\theta^H + \theta^L = 1$ and $\Omega_t > 0$. Proof: see Appendix A.3.

The optimal commodity taxation system reflects two main objectives: relaxation of the self-selection constraint and externality correction. The first and second terms are due to the self-selection constraint. These components would also be present in the absence of externalities, since commodity tax instruments are used for redistributive purposes under asymmetric information. Their signs depend on how the consumption pattern varies with leisure time, ceteris paribus. Under the same disposable income for a low-ability type and a potential mimicker, the only difference is that the mimicker has more leisure time. It drives a possible difference in consumption behavior.\footnote{If leisure is weakly separable from the other goods in the utility function, then $y^L_1, t = \tilde{y}^L_1, t$ and $x^H_1, t = x^H_1, t$, thus implying that the first and second terms on the right-hand side of equation (29) and (30), respectively, vanish. The outcome is consonant with what is derived by Atkinson and Stiglitz (1976).}

This difference in consumption behavior only refers to the first period of life, since the utility function is time-separable.

The third term in (30) reflects the purpose of correcting for the positional externality. The corrective term $\frac{\delta}{\delta H} \eta \left( \alpha^L \theta^L + \alpha^H \theta^H \right)$ is a weighted average of the social value of the positional externality generated by each ability type, since the two types of consumers contribute differently at the margin. Linear taxation of the positional good will, therefore, not be efficient in internalizing the social cost of positional externalities. By the implication in (30), the difference between the tax on the positional good and the type-specific marginal contribution to the positional externality, $\tau_{y,t} + m - \eta \alpha^L \theta^L / \eta^L_{y+m},$ is directly dependent on $\delta_{t+1} / \gamma_{t+1}$ from time $t$ and onwards. The incapability of $\tau_{y,t}$ to fully correct the positional externality in the same period implies that the other tax instruments will be modified in order to reflect the shadow price of the positional externality.

The straightforward effect on the dirty good tax is the appearance of the fourth and fifth terms in (29), indicating that $\tau_{x,t}$ directly corrects the positional externality, through the possible substitutability/supplementary relationship between the dirty good and the positional good $\partial x^t / \partial q_{x,t} > ( < ) 0, \partial y^t / \partial q_{x,t} > ( < ) 0$. The sign of the sum of the fourth and fifth terms determines how the dirty good tax partially corrects the positional externality, $(\alpha^L - \alpha^H)$ and

$$
\left( n^L H \frac{\partial^2 y^L}{\partial q_{x,t}^2} + n^L_{t-1} \frac{\partial x^L}{\partial q_{x,t}} \right) \left( n^L H \frac{\partial^2 x^L}{\partial q_{x,t}^2} + n^L_{t-1} \frac{\partial x^L}{\partial q_{x,t}} \right) - \left( n^H L \frac{\partial^2 y^H}{\partial q_{x,t}^2} + n^L_{t-1} \frac{\partial x^H}{\partial q_{x,t}} \right) \left( n^L H \frac{\partial^2 x^H}{\partial q_{x,t}^2} + n^L_{t-1} \frac{\partial x^H}{\partial q_{x,t}} \right)
$$

jointly affect the sign.\footnote{One similar result can be seen in Aronsson and Sjögren (2018). The commodity tax on the clean good whose consumption does not generate any environmental externality contains the
sitional and environmental externalities in Proposition 1. Sandmo’s additivity is violated as a result. Micheletto (2008) points out an exception where a non-atmospheric externality does not invalidate the additivity property in the commodity tax system. Recovering the property requires the cross-substitution effects to be zero, i.e., $\frac{\partial \tilde{x}_t}{\partial q_{x,t}} = \frac{\partial \tilde{y}_t}{\partial q_{x,t}} = 0$. The fourth and fifth terms in (29) will, therefore, disappear. However, it cannot recover the additivity property when the direct interaction between the positional and environmental externalities still holds, where $\mu_t/\gamma_t$ in (29) also reflects the value of correcting the positional externality. In Micheletto’s study, there is only one externality-generating commodity which is non-atmospheric, without any possible direct interaction between externalities.

In the special case where the positional externality is atmospheric such that $\alpha^H = \alpha^L$, the fourth and fifth terms on the right-hand side of equation (29), as well as the direct interaction between the two externalities, would vanish. Therefore, in this particular case, the commodity tax system would satisfy the additivity property.

The coming proposition specifies how the marginal income taxes are used as supplemental instruments for correcting the positional externality. The appearance of corrective purposes in these non-linear taxes implies that the principle of targeting does not apply.

6 Optimal marginal labor income and capital income taxation

There are three purposes for the marginal labor and capital income taxes to serve: (i) relaxing the self-selection constraint, (ii) compensating the consumers for distortions created by the commodity taxes due to that the commodity taxes are usually distortionary at the second-best optimum, where the role of marginal labor and capital income taxes can also be found in earlier studies (Aronsson and Mannberg, 2015; Aronsson and Sjögren, 2018).\(^{17}\) and (iii) correcting for the positional externality.

---

\(^{17}\) Aronsson and Mannberg (2015) look into how marginal labor and capital income taxation jointly work as supplementary instruments to correct for the positional externality raised by housing consumption when the tax rate on housing wealth is lower than optimal to fully correct the positional externality. However, they do not consider the case if the positional externality by consuming housing is non-atmospheric, nor the coexistence of other externalities. In addition, there is no potential mimicker issue inside their model, which frees the purpose of relaxing the self-selection constraint.
Proposition 3 The policy rules for marginal labor income taxation can be written

\[ T'_t (w_l^t H^t, l_{l,t}) = \frac{1}{w_l^t} \frac{\lambda^*_t}{n_l^t} \left( MRSL_t^L - \frac{\phi_l}{\phi_l} MRSL_t^H \right) + \frac{1}{w_l^t} \left( \tau_{x,t} - \frac{\mu_t}{\gamma_t} \frac{\partial \tilde{x}_{1,t}^L}{\partial z_{1,t}^L} + \left( \tau_{y,t} - \frac{\delta_t}{\gamma_t} \eta^L \right) \frac{\partial \tilde{y}_{1,t}^L}{\partial z_{1,t}^L} \right) \]

(31)

while the policy rules for marginal capital income taxation is

\[ \Phi'_{t+1} (r_t, v_t) = \frac{1 + r_{t+1}}{r_{t+1}} \left[ \left( \tau_{x,t} - \frac{\mu_t}{\gamma_t} \frac{\partial \tilde{x}_{1,t}^H}{\partial b_{1,t}^H} \right) \frac{\partial \tilde{y}_{1,t}^H}{\partial b_{1,t}^H} - \left( \tau_{y,t} - \frac{\delta_t}{\gamma_t} \eta^H \right) \frac{\partial \tilde{y}_{1,t}^H}{\partial b_{1,t}^H} \right] \]

(32)

Proof: See Appendix A.4, A.5.\(^\dagger\)

(i) The self-selection constraint relaxation. The first term in (31) works to increase the marginal labor income tax implemented for the low-ability type. This makes mimicking less attractive and opens up for more redistribution, because the low-ability type has a higher marginal utility of leisure than a potential mimicker (Stiglitz, 1982). The first term in (33) also serves to relax the self-selection constraint by exploiting differences in the intertemporal consumption tradeoff between the potential mimicker and the low-ability type. The sign of this component depends on how the marginal rate of substitution between present and future consumption varies with the labor supply, ceteris paribus.\(^\ddagger\)

\(^\dagger\)The marginal labor income tax implication follows the results in Cremer et al. (1998) analogously if time dimension is relaxed. The marginal capital income taxation can recover the results from Aronsson and Sjögren (2018) if the positional good and corresponding linear commodity taxation are dropped.

\(^\ddagger\)A saving tax(subsidy) on the low-ability type can relax the constraint by making potential mimickers suffer if the marginal rate of substitution between the present and future disposable income decreases(increases) with the leisure time.
One relevant concept to understand the role of the optimal marginal labor income taxation in purposes (ii) and (iii) is the effects of leisure time on the compensated conditional demands, where the definition follows equation (24),

$$\frac{\partial x^I_{1,t}}{\partial z^I_{1,t}} = \frac{\partial x^I_{1,t}}{\partial z^I_{1,t}} - \frac{\partial x^I_{1,t}}{\partial b^I_{1,t}} MRS^I_{z,b_1} \text{ and } \frac{\partial y^I_{1,t}}{\partial z^I_{1,t}} = \frac{\partial y^I_{1,t}}{\partial b^I_{1,t}} MRS^I_{z,b_1}.$$

(ii) Compensating the consumers for distortions created by the commodity taxes. Although the corrective part of the commodity tax ($\mu_t/\gamma_t$ and $\eta J_t/\gamma_t$) is not distortionary, the non-corrective part ($\tau_t - \mu_t/\gamma_t$ and $\tau_t - \eta J_t/\gamma_t$) distorts the resource allocation. The social planner relies on the non-linear taxation which is more flexible than the uniform commodity tax rate to offset the distortion. The compensation by labor income taxation is realized by affecting consumers’ decisions on labor supply, verified by the second and third terms in (31) together with (32). The second and third lines in (33) and (34) indicate that marginal capital income taxes partially compensate consumers for distortions caused by commodity taxation via affecting intertemporal disposable income.

(iii) Supplementing the commodity taxes as an instrument through which the government corrects for the positional externality. The distortionary form of the tax formula of the positional good $\tau_t - \mu_t/\gamma_t$ directly, meaning that the positional externality directly affects the policy rules for the non-linear tax policies. The corrective part in the marginal labor income taxation system is

$$(\alpha^H - \alpha^L) \theta^H \delta_{t} \frac{\partial y^H_{1,t}}{\partial z^H_{1,t}} \text{ in (31) and } (\alpha^L - \alpha^H) \theta^L \delta_{t} \frac{\partial y^L_{1,t}}{\partial z^L_{1,t}} \text{ in (32), respectively.}$$

The marginal labor income taxation $T_t(\cdot)$ corrects the positional externality by adjusting the compensated conditional demand for the positional good $\tilde{y}_{1,t}$ via leisure, if leisure varies with the compensated conditional demand for the positional good such that $\partial \tilde{y}_{1,t}/\partial z_{1,t} \neq 0$.

---

20To get the corrective parts in (31) and (32), I need the expressions $\tau^H_t - \alpha^L \eta \delta_{t}/\gamma_t$ and $\tau^H_t - \alpha^L \eta \delta_{t}/\gamma_t = (\alpha^H - \alpha^L) \theta^H \delta_{t} \frac{\partial y^H_{1,t}}{\partial z^H_{1,t}} \text{ in (31) and } (\alpha^L - \alpha^H) \theta^L \delta_{t} \frac{\partial y^L_{1,t}}{\partial z^L_{1,t}} \text{ in (32), respectively.}$

With an analogous argument, I difference $\alpha^H \eta \delta_{t}/\gamma_t$ on the both sides of (30). Therefore, the term that contains $\delta_{t}/\gamma_t$ in $(\tau^H_t - \alpha^L \eta \delta_{t}/\gamma_t) \frac{\partial y^H_{1,t}}{\partial z^H_{1,t}} (\alpha^H - \alpha^L) \theta^H \delta_{t} \frac{\partial y^H_{1,t}}{\partial z^H_{1,t}}$.

The corrective term in (32) is thus given by $(\alpha^L - \alpha^H) \theta^L \delta_{t} \frac{\partial y^L_{1,t}}{\partial z^L_{1,t}}$.
The corrective elements in the bottom line of the marginal capital income taxation formulas are \((\alpha^H - \alpha^L) \theta^L \frac{\delta_{+e}}{\gamma^e} \eta \frac{\partial y^L_{t+e,t+e}}{\partial b^L_{t+e,t+e}}\) in (33) and 
\((\alpha^L - \alpha^H) \theta^H \frac{\delta_{+e}}{\gamma^e} \eta \frac{\partial y^H_{t+e,t+e}}{\partial b^H_{t+e,t+e}}\) in (34) for \(e = 0, 1\).\(^\text{21}\) Compared to marginal labor income taxation, the corrective purpose in marginal capital income taxation that adjusts intertemporal disposable income is realized by the direct income effect on consumer demand for the positional good \(\partial y / \partial b\). The relative sizes of the positional externality at time \(t\) and that at time \(t+1\) determines how \(\Phi'_{t+1}\) corrects for the intertemporal positional externalities. The social marginal value of the positional externality at time \(t\) and \(t+1\) affects the marginal capital income tax in the opposite direction, which is analogous to Aronsson and Johansson-Stenman (2010).\(^\text{22}\)

The corrective aspects of marginal income taxation would vanish in the special case where the positional externality is atmospheric. This means that the additivity property would apply: a specific externality leads to a corrective term in the commodity tax for the externality-generating good, while it has no direct influence on the policy rules for the other tax instruments.

7 Results and Discussion

In an intertemporal model with multiple stock-externalities where the positional externality is non-atmospheric, Sandmo’s additivity property usually does not apply.\(^\text{23}\) First, the policy rule for the tax on the dirty good is modified to reflect the positional externality. This is interpretable in terms of a dynamic interaction between environmental and positional externalities and a direct inclusion of the elements that correct for the positional externality in the policy rule for the tax on the dirty good. Second, the policy rules for marginal labor and capital income taxation directly depend on the shadow price of the positional externality, meaning that income taxation also supplements the tax on the positional good for purposes of externality correction.

In a special case where the positional externalities are atmospheric \(\alpha^L = \alpha^H\), the tax system satisfies the additivity property. The commodity instruments now perfectly target the externalities. The corrective components in the commodity tax formulas do not interact directly either. Furthermore, the

\(^{21}\) I substitute \(\tau_{y,t} - \alpha^L \eta \delta_t / \gamma_t\) and \(\tau_{y,t} - \alpha^H \eta \delta_t / \gamma_t\), which are derived in the same way as footnote 22, back into (33) and (34) to get the terms that contain \(\delta_t / \gamma_t\) in the expressions of marginal capital income taxation.

\(^{22}\) An illustrative example for corrective purposes can be seen in Appendix A.6. The example is helpful in understanding how these tax instruments interact directly in corrective purposes.

\(^{23}\) Diamond (1973), Micheletto (2008) and Eckerstorfer and Wendner (2013) verify similar results but in a different static framework, where none of these studies deals with stock externalities.
policy rules for marginal labor income taxation and marginal capital income taxation do not directly depend on the shadow prices of the externalities.

The optimal taxation in this paper takes place in an autarky economy. A promising avenue for future research would be to extend this analysis to an open economy. The international economy makes environmental issues and social comparisons transboundary, leading to a direct interaction between externalities across national borders over time. A hypothetical global social planner becomes essential to coordinate taxes among different countries as in Aronsson and Johansson-Stenman (2015). This echoes the importance of international cooperation to confront current global warming and widespread anxiety and stress due to positional effects in the long run.

Meanwhile, capital mobility is relatively less affected by the problematic situation. The open economy may also affect the interest rate through capital mobility in this research framework. Consumers’ attitudes towards the future can be altered, affecting how marginal capital income taxation is designed when weighing the value of correcting the positional externality at different periods. The current research lays a good foundation for understanding the extension to an open economy.

To be specific in designing capital income taxation to correct the positional externality, I need to compare the socially marginal value of correcting the positional externality in different periods, i.e. the sign for the bracket in the third line of (33). In addition to the scale of the positional externality at different periods, \( \frac{\text{MRS}^{L}_{b_1, b_2}}{1 + r_{t+1}} \) determines how the social value of correcting the positional externality at \( t + 1 \) is discounted at \( t \).
References

Abel, A. B. (1990). *Asset prices under habit formation and catching up with the Joneses*.


A Appendix

To derive optimal conditions for policy design, we first need to get the first-order conditions for the social planner's Lagrangean

\[ L_0 = W \left( \sum_{j} n_0^j V_0^j, \sum_{j} n_1^j V_1^j \cdots \sum_{j} n_t^j V_t^j, \sum_{j} n_{t+1}^j V_{t+1}^j \cdots \right) + \sum_{t} \lambda_t (V_t^H - \hat{V}_t^H) \]

\[ + \sum_{t} \mu_t \left[ E_t - (1 - \rho)E_{t-1} - \sum_{j} (n_t^j x_{1,t} + n_{t-1}^j x_{2,t}^j) \right] \]

\[ + \sum_{t} \delta_t \left[ \tilde{y}_t - (1 - \eta) \tilde{y}_{t-1} - \eta \sum_{j} \alpha^j \left( n_t^j y_{1,t}^j + n_{t-1}^j y_{2,t}^j \right) \right] \]

\[ + \sum_{t} \gamma_t \left[ K_t (1 + r_t) - K_{t+1} + \sum_{j} \left( n_t^j (x_{1,t}^j \tau_{x,t} + y_{1,t}^j \tau_{y,t} + w_{1,t}^j l_{1,t}^j - b_{1,t}^j) + n_{t-1}^j (x_{2,t}^j \tau_{x,t} + y_{2,t}^j \tau_{y,t} - b_{2,t}^j) \right) \right]. \]

A.1 First-order conditions of the government’s maximization problem at an arbitrary time \( t \)

The first-order conditions of \( L_0 \) with respect to the optimal bundles \( (b_{1,t}^j, b_{2,t+1}^j, l_{1,t}^j) \) \( j = L, H \) read

\[ \frac{\partial W}{\partial V_t^L} \frac{\partial \nu_{1,t}^L}{\partial b_{1,t}^L} - \lambda_t \frac{\partial \nu_t^H}{\partial b_{1,t}^L} - \mu_t n_t^L \frac{\partial \nu_{1,t}^L}{\partial b_{2,t}^L} - \delta_t \eta n_t^L \frac{\partial \nu_{1,t}^L}{\partial b_{1,t}^L} + \gamma_t n_t^L \left( \tau_{x,t} \frac{\partial \nu_{1,t}^L}{\partial b_{1,t}^L} + \tau_{y,t} \frac{\partial \nu_{1,t}^L}{\partial b_{1,t}^L} - 1 \right) = 0 \] \hspace{1cm} (A1)

\[ \frac{\partial W}{\partial V_{t+1}^L} \frac{\partial \nu_{2,t+1}^L}{\partial b_{2,t+1}^L} - \lambda_t \frac{\partial \nu_{2,t+1}^L}{\partial b_{2,t+1}^L} - \mu_{t+1} n_t^L \frac{\partial \nu_{2,t+1}^L}{\partial b_{2,t+1}^L} - \delta_{t+1} \eta n_t^L \frac{\partial \nu_{2,t+1}^L}{\partial b_{2,t+1}^L} + \gamma_{t+1} n_t^L \left( \tau_{x,t+1} \frac{\partial \nu_{2,t+1}^L}{\partial b_{2,t+1}^L} + \tau_{y,t+1} \frac{\partial \nu_{2,t+1}^L}{\partial b_{2,t+1}^L} - 1 \right) = 0 \] \hspace{1cm} (A2)

\[ - \frac{\partial W}{\partial V_t^L} \frac{\partial \nu_{1,t}^L}{\partial z_{1,t}^L} + \lambda_t \frac{\partial \nu_t^H}{\partial z_{1,t}^L} \phi + n_t^L \mu_t \frac{\partial \nu_{1,t}^L}{\partial z_{1,t}^L} + \delta_t \eta n_t^L \frac{\partial \nu_{1,t}^L}{\partial z_{1,t}^L} \]

\[ + n_t^L \gamma_t \left( w_t^L - \tau_{x,t} \frac{\partial \nu_{1,t}^L}{\partial z_{1,t}^L} - \tau_{y,t} \frac{\partial \nu_{1,t}^L}{\partial z_{1,t}^L} \right) = 0 \] \hspace{1cm} (A3)

46
The first-order conditions of $L_0$ with respect to commodity tax rates $t_{x,t}$ and $t_{y,t}$ are given by

\[
\sum_j \beta \frac{\partial W}{\partial V^H_j} \frac{\partial \gamma^j_{x,t}}{\partial q_{x,t}} + \sum_j \frac{\partial W}{V^H_j} \frac{\partial \gamma^j_{y,t}}{\partial q_{y,t}} + \lambda \left( \frac{\partial \alpha^H_{x,t}}{\partial q_{x,t}} - \frac{\partial \alpha^H_{y,t}}{\partial q_{y,t}} \right) + \beta \lambda_{-1} \left( \frac{\partial \alpha^H_{x,t}}{\partial q_{x,t}} - \frac{\partial \alpha^H_{y,t}}{\partial q_{y,t}} \right) - \mu \sum_j \left( \frac{\partial x^j_{x,t}}{\partial q_{x,t}} + \frac{\partial x^j_{y,t}}{\partial q_{x,t}} \right) - \delta \eta \sum_j \alpha^j \left( \frac{\partial x^j_{x,t}}{\partial q_{x,t}} + \frac{\partial x^j_{y,t}}{\partial q_{x,t}} \right) + \gamma \sum_j \left[ n^j_{x,t} \frac{\partial x^j_{x,t}}{\partial q_{x,t}} + n^j_{x,t-1} \frac{\partial x^j_{y,t}}{\partial q_{x,t}} \right] = 0, \tag{A7}
\]

\[
\sum_j \beta \frac{\partial W}{\partial V^H_j} \frac{\partial \gamma^j_{x,t}}{\partial q_{y,t}} + \sum_j \frac{\partial W}{V^H_j} \frac{\partial \gamma^j_{y,t}}{\partial q_{y,t}} + \lambda \left( \frac{\partial \alpha^H_{x,t}}{\partial q_{x,t}} - \frac{\partial \alpha^H_{y,t}}{\partial q_{y,t}} \right) + \beta \lambda_{-1} \left( \frac{\partial \alpha^H_{x,t}}{\partial q_{x,t}} - \frac{\partial \alpha^H_{y,t}}{\partial q_{y,t}} \right) - \mu \sum_j \left( \frac{\partial x^j_{x,t}}{\partial q_{x,t}} + \frac{\partial x^j_{y,t}}{\partial q_{x,t}} \right) - \delta \eta \sum_j \alpha^j \left( \frac{\partial x^j_{x,t}}{\partial q_{x,t}} + \frac{\partial x^j_{y,t}}{\partial q_{x,t}} \right) + \gamma \sum_j \left[ n^j_{x,t} \frac{\partial x^j_{x,t}}{\partial q_{y,t}} + n^j_{x,t-1} \frac{\partial x^j_{y,t}}{\partial q_{y,t}} \right] = 0. \tag{A8}
\]
The first-order conditions of $L_0$ with respect to environmental $E_t$ and positional externalities $\bar{y}_t$ are

$$
\sum_j \beta \frac{\partial W}{\partial V_{t-1}^j} \frac{\partial v_{t}^j}{\partial E_t} + \sum_j \frac{\partial W}{\partial V_t^j} \frac{\partial v_t^j}{\partial E_t} + \lambda_t \left( \frac{\partial v_t^H}{\partial E_t} - \frac{\partial \hat{v}_t^H}{\partial E_t} \right) + \beta \lambda_{t-1} \left( \frac{\partial \hat{v}_t^H}{\partial E_t} - \frac{\partial \hat{v}_{t-1}^H}{\partial E_t} \right) \\
+ \mu_t \left[ 1 - \sum_j \left( n_t^j \frac{\partial x_{1,t}^j}{\partial E_t} + n_{t-1}^j \frac{\partial x_{2,t}^j}{\partial E_t} \right) \right] - (1 - \rho) \mu_{t+1} \\
- \delta_t \eta \sum_j \alpha^j \left( n_t^j \frac{\partial y_{1,t}^j}{\partial E_t} + n_{t-1}^j \frac{\partial y_{2,t}^j}{\partial E_t} \right) \\
+ \gamma \sum_j \left[ \tau_{x,t} \left( n_t^j \frac{\partial x_{1,t}^j}{\partial E_t} + n_{t-1}^j \frac{\partial x_{2,t}^j}{\partial E_t} \right) + \tau_{y,t} \left( n_t^j \frac{\partial y_{1,t}^j}{\partial E_t} + n_{t-1}^j \frac{\partial y_{2,t}^j}{\partial E_t} \right) \right] = 0, \quad (A9)
$$

$$
\sum_j \beta \frac{\partial W}{\partial V_{t-1}^j} \frac{\partial v_{t}^j}{\partial \bar{y}_t} + \sum_j \frac{\partial W}{\partial V_t^j} \frac{\partial v_t^j}{\partial \bar{y}_t} \\
+ \lambda_t \left( \frac{\partial v_t^H}{\partial \bar{y}_t} - \frac{\partial \hat{v}_t^H}{\partial \bar{y}_t} \right) + \beta \lambda_{t-1} \left( \frac{\partial \hat{v}_t^H}{\partial \bar{y}_t} - \frac{\partial \hat{v}_{t-1}^H}{\partial \bar{y}_t} \right) \\
- \mu_t \sum_j \left( n_t^j \frac{\partial x_{1,t}^j}{\partial \bar{y}_t} + n_{t-1}^j \frac{\partial x_{2,t}^j}{\partial \bar{y}_t} \right) \\
+ \delta_t \left[ 1 - \eta \sum_j \alpha^j \left( n_t^j \frac{\partial y_{1,t}^j}{\partial \bar{y}_t} + n_{t-1}^j \frac{\partial y_{2,t}^j}{\partial \bar{y}_t} \right) \right] - \delta_{t+1} (1 - \eta) \\
+ \gamma \sum_j \left[ \tau_{x,t} \left( n_t^j \frac{\partial x_{1,t}^j}{\partial \bar{y}_t} + n_{t-1}^j \frac{\partial x_{2,t}^j}{\partial \bar{y}_t} \right) + \tau_{y,t} \left( n_t^j \frac{\partial y_{1,t}^j}{\partial \bar{y}_t} + n_{t-1}^j \frac{\partial y_{2,t}^j}{\partial \bar{y}_t} \right) \right] = 0, \quad (A10)
$$

The first-order condition of $L_0$ with respect to physical capital $K_{t+1}$ is

$$
-\gamma + \gamma_{t+1} (1 + r_{t+1}) = 0. \quad (A11)
$$

A.2 The proof for Proposition 1

To derive the shadow price of environmental externalities $\mu_t/\gamma$, we rewrite (A1), (A2), (A4) and (A5) as

$$
\frac{\partial W}{\partial V_t^L} \frac{\partial v_t^L}{\partial b_{1,t}^L} = \lambda_t \frac{\partial v_t^H}{\partial b_{1,t}^L} + \mu_t n_t^L \frac{\partial x_{1,t}^L}{\partial b_{1,t}^L} + \delta_t \eta \alpha^L n_t^L \frac{\partial y_{1,t}^L}{\partial b_{1,t}^L} - \gamma n_t^L \left( \tau_{x,t} \frac{\partial x_{1,t}^L}{\partial b_{1,t}^L} + \tau_{y,t} \frac{\partial y_{1,t}^L}{\partial b_{1,t}^L} - 1 \right), \quad (A1a)
$$
Lagging (A2a) and (A5a) one period, respectively, generates

\[
\frac{\partial W}{\partial V^L_t} \frac{\partial v^L_{2,t+1}}{\partial b^L_{2,t+1}} = \lambda_t \frac{\partial v^H_{2,t+1}}{\partial b^L_{2,t+1}} \beta + \mu_t n^L_t \frac{\partial x^L_{2,t+1}}{\partial b^L_{2,t+1}} + \delta + n^L_t \alpha^\prime n^L_t \frac{\partial y^L_{2,t+1}}{\partial b^L_{2,t+1}}
\]

\[
- \gamma + n^L_t \left( \tau_{x,t+1} \frac{\partial x^L_{2,t+1}}{\partial b^L_{2,t+1}} + \tau_{y,t+1} \frac{\partial y^L_{2,t+1}}{\partial b^L_{2,t+1}} - 1 \right)
\]

\[
\left( \frac{\partial W}{\partial V^H_t} + \lambda_t \right) \frac{\partial v^H_{2,t+1}}{\partial b^H_{2,t+1}} = \mu_t n^H_t \frac{\partial x^H_{2,t+1}}{\partial b^H_{2,t+1}} + \delta + n^H_t \alpha^\prime n^H_t \frac{\partial y^H_{2,t+1}}{\partial b^H_{2,t+1}}
\]

\[
- \gamma n^H_t \left( \tau_{x,t+1} \frac{\partial x^H_{2,t+1}}{\partial b^H_{2,t+1}} + \tau_{y,t+1} \frac{\partial y^H_{2,t+1}}{\partial b^H_{2,t+1}} - 1 \right)
\]

\[
\left( \frac{\partial W}{\partial V^H_{t-1}} + \lambda_{t-1} \right) \frac{\partial v^H_{2,t+1}}{\partial b^H_{2,t+1}} = \mu_t n^H_{t-1} \frac{\partial x^H_{2,t+1}}{\partial b^H_{2,t+1}} + \delta + n^H_{t-1} \alpha^\prime n^H_{t-1} \frac{\partial y^H_{2,t+1}}{\partial b^H_{2,t+1}}
\]

\[
- \gamma n^H_{t-1} \left( \tau_{x,t} \frac{\partial x^H_{2,t+1}}{\partial b^H_{2,t+1}} + \tau_{y,t} \frac{\partial y^H_{2,t+1}}{\partial b^H_{2,t+1}} - 1 \right)
\]

Lagging (A2a) and (A5a) one period, respectively, generates

We rewrite (A9) as

\[
\frac{\partial W}{\partial V^L_t} \frac{\partial v^L_t}{\partial E_t} - \lambda_t \frac{\partial v^H_t}{\partial E_t} + \left( \frac{\partial W}{\partial V^H_t} + \lambda_t \right) \frac{\partial v^H_t}{\partial E_t}
\]

\[
+ \beta \frac{\partial W}{\partial V^L_{t-1}} \frac{\partial v^L_{2,t}}{\partial E_t} - \lambda_{t-1} \frac{\partial v^H_{2,t}}{\partial E_t} + \left( \frac{\partial W}{\partial V^H_{t-1}} + \lambda_{t-1} \right) \frac{\partial v^H_{2,t}}{\partial E_t}
\]

\[
+ \mu_t \left[ 1 - \sum_j \left( n^j_t \frac{\partial x^j_{1,t}}{\partial E_t} + n^j_{t-1} \frac{\partial x^j_{2,t}}{\partial E_t} \right) \right] - (1 - \rho) \mu_{t+1} - \delta \eta \sum_j \alpha^j \left( n^j_t \frac{\partial y^j_{1,t}}{\partial E_t} + n^j_{t-1} \frac{\partial y^j_{2,t}}{\partial E_t} \right)
\]

\[
+ \gamma \sum_j \left[ \tau_{x,t} \left( n^j_t \frac{\partial x^j_{1,t}}{\partial E_t} + n^j_{t-1} \frac{\partial x^j_{2,t}}{\partial E_t} \right) + \tau_{y,t} \left( n^j_t \frac{\partial y^j_{1,t}}{\partial E_t} + n^j_{t-1} \frac{\partial y^j_{2,t}}{\partial E_t} \right) \right] = 0
\]
and substitute (A1a), (A2a’), (A4a), (A5a’) and (A11) together with the expression for the marginal willingness to pay for pollution abatement $\frac{\partial v_{1,t}^j}{\partial E_t} = - \frac{\partial v_{1,t}^j}{\partial b_{1,t}} MW_{j_{Eb,1,t}}^j + \frac{\partial v_{1,t}^j}{\partial \gamma_{t+1}} MW_{j_{Eb,2,t}}^j$, $\frac{\partial v_{1,t}^j}{\partial E_t} = - \frac{\partial v_{1,t}^j}{\partial b_{2,t}} MW_{j_{Eb,1,t}}^j + \frac{\partial v_{1,t}^j}{\partial \gamma_{t+1}} MW_{j_{Eb,2,t}}^j$ into (A9a). To divide $\gamma_t$ on both sides of the resulting expression, we reach

$$\frac{\mu_t}{\gamma_t} = \sigma_t^E \sum_j \left( n_{t,1}^j MW_{j_{Eb,1,t}}^j + n_{t-1}^j MW_{j_{Eb,2,t}}^j \right) + \sigma_t^E \lambda_{t,1}^* \left( MW_{j_{Eb,1,t}}^L - MW_{j_{Eb,1,t}}^H \right)$$

$$- \sigma_t^E \tau_{t,j} \left( n_{t,1}^j \frac{\partial x_{1,t}}{\partial E_t} + n_{t-1}^j \frac{\partial x_{2,t}}{\partial E_t} \right) - \sigma_t^E \sum_j \left( \tau_{j,t} - \alpha^j \frac{\delta_t}{\gamma_t} \left( n_{t,1}^j \frac{\partial y_{1,t}}{\partial E_t} + n_{t-1}^j \frac{\partial y_{2,t}}{\partial E_t} \right) \right)$$

$$+ \sigma_t^E \frac{\mu_{t+1}}{\gamma_{t+1}} \frac{1 - \rho}{1 + r_{t+1}}$$

(A12)

where $\lambda_{t,1}^* = \frac{\lambda_{t,1}^*}{n_t}$ and $\frac{1}{\sigma_t^E} = 1 - \sum_j \left( n_{t,1}^j \frac{\partial x_{1,t}}{\partial E_t} + n_{t-1}^j \frac{\partial x_{2,t}}{\partial E_t} \right)$.

Forwarding (A12) by one period and substituting $\mu_{t+1}/\gamma_{t+1}$ back into (A12) yield

$$\frac{\mu_t}{\gamma_t} = \sigma_t^E \sum_j \left( n_{t,1}^j MW_{j_{Eb,1,t}}^j + n_{t-1}^j MW_{j_{Eb,2,t}}^j \right)$$

$$+ \sigma_t^E \tau_{t+1} \sum_j \left( n_{t,1}^j \frac{\partial x_{1,t+1}}{\partial E_{t+1}} + n_{t-1}^j \frac{\partial x_{2,t+1}}{\partial E_{t+1}} \right)$$

$$- \sigma_t^E \tau_{t,1} \sum_j \left( n_{t,1}^j \frac{\partial x_{1,t+1}}{\partial E_{t+1}} + n_{t-1}^j \frac{\partial x_{2,t+1}}{\partial E_{t+1}} \right) \frac{1 - \rho}{1 + r_{t+1}}$$

$$- \sigma_t^E \sum_j \left( \tau_{j,t} - \alpha^j \frac{\delta_t}{\gamma_t} \left( n_{t,1}^j \frac{\partial y_{1,t}}{\partial E_t} + n_{t-1}^j \frac{\partial y_{2,t}}{\partial E_t} \right) \right)$$

$$- \sigma_t^E \tau_{t,1} \sum_j \left( \tau_{j,t+1} - \alpha^j \frac{\delta_t}{\gamma_t+1} \left( n_{t,1}^j \frac{\partial y_{1,t+1}}{\partial E_{t+1}} + n_{t-1}^j \frac{\partial y_{2,t+1}}{\partial E_{t+1}} \right) \right) \frac{1 - \rho}{1 + r_{t+1}}$$

$$+ \sigma_t^E \frac{\mu_{t+2}}{\gamma_{t+2}} \frac{1 - \rho}{1 + r_{t+2}}.$$  

(A13)
Summing (A13) over periods and the assumption that time has been long enough such that \(\frac{(1-\rho)^n\Pi_{t=1}^n\sigma_{\bar{t},t}^y}{\Pi_{t=1}^n(1+r_{t+1})} \to 0\) indicates equation (25).

The proof of the shadow price of positional externality at time \(t\), \(\delta_t/\gamma_t\) follows the same method as above. The distinction is the use of the marginal willingness to pay for alleviating social status \(MW P_j^{\bar{y},t}\) and feedback effects of the positional externality \(\sigma_{\bar{t},t}^\gamma\).

\[\square\]

A.3 The proof of Proposition 2

To derive the formulas for commodity taxation, we need to multiply (A1) by \(x_{1,t}\) and (A4) by \(x_{2,t}^H\), respectively. We lag (A2) and (A5) one period behind and multiply by \(x_{1,t}^L\) and \(x_{2,t}^H\), respectively. Adding these outcomes to (A7) gives the following result

\[
\tau_{x,t} \sum_j \left( n_j \frac{\partial \bar{\gamma}_{1,j}^L}{\partial q_{x,t}} + n_j' \frac{\partial \bar{\gamma}_{2,j}^L}{\partial q_{x,t}} \right) + \tau_{y,t} \sum_j \left( n_j \frac{\partial \bar{\gamma}_{1,j}^L}{\partial q_{y,t}} + n_j' \frac{\partial \bar{\gamma}_{2,j}^L}{\partial q_{y,t}} \right) = \\
\lambda \frac{\partial \bar{\gamma}_{1,t}^H}{\partial b_{1,t}} (x_{1,t}^L - x_{1,t}^H) + \mu \frac{\partial \bar{\gamma}_{1,t}^L}{\partial b_{1,t}} (x_{1,t}^L - x_{1,t}^H) + \gamma \frac{\partial \bar{\gamma}_{2,t}^L}{\partial q_{y,t}} (x_{2,t}^L - x_{2,t}^H) + \eta \frac{\partial \bar{\gamma}_{2,t}^H}{\partial q_{y,t}} (x_{2,t}^L - x_{2,t}^H)
\]

(A14)

where we have used the property of Roy’s identity

\[
\frac{\partial v_{1,t}^L}{\partial q_{x,t}} = -x_{1,t}^L \frac{\partial v_{2,t}^L}{\partial q_{x,t}} = -x_{2,t}^L \frac{\partial v_{2,t}^L}{\partial q_{x,t}} = -x_{2,t}^L \frac{\partial v_{2,t}^L}{\partial q_{x,t}} = -x_{2,t}^L \frac{\partial v_{2,t}^L}{\partial q_{x,t}}
\]

and \(x_{1,t}^L = x_{2,t}^H\) for which the low-ability type and the mimicker face the same allocation in their second period of life. Similarly, we switch the multiplying factor \(x\) with \(y\) while manipulating (A1), (A4), (A2), and (A5) as before. The summation of them is made to (A8) instead. Then we reach

\[
\tau_{x,t} \sum_j \left( n_j \frac{\partial \bar{\gamma}_{1,j}^L}{\partial q_{x,t}} + n_j' \frac{\partial \bar{\gamma}_{2,j}^L}{\partial q_{x,t}} \right) + \tau_{y,t} \sum_j \left( n_j \frac{\partial \bar{\gamma}_{1,j}^L}{\partial q_{y,t}} + n_j' \frac{\partial \bar{\gamma}_{2,j}^L}{\partial q_{y,t}} \right) = \\
\lambda \frac{\partial \bar{\gamma}_{1,t}^H}{\partial b_{1,t}} (y_{1,t}^L - y_{1,t}^H) + \mu \frac{\partial \bar{\gamma}_{1,t}^L}{\partial b_{1,t}} (y_{1,t}^L - y_{1,t}^H) + \gamma \frac{\partial \bar{\gamma}_{2,t}^L}{\partial q_{y,t}} (y_{2,t}^L - y_{2,t}^H) + \eta \frac{\partial \bar{\gamma}_{2,t}^H}{\partial q_{y,t}} (y_{2,t}^L - y_{2,t}^H)
\]

(A15)

where we have applied Roy’s identity to the consumption of good \(y\)

\[
\frac{\partial v_{1,t}^L}{\partial q_{y,t}} = -y_{1,t}^L \frac{\partial v_{2,t}^L}{\partial q_{y,t}} = -y_{2,t}^L \frac{\partial v_{2,t}^L}{\partial q_{y,t}} = -y_{2,t}^L \frac{\partial v_{2,t}^L}{\partial q_{y,t}} = -y_{2,t}^L \frac{\partial v_{2,t}^L}{\partial q_{y,t}}
\]

and \(y_{1,t}^L = \bar{y}_{1,t}^L\) as the low-ability type and the potential mimicker in the second stage do not differ in terms of their consumption behavior (nor their marginal willingness to pay to avoid pollution/ social comparison). The commodity
Cramer’s rule is helpful in deriving the explicit solutions for \( \tau_{x,t} \) and \( \tau_{y,t} \) as equations (29) and (30) in Proposition 2. To have a clear expression for the commodity taxation in a second-best setting, following Aronsson and Sjögren (2018), we define the weight of price effects on compensated conditional demand over two types of consumers as

\[
\Theta^L = \left( n^L_i \frac{\partial x^L_i}{\partial q_{1,t}} + n^L_i \frac{\partial x^L_i}{\partial q_{2,t}} \right) \sum_j \left( n^L_i \frac{\partial q_{1,t}}{\partial q_{1,t}} + n^L_i \frac{\partial q_{2,t}}{\partial q_{2,t}} \right) \left( n^L_i \frac{\partial q_{1,t}}{\partial q_{1,t}} + n^L_i \frac{\partial q_{2,t}}{\partial q_{2,t}} \right) \sum_j \left( n^L_i \frac{\partial x^L_i}{\partial q_{1,t}} + n^L_i \frac{\partial x^L_i}{\partial q_{2,t}} \right)
\]

\[
\Theta^H = \left( n^H_i \frac{\partial x^H_i}{\partial q_{1,t}} + n^H_i \frac{\partial x^H_i}{\partial q_{2,t}} \right) \sum_j \left( n^H_i \frac{\partial q_{1,t}}{\partial q_{1,t}} + n^H_i \frac{\partial q_{2,t}}{\partial q_{2,t}} \right) \left( n^H_i \frac{\partial q_{1,t}}{\partial q_{1,t}} + n^H_i \frac{\partial q_{2,t}}{\partial q_{2,t}} \right) \sum_j \left( n^H_i \frac{\partial x^H_i}{\partial q_{1,t}} + n^H_i \frac{\partial x^H_i}{\partial q_{2,t}} \right)
\]

with \( \Theta^H + \Theta^L = 1 \), where \( \Theta = \sum_j \left( n^L_i \frac{\partial x^L_i}{\partial q_{1,t}} + n^L_i \frac{\partial x^L_i}{\partial q_{2,t}} \right) \sum_j \left( n^L_i \frac{\partial q_{1,t}}{\partial q_{1,t}} + n^L_i \frac{\partial q_{2,t}}{\partial q_{2,t}} \right) \left( n^L_i \frac{\partial q_{1,t}}{\partial q_{1,t}} + n^L_i \frac{\partial q_{2,t}}{\partial q_{2,t}} \right) \). Note that \( \Theta > 0 \) because we suppose that the own-price effects on the compensated conditional demand, which are negative under second-order conditions, dominate the cross-price effects.

A.4 The proof of Proposition 3 (the first part)

To derive the optimal marginal labor income tax rates, we need to rewrite (A1) and (A3) as

\[
\frac{\partial W}{\partial V^L_t} \frac{\partial v^L_{t,t}}{\partial b^L_{t,t}} = \lambda_t \frac{\partial H^L_t}{\partial b^L_{t,t}} + \mu_t n^L_t \frac{\partial x^L_t}{\partial b^L_{t,t}} + \delta \eta \alpha_t n^L_t \frac{\partial y^L_t}{\partial b^L_{t,t}} - \gamma n^L_t \left( \tau_{x,t} \frac{\partial x^L_t}{\partial b^L_{t,t}} + \tau_{y,t} \frac{\partial y^L_t}{\partial b^L_{t,t}} - 1 \right)
\]

\[
\frac{\partial W}{\partial V^L_t} \frac{\partial v^L_{t,t}}{\partial q^L_{t,t}} = \lambda_t \frac{\partial H^L_t}{\partial q^L_{t,t}} + \mu_t n^L_t \frac{\partial x^L_t}{\partial q^L_{t,t}} + \delta \eta \alpha_t n^L_t \frac{\partial y^L_t}{\partial q^L_{t,t}} + \gamma n^L_t \left( w^L_t - \tau_{x,t} \frac{\partial x^L_t}{\partial q^L_{t,t}} - \tau_{y,t} \frac{\partial y^L_t}{\partial q^L_{t,t}} \right)
\]
The rearrangement of (A19) creates a connection with the individual optimal-labor income taxation on the low-ability type reads

\[
\frac{\partial v_{L,t}}{\partial z_{L,t}} = MRS_{z,t}^L = \frac{\lambda_t \phi \frac{\partial v^H_{1,t}}{\partial b_{1,t}^L} + \mu_t \frac{\partial v^L_{1,t}}{\partial b_{1,t}^L} + \delta_t \eta \alpha_L \frac{\partial v^L_{1,t}}{\partial b_{1,t}^L} + \gamma_t \left[ w_{L,t} - \tau_{x,t} \frac{\partial x_{L,t}}{\partial z_{L,t}} - \tau_{y,t} \frac{\partial y_{L,t}}{\partial z_{L,t}} \right] + \gamma_t \left[ \tau_{x,t} \frac{\partial x_{L,t}}{\partial b_{1,t}^L} + \tau_{y,t} \frac{\partial y_{L,t}}{\partial b_{1,t}^L} + 1 \right]}{\lambda_t \phi \frac{\partial v^H_{1,t}}{\partial b_{1,t}^L} + \mu_t \frac{\partial v^L_{1,t}}{\partial b_{1,t}^L} + \delta_t \eta \alpha_L \frac{\partial v^L_{1,t}}{\partial b_{1,t}^L} - \gamma_t \left( \tau_{x,t} \frac{\partial x_{L,t}}{\partial b_{1,t}^L} + \tau_{y,t} \frac{\partial y_{L,t}}{\partial b_{1,t}^L} - 1 \right)}.
\]

(A19)

The rearrangement of (A19) creates a connection with the individual optimality (12)

\[
\frac{\lambda_t}{\gamma_t} \frac{1}{n_t^L} \left( \frac{\partial v^H_{1,t}}{\partial b_{1,t}^L} MRS_{z,b}^L - \phi \frac{\partial v^H_{1,t}}{\partial z_{1,t}^L} \right) + \frac{\partial v^L_{1,t}}{\partial z_{1,t}^L} (\tau_{x,t} - \mu_t) + \frac{\partial y^L_{1,t}}{\partial z_{1,t}^L} (\tau_{y,t} - \alpha_L \frac{\delta_t}{\gamma_t} \eta) MRS_{z,b}^L = w_{L,t} - MRS_{z,b}^L.
\]

(A20)

With \( MRS_{z,b_1}^{H,t} = \frac{\partial v^H_{1,t}}{\partial z_{1,t}^L} / \frac{\partial v^H_{1,t}}{\partial b_{1,t}^L} \) and \( w_{L,t} - MRS_{z,b}^L = T_{L,t} \left( w_{L,t} l_{1,t} \right) \), the marginal labor income taxation on the low-ability type reads

\[
T_{L,t} \left( w_{L,t} l_{1,t} \right) = \frac{1}{w_{L,t}} \frac{\lambda_t}{\gamma_t} \frac{1}{n_t^L} \frac{\partial v^H_{1,t}}{\partial b_{1,t}^L} \left( MRS_{z,b_1}^L - \phi \overline{MRS}_{z,b_1}^{H,t} \right) + \frac{1}{w_{L,t}} \left( \frac{\partial v^L_{1,t}}{\partial z_{1,t}^L} - \frac{\partial v^L_{1,t}}{\partial b_{1,t}^L} MRS_{z,b_1}^L \right) \left( \tau_{x,t} - \mu_t \gamma_t \right) + \frac{1}{w_{L,t}} \left( \frac{\partial v^L_{1,t}}{\partial z_{1,t}^L} - \frac{\partial v^L_{1,t}}{\partial b_{1,t}^L} MRS_{z,b_1}^L \right) \left( \tau_{y,t} - \alpha_L \frac{\delta_t}{\gamma_t} \eta \right)
\]

(A21)

Applying the compensated conditional demands generates the expression (33)

in Proposition 3, where we use the notation \( \lambda^*_t l_{1,t+m} = \frac{\lambda^*_t}{\gamma_t+m} \frac{\partial v^H_{1,t+m}}{\partial b_{1,t+m}^L} \). The derivation for the optimal marginal labor income tax rate for the high-ability types is analogous.

A.5 The proof of Proposition 3 (the second part)

To derive the marginal capital income tax rate for the low ability type, we divide (A1a) by (A2a) to form the marginal rate of substitution between in-
tertemporal disposable income

\[ MRS_{L_t}^{L_t} = \frac{\lambda_t}{\gamma_t} \frac{\partial v^H_t}{\partial b_{L_t}^H} + \mu_t \frac{\partial x^L_t}{\partial b_{L_t}^L} + \delta_t \alpha_t \frac{\partial y^L_t}{\partial b_{L_t}^L} + \left( 1 - \tau_{y,t}^L \frac{\partial x^L_t}{\partial b_{L_t}^L} - \tau_{y,t}^L \frac{\partial y^L_t}{\partial b_{L_t}^L} \right) \gamma_t \]

\[ \frac{\lambda_t}{\gamma_t} \frac{\partial v^H_{t+1}}{\partial b_{L_{t+1}}^H} + \mu_{t+1} \frac{\partial x^L_{t+1}}{\partial b_{L_{t+1}}^L} + \delta_{t+1} \alpha_t \frac{\partial y^L_{t+1}}{\partial b_{L_{t+1}}^L} + \left( 1 - \tau_{y,t+1}^L \frac{\partial x^L_{t+1}}{\partial b_{L_{t+1}}^L} - \tau_{y,t+1}^L \frac{\partial y^L_{t+1}}{\partial b_{L_{t+1}}^L} \right) \gamma_{t+1} \]

By rearrangement and use of (A11), (A22) updates as

\[ \gamma_t - \gamma_{t+1} MRS_{b_1,b_2}^{L_t} = \frac{\lambda_t}{\gamma_t} \frac{\partial v^H_{2,t+1}}{\partial b_{L_{2,t+1}}^H} MRS_{b_1,b_2}^{L_t} B - \frac{\lambda_t}{\gamma_t} \frac{\partial v^H_t}{\partial b_{L_t}^H} - (\mu_t - \gamma_t \tau_{y,t}) \frac{\partial x^L_t}{\partial b_{L_t}^L} + (\mu_{t+1} - \gamma_{t+1} \tau_{y,t+1}) \frac{\partial x^L_{t+1}}{\partial b_{L_{t+1}}^L} MRS_{b_1,b_2}^{L_t} \]

\[ - (\delta_t \alpha_t - \gamma_t \tau_{y,t}) \frac{\partial y^L_t}{\partial b_{L_t}^L} + (\delta_{t+1} \alpha_t \gamma_t - \gamma_{t+1} \tau_{y,t+1}) \frac{\partial y^L_{t+1}}{\partial b_{L_{t+1}}^L} MRS_{b_1,b_2}^{L_t} \]

Replacing \( \frac{\partial v^H_t}{\partial b_{L_t}^L} \) with \( MRS_{b_1,b_2}^{H_t} \frac{\partial v^H_{2,t+1}}{\partial b_{L_{2,t+1}}^H} \) and dividing \( \gamma_t \) for (A23) on both sides generate

\[ 1 - \frac{1}{1 + r_{t+1}} MRS_{b_1,b_2}^{L_t} = \frac{\lambda_t}{\gamma_t} \frac{\beta}{n_t} \frac{\partial v^H_{2,t+1}}{\partial b_{L_{2,t+1}}^H} \left( MRS_{b_1,b_2}^{L_t} - MRS_{b_1,b_2}^{H_t} \right) \]

\[ + \left( \tau_{y,t} - \frac{\mu_t}{\gamma_t} \right) \frac{\partial x^L_t}{\partial b_{L_t}^L} - \frac{1}{1 + r_{t+1}} \left( \tau_{y,t+1} - \frac{\mu_{t+1}}{\gamma_{t+1}} \right) \frac{\partial x^L_{t+1}}{\partial b_{L_{t+1}}^L} \]

\[ + \left( \tau_{y,t} - \delta_t \frac{\delta_t}{n_t} \right) \frac{\partial y^L_t}{\partial b_{L_t}^L} - \frac{1}{1 + r_{t+1}} \left( \tau_{y,t+1} - \delta_{t+1} \frac{\delta_{t+1}}{n_{t+1}} \right) \frac{\partial y^L_{t+1}}{\partial b_{L_{t+1}}^L} \]

(A24)

The combination with the private optimality condition (13) gives equation (33). The derivation of equation (34) follows the same method.

\[ \square \]

A.6 An illustrative example for explaining the corrective purposes in marginal labor and capital income taxation

An illustrative example follows for better intuition on the specific sign of the corrective parts in the non-linear taxation system. Suppose that (i) \( \alpha^H > \alpha^L \) meaning that high-ability people always contribute to the positional externality more than low-ability people at the margin, (ii) \( \partial y^L_{1,t} / \partial b_{1,t}^L > 0 \) meaning that the positional good is a normal good and (iii) \( \partial y^L_{1,t} / \partial z^j_{1,t} < 0 \) meaning that leisure is substitutable for the positional good. The first assumption maintains that the corrective part in \( \tau_{y,t} \) exceeds the low-ability types’ marginal contribution to the comparison reference while it falls short of the marginal damage made by high-ability types’ positional consumption. The second and third
assumptions generate the substitution effect between leisure and the compensated conditional demand for the positional good $\partial \tilde{y}_{1,t} / \partial z_{1,t} < 0$.

Under assumptions (i) and (ii), the government is motivated to raise low-ability types’ consumption of the positional commodity by leaving more disposable income at each period. However, the marginal capital income taxation yields a tradeoff effect on intertemporal disposable income and is invalid to continuously raise their positional consumption. The sign of the third line in (33) will be negative if the socially marginal value of correcting the positional externality at time $t + 1$ matters more than that at time $t$. The marginal capital income tax for the low-ability types $\Phi_{t+1}$ will be set lower to solve the positional externality at period $t + 1$, by raising low-ability types’ disposable income via savings and more consumption on the positional good when they are old at $t + 1$. It will worsen the situation that low-ability consumers have less positional consumption than the socially desired at time $t$.

The marginal labor income taxation will be useful for correcting the positional externality when the commodity and the marginal capital income taxation fail to fully correct the positional externality at time $t$. The combination of different tax instruments maintains the system efficiently, explaining the critical way these instruments interact directly in corrective purposes. The assumptions above yield a lower labor income tax for the low-ability types $T_L^t(\cdot)$ to raise low-ability types’ positional consumption, because the corrective part $(\alpha^H - \alpha^L) \phi^H \frac{\delta}{\theta} \eta \frac{\partial \tilde{y}_{1,t}}{\partial z_{1,t}}$ in (31) is negative. A lower $T_L^t(\cdot)$ motivates them to cut leisure time and stimulate more positional consumption due to the substitution effect $\partial \tilde{y}_{1,t} / \partial z_{1,t} < 0$.

\footnote{Under the assumptions of the illustrative example, the social planner is motivated to lower high-ability types’ consumption of the positional good in each period, as the corrective part from the optimal commodity taxation on the positional good falls short of their social damage at the margin. If the value of rectifying the positional externality next period outweighs that for the current period, the capital tax for the high-ability types will instead be set higher to suppress their consumption of the positional good at time $t + 1$ while leaving their excessive positional consumption unsolved at $t$.}
Essay II. Time Capacity and Firms’ Economic Stagnation

Acknowledgements: I would like to extend my sincere gratitude to Jonas Hjort, Vinayak Iyer, and Golvine de Rochambeau for their insightful comments. I also owe special thanks to Daniel Spiro, Hongning Huang, and Xuemeng Wang for their generous assistance. Additionally, I deeply appreciate the valuable feedback provided by Yikai Wang, Per Engström, and Chuanzhong Li. I gratefully acknowledge the funding from the project “The Economics of Planetary Boundaries”, FORMAS grant 2020-00371, hosted by Daniel Spiro.
1 Introduction
At the firm level, there exist various sources that can hinder economic growth. One such mechanism is a time capacity constraint, referring to the limited amount of time available to process ongoing tasks and information.1 As highlighted by DellaVigna (2009), firms can face (i) a shortage of time and (ii) limited tasks that can be handled. Firms may, of course, specialize, recruit consultants, or expand the workforce, but doing this requires a time investment from management.2 The challenge of time allocation becomes particularly pronounced for small firms, especially those in developing countries, which usually operate in an unfavorable socioeconomic context. Often comprised of a manager or a small group of individuals, these firms confront the pressing dilemma of distributing limited daily hours across a myriad of tasks.

This paper serves as a companion paper to the work of Hjort et al. (2020). While their study performs a field experiment, the current paper constructs a theoretical model that fits their setting, and that is able to rationalize their findings. Hjort et al. (2020) study small and medium-sized firms in Liberia, where the sampled firms primarily comprise two to twenty employees, including at least one employee in addition to the owner. Liberia is characterized by its post-conflict economy, narrow industrial base, and the dominance of firms with limited size. One of the critical challenges these firms face is market access, often constrained by insufficient information about opportunities beyond their local environment.

Hjort et al. (2020) delivered a specialized training program, aiming to assess the training’s direct impact and explore how information frictions might deter firms from engaging in profitable ventures. To this end, a randomized controlled trial (RCT) was carried out: a total of 772 firms were randomly assigned to the treatment group, while another 420 firms were placed in the control group. Those in the treatment group were trained in essential skills for navigating the tender application process successfully.3

---

1 This also has important implications for understanding various research questions at the individual level such as income distribution (Banerjee and Mullainathan, 2008), precautionary saving motives for wealthier people (Yin, 2021), consumption dynamics and equity return (Lynch, 1996; Gabaix and Laibson, 2001), and corporate finance (Hirshleifer et al., 2002).
2 One phenomenon that is relevant to the theory of firms’ capacity constraint can be a crowd-out effect, which means that the increase in one business might squeeze out the other one, such as a sale drop in the private sector due to access to public procurement verified by (Giovanni et al., 2022) (a short-run effect) and (Hoekman et al., 2022) (both short and long-run effects) while Ferraz et al. (2015) does not verify the effect in Brazil.
3 Randomization was conducted at the firm level using a computerized process within the office. To ensure a balanced sample across both groups, the firms were categorized into strata based on sector, geographical zone, and number of employees. Randomization was then executed within each stratum. Of the 1,192 firms initially included in the sample, the researchers successfully conducted follow-up surveys with 789 firms for long-run effects: 284 in the control group and 505 in the treatment group.
Hjort et al. (2020)’s primary outcome variables focus on the firms’ abilities to both apply for and win tenders, measured by the volume of bids firms submit and the number of contracts they subsequently win.\(^4\) In the treatment group, firms’ participation in the training program notably enhanced their ability to secure contracts. Before the treatment, the average number of bid submissions was 1.10 (SD = 2.9), with firms winning approximately 0.43 (SD = 1.77) of these bids over a six-month period. However, after the treatment, firms experienced a significant increase in activity. They submitted an additional 0.55 bids, raising the average to 1.65, and increased their wins by 0.33, boosting the average to 0.76. This represents a substantial rise of over 150 percent in comparison to firms in the control group.

Within the treatment group, the impact varied across firms regarding effects after one year. Firms with prior participation in bidding contracts experienced a threefold increase in formal contracts in one-year out. The estimated treatment effects are both large in magnitude and statistically significant, but the training does not yield significant benefits for firms that had no bidding experience. Three years into the training program, outcomes varied within the treatment group as well. Firms that secured significant contracts in the first year continued to reap benefits, while those that gained no contracts experienced no such advantages. There was a notable uptick in their market reach, sales, and profitability. Meanwhile, the firms that derived benefits from the training program demonstrated proficient utilization of the internet as a distinguishing characteristic, reflecting an advantage in processing information.\(^5\)

The current paper builds a model of a small firm’s decision of which types of contracts to bid for and how this depends on their information processing ability and their prior experience.\(^6\) It explains the following findings of Hjort et al. (2020):

- After the treatment, which is the training program that enhances firms’ ability to process information, firms with prior bidding participation

\(^4\)The study also incorporates a comprehensive evaluation of overall firm performance, assessing aspects such as the number and roles of employees and revenue and profit metrics. Treated firms that participate in the ‘Winning-contracts’ training see a substantial rise in the total value of contracts won, amounting to approximately USD 10,000—a 200 percent increase. Furthermore, their findings indicate a significant impact on employment, with the training contributing to a 400 percent surge in the workforce. This upturn represents an addition of four employees, expanding from an average of one, necessary to fulfill the demands of their formal contracts. However, my paper primarily focuses on the bidding for tenders, given their higher profitability and greater impact on a firm’s overall performance.

\(^5\)A reliable and extensive internet access is considered to be closely linked to a firm’s ability to process information efficiently.

\(^6\)In the baseline, there is a noticeable divergence in business types among small firms: some firms actively pursue tender contracts, while others refrain. This variation can largely be attributed to what I will categorize as information processing abilities, a collective term that reflects a company’s resourcefulness and operational efficiency, significantly influenced by internet access and workforce size etc.
show a surge in bidding activity in the short term. These firms sustain heightened performance in bidding for contracts in the long run.

- The training program fails to stimulate similar improvements among firms without previous bidding experience, neither in the short nor the long term.

2 Related Literature

2.1 Theoretical poverty traps

To the best of my knowledge, there is no existing theoretical framework that addresses a poverty trap at the firm level from the demand side (weak access to potential buyers) to elucidate challenges to a firm’s growth potential. Predominantly, such traps are discussed in socioeconomic contexts, in relation to production technology, and at the individual level.

The literature on the poverty trap can be categorized based on different levels, considering the extent to which individual decisions contribute to the formation of a poverty trap. I start the literature from the macro factors rooted in the socioeconomic environment, typically beyond individual control.

2.1.1 Macro social background

*Infrastructure and Public Goods.* The lack of essential infrastructure can lead to regional poverty traps. Sachs (2005) models infrastructure as a factor that enhances the returns on private investment. In this vein, some poverty trap models that incorporate endogenous TFP highlight the inability to adopt the most efficient technologies as a reason for low TFP and income in impoverished countries (Murphy et al., 1989; Ciccone and Matsuyama, 1996).

*Political Systems and Governance.* Corruption and inefficiency, which can be understood as additional investment burdens, can stifle opportunities and perpetuate inequalities. A more nuanced categorization includes (i) corruption and rent-seeking (Murphy et al., 1993; Tirole, 1996; Bardhan, 1997); and (ii) kinship systems (Baker, 2004; Sen and Hoff, 2006).7

*Liberalization and Globalization.* According to Matsuyama (2004), in a world with many such nations, autarky’s steady state becomes volatile with open international markets. Post-liberalization, minor disruptions can cause instability. Countries hit by shocks might see their loan guarantee capabilities

---

7Baker (2004) presents an interpretation of Africa’s lackluster growth, attributing it to a hindrance in technology diffusion resulting from institutional barriers. Sen and Hoff (2006) study the migration of kin members from rural to modern-sector jobs, emphasizing the role of network externalities in migration decisions. They suggest kin groups may set up barriers to retain their key members. Interestingly, even when kin decisions are majority-driven, these barriers can be misaligned with the group’s collective welfare.
shift, changing their standing in global finance. This can polarize the global economy, dividing it into rich and poor nations.\footnote{In the model, agents, facing credit market imperfections, work in their first period of life and consume in the second. They can invest in capital markets or indivisible projects, the latter requiring borrowing beyond wages. Given risk neutrality, agents choose projects based on profitability and borrowing limits. When borrowing constraints exist, they might hinder project initiation.}

**Social and Cultural Norms.** Some societies’ beliefs restrict access to resources for specific groups, perpetuating poverty, e.g., Platteau (2015) incorporates social penalties in utility functions for actions against societal norms. In my framework, small firms in developing countries operate within a shared socioeconomic landscape. They consistently grapple with these issues. However, these overarching challenges are not my main concern. Drawing parallels to macro factors but with more relevance to micro-level decisions, I subsequently investigate how external frictions can engender a poverty trap on a smaller scale.

### 2.1.2 External frictions

External frictions are usually connected with *indivisible technology with an increasing return to inputs*. There is extensive literature on the increasing returns to input factors. In conjunction with the concept of a poverty trap, Ray (1998) and Carter and Barrett (2005) posit that when initial assets fall below a specific threshold, households are unable to invest productively, consequently remaining ensnared in poverty.\footnote{Similarly, Dasgupta and Ray (1986) proposed a minimum consumption requirement for sustaining optimal productivity, akin to the nutrition-based efficiency wage argument. Criticisms related to the nutrition poverty trap were discussed in the previous section.} Banerjee et al. (2015) show that poor households often do not invest in potentially profitable technologies because they lack the necessary capital. Their work is underscored by S-shaped production functions where early investments yield low returns until a certain threshold is reached. In a similar way, there is a theory for *Nutrition-Based Traps*.\footnote{One empirical research from by Bloom and Canning (2000) centers on Health-Related Traps. The central idea emphasizes how poor health can constrain work and educational opportunities, while poverty, in turn, can hinder access to healthcare. The primary insight from their model is the cyclical nature of the relationship: poor health can lead to diminished income, which then adversely impacts health.} Limited income is thought to cause malnutrition, reducing work capacity and further decreasing income. Dasgupta (1997) established a feedback loop linking nutrition and labor productivity. Dasgupta’s model demonstrates two equilibria: high productivity with good nutrition and low productivity with malnutrition. Critiques, such as that by Srinivasan (1994) have proposed that...
the connection between nutrition and productivity, and vice versa, might be too tenuous to account for more than a minor aspect of the overarching narrative. The health status of any specific employee is unlikely to significantly impact the firm’s overall performance. In developing countries, it is common for small firms to replace employees with poor health, regardless of the underlying reasons, such as malnutrition or general health conditions. Thus, this perspective may not fit well within my current research framework.

In the presence of such input indivisibility, limited access to the credit market becomes a significant issue perpetuating poverty, as the poor face challenges accessing loans due to lacking collateral or credit histories (Besley and Burgess, 2003; Udry, 1994). Azariadis and Stachurski (2005) look into the interaction between credit constraint with (i) human capital and (ii) risk attitudes by considering partial credit imperfections in forming a possible poverty trap.

(i) Credit constraint and human capital. In an economic model with zero interest rates, agents can work using a standard method or start a project that requires investment. They can borrow for this at higher rates due to credit imperfections. While the project’s return is often higher than regular wages, its costs might deter those with lower incomes. Over time, agents pass on part of their wealth to their children, with various factors influencing economic outcomes. This project can represent a business or personal investment like education. However, upfront costs and borrowing rates can hinder the less affluent from leveraging growth opportunities.

Ghatak (2015) builds upon this theme, suggesting that in the presence of non-convex technology, intertemporal borrowing constraints make investments in subsequent periods suboptimal, which can also be seen in Galor and Zeira (1993). Diving deeper into this idea, Ghatak (2015) models a production function with varied productivity determined both by physical and human capital, leading to a distinct surge in output. He further explores the imperfections in the human capital market. Beyond just physical capital, Ghatak (2015) emphasizes human capital as an integral component of the production function. However, given the absence of a tangible market for human capital, its accumulation through education and skills hinges solely on initial investments. Consequently, a poverty trap emerges, anchored not just by physical capital completeness but more critically by the foundational value of human capital.

(ii) Credit constraints and risk attitudes. In a framework similar to that of (i) credit constraints and human capital, Azariadis and Stachurski (2005) posit that there are shocks to both wage and project returns, characterized by a joint distribution. Consequently, the economically disadvantaged often abstain from borrowing to fund projects that could yield higher average returns. This behavior tends to depress mean income and perpetuate long-term poverty. Acemoglu and Zilibotti (1997) offer another specific connection with risk and economic growth. Their work indicates that technological indivisibilities link income diversification possibilities. Essentially, when investment rises, it increases output, enhancing diversification. Given that individuals tend to be
risk-averse, heightened diversification further motivates investment. However, the decentralized nature of this system often results in underinvestment, as individuals overlook the broader impact of their investment on others’ diversification options. As a result, the poor with limited financial means cannot afford to insure themselves, leading to less risky but lower average income. The effect is to reinforce poverty.

However, all small firms have credit constraints in the market. This notion aligns with discussions related to macro-level poverty traps. In addition, this strand of literature primarily concentrates on the physical aspects of production from a supply perspective. My focus lies on the demand side, specifically examining non-physical factors at the firm level that cause them to lose their market foothold.

2.1.3 Scarcity-driven poverty traps
Next, I examine poverty traps more closely aligned with micro-level foundations by looking into scarcity-driven poverty traps.

*Self-control issues.* Prolonged poverty can induce stress or hopelessness, affecting decisions and maintaining poverty. Haushofer and Fehr (2014) modify utility functions based on current conditions or mental states, like stress influencing discount rates. Along this line, Bernheim et al. (2015) look into the challenges people face when making decisions that trade off current enjoyment for future benefits. They describe this as a battle of self-control. The constant decision-making pressure and limited cognitive bandwidth in poverty lead to impulsive spending and missed investment opportunities. As people spend on immediate pleasures rather than long-term gains, they forgo chances to escape poverty, such as through education or business investments. Over time, this creates a cycle where immediate needs and wants take precedence, while the opportunities to break free from poverty diminish.

*Temptation.* Along with such self-control issues, Banerjee and Mullainathan (2007), (2010) look into the interaction with temptation by modeling non-homothetic preference over temptation goods. They explore how temptations, even for basic comforts, can sidetrack long-term financial goals. The struggle between giving in to temptations and exercising restraint can dictate many economic choices. The poor might spend on non-essentials, even if basic needs are not fulfilled.

*The demand for conspicuous consumption.* The desire to attain or maintain social status can lead to overconsumption of status goods, even when income is low. This can result in under-investment in productive assets or essentials, perpetuating a cycle of poverty. Instead of investing in things that could elevate their long-term income (like education or health), individuals might spend their resources on short-term status symbols, making it hard for them to escape poverty.

Moav and Neeman (2010) investigate the relationship between social status-seeking behaviors (via conspicuous consumption) and persistent poverty. In-
individuals often spend on visible goods to signal their social status. In certain societies, individuals may overspend on these goods, even at the expense of essential items or savings, in order to maintain or improve their perceived social status. Omer and Neeman (2012) explore deeper into the determinants of saving rates, focusing on the interplay between conspicuous consumption and human capital accumulation as drivers of persistent poverty. The paper suggests that in societies with high emphasis on conspicuous consumption, the desire to display status can reduce saving rates. Lower savings, in turn, diminish the resources available to invest in human capital. Without investments in human capital, individuals are less likely to access higher-paying opportunities, leading them to remain in poverty.

The aforementioned considerations do not easily translate to the economic decisions of firms. Firms strive to maximize profits, often taking a more strategic and rational approach to their choices. While individual consumers might be enticed by luxury or temptation goods, these types of expenditures are not usually on the radar of business managers or entities, regardless of their size.

Models that bear partial relevance to my framework include Banerjee and Mullainathan (2008), which centers on cognitive capacity constraints, and Ghatak (2015), which discusses individual time capacity constraints.

Banerjee and Mullainathan (2008) dive into the relationship between cognitive attention constraints and income distribution. Attention is a scarce cognitive asset that impacts income and productivity. Within their model, individuals must strategically allocate attention between work and personal demands. The trade-off is that prioritizing personal matters can reduce work productivity and vice versa. A notable element they introduce is distraction-saving goods and services—resources that alleviate home distractions. Possessing these resources allows the more affluent to focus better at work, leading to enhanced productivity and income.

On the subject of income dynamics, they utilize an overlapping generational model, where a fraction of resources (food in their model) is passed to offspring, suggesting that resources used by one generation can shape the human capital of the next. The study also touches on how attention constraints guide job choices. Depending on their cognitive capacity and distractions, individuals might opt for high-responsibility jobs requiring constant attention or more lenient, low-responsibility roles. This interplay between attention and job choice further highlights the complex relationship between cognitive constraints and income distribution.

This modeling approach resembles family-centered agent decision-making in overlapping generation models without anticipating how human capital will evolve over time. Ceroni (2001) in a similar strand considers the interrelation between human capital growth and persistent poverty traps. In her overlapping generation model, parents, at optimum, allocate a portion of their income to the next generation, which can be used for personal consumption or educational input to boost human capital and, subsequently, income. The foun-
ational assumption hinges on the evolution of human capital: it advances in a convex trajectory once a specific threshold is crossed. Before this point, everyone is endowed with a consistent human capital level, irrespective of educational input. Her research emphasizes the disparities in human capital investments, highlighting the entrenched and intergenerational patterns that perpetuate poverty across familial and communal lines.

Firms operate differently than entities in overlapping generation scenarios. They must constantly maintain viability in the market or risk being eliminated. Their primary motive is profit maximization, uninfluenced by altruistic considerations.

Ghatak (2015) presents an individual time capacity model, proposing that time is a scarce resource—a perspective that aligns with my focal point. In his model, while certain activities yield high utilities, others optimize time use in subsequent periods. However, the agent in this framework maximizes the immediate utility without any foresight into the future. The benefits accrued from time invested in present activities are seen as unanticipated returns. This approach might resonate with the decisions of impoverished individuals who often prioritize immediate needs over future considerations. Yet, applying this perspective to firms, especially when making decisions, becomes challenging. While firms, particularly smaller ones, may not always project too far into the future due to the pressing need to remain viable in the market, they typically weigh the implications of present decisions on near-future profits and capability evolution.

My study makes a contribution to the scarcity-driven poverty trap model by examining a firm’s anticipation of the ability-accumulation process and intertemporal profit maximization. The study emphasizes time efficiency as a critical factor in the economic underperformance of small firms. The pivotal variable in my model pertains to the ability to process information, which determines a firm’s efficiency in time utilization. While this variable operates similarly to human capital in its function, there are three significant shortcomings in the current human capital model with respect to poverty traps, as summarized from the earlier discussions. (i) It predominantly pertains to physical production limitations at the firm level. The model falls short in explaining how firms might lose access to significant buyers from a demand perspective. (ii) In the context of a family-based overlapping generation model, a portion of income/resources is inherited and can be leveraged to enhance human capital, thus impacting both income and consumption. The maximization problem in this model incorporates altruistic considerations, which significantly diverge from a typical firm’s considerations. (iii) Some literature omits the evolution of the ability function or human capital function during decision-making processes. The resulting ability or human capital is often viewed as an unforeseen yield, which is at odds with a firm’s typical forward-looking considerations.
2.2 Policy instruments to solve poverty traps

Firms can face challenges in overcoming poverty traps like poor individuals. Without external intervention or support, firms in poverty traps may struggle to improve their situation and may continue to experience limited economic opportunities.

In response to poverty traps, Sherraden (2000) argues that asset-based policies, such as savings accounts, can help break the cycle of poverty at individual levels. The author highlights the importance of providing assets and opportunities to low-income individuals and families to enable them to build wealth over time. This has been demonstrated in cash and asset transfer programs (Blattman et al., 2014; Haushofer and Shapiro, 2018; Blattman et al., 2020; Banerjee et al., 2015; Bandiera et al., 2017; Banerjee et al., 2021; Balboni et al., 2022).

The evidence on the effectiveness of microfinance programs for firms in reducing poverty and promoting growth is mixed. Some studies have found that microfinance programs can have a limited impact on firm growth and profitability, and may even lead to increased indebtedness and financial instability (Mersland and Strøm, 2009; Armendariz de Aghion and Morduch, 2010).

In addition to microfinance, some studies verified that providing technology aid alone may not be sufficient to improve firm performance, as it may not address the broader institutional and social barriers that can impede growth and development (Arora and Gambardella, 2010; Oyelaran-Oyeyinka, 2011). Therefore, complementing technology aid with other forms of support, such as training, infrastructure, and access to markets, becomes important. Removing informational barriers is thus expected to significantly affect the economic performance of firms, if the fact that capital or technology aid is relatively costly and improving physical productivity or organizational design takes time is considered.

My model shows that it is not sufficient to realize sustainable growth by starting the more-productive business that requires surpassing the ability threshold of processing information. Instead, passing the more demanding ability threshold to be proficient in the more-productive business is crucial for achieving sustained growth. These findings explain the main empirical results in Hjort et al. (2020), which suggest that firms with prior experience in more-

---

11 Fields and Pfeffermann (2003) examine the role of private firms in promoting economic mobility and reducing poverty in developing countries. The authors explore the factors that contribute to the success of private firms in creating pathways out of poverty, including innovation, entrepreneurship, access to finance, and market opportunities. They also discuss the challenges and barriers that firms face in promoting economic mobility and suggest policy interventions to address these challenges.

12 Some studies have found that microfinance programs may not be effective in reducing poverty and can even lead to increased debt burdens for poor individuals and households. (Bateman and Chang, 2012; Roodman, 2012)
productive business and good use of the internet are more likely to achieve both short and long-term growth from the training program.

3 The model

I start with small firms’ optimization problems confronting time capacity constraints. This concerns the limited hours available in a day for a manager or a self-employed individual. It is about how a self-employed person decides to allocate their time. In developing countries, many small firms essentially operate with only a limited number of employees. How do these individuals prioritize their tasks? This is a decision everyone faces. Firms can be thought of as, in a two-period economy, to decide the allocation of time between two types of businesses: serving the final consumer (S) and bidding for contracts (B). The firm’s profits at time $t$ and $t + 1$ are as follows:

$$\Pi = R_S \sqrt{s_t} + R_B \sqrt{b_t} + R_S \sqrt{s_{t+1}} + R_B \sqrt{b_{t+1}},$$ (1)

where I abstract from discounting effects for simplicity. $R_S$ represents the return by investing time on a basic business $s_t$, i.e., serving final consumers, at the margin. $R_B$ represents the return by spending time on the more-productive business $b_t$, i.e., bidding for contracts, at the margin, with $R_B > R_S$. The square-root function describes an increasing and concave relationship between the time input and return, for instance, when four hours of input are applied, the resulting output is two units of return.

In this context, the firm encounters a time capacity constraint on the amount of time allocated to serve final consumers and bid on contracts in both time periods $t$ and $t + 1$. The time constraint is a piecewise function, depending on whether the firm engages in the more-productive business. It will be

$$\begin{cases} 
  s_t + P (b_t + \beta) \leq h_t T & \text{if } b_t > 0 \\
  s_{t+1} + P (b_{t+1} + \beta) \leq h_{t+1} T & \text{if } b_{t+1} > 0 
\end{cases},$$ (2)

and otherwise,

$$\begin{cases} 
  s_t \leq h_t T & \text{if } b_t = 0 \\
  s_{t+1} \leq h_{t+1} T & \text{if } b_{t+1} = 0
\end{cases}.$$ (3)

The terms $h_t T, h_{t+1} T$ on the right-hand side of (2) and (3) denote the aggregate time capacity for firms. Time endowment $T$ is identical across all firms, but the time capacity is heterogeneous due to the difference in $h$, which is interpreted as the ability to process information that can be reflected as internet access, the scales of skillful managers and so on.\textsuperscript{13} Specifically, firms with

\textsuperscript{13}The separation of the ability to process information from physical productivity highlights the potential dilemma for productive firms that lack information handling capabilities.
a stronger ability to process information are more efficient in utilizing their given hours.\textsuperscript{14}

From equation (2), operating the more-productive business is more demanding. The parameter $\beta > 0$ denotes the fixed time required when a firm operates the more-productive business. Additionally, $P > 1$ refers to an augmented time factor in the more-productive business, as the marginal time cost of maintaining the more-productive business is higher than the basic business. The additional costs that include $P$ and $\beta$ differentiate the time cost function (2) from (3). Equation (3) states that, if a firm does not operate any more-productive business, $b_t = 0, b_{t+1} = 0$, the additional fixed time inputs $\beta$ and augmented factor $P$ would no longer have any effect on the time capacity of the firm.

Investing in $b_t$ can enhance the information processing ability $h_{t+1}$ in the next period. I suppose that bidding in contracts positively interacts with firms’ ability, indicating that a higher ability firm tends to gain more ability improvement and once a firm starts to bid a contract, the ability can be maintained without worsening off than before they start. Therefore I set the accumulation process as

$$h_{t+1} = f (b_t(h_t), h_t) = \begin{cases} \frac{c(h_t)}{T} b_t + \frac{P\beta}{T}, & \text{if } b_t > 0 \\ h_t, & \text{if } b_t = 0 \end{cases}$$ \textsuperscript{15}

(4)

The first term in the first line of equation (4), where $c(h_t) > 0$, indicates that investing in the more-productive business in the first period with $b_t$ can enhance a firm’s ability $h_{t+1}$ to process information in the next period. For example, companies can leverage highly productive operations to gain insights into the current preferences of major clients and their bidding tactics. A firm with superior information processing capabilities is better equipped to gather valuable data when involved in high-performing activities, like contract bidding. Suppose $c(h_t) = h_t - \frac{P\beta}{T}$, where $\frac{P\beta}{T}$ is a ratio of augmented fixed cost to the total time endowment from (2) and (3). To ensure the existence of an interaction between $h_t$ and $b_t$, the function $c$ must be positive, indicating that the firm’s ability should be sufficiently large such that $h_t > \frac{P\beta}{T}$.

The second term is an additive term $\frac{P\beta}{T}$. It indicates that once small firms start the more-productive business at current time period ($b_t > 0$), the ability can be maintained without worsening than the situation before a firm starts bidding for contracts.

The second line of equation (4) indicates that if a firm does not engage in the more-productive business, its ability to process information in the second

\textsuperscript{14}For example, a management group in the firm with a high $h$ is expected to be more efficient in handling different tasks at a lower time cost; meanwhile, it has limited time to deal with everything ($T$ is finite).

\textsuperscript{15}The accumulation process could be other function forms as well. This simplification is for a straightforward analytical solution.
period remains unchanged. This implies that firms may have limited access to new information from international buyers, making it difficult for them to remain competitive in acquiring information. They have to continue to serve local final consumers without making sufficient progress.

4 Results

The interior solutions for \( b_t \) and \( b_{t+1} \) that maximize a firm’s profits (1) subject to the time cost (2) and the ability-accumulation process (4) are

\[
\begin{align*}
    b_{t+1} &= \frac{R_B^2}{PR_B^2 + P^2R_S^2}h_{t+1}T - \frac{R_B^2}{PR_B^2 + P^2R_S^2}P\beta \\
    b_t &= \frac{h_tT - P\beta}{\frac{1}{N} + P} \\
    \text{with } N &= \left[ \frac{R_B + \sqrt{h_tT - P\beta R_S} \sqrt{1 + \left( \frac{R_B}{R_S} \right)^2} \frac{1}{P} }{PR_S} \right]^2.
\end{align*}
\]

See derivations in Appendix A.1.

There is an operation of the more-productive business when \( b_t > 0 \), while its absence is represented by \( b_t = 0 \). Equations (5) and (6) suggest that in order for firms to continue operating the more-productive business, i.e., maintain \( b_t > 0 \) and \( b_{t+1} > 0 \), their ability to process information must exceed \( \frac{P\beta}{T} \). Moreover, as indicated by equation (6), \( b_t \) increases with \( h_t \). This implies that a higher ability to process information in the first period allows firms to create more space to operate the more-productive business and generate higher profits, as they are more efficient in overcoming the additional time capacity requirement.

I plug the solution for \( b_t \) in (6) into the accumulation process (4) and generate that \( h_{t+1} \) is a function of \( h_t \) when firms operate the more-productive business, \( b_t > 0 \).

\[
h_{t+1} = \frac{(h_t - \frac{P\beta}{T})^2}{\frac{1}{N} + P} + \frac{P\beta}{T}.
\]

**Lemma 1** The evolution function \( h_{t+1} \) in (7) is increasing and convex in \( h_t \) when \( h_t > \frac{P\beta}{T} \) (i.e. \( \frac{\partial h_{t+1}}{\partial h_t} > 0 \) and \( \frac{\partial^2 h_{t+1}}{\partial h_t^2} > 0 \) if \( h_t > \frac{P\beta}{T} \)).

Proof: See Appendix A.2. □

The first part \( \frac{\partial h_{t+1}}{\partial h_t} > 0 \) suggests that higher ability in the first period \( t \) leads to an expansion of ability in the next period, resulting in an increase in time
capacity for both periods. This is based on equation (2). Furthermore, according to equation (6), an increase in $h_t$ is associated with an increase in $b_t$, implying that firms can enhance their more-productive business operations by improving their information processing ability.

The second part, $\frac{\partial^2 h_{t+1}}{\partial (h_t)^2} > 0$, indicates that $h_{t+1}$ grows in $h_t$ at an increasing rate. This can be explained by equation (4), where $h_{t+1} = \frac{c(h_t)h_t}{T} + \frac{P\beta T}{T}$. The interaction of the increased firms’ intensity in the more-productive business and their ability in the time $t$, would enhance their information processing ability in period $t+1$. This holds true as long as their information processing ability is sufficiently high, that is, $h_t > \frac{P\beta}{T}$.

**Lemma 2** The evolution function $h_{t+1}$ in (7) generates two different steady states: a stable one $h_t^* = \frac{P\beta}{T}$ and an unstable one $h_t^{**} > \frac{P\beta}{T}$.

Proof: See Appendix A.3. □

The steady state $h_t^* = \frac{P\beta}{T}$ is stable, and serves as an ability threshold that distinguishes a firm’s time between the more-productive and basic business. When $h_t$ is below $h_t^*$, the firm will allocate all of its time to business $s$, serving final consumers. When $h_t$ exceeds $h_t^*$, the firm allocates time between business $s$ and business $b$, thereby diversifying its operations and generating profits through the more-productive business.

The steady state $h_t^{**}$ is unstable and acts as the other ability threshold as a proficiency requirement in operating the more-productive business. It determines whether a firm can overcome self-stagnation. If a firm stands above $h_t^*$ but below $h_t^{**}$ for engaging in the more productive business, it cannot efficiently grow as the accumulation process falls below the 45-degree line. However, if $h_t > h_t^{**}$, the firm’s ability next period $h_{t+1}$ will always be higher than $h_t$, because the accumulation process is above the 45-degree line. It implies that the firm experiences growth in the more-productive business.

**Proposition 1** There is an economic stagnation system (a poverty trap where $h_{t+1} \leq h_t$) at a firm level due to limited time capacity when $h_t \leq h_t^{**}$, where (i) a firm with $h_t < h_t^*$ only serves final consumers ($s > 0, b = 0$) and stays in its initial ability level ($h_{t+1} = h_t$); (ii) a firm with $h_t^* < h_t \leq h_t^{**}$ operates the two types of businesses ($s > 0, b > 0$), and will slide towards $h_t^*$ ($h_{t+1} < h_t$).

Moreover, a firm with $h_t > h_t^{**}$ that operates in the two types of businesses ($s > 0, b > 0$) can experience a self-growth in the ability to process information ($h_{t+1} > h_t$).

Proof: Follows from Lemmas 1, 2 and the results from the second line of equation (4). □
Getting rid of the economic self-stagnation system requires firms to have enough intensity in their more-productive business to overcome the barrier, which means surpassing $h_t^{*}$. Once a firm’s ability to process information surpasses $h_t^{*}$, $h_{t+1}$ will always remain above the 45-degree line, resulting in growth in the firm’s ability to process information. I define this as the long-run effects. This desirable outcome presents an expansion in the firm’s time capacity (i.e., $hT$) due to a positive correlation between $h_t$ and $b_t$, indicating an increase in the intensity of the more-productive business operations.

5 Aid

When a firm falls into the self-stagnation system $h_t < h_t^{*}$, an exogenous intervention is required. This intervention, in the form of aid including the training program, aims to enhance the firm’s information processing ability $h_t$ in the baseline and expand the firm’s capacity. However, it is important to note that the effects of such aid will generate heterogeneous effects on the increased intensity of the more-productive business among firms. To comprehend these effects, I introduce a concept that the magnitude of the increase in $h_t$ resulting from the aid is represented by $d > 0$. The increased ability $h_t + d$ expands firm’s time capacity via equation (2) and (3), where the term on the right hand side of equation (2) and (3) grow. As a result, the increased ability affects firm’s intensity in the productive business by equation (6) and firm’s ability accumulation process (7) if $h_t + d$ is above $h_t^{*}$.

Corollary 1 (1) The short-run effects:

(1i) For a firm with a high ability ($h_t > h_t^{*}$), aid will generate a stronger increase in the intensity of the more-productive business ($b_t$) than a low ability firm in the short-run.
(1ii) For a firm with a low ability \((0 < h_t < h^*_t)\), the impact of aid will be the same as described in (1i) if \(h^*_t - d \leq h_t < h^*_t\) but it will have no effect on a firm if \(h_t < h^*_t - d\).

(2) Aid allows firms above \(h^*_t - d\) to experience a self-growth \(h_{t+1} > h_t\), which is summarized as long-run effects.

Proof: see Appendix A.4.

In the short run, higher-ability firms benefit more from the program in terms of growing intensity in the more-productive business \((b_t)\). This can be observed through the increasing and convex trajectory of \(b_t\) in relation to \(h_t\). In other words, the same amount of increase in ability resulting from aid will lead to a greater increase in the intensity of the more-productive business for firms with higher abilities. For a firm with higher abilities, the same increase in ability resulting from the aid program leads to a greater improvement in the intensity of their more-productive business operations. This greater improvement can be attributed to the positive interaction between a firm’s ability to process information \(c(h_t)\) and its bidding intensity for contracts \(b_t\), as indicated by the first line of equation (4).

For a firm whose ability falls within the range of \((h^*_t - d, h^*_t)\), aid can help the firm surpass the ability threshold of \(h^*_t\) to initiate the more-productive business temporarily. The extent of the improvement in intensity will be lesser compared to firms with higher abilities (above \(h^*_t\)).

Improving a firm’s ability to process information through aid may not always be sufficient to overcome the barriers to operating the more-productive business, especially for a firm with very low ability, e.g., \(h_t < h^*_t - d\). Aid cannot push them over the threshold of \(h^*_t\), to initiate the more-productive business.

The long-term impact of aid depends on whether a firm’s information processing ability can surpass the threshold of \(h^*_t\), which is the second ability threshold in this paper. Surpassing this threshold implies that \(h_{t+1}\) will exceed \(h_t\), above the 45-degree line. A firm can experience self-growth in the ability to process information, which allows a firm to further increase the intensity of the more-productive business. Firms that are closer to this threshold have a higher likelihood of further improving their information processing ability as a result of the aid intervention. Specifically, a firm in \((h^*_t - d, h^*_t)\) that is currently stuck with further growth can derive the long-run effects from the aid. This observation aligns with the findings in Hjort et al. (2020), where a firm that achieved significant outcomes was often more experienced in operating the more-productive business and using the internet.

However, for a firm that is further away from the second threshold (below \(h^*_t - d\)), the impact of the aid will be limited in the prospects for long-term growth. In other words, growth in the ability to process information requires a high intensity in the more-productive business \((b)\), which demands a substantial value of \(h\).
6 Conclusions

I have developed a model based on time capacity constraints to explain why small firms may struggle to achieve further growth in the more-productive business, where a firm’s time capacity plays a vital role. The model suggests that aid like a training program, which focuses on improving information processing ability and time capacity, tends to benefit firms with prior experience in operating the more-productive business and good internet access, both in the short and long term. The model identifies the presence of two crucial thresholds that firms need to surpass in order to attain lasting enhancements in the short and long term through the aid. Meeting these thresholds is essential for achieving sustainable improvements in firm performance.
References


75


Appendix A  Appendix

A.1 The proof for equation (5) and (6) - interior solutions

For an interior solution $b_t > 0$ and $b_{t+1} > 0$, we substitute all constraints into the profit function (1) that a firm faces. The optimization problem becomes

$$\text{Max}_{b_t, b_{t+1}} \Pi = R_S \sqrt{h_t T - P (b_t + \beta)} + R_B \sqrt{b_t}$$
$$+ R_S \sqrt{h_{t+1} (b_t) T - P (b_{t+1} + \beta)} + R_B \sqrt{b_{t+1}}.$$  \hspace{1cm} (A1)

The first-order condition with respect to $b_t$ yields

$$\frac{PR_S}{\sqrt{h_t T - P (b_t + \beta)}} = \frac{R_B}{\sqrt{b_t}} + \frac{R_S}{\sqrt{h_{t+1} (b_t) T - P (b_{t+1} + \beta)}} \frac{T \partial h_{t+1}}{\partial b_t}.$$  \hspace{1cm} (A2)

The first-order condition with respect to $b_{t+1}$ is

$$\frac{R_S P}{\sqrt{h_{t+1} (b_t) T - P (b_{t+1} + \beta)}} = \frac{R_B}{\sqrt{b_{t+1}}}.$$  \hspace{1cm} (A3)

By operation, (A3) generates

$$b_{t+1} = \frac{h_{t+1} T (R_B)^2 - \beta P (R_B)^2}{(R_S P)^2 + P (R_B)^2}.$$  \hspace{1cm} (A.3a)

We substitute (A.3a) into (A2) to generate

$$\frac{PR_S}{\sqrt{h_t T - P (b_t + \beta)}} = \frac{R_B}{\sqrt{b_t}} + \sqrt{\frac{(R_S P)^2 + P (R_B)^2}{(R_S P)^2}} \frac{R_S}{\sqrt{h_{t+1} T - P \beta}} \frac{T \partial h_{t+1}}{\partial b_t}.$$  \hspace{1cm} (A.2a)

Then we apply the ability accumulation function (4) to (A.2a) to get

$$\frac{PR_S}{\sqrt{h_t T - P (b_t + \beta)}} = \frac{R_B}{\sqrt{b_t}} + \sqrt{\frac{(R_S P)^2 + P (R_B)^2}{(R_S P)^2}} \frac{R_S}{\sqrt{b_t}} \frac{\sqrt{c(h_t)}}{\sqrt{b_t}}.$$  \hspace{1cm} (A.2b)

The operation of (A.2b) generates the result in equation (5). To verify the argument we got to be optimal, we check the second-order condition. The Hessian matrix will be

$$H = \begin{bmatrix}
\frac{\partial^2 \Pi}{\partial b_t^2} & \frac{\partial^2 \Pi}{\partial h_t \partial b_{t+1}} \\
\frac{\partial^2 \Pi}{\partial h_t \partial b_t} & \frac{\partial^2 \Pi}{\partial b_{t+1}^2}
\end{bmatrix}$$

since we have variables $b_t$ and $b_{t+1}$ in this optimization. The first-order minors are

$$\frac{\partial^2 \Pi}{\partial b_t^2} = -\frac{P^2 R_S}{4} (h_t T - P (h_t + \beta))^{-\frac{3}{2}} - \frac{R_B}{4} b_t^{-\frac{3}{2}} - \frac{R_S}{4} c^2 (h_t) (c (h_t) b_t - P b_{t+1})^{-\frac{3}{2}} < 0,$$
\[
\frac{\partial^2 \Pi}{\partial b_{t+1}^2} = - \frac{R_S P^2}{4} (c(h_t) b_t - P b_{t+1})^{-\frac{3}{2}} - \frac{R_B}{4} b_{t+1}^{-\frac{3}{2}} < 0.
\]

The second-order minor is
\[
\frac{\partial^2 \Pi}{\partial b_t^2} \frac{\partial^2 \Pi}{\partial b_{t+1}^2} - \left[ \frac{\partial^2 \Pi}{\partial b_t \partial b_{t+1}} \right]^2 = \frac{1}{16} P^4 R_S^2 (h_t T - P (h_t + \beta))^{-\frac{3}{2}} (c(h_t) b_t - P b_{t+1})^{-\frac{3}{2}}
+ \frac{1}{16} P^2 R_S R_B b_{t+1}^{-\frac{3}{2}} (h_t T - P (h_t + \beta))^{-\frac{3}{2}} + \frac{1}{16} R_B R_S c^2 (h_t) b_{t+1}^{-\frac{3}{2}} (c(h_t) b_t - P b_{t+1})^{-\frac{3}{2}}
+ \frac{1}{16} R_B b_t^{-\frac{3}{2}} R_S P^2 (c(h_t) b_t - P b_{t+1})^{-\frac{3}{2}} + \frac{1}{16} R_B b_t^{-\frac{3}{2}} b_{t+1}^{-\frac{3}{2}} > 0.
\]

where
\[
\frac{\partial^2 \Pi}{\partial b_{t+1} \partial b_t} = \frac{1}{4} R_S P c (h_t) (c(h_t) b_t - P b_{t+1})^{-\frac{3}{2}} > 0.
\]

Therefore, we can confirm
\[
(-1)^n |H_n| > 0
\]
indicating that the Hessian matrix is negative definite and the solution we derived maximizes firms’ profits. \(\square\)

### A.2 The proof for Lemma 1

We first show \(\frac{\partial h_{t+1}}{\partial h_t} > 0\). With
\[
h_{t+1} = \frac{(h_t - \frac{P \beta}{T})^2}{\frac{1}{N} + P} + \frac{P \beta}{T},
\]
we have
\[
\frac{\partial h_{t+1}}{\partial h_t} = \frac{2 \left( h_t - \frac{P \beta}{T} \right)}{\frac{1}{N} + P} + \frac{\left( h_t - \frac{P \beta}{T} \right)^2}{(1 + N P)^2} \frac{\partial N}{\partial h_t}.
\]

With \(N = \left[ \frac{R_B + \sqrt{c R_S}}{P R_S} \right]^{\frac{1}{2}}\) and \(c = h_t - \frac{P \beta}{T}\), we have
\[
\frac{\partial N}{\partial h_t} = \left[ \frac{R_B}{P R_S} \sqrt{h_t - \frac{P \beta}{T}} + \sqrt{1 + \left( \frac{R_B}{R_S} \right)^2 \frac{1}{P}} \right] \sqrt{1 + \left( \frac{R_B}{R_S} \right)^2 \frac{1}{P}} \hspace{1cm} (A5)
\]
and then
\[
\frac{\partial h_{t+1}}{\partial h_t} = \frac{2 \left( h_t - \frac{P\beta}{T} \right)}{N + P} + \left( h_t - \frac{P\beta}{T} \right)^2 \frac{R_B}{(1 + NP)^2} \left[ \frac{P}{R_S} \frac{\sqrt{h_t - \frac{P\beta}{T}}}{P} \right] \sqrt{1 + \left( \frac{R_B}{R_S} \right)^2 \frac{P}{T}}. 
\] (A6)

When \( h_t > \frac{P\beta}{T} \), we have \( \frac{\partial N}{\partial h_t} > 0 \), then \( \frac{\partial h_{t+1}}{\partial h_t} > 0 \).

We could also get
\[
\lim_{h_t \to \frac{P\beta}{T}} \frac{\partial h_{t+1}}{\partial h_t} = 0, 
\] (A7)
because the second term by L’Hopital rule is zero.

We now turn to show \( \frac{\partial^2 h_{t+1}}{\partial (h_t)^2} > 0 \).

\[
\frac{\partial^2 h_{t+1}}{\partial (h_t)^2} = 
\frac{-(L-h_t)^2Q}{2} \left( \frac{Q}{L-h_t} + \frac{R_B + Q\sqrt{h_t-L}}{(h_t-L)^{3/2}} \right) + 4Q (h_t - L) \frac{R_B + Q\sqrt{h_t-L}}{\sqrt{h_t-L}} + 2 \left( Q + Q\sqrt{h_t-L} \right)^2 \frac{d^2}{d^2}
\]

where we use several notations to simplify the expression
\[
d = PR_S > 0, \\
Q = R_S \sqrt{1 + \left( \frac{R_B}{R_S} \right)^2 \frac{P}{T}} > 0, \\
L = P\beta / T > 0.
\]

The dominator \( d^2 \) can be skipped when we decide the sign of \( \frac{\partial^2 h_{t+1}}{\partial (h_t)^2} \). The numerator by operation, reads
\[
\frac{7Q^2R_B}{2} \sqrt{h_t-L} + 4Q^2 (h_t - L) + 2 \left( R_B + Q\sqrt{h_t-L} \right)^2 > 0
\]
for \( h_t > L = P\beta / T \). Therefore \( \frac{\partial^2 h_{t+1}}{\partial (h_t)^2} > 0 \).

A.3 The proof for Lemma 2
We connect the evolution function (7) with the 45-degree line to get
\[
h_t = \left( h_t - \frac{P\beta}{T} \right)^2 \frac{1}{N + P} + \frac{P\beta}{T}, 
\] (A8)
which can be written as

\[
\left( h_t - \frac{P\beta}{T} \right) \left[ 1 - \frac{h_t - \frac{P\beta}{T}}{\frac{1}{N} + P} \right] = 0.
\]

The first solution is \( h_t = \frac{P\beta}{T} \).

In order to solve for a solution that is not equal to \( \frac{P\beta}{T} \), the equation can be rewritten as

\[
\frac{1}{N} + P = h_t - \frac{P\beta}{T}.
\]

With \( N = \left[ \frac{R_B + \sqrt{cR_S}}{P_{RS}} \sqrt{1 + \left( \frac{R_B}{R_S} \right)^2 \frac{1}{P}} \right]^2 \) and \( c = h_t - \frac{P\beta}{T} \), we have

\[
\left( PR_S \right)^2 + \frac{P = h_t - \frac{P\beta}{T}}{R_B + R_S \sqrt{x} \sqrt{1 + \left( \frac{R_B}{R_S} \right)^2 \frac{1}{P}}}.
\]

Define \( x = \left( h_t - \frac{P\beta}{T} \right) > 0 \), we can update the expression to

\[
\left( PR_S \right)^2 \frac{P = x}{R_B + R_S \sqrt{x} \sqrt{1 + \left( \frac{R_B}{R_S} \right)^2 \frac{1}{P}}}.
\]

Define

\[
g(x) = \frac{(PR_S)^2}{R_B + R_S \sqrt{x} \sqrt{1 + \left( \frac{R_B}{R_S} \right)^2 \frac{1}{P}}} \]

which is the LHS. Define \( l(x) = x \), and this is the RHS. We find that

\[
g'(x) = - (PR_S)^2 \left[ R_B + R_S \sqrt{x} \sqrt{1 + \left( \frac{R_B}{R_S} \right)^2 \frac{1}{P}} \right]^{-3} R_S \left[ \sqrt{1 + \left( \frac{R_B}{R_S} \right)^2 \frac{1}{P}} \right] x^{-\frac{1}{2}} < 0.
\]

For \( x \geq 0 \), \( g(x) \) is continuous and monotonically decreasing and \( l(x) \) is continuous and monotonically increasing. Also, since \( g(0) = \left( \frac{PR_S}{R_B} \right)^2 > 0 \), \( l(0) = 0 \), \( \lim_{x \to \infty} g(x) = 0 \), and \( \lim_{x \to \infty} l(x) = \infty \). Hence, there exists an intersection when \( x > 0 \) and that is \( h > \frac{P\beta}{T} \). See the figure.
In combination with Lemma 1 that features the trajectory of $h_{t+1}$, we can confirm that $h^*_t$ is stable while $h^*_t$ is not.

\[\square\]

A.4 The proof for Corollary 1

A.4.1 The short-run effect

For firms with $h_t > \frac{P\beta}{T}$, we would like to prove $\frac{\partial b_t}{\partial h_t} > 0$ and $\frac{\partial^2 b_t}{\partial (h_t)^2} > 0$. The first-order derivative of $b_t$ with respect to $h_t$ is

\[
\frac{\partial b_t}{\partial h_t} = \frac{T}{N + P} + \frac{(h_T - P\beta)}{(1 + NP)^2} \frac{\partial N}{\partial h_t} > 0,
\]

where $\frac{\partial N}{\partial h_t} > 0$ and $h_t > \frac{P\beta}{T}$.

The second-order derivative of $b_t$ with respect to $h_t$ is

\[
\frac{\partial^2 b_t}{\partial h_t^2} = \frac{T b (3a b^2 x + (6a^2 + 4P) b \sqrt{x} + 3a^3 + 3Pa)}{2\sqrt{x} (b^2 x + 2ab \sqrt{x} + a^2 + P)^3} > 0,
\]

where we use several notations to simplify the expression

\[
x = h_t - \frac{P\beta}{T} > 0
\]

\[
a = \frac{R_B}{R_s} > 1
\]

\[
b = \frac{\sqrt{1 + \left(\frac{R_B}{K_c}\right)^2}}{\sqrt{T}} \frac{1}{P} > 0.
\]

Therefore, $b_t$ grows in $h_t$ at an increasing rate. Given the same amount of push in $h_t$, a more substantial effect in terms of an instantaneous improvement happens to the firms with stronger abilities.
For firms $h_t < h_t^*$, there will be no increase in $b_t$ if they are still below $h_t^*$ after aid. If they pass the threshold after aid, the results for $h_t > h_t^*$ apply to them in the increment of $b_t$.

A.4.2 The long-run effect

The long-run effects depend on whether firms can realize a self-growth such that $h_{t+1} > h_t$. Therefore it hinges on whether a firm stands above $h_t^{u*}$. Given the assumption that aid increases $h_t$ by $d$, the long-run effects due to aid do not take place to firms below $h_t^{u*} - d$. 
Essay III. Optimal Aid Distribution and Poverty Traps

Acknowledgements: I am deeply grateful for the generous assistance and insightful comments provided by Daniel Spiro, Thomas Aronsson, Chuanzhong Li, Andreas Kotsadam, Fredrik Armerin, and Hongning Huang, Erlend Magnusen Fleisje, Sergei Jegorov as well as all participants in UCFS and others who have offered feedback on this paper. I extend special thanks to Yikai Wang and Per Engström for their invaluable support and guidance throughout this research. Gratitude for the financial support provided by the project “The Economics of Planetary Boundaries”, FORMAS grant 2020-00371, hosted by Daniel Spiro.
1 Introduction

A poverty trap is characterized by a self-reinforcing mechanism such that initial poverty implies poverty in the future, which exacerbates the existing inequality and underutilization of talents for individuals. The existence of a poverty trap is supported by the empirical evidence of Balboni et al. (2022), where poverty is generated by a lack of a better opportunity instead of traits that are required to achieve higher economic well-being.\(^1\) Based on global evidence by United Nations Development Programme (UNDP) (2023), the sum of poor individuals measured by multiple dimensional poverty index across 110 developing countries constitutes 1.1 billion out of 6.1 billion. However, the number of poor people may very well be underestimated due to previous shocks from the COVID-19 pandemic, the war between Russia and Ukraine, and the global climate crisis, indicating the importance of poverty alleviation.

“A big push”, meaning that a sufficient amount of money is allocated to recipients inside a poverty trap to solve poverty once and for all, has been verified with substantial effects by a strand of literature (Murphy et al., 1989; Banerjee et al., 2021; Balboni et al., 2022). It was formulated in response to poverty recurrence by current aid policies. Most current programs do not allocate a sufficient amount to ultra-poor people, who are relatively worse off economically, preventing them from escaping poverty permanently despite momentary improvements.\(^2\) Aizer et al. (2016) show that a fair distribution of money transfers will facilitate poor people’s lives directly through more consumption, but it does not affect their indigent lives in the long term. However, there are still three parts that “Big pushes” miss. Firstly, available funds for poverty alleviation are limited. The big push to all ultra-poor people demands unrealistically large amounts of funds to be practically feasible. The relevant policy question is how to maximize the marginal social welfare gain of the last dollar by designing prioritization and optimal quantity conditional on limited funds. Secondly, the maximization of social welfare in the presence of poverty traps and insufficient funds is not at odds with leaving some receivers in poverty, because the value of eradicating poverty might not be relatively low compared to the costs. Thirdly, the ultra-poor individuals are targeted by the “Big pushes” in many cases, but other poor people who are slightly better off, especially implicitly poor individuals, are often ignored.

The implicit poor in this research, from a theoretical perspective, refers to individuals who are closer to the threshold of surpassing a poverty trap. They

---

\(^1\) A poverty trap can be caused by a number of factors such as poor education, lack of job training and skills, limited access to financial resources, and systemic inequalities. In this study, I mainly focus on the indivisibility of inputs that generates productivity differences.

\(^2\) There are multiple programs for saving the extremely poor to fight poverty, malnutrition, and inequality, such as Targeting the Ultra-Poor (TUP) Program in Bangladesh (BRAC), The Bolsa Familia in Brazil and Progressa in Mexico, Bantuan Siswa Miskin in Indonesia, Giving Money to Very Poor People Will Make Their Lives Better — Just Ask Ecuador, Targeted Poverty Alleviation in China that declares victory in ending extreme poverty.
have more initial endowment compared to the ultra-poor that situate far away from the threshold. Haushofer et al. (2022) discuss aid for these individuals to be more cost-efficient, as less redistribution is needed to surpass the poverty threshold. Barrett et al. (2008) mention that targeting the achievement of critical asset thresholds with a gentle push could have substantial long-term effects on poverty alleviation. In economic reality, such poor groups are not very noticeable in terms of poverty by the public but are at a high risk of eventually becoming ultra-poor people. It will require much more than they would have otherwise needed to solve poverty if they slide to the bottom. The term “Working Poor” can be used to describe implicitly poor individuals. For example, the Chinese born in the 1990s, on average, have approximately 17,433 US dollars in debt, of which consumption loans dominate. Quite a few of them suffer chronic poverty.

Mainly, the motivation for this paper is to consider the entire poor population besides the ultra-poor as potential aid recipients for social welfare maximization, when the restriction on funds appears in a poverty trap model. To be clear, I focus less on motivating why there is a poverty trap since there are many different mechanisms, but more on a parsimonious model to obtain analytical results for welfare comparison. To this end, I establish a simplified poverty trap model with a threshold externality in a Solow-like model inspired by Ghatak (2015), where all individuals are identical except for their initial endowment - the income at the beginning- since individual characteristics that might lead to poverty such as aspiration, diligence, intelligence, etc., are abstracted away. The modified Solow model is closer to reality in terms of

---

3 The research on individuals who are implicitly poor is limited, and only a few exceptions exist, such as a concept of “relatively poor” in Decerf (2017). In contrast to absolutely poor individuals who struggle to meet basic needs, relatively poor individuals have insufficient income to cover the cost of social life, such as attending social events. While Decerf’s aid policy design differentiates between the two types of poverty, there is room for further extension of the research framework by considering income threshold effects and poverty transmission.

4 According to a recent HSBC survey conducted in 2019, the debt-to-income ratio of China’s post-90s generation has reached a staggering 1,850 percent, with the average debt owed to lending and credit-issuing institutions exceeding 120,000 CNY (17,433 US dollars). In 2019, a study of Chinese white-collar workers found that 34.6% of respondents were in debt without any balance, with more than 50% of those respondents having a balance below 10,000 CNY (approximately 1,500 US dollars), often due to low-paying jobs. New Chinese university graduates frequently fall into this group.

Meanwhile, in industrialized countries, student debt is a major concern. A survey conducted in June 2020 found that the average household with student debt owed about 48,000$, and 60% of borrowers reported feeling like they were living paycheck to paycheck. T.D. Bank estimates that student debt holders spend 20% of their take-home pay on loans. The impact of student debt can be significant, with 21% of borrowers delaying getting married, 26% delaying having children, and 36% delaying buying a home. A financial planning firm, Stash Wealth, reported that their visitors have an average of 80,000$ in student debt.

5 Poverty traps can also be driven by behavioral biases e.g., a failure in aspiration (Beaman et al., 2012), non-homothetic preference over temptation goods (Banerjee and Mullainathan, 2007)
poor individual behaviors despite lacking an individual optimization process. Firstly, it is hard for a policymaker to observe the exact consumption and investment behavior of individuals at a micro level, but the income or debt in each period is relatively easier to observe. Secondly, individuals can not alter their worsening economic situations despite optimal investment and consumption decisions if they started in poverty.

The threshold externality in this model means that productivity, which is an exogenous technology term in a production function, jumps substantially once an input reaches a certain level, as in Azariadis (1996). Such indivisible technology would capture the potential gains of getting rid of stagnation and the cost of degrading economic conditions because there is no borrowing behavior in the Solow model to dampen the existence of a poverty trap. The input threshold, which demarcates differentiated productivity, assigns two stable steady states: a desired one - a wealthy state, and an undesired one - extreme poverty. The interval between the undesired steady state and the threshold can be characterized as a poverty trap in this research. Without any intervention, the people inside the poverty trap will converge to extreme poverty, and the individuals above the threshold will converge to the wealthy state. This creates a possibility of solving poverty by relocating the position where poor ones are initially situated. The most straightforward policy is a one-off transfer, which could be understood as an external asset infusion to a company in a financial dilemma or direct support for students’ maintenance.

Inside the poverty trap, the funds for one-off transfers are divisible between a top-down and a bottom-up policy. A top-down policy yields permanent poverty alleviation for its receivers, where the prioritization of receiving assistance positively correlates to initial endowment, so the implicitly poor are favored. The amount is exactly enough for recipients to jump out of the poverty trap, as exiting the trap with a margin turns out inefficient, which is reminiscent of “A big push” if recipients are in great financial deprivation. A bottom-up policy indicates an opposite prioritization and thus favors ultra-poor individuals. The recipients share the funds to reach the same level for a mo-

and excessive conspicuous consumption leading to low savings (Moav and Neeman, 2010) or human capital (Omer and Neeman, 2012).

A friction-driven poverty trap usually occurs when non-convex technology, such as indivisible inputs, and a borrowing constraint prevent individuals from escaping poverty on their own. In models such as Ramsay or DSGE, which separate consumption and investment behaviors, a borrowing constraint is necessary. Without such a constraint, individuals could overcome non-convexity technology on their own by borrowing money.

There are various policies available for alleviating poverty when a poverty trap exists, including lowering the barrier to a non-poor state and transforming the system from an ecological and sociological perspective, as proposed by Lade et al. (2017). However, these methods are not considered in this analysis due to their demanding nature and the assumption of an exogenously existing poverty trap.

The help to these implicitly poor can be understood as a complement to those who are working poor, instead of a substitute i.e., cancellation of the subsidy once they are employed.
mentary improvement. Essentially, when the social planner values each individual’s utility equally, the social welfare maximization problem becomes the optimal allocation of funds to trade off long-term poverty eradication through productivity improvements and temporary amelioration driven by scarcity effects.

At optimum, given an exogenous amount of funds, the welfare effect for the last dollar should be maximized, echoing the reality that one-off transfers are irreversible. The combination of available funds and the current value of productivity improvement from lifting poverty determines the optimal aid design. I, therefore, characterize the conditions for (i) an optimal pure policy where the top-down policy gets all available funds and (ii) an optimal policy mix where two policies share all available funds. This suggests that the top-down policies consistently receive priority and the amount of funds prioritized to ensure this will be contingent upon productivity improvement.

In addition, the method of optimal assistance design exhibits a minor connection to the distribution of initial income. I use a uniform distribution over individual initial income to derive an analytical solution. But the welfare analysis can apply to a different distribution to form a specific aid distribution, because it roots in hypothetical initial endowment. A different probability density function might change the expense and the marginal welfare effects of funding bottom-up and top-down policies.

**Related literature and contributions.** One-off transfers are expected to have persistent effects if the amount is sufficient for individuals to jump over a poverty trap. The predicted effects are verified in cash transfers (Haushofer and Shapiro, 2018) despite the mixed evidence of cash transfers from emerging literature, and asset transfers (Blattman et al., 2014; Banerjee et al., 2015; Bandiera et al., 2017; Blattman et al., 2020; Banerjee et al., 2021; Balboni et al., 2022). Balboni et al. (2022) state that a one-off transfer of productive assets can push the extremely poor out of a poverty trap if this assistance is large enough. The benefits of skipping poverty traps for impoverished individuals are estimated to outweigh the cost of one-off transfers around fifteen times. Banerjee et al. (2021) verified that a one-time transfer had a sustaining effect. The one-off financial assistance made ultra-poor people in the program improve their income level and living conditions. The positive impact was still visible more than seven years after the program had ended. The benefited groups became more capable of taking advantage of labor market opportunities elsewhere and therefore migrated further to earn higher labor income.

The design of aid distribution in this research includes what Balboni et al. (2022), Banerjee et al. (2021), and other research mentioned above conclude on the substantial effects of one-off transfers. It also extends potential recipients to all poor people in a poverty trap, where the implicitly poor are considered in line with Haushofer et al. (2022) and Barrett et al. (2008). I also model how the social welfare effects of alleviating poverty and the cost form optimal
assistance strategies.\textsuperscript{9} The results sort out prioritization and optimal amounts and characterize a possible optimal policy mix of top-down and bottom-up.

2 Model Setup
Consider an individual $j$ who lives one period and cares about descendants for all periods, indicating Barro-Becker Altruistic preferences.\textsuperscript{10} Individual’s utility relies on the present value of private consumption across all generations

$$U^j = \sum_{t=0}^{\infty} \beta^t \ln c^j_t$$  \hspace{1cm} (1)

where $\beta \in (0, 1)$ is a discount factor and $c$ is a general consumption representing expenditure in all possible aspects, such as necessities, education, medicine and other services. The instantaneous utility is logarithmic such that consumers’ welfare is continuously increasing and concave in their consumption. The financial budget constraint without access to intertemporal borrowing at each period is

$$c^j_t \leq y^j_t \text{ for } t = 0, 1, \ldots, T,$$  \hspace{1cm} (2)

meaning that the consumption does not exceed the income $y$ in each period. The tight bundle implies that all income will be used entirely. The initial income $y^j_0$ is exogenously given and is defined as initial endowment. The endowment for one individual can be multidimensional, which could be a combination of human capital, physical capital, cash and savings, financial assets like bonds, stocks, and insurance. Nevertheless, I abstract and reduce the multidimensional entity to a scalar $y$ for simplicity. A higher $y$, therefore, indicates a wealthier individual. The income trajectory is assumed to be expressed by the Solow-like model with differentiated technologies

$$y_{t+1} = P^\alpha y_t = \begin{cases} Ay_t^\alpha, & \text{if } y_t \geq \bar{y} \\ By_t^\alpha, & \text{otherwise} \end{cases} \text{ for } A > B \text{ and } t \geq 0$$  \hspace{1cm} (3)

where $\alpha \in (0, 1)$ represents intertemporal elasticity, describing the sensitivity of how current inputs change the income next period.\textsuperscript{11} Future income is increasing and (locally) concave with respect to current income, depending on

\textsuperscript{9}In the analysis of the current value of productivity improvement by lifting poverty, a discount factor is specifically considered in a poverty trap affects policy design, since it matters for the marginal welfare effects that a top-down policy can reach. The importance of patience is also discussed in Sunde et al. (2021) as relevant for the growth of capital, assets, etc. at both macro and micro levels.

\textsuperscript{10}This can also be interpreted as a representative individual lives infinite periods and maximizes the utilities of his own life span.

\textsuperscript{11}In the Solow-like model used in this study, there are no saving behaviors considered to reflect the fact that poor individuals have to spend all of their income to meet their basic needs. Therefore, the saving rate in the model is a constant of zero.
individual trajectories. The Inada condition holds, which means that when income is low, the marginal effect of income (as an input) is high. Productivity $P$ is lower ($B < A$) if the income is below the threshold $\bar{y}$,\footnote{Ghatak (2015) lists several external frictions that can cause non-convergence in the Solow model due to non-convexity in the production technology. For example, (i) capital market imperfections can lead to non-convexity of the production function: since subsistence activities require no capital, poor people cannot borrow money at a fair price in an imperfect capital market. (ii) Frictions in human capital can also create non-convexity when the market for human capital is not perfect. Dasgupta and Ray (1986) focus on the nutrition-based efficiency wage, which states that people cannot survive or be productive if their income or consumption falls below a certain level.} indicating an indivisible technology of inputs. Compared to an S-shaped technology poverty trap, the jumping-up point from indivisible technology captures an instant improvement in productivity. For example, cows, as expensive productive assets compared to poultry, are more productive in agricultural yields. Wages can experience substantial growth after migrating to a new system (Clemens et al., 2008), although it is hard to achieve.

![Figure 1. A depiction of the poverty trap](image)

In Figure 1, the differentiated productivity by an income threshold $\bar{y}$ yields different accumulation pathways with two steady states. One is a suboptimal steady state $y^*_{l}$ which represents extreme poverty, and the other is an optimal steady state $y^*_h$ which represents desired wealth levels. The interval between the suboptimal steady state and the income threshold $[y^*_{l}, \bar{y}]$ refers to the poverty trap. Without any intervention, the individual staying in the poverty trap will converge to $y^*_{l}$, and the one whose income exceeds the threshold $\bar{y}$
will converge to the desired equilibrium $y^*_h$. More specifically, if $y < \bar{y}$, the income next period becomes lower than the current income, as the productivity $B$ is not high enough to maintain a net increase. The individuals will stay in self-sustaining poverty.

The poverty trap in terms of endowed income can be alternatively written as $\rho \bar{y}$ by normalization where $\rho \in [\rho, 1)$ for easier analysis, because the relative distance to the threshold for poverty alleviation is the focus. Inside the poverty trap, the recipients can be divided into those who are closer to $\bar{y}$ e.g., the implicitly poor, and those who are closer to $\rho$, e.g. the ultra-poor, where $0 < \rho \leq 1$. The lower bound $\rho > 0$ represents positive despite small initial endowment. The former group, therefore, requires fewer resources to escape poverty, while the latter group experiences stronger scarcity-driven welfare effects for each unit of aid they receive. Similarly, individuals whose initial income exceeds the threshold $\bar{y}$ can be represented by $\rho^*\bar{y}$, where $\rho^* \geq 1$.

Given the convergence theory, which states it takes an infinite time to reach a steady state in this system, the terminal period $T$ is set to be positive infinite. The utility for a household $j$ over all periods is expressed by

$$U^j = \frac{1}{1-a\beta} \left[ \ln(\rho^j\bar{y}) + \frac{\beta}{1-\beta}\ln P \right]$$

for $P = \begin{cases} A & \rho^j \geq 1 \\ B & \rho \leq \rho^j < 1 \end{cases}$. (4)

The derivation can be seen in Appendix A.1.

$U^j$ depends on the net discount factor $1 - a\beta$ which is formed by output elasticity with respect to inputs in the production function $\alpha$ and discount factor $\alpha$, the initial income level $\rho^j$, the productivity $P$ in the convergence process. This expression shows how one-off transfers possibly affect individual utility. The assistance infuses an exogenous increase to a recipient’s initial income and possibly alters the technology trajectory if the increased income is sufficient to wedge the gap to the income threshold.

3 Policy Design

3.1 The social planner

The social planner is characterized as a utilitarian and aims to maximize the overall social welfare across all individuals who care about utility over all periods. Each individual is equally important in distribution $G(\rho)$, with an equalized weight across all individuals in the following social welfare function

$$W = \int U(\rho) dG(\rho).$$

The recipients below $y^*_l$ are not included in the analysis by the assumption that there is no population whose endowment is zero, meaning that the ultra-poor individuals still have some resources, albeit limited. In this line, Briggs (2021) shows that assistance usually does not reach the remotest areas for the poorest people due to impoverished infrastructure, etc.
Suppose that a one-off transfer is the only policy instrument at hand, where the funds \( F \) are exogenously given, e.g., international aid, charity donation, and balance from last year. The social planner has all the endowment information and can decide who will get the funds.\(^{14}\)

3.2 Aid policy inside the poverty trap

I formulate Lemma 1 to explain why the focus should be given to the individuals inside the poverty trap.

**Lemma 1** Individuals above the poverty trap are ineligible for aid when the available funds are insufficient to elevate everyone out of the poverty trap \( F < F^{all} \).

Proof: See Appendix A.2. \( \square \)

\( F^{all} \) denotes the total funds needed by a social planner to extricate everyone from the poverty trap. The analysis is conducted under the premise that there are insufficient funds to eradicate poverty for all. Given the constraint of limited financial resources, the planner strives for the most efficient allocation of these funds. By expression (4), there are two channels for a single aid receiver through which a one-off transfer can work: the income part and the possible change in productivity if the one-off transfer is sufficient to fill the income gap to lift poverty. Suppose a marginal fund is insufficient to help anyone surpass the poverty trap except the one who is infinitely close to the threshold. Therefore, irrespective of receiving assistance, the group inside the poverty trap and the group above the poverty trap stay in their productivity trajectories, respectively. Given higher initial income and productivity, the absolute utility level for the individuals above the threshold is strictly higher. It implies a strictly lower marginal utility of receiving an equivalent amount of funds. Lemma 1 also implies that any assistance beyond the threshold is inefficient, given a lack of funds to raise everyone out of the poverty trap. The amount of aid should be limited to what is required for reaching the poverty threshold \( \bar{y} \).

The main problem reduces to the internal allocation among poor individuals. Consider a large number of poor individuals uniformly distributed over a normalized income interval of the poverty trap \( \rho \in [\rho, 1) \), by normalizing the population to unity, the probability density function is

\[
f(\rho) = \frac{1}{1 - \rho},\quad (6)
\]

\(^{14}\)The funds are assumed to be completely transferred to recipients without delivery and implementation issues. It potentially can be overcome by information from the local community and help workers, etc.

\(^{15}\)The uniform distribution appears as a strong assumption, as it might require some shocks to disturb the bimodal income distribution at the equilibrium. However, the main focus is on
I take recipients as one group (called One) that is closer to the income threshold \( \bar{y} \), including the implicitly poor, and another group (called Two) that is further below in income distribution, including the ultra-poor.

### 3.2.1 A top-down policy

A top-down policy, which is called Scenario One for responding to the receivers from Group One, aims to aid \( \rho \in [x_1, 1) \) to get out of the poverty trap completely for long-run welfare growth, leading to productivity improvement. \( x_1 \) is the lower bound to be covered by the top-down policy. This aid design favors less disadvantaged individuals i.e. implicitly poor individuals.

A top-down policy is described by its budget constraint and welfare effects. The budget constraint

\[
\int_{x_1}^{1} (1 - \rho) f(\rho) \, d\rho = F_1
\]

means that cost of raising all in \([x_1, 1)\) should not exceed the available funds \( F_1 \), where \( f(\rho) \) is the density function from a uniform distribution (6). It is also responsive to the implications from Lemma 1 that no margins are needed to surpass the threshold. The welfare effect given an exogenous amount of funds \( F_1 \) is

\[
W_1(F_1) = \frac{1-x_1}{1-\rho} U_j(\rho = 1). 
\]

Expression (8) refers to that \( \frac{1-x_1}{1-\rho} \) individuals in the distribution are experiencing the utility of lifting poverty permanently \( U^j(\rho = 1) \). By substituting the utility function (4) that individual \( j \) can derive over all periods and solving the budget constraint, Scenario One is explicitly described by

\[
W_1(F_1) = \frac{1-x_1}{1-\rho} \left[ \ln \bar{y} + \frac{\beta}{1-\beta} \ln A \right] \\
\text{s.t. } \frac{(1-x_1)^2}{2(1-\rho)} = F_1.
\]

### 3.2.2 A bottom-up policy

A bottom-up policy, which is called Scenario Two, targets \([\rho, x_2]\) for a momentary improvement since they will converge to \( y^* \) in any case, where \( x_2 < 1 \)
is the upper bound for the bottom-up policy. Each receiver below $x_2$ in a bottom-up policy is expected to reach the same income level $x_2$ after the aid. This upper bound could be interpreted as the income limit to receive assistance for an urgent situation. This policy favors the individuals who are sitting at the bottom of the income distribution.

The sum of total costs in the bottom-up policy should not exceed the available funds

$$\int_{\rho}^{x_2} (x_2 - \rho) f(\rho) d\rho = F_2.$$  \hfill (10)

The welfare effects are the number of individuals who experience the utility at the income level $x_2$

$$W_2(F_2) = \frac{x_2 - \rho}{1 - \rho} U_j\left(\rho_j = x_2\right).$$  \hfill (11)

Solving the budget constraint allows to describe Scenario Two as

$$W_2(F_2) = \frac{x_2 - \rho}{1 - \rho} \frac{1}{1 - a\beta} \left[ \ln(x_2) + \frac{\beta}{1 - \beta} \ln B \right]$$

s.t. $\frac{(x_2 - \rho)^2}{2(1 - \rho)} = F_2$.  \hfill (12)

**Lemma 2** Bottom-up and top-down policies are the only potential best instruments.

Proof: See Appendix A.3. \hfill \Box

The term ’best instruments’ refers to the most efficient methods of financial expenditure. I will explore expenditure policy design without altering the nature or characteristics of the poverty trap itself. When there are limited resources, such as a small budget, deciding how to prioritize the spending can be challenging. A top-down policy aims to lift individuals out of poverty starting from the wealthiest ones, while a bottom-up policy focuses on providing temporary relief to the poorest individuals.

Under a bottom-up policy, providing assistance to those significantly below the poverty trap line can have a greater impact on overall welfare than helping those just slightly below it, due to the limited effects of initial income. Prioritizing assistance in this way can have a global impact, making it the best policy when recipients do not surpass the poverty trap. A top-down policy can increase income and productivity for its recipients, but it becomes increasingly expensive for those further below the poverty line. The impact of this prioritization is locally continuous if future productivity gains are modest; otherwise, it is globally continuous, meaning that aiding someone higher up in the income distribution to get out of the poverty trap is always more beneficial.
3.2.3 The group without any aid
The possible space between \( x_2 \) and \( x_1 \) for \( x_1 \neq x_2 \) is for individuals who are not targeted by any aid scenario due to limited funds (Group Three). Such a group exists because its marginal utility of getting the aid is lower than either the high-income group or the low-income group since they need more than Group One to lift poverty and have fewer scarcity effects than Group Two, as stated in Lemma 2: no policy design starts with anyone in between. If total funds are abundant, the third group is included implicitly as \( x_2 \) and \( x_1 \) will coincide. Otherwise, the third group in \( [x_2,x_1] \) that got no aid populates with a size \( \frac{x_1-x_2}{1-\rho} \). The total welfare for the third group is

\[
W_3 = \int_{x_2}^{x_1} U(\rho)f(\rho)d\rho = \frac{1}{1-\rho} \int_{x_2}^{x_1} \left[ \ln(\rho \bar{y}) + \frac{\beta}{1-\beta} \ln B \right] d\rho. \tag{13}
\]

3.3 Aggregate welfare analysis
The social planner aims to maximize aggregate social welfare for the three groups on top of expression (5) when facing a finite fund \( F \) in the sense that the total available funds \( F \) are no more than the amount to finalize the aid \( F_m \), which allows everyone inside the poverty trap to get assistance under a hybrid of top-down and bottom-up policies.\(^\text{16}\) By the assumption that \( F \) is divisible between the two different policies, the planner needs to determine the allocation of \( F \) between \( F_1 \) and \( F_2 \) to solve the following optimization problem

\[
\begin{align*}
\text{Max} & \quad W = W_1(x_1) + W_2(x_2) + W_3(x_1,x_2) \\
\text{s.t.} & \quad F_1 + F_2 \leq F \\
& \quad \frac{(1-x_1)^2}{2(1-\rho)} = F_1 \\
& \quad \frac{(x_2-\rho)^2}{2(1-\rho)} = F_2 \\
& \quad x_2 \leq x_1.
\end{align*} \tag{14}
\]

The first budget constraint indicates that the cost of all assistance should not exceed the total available amount of funds \( F \). The second constraint implies that the lower bound determines the expenses of the top-down policy. Conversely, the third constraint sets a limit on the funding allocated under the bottom-up policy. It depends on the upper bound for the bottom-up policy. The boundary restriction \( x_2 \leq x_1 \) means that the optimal upper bound of the bottom-up policy should never surpass the optimal lower bound of the top-down policy, because the minimum income after the aid by the bottom-up

\(^\text{16}\)With \( F_m \), some receivers can get a small portion that is not enough to pass the poverty trap. A detailed discussion can be seen in Proposition 1(2b). \( F_m \) is the purple curvature \( RNZ \) in Figure 2 at which \( x_1^* = x_2^* \). Its proof can be seen in Appendix A.6.4. \( F_m \) is weakly less than \( F_{all} = (1-\rho)\bar{y}/2 \) that can raise everyone out of the poverty trap.

96
policy becomes \( x_2 \), there is no individual populating further below \( x_2 \) to be aided by the top-down policy.

The optimal \( F_1 \) and \( F_2 \) base on maximizing the efficiency of available funds, by trading off the long-run effects for those who are closer to the income threshold to lift poverty completely and the short-run effect for those who are distributing at the bottom of the poverty trap to experience a temporary improvement.\(^\text{17}\) The distribution of funds also affects the utility of Group Three who will not receive any aid by affecting the exact interval for the untargeted group to locate. Accordingly, the marginal social welfare of funding the top-down policy and the bottom-up policy by dropping the common factor \( \frac{1 - \rho}{1 - \alpha \beta} \) can be expressed as

\[
\frac{\partial W}{\partial F_1} = \left( \frac{\beta}{1 - \beta} \ln \frac{A}{B} - \ln x_1 \right) \frac{1}{2} \sqrt{\frac{2(1 - \rho)}{F_1}} \tag{15}
\]

\[
\frac{\partial W}{\partial F_2} = \frac{1 - \rho}{\rho + \sqrt{2(1 - \rho)} F_2}. \tag{16}
\]

Proof: See Appendix A.4.1. \( \square \)

The marginal welfare effects of funding the top-down policy can be divided into two main components: The intensive margin: the value derived from surpassing the poverty trap at the threshold \( \frac{\beta}{1 - \beta} \ln \frac{A}{B} - \ln x_1 \). This first component encapsulates both (i) the current value of productivity improvement \( \frac{\beta}{1 - \beta} \ln \frac{A}{B} \). Let \( v \) denote such value. It is jointly formed by (increasing in) patience \( \beta \) and the productivity difference before and after jumping out of the poverty trap \( A/B \). It is straightforward that an increase in \( v \) raises the marginal welfare effects of funding the top-down policy, as it becomes more beneficial to surpass the poverty trap. (ii) and the effects of income increments \( - \ln x_1 \), as delineated in equation (15). This term can be further expressed as the income increment to reach the income threshold to lift poverty \( + (\ln 1 - \ln x_1) \).

The extensive margin: the marginal growth rate associated with the number of recipients \( \frac{1}{2} \sqrt{\frac{2(1 - \rho)}{F_1}} \). It captures the marginal effects describing the number of new beneficiaries eligible to join the top-down aid program.

Funding the bottom-up policy has implications on two fronts as well: it results in both intensive and extensive margin effects. When comparing this to the top-down policy, one key difference emerges: the intensive margin in

\(^{17}\)Equivalently, the social planner needs to choose an optimal lower bound \( x_1^* \) for a top-down policy and an optimal upper bound \( x_2^* \) for a bottom-up policy under the constraints. In some expressions, \( x_i \) and \( F_i \) for \( i = 1, 2 \) appear simultaneously for a good intuition instead of complicating the notations. For example, \( x_1 < 1 \) in equation (15), where \( F_1 \) also shows up, directly indicates that the marginal welfare effects of funding the top-down policy are strictly positive.
the bottom-up approach does not come with the added benefit of productivity improvement. Consequently, to understand the net welfare effects of funding the bottom-up policy, one should compare the increase in income effects with the rising rate of new recipients.

**Lemma 3**

(i) The welfare effects of funding a top-down policy increase with the funds, but the marginal welfare effects decrease with the funds. \( W \) is increasing and concave in \( F_1 \)

(ii) The welfare effects of funding a bottom-up policy increase with the funds, but the marginal welfare effects decrease with the funds. \( W \) is increasing and concave in \( F_2 \)

The proof is in A.4.2 and A.4.3.

According to Lemma 3(i), allocating more funds to the top-down scenario will allow more individuals below the threshold \( \bar{y} \) to lift themselves out of poverty, resulting in an overall increase in welfare through funding the top-down policy. However, it becomes increasingly costly to help those who are further below the poverty threshold, resulting in weaker marginal welfare effects as more funds are contributed to the top-down policy.

The second part of Lemma 3 suggests that having more available funds in the bottom-up scenario can increase aggregate welfare effects by allowing more individuals to receive aid and reach higher income levels. However, it also implies an increasing cost associated with maintaining a bottom-up policy since the marginal increase for the upper bound requires all individuals below to reach \( x_2 \). Additionally, the scarcity effects become weaker for new recipients who are less poor and added to the bottom-up policy.

**Lemma 4**

When available funds are scarce, the optimal aid always prioritizes the top-down policy

\[
\left( \frac{\partial W}{\partial F_1} - \frac{\partial W}{\partial F_2} \right) \bigg|_{F_1=F_2=\zeta \to 0^+} > 0.
\]

Proof: See Appendix A.5.1.

Rewriting (15) as an expression of \( F_1 \) and separating the terms embedded by the intensive margins give

\[
\frac{\partial W}{\partial F_1} = \frac{1}{2} \sqrt{\frac{2(1-\rho)}{\sqrt{F_1}}} v - \frac{\sqrt{2(1-\rho)}}{2} \ln \left[ 1 - \frac{2(1-\rho)F_1}{\sqrt{F_1}} \right]. \tag{15a}
\]

In situations where \( F_1 \) is extremely small, such that \( F_1 \to 0^+ \), then \( \frac{\partial W}{\partial F_1} \to +\infty \). This shift towards positive infinity is attributed to the marginal effects stemming from an uptick in potential beneficiaries and the existence of the additional value \( v > 0 \) by jumping out of the poverty trap. Within the framework of the top-down policy, the number of recipients can span from none to a certain
limit, resulting in an effectively infinite growth rate. The marginal increasing effect in the number of recipients magnify the marginal welfare effects of funding the top-down policy tremendously $\sqrt{2(1-\rho)}v \to +\infty$.

The second term is the compound effect consisting of the income effect $\ln \left[ 1 - \sqrt{2(1-\rho)F_1} \right]$ which diminishes to zero, and the effect of increasing the number of recipients $\sqrt{2(1-\rho)}/\sqrt{F_1}$. The second term converges to a fixed value $1-\rho$ if $F_1 \to 0^+$.

Readers may have noticed that $\frac{\partial W}{\partial F_2}$ does not approach positive infinity when $\rho = 0$, given $F_2 \to 0^+$. Even though there is a pronounced marginal increase in the number of aid recipients under the bottom-up policy, the constrained income effects $x_2 - \rho$ neutralize it, as evidenced by the third line of equation (7).\(^{18}\)

Lemma 4 indicates that with limited funds, the marginal welfare effects of funding a bottom-up policy are always relatively weak, since the available funds could be more beneficial in lifting productivity for individuals who are closest to the poverty threshold, ultimately generating a long-term welfare increase. In this case, individuals who have higher initial endowment are prioritized.

**Lemma 5** There exists a threshold value, $\bar{v}$, such that $\frac{\partial W}{\partial F_1} \bigg|_{F_1 \to F_{all}} > \frac{\partial W}{\partial F_2} \bigg|_{F_2 \to 0^+}$ when $v > \bar{v}$.

Proof: see appendix A.5.2.

The value of $\bar{v}$ is derived from comparing the marginal welfare effects of funding the top-down policy—especially when the funds are near the threshold needed to lift everyone out of poverty—to the marginal welfare effects of financing the bottom-up policy in situations where its funding is limited. This suggests that if the perceived benefits from productivity enhancements to alleviate poverty are sufficiently high, all available funds could be allocated to the top-down policy.

Based on Lemma 4 and 5, I can characterize a number of funds $F_1^p$ to ensure the top-down policy financial support. The number can range from infinitesimal to all available funding.

**Lemma 6** When $0 < v < \bar{v}$, $F_1^p$ is a function of $v$. $F_1^p(v)$ grows at an increasing rate as $v$ increases, and drops at a decreasing rate as $v$ decreases. $(\partial F_1^p(v)/\partial v > 0, \partial^2 F_1^p(v)/\partial v^2 > 0)$.

Proof: See Appendix A.5.3.

\(^{18}\)However, when $\rho = 0$, $\frac{\partial W}{\partial F_2}$ approaches positive infinity under $F_2 \to 0^+$. This is because, when the poorest recipient has nothing, the marginal welfare impact of financing the bottom-up policy, which caters to such individuals, is immense.
When \(0 < v < \bar{v}\), the funds required to exclusively support a top-down policy favoring those in higher income brackets depend on the value of \(v\), denoted by \(F^p_1(v)\). If \(v\) is not high enough to make the top-down policy dominant everywhere, it is optimal to prioritize the top-down policy with a certain sum of funds. It means that all of the available funds \(F\), when \(F \leq F^p_1(v)\), will be given to the policy that favors the implicitly poor group.

For the trajectories of \(F^p_1(v)\), as the value from enhanced productivity (\(v\)) increases, the marginal welfare of funding the top-down policy also increases. This creates more space for carrying out the policy. It also indicates that the funds needed to prioritize the top-down policy do not increase proportionally, but at a higher rate. This is because the decreasing rate for \(\frac{\partial W}{\partial F_1}\) goes down as more funds are allocated.

3.4 The optimal aid distribution

I will consider differentiated amounts of funds for assistance in this section as a general case, as the current value of productivity improvement by lifting poverty permanently and available funds can jointly form the optimal aid distribution for the recipients inside the poverty trap. The complete policy design can be seen in the following diagram.

*Figure 2. The optimal aid design under different \(F\) and \(v\)*
Proposition 1 The optimal aid distribution always prioritizes the top-down policy $F_1 > 0$.

1. If the current value of productivity improvement from alleviating poverty is high enough ($\bar{v}$), the top-down policy takes precedence in all scenarios.
   \[ F = F_1 > 0, F_2 = 0 \] (zone I)

2. If the current value of productivity improvement by lifting poverty is moderate ($0 < v < \bar{v}$), the available funds $F$ will be given to
   a) the top-down policy if $F_1 > 0, F_2 = 0$ (zone II);
   b) the top-down and the bottom-up policy if $F > F^*_{10}$ according to the equalization of marginal welfare across the two policies by equation (17).
   \[ F_1 > 0, F_2 > 0 \] (zone III).

Proof: See Appendix A.6.

One key finding is that aid always begins with those closest to the threshold of escaping the poverty trap. This aligns with the notion that those with the most potential to improve their situation should be assisted first, especially since the cost of helping them is low. Consequently, the implicitly poor warrant first assistance, even when there are ultra-poor individuals still awaiting aid. The amount of funds earmarked specifically for the top-down policy can vary from an infinitesimal sum to an amount sufficient to elevate everyone, depending on the current value of productivity improvement.

Given that the marginal welfare effects of funding the top-down policy are diminishing, and both $F$ and $v$ jointly influence aid distribution, the first part of Proposition 1 posits that a decline in these marginal welfare effects does not compromise its dominance, provided $v$ remains high, irrespective of available funding levels. Here, the diminishing marginal welfare effects from continuously funding a top-down policy still exceed those from supporting a bottom-up approach. The considerable gains in productivity after escaping poverty can offset the rising opportunity costs of forgoing the bottom-up policy. Therefore, all available resources should be channeled into the top-down policy to effectively alleviate poverty. This aligns with prior literature, which promotes “A big push” to achieve a significant impact within a poverty trap.

This aid policy does not favor the more disadvantaged if the total amount of funds is not sufficient for everyone to surpass the poverty trap. The top-down policy continues over the rest of the ultra-poor if more funds are available. The lower bound of a top-down policy $x^*_1$ approaches $\rho$ if the available funds increase $F \rightarrow F^{all}$ to cover everyone to escape poverty.

This policy design requires exceptional improvement in productivity after surpassing the trap (high $A/B$) and a discount factor (high $\beta$). Intuitively, staying out of the poverty trap will result in a great advantage of accumulating
wealth, and the individual is very patient with such an inflow of welfare in the future driven by better productivity.

As \( v \) decreases, \( F^p_1(v) \) contracts its space to prioritize the top-down policy. The second section posits that a moderate \( v \) can lead to a potential equalization of marginal welfare effects. The actualization of this equalization hinges on whether the available funds surpass the amount needed to prioritize the top-down policy \( F^p_1(v) \).

(2a) posits that the funding amount \( F \) does not exceed \( F^p_1(v) \) in order to achieve equalized marginal welfare between \( \frac{\partial W}{\partial F_1} \) and \( \frac{\partial W}{\partial F_2} \). This is because, for \( 0 < v < \bar{v} \), \( \frac{\partial W}{\partial F_1} \geq \frac{\partial W}{\partial F_2} \) holds true when \( F \leq F^p_1(v) \). This suggests that financing the top-down policy yields welfare improvements that are at least as great as those from the bottom-up policy.

(2b) asserts that if the available funds fall within the range \( F^p_1(v) < F \leq F^m \), both policies receive support. The top-down policy receives priority funding of \( F^p_1(v) \) because

\[
\frac{\partial W}{\partial F_1} < \frac{\partial W}{\partial F_2} \quad \text{for } F^p_1(v) \leq F \leq F^m
\]

and the remaining amount \( F - F^p_1(v) \) is then allocated between the two policies.

After prioritizing the better-off individuals inside the poverty trap with funds \( F^p_1(v) \), exclusive funding to the top-down policy will not be optimal because the marginal welfare effects from funding a bottom-up policy, which are relatively lower when \( F < F^p_1 \), become equally beneficial when \( F \geq F^p_1(v) \). A momentary improvement for a mass of ultra-poor people \( [\rho, x^*_2] \) will become as equally crucial as raising the poor distributing in a higher income level \( [x^*_1, 1] \) to lift poverty permanently. \( (F - F^p_1(v)) \) will be divided between the two scenarios for maintaining the equalization of marginal welfare effects, which are descending though. As a result, both top-down and bottom-up policies have their spaces, reflected by lowering the optimal lower bound of a top-down policy \( x^*_1 \) and raising the optimal upper bound of a top-down policy \( x^*_2 \).

The optimal aid distribution in (2b) is a policy mix where the top-down and bottom-up policies coexist. It is the interior solution to the optimization problem, where the social welfare maximization is realized by equalizing marginal social welfare across the two different aid policies, i.e. \( \frac{\partial W}{\partial F_1} = \frac{\partial W}{\partial F_2} \). Therefore, I carry out first-order conditions with respect to the arguments. The optimality condition can be expressed by the optimal lower bound of the top-down policy \( x^*_1 \) and optimal upper bound of the bottom-up policy \( x^*_2 \) for convenience

\[
v = \ln x^*_1 + \frac{1 - x^*_1}{x^*_2}
\]

with the budget constraint

\[
\frac{(1 - x^*_1)^2}{2(1 - \rho)} + \frac{(x^*_2 - \rho)^2}{2(1 - \rho)} = F.
\]
See derivation in Appendix A.6.3.

The two conditions indicate that \( v \) in conjunction with \( F \) determines where \( x_1^* \) and \( x_2^* \) locate after the aid, implying (a) how many recipients will escape poverty, (b) how many will reach \( x_1^* \) temporarily, and (c) how large the vacuum space is for those who do not get any aid. Equation (17) states that \( v \) determines the internal allocation between the two different assistance scenarios given \( F \). Equation (18) implies that how far a policy can reach depends on the available funds given \( v \). Based on Proposition 1(2b), I proceed to Corollary 1 for comparative statics of the optimal boundaries \( x_1^* \) and \( x_2^* \).

**Corollary 1** Given an exogenous amount of funds \( F^p(v) < F \leq F^m \) and the current value of lifting poverty from productivity improvement \( 0 < v < \bar{v} \)

(i) the higher \( v \),
the lower the optimal upper bound of the bottom-up policy is \( (\partial x_2^* / \partial v < 0) \),
the lower the optimal lower bound of the top-down policy is \( (\partial x_1^* / \partial v < 0) \).
The opposite is true for the lower \( v \).
(ii) More funds benefit the bottom-up policy and the top-down policy
\( (\partial x_2^* / \partial F > 0 \text{ and } \partial x_1^* / \partial F < 0) \).

The proof can be shown straightforwardly through comparative statics on expressions (17) and (18).

By (i) in Corollary 1, a higher current value in productivity improvement via lifting poverty permanently indicates that there will be more space for a top-down policy and less space for a bottom-up policy. It lends support to the top-down policy that prioritizes those who have more initial income to get rid of the poverty trap, by infusing returns to the scenario, because the marginal welfare effects are more potent by helping people obliterate poverty. A lower value of removing poverty will give more space to a bottom-up policy, indicating that getting out of the poverty trap might not always be optimal for social welfare maximization. The second part shows that an increase in funds creates the possibility of expanding recipients.

### 4 Robustness

*Population distributions.* The optimal distribution of aid represents the maximum marginal welfare effects for different levels of funding and the value of poverty alleviation. When the population is uniformly distributed in the poverty trap, the optimal allocation of funds between different aid policies can be easily determined through analytical solutions. The previous propositions remain valid and provide practical policy guidance without such a uniform distribution. However, the specific amounts of funding required by an aid program and relevant marginal welfare changes may differ, depending on the number of individuals in each income interval which applies to a country that has specific income distribution in designing an optimal aid distribution.
Shocks that affect specific productivity and patience. The analysis is based on the comparative statics of relative productivity advantages and discounting effects. In the model, there are no restrictions on productivity after surpassing the poverty trap, leaving room for the possibility of progress or setbacks due to exogenous shocks. For instance, in the early stages of market and trade liberalization in developing countries, such as the economic reform in China during the 1980s, there was significant space for funding a top-down policy as increased productivity led to substantial improvements in wealth. Similarly, if there is an increase in individuals’ patience, which can be achieved through better access to education and positive social values, the outcome will also be favorable for a top-down policy. However, the opposite can also be true in the aftermath of natural disasters, wars, pandemics, and other crises that cause a drop in productivity and an increase in impatience due to the chaotic social environment. In such cases, a top-down policy may have less space to operate effectively, coexisting with a bottom-up policy.

Utility function forms. The trade-off effects between funding top-down and bottom-up policies are determined by a concave utility function. If such a function is not present, funding someone in a higher income distribution is always preferred. However, for specific forms of concave utility functions, additional assumptions may be required to formulate analytical solutions.

5 Conclusions

A poverty trap is a situation where people are stuck in an ongoing cycle of poverty and are unable to break out of it. This paper gives an analytical implication on one-off transfer assistance in the case of a friction-based poverty trap that has nothing to do with individual faults. It summarizes the effect of one-off transfers by considering different quantities of assistance, where the entirety of poor individuals inside the poverty trap are considered. Besides ultra-poor people, the rest of the poor, including the implicitly poor, are considered recipients due to their lack of attention from the public and potential welfare, which a less costly push can realize.

Alleviating poverty is complicated and multidimensional. The complexity can be summarized as a lack of funds to lift poverty for everyone, where information frictions, moral hazard, rent-seeking, etc., can be generalized as an expenditure for available funds. The funding problems are widely spread across the world. Otherwise, poverty would not be a severe issue. The restriction on available funds motivates a social planner to make trade-offs and priorities in spending decisions. In this context, the current value of productivity improvement by lifting poverty is deterministic. When funds start with a significant limitation, the optimal assistance always prioritizes those who are better off in initial endowment to jump out of the poverty trap. When there are more funds at the disposal, there is a possibility to form an optimal policy
mix by funding two policies simultaneously. To my best knowledge, this possibility contributes to forming a comprehensive policy for alleviating poverty when funds are lacking.

One concern would be to find the threshold that separates implicitly poor and wealthy individuals, even if Balboni et al. (2022) empirically found a poverty trap based on asset levels in Bangladesh. This impedes the exact allocation of funds to poor individuals to circumvent a poverty trap. One way is to add uncertainty of the threshold in the model. The volatility, if severe, would demand more funds in expectation. It, therefore, raises the opportunity costs of giving up a group of ultra-poor people who can be sure to improve their economic situation with such increased funds.
References


Appendix A  Appendix

A.1 The proof for the sum of discounted utilities over time for one individual

The sum of discounted utility in a convergence path is

\[ U^j = \ln c^j_0 + \beta \ln c^j_1 + \beta^2 \ln c^j_2 + \ldots + \beta^T \ln c^j_T. \]

With the tight individual budget constraint, the utility can be translated into the expression of income \( y \)

\[ U^j = \ln y^j_0 + \beta \ln y^j_1 + \beta^2 \ln y^j_2 + \ldots + \beta^T \ln y^j_T. \]  \( \text{(A1)} \)

By substituting the income at each period expressed by initial income, productivity, and intertemporal income elasticity, we can update the aggregated utility as

\[ U^j = \ln (\rho^j \bar{y}) + \beta \ln \left[ \prod (\rho^j \bar{y})^a \right] + \beta^2 \ln \left[ \prod^{a+1} (\rho^j \bar{y})^{a^2} \right] + \ldots \]

\[ + \beta^T \ln \left[ \prod^{aT-1+aT-2+\ldots+a+1} (\rho^j \bar{y})^{a^T} \right], \]  \( \text{(A2)} \)

where the income timeline follows

at time 0/ beginning \( y^j_0 = \rho^j \bar{y} \),

at time 1 \( y^j_1 = \prod (\rho^j \bar{y})^a \),

at time 2 \( y^j_2 = \prod^{a+1} (\rho^j \bar{y})^{a^2} \),

at time 3 \( y^j_3 = \prod^{a^2+a+1} (\rho^j \bar{y})^{a^3} \),

\ldots

at the terminal time \( T \)

\( y^j_T = \prod^{aT-1+aT-2+\ldots+a+1} (\rho^j \bar{y})^{a^T} \).

With geometric progression and rearrangement

\[ U^j = \frac{1 - (a\beta)^{T+1}}{1 - a\beta} \ln (\rho^j \bar{y}) + \left( \frac{\beta - aT^T \beta^{T+1}}{1 - a \beta} - \frac{1 - aT^T}{1 - a} \beta^{T+1} \right) \frac{\ln P}{1 - \beta}. \]  \( \text{(A3)} \)

With the terminal period \( T \) going to positive infinite, \( \beta^T \rightarrow 0 \), the utility that individual \( j \) can get over all periods converges to

\[ U^j = \frac{1}{1 - a\beta} \left[ \ln (\rho^j \bar{y}) + \frac{\beta}{1 - \beta} \ln P \right]. \]
A.2 The proof for Lemma 1

We will compare the marginal welfare effects with respect to funds between one individual above the threshold $\rho^*$ and one inside the poverty trap $\rho$. Suppose that $F$ is a marginal amount of funds for the take-it-or-leave-it comparison, the utility that $\rho^*$ gets the aid becomes

$$U(\rho^* \bar{y} + F) = \frac{1}{1 - \alpha \beta} \ln(\rho^* \bar{y} + F) + \frac{\beta}{1 - \alpha \beta} \ln A,$$

(A4)

based on individual utility inside the system-equation (4). Then we differentiate (A4) with respect to the funds $F$ to get the marginal utility

$$\frac{\partial U(\rho^* \bar{y} + F)}{\partial F} = \frac{1}{1 - \alpha \beta} \frac{1}{\rho^* \bar{y} + F}.$$

(A5)

The equivalent amount of fund $F$ can alternatively be allocated to one poor individual inside the poverty trap. The utility that $\rho$ can get by social assistance is

$$U(\rho \bar{y} + F) = \frac{1}{1 - \alpha \beta} \ln(\rho \bar{y} + F) + \frac{\ln B}{1 - \beta} \frac{\beta}{1 - \alpha \beta}.$$

(A6)

The relevant marginal utility effect for (A6) is

$$\frac{\partial U(\rho \bar{y} + F)}{\partial F} = \frac{1}{1 - \alpha \beta} \frac{1}{\rho \bar{y} + F}.$$

(A7)

With $\rho^* > 1 > \rho$, the denominator of (A5) is higher than that of (A7). We can get

$$\frac{\partial U(\rho^* \bar{y} + F)}{\partial F} < \frac{\partial U(\rho \bar{y} + F)}{\partial F}. \quad \text{1}$$

Consequently, funding individuals inside the poverty trap always yields more social welfare than funding individuals above the poverty trap. □

A.3 The proof that bottom-up and top-down policies are the only potential best instruments.

When we have limited resources, such as a small amount of money to spend, it can be challenging to decide how to prioritize spending. Why do we prioritize either a top-down that pushes individuals out of the poverty trap from the richest individuals (Section A.3.2) or a bottom-up policy that aids in a momentary improvement from the poorest individuals (Section A.3.1), instead of other

\footnote{The effect of productivity differentiation is smoothed out in this comparison. The comparison also rules out the effect of convergence speed which was thought potentially beneficial for those who are above the threshold, since the terminal period goes infinite and the net discount factor converges to $1/(1 - \alpha \beta)$ as an identical term for all potential recipients. The key difference is the scarcity effect determined by initial income.}
strategies, for example, giving the limited amount of funds to any arbitrary individuals e.g., someone in the middle of income distribution? The prioritization we will prove states that two policy instruments are continuous in design. To prove this, we need the individual utility function in the economic system

\[ U^j = \frac{1}{1 - a\beta} \left[ \ln \left( \frac{\rho^j}{\bar{y}} \right) + \frac{\beta}{1 - \beta} \ln P \right] \text{ for } P = \begin{cases} A & \rho^j \geq 1 \\ B & \rho \leq \rho^j < 1 \end{cases} \text{ where } A > B. \]

for welfare analysis. There is a threshold externality after jumping out of \( \rho \leq \rho < 1 \). Accordingly, the increased utility from momentary improvement to reach the income level \( x_2 < 1 \) for an individual is

\[ \Delta U_2 = U(\rho = x_2) - U(\rho) = \frac{1}{1 - a\beta} \left( \ln \frac{x_2}{\rho} \right). \quad (A8) \]

The increased utility is driven by how scarce the receiver’s initial income \( \rho \) is and the income after the aid \( x_2 \). Similarly, the increased utility of lifting poverty for an individual is

\[ \Delta U_1 = U(\rho = 1) - U(\rho < 1) = \frac{1}{1 - a\beta} \left( \ln \frac{1}{\rho} + \frac{\beta}{1 - \beta} \ln \frac{A}{B} \right). \quad (A9) \]

The increased welfare by pushing individuals out of the poverty trap has two aspects:
(i) the first term summarizes the scarcity effects of reaching the threshold, which is dependent on the initial income and leads to a corresponding increase in utility;
(ii) the second term \( v = \frac{\beta}{1 - \beta} \ln \frac{A}{B} \) represents the current value gained from improvements in productivity.

A.3.1 The bottom-up policy
The marginal increment in utility for an individual without eradicating poverty completely is

\[ \frac{\partial \Delta U_2}{\partial \rho} = -\frac{1}{1 - a\beta} \frac{1}{\rho} < 0 \quad (A10) \]

indicating that the increased welfare is decreasing in \( \rho \). It suggests that when we help someone who is further below, it is more likely to result in a greater increase in overall welfare than if we help someone who is only slightly above, when everyone can have an equally small increase in their initial income. Such prioritization supports a bottom-up policy as a possible instrument when funds are not sufficient for everyone to surpass the poverty trap.

A.3.2 The top-down policy
The objective is to demonstrate that prioritizing a top-down approach is locally advantageous but can be globally advantageous if \( v \) is high enough. This
approach involves lifting individuals out of poverty starting from a higher income group rather than a lower one to achieve a discrete jump in welfare. If pushing individuals who are from a higher income group is not more advantageous than those in a lower one, the welfare increment achieved by funding the latter group will be less than the welfare increment achieved by funding those who are further below for a temporary improvement, which is the main goal of a bottom-up policy.

Let us start with some notations and expressions. The aggregate welfare increment to aid individuals in an arbitrary interval \([b, a]\) to surpass the poverty trap is

\[
\Delta W_{1,ab} = \int_b^a \Delta U_1 f(\rho) d\rho = \frac{1}{1 - \alpha\beta} \frac{1}{1 - \rho} \int_b^a \left[ \ln \left( \frac{1}{\rho} \right) + v \right] d\rho. \tag{A11}
\]

By calculation,

\[
\Delta W_{1,ab} = \frac{1}{1 - \alpha\beta} \frac{1}{1 - \rho} [ (a-b)(v+1) - a \ln a + b \ln b]. \tag{A.11a}
\]

Suppose an amount of funds \(F_{ab}\) that is exactly enough to push individuals populating in \([b, a]\) out of the poverty trap. The budget constraint specifies as

\[
\int_b^a (1 - \rho) f(\rho) d\rho = F_{ab} \text{ with } 0 < \rho \leq b < a < 1. \tag{A12}
\]

By operation

\[
(b-1)^2 - (a-1)^2 = 2(1-\rho)F_{ab}. \tag{A.12a}
\]

To depict a small number of funds at the disposal, we let \(b = a - d\) for \(d > 0\) as a small positive number. The welfare increment updates as

\[
\Delta W_{1,a,a-d} = \frac{1}{1 - \alpha\beta} \frac{1}{1 - \rho} \left[ d(v+1) - a \ln a + (a-d) \ln(a-d) \right] \tag{A11b}
\]

s.t. \((a-d-1)^2 - (a-1)^2 = 2(1-\rho)F_{a,a-d}.

The budget constraint can be rewritten as

\[
d(2-2a+d) = 2(1-\rho)F_{a,a-d}. \tag{A.12b}
\]

Let us define the welfare increment by raising the individuals populating in a higher income distribution to jump out of the poverty trap as \(\Delta W_{1,mn}\), where we pick up the individuals populating between \([n, m]\) for \(a < m < 1\). By applying to (A11), the welfare increment specifies

\[
\Delta W_{1,mn} = \frac{1}{1 - \alpha\beta} \frac{1}{1 - \rho} [(m-n)(v+1) - m \ln m + n \ln n] \tag{A13}
\]
and corresponding budget constraint
\[(n - 1)^2 - (m - 1)^2 = 2(1 - \rho)F_{mn}. \quad \text{(A14)}\]

With the equivalent amount of disposable funds \(F_{a,a-d} = F_{mn}\), we get the expression of the lower bound \(n\) expressed by the available funds and the upper bound \(m\)
\[1 - n = \sqrt{d(2 - 2a + d) + (m - 1)^2} \text{ and } n = 1 - \sqrt{d(2 - 2a + d) + (m - 1)^2}. \quad \text{(A15)}\]

The welfare increment by funding \([n, m]\) to jump out of the poverty trap by plugging (A15) into (A13) becomes
\[\Delta W_{1, mn} = \frac{1}{1 - \alpha \beta} \frac{1}{1 - \rho} \left\{ \ln \left( \frac{x_2}{\rho} \right) f(\rho) \right\} \text{ s.t. } (x_2 - q) = \sqrt{\frac{2F_2}{f(\rho)}}. \quad \text{(A17a)}\]

Next, we define the welfare increment from a temporary improvement \(\Delta W_2\) which is part of a bottom-up policy that covers those who are further below than \(a\). Let us take the integral over \(q\) to \(x_2\) with the lower bound \(q < a\),
\[\Delta W_2 = \int_q^{x_2} \ln \left( \frac{x_2}{\rho} \right) f(\rho) d\rho \quad \text{(A17)}\]
under the budget constraint
\[\int_q^{x_2} (x_2 - \rho) f(\rho) d\rho = F_2. \quad \text{(A18)}\]
Solving the integrals yields
\[\Delta W_2 = \frac{1}{1 - \alpha \beta} \frac{1}{1 - \rho} [(x_2 - q) - q(\ln x_2 - \ln q)] \quad \text{(A17a)}\]
s.t. \((x_2 - q) = \sqrt{\frac{2F_2}{f(\rho)}}. \quad \text{(A18a)}\]
With an equivalent amount of funds \(F_2 = F_{a,a-d}, \ x_2 = q + \sqrt{d(2 - 2a + d)}\)
\(\Delta W_2\) updates as
\[\Delta W_2 = \frac{1}{1 - \alpha \beta} \frac{1}{1 - \rho} \left\{ \sqrt{d(2 - 2a + d)} - q[\ln(q + \sqrt{d(2 - 2a + d)}) - \ln q] \right\}. \quad \text{(A18b)}\]

In what follows, we will show if \(\Delta W_{1,mn} - \Delta W_{1,a,a-d} < 0\) when \(d \to 0^+\), then \(\Delta W_2 - \Delta W_{1,a,a-d} > 0\), meaning that \(\Delta W_{1,a,a-d} < \max \{\Delta W_{1,mn}, \Delta W_2\}\) for \(\rho \leq \rho^*\).
With (A21), we can get

\[ v \]

below. This requires a policy instrument than a temporary improvement for those who are further away. The top-down prioritization is locally continuous, but an arbitrary target to lift poverty for someone in a lower income group is never best. This implies that top-down prioritization is locally continuous, but an arbitrary target to lift poverty for someone in a higher income distribution out of the poverty trap is always more beneficial. Otherwise, we will look into the second case.

\[ \text{Case 1. If } \Delta W_{1,mn} > \Delta W_{1,a,a-d}, \text{ for } d \to 0^+, \text{ top-down prioritization is continuous everywhere.} \]

\[ \frac{\partial (\Delta W_{1,mn} - \Delta W_{1,a,a-d})}{\partial d} = \left[ \frac{1-a+d}{\sqrt{d(2-2a+d)+(m-1)^2}} - 1 \right] v \]

\[ = \frac{1-a+d}{\sqrt{d(2-2a+d)+(m-1)^2}} \ln \left( 1 - \sqrt{d(2-2a+d)+(m-1)^2} \right) + \ln(a-d). \]

\[ \lim_{d \to 0^+} \frac{\partial (\Delta W_{1,mn} - \Delta W_{1,a,a-d})}{\partial d} = \frac{m-a}{1-m} - \frac{1-a}{1-m} \ln m + \ln a > 0 \]

if \( v > \frac{(1-m)(1-a)}{m-a} \left( \frac{\ln m}{1-m} - \frac{\ln a}{1-a} \right). \) If the condition that \( v \) is high enough can hold, raising someone in a higher income distribution out of the poverty trap is always more beneficial. The top-down prioritization is globally continuous. Otherwise, we will look into the second case.

\[ \text{Case 2. If } \Delta W_{1,mn} < \Delta W_{1,a,a-d}, \text{ for } d \to 0^+, \Delta W_2 > \Delta W_{1,a,a-d}. \]

This implies that top-down prioritization is locally continuous, but an arbitrary target to lift poverty for someone in a lower income group is never best the policy instrument than a temporary improvement for those who are further away. This requires

\[ v < \frac{(1-m)(1-a)}{m-a} \left( \frac{\ln m}{1-m} - \frac{\ln a}{1-a} \right). \]  

The difference between \( \Delta W_2 \) and \( \Delta W_{1,a,a-d} \) is

\[ \Delta W_2 - \Delta W_{1,a,a-d} = \frac{1}{1-\alpha \beta} \left\{ \frac{1}{1-\beta} \left\{ \sqrt{d(2-2a+d)} - q \ln(q + \sqrt{d(2-2a+d)}) \right\} + \ln q - d(v+1) + a \ln a - (a-d) \ln(a-d) \right\}. \]

With (A21), we can get

\[ -d(v+1) > -d \left[ \frac{(1-m)(1-a)}{m-a} \left( \frac{\ln m}{1-m} - \frac{\ln a}{1-a} \right) + 1 \right]. \]  

(A21a)

(q < a < m < 1). To this end, we will use the welfare comparison between \( \Delta W_{1,mn} \) and \( \Delta W_{1,a,a-d} \) by dropping the common factor
The substitution of (A21) into (A22) generates the following inequation

\[
\Delta W_2 - \Delta W_{1,a,a-d} > \frac{1}{1 - \alpha} \frac{1}{1 - \beta} \times \left\{ \sqrt{d(2 - 2a + d)} - q \ln(q + \sqrt{d(2 - 2a + d)}) + \ln q + a \ln a \\
- d \left[ (1-m)(1-a) \left( \frac{\ln m}{1-m} - \frac{\ln a}{1-a} \right) + 1 \right] - (a - d) \ln(a - d) \right\}.
\]

\[ \text{(A23)} \]

Let us define the terms inside the bracket as \( G(d) \). Now we would like to show \( G(d) > 0 \) for \( d \to 0^+ \).

\[
\frac{\partial G(d)}{\partial d} = \frac{1 - a + d}{q + \sqrt{d(2 - 2a + d)}} - \frac{(1 - m)(1 - a)}{m - a} \left( \frac{\ln m}{1-m} - \frac{\ln a}{1-a} \right) + \ln(a - d).
\]

\[ \text{(A24)} \]

By taking the limit

\[
\lim_{d \to 0^+} \frac{\partial G(d)}{\partial d} = \frac{1 - a}{q} - \frac{(1 - m)(1 - a)}{m - a} \left( \frac{\ln m}{1-m} - \frac{\ln a}{1-a} \right) + \ln a.
\]

By rearrangement,

\[
\lim_{d \to 0^+} \frac{\partial G(d)}{\partial d} = (1 - a) \left( \frac{1}{q} - \frac{\ln m - \ln a}{m - a} \right) > 0,
\]

\[ \text{(A25)} \]

where the sign of the second parenthesis is positive because the slope of \( \ln(\cdot) \) is steeper at \( q \), than the slope of the segment line between \( a \) and \( m \) if \( q < a < m \).

In the following figure, it is clear that the slope of the green line is steeper than that of the red line.

\[ \text{Figure A.1. The graphical explanation for Case 2} \]
Therefore we can find \( d = \varepsilon > 0 \), so that \( \Delta W_2 - \Delta W_{1,a,a-d} \) is increasing in \( d \), making \( G(\varepsilon) > 0 \) and thus \( \Delta W_2 > \Delta W_{1,a,a-d} \) for \( q < a \).

Conclusively, even if the value in productivity improvement by jumping out of the poverty trap may not be high enough to ensure a global prioritization of lifting poverty for someone in a higher income group \( \Delta W_{1,mm} < \Delta W_{1,a,a-d} \), but the scenario of funding someone in a lower income group will not get any small amount of funds in practice, because to help those who are further below for a temporary improvement can generate more welfare \( \Delta W_2 > \Delta W_{1,a,a-d} \). That is \( \Delta W_{1,a,a-d} < \max \{ \Delta W_{1,mm}, \Delta W_2 \} \) for \( 0 < q < a < m < 1 \). Only top-down and bottom-up policies can be potentially the best policy instruments in aiding the individuals inside the poverty trap. Meanwhile, we can also show that for any positive value of \( v > 0 \), assisting the group closer to the threshold \( \bar{y} \) is always more advantageous than funding any group in a lower distribution in Appendix B.

We can also rule out any arbitrary policy for assisting individuals with a discrete amount of funds that are not sufficient to jump out of the poverty trap, which could be considered an insufficient push scenario. Allocating a small amount of funds to a single individual is less effective than distributing the funds to those who are further below. This is because the welfare effect is trivial since in a continuous interval, the integral value of an individual is zero. This approach is not included in aid policy design because bottom-up policies have proven to be the most effective tool for temporary improvement. We do not need to compare the welfare effects of lifting someone else above. \( \square \)

A.4 The proof related to marginal welfare effects in funding different policies

A.4.1 The proof for marginal welfare effects of funding the top-down and the bottom-up policy

To derive the marginal welfare effects with respect to different funding scenarios, \( \partial W / \partial F_1 \) and \( \partial W / \partial F_2 \), we need the aggregation welfare function

\[
W = \left[ \ln \bar{y} + \frac{\beta}{1-\beta} \ln A \right] (1-x_1) + \left[ \ln (x_2 \bar{y}) + \frac{\beta}{1-\beta} \ln B \right] (x_2 - \rho)
\]

\( (A26) \)

where we drop the joint scalar, a net discounting effect in a dynamic framework \( 1/(1 - \alpha \beta) \), and the density function \( 1/(1 - \rho) \), since they are identical for the three welfare components.

The marginal welfare effects of funding the top-down policy \( \partial W / \partial F_1 \) by chain rules in derivatives can be given by

\[
\frac{\partial W}{\partial F_1} = \frac{\partial W}{\partial x_1} \frac{\partial x_1}{\partial F_1}.
\]
With (A26),
\[
\frac{\partial W}{\partial x_1} = \ln x_1 - \frac{\beta}{1 - \beta} \ln \frac{A}{B}.
\]

By solving the budget constraint for Scenario One, we can arrive at
\[
x_1 = 1 - \sqrt{2(1 - \rho)F_1}.
\]

So
\[
\frac{\partial x_1}{\partial F_1} = -\frac{1}{2} \frac{\sqrt{2(1 - \rho)}}{\sqrt{F_1}}.
\]

We, therefore, get the expression (15).

The marginal welfare effects of funding the bottom-up policy \(\frac{\partial W}{\partial F_2}\) can also be derived from chain rules
\[
\frac{\partial W}{\partial F_2} = \frac{\partial W}{\partial x_2} \frac{\partial x_2}{\partial F_2}.
\]

The derivative of equation (A26) with respect to \(x_2\) yields
\[
\frac{\partial W}{\partial x_2} = \frac{x_2 - \rho}{x_2}.
\]

The budget constraint for the second scenario yields
\[
x_2 = \rho + \sqrt{2(1 - \rho)F_2}.
\]

\[
\frac{\partial x_2}{\partial F_2} = \frac{1}{2} \frac{\sqrt{2(1 - \rho)}}{\sqrt{F_2}}
\]

is derived accordingly. Thus
\[
\frac{\partial W}{\partial F_2} = \left( \frac{x_2 - \rho}{x_2} \right) \frac{1}{2} \frac{\sqrt{2(1 - \rho)}}{\sqrt{F_2}}.
\]

By iteration of \(x_2 = \rho + \sqrt{2(1 - \rho)F_2}\), we can get the expression (16). \(\square\)

A.4.2 The first part of Lemma 3
We can get \(\frac{\partial W}{\partial F_1} > 0\) for \(0 < x_1 < 1\) in (15). Rewriting \(\frac{\partial W}{\partial F_1}\) as
\[
\frac{\partial W}{\partial F_1} = \frac{\sqrt{2(1 - \rho)}}{2} \left[ v - \ln \left( 1 - \sqrt{2(1 - \rho)F_1} \right) \right] \frac{1}{\sqrt{F_1}}
\]
by substituting \( x_1 = 1 - \sqrt{2(1 - \rho)F_1} \) to derive \( \frac{\partial^2 W}{\partial (F_1)^2} \).

\[
\frac{\partial^2 W}{\partial (F_1)^2} = \frac{\sqrt{2(1 - \rho)}}{4} \left\{ \frac{\sqrt{2(1 - \rho)}}{1 - \sqrt{2(1 - \rho)F_1}} \frac{1}{F_1} - \frac{v - \ln \left(1 - \sqrt{2(1 - \rho)F_1}\right)}{(F_1)^{\frac{3}{2}}} \right\}.
\]

By rearrangement,

\[
\frac{\partial^2 W}{\partial (F_1)^2} = \frac{\sqrt{2(1 - \rho)}}{4} \frac{1}{(F_1)^{\frac{3}{2}}} \left\{ \frac{1 - x_1}{x_1} + \ln (x_1) - v \right\}. \quad (A27)
\]

If \( v \) is very high, \( \frac{\partial^2 W}{\partial (F_1)^2} \) will always be negative. It means that top-down prioritization is global, as what we state in the first case of Appendix A.3. Otherwise, a moderate value of \( v \) yields a locally continuous top-down prioritization as we proved in the second case and the equalization between \( \frac{\partial W}{\partial F_1} \) and \( \frac{\partial W}{\partial F_2} \) that is \( \ln (x_1) - \frac{\beta}{1 - \rho} \ln \frac{A}{B} = -\frac{1 - x_1}{x_2} \), (A27) can be updated as

\[
\frac{\partial^2 W}{\partial (F_1)^2} = \frac{\sqrt{2(1 - \rho)}}{4} \frac{1}{(F_1)^{\frac{3}{2}}} \left\{ \frac{1 - x_1}{x_1} - \frac{1 - x_1}{x_2} \right\}. \quad (A27a)
\]

By arrangement,

\[
\frac{\partial^2 W}{\partial (F_1)^2} = \frac{\sqrt{2(1 - \rho)}}{4} \frac{1}{(F_1)^{\frac{3}{2}}} \frac{1 - x_1}{x_1} x_2 \leq 0, \text{ given } \rho \leq x_2 \leq x_1 < 1
\]

where \( \frac{\partial^2 W}{\partial (F_1)^2} = 0 \) when \( x_2 = x_1 \).

**A.4.3 The second part of Lemma 3**

\( \frac{\partial W}{\partial F_2} > 0 \) for \( 0 < \rho < 1 \) in (16). The derivative of \( \frac{\partial W}{\partial F_2} \) with respect to \( F_2 \) yields

\[
\frac{\partial^2 W}{\partial (F_2)^2} = -\frac{(1 - \rho)}{\left(\rho + \sqrt{2(1 - \rho)F_2}\right)^2} \frac{\sqrt{2(1 - \rho)}}{2\sqrt{F_2}}. \quad (A28)
\]

\( \frac{\partial^2 W}{\partial (F_2)^2} < 0 \) for \( 0 < \rho < 1 \).

**A.5 The proof related to Figure 2**

This section provides the proofs of the proposition and lemmas that are relevant to Figure 2. More specifically, it contains the proofs for the existence of threshold values \( \bar{v} \), the curvature of \( F_1^p(v) \) and RNZ.
A.5.1 The proof for Lemma 4

The section aims to prove that given the existence of current value of productivity improvement $v > 0$ and when funds are very limited, the top-down policy is always prioritized, we can carry out marginal welfare comparison between $\frac{\partial W}{\partial F_1}$ and $\frac{\partial W}{\partial F_2}$ when funds are minimal, where we are doing a keep-it-leave-it comparison to decide which group deserves the limited funds exclusively. Suppose that $F_1 = F_2 = \zeta \to 0^+$, we would like to prove

$$\lim_{\zeta \to 0^+} \left[ \frac{\partial W}{\partial F_1} - \frac{\partial W}{\partial F_2} \right]_{F_1 = F_2 = \zeta} > 0$$

such that the top-down policy always gets prioritized when funds are limited. The difference in marginal welfare of funding Scenarios One and Two is

$$\left( \frac{\partial W}{\partial F_1} - \frac{\partial W}{\partial F_2} \right)_{F_1 = F_2 = \zeta} = \frac{1}{2} \sqrt{\frac{2(1-\rho)}{\zeta}} \left( v - \ln(1 - \sqrt{2(1-\rho)}) \zeta \right) - \frac{(1-\rho)}{\rho + \sqrt{2(1-\rho)} \zeta}.$$  \hspace{1cm} (A29)

$$\lim_{\zeta \to 0^+} \left[ \frac{\partial W}{\partial F_1} - \frac{\partial W}{\partial F_2} \right]_{F_1 = F_2 = \zeta} = \lim_{\zeta \to 0^+} \frac{\sqrt{2(1-\rho)}}{2} \frac{v}{\sqrt{\zeta}} - \sqrt{2(1-\rho)} \lim_{\zeta \to 0^+} \frac{\ln(1 - \sqrt{2(1-\rho)} \zeta)}{\sqrt{\zeta}} - \lim_{\zeta \to 0^+} \frac{(1-\rho)}{\rho + \sqrt{2(1-\rho)} \zeta},$$  \hspace{1cm} (A30)

where

$$\lim_{\zeta \to 0^+} \frac{\ln(1 - \sqrt{2(1-\rho)} \zeta)}{\sqrt{\zeta}} = \lim_{\zeta \to 0^+} \frac{-\sqrt{2(1-\rho)}}{1 - \sqrt{2(1-\rho)} \zeta} \frac{1}{2\sqrt{\zeta}} = \lim_{\zeta \to 0^+} \frac{-\sqrt{2(1-\rho)}}{1 - \sqrt{2(1-\rho)} \zeta} = -\sqrt{2(1-\rho)}.$$  \hspace{1cm} (A31)

Therefore,

$$\lim_{\zeta \to 0^+} \left[ \frac{\partial W}{\partial F_1} - \frac{\partial W}{\partial F_2} \right]_{F_1 = F_2 = \zeta} = \lim_{\zeta \to 0^+} \frac{\sqrt{2(1-\rho)}}{2} \frac{v}{\sqrt{\zeta}} + (1-\rho) - \frac{(1-\rho)}{\rho} > 0,$$  \hspace{1cm} (A31)

because the first term goes to positive infinite when $\zeta \to 0^+$. \hspace{1cm} \Box

A.5.2 The proof for the existence of $\bar{v}$

We are looking into the marginal welfare effects of funding the top-down policy when the amounts are close to pushing everyone out of the poverty trap
\[
\frac{\partial W}{\partial F_1} \bigg|_{F_1 \rightarrow F^{all}} \quad \text{and the marginal welfare effects of funding the bottom-up policy}
\]
when the amounts are scarce \( \frac{\partial W}{\partial F_2} \bigg|_{F_2 \rightarrow 0^+} \). In line with the methods in A.5.1,

\[
\frac{\partial W}{\partial F_1} \bigg|_{F_1 \rightarrow F^{all}} = \left( \frac{\beta}{1 - \beta} \ln \frac{A}{B} - \ln \rho \right) \frac{1}{2} \sqrt{\frac{2(1 - \rho)}{(1 - \rho)^2}} = v - \ln \rho
\]

\[
\frac{\partial W}{\partial F_2} \bigg|_{F_2 \rightarrow 0^+} = \frac{1 - \rho}{\rho}.
\]

For \( \frac{\partial W}{\partial F_1} \bigg|_{F_1 \rightarrow F^{all}} > \frac{\partial W}{\partial F_2} \bigg|_{F_2 \rightarrow 0^+} \), we require \( v > \ln \rho + \frac{1}{\rho} - 1 \). The following graph depicts the trajectory of

\[
g(\rho) = \ln \rho + \frac{1}{\rho} - 1.
\]

The function \( g(\rho) = \ln \rho + \frac{1}{\rho} - 1 \) decreases in \( \rho \) when \( \rho \in (0, 1) \). For any finite \( \rho \), we confirm a fixed value of \( g(\rho) \) which we defined as \( \bar{v} \) such that

\[
\frac{\partial W}{\partial F_1} \bigg|_{F_1 \rightarrow F^{all}} > \frac{\partial W}{\partial F_2} \bigg|_{F_2 \rightarrow 0^+} \quad \text{if} \quad v \ > \ \bar{v}.
\]

This suggests that if the current value of alleviating poverty is extremely significant, the top-down policy receives all the available funds. The graphical explanation can be seen in the following

\[
\text{Figure A.2. The function } g(\rho) = \ln \rho + \frac{1}{\rho} - 1.
\]
A.5.3 The proof for the curvature $F_1^p(v)$

To depict the curvature of $F_1^p(v)$, we need to look into the graph of $\partial W/\partial F_1$ and $\partial W/\partial F_2$ with respect to different amounts of funds and different value of lifting poverty. Therefore we check their first-order derivatives, which we have derived above $\left(\partial^2 W/\partial (F_1)^2 \text{ and } \partial^2 W/\partial (F_2)^2\right)$, and the second-order derivatives $\left(\partial^3 W/\partial (F_1)^3 \text{ and } \partial^3 W/\partial (F_2)^3\right)$. Let us write down the expression $\partial^2 W/\partial (F_1)^2$ and $\partial^2 W/\partial (F_2)^2$

$$\frac{\partial^2 W}{\partial (F_1)^2} = \sqrt{2(1-\rho)} \frac{1}{2} \frac{1}{(F_1)^2} \left\{ \frac{1-x_1}{x_1} + \ln(x_1) - \frac{\beta}{1-\beta} \ln \frac{A}{B} \right\},$$

$$\frac{\partial^2 W}{\partial (F_2)^2} = -\frac{(1-\rho)}{(\rho + \sqrt{2(1-\rho)F_2})^2} \frac{2 \sqrt{2(1-\rho)}}{2\sqrt{F_2}}.$$

The second-order derivative of $\partial W/\partial F_1$ with respect to $F_1$ is

$$\frac{\partial^3 W}{\partial (F_1)^3} =$$

$$\sqrt{2(1-\rho)} \frac{1}{4} \frac{1}{(F_1)^2} \frac{1}{x_1} \left(1 - \frac{1}{x_1}\right) \frac{1}{1-\rho} \frac{1}{(F_1)^2} \left\{ \frac{1-x_1}{x_1} + \ln(x_1) - \frac{\beta}{1-\beta} \ln \frac{A}{B} \right\}.$$

With $\frac{\partial^2 W}{\partial (F_1)^2} < 0$, the second term inside the bracket is positive. With $\frac{\partial x_1}{\partial F_1} < 0$ and $0 < x_1 < 1$, the first term inside the bracket is also positive. The two terms jointly make $\frac{\partial^3 W}{\partial (F_1)^3} > 0$.

We rewrite $\frac{\partial^2 W}{\partial (F_2)^2}$ as

$$\frac{\partial^2 W}{\partial (F_2)^2} = -\frac{1}{2} \frac{(1-\rho)}{(x_2)^2} \frac{1}{\sqrt{2(1-\rho)}} \frac{1}{F_2}.$$

The second-order derivative of $\partial W/\partial F_2$ with respect to $F_2$ is

$$\frac{\partial^3 W}{\partial (F_2)^3} = \frac{(1-\rho)}{2} \frac{2(1-\rho)}{(x_2)^3} \frac{1}{\sqrt{2(1-\rho)}} \left\{ \frac{\partial x_2}{\partial F_2} + \frac{1}{2} \frac{1}{(x_2)^2} \frac{1}{(F_2)^2} \right\} > 0$$

for $\frac{\partial x_2}{\partial F_2} > 0$.

With $\frac{\partial^3 W}{\partial (F_1)^3} > 0$ and $\frac{\partial^3 W}{\partial (F_2)^3} > 0$, we can depict the curvature of $\partial W/\partial F_1$ and $\partial W/\partial F_2$ convex in its funds, respectively, in Figures A.3. and A.4.
In Figure A.3, with a lower value of lifting poverty when $0 < v < \bar{v}$ still holds, the funds to prioritize the top-down policy $F_1^p$ decrease but at a decreasing speed that $\Delta F_1^p$ becomes smaller and smaller. A higher value of $v$ raises the flat decreasing part and creates more space for $\partial W/\partial F_1 > \partial W/\partial F_2$, and vice versa.

\[ \square \]

A.6 The proof for Proposition 1

A.6.1 The first part of proposition 1

If the current value of productivity improvement by lifting poverty is substantially high $v \geq \bar{v}$, the marginal welfare effects of funding the top-down policy are always higher than those of funding the bottom-up policy $\partial W/\partial F_1 \geq \partial W/\partial F_2$. The intuition can be seen in the following graph.
With more funds available, $x_1^*$ tends to move down and eventually reaches $\rho = x_2^* = x_1^* < 1$. □

A.6.2 The second part of Proposition 1(2a)

The moderate value of $v$: $0 < v < \bar{v}$ yields that

$$\left. \frac{\partial W}{\partial F_1} \right|_{F_1 = F_1^p(v)} \geq \left. \frac{\partial W}{\partial F_2} \right|_{F_2 \to 0^+}$$

for $F \leq F_1^p(v)$. There does not exist any equalization of marginal welfare between funding the top-down and the bottom-up policy when $F < F_1^p(v)$, which is the part above the black line in Figure A.5.

![Figure A.5. The marginal welfare comparison when $0 < v < \bar{v}$](image)
A.6.3 The proof for the optimality condition in the mixed policy design Proposition 1 (2b)

In line with the proof for Proposition 1 (ii) in Appendix A.6.2, 0 < v < \bar{v} indicates a prioritization for funding the top-down policy. A bottom-up policy can also be funded if the funds are not too scarce F > F_1^p(v). Because by Lemma 3, the marginal welfare effects in the two policies are decreasing with the input funds. It creates the possibility of equalized marginal welfare effects across the two different aiding policies.

Therefore, after prioritizing the individuals who are relatively better off, the remaining funds (F - F_1^p(v)) is shared between the two aiding scenarios. The black line—the equalization of marginal welfare across the two different policies— in Figure A.5 moves down until passing through the point, where ρ < x^*_2 = x^*_1 < 1.

The interior solutions allow both F_1 > 0 and F_2 > 0. The substitution of F_1 > 0 and F_2 > 0 by the optimal boundaries x_1 and x_2 updates the social welfare maximization problem

\[
\text{Max}_{x_2} W = W_1(x_1(x_2)) + W_2(x_2) + W_3(x_1(x_2), x_2)
\]

s.t. \(\frac{(1-x_1)^2}{2(1-\rho)} + \frac{(x_2-\rho)^2}{2(1-\rho)} = F\)  \((A34)\)

where the budget constraint with implicit function theorem can imply a reaction function between the two boundaries

\[
\frac{\partial x_1}{\partial x_2} = \frac{x_2 - \rho}{1 - x_1}. \quad (A35)
\]

We rewrite the target function as the optimization problem equivalent to maximize

\[
W = \left[\ln \bar{y} + \frac{\beta}{1-\beta} \ln A\right] (1 - x_1(x_2)) + \left[\ln (x_2\bar{y}) + \frac{\beta}{1-\beta} \ln B\right] (x_2 - \rho)
\]

\[
+ \int_{x_1(x_2)}^{x_2} \left[\ln(\rho\bar{y}) + \frac{\beta}{1-\beta} \ln B\right] d\rho
\]

by taking optimal x_2. The first-order derivative for each term inside the social welfare function is

\[
\frac{\partial W_1}{\partial x_2} = -\frac{\partial x_1}{\partial x_2} \left[\ln \bar{y} + \frac{\beta}{1-\beta} \ln A\right]
\]

\[
\frac{\partial W_2}{\partial x_2} = \left[\ln (x_2\bar{y}) + \frac{\beta}{1-\beta} \ln B\right] + \frac{x_2 - \rho}{x_2}
\]

\[
\frac{\partial W_3}{\partial x_2} = \frac{\partial x_1}{\partial x_2} \left[\ln x_1\bar{y} + \frac{\beta}{1-\beta} \ln B\right] - \left[\ln x_2\bar{y} + \frac{\beta}{1-\beta} \ln B\right]
\]
where

\[
W_3 = \int_{x_2}^{x_1} \left( \ln \rho + \ln \bar{y} + \frac{\beta}{1 - \beta} \ln B \right) d\rho
\]

\[
= \rho \left( \ln \bar{y} + \frac{\beta}{1 - \beta} \ln B \right) \bigg|_{x_2}^{x_1} + \int_{x_2}^{x_1} \ln \rho \, d\rho
\]

\[
= (x_1 - x_2) \left( \ln \bar{y} + \frac{\beta}{1 - \beta} \ln B \right) + \rho \ln \rho \bigg|_{x_2}^{x_1} - \int_{x_2}^{x_1} 1 \, d\rho
\]

\[
= (x_1 - x_2) \left( \ln \bar{y} + \frac{\beta}{1 - \beta} \ln B \right) + (x_1 \ln x_1 - x_2 \ln x_2) - (x_1 - x_2).
\]

Inside the expression \( \frac{\partial W_3}{\partial x_2} \), the first term on the RHS describes how changes in the upper bound of the middle group \( x_1 \) due to the change of the lower bound \( x_2 \) affect the utility. The second term which is the minimum utility for the group is the comparison benchmark for the welfare changes.

By rearrangement of the first-order derivative results, we reach

\[
- \frac{\partial x_1}{\partial x_2} \frac{\beta}{1 - \beta} \ln \frac{A}{B} + \frac{\partial x_1}{\partial x_2} \ln x_1 + \frac{x_2 - \rho}{x_2} = 0. \quad (A36)
\]

The substitution of \( \frac{\partial x_1}{\partial x_2} = \frac{x_2 - \rho}{1 - x_1} \) into expression (A36) and dividing \( x_2 - \rho \) on both sides yield the expression (17).

**A.6.4 The proof for curvature RNZ for \( F^m \)**

The purple curvature RNZ refers to the amount of funds to finalize any mixed policies

\[
F^m = \frac{(1 - x_1^*)^2}{2(1 - \rho)} + \frac{(x_1^* - \rho)^2}{2(1 - \rho)} \quad (A37)
\]

where \( x_2^* = x_1^* \). On the purple curvature RNZ, the equalization of marginal welfare across the two scenarios ends at \( x_2^* = x_1^* \). On the left-hand side of RNZ, the equalized marginal welfare happens to \( x_2^* < x_1^* \). The non-linear shape of RNZ also characterizes the difference between \( F^m \) and \( F^{all} \). The difference in the required amount of funds is

\[
F^{all} - F^m = \frac{(1 - \rho)}{2} - \left( \frac{(1 - x_1^*)^2}{2(1 - \rho)} + \frac{(x_1^* - \rho)^2}{2(1 - \rho)} \right). \quad (A38)
\]

The edge values \( (F^{all} - F^m)|_{x_1^* \to 1} = 0 \) and \( (F^{all} - F^m)|_{x_1^* \to \rho} = 0 \) coincide with the required amount of funds to finalize a pure bottom-up or a pure top-down policy, respectively. The first-order derivative of \( F^{all} - F^m \) with respect
to \(x_1^*\) is
\[
\frac{\partial (F^{\text{all}} - F^m)}{\partial x_1^*} = -\frac{4x_1^* - 2(\rho + 1)}{2(1 - \rho)}.
\] (A39)

Accordingly, \(\frac{\partial (F^{\text{all}} - F^m)}{\partial x_1^*} > 0\) if \(x_1^* < \frac{\rho}{2}\) and \(\frac{\partial (F^{\text{all}} - F^m)}{\partial x_1^*} < 0\) if \(x_1^* > \frac{\rho}{2}\).

So \(F^{\text{all}} - F^m\) grows in \(x_1^*\) before \(\frac{\rho}{2}\), and reaches its peak at \(\frac{\rho}{2}\). Then it goes down when \(x_1^*\) moves to 1. The curvature can be seen in Figure A.6.

\[
\text{Figure A.6. The curvature of RNZ in Figure 2}
\]

Let us start the analysis with \(v = \bar{v}\) at point \(R\) in Figure 2 where \(x_1^* = \rho\). If \(v\) goes down, \(x_1^*\) goes up as \(\frac{\partial v}{\partial x_1^*} = \frac{1}{x_1} - \frac{1}{x_2} < 0\) by Corollary 1. The increase happens to a higher speed forward \(\frac{\rho}{2}\) due to \(\frac{\partial^2 v}{\partial (x_1^*)^2} = -\frac{1}{(x_1^*)^2} < 0\), which derived from \(\frac{\partial v}{\partial x_1^*} = \frac{1}{x_1} - \frac{1}{x_2}\).

\[\square\]

Appendix B Supplementary materials to understand top-down prioritization

We will examine a specific case to illustrate why it is more effective to start pushing individuals out of the poverty trap from the top. This connects to the argument for why a top-down policy is an exclusively optimal approach for lifting poverty completely. In a top-down policy, aid is prioritized for individuals from the top to the lower bound \(x_1\). The objective is to determine whether the welfare effects would be greater if the funds \(F_{a,a-d}\) were allocated to the segment of the population. By setting \(a = 1\) and \(b = x_1\) in equations
(A.11b) and (A.12b), we can calculate the corresponding welfare increment and budget constraint for the top-down policy
\[
\Delta W_1 = \frac{1}{1 - \alpha \beta} \frac{1}{1 - \rho} [(1 - x_1) (v + 1) + x_1 \ln x_1]
\]
(B1)
and
\[
(x_1 - 1)^2 = d(2 - 2a + d).
\]
(B2)
We can therefore get
\[
1 - x_1 = \sqrt{d(2 - 2a + d)} \quad \text{and} \quad x_1 = 1 - \sqrt{d(2 - 2a + d)}.
\]
(B3)
With (B2), \(F_{a,a-d}\) can therefore generate the following welfare effects in the top-down policy
\[
\Delta W_1 = \frac{1}{1 - \alpha \beta} \frac{1}{1 - \rho} \times [\sqrt{d(2 - 2a + d)}(v + 1) + (1 - \sqrt{d(2 - 2a + d)}) \ln(1 - \sqrt{d(2 - 2a + d)})].
\]
(B4)
The comparison in welfare increment between funding the top-down policy and an arbitrary policy that lifts poverty for someone else further below is
\[
\Delta W_1 - \Delta W_{1,a,a-d} = [\sqrt{d(2 - 2a + d)} - d](v + 1) + a \ln a
+ (1 - \sqrt{d(2 - 2a + d)}) \ln(1 - \sqrt{d(2 - 2a + d)})
- (a - d) \ln(a - d)
\]
(B5)
where I drop \(\frac{1}{1 - \alpha \beta} \frac{1}{1 - \rho}\) for convenience, as it does not affect the sign of the expression. The derivative of (B5) for \(d\) is
\[
\frac{\partial}{\partial d} \left( \Delta W_1 - \Delta W_{1,a,a-d} \right) = v \left[\frac{1 - a + d}{\sqrt{d(2 - 2a + d)}} - 1\right]
- \frac{1 - a + d}{\sqrt{d(2 - 2a + d)}} \ln(1 - \sqrt{d(2 - 2a + d)}) + \ln(a - d).
\]
(B6)
We can get
\[
\lim_{d \to 0^+} \frac{\partial}{\partial d} \left( \Delta W_1 - \Delta W_{1,a,a-d} \right) > 0
\]
(B7)
for \(v > 0\), because
(i) the first term inside the bracket goes to positive infinite when \(d\) goes to a small positive number;
(ii) the summation of the rest terms \(1 - a + d + \ln(a - d)\) converges to \(1 - a + \ln a\) which is a finite negative number due to \(0 < \rho \leq a < 1\), where we use \(\ln(1 - \sqrt{d(2 - 2a + d)}) \sim -\sqrt{d(2 - 2a + d)}\) when \(d \to 0^+\).
It indicates that when $d$ is a small positive number, $(\Delta W_1 - \Delta W_{1,a,a-d})$ increases in $d$, leading to $\left(\Delta W_1 - \Delta W_{1,a,a-d}\right)|_{d>0} > \left(\Delta W_1 - \Delta W_{1,a,a-d}\right)|_{d=0} = 0$. Therefore, we expect a small positive number $d = \varepsilon > 0$ such that $\Delta W_1 - \Delta W_{1,a,a-d} > 0$. $\varepsilon$ represents that the arbitrary interval we mentioned above is small, echoing a small discrete amount of funds under consideration. In this context, we show that when we have small funds, it is always preferable to fund a group close to the income threshold $\bar{y}$, which is normalized to one over any other group.
## Economic Studies

1987:1 Haraldson, Marty. To Care and To Cure. A linear programming approach to national health planning in developing countries. 98 pp.


1989:3 Choe, Byung-Tae. Some Notes on Utility Functions Demand and Aggregation. 39 pp.


<table>
<thead>
<tr>
<th></th>
<th>Author</th>
<th>Title</th>
<th>Year</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>Žamac, Jovan</td>
<td>Education, Pensions, and Demography.</td>
<td>2007</td>
<td>105</td>
</tr>
<tr>
<td>101</td>
<td>Post, Erik</td>
<td>Macroeconomic Uncertainty and Exchange Rate Policy.</td>
<td>2007</td>
<td>129</td>
</tr>
<tr>
<td>102</td>
<td>Nordberg, Mikael</td>
<td>Allies Yet Rivals: Input Joint Ventures and Their Competitive Effects.</td>
<td>2007</td>
<td>122</td>
</tr>
<tr>
<td>103</td>
<td>Johansson, Fredrik</td>
<td>Essays on Measurement Error and Nonresponse.</td>
<td>2007</td>
<td>130</td>
</tr>
<tr>
<td>104</td>
<td>Haraldsson, Mattias</td>
<td>Essays on Transport Economics.</td>
<td>2007</td>
<td>104</td>
</tr>
<tr>
<td>105</td>
<td>Edmark, Karin</td>
<td>Strategic Interactions among Swedish Local Governments.</td>
<td>2007</td>
<td>141</td>
</tr>
<tr>
<td>106</td>
<td>Oreland, Carl</td>
<td>Family Control in Swedish Public Companies. Implications for Firm Performance, Dividends and CEO Cash Compensation.</td>
<td>2007</td>
<td>121</td>
</tr>
<tr>
<td>107</td>
<td>Andersson, Christian</td>
<td>Teachers and Student Outcomes: Evidence using Swedish Data.</td>
<td>2007</td>
<td>154</td>
</tr>
<tr>
<td>108</td>
<td>Kjellberg, David</td>
<td>Expectations, Uncertainty, and Monetary Policy.</td>
<td>2007</td>
<td>132</td>
</tr>
<tr>
<td>110</td>
<td>Selin, Håkan</td>
<td>Four Empirical Essays on Responses to Income Taxation.</td>
<td>2008</td>
<td>133</td>
</tr>
<tr>
<td>111</td>
<td>Lindahl, Erica</td>
<td>Empirical studies of public policies within the primary school and the sickness insurance.</td>
<td>2008</td>
<td>143</td>
</tr>
<tr>
<td>112</td>
<td>Liang, Che-Yuan</td>
<td>Essays in Political Economics and Public Finance.</td>
<td>2008</td>
<td>125</td>
</tr>
<tr>
<td>113</td>
<td>Elinder, Mikael</td>
<td>Essays on Economic Voting, Cognitive Dissonance, and Trust.</td>
<td>2008</td>
<td>120</td>
</tr>
<tr>
<td>114</td>
<td>Grönqvist, Hans</td>
<td>Essays in Labor and Demographic Economics.</td>
<td>2009</td>
<td>120</td>
</tr>
<tr>
<td>115</td>
<td>Bengtsson, Niklas</td>
<td>Essays in Development and Labor Economics.</td>
<td>2009</td>
<td>93</td>
</tr>
<tr>
<td>117</td>
<td>Liu, Qian</td>
<td>Essays on Labor Economics: Education, Employment, and Gender.</td>
<td>2009</td>
<td>133</td>
</tr>
<tr>
<td>118</td>
<td>Glans, Erik</td>
<td>Pension reforms and retirement behaviour.</td>
<td>2009</td>
<td>126</td>
</tr>
<tr>
<td>119</td>
<td>Douhan, Robin</td>
<td>Development, Education and Entrepreneurship.</td>
<td>2009</td>
<td></td>
</tr>
</tbody>
</table>


156 Gabriella Chirico Willstedt. Demand, Competition and Redistribution in Swedish Dental Care. 2015. 119 pp.


