Kobe University Repository : Kernel

PDF issue: 2026-11-27

The Theoretical Analysis of Economic Behaviors with Psychological Factors

熊代, 和樹

(Degree)
博士（経済学）

(Date of Degree)
2017-03-25

(Date of Publication)
2019-03-25

(Resource Type)
doctoral thesis

(Report Number)
甲第6821号

(URL)
https://hdl.handle.net/20.500.14094/D1006821

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。

博士論文

平成 28 年 12 月

神戸大学大学院経済学研究科

経済学専攻

指導教員　　宮川　栄一

熊代　和樹

博士論文

The Theoretical Analysis of Economic Behaviors with Psychological Factors
心理的要因の下での経済行動に関する理論的分析

平成 28 年 12 月

神戸大学大学院経済学研究科
経済学専攻
指導教員　　宮川　栄一
熊代　和樹

Contents

1 Preface **1**

2 Optimal Funded Pension for Consumers with Heterogeneous Self-Control **2**
2.1 Introduction . . . 2
2.2 Model . . . 4
2.2.1 Budget Constraint . . . 4
2.2.2 Preference . . . 5
2.2.3 Consumption . . . 6
2.2.4 Government . . . 7
2.3 Benchmark . . . 7
2.3.1 Identical-type consumers . . . 7
2.3.2 Two types of consumers . . . 14
2.4 Generalization . . . 20
2.4.1 Finite types . . . 20
2.4.2 Continuous types . . . 22
2.5 Discussion . . . 24
2.5.1 The effect of a borrowing constraint . . . 24
2.5.2 Income diversity . . . 24
2.6 Conclusion . . . 29

3 Consumer Inertia and Competitive Behavior between Firms **31**
3.1 Introduction . . . 31
3.2 Model . . . 32
3.3 Equilibria . . . 33
3.3.1 Separation over qualities . . . 33
3.3.2 Separation over costs . . . 34
3.4 Conclusion . . . 42

4 The Economic Analysis of Psychological Reactance **43**
4.1 Introduction . . . 43
4.2 Discrete Precision Model . . . 44
4.2.1 Equilibrium . . . 45
4.2.2 An alternative equilibrium . . . 54
4.3 Discussion . . . 55
4.3.1 Independence from Self-Evaluation . . . 55
4.4 Conclusion . . . 57

Appendices **58**

Chapter 1

Preface

Over the past few decades, the effects of various psychological factors to decision makings have been studied in behavioral economics. However, the field is yet developing. The purposes of this thesis are to contribute the development of theoretical aspect of behavioral economics and to propose new notable phenomenon in decision makings.

The construction of this thesis is as follows. In chapter 2, we study the consumption-saving decision by consumers facing to temptation that leads them to over-consumption. Consumers can resist the temptation but it is psychologically costly. Under the assumption that consumption utility is log, we propose a funded pension scheme as a commitment device, which maximizes social welfare, which takes account of self-control cost. In chapter 3, we study a price competition between two firms, where the information about the qualities are asymmetric among consumers who feels inertia to their default product. We focus to the endogenous information acquisition of consumers and investigate some separation equiribria. In chapter 4, we propose a formal model that can describe a phenomenon called psychological reactance. We show that the way of the decision maker's evaluation of reputation is important factor.

Acknowledgement

I would like to thank Eiichi Miyagawa for his invaluable comments and encouragements. In addition, I am grateful to Kazuhiko Hashimoto, Yasuyuki Miyahara, Shuhei Morimoto, Takeshi Murooka, Noritsugu Nakanishi, Takashi Shimizu, Tomoya Tajika, Yoshikatsu Tatamitani, Hirofumi Yamamura and participants of seminars at Kobe university, Japan Economic Association meeting at Doshisha university for their many constructive comments. This matter in chapter 2 is the extension of author's master thesis, permitted by graduate school of economics, Kobe university in 2013 and it was selected as one of the three candidates of 18th Osaka University Institute of Social and Economic Research/Moriguchi Prize. I thank to judges and two anonymous referees for the prize. All errors remain my own.

Chapter 2

Optimal Funded Pension for Consumers with Heterogeneous Self-Control

2.1 Introduction

We consider a public funded pension scheme. Social security policies including a funded pension have some objectives. One of them is income redistribution. This is common objective in almost all social security policies such as funded pensions, social aid by taxation and pay-as-you-go pensions. Other important role of social securities is that as forced saving devices, which is specific to funded pensions. This is because funded pensions assure subscribers that they can receive retirement incomes at least their contribution. However, it is well known that for consumers in standard economic models, funded pensions do not improve social welfare (e.g. Samuelson, 1975). For example, suppose that a consumer found that his optimal amount of saving is \$1000 per month. When there is no pension policy, he privately saves \$1000. When a pension is introduced and the premium that he pays is \$500 per month, he pays it and privately saves \$500 because his optimal saving is \$1000. Since his total amount of saving does not change between before and after the introducing of the pension, so does his lifetime utility. Of course, if the premium is larger than \$1000, he goes into oversaving that decreases his lifetime utility. Thus the pension does not improve social welfare.

One of factors to justify funded pensions is temptation that leads consumers to overconsumption (Diamond, 1977). Even if a consumer knows how much to save his income, he often cannot help but overspending more than the amount when he has disposable income. Once he faces this temptation, he feels psychological cost to self-control his wasteful spending. So it is ideal for him not to face temptation by a valid way for commitment. This is why we need funded pensions as a forced saving policy.

However, in the context of temptation, most of pension schemes presently used are not suitable. It is natural that the degrees of temptation are not the same for all consumers. While there are consumers that spend much of their income as soon as they earns them and regret their myopic behavior, there are also consumers that can spend their income farsightedly. Since pension schemes affect the behavior of consumers, they should be designed with thinking of this point. However, in present pension schemes, a premium that each consumer pays is determined mainly by his income level, not by his degree of temptation. Our purpose is to design a pension scheme that maximizes social welfare considering the difference in the degree of temptation.

In our model, the government proposes a pension schedule to consumers, where pension schedule is a list of pension plans, a pair of pension premium and its payout. Each consumer choose a plan from the schedule and pay the premium according to the plan before his consumption, that is, before he faces to temptation. This is why the pension works as forced saving [1]. However, we allow the consumer to borrow money by collateralizing his future payout.

Some literature studied policies as a way to resist temptation. It was shown that funded pensions improve social welfare (Gul and Pesendorfer, 2004). This is because, as stated above, they enable consumers to avoid temptation with less self-control cost. In addition, pay-as-you-go pension is also able to improve social welfare if temptation is sufficiently large (Kumru and Thanopoulos, 2008). Other way to relieve self-control cost is to make consumption relatively unattractive. This works because the cost arises from the gap between the attractiveness of normatively desirable alternatives and that of tempting alternative. Krusell et al. (2010) shows that subsidies for savings improve social welfare for this reason. Though these studies are important, their research objects are economies with consumers whose degrees of temptation are the same for all of them. On the other hand, our model can analyze the situation in which the degrees are different for different consumers.

Galperti (2015) investigated an optimal contract as a commitment device. The agent is either consistent or inconsistent, which is randomly determined and is private information of the agent. If he is inconsistent, he values his consumption in the second period smaller than that in the first period. The characteristic of the model is that it takes in the preference for flexibility. It is an important stuff especially when the problem has long period. On the other hand, it is assumed that no third party can offer the agent contracts that might interfere with the provider's devices (Galperti, 2015:p.1431). It is the advantage of our model that agents (consumers) are allowed to make debt after they choose their devices.

Since we need a model that describes decisions with temptation and self-control, we apply the model proposed by Gul and Pesendorfer (2001), henceforth GP. They considered two kinds of preferences over alternatives, normative preferences and temptation preferences, which are represented by functions u and v, respectively. Using u and v, the valuation of menu (a set of alternatives) M is defined by

$$W(M) \equiv \max_{x \in M} [u(x) + v(x)] - \max_{y \in M} v(y). \tag{2.1}$$

GP assumes that, for a given menu M, a decision maker chooses an alternative $x \in M$ that maximizes $u(x) + v(x)$. The intuition is that he chooses a compromise between normative desirability and temptation desirability after facing temptation. Let $x^* \in M$ be an alternative that maximizes $u(x) + v(x)$. Then (2.1) is rewritten as

$$W(M) = u(x^*) - \left\{ \max_{y \in M} v(y) - v(x^*) \right\}.$$

The term $\max_{y \in M} v(y) - v(x^*)$ represents self-control cost. For instance, suppose that a consumer faces a menu $M = \{s, h\}$, where s and h represents salad and hamburger, respectively. And suppose that he wants to diet but he likes meat. In GP model, u corresponds to a desirability of health and v does to a desirability of having what he likes. Thus $u(s) > u(h)$ and $v(h) > u(s)$ holds. If $u(s) + v(s) > u(h) + v(h)$, he chooses salad for

[1] We focus on sophisticated consumers, that is, each consumer knows that he will face to temptation at the decision of consumption.

his lunch. Note that salad does not maximize v, thus he gives up satisfaction of temptation as much as $v(h) - v(s)$. GP defined self-control cost by this difference. Generalizing this concept, we have the expression above.

Menus correspond to budget sets in consumption choice problems. If a pension premium can shrink the budget sets, it works as a commitment device and decreases self-control cost. Formally, assume that a pension premium makes a menu M shrink to $M' \subset M$. Then we have $\max_{y \in M'} v(y) \leqslant \max_{y \in M} v(y)$, thus $\max_{y \in M'} v(y) - v(x^*) \leqslant \max_{y \in M} v(y) - v(x^*)$ follows. Especially, if the inequality holds strictly, the self-control cost decreases strictly.

As an alternative approach to analyze myopic behaviors, we may use (quasi) hyperbolic discounting model provided by Laibson (1997). Actually, some literature on economic policies employ this approach (e.g. Roeder, 2014). The model describes the situation in which preferences are time inconsistent. However, the model cannot describe self-control cost explicitly. Since we want to treat effects of temptation and self-control separately, we employ GP model that is proper for our objective.

Our main contribution is presenting a concrete way to design the optimal pension scheme under the assumption that normative utility is $\log(c)$. This scheme has some interesting character. First, the optimal schedule consists of plans that return just as managed premium paid by a consumer that chose the plan. This implies that no income redistribution is made. Second, the consumer with smaller self-control cost chooses a smaller premium in an equilibrium. If he pays the large amount of premium, he faces larger temptation of borrowing since the premium brings him a large borrowing facility. So he chooses a small premium to avoid it. If consumers cannot borrow money, the optimal schedule has unique plan whose premium is equal to the amount of saving that consumers choose when there is no temptation and no pension scheme. Third, the optimal schedule does not depend on the information of types such as its distribution and even what types there are. This means that the government does not need to know these information. This is a beneficial feature from a practical viewpoint.

Section 2 introduces notations and assumptions used in the study and looks at the consumption-saving decision of a consumer with a self-control preference. Section 3 studies identical-type and two-type economies as benchmarks and shows the monotonicity of an optimal pension. In section 4, we generalize the results in section 3 to finitely many types and continuous-types model. We show that the results are robust for generalization. In section 5, we discuss the effect of a borrowing constraint and extend the model to analyze income diversity.

2.2 Model

The only difference between standard models of pension and our model is that consumers have preference with temptation and self-control. There are consumers and an government. At first, the government offers a set of pairs of pension payout and pension premium to each consumer before his consumption decision in working age. Then the consumers choose one of pairs from the set. After a payment of premium the consumer has chosen, he decides how much to consume in working age.

2.2.1 Budget Constraint

We standardize the population of consumer to be 1. Each consumer is endowed with an identical income of $I \in \mathbb{R}_{++}$. Let $R \in \mathbb{R}_{+}$ and $P \in [0, I]$ be a pension payout and a

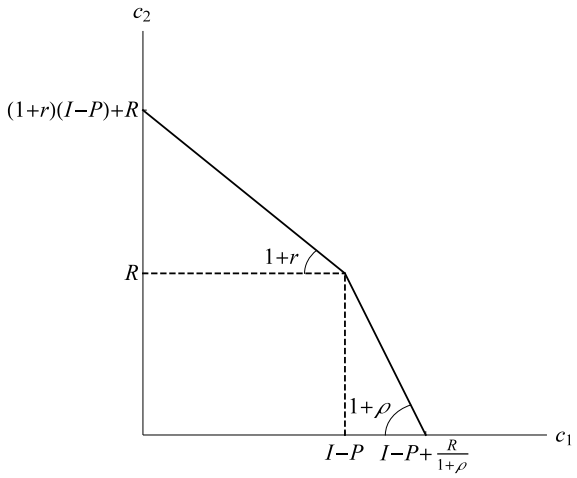

Figure 2.1: Budget constraint

pension premium, respectively. We assume the upper bound of P to be I to rule out situations in which consumers borrow money to pay the premiums. We call a pair of R and P *pension plan*. For the simplicity of notation, we define the set of possible pension plans as $T \equiv \mathbb{R}_+ \times [0,1]$. A *pension schedule* is a set of pension plans.

In period 0, each consumer chooses a pension plan $\tau \in T$, from a pension schedule $S \subseteq T$. At the same time, he has to pay the government the pension premium he has chosen. In period 1, he decides the amounts of consumption $c_1 \in \mathbb{R}_+$ and saving $I-P-c_1$ in working age. We assume that consumers can also borrow money in period 1 with putting up their pension income as collateral. Thus, if $I-P-c_1$ is strictly greater (less) than 0, it represents the amount of saving (borrowing). Let r, $\rho \in \mathbb{R}_{++}$ be the interest rate for saving and borrowing, respectively. Thus if a consumer chose $\tau = (R,P)$, he can borrow up to $\frac{R}{1+\rho}$. About interest rates, we assume that $\rho > r > 0$. This assumption is natural from a practical standpoint. In period 2, he receives a pension income R and decides consumption in old age $c_2 \in \mathbb{R}_+$.

The budget set for consumers having chosen $\tau = (R,P)$ is summarized as follows.

$$B(\tau) \equiv \left\{ (c_1,c_2) \in \mathbb{R}_+^2 : \begin{array}{l} c_1 \leqslant I - P + \frac{R}{1+\rho}, \\ I-P-c_1 \geqslant 0 \Rightarrow c_2 \leqslant (1+r)(I-P-c_1)+R, \\ I-P-c_1 < 0 \Rightarrow c_2 \leqslant (1+\rho)(I-P-c_1)+R \end{array} \right\}.$$

Denote the set of all possible $B(\tau)$ by $\mathscr{B}$. Figure shows this budget constraint. As you see, the budget line kinks at $(c_1,c_2) = (I-P,R)$.

2.2.2 Preference

Consumers have self-control preference introduced by GP. We begin with defining two kinds of utility functions representing preference over consumption. One is normative utility function $u\colon \mathbb{R}_+ \to \mathbb{R}$ and the other is temptation utility function $v\colon \mathbb{R}_+ \to \mathbb{R}$. To express the temptation, v is specialized as $v(\cdot,\cdot) \equiv \lambda u(\cdot,\cdot)$, where $\lambda \in \Lambda \subseteq \mathbb{R}_+$ denotes the strength of the temptation. This form of temptation utility is also used in Gul and Pesendorfer (2004). We assume the value λ differs across agent and is private information. For any $\lambda \in \Lambda$, let n_λ be a proportion of consumer, that is, n_λ satisfies $0 \leqslant n_\lambda \leqslant 1$ and $\sum_{\lambda\in\Lambda} n_\lambda = 1$. In section 4.2, we consider continuous Λ. Then the proportion is represented by a distribution function $F\colon \Lambda \to [0,1]$ with density function $f\colon \Lambda \to \mathbb{R}_+$.

We impose a few assumptions about the normative utility function.

Assumption 2.2.1. *u is twice continuously differentiable.*

Assumption 2.2.2. *u is strictly concave and satisfies $u'(c) > 0$ and $u''(c) < 0$.*

Assumption 2.2.3. $\lim_{c\to 0} u'(c) = \infty$.

For example, these assumption is satisfied if normative utility functions is $u(c) = \log(c)$ or $u(c) = c^\alpha$ with $\alpha \in (0, 1)$. In the latter half of this chapter, we assume $u(c) \equiv \log(c)$ to derive sharper results. Applying the utility function by Gul and Pesendorfer (2001), the preference over budget sets is represented by the function $W\colon \mathscr{B} \times \mathbb{R}_+ \to \mathbb{R}$ satisfies that

$$\begin{aligned} W(B(\tau);\lambda) &= \max_{(c_1,c_2)\in B(\tau)} \left[u(c_1) + \delta u(c_2) + \lambda u(c_1)\right] - \max_{(c_1,c_2)\in B(\tau)} \lambda u(c_1) \\ &= \max_{(c_1,c_2)\in B(\tau)} \left[(1+\lambda)u(c_1) + \delta u(c_2)\right] - \max_{(c_1,c_2)\in B(\tau)} \lambda u(c_1), \end{aligned}$$

where $\delta \in (0, 1)$ is a discount factor.

Using assumption 2.2.2, we have the following claim.

Claim 2.2.1. $\max_{(c_1,c_2)\in B(\tau)} \lambda u(c_1) = \lambda u\left(I - P + \frac{R}{1+\rho}\right)$.

This is obvious by $u'(c) > 0$ (assumption 2.2.2) and the constraint for c_1 is $c_1 \leqslant I - P + \frac{R}{1+\rho}$.

2.2.3 Consumption

In this section, we consider the consumption problem contained in $W(B(\tau);\lambda)$, that is, $\max_{(c_1,c_2)\in B(\tau)} \left[(1+\lambda)u(c_1)+\delta u(c_2)\right]$. When the consumer of type $\lambda \in \Lambda$ chooses a pension plan $\tau \in T$, let his optimal consumption in period t be $c_t^\lambda(\tau)$.

There can be the three types of consumption: positive saving, zero saving (or balanced) and negative saving (or borrowing). The type of optimal consumption depends on the pension plan that a consumer has chosen. Intuitively, since a large premium brings too little disposable income at a working age, it may be necessary for him to borrow to achieve his desired level of consumption. Formally, we separate T as

$$\begin{aligned} T_s^\lambda &\equiv \{\tau \in T\colon c_1^\lambda(\tau) < I - P\} \\ T_b^\lambda &\equiv \{\tau \in T\colon c_1^\lambda(\tau) = I - P\} \\ T_d^\lambda &\equiv \{\tau \in T\colon c_1^\lambda(\tau) > I - P\}. \end{aligned}$$

τ_s^λ, τ_b^λ and τ_d^λ denote the typical elements of T_s^λ, T_b^λ and T_d^λ, respectively. λ in superscript is omitted if there is no threat of confusion. Note that if $\tau \in T_b$, $c_2^\lambda(\tau) = R$, since $u'(c) > 0$ and income in period 2 is only pension income R.

We have the necessary conditions for optimality in different form for τ_s and τ_d since the budget constraint in period 2 differs. In this consumption optimization problem, an objective function is $U(c_1, c_2) \equiv (1 + \lambda)u(c_1) + \delta u(c_2)$. So (the absolute value of) the marginal rate of substitution of c_1 for c_2 is

$$\frac{\frac{\partial U(c_1,c_2)}{\partial c_1}}{\frac{\partial U(c_1,c_2)}{\partial c_2}} = \frac{(1+\lambda)u'(c_1)}{\delta u'(c_2)}.$$

If $\tau \in T_s$, the budget line has the slope of $1+r$. So the necessary condition for τ_s is [2]

$$\begin{aligned}&\frac{(1+\lambda)u'(c_1(\tau_s))}{\delta u'(c_2(\tau_s))} = 1+r\\ \Leftrightarrow &(1+\lambda)u'(c_1(\tau_s)) = (1+r)\delta u'(c_2(\tau_s)).\end{aligned} \tag{2.2}$$

On the other hand, if $\tau \in T_d$, the budget line has a slope of $1+\rho$. So the necessary condition for τ_d is

$$\begin{aligned}&\frac{(1+\lambda)u'(c_1(\tau_d))}{\delta u'(c_2(\tau_d))} = 1+\rho\\ \Leftrightarrow &(1+\lambda)u'(c_1(\tau_d)) = (1+\rho)\delta u'(c_2(\tau_d)).\end{aligned} \tag{2.3}$$

Since a budget line kinks at the consumption vector for τ_b, the slope of the line there is not defined. Thus we cannot derive a condition as in other cases. However, we have at least the following inequality:

$$\begin{aligned}&1+r < \frac{(1+\lambda)u'(c_1(\tau_b))}{\delta u'(c_2(\tau_b))} < 1+\rho\\ \Leftrightarrow &(1+r)\delta u'(c_2(\tau_b)) < (1+\lambda)u'(c_1(\tau_b)) < (1+\rho)\delta u'(c_2(\tau_b)).\end{aligned}$$

2.2.4 Government

The government offers consumers a pension schedule $S \equiv \{\tau_\lambda \in T : \lambda \in \Lambda\} \subset T$, a set of pension plans. The objective of the government is to maximize the (expected) aggregated welfare of consumers while controlling S to satisfy the following conditions:

$$W(B(\tau_\lambda);\lambda) \geqslant W(B(\tau');\lambda),\ \forall\lambda \in \Lambda,\ \forall\tau' \in S \tag{IC}$$

$$(1+r)\sum_{\lambda\in\Lambda} n_\lambda P_\lambda \geqslant \sum_{\lambda\in\Lambda} n_\lambda R_\lambda. \tag{FB}$$

The first condition states that the consumers of type $\lambda \in \Lambda$ will choose a plan $\tau(\lambda)$ at their own initiative; in other words, this is the condition for not giving consumers any incentive to report their types untruthfully. Note that this is not necessary when types are public information since then government can force any plan on consumers.

The second condition states that S is feasible. We assume that pension management interest rate r is the same as that of the private saving. This means that there is no difference between the government and private banks in the ability to manage assets.

We also assume that we can ignore an individual rationality condition since the pension is managed by the government, which has the power of consumers' participation.

2.3 Benchmark

2.3.1 Identical-type consumers

In this section, we consider the simple case in which all consumers have identical type, that is, $\Lambda = \{\lambda\}$ and this is common knowledge. Then the condition (IC) can be ignored. Moreover, since λ is identical, the condition (FB) is rewritten as private condition, that is, $(1+r)P(\lambda) \geqslant R(\lambda)$.

[2] There is not any corner solution because of assumption 2.2.3.

Consider the individual welfare function $W(\tau;\lambda)$. Define $\sigma_\lambda : T \to \mathbb{R}$ to be the marginal rate of substitution of R for P on $\tau \in T$, that is,

$$\sigma_\lambda(\tau) = -\frac{\frac{\partial W(\tau;\lambda)}{\partial R}}{\frac{\partial W(\tau;\lambda)}{\partial P}}.$$

Since we cannot represent the optimal consumption in a general form, we have to consider the optimal pension plan separately. However, the following lemma makes it easy to analyze.

Lemma 2.3.1. *For any individual welfare level, an indifference curve corresponding to the level in R-P plane satisfies following properties.*

(i) $\sigma_\lambda(\tau_s)$ *is greater than or equal to* $\frac{1}{1+r}$*, where equality holds if and only if* $\lambda = 0$*.*

(ii) $\sigma_\lambda(\tau_d)$ *is equal to* $\frac{1}{1+\rho} < \frac{1}{1+r}$*.*

(iii) Every indifference curve is differentiable.

Proof of property (i). Consider arbitrary $\tau \in T_s$. Then we have

$$W(B(\tau);\lambda) = (1+\lambda)u(c_1(\tau)) + \delta u\big(c_2(\tau)\big) - \lambda u\left(I - P + \frac{R}{1+\rho}\right).$$

The marginal welfare of R is calculated as follows:

$$\begin{aligned}\frac{\partial W(B(\tau);\lambda)}{\partial R} =& (1+\lambda)u'(c_1(\tau))\frac{\partial c_1(\tau)}{\partial R} + \delta u'\big(c_2(\tau)\big)\big)\frac{\partial c_2(\tau)}{\partial R} \\ & - \frac{\lambda}{1+\rho}u'\left(I - P + \frac{R}{1+\rho}\right)\end{aligned}$$

Since $c_2(\tau_s) = (1+r)(I - P - c_1(\tau_s)) + R$, $\frac{\partial c_2(\tau_s)}{\partial R} = -(1+r)\frac{\partial c_1(\tau_s)}{\partial R} + 1$. Hence we have

$$\begin{aligned}\frac{\partial W(B(\tau);\lambda)}{\partial R} &= \Big\{(1+\lambda)u'(c_1(\tau)) - (1+r)\delta u'(c_2(\tau))\Big\}\frac{\partial c_1(\tau)}{\partial R} \\ &\quad + \delta u'(c_2(\tau)) - \frac{\lambda}{1+\rho}u'\left(I - P + \frac{R}{1+\rho}\right) \\ &= \delta u'(c_2(\tau)) - \frac{\lambda}{1+\rho}u'\left(I - P + \frac{R}{1+\rho}\right).\end{aligned}$$

The second equality follows by the first order condition for optimal consumption, that is, $(1+\lambda)u'(c_1(\tau)) = (1+r)\delta u'(c_2(\tau))$. Moreover, the last line can be rewritten as

$$\frac{1+\lambda}{1+r}u'(c_1(\tau)) - \frac{\lambda}{1+\rho}u'\left(I - P + \frac{R}{1+\rho}\right),$$

and this is strictly positive. To see this, use assumption 2.2.2. Since $c_1(\tau) \leqslant I - P + \frac{R}{1+\rho}$ and $u''(c) < 0$, it holds that $u'(c_1(\tau)) \geqslant u'\left(I - P + \frac{R}{1+\rho}\right)$. By $\rho > r$ and $\lambda \geqslant 0$, we have $\frac{\partial W(\tau;\lambda)}{\partial R} > 0$.

Next, the marginal welfare of P is

$$\begin{aligned}\frac{\partial W(B(\tau);\lambda)}{\partial P} &= (1+\lambda)u'(c_1(\tau))\frac{\partial c_1(\tau)}{\partial P} - \delta u'(c_2(\tau))(1+r)\left(1+\frac{\partial c_1(\tau)}{\partial P}\right) \\ &\quad + \lambda u'\left(I-P+\frac{R}{1+\rho}\right) \\ &= \left\{(1+\lambda)u'(c_1(\tau)) - (1+r)\delta u'(c_2(\tau))\right\}\frac{\partial c_1(\tau)}{\partial P} \\ &\quad - (1+r)\delta u'(c_2(\tau)) + \lambda u'\left(I-P+\frac{R}{1+\rho}\right) \\ &= -(1+\lambda)u'(c_1(\tau)) + \lambda u'\left(I-P+\frac{R}{1+\rho}\right).\end{aligned}$$

By the same ways as in the previous paragraph, the equalities follow and the last line is strictly negative.

Thus, $\sigma_\lambda(\tau)$ for $\tau \in T_s$ is

$$\sigma_\lambda(\tau) = -\frac{\frac{\partial W(B(\tau);\lambda)}{\partial R}}{\frac{\partial W(B(\tau);\lambda)}{\partial P}} = \frac{-\delta u'(c_2(\tau)) + \frac{\lambda}{1+\rho}u'\left(I-P+\frac{R}{1+\rho}\right)}{-(1+\lambda)u'(c_1(\tau)) + \lambda u'\left(I-P+\frac{R}{1+\rho}\right)}.$$

We show that this is greater than or equal to $\frac{1}{1+r}$. Since $\rho > r$, we have

$$\begin{aligned}&-(1+\lambda)u'(c_1(\tau)) + \frac{1+r}{1+\rho}\lambda u'\left(I-P+\frac{R}{1+\rho}\right) \\ &\qquad \leqslant -(1+\lambda)u'(c_1(\tau)) + \lambda u'\left(I-P+\frac{R}{1+\rho}\right),\end{aligned}$$

where equality holds if $\lambda = 0$. Furthermore, since $u'\left(I-P+\frac{R}{1+\rho}\right) > 0$ it holds only if $\lambda = 0$. Using the first order condition,

$$\begin{aligned}&-(1+r)\delta u'(c_2(\tau)) + \frac{1+r}{1+\rho}\lambda u'\left(I-P+\frac{R}{1+\rho}\right) \\ &\qquad \leqslant -(1+\lambda)u'(c_1(\tau)) + \lambda u'\left(I-P+\frac{R}{1+\rho}\right).\end{aligned}$$

Note that the right-hand side is strictly negative. Thus this inequality is rewritten as,

$$\begin{aligned}&\frac{-\delta u'(c_1(\tau)) + \frac{\lambda}{1+\rho}u'\left(I-P+\frac{R}{1+\rho}\right)}{-(1+\lambda)u'(c_1(\tau)) + \lambda u'\left(I-P+\frac{R}{1+\rho}\right)} \geqslant \frac{1}{1+r} \\ \Leftrightarrow &\sigma_\lambda(\tau) \geqslant \frac{1}{1+r}.\end{aligned}$$

□

Proof of Property (ii). Calculating $\sigma_\lambda(\tau)$ for $\tau \in T_d$, we have

$$\begin{aligned}\sigma_\lambda(\tau) &= \frac{-\delta u'(c_2(\tau)) + \frac{\lambda}{1+\rho}u'\left(I-P+\frac{R}{1+\rho}\right)}{-(1+\lambda)u'(c_1(\tau)) + \lambda u'\left(I-P+\frac{R}{1+\rho}\right)} \\ &= \frac{-(1+\rho)\delta u'(c_2(\tau)) + \lambda u'\left(I-P+\frac{R}{1+\rho}\right)}{(1+\rho)\left\{-(1+\lambda)u'(c_1(\tau)) + \lambda u'\left(I-P+\frac{R}{1+\rho}\right)\right\}}.\end{aligned}$$

By the first order condition for $\tau \in T_d$ in the consumption problem, we have

$$(1+\rho)\delta u'(c_2(\tau)) = (1+\lambda)u'(c_1(\tau)).$$

Therefore $\sigma_\lambda(\tau) = \frac{1}{1+\rho} < \frac{1}{1+r}$. □

Proof of Property (iii). It is enough to show that there exist $\tau \in T_s$ and $\tau \in T_d$ such that $\sigma_\lambda(\tau)$ is equal to that for $\tau \in T_b$, respectively. We have found that the slopes are

$$\frac{-\delta u'(c_2(\tau)) + \frac{\lambda}{1+\rho}u'\left(I - P + \frac{R}{1+\rho}\right)}{-(1+\lambda)u'(c_1(\tau)) + \lambda u'\left(I - P + \frac{R}{1+\rho}\right)}, \text{ if } \tau \in T_s \tag{2.4}$$

$$\frac{-\delta u'(R) + \frac{\lambda}{1+\rho}u'\left(I - P + \frac{R}{1+\rho}\right)}{-(1+\lambda)u'(I-P) + \lambda u'\left(I - P + \frac{R}{1+\rho}\right)}, \text{ if } \tau \in T_b \tag{2.5}$$

$$\frac{-\delta u'(c_2(\tau)) + \frac{\lambda}{1+\rho}u'\left(I - P + \frac{R}{1+\rho}\right)}{-(1+\lambda)u'(c_1(\tau)) + \lambda u'\left(I - P + \frac{R}{1+\rho}\right)}, \text{ if } \tau \in T_d. \tag{2.6}$$

Fix arbitrary $\hat{P} \in [0, I]$ and choose any pair $(R, \hat{P})$ such that $(R, \hat{P}) \in T_s$. Similarly, choose any pair $(R', \hat{P})$ such that $(R', \hat{P}) \in T_b$. Denote $\sup_{(R,\hat{P})\in T_s} R$ by $\overline{R}(\hat{P})$. Note that consumption function is continuous. According to the definitions of T_s and T_b, it follows that

$$\begin{aligned}
&c_1(R, \hat{P}) \to I - \hat{P},\ c_2(R, \hat{P}) \to \overline{R}(\hat{P}) \\
&c_1(R'\hat{P}) \to I - \hat{P},\ c_2(R', \hat{P}) \to \overline{R}(\hat{P}) \\
&\text{as } R \to \overline{R}(\hat{P}) \text{ and } R' \to \overline{R}(\hat{P}).
\end{aligned}$$

Then we have

$$\begin{aligned}
\lim_{R\to\overline{R}(\hat{P})} \sigma_\lambda(R, \hat{P}) &= \frac{-\delta u'(\overline{R}(\hat{P})) + \frac{\lambda}{1+\rho}u'\left(I - \hat{P} + \frac{\overline{R}(\hat{P})}{1+\rho}\right)}{-(1+\lambda)u'(I-\hat{P}) + \lambda u'\left(I - \hat{P} + \frac{\overline{R}(\hat{P})}{1+\rho}\right)} \\
\lim_{R'\to\overline{R}(\hat{P})} \sigma_\lambda(R', \hat{P}) &= \frac{-\delta u'(\overline{R}(\hat{P})) + \frac{\lambda}{1+\rho}u'\left(I - \hat{P} + \frac{\overline{R}(\hat{P})}{1+\rho}\right)}{-(1+\lambda)u'(I-\hat{P}) + \lambda u'\left(I - \hat{P} + \frac{\overline{R}(\hat{P})}{1+\rho}\right)},
\end{aligned}$$

that is, $\lim_{R\to\overline{R}(\hat{P})} \sigma_\lambda(R, \hat{P}) = \lim_{R'\to\overline{R}(\hat{P})} \sigma_\lambda(R', \hat{P})$. Therefore the indifference curve is smoothly continuous at $(\overline{R}(\hat{P}), \hat{P})$. Similarly we can prove the latter half of property (iii). □

Even though the consumption patterns are different for τ, this lemma states that indifference curves for fixed welfare are smooth. Thus we can use the standard method of welfare maximization. The indifference curve is shown in Figure 2.2. Here, the pension plan that satisfies feasibility is in the upper-left area of the line $P = \frac{R}{1+r}$. Hence we find that optimal plans are determined at the tangent point of the indifference curve and the line of $P = \frac{R}{1+r}$. According to the lemma 2.3.1, the indifference curve does not have any tangent to the line $P = \frac{R}{1+r}$ at $\tau \in T_s$ and $\tau \in T_d$ if $\lambda > 0$. Thus, if τ is optimal, it is included in T_b except the case of $\lambda = 0$. If $\lambda = 0$, since $\sigma_\lambda(\tau_s) = \frac{1}{1+r}$, all $\tau_s \in T_s$ are optimal.

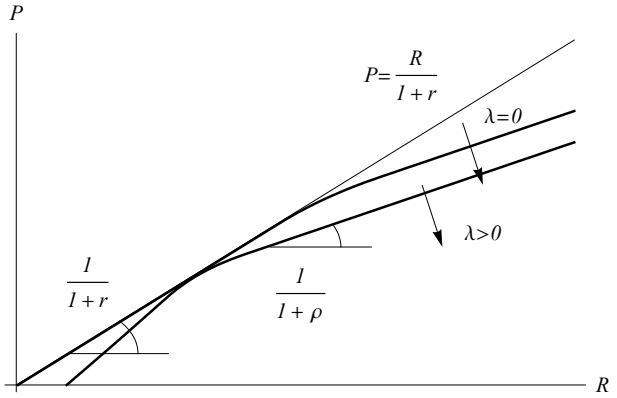

Figure 2.2: Indifference curves

Let us consider optimal τ. First, we assume that $\lambda > 0$. As we discussed, it is enough to obtain optimal plans to consider $\tau \in T_b$. The optimal plan is determined at the tangent point of the indifference curve and $P = \frac{R}{1+r}$. Here, we have

$$\sigma_\lambda(\tau_b) = \frac{-\delta u'(R) + \frac{\lambda}{1+\rho} u'\left(I - P + \frac{R}{1+\rho}\right)}{-(1+\lambda)u'(I-P) + \lambda u'\left(I - P + \frac{R}{1+\rho}\right)}.$$

Thus the necessary condition is

$$\frac{-\delta u'(R) + \frac{\lambda}{1+\rho} u'\left(I - P + \frac{R}{1+\rho}\right)}{-(1+\lambda)u'(I-P) + \lambda u'\left(I - P + \frac{R}{1+\rho}\right)} = \frac{1}{1+r}.$$

This can be rearranged as

$$\lambda = \frac{(1+r)\delta u'(R) - u'(I-P)}{u'(I-P) - \frac{\rho - r}{1+\rho} u'\left(I - P + \frac{R}{1+\rho}\right)}.$$

The denominator on the right-hand side is obviously positive since $I - P \leqslant I - \frac{\rho - r}{1+\rho} P$ and $\frac{\rho - r}{1+\rho} < 1$. However, we cannot state whether the numerator is positive or negative. Indeed, $R = (1+r)P$ in the solution then the numerator is $(1+r)\delta u'((1+r)P) - u'(I-P)$. This is positive for a sufficiently small P but negative for a sufficiently large P. Since λ is greater than 0, in order to characterize the optimal plan, we consider only P such that $(1+r)\delta u'((1+r)P) - u'(I-P) > 0$. Define $Q\colon \mathscr{P} \to \Lambda$ as

$$Q(P) \equiv \frac{(1+r)\delta u'((1+r)P) - u'(I-P)}{u'(I-P) - \frac{\rho - r}{1+\rho} u'\left(I - \frac{\rho - r}{1+\rho} P\right)}, \tag{2.7}$$

where $\mathscr{P} \equiv \{P \in [0, I]\colon (1+r)\delta u'((1+r)P) - u'(I-P) > 0\}$.

Let us see what properties Q has. Importantly, the following lemma says that the optimal τ is unique for each $\lambda > 0$, that is, there exists an inverse function for Q.

Lemma 2.3.2. *Q satisfies the following properties.*

(i) Q is strictly decreasing.

(ii) $Q(P) \to \infty$ *as* $P \to 0$.

(iii) $Q(P) = 0$ *if* P *satisfies* $(1+r)\delta u'((1+r)P) - u'(I-P) = 0$.

Proof of property (i). Let P and P' be arbitrary premiums such that $P' > P$. By the assumptions of $u'(c) > 0$ and $u''(c) < 0$, we have

$$u'(I-P') - \frac{\rho - r}{1+\rho} u'\left(I - \frac{\rho - r}{1+\rho} P'\right) > u'(I-P) - \frac{\rho - r}{1+\rho} u'\left(I - \frac{\rho - r}{1+\rho} P\right)$$

and

$$(1+r)\delta u'((1+r)P') - u'(I-P') < (1+r)\delta u'((1+r)P) - u'(I-P).$$

So it follows that

$$\begin{aligned} Q(P) &= \frac{(1+r)\delta u'((1+r)P) - u'(I-P)}{u'(I-P) - \frac{\rho-r}{1+\rho} u'\left(I - \frac{\rho-r}{1+\rho} P\right)} > \frac{(1+r)\delta u'((1+r)P) - u'(I-P)}{u'(I-P') - \frac{\rho-r}{1+\rho} u'\left(I - \frac{\rho-r}{1+\rho} P'\right)} \\ &> \frac{(1+r)\delta u'((1+r)P') - u'(I-P')}{u'(I-P') - \frac{\rho-r}{1+\rho} u'\left(I - \frac{\rho-r}{1+\rho} P'\right)} = Q(P'). \end{aligned}$$

Therefore, Q is strictly decreasing. □

Proof of property (ii). By assumption 2.2.3, we have

$$\lim_{P \to 0} Q(P) = \lim_{P \to 0} \left[\frac{(1+r)\delta u'((1+r)P) - u'(I-P)}{u'(I-P) - \frac{\rho-r}{1+\rho} u'\left(I - \frac{\rho-r}{1+\rho} P\right)} \right] = \infty.$$

□

Proof of property (iii). Since $u'(I-P) - \frac{\rho-r}{1+\rho} u'\left(I - \frac{\rho-r}{1+\rho} P\right)$ is positive for any $P \leqslant I < \infty$, the property is obvious. □

The following theorem characterizes the optimal pension plan.

Theorem 2.3.1. *The optimal pension plan is determined by the inverse function of* Q.

Proof. Q corresponds to an arbitrary pension premium $P \in \mathscr{P}$ to type $\lambda \in \Lambda$ whose optimal pension premium is P. By the lemma 2.3.2, Q is one to one function. In addition, properties (i) and (ii) say that Q is onto function. Therefore, for Q, there exists an inverse function that determines the optimal pension premium for any $\lambda \in \Lambda$. □

Theorem 2.3.1 together with property (i) in the lemma 2.3.2 says that it is optimal to set a lower premium for consumers that feel greater temptation. This result can be understood by the following discussion. Consider the situation when a consumer does not save and not borrow at all. His welfare function is rewritten as

$$\begin{aligned} W(B(\tau);\lambda) &= (1+\lambda)u(I-P) + \delta u((1+r)P) - \lambda u\left(I - \frac{\rho - r}{1+\rho} P\right) \\ &= u(I-P) + \delta u((1+r)P) - \lambda \left[u\left(I - \frac{\rho - r}{1+\rho} P\right) - u(I-P) \right]. \end{aligned}$$

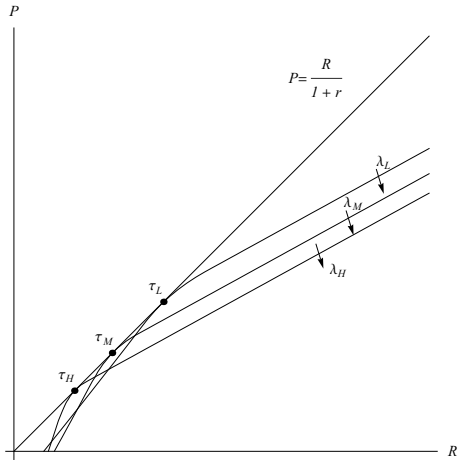

Figure 2.3: Optimal plans

The summation of the first and second terms represents the utility of consumption and the third term represents the cost of self-control. Differentiate both sides with P and we have the marginal utility minus the marginal cost of P,

$$\begin{aligned}\frac{\partial W(B(\tau);\lambda)}{\partial P} &= -u'(I-P) + (1+r)\delta u'((1+r)P) \\ &\quad - \lambda\left[-u'\left(I - \frac{\rho - r}{1+\rho}P\right)\left(\frac{\rho - r}{1+\rho}\right) + u'(I-P)\right].\end{aligned}$$

The first-order condition says that the marginal utility equals the marginal cost at optimal P. Differentiate this again with P and we have

$$\begin{aligned}\frac{\partial^2 W(B(\tau);\lambda)}{\partial P^2} =& u''(I-P) + (1+r)^2\delta u''((1+r)P) \\ &\quad - \lambda\left[u''\left(I - \frac{\rho - r}{1+\rho}P\right)\left(\frac{\rho - r}{1+\rho}\right)^2 - u''(I-P)\right].\end{aligned}$$

By assumption 2.2.2, the marginal utility decreases as P increases. And in the proof of the lemma 2.3.2, we have seen that the marginal cost increases as P does. Note that the increase in λ raises the marginal cost and we can see that P, at the point where marginal utility equals marginal cost, moves lower. Therefore, higher λ implies lower $P(\lambda)$. Intuitively, the increase in P has two effects. One effect reduces the amount of money the consumer can use when she is young. For the optimal plan, it holds that $R = (1+r)P$. Hence the increase of P means the increase of R. Thus another effect enlarges the amount of money that the consumer can borrow when she is young and it increases the temptation of borrowing. For consumers with larger λ, the latter effect is stronger. Therefore to avoid a high self-control cost, it is not proper to apply a large pension premium to a consumer with large λ. This result can be understood also by a picture. In Figure 2.3, each of three indifference curves for τ_L, τ_M, τ_H, where ,$\tau_L < \tau_M < \tau_H$, is a tangent to line $P = \frac{R}{1+r}$. The tangent points are the optimal plans. Obviously, as λ increases, the optimal plan is determined in the lower-left area of the plane.

At the end of this subsection, we consider the value of $(1+r)\delta u'((1+r)P) - u'(I-P)$ in property (iii). It is equivalent to the amount of saving when $\lambda = 0$ and there is no pension policy. We have already shown that any $\tau \in T_s^0$ is optimal for a consumer with $\lambda = 0$. That is, when a consumer does not feel any temptation, it is optimal to impose

an arbitrary premium between 0 and the optimal amount of saving. This is a well-known result: if the pension earns the same interest rate as private savings, these two ways of saving are indifferent and so the pension does not improve welfare (For example, see Samuelson, 1975). Interestingly, however, when a consumer has $\lambda > 0$, this result does not hold. Even if the interest rates are equal to each other, the private saving and the pension are not equivalent.

2.3.2 Two types of consumers

In this section, we consider two types of heterogeneous consumers. The assumption of complete information is a useful benchmark. So we begin with the situation where the government knows what type each consumer has. Denote the types as λ_L and λ_H, where $0 < \lambda_L < \lambda_H$. And the population of λ_L and λ_H are n_L and n_H, respectively, where we assume that $n_L > 0$, $n_H > 0$ and $n_L + n_H = 1$. Note that, unlike in the identical-type case, we cannot state that $R = (1+r)P$ is always satisfied at optimal schedules since there may be transfer between the types.

Remember that the marginal welfare of P for each agent is negative. Hence the pension schedule $S \equiv \{\tau_L, \tau_H\}$ such that $P_L > \frac{R_L}{1+r}$ and $P_H > \frac{R_H}{1+r}$ is not optimal since the schedule satisfies the feasibility constraint slackly, so that a sufficiently small decrease in P is feasible and improves social welfare. It is also clear that the pension schedule such that $P_L < \frac{R_L}{1+r}$ and $P_H < \frac{R_H}{1+r}$ is not feasible.

In addition, the following lemma holds.

Lemma 2.3.3. *If a pension schedule $\{\hat{\tau}_L, \hat{\tau}_H\}$ is feasible, then the schedules $\{\tau_L, \tau_H\}$ such that*

$$\left[P_L \geqslant \frac{R_L - \hat{R}_L}{1+r} + \hat{P}_L, \tau_H = \hat{\tau}_H \right] \tag{2.8}$$

or

$$\left[P_H \geqslant \frac{R_H - \hat{R}_H}{1+r} + \hat{P}_H, \tau_L = \hat{\tau}_L \right] \tag{2.9}$$

are also feasible.

Proof. Without loss of generality, we show that the schedule that satisfies (2.8) is feasible. Suppose that $\{\hat{\tau}_L, \hat{\tau}_H\}$ is feasible; that is,

$$\begin{aligned} &(1+r)(n_L\hat{P}_L + n_H\hat{P}_H) \geqslant n_L\hat{R}_L + n_H\hat{R}_H \\ \Leftrightarrow &(1+r)n_L\hat{P}_L \geqslant -(1+r)n_H\hat{P}_H + n_L\hat{R}_L + n_H\hat{R}_H. \end{aligned} \tag{2.10}$$

Choose a schedule such that $P_L \geqslant \frac{R_L - \hat{R}_L}{1+r} + \hat{P}_L, \tau_H = \hat{\tau}_H$ arbitrarily. Then it follows that

$$\begin{aligned} P_L &\geqslant \frac{R_L - \hat{R}_L}{1+r} + \hat{P}_L \\ \Leftrightarrow &(1+r)n_L P_L - n_L(R_L - \hat{R}_L) \geqslant (1+r)n_L\hat{P}_L. \end{aligned}$$

Together with (2.10), we have

$$\begin{aligned} &(1+r)n_L P_L - n_L(R_L - \hat{R}_L) + (1+r)n_H\hat{P}_H \geqslant n_L\hat{R}_L + n_H\hat{R}_H \\ \Leftrightarrow &(1+r)(n_L P_L + n_H\hat{P}_H) \geqslant n_L R_L + n_H\hat{R}_H. \end{aligned}$$

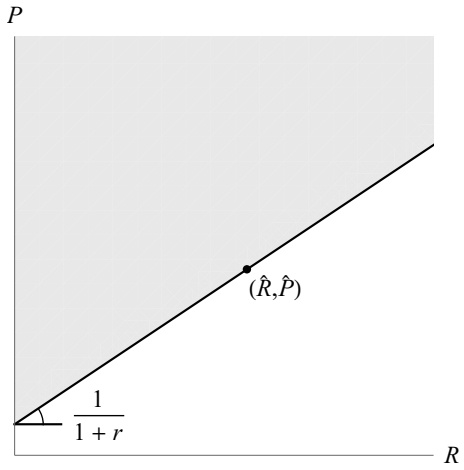

Figure 2.4: Individually feasible schedules for $(\hat{R}, \hat{P})$

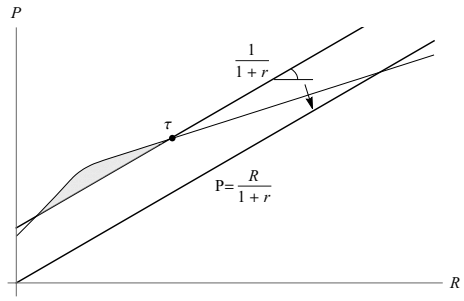

Figure 2.5: τ is not optimal plan

By $\tau_H = \hat{\tau}_H$, it follows that

$$(1+r)(n_L P_L + n_H P_H) \geqslant n_L R_L + n_H R_H.$$

Thus, the schedule $\{\tau_L, \tau_H\}$ is feasible. We can show the feasibility of the schedule that satisfies (2.9). □

For a given plan $(\hat{\tau}_\lambda)$, we call the plans τ_λ that satisfy $P_\lambda \geqslant \frac{R_\lambda - \hat{R}_\lambda}{1+r} + \hat{P}_\lambda$ individually feasible. Figure 2.4 depicts the set of schedules that is individually feasible for $(\hat{R}, \hat{P})$. The lemma 2.3.3 is useful for searching for solutions. For example, in Figure 2.5, $\sigma_\lambda(\tau)$ is lower than $\frac{1}{1+r}$. Note that with complete information, we can ignore the incentive compatibility conditions. Hence a pension plan for another type can be fixed arbitrarily. Then there is room for the improvement. Indeed, obviously τ_λ is not optimal since we can find individually feasible and more preferable plans in the hatched area in the Figure. The point is that if this improvement is available, it works individually; that is, it does not need transfer between types. Thus this is a Pareto improvement.

The next lemma is important for proving the lemma 2.3.5.

Lemma 2.3.4. *For any $\lambda \in \Lambda \subseteq \mathbb{R}_+$, τ such that $\sigma_\lambda(\tau) = \frac{1}{1+r}$ satisfies $R > 0$.*

Proof. Fix arbitrary $\lambda \in \Lambda \subseteq \mathbb{R}_+$ and choose τ such that $\sigma_\lambda(\tau) = \frac{1}{1+r}$. Note that

$\sigma_\lambda(\tau) = \frac{1}{1+r}$ implies $\tau \in T_b$. $\sigma_\lambda(\tau)$ is equal to

$$\frac{-\delta u'(R) + \frac{\lambda}{1+\rho} u'\left(I - P + \frac{R}{1+\rho}\right)}{-(1+\lambda)u'(I-P) + \lambda u'\left(I - P + \frac{R}{1+\rho}\right)}.$$

Then, by assumption 2.2.3,

$$\begin{aligned}
&\lim_{R\to 0} \frac{-\delta u'(R) + \frac{\lambda}{1+\rho} u'\left(I - P + \frac{R}{1+\rho}\right)}{-(1+\lambda)u'(I-P) + \lambda u'\left(I - P + \frac{R}{1+\rho}\right)} \\
=&\frac{-\lim_{R\to 0} \delta u'(R) + \frac{\lambda}{1+\rho} u'(I-P)}{-u'(I-P)} \\
=&\frac{-\infty}{-u'(I-P)} = \infty.
\end{aligned}$$

Since $\sigma_\lambda(\tau)$ is decreasing in R, τ such that $\sigma_\lambda(\tau) = \frac{1}{1+r}$ satisfies $R > 0$. □

The following lemma, derived from the previous lemma gives the condition for optimality.

Lemma 2.3.5. *In the optimal schedule with complete information $\{\tau_L, \tau_H\}$, both $\sigma_L(\tau_L)$ and $\sigma_H(\tau_H)$ are equal to $\frac{1}{1+r}$.*

Proof. Suppose that either $\sigma_L(\tau_L)$ or $\sigma_H(\tau_H)$ is not equal to $\frac{1}{1+r}$. Without loss of generality, suppose that $\sigma_L(\lambda_L) \neq \frac{1}{1+r}$.

(i) $\sigma_L(\tau_L) > \frac{1}{1+r}$

Consider ε_1, $\varepsilon_2 > 0$ such that $\varepsilon_2 = \frac{\varepsilon_1}{1+r}$. For such ε_1 and ε_2, let $R' \equiv R_L + \varepsilon_1$ and $P' \equiv P_L + \varepsilon_2$. Then it follows that

$$\begin{aligned}
&\varepsilon_2 = \frac{\varepsilon_1}{1+r} \\
\Leftrightarrow &P_L + \varepsilon_2 = \frac{R_L + \varepsilon_1 - R_L}{1+r} + P_L \\
\Leftrightarrow &P' = \frac{R' - R_L}{1+r} + P_L,
\end{aligned}$$

that is, $\tau' = (R', P')$ is individually feasible by lemma 2.3.3. The line represented by this equation has the slope of $\frac{1}{1+r}$ on R-P plane and passes through (R_L, P_L). Furthermore, the marginal rate of substitution at τ_L is $\sigma_L(\tau_L) > \frac{1}{1+r}$. Thus for sufficiently small ε_1 and ε_2, τ' brings more individual welfare. Since there exists an another schedule that is feasible and preferable to $\{\tau_L, \tau_H\}$, $\{\tau_L, \tau_H\}$ is not optimal.

(ii) $\sigma_L(\tau_L) < \frac{1}{1+r}$

Consider ε_1, $\varepsilon_2 > 0$, which satisfies that $\varepsilon_2 = \frac{\varepsilon_1}{1+r}$. For such ε_1 and ε_2, let $R' \equiv R_L - \varepsilon_1$ and $P' \equiv P_L - \varepsilon_2$. Here, for τ_L such that $\sigma_L(\tau_L) < \frac{1}{1+r}$, it is satisfied that $R > 0$ by the lemma 2.3.4. Then, similar to the previous case, it is satisfied that

$$\begin{aligned}
&\varepsilon_2 = \frac{\varepsilon_1}{1+r} \\
\Leftrightarrow &P_L - \varepsilon_2 = \frac{R_L - \varepsilon_1 - R_L}{1+r} + P_L \\
\Leftrightarrow &P' = \frac{R' - R_L}{1+r} + P_L,
\end{aligned}$$

that is, $\tau' = (R', P')$ is individually feasible by the lemma 2.3.3. This line has the slope of $\frac{1}{1+r}$ and passes through (R_L, P_L). Furthermore, the marginal rate of substitution at τ_L is $\sigma_L(\tau_L) < \frac{1}{1+r}$. Thus, for sufficiently small ε_1 and ε_2, τ' brings more individual welfare. Since there exists another schedule that is feasible and preferable to $\{\tau_L, \tau_H\}$, $\{\tau_L, \tau_H\}$ is not optimal. □

By the lemma 2.3.5, if a pension schedule is optimal, it holds that $\sigma_\lambda(\tau) = \frac{1}{1+r}$ for all $\lambda \in \Lambda$. Together with the lemma 2.3.4, if a pension schedule is optimal, we have $R_\lambda > 0$ for all $\lambda \in \Lambda$. That is, if a consumer feels temptation, the funded pension improves social welfare.

Here, we have the following result.

Theorem 2.3.2. *If $\lambda \neq \lambda'$, it is not optimal that $\tau_\lambda = \tau_{\lambda'}$.*

Proof. Consider an arbitrary λ, λ'. By the lemma 2.3.5, if the singleton schedule of $\{\tau\}$ is optimal, it is satisfied that

$$\frac{-\delta u'(R) + \frac{\lambda}{1+\rho} u'\left(I - P + \frac{R}{1+\rho}\right)}{(1+\lambda)u'(I-P) + \lambda u'\left(I - P + \frac{R}{1+\rho}\right)} = \frac{1}{1+r}$$

and

$$\frac{-\delta u'(R) + \frac{\lambda'}{1+\rho} u'\left(I - P + \frac{R}{1+\rho}\right)}{(1+\lambda')u'(I-P) + \lambda' u'\left(I - P + \frac{R}{1+\rho}\right)} = \frac{1}{1+r},$$

then it must be follow that

$$\frac{-\delta u'(R) + \frac{\lambda}{1+\rho} u'\left(I - P + \frac{R}{1+\rho}\right)}{(1+\lambda)u'(I-P) + \lambda u'\left(I - P + \frac{R}{1+\rho}\right)} = \frac{-\delta u'(R) + \frac{\lambda'}{1+\rho} u'\left(I - P + \frac{R}{1+\rho}\right)}{(1+\lambda')u'(I-P) + \lambda' u'\left(I - P + \frac{R}{1+\rho}\right)}. \tag{2.11}$$

This is a necessary condition for τ is optimal. For simplicity, we abbreviate some parts of (2.11) as follows:

$$\begin{aligned}
A &\equiv -\delta u'(R) \\
B &\equiv \frac{1}{1+\rho} u'\left(I - P + \frac{R}{1+\rho}\right) \\
C &\equiv u'(I-P) \\
D &\equiv u'(I-P) + u'\left(I - P + \frac{R}{1+\rho}\right).
\end{aligned}$$

Then (2.11) is rewritten to,

$$\begin{aligned}
&\frac{A + \lambda B}{C + \lambda D} = \frac{A + \lambda' B}{C + \lambda' D} \\
\Leftrightarrow &(A + \lambda B)(C + \lambda' D) = (A + \lambda' B)(C + \lambda D) \\
\Leftrightarrow &(\lambda - \lambda')(AD - BC) = 0.
\end{aligned}$$

By assumption 2.2.2, $AD = -\delta u'(R)\left[u'(I-P) + u'\left(I - P + \frac{R}{1+\rho}\right)\right]$ is strictly negative and $BC = \frac{1}{1+\rho} u'\left(I - P + \frac{R}{1+\rho}\right) u'(I-P)$ is strictly positive, it must be that $\lambda = \lambda'$. This implies (2.11), necessary condition for singleton schedule to be optimal, is satisfied only if $\lambda = \lambda'$. □

That is, if consumers have different strength of temptation, applying a common pension plan is not optimal. This is consistent with our first intuition.

Next we construct the optimal schedule in this situation. To obtain a precise result, henceforth we specialize in the normative utility function as $u(c) = \log c$ for all $c \in \mathbb{R}_+$. $\log c$ satisfies assumptions 2.2.2 and 2.2.3. For this specialized utility function, we have the following consumptions:

$$c_1(\tau) = \begin{cases} \frac{(1+\lambda)\left((1+r)(I-P)+R\right)}{(1+r)(1+\delta+\lambda)} & \text{if } 0 \leqslant P < I - \frac{(1+\lambda)R}{(1+r)\delta} \\ I - P & \text{if } I - \frac{(1+\lambda)R}{(1+r)\delta} \leqslant P \leqslant I - \frac{(1+\lambda)R}{(1+\rho)\delta} \\ \frac{(1+\lambda)\left((1+\rho)(I-P)+R\right)}{(1+\rho)(1+\delta+\lambda)} & \text{otherwise,} \end{cases}$$

$$c_2(\tau) = \begin{cases} \frac{\delta\left((1+r)(I-P)+R\right)}{1+\delta+\lambda} & \text{if } 0 \leqslant P < I - \frac{(1+\lambda)R}{(1+r)\delta} \\ R & \text{if } I - \frac{(1+\lambda)R}{(1+r)\delta} \leqslant P \leqslant I - \frac{(1+\lambda)R}{(1+\rho)\delta} \\ \frac{\delta\left((1+\rho)(I-P)+R\right)}{1+\delta+\lambda} & \text{otherwise.} \end{cases}$$

Next we consider the locus of points at which $\sigma_\lambda(\tau)$ equals $\frac{1}{1+r}$. According to the lemma 2.3.1, it is enough to obtain the locus to consider τ such that $I - \frac{(1+\lambda)R}{(1+r)\delta} \leqslant P \leqslant I - \frac{(1+\lambda)R}{(1+\rho)\delta}$. Consider an arbitrary $\lambda \in \Lambda$. Then the individual welfare for such τ is

$$W(B(\tau);\lambda) = (1+\lambda)\log(I-P) + \delta\log(R) - \lambda\log\left(I - P + \frac{R}{1+\rho}\right).$$

We have

$$\sigma_\lambda(\tau) = \frac{-\frac{\delta}{R} + \frac{\lambda}{(1+\rho)(I-P)+R}}{-\frac{1+\lambda}{I-P} + \frac{(1+\rho)\lambda}{(1+\rho)(I-P)+R}}.$$

The relation between P and R such that $\sigma_\lambda(\tau) = \frac{1}{1+r}$ is as follows:

$$P = I - K_\lambda R,$$

$$K_\lambda \equiv \frac{1+\rho+(1+r)(\lambda-\delta) + \sqrt{(1+r)^2\delta^2 + (1+\rho+(1+r)\lambda)^2 + 2\delta(1+r)(1+(1-r)\lambda+(1+2\lambda)\rho)}}{2\delta(1+r)(1+\rho)}.$$

Proposition 2.3.1. *For any $\lambda > 0$ and $\lambda' > 0$ such that $\lambda \neq \lambda'$, if $\sigma_\lambda(\tau) = \sigma_{\lambda'}(\tau') = \frac{1}{1+r}$, then $\tau \neq \tau'$.*

Proof. Fix arbitrarily $\lambda > 0$ and $\lambda' > 0$ such that $\lambda \neq \lambda'$. Suppose that for τ and τ', $\sigma_\lambda(\tau) = \sigma_{\lambda'}(\tau') = \frac{1}{1+r}$. Then it holds that $P = I - K_\lambda R$ and $P' = I - K_{\lambda'}R'$. Note that the relation between P and R and that between P' and R' are linear with the slopes of $-K_\lambda$ and $-K_{\lambda'}$, respectively. We show that if $\lambda > \lambda'$, then $K_\lambda > K_{\lambda'}$. Calculating $K_\lambda - K_{\lambda'}$, we have

$$\begin{aligned} &K_\lambda - K_{\lambda'} \\ &= \frac{1+\rho+(1+r)(\lambda-\delta) + \sqrt{(1+r)^2\delta^2 + (1+\rho+(1+r)\lambda)^2 + 2\delta(1+r)(1+(1-r)\lambda+(1+2\lambda)\rho)}}{2\delta(1+r)(1+\rho)} \\ &- \frac{1+\rho+(1+r)(\lambda'-\delta) + \sqrt{(1+r)^2\delta^2 + (1+\rho+(1+r)\lambda')^2 + 2\delta(1+r)(1+(1-r)\lambda'+(1+2\lambda')\rho)}}{2\delta(1+r)(1+\rho)} \\ &= \frac{(1+r)\lambda + \sqrt{(1+r)^2\delta^2 + (1+\rho+(1+r)\lambda)^2 + 2\delta(1+r)(1+(1-r)\lambda+(1+2\lambda)\rho)}}{2\delta(1+r)(1+\rho)} \\ &- \frac{(1+r)\lambda' + \sqrt{(1+r)^2\delta^2 + (1+\rho+(1+r)\lambda')^2 + 2\delta(1+r)(1+(1-r)\lambda'+(1+2\lambda')\rho)}}{2\delta(1+r)(1+\rho)}. \end{aligned}$$

Comparing inside the square roots, the former is larger than the latter. Thus we have $K_\lambda > K_{\lambda'}$. Though $\tau = \tau'$ holds only if $R = R' = 0$, $R > 0$ and $R' > 0$ by the lemma 2.3.4. Therefore $\tau \neq \tau'$. □

This proposition states that the point at which the indifference curve of each consumer is tangential to the feasibility frontier differs according to the type.

By the lemma 2.3.5, the optimal schedule satisfies that $\sigma_L(\tau_L) = \sigma_H(\tau_H) = \frac{1}{1+r}$ so we have

$$\begin{aligned} P_L &= I - K_L R_L \\ P_H &= I - K_H R_H. \end{aligned}$$

As we have mentioned, the feasibility condition is satisfied with equality for the optimal schedule. Hence, according to the feasibility condition, we have

$$\begin{aligned} &(1+r)[n_L P_L + n_H P_H] = n_L R_L + n_H R_H \\ \Leftrightarrow &(1+r)[n_L(I - K_L R_L) + n_H(I - K_H R_H)] = n_L R_L + n_H R_H \\ \Leftrightarrow &(1+r)[I - n_L K_L R_L - n_H K_K R_H] = n_L R_L + n_H R_H \\ \Leftrightarrow &(1 + (1+r)K_H) n_H R_H = (1+r)I - (1 + (1+r)K_L) n_L R_L \\ \Leftrightarrow &R_H = \frac{(1+r)I}{(1 + (1+r)K_H) n_H} - \frac{(1 + (1+r)K_L) n_L}{(1 + (1+r)K_H) n_H} R_L. \end{aligned}$$

Thus the summarized problem is that

$$\begin{aligned} \max_{\{\tau_\lambda\}_{\lambda = L,H}} \; & n_L \left[(1+\lambda_L) \log(I - P_L) + \delta \log R_L - \lambda_L \log \left(I - P_L + \frac{R_L}{1+r\rho} \right) \right] \\ & + n_H \left[(1+\lambda_H) \log(I - P_H) + \delta \log R_H - \lambda_H \log \left(I - P_H + \frac{R_H}{1+r\rho} \right) \right] \\ \text{s.t. } & P_L = I - K_L R_L, \; P_H = I - K_H R_H \\ & R_H = \frac{(1+r)I}{(1 + (1+r)K_H) n_H} - \frac{(1 + (1+r)K_L) n_L}{(1 + (1+r)K_H) n_H} R_L. \end{aligned}$$

Substituting P_L, P_H, and R_H, this problem can be seen as simply one variable problem. The first-order condition is

$$\begin{aligned} \frac{\partial W}{\partial R_L} =& \frac{n_L[(n_H + n_L)(1 + (1+r)K_L) R_L - (1+r)I](1+\delta)}{R_L(n_L(1 + (1+r)K_L) R_L - (1+r)I)} = 0 \\ \Leftrightarrow & R_L = \frac{(1+r)I}{1 + (1+r)K_L}. \end{aligned}$$

Then we have

$$\begin{aligned} P_L &= I - \frac{K_L(1+r)I}{1 + (1+r)K_L} = \frac{I}{1 + (1+r)K_L} \\ R_H &= \frac{(1+r)I - (1 + (1+r)K_L) n_L \frac{(1+r)I}{(1+(1+r)K_L)}}{(1 + (1+r)K_H) n_H} = \frac{(1+r)I}{1 + (1+r)K_H}. \\ P_H &= I - K_H \frac{(1+r)I}{1 + (1+r)K_H} = \frac{I}{1 + (1+r)K_H}. \end{aligned}$$

Seeing this result, we find that the optimal schedule satisfies $R_\lambda = (1+r)P_\lambda$ for $\lambda \in \Lambda$. In this situation, the incentive compatibility condition is strictly satisfied. Therefore, we have the following theorem.

Theorem 2.3.3. *Assume $u(c) = \log c$ and $|\Lambda| = 2$. The optimal pension schedule has the following form: for any $\lambda \in \Lambda$,*

$$P_\lambda = \frac{I}{1 + (1+r)K_\lambda},$$
$$R_\lambda = (1+r)P_\lambda.$$

We can see some things from the result. First, there is no monetary transfer among types since the optimal plans are balanced for each type. Second, similar to the identical-type case, there are no private saving and borrowing at the optimal schedule. This implies that the monetary market is balanced. Third, and importantly, the optimal schedule does not depend on the distribution of types.Indeed, it does not contain n_L and n_H. Furthermore, consider the following mechanism. The government shows a pension schedule such that

$$\left\{(R, P) \in T \colon P = \frac{R}{1+r}\right\}$$

and let each consumer announce the maximal amount of pension she foresees wanting in old age. Then the consumer is enrolled in the pension plan that is in accord with her decision. Though formal proof is omitted, this mechanism implements the optimal schedule with a weakly dominant strategy. This follows by the convexity of individual welfare and not by the existing externality between types. Note that this mechanism does not need any information about types, such as the distribution and even what types there are.

2.4 Generalization

In this section, we generalize our model. One generalization is about the number of types. First, we consider the case of many finite types. Then we consider the case of continuous types. As in the previous section, we specialize the normative utility function as a logarithm function.

2.4.1 Finite types

Consider the set of types $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_m\}$, where $m < \infty$. Denote the population of type λ_i by n_i, where it is assumed that $\sum_{i=1}^m n_i = 1$ and $n_i > 0$ for all $\lambda_i \in \Lambda$. The problem is that

$$\max_S \sum_{i=1}^m n_i \left\{(1+\lambda_i)\log(I - P_i) + \delta \log R_i - \lambda_i \log\left(I - P_i + \frac{R_i}{1+\rho}\right)\right\}$$
$$\text{s.t. } (1+r)\sum_{i=1}^m n_i P_i \geqslant \sum_{i=1}^m n_i R_i.$$

We can apply the lemma 2.3.5 also in this model. Thus, by substituting $P_i = I - K_i R_i$ in the problem, we have

$$\max_S \sum_{i=1}^m n_i \left\{(1+\lambda_i)\log K_i R_i + \delta \log R_i - \lambda_i \log\left(\left(K_i + \frac{1}{1+\rho}\right) R_i\right)\right\}$$
$$\text{s.t. } (1+r)\sum_{i=1}^m n_i (I - K_i R_i) \geqslant \sum_{i=1}^m n_i R_i.$$

The associated Lagrangian is

$$\begin{aligned}\mathscr{L} \equiv & \mu_0 \left[\sum_{i=1}^{m} n_i \left\{ (1+\lambda_i) \log K_i R_i + \delta \log R_i - \lambda_i \log \left(\left(K_i + \frac{1}{1+\rho} \right) R_i \right) \right\} \right] \\ & + \mu_1 \left[(1+r) \sum_{i=1}^{m} n_i (I - K_i R_i) - \sum_{i=1}^{m} n_i R_i \right].\end{aligned}$$

The necessary condition is

$$\begin{aligned}&\frac{\partial \mathscr{L}}{\partial R_j} = \mu_0 \left[n_j \left\{ \frac{1+\lambda_i}{R_j} + \frac{\delta}{R_j} - \frac{\lambda_j}{R_j} \right\} \right] + \mu_1 [-(1+r) n_j K_j - n_j] = 0, \ \forall \lambda \in \Lambda \\ &(\mu_0, \mu_1) \geqslant 0 \\ &(\mu_0, \mu_1) \neq 0 \\ &\mu_1 \left[(1+r) \sum_{i=1}^{m} n_i (I - K_i R_i) - \sum_{i=1}^{m} n_i R_i \right] = 0.\end{aligned}$$

Note that $[-(1+r)n_j K_j - n_j] < 0$. Then if $\mu_0 = 0$, the first order condition is satisfied only if $\mu_1 = 0$. This contradicts the non-zero condition of Lagrange multipliers. Thus $\mu_0 > 1$ and we can standardize this as $\mu_0 = 1$. Then the first-order conditions are rearranged as

$$n_j \frac{1+\delta}{R_j} - \mu_1 [(1+r) n_j K_j - n_j] = 0, \ \forall \lambda \in \Lambda. \tag{2.12}$$

μ_1 is not zero since the first term is strictly positive. So the feasibility condition is satisfied with equality. By (2.12), we have

$$\mu_1 = \frac{1}{(1+r)K_j + 1} \left(\frac{1+\delta}{R_j} \right).$$

Thus for any $\lambda_j \in \Lambda$ and $\lambda_k \in \Lambda$, it follows that

$$\begin{aligned}&\frac{1}{(1+r)K_j + 1} \left(\frac{1+\delta}{R_j} \right) = \frac{1}{(1+r)K_k + 1} \left(\frac{1+\delta}{R_k} \right) \\ \Leftrightarrow &[(1+r)K_j + 1] R_j = [(1+r)K_k + 1] R_k. \end{aligned} \tag{2.13}$$

By the feasibility condition with equality, we have

$$\begin{aligned}&(1+r) \sum_{i=1}^{m} n_i (I - K_i R_i) - \sum_{i=1}^{m} n_i R_i = 0 \\ \Leftrightarrow &\sum_{i=1}^{m} n_i ((1+r)K_i + 1) R_i = (1+r) I \sum_{i=1}^{m} n_i = (1+r) I.\end{aligned}$$

Fixing arbitrarily $\lambda_j \in \Lambda$ and using (2.13), it follows that

$$\begin{aligned}(1+r) I &= \sum_{i=1}^{m} n_i ((1+r)K_i + 1) R_i = ((1+r)K_j + 1) R_j \sum_{i=1}^{m} n_i \\ &= ((1+r)K_j + 1) R_j \\ \Leftrightarrow R_j &= \frac{(1+r) I}{(1+r)K_j + 1}.\end{aligned}$$

Substituting this to $P_j = I - K_j R_j$, we have

$$P_j = I - K_j \frac{(1+r)I}{(1+r)K_j + 1} = \frac{((1+r)K_j + 1)I - (1+r)K_j I}{(1+r)K_j + 1} = \frac{I}{(1+r)K_j + 1}.$$

Thus in the solution $R_j = (1+r)P_j$ is satisfied for any type $\lambda_j \in \Lambda$. This is the same result as in the two types case.

2.4.2 Continuous types

Let $\Lambda = [\underline{\lambda}, \overline{\lambda}]$, where $\underline{\lambda} \geqslant 0$. $\lambda \in \Lambda$ are distributed according to the distribution function F. $(R(\lambda), P(\lambda))$ denotes the pension plan for the consumer whose type is $\lambda \in \Lambda$. We assume that $R(\lambda)$ and $P(\lambda)$ are continuous and differentiable.

Similarly, considering the problem with complete information, it follows that $P(\lambda) = I - K(\lambda)R(\lambda)$, $c_1^\lambda = I - P(\lambda)$ and $c_2^\lambda = R(\lambda)$ for each λ, where $K(\lambda)$ is the constant that has the same form as K_λ in the previous section. The problem is that

$$\begin{aligned}
&\max_S \int_{\underline{\lambda}}^{\overline{\lambda}} \left[(1+\lambda)\log(I - P(\lambda)) + \delta \log(R(\lambda)) - \lambda \log\left(I - P(\lambda) + \frac{R(\lambda)}{1+\rho} \right) \right] dF(\lambda) \\
&\text{s.t.} \int_{\underline{\lambda}}^{\overline{\lambda}} \left[(1+r)P(\lambda) - R(\lambda) \right] dF(\lambda) = 0 \\
&\quad P(\lambda) = I - K(\lambda)R(\lambda),
\end{aligned}$$

where we can assume that the solution satisfies the feasibility condition with equality for the same reason as in the previous section. Since the feasibility constraint includes a integration, we rewrite it. We define a new state variable h as

$$h(\lambda) \equiv \int_{\underline{\lambda}}^{\lambda} \left[(1+r)P(x) - R(x) \right] dF(x).$$

Then, h satisfies the following conditions, conversely implying the feasibility condition.

$$\begin{aligned}
h'(\lambda) &= (1+r)P(\lambda) - R(\lambda), \\
h(\underline{\lambda}) &= 0 \text{ and } h(\overline{\lambda}) = 0.
\end{aligned}$$

Thus, we have the rewritten problem:

$$\begin{aligned}
&\max_S \int_{\underline{\lambda}}^{\overline{\lambda}} \left[(1+\lambda)\log(I - P(\lambda)) + \delta \log(R(\lambda)) - \lambda \log\left(I - P(\lambda) + \frac{R(\lambda)}{1+\rho} \right) \right] dF(\lambda) \\
&\text{s.t. } h'(\lambda) = (1+r)P(\lambda) - R(\lambda), \\
&\quad h(\underline{\lambda}) = 0 \text{ and } h(\overline{\lambda}) = 0 \\
&\quad P(\lambda) = I - K(\lambda)R(\lambda).
\end{aligned}$$

The associated Hamiltonian is,

$$\begin{aligned}
\mathscr{H} \equiv &(1+\lambda)\log(I - P(\lambda)) + \delta \log(R(\lambda)) - \lambda \log\left(I - P(\lambda) + \frac{R(\lambda)}{1+\rho} \right) + \mu(\lambda)h'(\lambda) \\
&= (1+\lambda)\log(I - P(\lambda)) + \delta \log(R(\lambda)) - \lambda \log\left(I - P(\lambda) + \frac{R(\lambda)}{1+\rho} \right) \\
&+ \mu(\lambda)\left[(1+r)P(\lambda) - R(\lambda) \right],
\end{aligned}$$

where μ is the co-state variable. Substituting $P(\lambda) = I - K(\lambda)R(\lambda)$, we have,

$$\begin{aligned}\mathscr{H} &= (1+\lambda)\log(K(\lambda)R(\lambda)) + \delta\log(R(\lambda)) - \lambda\log\left(\left(K(\lambda)+\frac{1}{1+\rho}\right)R(\lambda)\right) \\ &\quad + \mu(\lambda)\left[(1+r)I - ((1+r)K(\lambda)+1)R(\lambda)\right].\end{aligned}$$

By Pontryagin's principle,

$$\begin{aligned}\frac{\partial\mathscr{H}}{\partial R(\lambda)} &= \frac{1+\lambda}{R(\lambda)} + \frac{\delta}{R(\lambda)} - \frac{\lambda}{R(\lambda)} - \mu(\lambda)((1+r)K(\lambda)+1)R(\lambda) \\ &= \frac{1+\delta}{R(\lambda)} - \mu(\lambda)((1+r)K(\lambda)+1 = 0 \qquad (2.14) \\ \mu'(\lambda) &= -\frac{\partial\mathscr{H}}{\partial h(\lambda)}.\end{aligned}$$

Note that $-\partial\mathscr{H}/\partial h(\lambda) = 0$, so $u'(\lambda) = 0$. This implies that $u(\lambda)$ does not depend on λ. Hence we can simply wreeite $\mu(\lambda) = \mu$. (2.14) can be rearranged as,

$$R(\lambda) = \frac{1+\delta}{((1+r)K(\lambda)+1)\mu}.$$

By the feasibility condition,

$$\begin{aligned}&\int_{\underline{\lambda}}^{\overline{\lambda}} (1+r)(I - K(\lambda)R(\lambda)) - R(\lambda)dF(\lambda) \\ &= \int_{\underline{\lambda}}^{\overline{\lambda}} (1+r)I - ((1+r)K(\lambda)+1)\,R(\lambda)dF(\lambda) \\ &= \int_{\underline{\lambda}}^{\overline{\lambda}} (1+r)I - ((1+r)K(\lambda)+1)\,\frac{1+\delta}{((1+r)K(\lambda)+1)\,\mu}dF(\lambda) \\ &= \left((1+r)I - \frac{1+\delta}{\mu}\right)\int_{\underline{\lambda}}^{\overline{\lambda}} dF(\lambda) = (1+r)I - \frac{1+\delta}{\mu} = 0 \\ \Leftrightarrow\ &\mu = \frac{1+\delta}{(1+r)I}\end{aligned}$$

Therefore we have,

$$\begin{aligned}R(\lambda) &= \frac{1+\delta}{((1+r)K(\lambda)+1)\mu} = \frac{1+\delta}{((1+r)K(\lambda)+1)\frac{1+\delta}{(1+r)I}} = \frac{(1+r)I}{(1+r)K(\lambda)+1}, \\ P(\lambda) &= I - K(\lambda)R(\lambda) = I - \frac{(1+r)K(\lambda)I}{(1+r)K(\lambda)+1} = \frac{I}{(1+r)K(\lambda)+1}.\end{aligned}$$

Again we obtained the relation $R(\lambda) = (1+r)P(\lambda)$. For the same reason as in the previous section, if we construct the pension schedule of such plans, the IC condition is strictly satisfied. The following theorem summarizes this section.

Theorem 2.4.1. *Assume $u(c) = \log c$ and $\Lambda = [\underline{\lambda}, \overline{\lambda}]$. The optimal pension schedule has the following form: for any $\lambda \in \Lambda$,*

$$\begin{aligned}P(\lambda) &= \frac{I}{(1+r)K(\lambda)+1} \\ R(\lambda) &= (1+r)P(\lambda).\end{aligned}$$

2.5 Discussion

2.5.1 The effect of a borrowing constraint

So far, we have assumed that consumers are allowed to borrow money in the first period. In this section, we consider the special case in which consumers can borrow no money; that is, we restrict $s \geqslant 0$[3]. In the context of self-control preference, this assumption has an important meaning. The impossibility of borrowing after paying a premium strengthens the funded pension scheme as a commitment device.

With the borrowing constraint in place, the budget constraint is simply

$$B(\tau) = \{(c_1, c_2) \in \mathbb{R}^2_+ \colon c_1 + s \leqslant I - P, c_2 \leqslant (1+r)(I - c_1) + R\}.$$

For simplicity, we specialize a normative utility function as $u(c) = \log(c)$. Then the consumption in period 1 is

$$c_1(\tau) = \begin{cases} \frac{(1+\lambda)[(1+r)(I-P)+R]}{(1+r)(1+\delta+\lambda)} & \text{if } 0 \leqslant P < I - \frac{(1+\lambda)R}{\delta(1+r)} \\ I - P & \text{if } I - \frac{(1+\lambda)R}{\delta(1+r)} \leqslant P \leqslant I. \end{cases}$$

Similar to the case in section 3.1, we assume the identical type. Then $R = (1+r)P$ follows.

Calculating an optimal pension plan, we have

$$P_\lambda = \begin{cases} \text{any number } P \in [0, \frac{\delta I}{1+\delta}] & \text{if } \lambda = 0 \\ \frac{\delta I}{1+\delta} & \text{if } \lambda > 0 \end{cases}$$
$$R_\lambda = (1+r)P_\lambda, \ \forall \lambda \geqslant 0.$$

Note that P does not depend on the type if $\lambda > 0$. This optimal P is equal to optimal saving when there is no temptation and no pension policy. Intuitively, by the consumption decision above, a consumer who paid $\frac{\delta I}{1+\delta}$ consumes all of remaining money in period 1. Since $\frac{\delta I}{1+\delta}$ is the optimal saving, the optimal consumption includes all of the remaining money. Thus, naturally, welfare is maximized without the harm of temptation. The interest rate for the pension is the same as that for private saving, so there is no difference in the amount of payout between the pension and saving. However, the pension, which make her pay a premium in advance has a role as a commitment device. Saving does not have the role since a consumer decide how much to save after she faces the temptation. Furthermore, importantly, now the consumer is not allowed to borrow, so the effect of an increase in the premium further strengthens the budget set: there is no temptation to borrow. This result is very different from that in sections 3 and 4.

2.5.2 Income diversity

In this section, we study the case when the income can be differ for each consumer. The income is one of the elements in $\mathcal{I} = \{I_1, I_2, \ldots, I_m\}$ and the degree of temptation is drawn from $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_k\}$, where $0 < I_1 < I_2 < \cdots < I_m < \infty$ and $0 < \lambda_1 < \lambda_2 < \cdots < \lambda_k < \infty$. Consumers are characterized by the pair $(I_s, \lambda_t) \in \Theta \subseteq \mathcal{I} \times \Lambda$. n_{st} denotes the proportion of the consumers having $(I_s, \lambda_t) \in \Theta$. A pension plan for consumer with (I_s, λ_t) is the pair (R_{st}, P_{st}). We assume that R_{st} and P_{st} are weakly positive and that consumers have to be able to make their payment for the pension, that is, $P_{st} \leqslant I_s$.

[3]There may be various strengths of the constraint, but here we consider only the strongest borrowing constraint.

Differently from the degree of temptation, it is natural that the government can observe the income of consumers. In fact, the government uses the information for an income tax imposition. Thus we assume that the government can observe the income of consumers but does not know the degree of temptation. However, as in the previous section, we consider the problem with complete information at first. Then we can use lemma 2.3.5 again because the result does not depend on income, so we can focus on the pension plans which satisfies $P_{st} = I_s - K_t R_{st}$. Using this, the constraint that $0 \leqslant P_{st} \leqslant I_s$ is rewritten as $0 \leqslant R_{st} \leqslant I_s/K_t$. The feasibility condition can also be rearranged as

$$(1+r)\sum_{(s,t)\in\Theta} n_{st}P_{st} \geqslant \sum_{(s,t)\in\Theta} n_{st}R_{st} \iff (1+r)\sum_{(s,t)\in\Theta} n_{st}(I_s - K_t R_{st}) \geqslant \sum_{(s,t)\in\Theta} n_{st}R_{st}$$
$$\iff \sum_{(s,t)\in\Theta} n_{st}\left\{(1+r)I_s - (1+(1+r)K_t)R_{st}\right\} \geqslant 0.$$

The solution satisfies the following property.

Lemma 2.5.1. *If a pension schedule is optimal, it holds that $R_{st} > 0$ for all s and t.*

Proof. At first, we show that there is at least one type whose pension return is strictly positive. Suppose that there exists consumer with (I_s, λ_t) whose pension plan is $R_{st} = 0$. Then private welfare of the consumer is

$$\begin{aligned}&(1+\lambda_t)\log(I_s - P_{st}) + \delta \log R_{st} - \lambda_t \log\left(I_s - P_{st} + \frac{R_{st}}{1+\rho}\right)\\ &= (1+\lambda_t)\log(I_s - P_{st}) + \delta\log 0 - \lambda_t \log(I_s - P_{st}) = \log(I_s - P_{st}) + \delta \log 0 = -\infty.\end{aligned}$$

Note that private welfare is bounded above by the constraints. Then the social welfare in this case is $-\infty$. On the other hand, let $(R'_{st}, P'_{st}) = (\varepsilon, \varepsilon/(1+r))$ for all s and t with $\varepsilon > 0$ that satisfies $\varepsilon/(1+r) < I_1$. A pension schedule composed of this plan is feasible. In fact, it holds that

$$\sum_{(s,t)\in\Theta} n_{st}\left\{(1+r)P'_{st} - R_{st}\right\} = \sum_{(s,t)\in\Theta} n_{st}\left\{\varepsilon - \varepsilon\right\} = 0.$$

The social welfare of this schedule is

$$\sum_{(s,t)\in\Theta} n_{st}\left\{(1+\lambda_t)\log\left(I_s - \frac{\varepsilon}{1+r}\right) + \delta\log\varepsilon - \lambda_t \log\left(I_s - \frac{\varepsilon}{1+r} + \frac{\varepsilon}{1+\rho}\right)\right\}$$

This is strictly larger than $-\infty$ since the antilogarithm of the first and the second term is greater than 0 and the third term is bounded above by $\varepsilon/(1+r) < I_1 < \infty$ and $\lambda_k < \infty$. Therefore, the schedule that assign $R_{st} = 0$ for all consumers is not a solution.

Next suppose that $R_{st} = 0$ for some s, t in the optimal shcedule. Z denote the set of these (s,t). By the discussion above, there is at least one $\tilde{s}$, $\tilde{t}$ whose plan satisfies $R_{\tilde{s}\tilde{t}} > 0$. As we saw, social welfare when there exist some consumers whose pension return is 0 is $-\infty$. Let $R'_{\tilde{s}\tilde{t}}$ be $\varepsilon \in (0, R_{st})$ and the return for consumers whose initial return is 0 be $R'_{\tilde{s}\tilde{t}}$ that satisfies

$$\sum_{(s,t)\in Z} n_{st}R'_{st} = n_{\tilde{s}\tilde{t}}(R_{\tilde{s}\tilde{t}} - \varepsilon).$$

For other consumers that does not included in $Z \cup \{\tilde{s}, \tilde{t}\}$, their plans are not be changed from initial plans. This assures that the feasibility condition holds. Actually,

$$\begin{aligned}
&\sum_{(s,t)\in\Theta} n_{st}\{(1+r)P'_{st} - R'_{st}\} \\
&= \sum_{(s,t)\in Z} n_{st}\{(1+r)P_{st} - R'_{st}\} + \sum_{(s,t)\notin Z\cup\{(\tilde{s},\tilde{t})\}} n_{st}\{(1+r)P_{st} - R_{st}\} + n_{\tilde{s}\tilde{t}}\{(1+r)P_{\tilde{s}\tilde{t}} - \varepsilon\} \\
&= \sum_{(s,t)\in Z} n_{st}(1+r)P_{st} + \sum_{(s,t)\notin Z\cup\{(\tilde{s},\tilde{t})\}} n_{st}\{(1+r)P_{st} - R_{st}\} + n_{\tilde{s}\tilde{t}}\{(1+r)P_{\tilde{s}\tilde{t}} - R_{\tilde{s}\tilde{t}}\} \\
&= \sum_{(s,t)\in\Theta} \{(1+r)P_{st} - R_{st}\} \geqslant 0
\end{aligned}$$

since the last line is feasibility condition for initial schedule. By the same reason as the discussion above, the social welfare of the revised schedule is strictly larger than $-\infty$. Therefore, the schedule that assign zero return to more than one consumer is not optimal. □

We use this result in the analysis below. The government's problem is as follows:

$$\begin{aligned}
\max_{S} \quad & \sum_{(s,t)\in\Theta} n_{st}\left\{(1+\lambda_t)\log K_t R_{st} + \delta \log R_{st} - \lambda_t \log\left(K_t + \frac{1}{1+\rho}\right) R_{st}\right\} \\
\text{s.t} \quad & \sum_{(s,t)\in\Theta} n_{st}\left\{(1+r)I_s - (1+(1+r)K_t)R_{st}\right\} \geqslant 0 \\
& 0 \leqslant R_{st} \leqslant I_s/K_t, \ \forall (s,t) \in \Theta.
\end{aligned}$$

The associated Lagrangian is

$$\begin{aligned}
\mathcal{L} = \mu_0 & \sum_{(s,t)\in\Theta} n_{st}\left[(1+\lambda_t)\log K_t R_{st} + \delta \log R_{st} - \lambda_t \log\left(K_t + \frac{1}{1+\rho}\right) R_{st}\right] \\
+ \mu_1 & \sum_{(s,t)\in\Theta} n_{st}\left[(1+r)I_s - (1+(1+r)K_t)R_{st}\right] + \sum_{(s,t)\in\Theta} \mu_{2st} R_{st} + \sum_{(s,t)\in\Theta} \mu_{3st}\left(\frac{I_s}{K_t} - R_{st}\right),
\end{aligned}$$

where μ_0, μ_1, μ_{2st} and μ_{3st} are Lagrange multipliers. Note that $\mu_{2st} = 0$ for all (s,t) in the solution because of the complementary slackness conditions for the constraint $R_{st} \geqslant 0$ and lemma 2.5.1. Then the first order conditions for R_{st} are

$$\frac{\partial \mathcal{L}}{\partial R_{st}} = \mu_0 n_{st}\frac{1+\delta}{R_{st}} - \mu_1 n_{st}(1+(1+r)K_t) - \mu_{3st} = 0. \tag{2.15}$$

We can see that μ_0 is strictly positive in the solution. Suppose that $\mu_0 = 0$. Then $\mu_1 = 0$ and $\mu_{3st} = 0$ must hold for all (s,t) by (2.15). However, this contradicts non-zero condition of Lagrange multipliers. Hence we can standardize μ_0 as 1. Then (2.15) is rewritten as

$$\mu_1 = \left(\frac{1+\delta}{R_{st}} - \frac{\mu_{3st}}{n_{st}}\right)\frac{1}{1+(1+r)K_t}. \tag{2.16}$$

The left hand side is independent of (s,t), the right hand side is equivalent for all (s,t).

We can exclude a schedule with $R_{st} = I_s/K_t$ for all (s,t) from a candidate of solution since this schedule is infeasible. In addition, here we show the following lemma.

Lemma 2.5.2. *For all* $(s,t) \in \Theta$, $R_{st} < I_s/K_t$.

Proof. To show this, we assume that $R_{st} = I_s/K_t$ for some $(s,t) \in \Theta$ and derive optimal plans for other $(\tilde{s},\tilde{t}) \neq (s,t)$. We propose a contradiction between a condition that $R_{\tilde{s}\tilde{t}} > 0$ and a condition that $\mu_{3st} \geqslant 0$. Define C to be $C \equiv \{(s,t) \in \Theta \colon R_{st} = I_s/K_t\}$. By the definition and the complementary slackness, $\mu_{3s't'} = 0$ for $(s',t') \notin C$, which implies two things. First, $\mu_1 > 0$ by (2.16), that is, the feasibility condition is satisfied with equality. Second, by (2.16) again, it holds that

$$\begin{aligned} &\frac{1+\delta}{R_{st}} \frac{1}{1+(1+r)K_t} = \frac{1+\delta}{R_{s't'}} \frac{1}{1+(1+r)K_{t'}} \\ \iff &(1+(1+r)K_t)R_{st} = (1+(1+r)K_{t'})R_{s't'} \end{aligned} \tag{2.17}$$

for any (s,t), $(s',t') \notin C$. Using this relation, we have

$$\begin{aligned} &\sum_{(s,t)\in\Theta} n_{st}\left[(1+r)I_s - (1+(1+r)K_t)R_{st}\right] \\ &= \sum_{(s,t)\notin C} n_{st}\left[(1+r)I_s - (1+(1+r)K_t)R_{st}\right] \\ &\qquad + \sum_{(s,t)\in C} n_{st}\left[\frac{(1+r)K_tI_s - (1+(1+r)K_t)I_s}{K_t}\right] = 0 \\ \iff &(1+(1+r)K_{\tilde{t}})R_{\tilde{s}\tilde{t}} \sum_{(x,t)\notin C} n_{st} = (1+r)\sum_{(s,t)\notin C} n_{st}I_s - \sum_{(s,t)\in C} \frac{n_{st}I_s}{K_t} \\ \iff &R_{\tilde{s}\tilde{t}} = \frac{1}{(1+(1+r)K_{\tilde{t}})\sum_{(s,t)\notin C} n_{st}} \left[(1+r)\sum_{(s,t)\notin C} n_{st}I_s - \sum_{(s,t)\in C} \frac{n_{st}I_s}{K_t}\right] \end{aligned}$$

for any $(\tilde{s},\tilde{t}) \notin C$. This is positive if and only if

$$(1+r)\sum_{(s,t)\notin C} n_{st}I_s > \sum_{(s,t)\in C} \frac{n_{st}I_s}{K_t}. \tag{2.18}$$

On the other hand, it must hold that $\mu_{3st} \geqslant 0$ for $(s,t) \in C$. By (2.16), for any $(s,t) \in C$ and $(s',t') \notin C$, we have the following:

$$\begin{aligned} &\frac{1+\delta}{R_{s't'}} \frac{1}{1+(1+r)K_{t'}} = \left(\frac{1+\delta}{R_{st}} - \frac{\mu_{3st}}{n_{st}}\right) \frac{1}{1+(1+r)K_t} \\ \Rightarrow &\mu_{3st} = n_{st}(1+\delta)\left(\frac{K_t}{I_s} - \frac{(1+(1+r)K_t)\sum_{(s,t)\notin C} n_t^s}{(1+r)\sum_{(s,t)\notin C} n_t^s I_s - \sum_{(s,t)\in C}(n_{st}I_s/K_t)}\right). \end{aligned}$$

A sufficient and necessary condition for $\mu_{3st} \geqslant 0$ is

$$\begin{aligned} &\frac{K_t}{I_s} \geqslant \frac{(1+(1+r)K_t)\sum_{(s,t)\notin C} n_t^s}{(1+r)\sum_{(s,t)\in C} n_t^s I_s - \sum_{(s,t)\in C}(n_{st}I_s/K_t)} \\ \iff &\frac{I_s}{K_t} \frac{(1+(1+r)K_t)\sum_{(s,t)\notin C} n_t^s}{(1+r)\sum_{(s,t)\in C} n_t^s I_s - \sum_{(s,t)\in C}(n_{st}I_s/K_t)} \leqslant 0 \\ \iff &(1+r)\sum_{(s,t)\notin C} n_t^s I_s < \sum_{(s,t)\in C} \frac{n_{st}I_s}{K_t} \end{aligned} \tag{2.19}$$

Thus, (2.19) contradicts (2.18). □

By lemma, we can focus on an inner solution. Then by the complementary slackness conditions, $\mu_{3st} = 0$ for all $(s,t) \in \Theta$ and we can derive the relation (2.17) for any pair (s,t), $(s',t') \in \Theta$. Moreover, the feasibility condition is satisfied with equality by (2.16) and $\mu_{3st} = 0$. Substituting (2.17) to the feasibility condition, we have

$$R_{st} = \frac{(1+r)\sum_{(\tilde{s},\tilde{t})\in\Theta} n_{\tilde{s}\tilde{t}} I_{\tilde{s}}}{1+(1+r)K_t}$$

for all $(s,t) \in \Theta$. To derive P_{st}, substitute R_{st} to $P_{st} = I_s - K_t R_{st}$. Then

$$\begin{aligned}
P_{st} &= I_s - K_t \frac{(1+r)\sum_{(\tilde{s},\tilde{t})\in\Theta} n_{\tilde{s}\tilde{t}} I_{\tilde{s}}}{1+(1+r)K_t} \\
&= \frac{[1+(1+r)K_t]I_s - K_t(1+r)n_{st}I_s - K_t(1+r)\sum_{(s',t')\neq(s,t)} n_{s't'} I_{s'}}{1+(1+r)K_t} \\
&= \frac{I_s + (1+r)K_t \sum_{(s',t')\neq(s,t)} n_{s't'} I_s - (1+r)K_t \sum_{(s',t')\neq(s,t)} n_{s't'} I_{s'}}{1+(1+r)K_t} \\
&= \frac{I_s + (1+r)K_t \sum_{(s',t')\neq(s,t)} n_{s't'} (I_s - I_{s'})}{1+(1+r)K_t}.
\end{aligned}$$

We need to check whether consumers have incentive to untruthfully tell their degree of temptation. Since their optimal plans that we have derived is determined on the locus of $\sigma_{st}(\tau_{st}) = 1/(1+r)$, one sufficient condition for no-deviation is that plans for arbitrary two consumers with the same income are on the same line with the gradient of $1/(1+r)$. Actually, this is satisfied.

Proposition 2.5.1. *For any (s,t) and (s,t'), it holds that*

$$\frac{P_{st} - P_{st'}}{R_{st} - R_{st'}} = \frac{1}{1+r}.$$

Proof. We calculate $R_{st} - R_{st'}$ and $P_{st} - P_{st'}$ in practice.

$$\begin{aligned}
R_{st} - R_{st'} &= \frac{(1+r)\sum_{(\tilde{s},\tilde{t})\in\Theta} n_{\tilde{s}\tilde{t}} I_{\tilde{s}}}{1+(1+r)K_t} - \frac{(1+r)\sum_{(\tilde{s},\tilde{t})\in\Theta} n_{\tilde{s}\tilde{t}} I_{\tilde{s}}}{1+(1+r)K_{t'}} = \frac{(1+r)(K_{t'} - K_t)\sum_{(\tilde{s},\tilde{t})\in\Theta} n_{\tilde{s}\tilde{t}} I_{\tilde{s}}}{[1+(1+r)K_t][1+(1+r)K_{t'}]}, \\
P_{st} - P_{st'} &= (I_s - K_t R_{st}) - (I_s - K_t' R_{st'}) \\
&= \frac{K_{t'}((1+r)\sum_{(\tilde{s},\tilde{t})\in\Theta} n_{\tilde{s}\tilde{t}} I_{\tilde{s}})}{1+(1+r)K_t} - \frac{K_t((1+r)\sum_{(\tilde{s},\tilde{t})\in\Theta} n_{\tilde{s}\tilde{t}} I_{\tilde{s}})}{1+(1+r)K_{t'}} \\
&= \frac{K_{t'}[1+(1+r)K_{t'}] - K_t[1+(1+r)K_t]}{[1+(1+r)K_t][1+(1+r)K_{t'}]} \sum_{(\tilde{s},\tilde{t})\in\Theta} n_{\tilde{s}\tilde{t}} I_{\tilde{s}} \\
&= \frac{(K_{t'} - K_t)\sum_{(\tilde{s},\tilde{t})\in\Theta} n_{\tilde{s}\tilde{t}} I_{\tilde{s}}}{[1+(1+r)K_t][1+(1+r)K_{t'}]}.
\end{aligned}$$

Hence, we have the intended result. □

Therefore, we can conclude that the solution we have derived is not only for the problem with complete information but also for the problem with incomplete information.

Theorem 2.5.1. *In the unique optimal schedule, the plan of consumer with (I_s, λ_t) is*

$$\begin{aligned}
R_{st} &= \frac{(1+r)\sum_{(\tilde{s},\tilde{t})\in\Theta} n_{\tilde{s}\tilde{t}} I_{\tilde{s}}}{1+(1+r)K_t} \\
P_{st} &= \frac{I_s + (1+r)K_t \sum_{(s',t')\neq(s,t)} n_{s't'} (I_s - I_{s'})}{1+(1+r)K_t}.
\end{aligned}$$

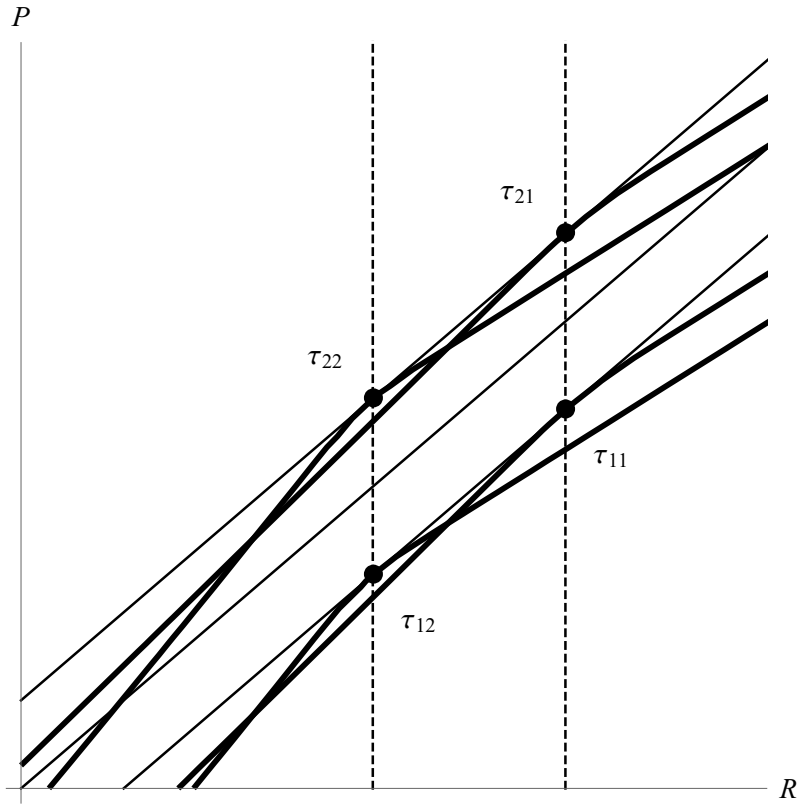

Figure 2.6: An optimal schedule

Figure 2.6 shows an example of the optimal schedule when $\mathcal{I} = \{I_1, I_2\}$ and $\Lambda = \{\lambda_1, \lambda_2\}$. Four points are the elements of optimal schedule. Note that the consumers can lie about only their degree of temptation. Thus the deviation that we have to think is between (I_1, λ_1) and (I_1, λ_2) and between (I_2, λ_1) and (I_2, λ_2). We can see that the deviation makes loss for any types.

The optimal schedule above has some characteristic property. First, clearly, this schedule is the generalization of the result in the previous section with identical income. You can see this by assuming $I_1 = I_2 = \cdots = I_m$ in the plan for (s, t). Second, the pension return is equivalent for two consumers if and only if their degree of temptation are equivalent. That is, the determination of the amount of return depends only on the degree of temptation and not on income. On the other hand, third, the payment for the pension fund depends both of income and the degree of temptation. Especially, consumers with relatively higher income pay more amount, since the summation in the numerator of P_{st} become larger when I_s is high. This implies that there is a monetary transfer from high-income consumers to low-income consumers through the differences of payments.

2.6 Conclusion

We have considered optimally funded pensions for consumers who face the temptation to overconsume and as well as for those who have enough self-control to withstand their temptation. Since funded pensions tighten consumers' budgets , they can serve as commitment devices to avoid overconsumption. We applied the pension to an economy in which consumers have heterogeneous self-control. We showed that funded pensions can

improve social welfare even if the interest rates they draw are the same as those for private saving. In addition, consumers do not save individually when they choose the pension plan that is optimal for them. Furthermore, interestingly, lower pension premium and lower pension payout are applied for a higher temptation economy. This result is related to borrowing constraints. In an identical-type economy, an increase in the premium leads to increased pension income, and this augments possibility of debt. Since consumers are tempted to overconsume, this works stronger for consumers who have strong temptations. This effect is greater than the benefit of strengthening the budget set. As a main result, we have considered an optimal pension schedule when there exist two or more types in the economy. In that situation, we show the necessary conditions for the optimal schedule. If the normative utility function is a logarithm, it is characterized in the same way as that for an economy with identical-type. We show that monetary transfer among types will not occur for the optimal schedule. An important result is that the optimal schedule does not depend on the distribution of types; that is, what the government has to know is only what types are in the economy. This make the operation of the pension policy easier.

In the future research, we face many problems. In the latter half of this chapter, we focused on the normative utility function as a logarithm. The first thing to do is to show that the result holds in a more general utility function. However, this challenge contains an analytical problem. We have not found a way to solve the problem explicitly for utility functions that are frequently used in economics, such as constant power, CRRA and CARA. Some novel methods are needed to work out this point.

In addition, we restrict the policy to funded pensions. But pay-as-you-go pension can be considered. By association, we can consider a model in which a consumer lives longer than the two periods used in this study. Then it may be natural to assume a consumer earns wages in each period. Moreover, if we assume multiple periods, consumers may have the opportunity to change their pension plans in each period. It will be interesting to see whether this change has a positive effect on social welfare. These problems may be complex and challenging.

Chapter 3

Consumer Inertia and Competitive Behavior between Firms

3.1 Introduction

When we buy extremely expensive goods such as a house and a car, we probably check the quality of them closely by making use of reviews in websites and word of mouth even if the survey of quality is costly. On the other hand, we do not check the qualities of inexpensive commodities. There are goods in the middle of these goods, which we have a conflict whether we do the survey or not. For example, each firm put out a new model of smartphones seasonally. In Japan, average age of use is 2.0-3.8 years and about 40% among the reasons of replacement are requirements for a high-end equipment.[1] When we replace these goods, in some vague way, we know the quality of a new product put out by the firm whose product is currently used. Though we can look at specs of other products, it is difficult to know actual quality unless we use them. Then, one option is to buy the product of familiar firm with out the survey. Another option is to survey the qualities of other product spending time and effort. It is not obvious whether consumers do the survey, which depends on the magnitude of the survey cost and how much they set store on the quality. We focus on the consumers' behavior of not surveying non-default product even if it may have higher quality. We call this behavior *inertia*. Our purpose is to study the inertial behaviors and price system which chosen in equilibria under inertia.

Some experimental and empirical studies show that individuals tend to choose their status quo alternatives (e.g. Samuelson and Zeckhauser, 1988). Frequently used concept to explain the bias is loss aversion. This is one special case of reference dependence preferences. Kahneman and Tversky (1984) formed a hypothesis that a value function is defined on gains and losses and it is steeper for losses than for gains.[2] It is well known that this value function can explain the loss aversive behaviors of individuals. As an other approach to the status quo bias, Masatlioglu and Ok (2005) proposed an rational choice theory which allows the existence of status quo bias and includes a standard choice model as a special case.

When consumers have default alternatives, their search of the prices and the qualities of other products are likely to be limited when the search is costly. Armstrong et al. (2009) studied the competition between firms when consumers are not completely informed

[1] Consumer Confidence Survey, Cabinet Office, Government of Japan (`http://www.esri.cao.go.jp/jp/stat/shouhi/shouhi.html`)

[2] Their one more hypothesis is that the value function is concave in the domain of gains and convex in the domain of losses.

about prices. In their model, the consumers decide whether to compare the prices or not endogenously. Though they are likely to save the expenditure, they may not compare when the comparison is costly. From the viewpoint of consumer protection policy, a government can make the upper bound of the price. However, the policy may end up a bad effect for consumers since it makes them feel ease and suppresses voluntary price comparisons.

We study the price competition between two firms. Each firm supplies one kind of product. Their prices are completely observed by consumers but their qualities are private information of the firm. Each consumer has his default option, which is one of the two product. The interpretation is the new product by the firm whose old product is used by the consumer. Each consumer can receive information about his default product without cost. If he embraces the cost of survey, he can receive additional information about the quality of non-default product. We investigate the effects of the ratio of default consumers and the survey cost to the behaviors of the consumers and the firms.

The construction of this chapter is as follows. In section 3.2, we pose a formal model of a market and check some basic property of the model. We consider two kinds of separating equilibria. In section 3.3.1, we show that there is no separating equilibrium over qualities. In section 3.3.2, we show that a separating equilibrium over costs can exist and study the property of the equilibrium.

3.2 Model

There are two firms whom indexed by 1 and 2. Firm $i = 1, 2$ sells its product with quality $q_i \in \{q_L, q_H\}$ (where $q_L < q_H$). The cost for firm i to sell one unit of goods is c_i. The quality and the cost are exogenously given by the restriction of technology, which is private information of firm i. Given them, each firm i chooses price p_i. Ordinal price competition models consider a continuous choice with using marginal analysis. However, we focus on discrete price choice since the objective function of firms in our model is not differentiable and we cannot enjoy the benefit of marginal analysis so much.

Consumers cannot observe directly the qualities of products and are risk averse for them. Each consumer initially thinks one of products as default. Let γ_i be the proportion of consumers whose default is product i. Though consumers cannot observe qualities, consumers whose default is i receive a signal $\sigma_i \in \{\sigma_H, \sigma_L\}$ about q_i exogenously. If a consumer surveys about the quality of the product that is not her default, she can receive an additional quality signal of it, where it costs her $K > 0$. For simplicity, we assume that signals are common among consumers that receive them regardless of the timing. This is alternatively implicated that consumers receive signals independently from identical distributions. Additionally, we assume that all consumers buy exactly one product. The prior distribution of qualities for any firm i is $\Pr[q_i = q_H] = a_q \in (0, 1)$ and signals are informative, that is, $\Pr[\sigma_H \mid q_H] \equiv r_{HH} > 1/2$ and $\Pr[\sigma_L \mid q_L] \equiv r_{LL} > 1/2$. The prior distribution of cost for any firm i is $\Pr[c_H] = a_c \in (0, 1)$.

The timing of the game is as follows.

1. Each firm learns its quality of the product and chooses price simultaneously.
2. Each consumer observes prices of all products and the quality of her default product.
3. Consumers decide whether she surveys the quality of another product. If she did, she receives its signal.
4. Consumers decide which product to buy.

Consumers evaluate the qualities of products differently for each. $\theta \in [0,1]$ denote the valuations. We assume that the expected utility functions of consumers are quasi-linear, that is,

$$u(x;\theta) = \mathbb{E}[\theta\sqrt{q_i} - p_i \mid \sigma_i, p] = \theta\mathbb{E}[\sqrt{q_i} \mid \sigma_i, p] - p_i$$

if the default product of a consumer is i and she buy it and

$$u(x_j;\theta) = \mathbb{E}[\theta\sqrt{q_j} - p_j \mid p] = \theta\mathbb{E}[\sqrt{q_j} \mid p] - p_j$$

if she buy product j without searching the quality, where $p = (p_1, p_2)$. For a simplicity, we assume that each consumer chooses default product if $u(x_i;\theta) = u(x_j;\theta)$.

3.3 Equilibria

3.3.1 Separation over qualities

Consider the following strategy of the firms:

$$p(q,c) = \begin{cases} p_H & \text{if } q = q_H \\ p_L & \text{if } q = q_L, \end{cases}$$

where $p_H > p_L$. When the firms follows this strategy, consumers can distinguish the quality of the firm. Then no consumer inquires about the quality of non-default product. In this case, default has a meaning on the decision when the both products are indifference.

We can derive a demand for each pair of price easily. Let $D_i(p_i, p_j)$ be the demand for the product i when the firms i and j choose prices p_i and p_j, respectively. If the both firms choose the same price, the demand for arbitrary product i is γ_i since the same prices implies the same quality. If the firm i chooses p_H and the firm j chooses p_L, the condition for consumers to buy the product i is,

$$\theta\sqrt{q_H} - p_H \geqslant \theta\sqrt{q_L} - p_L \iff \theta \geqslant \frac{p_H - p_L}{\sqrt{q_H} - \sqrt{q_L}}.$$

Note that the right hand side of this inequality is not necessary smaller than 1. Thus the demand for the product i is $1 - \min\left\{\frac{p_H - p_L}{\sqrt{q_H} - \sqrt{q_L}}, 1\right\}$ and the demand for the product j is $\min\left\{\frac{p_H - p_L}{\sqrt{q_H} - \sqrt{q_L}}, 1\right\}$.

Given these consumers' behaviors, we consider the condition for the firms do not deviate the strategy. Since each firm does not know the quality of the opponent before choosing the price, it thinks of the expected profit with respect to the probability a_q.

In fact, this class of strategy is not an equilibrium.

Proposition 3.3.1. *For any γ_i, the situation that both firms play the strategy*

$$s(q,c) = \begin{cases} p_H & \text{if } q = q_H \\ p_L & \text{if } q = q_L \end{cases}$$

cannot be an equilibrium.

Proof. Fix arbitrary $\gamma_i \in (0,1)$. The firm of type (q_H, c_H) does not deviate if

$$\begin{aligned}&(p_H - c_H)\big[a_q D_i(p_H, p_H) + (1 - a_q)D_i(p_H, p_L)\big] \\ &\geqslant (p_L - c_H)\big[a_q D_i(p_L, p_H) + (1 - a_q)D_i(p_L, p_L)\big].\end{aligned}$$

The left hand side is the expected profit when the firm i chooses p_H and the other side is the expected profit when it chooses p_L. The firm of type (q_L, c_H) does not deviate if

$$\begin{aligned}&(p_L - c_H)\big[a_q D_i(p_L, p_H) + (1-a_q)D_i(p_L, p_L)\big] \\ &\geqslant (p_H - c_H)\big[a_q D_i(p_H, p_H) + (1-a_q)D_i(p_H, p_L)\big].\end{aligned}$$

Getting together these conditions, we have

$$\begin{aligned}&(p_H - c_H)\big[a_q D_i(p_H, p_H) + (1-a_q)D_i(p_H, p_L)\big] \\ &= (p_L - c_H)\big[a_q D_i(p_L, p_H) + (1-a_q)D_i(p_L, p_L)\big] \\ \iff &p_H\big[a_q D_i(p_H, p_H) + (1-a_q)D_i(p_H, p_L)\big] - p_L\big[a_q D_i(p_L, p_H) + (1-a_q)D_i(p_L, p_L)\big] \\ &= c_H\left[a_q\left\{D_i(p_H, p_H) - D_i(p_L, p_H)\right\} + (1-a_q)\left\{D_i(p_H, p_L) - D_i(P_L, p_L)\right\}\right] \qquad (3.1)\end{aligned}$$

By no-deviation conditions for the firm of type (q_H, c_L) and (q_L, c_L), similarly, we can derive that

$$\begin{aligned}&p_H\big[a_q D_i(p_H, p_H) + (1-a_q)D_i(p_H, p_L)\big] - p_L\big[a_q D_i(p_L, p_H) + (1-a_q)D_i(p_L, p_L)\big] \\ &= c_L\left[a_q\left\{D_i(p_H, p_H) - D_i(p_L, p_H)\right\} + (1-a_q)\left\{D_i(p_H, p_L) - D_i(P_L, p_L)\right\}\right]. \qquad (3.2)\end{aligned}$$

Since the left hand sides of (3.1) and (3.2) are identical, it holds that

$$\begin{aligned}&c_H\left[a_q\left\{D_i(p_H, p_H) - D_i(p_L, p_H)\right\} + (1-a_q)\left\{D_i(p_H, p_L) - D_i(P_L, p_L)\right\}\right] \\ &= c_L\left[a_q\left\{D_i(p_H, p_H) - D_i(p_L, p_H)\right\} + (1-a_q)\left\{D_i(p_H, p_L) - D_i(P_L, p_L)\right\}\right] \\ \iff &c_H = c_L.\end{aligned}$$

This contradicts the assumption that $c_H > c_L$. □

3.3.2 Separation over costs

Let us consider the following strategy of the firms:

$$s(q, c) = \begin{cases} p_H & \text{if } c = c_H \\ p_L & \text{if } c = c_L, \end{cases}$$

where $p_H > p_L$. A price chosen by a firm can be a signal of the cost of the firm. On the other hand, it has no information about the quality. Actually, the prediction of the quality by consumers after observing the price is

$$\begin{aligned}\Pr[q_H \mid p_H] &= \frac{a_q a_c}{a_q a_C + (1-a_q)a_c} = a_q \\ \Pr[q_H \mid p_L] &= \frac{a_q(1-a_c)}{a_q(1-a_c) + (1-a_q)(1-a_c)} = a_q\end{aligned}$$

This implies a quality survey may be beneficial for the consumers. In what follows, we derive a demand for each pair of prices. At first, we check up which product is chosen by a consumer whose default is i in each case, that is, a case without a survey, a case with survey and received $\sigma_j = \sigma_H$ and a case with survey and received $\sigma_j = \sigma_L$. Then, we see whether the consumer surveys with comparing expected utility when she does not survey and that when she does.

When $p_i = p_j$, the decisions of the consumers depend only on the expected qualities of products. If a consumer whose default product is i received $\sigma_i = \sigma_H$, she necessarily buy i

since product j's expected quality cannot be strictly larger than i's regardless of whether she surveys the quality of j and what signals she received about j when she surveyed. Therefore, she chooses product i without the survey if $\sigma_i = \sigma_H$. On the other hand, she may choose i even if $\sigma_i = \sigma_L$. This is because both products are indifferent when $\sigma_j = \sigma_L$. In other cases, in which she does not survey j's quality and in which she did survey and received $\sigma_j = \sigma_H$, she buys j. The following condition describe the decision whether the survey is done:

$$\begin{aligned}
&\underbrace{[r_{LH}a_q + r_{LL}(1-a_q)]}_{\Pr[\sigma_L]}\theta\frac{r_{LH}a_q\sqrt{q_H} + r_{LL}(1-a_q)\sqrt{q_L}}{r_{LH}a_q + r_{LL}(1-a_q)} \\
&+ \underbrace{[r_{HH}a_q + r_{HL}(1-a_q)]}_{\Pr[\sigma_H]}\theta\frac{r_{HH}a_q\sqrt{q_H} + r_{HL}(1-a_q)\sqrt{q_L}}{r_{HH}a_q + r_{HL}(1-a_q)} - p_H - K \\
&\geqslant \theta[a_q\sqrt{q_H} + (1-a_q)\sqrt{q_L}] - p_H,
\end{aligned}$$

where $\Pr[\sigma_L]$ and $\Pr[\sigma_H]$ are the prior probabilities of consumers' signal. Actually, this inequality implies $K \leqslant 0$, which contradicts $K > 0$. Therefore, having received $\sigma_i = \sigma_L$, she chooses j without the survey.

Next we consider the case in which the firms choose different prices. Let $p_i = p_H$ and $p_j = p_L$. If a consumer receives $\sigma_i = \sigma_H$, the expected utility of product i is

$$\mathbb{E}[\theta\sqrt{q_i} - p_H \mid \sigma_H] = \theta\frac{r_{HH}a_q\sqrt{q_H} + r_{HL}(1-a_q)\sqrt{q_L}}{r_{HH}a_q + r_{HL}(1-a_q)} - p_H.$$

Thus she buys i when the survey of j's quality is not carried out if

$$\begin{aligned}
&\theta\frac{r_{HH}a_q\sqrt{q_H} + r_{HL}(1-a_q)\sqrt{q_L}}{r_{HH}a_q + r_{HL}(1-a_q)} - p_H \geqslant \theta[a_q\sqrt{q_H} + (1-a_q)\sqrt{q_L}] - p_L \\
\iff &\theta \geqslant \frac{\Pr[\sigma_H](p_H - p_L)}{a_q(1-a_q)(r_{HH} + r_{LL} - 1)(\sqrt{q_H} - \sqrt{q_L})} \equiv \overline{\theta}_{HL}.
\end{aligned}$$

Suppose that the consumer does the survey and receives $\sigma_j = \sigma_H$. In this case, the expectation of quality are identical but product j is strictly less expensive than i, so she buys j. If $\sigma_j = \sigma_L$, the product chosen by the consumer is not apparent. She buys i if

$$\begin{aligned}
&\theta\frac{r_{HH}a_q\sqrt{q_H} + r_{HL}(1-a_q)\sqrt{q_L}}{r_{HH}a_q + r_{HL}(1-a_q)} - p_H \geqslant \theta\frac{r_{LH}a_q\sqrt{q_H} + r_{LL}(1-a_q)\sqrt{q_L}}{r_{LH}a_q + r_{LL}(1-a_q)} - p_L \\
\iff &\theta \geqslant \frac{\Pr[\sigma_H][r_{LH}a_q + r_{LL}(1-a_q)](p_H - p_L)}{a_q(1-a_q)(r_{HH} + r_{LL} - 1)(\sqrt{q_H} - \sqrt{q_L})} \equiv \underline{\theta}.
\end{aligned}$$

Lemma 3.3.1. *It holds that* $\underline{\theta} < \overline{\theta}_{HL}$.

Proof. By the definitions of $\overline{\theta}_{HL} - \underline{\theta}$,

$$\begin{aligned}
\overline{\theta}_{HL} - \underline{\theta} &= \frac{\Pr[\sigma_H](p_H - p_L)}{a_q(1-a_q)(r_{HH} + r_{LL} - 1)(\sqrt{q_H} - \sqrt{q_L})}\left[1 - (r_{LH}a_q + r_{LL}(1-a_q))\right] \\
&= \frac{\Pr[\sigma_H](p_H - p_L)}{a_q(1-a_q)(r_{HH} + r_{LL} - 1)(\sqrt{q_H} - \sqrt{q_L})}\left[r_{HH}a_q + r_{HL}(1-a_q)\right] > 0.
\end{aligned}$$

□

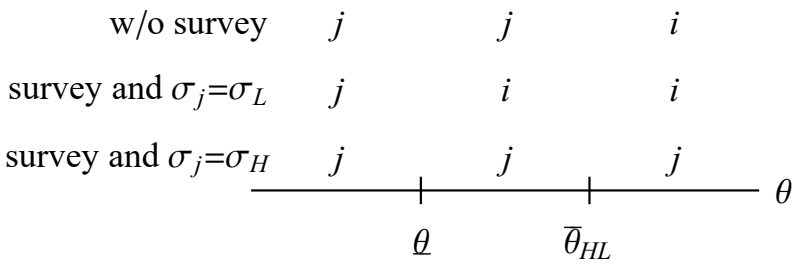

Figure 3.1: Decision by the consumer whose default is i when $p_i = p_H$ and $p_j = p_L$.

Figure 3.1 depicts the consumption by the consumer whose default is i. We can see that consumers with $\theta < \underline{\theta}$ do not survey since the choice does not change between before and after the survey. Intuitively, these consumers care of the price rather than the quality. Consumers with θ such that $\underline{\theta} < \theta < \overline{\theta}_{HL}$ may change their choice after observing the result of survey. Following this decision rule, their expected utilities of survey are

$$\begin{aligned}
&[r_{HH}a_q + r_{HL}(1-a_q)]\left[\theta\frac{r_{HH}a_q\sqrt{q_H} + r_{HL}(1-a_q)\sqrt{q_L}}{r_{HH}a_q + r_{HL}(1-a_q)} - p_L\right] \\
&+ [r_{LH}a_q + r_{LL}(1-a_q)]\left[\theta\frac{r_{HH}a_q\sqrt{q_H} + r_{HL}(1-a_q)\sqrt{q_L}}{r_{HH}a_q + r_{HL}(1-a_q)} - p_H\right] - K \\
&= \theta\frac{r_{HH}a_q\sqrt{q_H} + r_{HL}(1-a_q)\sqrt{q_L}}{r_{HH}a_q + r_{HL}(1-a_q)} \\
&\quad - (r_{HH}a_q + r_{HL}(1-a_q))p_L - (r_{LH}a_q + r_{LL}(1-a_q))p_H - K. \qquad (3.3)
\end{aligned}$$

Then the condition for them to do a survey is

$$\begin{aligned}
&\theta\frac{r_{HH}a_q\sqrt{q_H} + r_{HL}(1-a_q)\sqrt{q_L}}{r_{HH}a_q + r_{HL}(1-a_q)} \\
&\quad - (r_{HH}a_q + r_{HL}(1-a_q))p_L - (r_{LH}a_q + r_{LL}(1-a_q))p_H - K \\
&\geqslant \theta[a_q\sqrt{q_H} + (1-a_q)\sqrt{q_L}] - p_L \\
\iff \theta &\geqslant \frac{\Pr[\sigma_H]\left[\Pr[\sigma_L](p_H - p_L) + K\right]}{a_q(1-a_q)(r_{HH} + r_{LL} - 1)(\sqrt{q_H} - \sqrt{q_L})} \equiv \theta^S_{HL}.
\end{aligned}$$

Lemma 3.3.2. *There are consumers who do a survey in a range $\underline{\theta} < \theta < \overline{\theta}_{HL}$ if and only if $K < (r_{HH}a_q + r_{HL}(1-a_q))(p_H - p_L)$.*

Proof. It should be proven that $\underline{\theta} < \theta^S_{HL} < \overline{\theta}_{HL}$. At first, we show that $\theta^S_{HL} > \underline{\theta}$ for any $K > 0$. Choose arbitrary $K > 0$. By the definitions of θ^S_{HL} and $\underline{\theta}$,

$$\begin{aligned}
\theta^S_{HL} - \underline{\theta} &= \frac{\Pr[\sigma_H]\left[\Pr[\sigma_L](p_H - p_L) + K\right]}{a_q(1-a_q)(r_{HH} + r_{LL} - 1)(\sqrt{q_H} - \sqrt{q_L})} \\
&\quad - \frac{\Pr[\sigma_H]\Pr[\sigma_L](p_H - p_L)}{a_q(1-a_q)(r_{HH} + r_{LL} - 1)(\sqrt{q_H} - \sqrt{q_L})} \\
&= \frac{K}{a_q(1-a_q)(r_{HH} + r_{LL} - 1)(\sqrt{q_H} - \sqrt{q_L})} > 0.
\end{aligned}$$

Next we show that $\theta^S_{HL} < \overline{\theta}_{HL}$ if and only if $K < (r_{HH}a_q + r_{HL}(1-a_q))(p_H - p_L)$. By

the definitions of θ^S_{HL} and $\overline{\theta}_{HL}$, we have

$$\begin{aligned}
&\frac{\Pr[\sigma_H](p_H - p_L)}{a_q(1-a_q)(r_{HH}+r_{LL}-1)(\sqrt{q_H}-\sqrt{q_L})} \\
&> \frac{\Pr[\sigma_H]\left[\Pr[\sigma_L](p_H - p_L) + K\right]}{a_q(1-a_q)(r_{HH}+r_{LL}-1)(\sqrt{q_H}-\sqrt{qL})} \\
\iff & K < (r_{HH}a_q + r_{HL}(1-a_q))(p_H - p_L). \qquad (3.4)
\end{aligned}$$

□

Next we consider the consumers with $\theta \geqslant \overline{\theta}_{HL}$. In this case, the decision after a survey is the same as the one in previous case. Thus the consumer surveys if

$$\begin{aligned}
&\theta\frac{r_{HH}a_q\sqrt{q_H} + r_{HL}(1-a_q)\sqrt{q_L}}{r_{HH}a_q + r_{HL}(1-a_q)} \\
&\quad - (r_{HH}a_q + r_{HL}(1-a_q))p_L - (r_{LH}a_q + r_{LL}(1-a_q))p_H - K \\
&\geqslant \theta\frac{r_{HH}a_q\sqrt{q_H} + r_{HL}(1-a_q)\sqrt{q_L}}{r_{HH}a_q + r_{HL}(1-a_q)} - p_H \\
\iff & K \leqslant \Pr[\sigma_H](p_H - p_L).
\end{aligned}$$

Note that this is identical to (3.4). Hence we can conclude this case as follows.

Lemma 3.3.3. *Assume that $p_i = p_H$, $p_j = p_L$ and $\sigma_i = \sigma_H$. Then there exist consumers whose default is i, who surveys the quality of j if and only if $K \leqslant (r_{HH}a_q + r_{HL}(1 - a_q))(p_H - p_L)$.*

If the condition for K is not satisfied, no consumer does a survey. Especially, then consumers with $\theta \geqslant \overline{\theta}_{HL}$ buy default product, which is more expensive than non-default one. This means the height of survey cost leads them to inertial behavior. Additionally, since the threshold depends on the difference between p_H and p_L, the inertial behavior is likely to occur in a market in which firms choose similar prices.

Let $\sigma_i = \sigma_L$. In this case, the expectation of quality of product i is lower than that of j that is neutrally predicted. In addition, product i is expensive than j. Therefore, the consumers whose default is i choose j without a survey.

Similar discussion holds for the case in which $p_i = p_L$ and $p_j = p_H$. The consumption decision is depicted in Figure 3.2, where $\overline{\theta}_{LH}$ is defined to be

$$\overline{\theta}_{LH} \equiv \frac{\Pr[\sigma_L](p_H - p_L)}{a_q(1-a_q)(r_{HH}+r_{LL}-1)(\sqrt{q_H}-\sqrt{q_L})},$$

which is the solution of the equation that i and j are indifferent when a survey is not carried out. The definition of $\underline{\theta}$ is the same as above and it satisfies the following property.

Lemma 3.3.4. *It holds that $\overline{\theta}_{LH} > \underline{\theta}$.*

Proof. By the definitions of $\overline{\theta}_{LH}$ and $\underline{\theta}$, their difference is

$$\begin{aligned}
\overline{\theta}_{LH} - \underline{\theta} &= \frac{\Pr[\sigma_L](p_H - p_L)}{a_q(1-a_q)(r_{HH}+r_{LL}-1)(\sqrt{q_H}-\sqrt{q_L})} \\
&\quad - \frac{\Pr[\sigma_H]\Pr[\sigma_L](p_H - p_L)}{a_q(1-a_q)(r_{HH}+r_{LL}-1)(\sqrt{q_H}-\sqrt{q_L})} \\
&= \frac{(1-\Pr[\sigma_H])\Pr[\sigma_L](p_H - p_L)}{a_q(1-a_q)(r_{HH}+r_{LL}-1)(\sqrt{q_H}-\sqrt{q_L})}.
\end{aligned}$$

Since $\Pr[\sigma_H]$ is smaller than 1, this difference is strictly positive. □

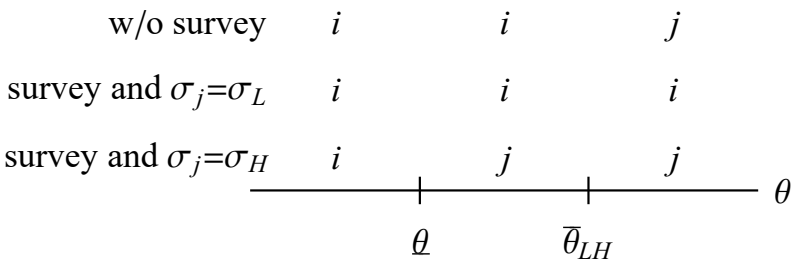

Figure 3.2: Decision by the consumer whose default is i when $p_i = p_L$ and $p_j = p_H$.

If $\theta \leqslant \underline{\theta}$, the consumer chooses i without a survey since her choice does not change before and after the survey. For θ such that $\underline{\theta} < \theta \leqslant \overline{\theta}_{LH}$, consumers do a survey if $\theta \geqslant \theta^S_{LH}$, where

$$\theta^S_{LH} \equiv \frac{\Pr[\sigma_L]\big[\Pr[\sigma_H](p_H - p_L) + K\big]}{a_q(1-a_q)(r_{HH} + r_{LL} - 1)(\sqrt{q_H} - \sqrt{q_L})}.$$

θ^S_{LH} is always bigger than $\underline{\theta}$ and it is smaller than $\overline{\theta}_{LH}$ if and only if $K \leqslant (r_{LH}a_q + r_{LL}(1-a_q))(p_H - p_L)$. This is equivalent to the condition for the existence of consumers who survey when $\theta \geqslant \overline{\theta}_{LH}$. Thus, as in lemma 3.3.3, the following lemma holds.

Lemma 3.3.5. *Assume that $p_i = p_L$, $p_j = p_H$ and $\sigma_i = \sigma_L$. Then there exist consumers whose default is i, who surveys the quality of j if and only if $K \leqslant (r_{LH}a_q + r_{LL}(1 - a_q))(p_H - p_L)$.*

Let us see the relations between $\overline{\theta}_{HL}$ and $\overline{\theta}_{LH}$ and between θ^R_{HL} and θ^R_{LH}.

Lemma 3.3.6. *For any $K > 0$, the following conditions are equivalent.*

(i) $\overline{\theta}_{HL} > \overline{\theta}_{LH}$

(ii) $\theta^S_{HL} > \theta^S_{LH}$

(iii) $\Pr[\sigma_H] > \Pr[\sigma_L]$

Proof. Suppose that $\overline{\theta}_{HL} > \overline{\theta}_{LH}$. This implies that

$$\begin{aligned}\overline{\theta}_{HL} - \overline{\theta}_{LH} &= \frac{(p_H - p_L)\{\Pr[\sigma_H] - \Pr[\sigma_L]\}}{a_q(1-a_q)(r_{HH} + r_{LL} - 1)(\sqrt{q_H} - \sqrt{q_L})} > 0 \\ &\iff \Pr[\sigma_H] > \Pr[\sigma_L].\end{aligned}$$

Thus (i) and (iii) are equivalent. Next suppose that $\theta^S_{HL} > \theta^S_{LH}$. This implied that

$$\begin{aligned}\theta^S_{HL} - \theta^S_{LH} &= \frac{K}{a_q(1-a_q)(r_{HH} + r_{LL} - 1)(\sqrt{q_H} - \sqrt{q_L})} \\ &\quad \times \{(r_{HH}a_q + r_{HL}(1-a_q)) - (r_{LH}a_q + r_{LL}(1-a_q))\} > 0 \\ &\iff r_{LL}(1-a_q) - r_{LH}a_q < \frac{1}{2}.\end{aligned}$$

Thus (ii) and (iii) are equivalent. Therefore, (i) and (ii) are equivalent. □

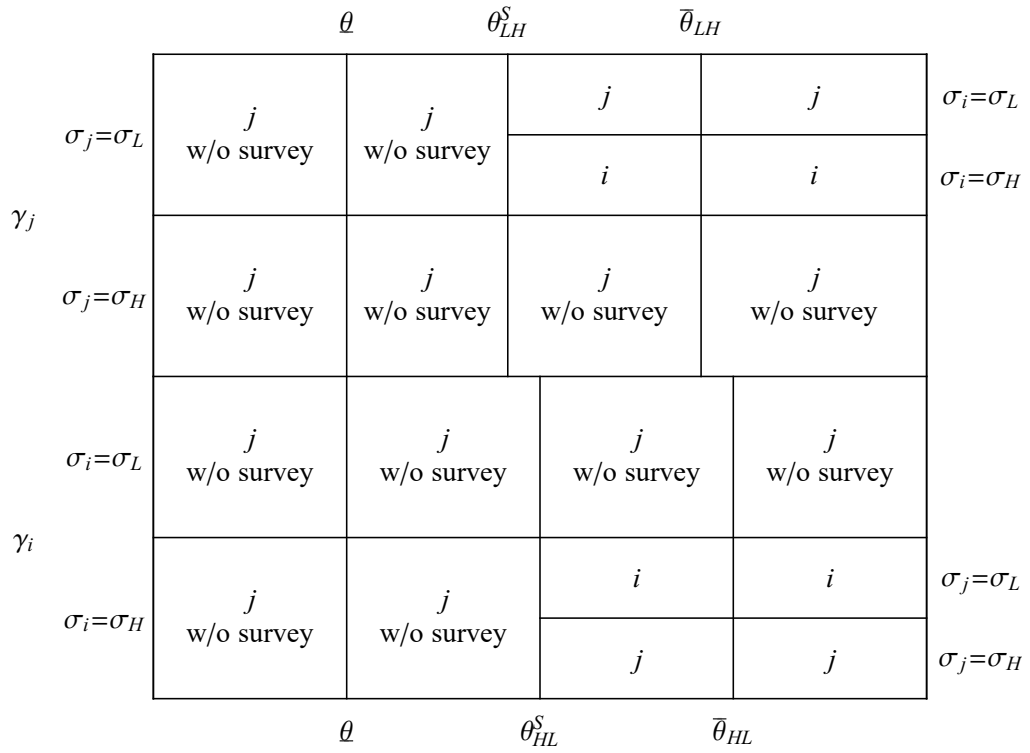

Figure 3.3: Choices by consumes when K is sufficiently small, $p_i = p_H$ and $p_j = p_L$.

The left hand side of (iii) in the statement is a probability with which consumers receives σ_L. That is, if the consumers expect that the survey is likely to bring them σ_H, θ^S_{LH}, the threshold of survey for consumers whose default is low price product, is lower than θ^S_{HL}.

Firms choose their price expecting the demand in light of the consumer behaviors discussed above. At first, we derive an expected demand when $p_i = p_j$. In this case, the decisions of consumers depend only on the signals they received as default. Among consumers whose default is i, those who received $\sigma_i = \sigma_H$ choose i. And among consumers whose default is j, those who receive $\sigma_j = \sigma_L$ choose j. Since firm i knows its quality, the prediction of its signal as a conditional probability, that is, $\Pr[\sigma_H \mid q_i] = r_{HH}$ if $q_i = q_H$ and $\Pr[\sigma_L \mid q_i] = r_{LH}$ if $q_i = q_L$. On the other hand, the prediction of signal of opponent cannot be updated since the quality of the opponent is unknown. Therefore, demands in this case is derived as follows

$$D_i(p_H, p_H; q_i) = D_i(p_L, p_L; q_i) = \gamma_i \Pr[\sigma_H \mid q_i] + \gamma_j \Pr[\sigma_L].$$

Next we derive demands when $p_i \neq p_j$. Figure 3.3 is useful to see the consumers' behavior. Lower (resp. upper) part of the figure corresponds to consumers whose default is product i (resp. j). Each of them are divided by a default signal received. Additionally, consumers who does a survey are divided by a signal received as a result of the survey. We can see that $\underline{\theta}$ itself does not divide the behavior of any consumers. It is, actually, the common lower bound of θ^S_{HL} and θ^S_{LH}.

Since θ is uniformly distributed, demands are calculated as areas of corresponding parts. Note that each threshold of θ is not necessarily smaller than 1. Thus, the demands

of product i when $p_i \neq p_j$ is

$$\begin{aligned}
D_i(p_H, p_L; q_i) &= \gamma_i \Pr[\sigma_H \mid q_i] \Pr[\sigma_L](1 - \min\{\theta^S_{HL}, 1\}) \\
&\quad + \gamma_j \Pr[\sigma_L] \Pr[\sigma_H \mid q_i] \left(1 - \min\{\theta^S_{LH}, 1\}\right) \\
&= \Pr[\sigma_H \mid q_i] \Pr[\sigma_L] \left(1 - \gamma_i \min\{\theta^S_{HL}, 1\} - \gamma_j \min\{\theta^S_{LH}, 1\}\right) \\
D_i(p_L, p_H; q_i) &= 1 - \gamma_i \Pr[\sigma_L \mid q_i] \Pr[\sigma_H](1 - \min\{\theta^S_{LH}, 1\}) \\
&\quad - \gamma_j \Pr[\sigma_H] \Pr[\sigma_L \mid q_i](1 - \min\{\theta^S_{HL}, 1\}) \\
&= 1 - \Pr[\sigma_L \mid q_i] \Pr[\sigma_H] \left(1 - \gamma_i \min\{\theta^S_{LH}, 1\} - \gamma_j \min\{\theta^S_{HL}, 1\}\right).
\end{aligned}$$

At first, We consider the case of $\theta^S_{HL} > \theta^S_{LH}$. We derive relationship of p_L and p_H that can let the strategy be an equilibrium for each case, $1 \leqslant \theta^S_{LH}$, $\theta^S_{LH} < 1 \leqslant \theta^S_{HL}$ and $\theta^S_{HL} < 1$.

Firms deviate the strategy if the deviation yields them more expected profit. Thus conditions for firm i with type (q_t, c_H) $(t \in \{H, L\})$ not to deviate is written as follows [3]:

$$\begin{aligned}
&(p_H - c_H) \left[a_c D_i(p_H, p_H; q_t) + (1 - a_c) D_i(p_H, p_L; q_t)\right] \\
&\geqslant (p_L - c_H) \left[a_c D_i(p_L, p_H; q_t) + (1 - a_c) D_i(p_L, p_L; q_t)\right], \\
\iff &p_H \left[a_c D_i(p_H, p_H; q_t) + (1 - a_c) D_i(p_H, p_L; q_t)\right] \\
&\qquad - p_L \left[a_c D_i(p_L, p_H; q_t) + (1 - a_c) D_i(p_L, p_L; q_t)\right] \\
&\geqslant c_H \left[(1 - a_c) D_i(p_H, p_L; q_t) - a_c D_i(p_L, p_H; q_t) + (2a_c - 1) D_i(p_H, p_H; q_t)\right], \qquad (3.5)
\end{aligned}$$

and the condition for firm i with type (q_t, c_L) is

$$\begin{aligned}
&(p_L - c_L) \left[a_c D_i(p_L, p_H; q_t) + (1 - a_c) D_i(p_L, p_L; q_t)\right] \\
&\geqslant (p_H - c_L) \left[a_c D_i(p_H, p_H; q_t) + (1 - a_c) D_i(p_H, p_L; q_t)\right] \\
\iff &p_H \left[a_c D_i(p_H, p_H; q_t) + (1 - a_c) D_i(p_H, p_L; q_t)\right] \\
&\qquad - p_L \left[a_c D_i(p_L, p_H; q_t) + (1 - a_c) D_i(p_L, p_L; q_t)\right] \\
&\leqslant c_L \left[(1 - a_c) D_i(p_H, p_L; q_t) - a_c D_i(p_L, p_H; q_t) + (2a_c - 1) D_i(p_H, p_H; q_t)\right]. \qquad (3.6)
\end{aligned}$$

To simplify, let X_t $(t \in \{H, L\})$ be a part that is multiplied to c_H and c_L in the right hand sides of (3.5) and (3.6), respectively. Here we have a property about expected revenue in the equilibrium.

Proposition 3.3.2. *In the equilibrium, the expected revenue of high cost firm is weakly smaller than that of low cost firm.*

Proof. Combining (3.5) and (3.6), we have

$$\begin{aligned}
c_H X_t &\leqslant p_H \left[a_c D_i(p_H, p_H; q_t) + (1 - a_c) D_i(p_H, p_L; q_t)\right] \\
&\qquad - p_L \left[a_c D_i(p_L, p_H; q_t) + (1 - a_c) D_i(p_L, p_L; q_t)\right] \leqslant c_L X_t.
\end{aligned}$$

By $c_H > c_L$, X_t must be weakly smaller than 0. Middle of the inequality is the difference between expected revenue when $p_i = p_H$, which is chosen by high cost firm, and that when $p_i = p_L$, which is chosen by low cost firm. Because left and right hand sides are negative, the middle is also negative, that is, this inequality requires that low cost farm yields more expected revenue than high cost firm in the equilibrium. □

[3] Note that $D_i(p_H, p_H; q_t) = D_i(p_L, p_L; q_t)$.

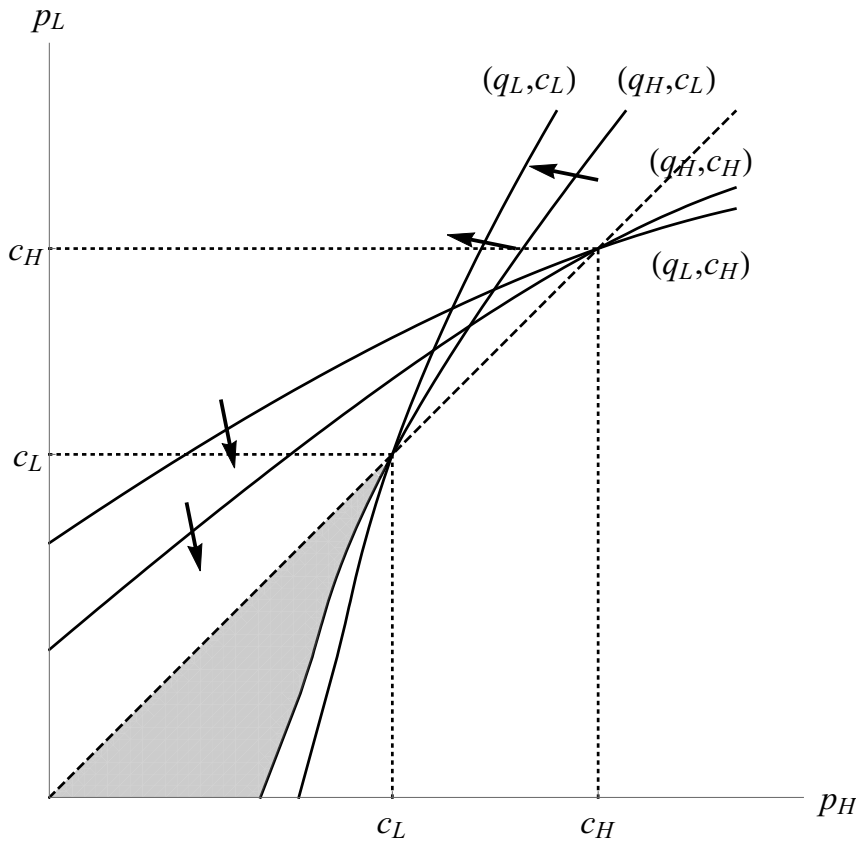

Figure 3.4: Pairs of prices satisfying no-deviation conditions

To obtain the properties of conditions, for t, $s \in \{H, L\}$, we define functions $F_{ts} \colon \mathbb{R}^2 \to \mathbb{R}$ as follows:

$$\begin{aligned} F_{ts}(p_H, p_L) \equiv (p_H - c_s)\left[a_c D_i(p_H, p_H; q_t) + (1 - a_c) D_i(p_H, p_L; q_t)\right] \\ - (p_L - c_s)\left[a_c D_i(p_L, p_H; q_t) + (1 - a_c) D_i(p_L, p_L; q_t)\right] \end{aligned}$$

Then, if $F_{ts}(p_H, p_L) = 0$ for some pairs of prices (p_H, p_L), choosing p_H and choosing p_L are indifferent for the firm with (q_t, c_s). Additionally, let $f_{ts} \colon \mathbb{R}_{++} \to \mathbb{R}$ be a function such that $F_{ts}(p_H, f_{ts}(p_H)) = 0$.

Lemma 3.3.7. *For any t, $s \in \{H, L\}$, f_{ts} has a fixed point $p^*_{ts} = c_s$.*

Proof. Fix arbitrary $t, s \in \{H, L\}$. Since expected demand is non-negative, if $p_H = p_L = p^*_{ts} = c_s$, we have $F(p^*_{ts}, p^*_{ts}) = 0$, that is, $p^*_{ts} = f_{ts}(p^*_{ts})$ by the definition of f_{ts}. Therefore p^*_{ts} is a fixed point of f_{ts}. □

This lemma states that the locus of (p_H, p_L) passes thorough the 45 degree line and the intersection point depends only on the cost.

Figure 3.4 illustrates an example of area in which no-deviation conditions are satisfied. Each four curves represents pairs of prices such that the condition for each type is satisfied with equality, that is, they are f_{ts}. The curves for the same cost crosses on the 45 degree line at the value of the cost because of lemma 3.3.7. Arrows indicates the direction that the condition is satisfied. In shadowed area, all conditions hold. You can see that maximal p_H that can be a equilibrium is $p_H = c_L$. Since $p_H > p_L$, this implies the following proposition.

Proposition 3.3.3. *In this separating equilibrium, no firm can earn positive profit.*

3.4 Conclusion

We have studied consumers' inertial behaviors and their endogenous information acquisitions. The quality and the cost of each firm are private information of the firm. Firms can send a signal about the information as the price. Each consumer has a default product and receives information about the quality of it more than other product. It costs them to receive additional information about non-default product. In this model, we analyzed two kinds of separation equilibria. The first one is separated over qualities. We that this equilibrium does not exist. The second one is separated over costs. It is found that the equilibrium can exist for the specific pairs of prices and no firm can earn positive profit in the equilibrium.

There are some problems. For example, so far it is unapparent that the effect of number of existing customer to firms' behavior. To analyze this point, we will consider a strategy that directly depends on the share of default consumers. Intuitively, it is natural that the firm with large share has strong attitude if the competition is one-shot. However, it is not necessarily obvious since the default consumers are likely to survey when the price is much higher than other product.

Chapter 4

The Economic Analysis of Psychological Reactance

4.1 Introduction

A worker has two projects A and B, and will choose which of these to execute. As a result of consideration in light of his experience in the field, he decided to promote project A. Then his new adviser headhunted recently came and told the worker to promote project A. The worker, who believed that he had more precise information than his adviser, is hostile to the advice and finally chose project B.

Like this example, there are situations in which a final chooser is reactive to persuasion that promote some actions even if they are actions that the chooser decided to choose originally. In psychology, this phenomenon is called *psychological reactance*. From Brehm (1966), literature that proposed this concept, a great number of experiments are done in various situations.

In fact, psychological reactance is broader concept than the example above. Brehm (1966) explain this phenomenon by the restoration of threatened freedom to choose. Note that "threatened" does not necessary mean that it disappears materially but means that some barriers' to choice emergent. In the example, the worker's freedom to choose by his own thought does not disappear but it is threatened by the advice of his adviser.

We study psychological reactance from the viewpoint of economics. To do this, we focus on the situation like the first example. Two players, adviser and chooser, face to the choice of alternative. Though one of this commonly benefits them, both players do not completely know which is the right alternative. Each player values alternatives using noisy signal. No one knows whether the precision of the chooser is higher than the adviser or not. The key of our model is the chooser's preference for high reputation. He wishes to be seen by the adviser that he has high precision. In addition, he values marginal increase of the reputation higher when the reputation is lower than his self-evaluation, that is, he has reference dependent preference, as seen in Tversky and Kahneman (1991), with self-evaluation as reference point. Intuitively, he does not like over-evaluation.

Levy (2004) investigates anti-herding behavior in the situation of career concern. Though the most basic setting of our model is like their model, there are some differences. In our model, we assume that the precisions of the chooser is discrete, whereas Levy (2004) assume continuous precision. We probably be able to extend our model about this point. A next difference is the chooser's knowledge about his precision. Levy (2004) assume that the chooser knows his own precision. In our model, however, it is uncertain even for himself. This implies that the advice of adviser affects not only to the

expectation of the reputation of the chooser's precision, but also his self-evaluation. Then the adviser chooses her action with considering this effect. Additionally, the feature of our model is the reference dependence of evaluation of reputation as stated above.

The construction of this chapter is as follows. In section 4.2, we propose the basic framework of our model. Then in 4.2.1, we investigate an equilibrium in which the reactive behavior of the chooser comes about.

4.2 Discrete Precision Model

We consider a game with two players, an adviser (A) and a chooser (C). There are two alternatives, 0 and 1. Just one of these is a *right* alternative. "Right" means that it yields higher material payoff to both players than the other. $x \in \{0, 1\}$ denotes right alternative and $r_k \in (0, 1)$ denotes a probability with which right alternative is $k \in \{0, 1\}$. Though both do not completely knows which is the right alternative, $i \in \{A, C\}$ receives signals s_i independently. We call them *alternative signals*. The alternative signal precision of the adviser is $q_A \in (1/2, 1)$ and this is common knowledge. q_A represents a probability that adviser receives alternative signal correctly, that is,

$$q_A \equiv \Pr[s_A = x \mid x].$$

On the other hand, the alternative signal precision of chooser, q_C, is $q_H \in (1/2, 1)$ in one case and in other case, it is $q_L \in (1/2, 1)$. The definitions of q_H and q_L are the same as q_A. We assume that $q_L < q_A < q_H$. So the adviser knows that there exists the case in which the chooser has higher precision about the right alternative. For $k \in \{H, L\}$, let $\alpha_k \in (0, 1)$ be the prior probability with which q_C is q_k. Both players do not completely know which precision are realized but they independently receive signals about chooser's precision. We call them *precision signals* and q_{Ci} denotes a precision signal that player $i \in \{A, C\}$ receives. The precision of precision signal of player $i \in \{A, C\}$ is $\varepsilon_i \in (0, 1/2)$. This represents the probability with which player i receives precision signal incorrectly, that is,

$$\varepsilon_i \equiv \Pr[q_{Ci} \neq q_C \mid q_C].$$

We do not consider mixed strategy. The set of actions that the adviser can choose is $\{0, 1, \phi\}$. 0 (1) represents an advice "you had better choose 0 (1)". ϕ represents saying nothing. Let $a \in \{0, 1, \phi\}$ be the action of the adviser. The adviser chooses one action after observing her signals, q_{CA} and s_A. The strategy of the adviser is a function that maps (q_{CA}, s_A) to a.

The set of actions that the chooser can choose is simply $\{0, 1\}$. An outcome depends on the choice $c \in \{0, 1\}$. The chooser chooses one action after observing q_{CC}, s_A and advice a. The strategy of the chooser is a function that maps (q_{CC}, s_A, a) to c.

The adviser is interested in material payoff that is yielded by the chooser's action and right alternative. Let the material payoff when the chooser chose right alternative be 1 and one when he did not choose right alternative be 0. Then the payoff of the adviser is simply the probability with which the chooser chooses right alternative.

The chooser also be interested in material. In addition, he cares about his image by the adviser. He wishes to be seen as he has high precision of alternative signal. He has self-image and he especially wants to avoid to seen lower than it. Formally, let $b_C(a, q_{CC})$ be a self-image of the chooser and this is defined to be,

$$b_c(a, s_C, q_{CC}) \equiv \Pr[q_C = q_H \mid a, s_C, q_{CC}].$$

Note that b_C depends on s_C. Intuitively, let the strategy of the adviser be $a = s_A$, then, if the chooser observes $q_{CC} = q_L$ and $s_C \neq a$, his alternative signal is more likely to be wrong. This causes decreasing in b_C.

Next let $b_A(a, c)$ be the chooser's prediction of "ex-post prediction of adviser's image to the chooser", that is,

$$b_A(a, c) = \Pr[q_C = q_H \mid a, c].$$

There are two points to note. The first is that since b_A depends on c, the chooser can affect the image by the adviser through c. The second is that b_A is independent on s_C, s_A, q_{CC}, and q_{CA}. This is because the adviser cannot use q_{CC} and s_C to make the prediction and the chooser cannot use q_{CA} and s_A to make the prediction of prediction. A function v represents the psychological payoff by the these images.

We define v to be

$$v(b_A, b_C) = \begin{cases} h(b_A - b_C) & \text{if } b_A \geqslant b_C \\ b_A - b_C & \text{otherwise} \end{cases},$$

where $h \in (0, 1)$. That is, v has a loss aversion form with reference point b_C. The whole payoff of the chooser is weighted sum of material and psychological payoff,

$$\Pr[x = c \mid a, c, q_{CC}, s_C] + \beta v(b_A, b_C).$$

4.2.1 Equilibrium

We consider the following strategies:

Adviser:

$$\begin{cases} a = \phi & \text{if } q_{CA} = q_H \\ a = s_A & \text{if } q_{CA} = q_L \end{cases} \tag{4.1}$$

Chooser:

$$\begin{cases} c = s_C & \text{if } a = \phi \\ c = a & \text{if } a \neq \phi \text{ and } q_{CC} = q_L \\ c \neq a & \text{if } a \neq \phi \text{ and } q_{CC} = q_H. \end{cases} \tag{4.2}$$

The adviser advises to the chooser when he has lower precision in order to avoid his wrong choice. On the other hand, if she predicts that the chooser has higher precision, she thinks better to leave the chooser to do what he likes. If the chooser follows this strategy, there exists a case in which he chooses $c \neq s_A$ even if he chooses the same action as s_A if the adviser says nothing. This is the psychological reactance. We show that this pair of strategies can be an equilibrium.

Consider the adviser's behavior. She is interested in the subjective probability with which the chooser chooses right alternative. For example, the probability when she observed $q_{CA} = q_L$ and $s_A = 0$ and she chose $a = 0$ is

$$\Pr[c = x \mid q_{CA} = q_L, s_A = 0, a = 0] = \frac{\begin{array}{c}\alpha_L(1-\varepsilon_A)(1-\varepsilon_C)r_0 q_A + \alpha_L(1-\varepsilon_A)\varepsilon_C r_1(1-q_A) \\ +\alpha_H \varepsilon_A \varepsilon_C r_0 q_A + \alpha_H \varepsilon_A (1-\varepsilon_C) r_1 (1-q_A)\end{array}}{\begin{array}{c}\alpha_L(1-\varepsilon_A)r_0 q_A + \alpha_L(1-\varepsilon_A)r_1(1-q_A) \\ +\alpha_H \varepsilon_A r_0 q_A + \alpha_H \varepsilon_A r_1 (1-q_A)\end{array}}.$$

When the chooser follows the above strategy, he chooses 0 if $q_{CC} = q_L$ and 1 if $q_{CC} = q_H$. The first and second terms in the denominator are the sum of the probabilities with which the true precision of the chooser is q_L and the adviser observes it correctly. The third and fourth terms in the denominator correspond to the case of the true precision is q_H and the adviser observes it incorrectly. Each term in the numerator contains the factor of the observation of the chooser, ε_C and $(1 - \varepsilon_C)$.

If the adviser chooses $a = \phi$, the choice of the chooser depends on the alternative signal he observes. For example, the subjective probability with which the chooser chooses right alternative when $q_{CA} = q_H$, $s_A = 0$ and $a = 0$ is,

$$\Pr[c = x \mid q_{CA} = q_H, s_A = 0, a = \phi] = \frac{\begin{array}{c}\alpha_L\varepsilon_A r_0 q_A q_L + \alpha_L\varepsilon_A r_1(1-q_A)q_L \\ +\alpha_H(1-\varepsilon_A)r_0 q_A q_H + \alpha_H(1-\varepsilon_A)r_1(1-q_A)q_H\end{array}}{\begin{array}{c}\alpha_L\varepsilon_A r_0 q_A + \alpha_L\varepsilon_A r_1(1-q_A) \\ +\alpha_H(1-\varepsilon_A)r_0 q_A + \alpha_H(1-\varepsilon_A)r_1(1-q_A)\end{array}}.$$

Henceforth, we assume that $r_0 = r_1 = 1/2$, that is, there is no bias in prior of right alternative. For simplicity, we define $\delta \equiv \varepsilon_C(1 - q_A) + (1 - \varepsilon_C)q_A$. The, obviously, $(1 - \varepsilon_C)(1 - q_A) + \varepsilon_C q_A = 1 - \delta$. In addition, the following property is satisfied.

Lemma 4.2.1. *For any $q_A \in (1/2, q_H)$ and $\varepsilon_C \in (0, 1/2)$, $1/2 < \delta < q_H$.*

Proof. Choose $\varepsilon_C \in (0, 1/2)$ arbitrarily. Then the derivative of δ for q_A is $1 - 2\varepsilon_C > 0$. Easily we can check that

$$\lim_{q_A \to 1/2} \left(\frac{1}{2} - \delta \right) = 0$$

and

$$\lim_{q_A \to q_H} (q_H - \delta) = 2q_H - \varepsilon_C > 0.$$

Therefore, we have $1/2 < \delta < q_H$ for any $q_A \in (1/2, q_H)$. □

Using this notation and assuming $r_0 = r_1 = 1/2$, the probabilities above are rewritten as

$$\Pr[c = x \mid q_{CA}, s_A = 0, a = 0] = \frac{\alpha_L(1-\varepsilon_A)\delta + \alpha_H\varepsilon_A(1-\delta)}{\alpha_L(1-\varepsilon_A) + \alpha_H\varepsilon_A}$$

$$\Pr[c = x \mid q_{CA} = q_H, s_A = 0, a = \phi] = \frac{\alpha_L\varepsilon_A q_L + \alpha_H(1-\varepsilon_A)q_H}{\alpha_L\varepsilon_A + \alpha_H(1-\varepsilon_A)}.$$

The probability of right choice under other conditions are in appendix A.1.

The conditions for the adviser's obeying the strategy are as follows:

$$\Pr[c = x \mid q_{CA} = q_H, s_A = 0, a = \phi] \geqslant \Pr[c = x \mid q_{CA} = q_H, s_A = 0, a = 0] \tag{4.3}$$
$$\Pr[c = x \mid q_{CA} = q_H, s_A = 0, a = \phi] \geqslant \Pr[c = x \mid q_{CA} = q_H, s_A = 0, a = 1] \tag{4.4}$$
$$\Pr[c = x \mid q_{CA} = q_H, s_A = 1, a = \phi] \geqslant \Pr[c = x \mid q_{CA} = q_H, s_A = 1, a = 0] \tag{4.5}$$
$$\Pr[c = x \mid q_{CA} = q_H, s_A = 1, a = \phi] \geqslant \Pr[c = x \mid q_{CA} = q_H, s_A = 1, a = 1] \tag{4.6}$$
$$\Pr[c = x \mid q_{CA} = q_L, s_A = 0, a = 0] \geqslant \Pr[c = x \mid q_{CA} = q_L, s_A = 0, a = 1] \tag{4.7}$$
$$\Pr[c = x \mid q_{CA} = q_L, s_A = 0, a = 0] \geqslant \Pr[c = x \mid q_{CA} = q_L, s_A = 0, a = \phi] \tag{4.8}$$
$$\Pr[c = x \mid q_{CA} = q_L, s_A = 1, a = 1] \geqslant \Pr[c = x \mid q_{CA} = q_L, s_A = 1, a = 0] \tag{4.9}$$
$$\Pr[c = x \mid q_{CA} = q_L, s_A = 1, a = 1] \geqslant \Pr[c = x \mid q_{CA} = q_L, s_A = 1, a = \phi] \tag{4.10}$$

Note that the prior probability of right alternative is neutral and 0 and 1 yields the same utility as long as it is the right, we can use some kind of symmetry on these conditional probabilities. Actually, (4.3) and (4.6), (4.4) and (4.5), (4.7) and (4.9), (4.8) and (4.10) are equivalent, respectively. Thus, we consider (4.3), (4.4), (4.7) and (4.8).

At first, substituting the probabilities in the both sides of (4.3), we have,

$$\begin{aligned}
&\Pr[c = x \mid q_{CA} = q_H, s_A = 0, a = \phi] \geqslant \Pr[c = x \mid q_{CA} = q_H, s_A = 0, a = 0] \\
\iff &\frac{\alpha_L \varepsilon_A q_L + \alpha_H (1 - \varepsilon_A) q_H}{\alpha_L \varepsilon_A + \alpha_H (1 - \varepsilon_A)} \geqslant \frac{\alpha_L \varepsilon_A \delta + \alpha_H (1 - \varepsilon_A)(1 - \delta)}{\alpha_L \varepsilon_A + \alpha_H (1 - \varepsilon_A)} \\
\iff &\alpha_L \varepsilon_A q_L + \alpha_H (1 - \varepsilon_A) q_H \geqslant \alpha_L \varepsilon_A \delta + \alpha_H (1 - \varepsilon_A)(1 - \delta) \\
\iff &\frac{\alpha_L \varepsilon_A}{\alpha_H (1 - \varepsilon_A)} \leqslant \frac{\delta - (1 - q_H)}{\delta - q_L}.
\end{aligned} \tag{4.11}$$

Similarly, (4.4) is,

$$\begin{aligned}
&\Pr[c = x \mid q_{CA} = q_H, s_A = 0, a = \phi] \geqslant \Pr[c = x \mid q_{CA} = q_H, s_A = 0, a = 1] \\
\iff &\frac{\alpha_L \varepsilon_A q_L + \alpha_H (1 - \varepsilon_A) q_H}{\alpha_L \varepsilon_A + \alpha_H (1 - \varepsilon_A)} \geqslant \frac{\alpha_L \varepsilon_A (1 - \delta) + \alpha_H (1 - \varepsilon_A) \delta}{\alpha_L \varepsilon_A + \alpha_H (1 - \varepsilon_A)} \\
\iff &\alpha_L \varepsilon_A q_L + \alpha_H (1 - \varepsilon_A) q_H \geqslant \alpha_L \varepsilon_A (1 - \delta) + \alpha_H (1 - \varepsilon_A) \delta \\
\iff &\frac{\delta - q_H}{\delta - (1 - q_L)} \leqslant \frac{\alpha_L \varepsilon_A}{\alpha_H (1 - \varepsilon_A)}.
\end{aligned}$$

Since the left hand side of this is strictly negative by lemma 4.2.1 and the other side is strictly positive, we can ignore this condition. Hence (4.11) is the condition for the adviser to interfere the chooser when she observed $q_{CA} = q_H$. α_H and q_H represent how much the chooser is worse to trust, so increases of these make the condition be easy to be satisfied.

Next we check up conditions for the adviser to advise the chooser when she observed $q_{CA} = q_L$. (4.7) is

$$\begin{aligned}
&\Pr[c = x \mid q_{CA} = q_L, s_A = 0, a = 0] \geqslant \Pr[c = x \mid q_{CA} = q_L, s_A = 0, a = 1] \\
\iff &\frac{\alpha_L (1 - \varepsilon_A) \delta + \alpha_H \varepsilon_A (1 - \delta)}{\alpha_L (1 - \varepsilon_A) + \alpha_H \varepsilon_A} \geqslant \frac{\alpha_L (1 - \varepsilon_A)(1 - \delta) + \alpha_H \varepsilon_A \delta}{\alpha_L (1 - \varepsilon_A) + \alpha_H \varepsilon_A} \\
\iff &\frac{\alpha_H \varepsilon_A}{\alpha_L (1 - \varepsilon_A)} \leqslant 1.
\end{aligned} \tag{4.12}$$

And (4.8) is

$$\begin{aligned}
&\Pr[c = x \mid q_{CA} = q_L, s_A = 0, a = 0] \geqslant \Pr[c = x \mid q_{CA} = q_L, s_A = 0, a = \phi] \\
\iff &\frac{\alpha_L (1 - \varepsilon_A) \delta + \alpha_H \varepsilon_A (1 - \delta)}{\alpha_L (1 - \varepsilon_A) + \alpha_H \varepsilon_A} \geqslant \frac{\alpha_L (1 - \varepsilon_A) q_L + \alpha_H \varepsilon_A q_H}{\alpha_L (1 - \varepsilon_A) + \alpha_H \varepsilon_A} \\
\iff &\frac{\alpha_H \varepsilon_A}{\alpha_L (1 - \varepsilon_A)} \leqslant \frac{\delta - q_L}{\delta - (1 - q_H)}.
\end{aligned} \tag{4.13}$$

In fact, (4.13) is the sufficient condition of (4.12) since $q_L > 1 - q_H$ by the assumption. Thus, the adviser tells her alternative signal to the chooser when (4.13) is satisfied and she observed $q_{CA} = q_L$. Large α_L and small q_L mean that the chooser tends to have low precision and the adviser will have an incentive to advise.

The following lemma summarizes above.

Lemma 4.2.2. *The adviser follows the strategy if and only if*

$$\frac{\alpha_L \varepsilon_A}{\alpha_H (1 - \varepsilon_A)} \leqslant \frac{\delta - (1 - q_H)}{\delta - q_L} \text{ and } \frac{\alpha_H \varepsilon_A}{\alpha_L (1 - \varepsilon_A)} \leqslant \frac{\delta - q_L}{\delta - (1 - q_H)}. \tag{4.14}$$

The following definitions simplifies the notations.

$$\begin{aligned}
&\sigma_H \equiv \alpha_H \varepsilon_A, \ \ \sigma_L \equiv \alpha_L(1-\varepsilon_A) \\
&q_{ij} \equiv q_i q_j + (1-q_i)(1-q_j) \\
&\bar{q}_{ij} \equiv q_i q_j - (1-q_i)(1-q_j) = q_i + q_j - 1 \\
&q'_{ij} \equiv q_i(1-q_j) + (1-q_i)q_j \\
&\bar{q}'_{ij} \equiv q_i(1-q_j) - (1-q_i)q_j = q_i - q_j
\end{aligned}$$

Lemma 4.2.3. *$q_{AH} > q_{AL}$ and $q'_{AL} > q'_{AH}$ if $q_H > q_L$.*

Proof. By the definitions of q_{AH} and q_{AL},

$$\begin{aligned}
q_{AH} - q_{AL} &= q_A q_H + (1-q_A)(1-q_H) - \{q_A q_L - (1-q_A)(1-q_L)\} \\
&= q_A(q_H - q_L) + (1-q_A)(q_H - q_L) = q_H - q_L > 0,
\end{aligned}$$

and

$$\begin{aligned}
q'_{AL} - q'_{AH} &= q_A(1-q_L) + (1-q_A)q_L - \{q_A(1-q_H) + (1-q_A)q_H\} \\
&= q_A(q_H - q_L) - (1-q_A)(q_H - q_L) = (2q_A - 1)(q_H - q_L) > 0
\end{aligned}$$

since $q_A > 1/2$. □

From here, we consider the condition that assures that the chooser follows the strategy. To see the decision of the chooser, we need to check up three predictions by the chooser: which is the right alternative, how much precise he is (b_C) and what prediction the adviser will have on the chooser's precision after observing his choice (b_A). Information that the chooser can use are q_{CC}, s_C and a. Under these information, the predictions are formed.

Consider the prediction of the right alternative. As an example, the probability of $x = 0$ after observing $a = 0$, $s_C = 0$ and $q_{CC} = q_H$ is,

$$\begin{aligned}
&\Pr[x = 0 \mid x = 0, a = 0, s_C = 0, q_{CC} = q_H] \\
&= \frac{\alpha_H \varepsilon_A (1-\varepsilon_C) r_0 q_A q_H + \alpha_L (1-\varepsilon_A)\varepsilon_C r_0 q_A q_L}{\alpha_H \varepsilon_A (1-\varepsilon_C) r_0 q_A q_H + \alpha_H \varepsilon_A (1-\varepsilon_C) r_1 (1-q_A)(1-q_H)} \\
&\qquad + \alpha_L (1-\varepsilon_A)\varepsilon_C r_0 q_A q_L + \alpha_L (1-\varepsilon_A)\varepsilon_C r_1 (1-q_A)(1-q_L) \\
&= \frac{\sigma_H (1-\varepsilon_C) q_A q_H + \sigma_L \varepsilon_C q_A q_L}{\sigma_H (1-\varepsilon_C) q_{AH} + \sigma_L \varepsilon_C q_{AL}}.
\end{aligned}$$

Note that the advice $a \neq \phi$ implies that $q_{CA} = q_L$ and $s_A = a$ in the equilibrium. On the other hand, $a = \phi$ is a signal that passes along only $q_{CA} = q_H$. The predictions under other conditions are in appendix A.2.

Next we consider the posterior of $q_C = q_H$ after observing (a, s_C, q_{CC}). We write this as $b_C(a, s_C, H)$ ($b_C(a, s_C, L)$) if $q_{CC} = q_H$ ($q_{CC} = q_L$). We can interpret this as a self-confidence of the chooser. It is worth to note that this depends on s_C and s_A indirectly. For example, when the chooser observed $q_{CC} = q_L$, $s_C = 0$ and $a = 0$, he may revise up the posterior since the adviser, who has comparatively high precision, observed the same alternative signal as him. In this case, the posterior is,

$$\begin{aligned}
&b_C(0, 0, L) \\
&= \frac{\alpha_H \varepsilon_A \varepsilon_C r_0 q_A q_H + \alpha_H \varepsilon_A \varepsilon_C r_1 (1-q_A)(1-q_H)}{\alpha_H \varepsilon_A \varepsilon_C r_0 q_A q_H + \alpha_H \varepsilon_A \varepsilon_C r_1 (1-q_A)(1-q_H)} \\
&\qquad + \alpha_L (1-\varepsilon_A)(1-\varepsilon_C) r_0 q_A q_L + \alpha_L (1-\varepsilon_A)(1-\varepsilon_C) r_1 (1-q_A)(1-q_L) \\
&= \frac{\sigma_H \varepsilon_C q_{AH}}{\sigma_H \varepsilon_C q_{AH} + \sigma_L (1-\varepsilon_C) q_{AL}}.
\end{aligned}$$

The posteriors for other observations are in appendix A.3.

Finally, we derive the chooser's predictions of posterior of q_C by the adviser. These represent the prediction of his feeling by the adviser. Though the chooser cannot use s_A and q_{CA} and the adviser cannot use s_C and q_{CC} directly, their actions, a and c, can play a role as the signals of these information. When $a = 0$, for example, the chooser sees that he is seen to have a low precision by the adviser. To this, he can argue that his precision is high by choosing $c = 1$, which is contrary to the advice. In this case, $b_A(0,0)$ is

$$\begin{aligned} b_A(0,0) = &\frac{\alpha_H \varepsilon_A \varepsilon_C r_0 q_A + \alpha_H \varepsilon_A \varepsilon_C r_1 (1-q_A)}{\alpha_H \varepsilon_A \varepsilon_C r_0 q_A + \alpha_H \varepsilon_A \varepsilon_C r_1 (1-q_A)} \\ &\qquad + \alpha_L (1-\varepsilon_A)(1-\varepsilon_C) r_0 q_A + \alpha_L (1-\varepsilon_A)(1-\varepsilon_C) r_1 (1-q_A) \\ &= \frac{\sigma_H \varepsilon_C}{\sigma_H \varepsilon_C + \sigma_L (1-\varepsilon_C)}. \end{aligned}$$

By the symmetric property of alternatives, $b_A(1,1)$ is the same as this.

Similarly, we can obtain the beliefs in other cases as

$$b_A(0,1) = b_A(1,0) = \frac{\sigma_H (1-\varepsilon_C)}{\sigma_H (1-\varepsilon_C) + \sigma_L \varepsilon_C}.$$

We now check that the reactive behavior of the chooser can revise up the reputation by the adviser.

Lemma 4.2.4. *It holds that* $b_A(0,1) > b_A(0,0)$ *and* $b_A(1,0) > b_A(1,1)$.

Proof. By the assumption, $\varepsilon_C < 1/2$. Then we have

$$\begin{aligned} &1 - 2\varepsilon_C + \varepsilon_C^2 > \varepsilon_C^2 \iff \sigma_H (1-\varepsilon_C)\varepsilon_C + \sigma_L (1-\varepsilon_C)^2 > \sigma_H (1-\varepsilon_C)\varepsilon_C + \sigma_L \varepsilon_C^2 \\ \iff &\frac{\sigma_H (1-\varepsilon_C)}{\sigma_H (1-\varepsilon_C) + \sigma_L \varepsilon_C} > \frac{\sigma_H \varepsilon_C}{\sigma_H \varepsilon_C + \sigma_L (1-\varepsilon_C)}. \end{aligned}$$

Thus, we have $b_A(0,1) > b_A(0,0)$ and $b_A(1,0) > b_A(1,1)$. □

The intuition is as follows. $\varepsilon_C < 1/2$ implies that the chooser is likely to receive his precision signal correctly. The chooser following the strategy chooses reactive choice, that is, $c \neq a$ for $a \neq \phi$ only if he received $q_{CC} = q_H$. Thus his choice plays a role as signal that his precision is high, which revises up the reputation.

We derive the condition under that the chooser follows the strategy. At first, let us consider the condition to choose $c = 1$ for $a = 0$, $s_C = 0$ and $q_{CC} = q_H$. If he chooses $c = 1$,

$$b_A(0,1) - b_C(0,0,H) = -\frac{\sigma_H \sigma_L (1-\varepsilon_C)\varepsilon_C (q_{AH} - q_{AL})}{\big[\sigma_H (1-\varepsilon_C) + \sigma_L \varepsilon_C\big]\big[\sigma_H (1-\varepsilon_C) q_{AH} + \sigma_L \varepsilon_C q_{AL}\big]}.$$

If he deviates and chooses $c = 0$,

$$b_A(0,0) - b_C(0,0,H) = -\frac{\sigma_H \sigma_L \big[(1-\varepsilon_C)^2 q_{AH} - \varepsilon_C^2 q_{AL}\big]}{\big[\sigma_H \varepsilon_C + \sigma_L (1-\varepsilon_C)\big]\big[\sigma_H (1-\varepsilon_C) q_{AH} + \sigma_L \varepsilon_C q_{AL}\big]}.$$

Since both are strictly negative, v values them with a weight β. Then, the condition is

$$\begin{aligned}
&\Pr[x=1 \mid a=0, s_C=0, q_{CC}=q_H] + \beta\big(b_A(0,1) - b_C(0,0,H)\big) \\
&\quad \geqslant \Pr[x=0 \mid a=0, s_C=0, q_{CC}=q_H] + \beta\big(b_A(0,0) - b_C(0,0,H)\big) \\
\iff &\frac{\sigma_H(1-\varepsilon_C)(1-q_A)(1-q_H) + \sigma_L\varepsilon_C(1-q_A)(1-q_L)}{\sigma_H(1-\varepsilon_C)q_{AH} + \sigma_L\varepsilon_C q_{AL}} \\
&\qquad - \beta\frac{\sigma_H\sigma_L(1-\varepsilon_C)\varepsilon_C(q_{AH}-q_{AL})}{\big[\sigma_H(1-\varepsilon_C)+\sigma_L\varepsilon_C\big]\big[\sigma_H(1-\varepsilon_C)q_{AH}+\sigma_L\varepsilon_C q_{AL}\big]} \\
&\quad \geqslant \frac{\sigma_H(1-\varepsilon_C)q_A q_H + \sigma_L\varepsilon_C q_A q_L}{\sigma_H(1-\varepsilon_C)q_{AH} + \sigma_L\varepsilon_C q_{AL}} \\
&\qquad - \beta\frac{\sigma_H\sigma_L\big[(1-\varepsilon_C)^2 q_{AH} - \varepsilon_C^2 q_{AL}\big]}{\big[\sigma_H\varepsilon_C+\sigma_L(1-\varepsilon_C)\big]\big[\sigma_H(1-\varepsilon_C)q_{AH}+\sigma_L\varepsilon_C q_{AL}\big]} \\
\iff &\beta \geqslant \frac{\sigma_H(1-\varepsilon_C)+\sigma_L\varepsilon_C}{\sigma_H\sigma_L} \times \frac{\sigma_H(1-\varepsilon_C)\bar{q}_{AH} + \sigma_L\varepsilon_C\bar{q}_{AL}}{\sigma_H(1-\varepsilon_C)q_{AH}+\sigma_L\varepsilon_C q_{AL}} \times \frac{\sigma_H\varepsilon_C+\sigma_L(1-\varepsilon_C)}{1-2\varepsilon_C}.
\end{aligned} \tag{4.15}$$

The condition to choose $c=1$ for $a=1$, $s_C=0$ and $q_{CC}=q_L$ is the same as this by the symmetry.

Next we consider the condition to choose $c=0$ for $a=0$, $s_C=1$ and $q_{CC}=q_L$. In this case, the deferences of b_A and b_C are

$$\begin{aligned}
b_A(0,1) - b_C(0,1,H) &= \frac{\sigma_H(1-\varepsilon_C)}{\sigma_H(1-\varepsilon_C)+\sigma_L\varepsilon_C} - \frac{\sigma_H(1-\varepsilon_C)q'_{AH}}{\sigma_H(1-\varepsilon_C)q'_{AH}+\sigma_L\varepsilon_C s'_{AL}} \\
&= \frac{\sigma_H(1-\varepsilon_C)\sigma_L\varepsilon_C(q'_{AL}-q'_{AH})}{\big[\sigma_H(1-\varepsilon_C)+\sigma_L\varepsilon_C\big]\big[\sigma_H(1-\varepsilon_C)q'_{AH}+\sigma_L\varepsilon_C q'_{AL}\big]} \\
b_A(0,0) - b_C(0,1,H) &= \frac{\sigma_H\varepsilon_C}{\sigma_H\varepsilon_C+\sigma_L(1-\varepsilon_C)} - \frac{\sigma_H(1-\varepsilon_C)q'_{AH}}{\sigma_H(1-\varepsilon_C)q'_{AH}+\sigma_L\varepsilon_C q'_{AL}} \\
&= \frac{\sigma_H\sigma_L\big[\varepsilon_C^2 q'_{AL} - (1-\varepsilon_C)^2 q'_{AH}\big]}{\big[\sigma_H\varepsilon_C+\sigma_L(1-\varepsilon_C)\big]\big[\sigma_H(1-\varepsilon_C)q'_{AH}+\sigma_L\varepsilon_C q'_{AL}\big]}.
\end{aligned}$$

These are strictly positive, so v values them with weight h. Thus the condition is

$$\begin{aligned}
&\Pr[x=0 \mid a=0, s_C=1, q_{CC}=q_L] + \beta h\big(b_A(0,0) - b_C(0,1,L)\big) \\
&\quad \geqslant \Pr[x=1 \mid a=0, q_C=1, q_{CC}] + \beta h\big(b_A(0,1) - b_C(0,1,L)\big) \\
\iff &\frac{\sigma_H\varepsilon_C q_A(1-q_H) + \sigma_L(1-\varepsilon_C)q_A(1-q_L)}{\sigma_H\varepsilon_C q'_{AH} + \sigma_L(1-\varepsilon_C)q'_{AL}} \\
&\qquad + \beta h\frac{\sigma_H\sigma_L\varepsilon_C(1-\varepsilon_C)(q'_{AL}-q'_{AH})}{\big[\sigma_H\varepsilon_C+\sigma_L(1-\varepsilon_C)\big]\big[\sigma_H\varepsilon_C q'_{AH}+\sigma_L(1-\varepsilon_C)q'_{AL}\big]} \\
&\quad \geqslant \frac{\sigma_H\varepsilon_C(1-q_A)q_H + \sigma_L(1-\varepsilon_C)(1-q_A q_L)}{\sigma_H\varepsilon_C q'_{AH} + \sigma_L(1-\varepsilon_C)q'_{AL}} \\
&\qquad + \beta h\frac{\sigma_H\sigma_L\big[(1-\varepsilon_C)^2 q'_{AL} - \varepsilon_C^2 q'_{AH}\big]}{\big[\sigma_H(1-\varepsilon_C)+\sigma_L\varepsilon_C\big]\big[\sigma_H\varepsilon_C q'_{AH}+\sigma_L(1-\varepsilon_C)q'_{AL}\big]} \\
\iff &\beta \leqslant \frac{\sigma_H(1-\varepsilon_C)+\sigma_L\varepsilon_C}{\sigma_H\sigma_L} \times \frac{\sigma_H\varepsilon_C\bar{q}'_{AH} + \sigma_L(1-\varepsilon_C)\bar{q}'_{AL}}{\sigma_H\varepsilon_C q'_{AH}+\sigma_L(1-\varepsilon_C)q'_{AL}} \\
&\qquad \times \frac{\sigma_H\varepsilon_C+\sigma_L(1-\varepsilon_C)}{1-2\varepsilon_C} \times \frac{1}{h}.
\end{aligned} \tag{4.16}$$

This is the same as the condition to choose $c = 1$ for $a = 1$, $s_C = 0$ and $q_{CC} = q_L$.

Next we consider the condition to choose $c = 1$ for $a = 0$, $s_C = 1$ and $q_{CC} = q_H$. The differences of beliefs are

$$\begin{aligned}
b_A(0,1) - b_C(0,1,H) &= \frac{\sigma_H(1-\varepsilon_C)}{\sigma_H(1-\varepsilon_C)+\sigma_L\varepsilon_C} - \frac{\sigma_H(1-\varepsilon_C)q'_{AH}}{\sigma_H(1-\varepsilon_C)q'_{AH}+\sigma_L\varepsilon_C s'_{AL}} \\
&= \frac{\sigma_H(1-\varepsilon_C)\sigma_L\varepsilon_C(q'_{AL}-q'_{AH})}{\big[\sigma_H(1-\varepsilon_C)+\sigma_L\varepsilon_C\big]\big[\sigma_H(1-\varepsilon_C)q'_{AH}+\sigma_L\varepsilon_C q'_{AL}\big]} \\
b_A(0,0) - b_C(0,1,H) &= \frac{\sigma_H\varepsilon_C}{\sigma_H\varepsilon_C+\sigma_L(1-\varepsilon_C)} - \frac{\sigma_H(1-\varepsilon_C)q'_{AH}}{\sigma_H(1-\varepsilon_C)q'_{AH}+\sigma_L\varepsilon_C q'_{AL}} \\
&= \frac{\sigma_H\sigma_L\big[\varepsilon_C^2 q'_{AL}-(1-\varepsilon_C)^2 q'_{AH}\big]}{\big[\sigma_H\varepsilon_C+\sigma_L(1-\varepsilon_C)\big]\big[\sigma_H(1-\varepsilon_C)q'_{AH}+\sigma_L\varepsilon_C q'_{AL}\big]}.
\end{aligned}$$

The former is always strictly positive. On the other hand, the sign of latter is not be determined. It is strictly negative if and only if $\varepsilon_C^2 q'_{AL} - (1-\varepsilon_C)^2 q'_{AH} < 0$. Hence, we need to consider the condition for no-deviation for every case. If $\varepsilon_C^2 q'_{AL} - (1-\varepsilon_C)^2 q'_{AH} \geqslant 0$, $b_A(0,0) - b_C(0,1,H) \geqslant 0$. So in this case, v values both beliefs with weight h. Thus the condition is

$$\begin{aligned}
&\Pr[x=1 \mid a=0, s_C=1, q_{CC}=q_H] + \beta h\big(b_A(0,1) - b_C(0,1,H)\big) \\
&\quad\geqslant \Pr[x=0 \mid a=0, s_C=1, q_{CC}=q_H] + \beta h\big(b_A(0,0) - b_C(0,1,H)\big) \\
\iff &\frac{\sigma_H(1-\varepsilon_C)(1-q_A)q_H + \sigma_L\varepsilon_C(1-q_A)q_L}{\sigma_H(1-\varepsilon_C)q'_{AH}+\sigma_L\varepsilon_C q'_{AL}} \\
&\qquad + \beta h \frac{\sigma_H(1-\varepsilon_C)\sigma_L\varepsilon_C(q'_{AL}-q'_{AH})}{\big[\sigma_H(1-\varepsilon_C)+\sigma_L\varepsilon_C\big]\big[\sigma_H(1-\varepsilon_C)q'_{AH}+\sigma_L\varepsilon_C q'_{AL}\big]} \\
&\quad\geqslant \frac{\sigma_H(1-\varepsilon_C)q_A(1-q_H) + \sigma_L\varepsilon_C q_A(1-q_L)}{\sigma_H(1-\varepsilon_C)q'_{AH}+\sigma_L\varepsilon_C q'_{AL}} \\
&\qquad + \beta h \frac{\sigma_H\sigma_L\big[\varepsilon_C^2 q'_{AL}-(1-\varepsilon_C)^2 q'_{AH}\big]}{\big[\sigma_H\varepsilon_C+\sigma_L(1-\varepsilon_C)\big]\big[\sigma_H(1-\varepsilon_C)q'_{AH}+\sigma_L\varepsilon_C q'_{AL}\big]} \\
\iff \beta \geqslant &\frac{\sigma_H(1-\varepsilon_C)+\sigma_L\varepsilon_C}{\sigma_H\sigma_L} \times \frac{\sigma_H(1-\varepsilon_C)\bar{q}'_{AH}+\sigma_L\varepsilon_C\bar{q}'_{AL}}{\sigma_H(1-\varepsilon_C)q'_{AH}+\sigma_L\varepsilon_C q'_{AL}} \\
&\qquad \times \frac{1}{\frac{1-2\varepsilon_C}{\sigma_H\varepsilon_C+\sigma_L(1-\varepsilon_C)} \times h}.
\end{aligned} \tag{4.17}$$

In the opposite case in which $\varepsilon_C^2 q'_{AL} - (1-\varepsilon_C)^2 q'_{AH} < 0$, the weight of $b_A(0,0) - b_C(0,1,H)$

is 1. Then the condition is

$$
\begin{aligned}
&\Pr[x=1 \mid a=0, s_C=1, q_{CC}=q_H] + \beta h\big(b_A(0,1) - b_C(0,1,H)\big) \\
&\geqslant \Pr[x=0 \mid a=0, s_C=1, q_{CC}=q_H] + \beta\big(b_A(0,0) - b_C(0,1,H)\big) \\
\iff &\frac{\sigma_H(1-\varepsilon_C)(1-q_A)q_H + \sigma_L\varepsilon_C(1-q_A)q_L}{\sigma_H(1-\varepsilon_C)q'_{AH} + \sigma_L\varepsilon_C q'_{AL}} \\
&\quad + \beta h \frac{\sigma_H(1-\varepsilon_C)\sigma_L\varepsilon_C(q'_{AL}-q'_{AH})}{\big[\sigma_H(1-\varepsilon_C)+\sigma_L\varepsilon_C\big]\big[\sigma_H(1-\varepsilon_C)q'_{AH}+\sigma_L\varepsilon_C q'_{AL}\big]} \\
&\geqslant \frac{\sigma_H(1-\varepsilon_C)q_A(1-q_H) + \sigma_L\varepsilon_C q_A(1-q_L)}{\sigma_H(1-\varepsilon_C)q'_{AH} + \sigma_L\varepsilon_C q'_{AL}} \\
&\quad + \beta \frac{\sigma_H\sigma_L\big[\varepsilon_C^2 q'_{AL} - (1-\varepsilon_C)^2 q'_{AH}\big]}{\big[\sigma_H\varepsilon_C+\sigma_L(1-\varepsilon_C)\big]\big[\sigma_H(1-\varepsilon_C)q'_{AH}+\sigma_L\varepsilon_C q'_{AL}\big]} \\
\iff \beta \geqslant &\frac{\sigma_H(1-\varepsilon_C)+\sigma_L\varepsilon_C}{\sigma_H\sigma_L} \times \frac{\sigma_H(1-\varepsilon_C)\bar{q}'_{AH} + \sigma_L\varepsilon_C\bar{q}'_{AL}}{\sigma_H(1-\varepsilon_C)q'_{AH} + \sigma_L\varepsilon_C q'_{AL}} \\
&\quad \times \frac{1}{\frac{1-2\varepsilon_C}{\sigma_H\varepsilon_C+\sigma_L(1-\varepsilon_C)} - \frac{(1-h)(1-\varepsilon_C)\varepsilon_C(q'_{AL}-q'_{AH})}{\sigma_H(1-\varepsilon_C)q'_{AH}+\sigma_L\varepsilon_C q'_{AL}}}.
\end{aligned} \tag{4.18}
$$

Actually, $\varepsilon_C^2 q'_{AL} - (1-\varepsilon_C)^2 q'_{AH} < 0$, the condition for $b_A(0,0) - b_C(0,1,H) < 0$, is equivalent to the condition that the third term of (4.18) is smaller than that of (4.17). And you can see that the first and second terms in them are the same. That is, the third term of the condition that β has to satisfy is smaller one of the third terms of (4.17) and (4.18). Therefore, they are summarized as

$$
\begin{aligned}
\beta \geqslant &\frac{\sigma_H(1-\varepsilon_C)+\sigma_L\varepsilon_C}{\sigma_H\sigma_L} \times \frac{\sigma_H(1-\varepsilon_C)\bar{q}'_{AH} + \sigma_L\varepsilon_C\bar{q}'_{AL}}{\sigma_H(1-\varepsilon_C)q'_{AH} + \sigma_L\varepsilon_C q'_{AL}} \\
&\times \min\left\{ \frac{1}{\frac{1-2\varepsilon_C}{\sigma_H\varepsilon_C+\sigma_L(1-\varepsilon_C)} \times h}, \frac{1}{\frac{1-2\varepsilon_C}{\sigma_H\varepsilon_C+\sigma_L(1-\varepsilon_C)} - \frac{(1-h)(1-\varepsilon_C)\varepsilon_C(q'_{AL}-q'_{AH})}{\sigma_H(1-\varepsilon_C)q'_{AH}+\sigma_L\varepsilon_C q'_{AL}}} \right\}.
\end{aligned} \tag{4.19}
$$

At last, we calculate the condition to choose $c=0$ for $a=0$, $s_C=0$ and q_{CC}. In this case, the differences of beliefs are

$$
\begin{aligned}
b_A(0,0) - b_C(0,0,L) &= \frac{\sigma_H\varepsilon_C}{\sigma_H\varepsilon_C+\sigma_L(1-\varepsilon_C)} - \frac{\sigma_H\varepsilon_C q_{AH}}{\sigma_H\varepsilon_C q_{AH}+\sigma_L(1-\varepsilon_C)q_{AL}} \\
&= \frac{\sigma_H\sigma_L(1-\varepsilon_C)\varepsilon_C(q_{AL}-q_{AH})}{\big[\sigma_H\varepsilon_C+\sigma_L(1-\varepsilon_C)\big]\big[\sigma_H\varepsilon_C q_{AH}+\sigma_L(1-\varepsilon_C)q_{AL}\big]} \\
b_A(0,1) - b_C(0,0,L) &= \frac{\sigma_H(1-\varepsilon_C)}{\sigma_H(1-\varepsilon_C)+\sigma_L\varepsilon_C} - \frac{\sigma_H\varepsilon_C q_{AH}}{\sigma_H\varepsilon_C q_{AH}+\sigma_L(1-\varepsilon_C)q_{AL}} \\
&= \frac{\sigma_H\sigma_L\big[(1-\varepsilon_C)^2 q_{AL} - \varepsilon_C^2 q_{AH}\big]}{\big[\sigma_H(1-\varepsilon_C)+\sigma_L\varepsilon_C\big]\big[\sigma_H\varepsilon_C q_{AH}+\sigma_L(1-\varepsilon_C)q_{AL}\big]}.
\end{aligned}
$$

Though the former is always strictly negative, the latter is undetermined. The condition for this to be strictly negative is $(1-\varepsilon_C)^2 q_{AL} < \varepsilon_C^2 q_{AH}$. By similar way in the previous case, we can state that the chooser does not deviate if

$$
\begin{aligned}
\beta \leqslant &\frac{\sigma_H\varepsilon_C+\sigma_L(1-\varepsilon_C)}{\sigma_H\sigma_L} \times \frac{\sigma_H\varepsilon_C\bar{q}_{AH} + \sigma_L(1-\varepsilon_C)\bar{q}_{AL}}{\sigma_H\varepsilon_C q_{AH} + \sigma_L(1-\varepsilon_C)q_{AL}} \\
&\times \max\left\{ \frac{1}{\frac{(1-2\varepsilon_C)h}{\sigma_H(1-\varepsilon_C)+\sigma_L\varepsilon_C} + \frac{(1-h)\varepsilon_C(1-\varepsilon_C)(q_{AH}-q_{AL})}{\sigma_H\varepsilon_C q_{AH}+\sigma_L(1-\varepsilon_C)q_{AL}}}, \frac{1}{\frac{1-2\varepsilon_C}{\sigma_H(1-\varepsilon_C)+\sigma_L\varepsilon_C}} \right\}.
\end{aligned} \tag{4.20}
$$

Therefore, the pair of strategies is an equilibrium if (4.14), (4.15), (4.16), (4.19) and (4.20) are satisfied. You can see that the conditions for the chooser's rebellious action, (4.15) and (4.19), require that β is sufficiently large. This is intuitive since the source of this kind of action is his self-confidence and β represents the weight of it. On the other hand, the conditions for the chooser's amenable action, (4.16) and (4.20), require that β is sufficiently small. In this strategy the chooser chooses amenable action if he received the signal saying his precision is low. If he behaves rebelliously in this situation, the outcome tends to be wrong. When β is sufficiently small, the chooser follows the advice by the adviser, who may have higher precision than him, in order to choose a right alternative rather than arguing that his precision is high.

We can confirm that there are tuples of parameter for which the conditions hold. For example, if $h = 0.7$, $\alpha_L = 0.75$, $\varepsilon_A = 0.01$, $\varepsilon_C = 0.01$, $q_L = 0.51$, $q_H = 0.9$ and $q_A = 0.8$, the first in (4.14) is $0.03 \leqslant 2.44$ and the second is $0.003 \leqslant 0.41$ for the adviser. The chooser's condition is also satisfied if $2.89 \leqslant \beta \leqslant 3.35$. Thus, we have the following proposition.

Proposition 4.2.1. *There are tuples of parameters for which the pair of strategies is an equilibrium.*

Of course, this is not the unique parameters for which the conditions are satisfied. If $\beta \leqslant 3$, for example, conditions hold for approximately 1 percent of parameters.

In fact, $h < 1$ is a necessary condition for the pair of strategies to be an equilibrium.

Proposition 4.2.2. *There is no tuple of parameters for which the pair of strategies being an equilibrium if $h = 1$.*

Proof. Suppose that $h = 1$. Then (4.20) is equivalent to

$$\beta \leqslant \frac{\sigma_H\varepsilon_C + \sigma_L(1-\varepsilon_C)}{\sigma_H\sigma_L} \times \frac{\sigma_H\varepsilon_C\bar{q}_{AH} + \sigma_L(1-\varepsilon_C)\bar{q}_{AL}}{\sigma_H\varepsilon_C q_{AH} + \sigma_L(1-\varepsilon_C)q_{AL}} \times \frac{1}{\frac{1-2\varepsilon_C}{\sigma_H(1-\varepsilon_C)+\sigma_L\varepsilon_C}}.$$

It is necessary for the pair of strategies to be an equilibrium that there exists β that satisfies at least this inequality and (4.15). To this, it must hold that

$$\begin{aligned}
&\frac{\sigma_H(1-\varepsilon_C) + \sigma_L\varepsilon_C}{\sigma_H\sigma_L} \times \frac{\sigma_H(1-\varepsilon_C)\bar{q}_{AH} + \sigma_L\varepsilon_C\bar{q}_{AL}}{\sigma_H(1-\varepsilon_C)q_{AH} + \sigma_L\varepsilon_C q_{AL}} \times \frac{\sigma_H\varepsilon_C + \sigma_L(1-\varepsilon_C)}{1-2\varepsilon_C} \\
&\leqslant \frac{\sigma_H\varepsilon_C + \sigma_L(1-\varepsilon_C)}{\sigma_H\sigma_L} \times \frac{\sigma_H\varepsilon_C\bar{q}_{AH} + \sigma_L(1-\varepsilon_C)\bar{q}_{AL}}{\sigma_H\varepsilon_C q_{AH} + \sigma_L(1-\varepsilon_C)q_{AL}} \times \frac{1}{\frac{1-2\varepsilon_C}{\sigma_H(1-\varepsilon_C)+\sigma_L\varepsilon_C}} \\
&\iff \bar{q}_{AH}q_{AL} \leqslant q_{AH}\bar{q}_{AL}.
\end{aligned}$$

By the definitions of q_{AH}, q_{AL}, $\bar{q}_{AH}$ and $\bar{q}_{AL}$,

$$\begin{aligned}
&\bar{q}_{AH}q_{AL} \leqslant q_{AH}\bar{q}_{AL} \\
\iff &\big[q_Aq_H - (1-q_A)(1-q_H)\big]\big[q_Aq_L + (1-q_A)(1-q_L)\big] \\
&\quad \leqslant \big[q_Aq_H + (1-q_A)(1-q_H)\big]\big[q_Aq_L - (1-q_A)(1-q_L)\big] \\
\iff &q_H(1-q_L) \leqslant q_L(1-q_H) \iff q_H \leqslant q_L.
\end{aligned}$$

This contradicts the assumption that $q_H > q_L$. □

4.2.2 An alternative equilibrium

It is difficult to find all equilibria since there probably be equilibria that players take behavioral strategies that attain positive probability for several actions. In this section, we investigate the following pair of strategies:

Adviser:

$$\begin{cases} a = \phi & \text{if } q_{CA} = q_H \\ a = s_A & \text{if } q_{CA} = q_L \end{cases} \tag{4.21}$$

Chooser:

$$\begin{cases} c = s_C & \text{if } a = \phi \text{ or } q_{CC} = q_H \\ c = a & \text{if } a \neq \phi \text{ and } q_{CC} = q_L. \end{cases} \tag{4.22}$$

The difference between this strategy and that in the previous section is the action of the chooser when $a \neq \phi$ and $q_{CC} = q_H$. Though the apparent actions of these two strategies may be identical, it is different that the chooser's emotions behind. When the chooser follows (4.2), he wants to show that his precision is high rather than to choose right alternative. On the other hand, when the chooser follows (4.22) cannot completely show that his precision is high since the adviser cannot distinguish his precision if $c = a$. Thus the chooser following (4.22) wants to balance choosing right alternative and showing his precision.

The way to show that this pair of strategies is an equilibrium is approximately the same as previous one. Especially, the prediction of right alternative and the self evaluations by the chooser do not change because the strategy of the adviser does not change. The predictions of right choice by the adviser are collected up in the appendix A.4.

The conditions for the adviser, who cares only the material outcome, to follow the strategy (4.21) are summarized as follows:

$$\begin{aligned} &\alpha_L \varepsilon_A (\delta_L - q_L) \leqslant \alpha_H (1 - \varepsilon_A)(\delta_H - (1 - q_H)) \\ &\alpha_H (1 - \varepsilon_A)(\delta_H - q_H) \leqslant \alpha_L \varepsilon_A (\delta_L - (1 - q_L)) \\ &\alpha_L (1 - \varepsilon_A)(\delta_L - (1 - \delta_L)) \geqslant \alpha_H \varepsilon_A (\delta_H - (1 - \delta_H)) \\ &\alpha_L (1 - \varepsilon_A)(\delta_L - q_L) \geqslant \alpha_H \varepsilon_A (\delta_H - (1 - q_H)), \end{aligned}$$

where $\delta_L \equiv \varepsilon_C (1 - q_A) q_L + (1 - \varepsilon_C) q_A$ and $\delta_H \equiv \varepsilon_C (1 - q_A) + (1 - \varepsilon_C) q_A q_H$. Since δ_L and δ_H do not have the property that is stated in lemma 4.2.1, these conditions do not imply each other.

The conditions for the chooser to follow (4.22) are [1]

$$\begin{aligned}
&\Pr[x=0 \mid a=0, s_C=0, q_{CC}=q_H] + \beta v\left(b_A(0,0) - b_C(0,0,H)\right) \\
&\geqslant \Pr[x=1 \mid a=0, s_C=0, q_{CC}=q_H] + \beta v\left(b_A(0,1) - b_C(0,0,H)\right) \\
&\Pr[x=0 \mid a=0, s_C=1, q_{CC}=q_L] + \beta v\left(b_A(0,0) - b_C(0,0,L)\right) \\
&\geqslant \Pr[x=1 \mid a=0, s_C=1, q_{CC}=q_L] + \beta v\left(b_A(0,1) - b_C(0,0,L)\right) \\
&\Pr[x=1 \mid a=0, s_C=1, q_{CC}=q_H] + \beta v\left(b_A(0,1) - b_C(0,1,H)\right) \\
&\geqslant \Pr[x=0 \mid a=0, s_C=1, q_{CC}=q_H] + \beta v\left(b_A(0,0) - b_C(0,1,H)\right) \\
&\Pr[x=0 \mid a=0, s_C=0, q_{CC}=q_L] + \beta v\left(b_A(0,0) - b_C(0,0,L)\right) \\
&\geqslant \Pr[x=1 \mid a=0, s_C=0, q_{CC}=q_L] + \beta v\left(b_A(0,1) - b_C(0,0,L)\right).
\end{aligned}$$

There are some tuples of parameters for which all conditions for the adviser and the chooser are satisfied, for example, $\beta = 3$, $h = 0.5$, $a_L = 0.5$, $e_A = 0.2$, $e_C = 0.2$, $q_L = 0.51$, $q_A = 0.75$, $q_H = 0.85$. Approximately 2.6 percent of parameters can satisfy the conditions when $\beta \leqslant 3$. Therefore, we have the following proposition.

Proposition 4.2.3. *There are situations in which the pair of strategies* (4.21) *and* (4.22) *is an equilibrium.*

It is worth to see that this pair of strategies can be an equilibrium even if $h = 1$. For example, when $\beta = 1.5$, $h = 1$, $a_L = 0.5$, $e_A = 0.2$, $e_C = 0.2$, $q_L = 0.51$, $q_A = 0.75$, $q_H = 0.85$, conditions are satisfied. Thus $h < 1$ is not necessary condition for the strategy being an equilibrium, which differs from the pair of strategy (4.1) and (4.2).

4.3 Discussion

4.3.1 Independence from Self-Evaluation

We have assumed that the chooser's valuation for reputation depends on his self-evaluation. Concretely, we have specialized the valuation function as

$$v(b_A, b_C) = \begin{cases} h(b_A - b_C) & \text{if } b_A \geqslant b_C \\ b_A - b_C & \text{otherwise} \end{cases},$$

where $h \in (0,1)$. You may think that this specialization is too atypical. Other simple and typical way to measure the value of reputation is to let v be a increasing function for b_A, which does not depend on chooser's self-evaluation. In fact, however, the pair of strategies (4.1) and (4.2) is not an equilibrium for this preference.

Proposition 4.3.1. *For any increasing function $v\colon [0,1] \to \mathbb{R}$, which does not depend on b_C, the pair of following strategy is not an equilibrium.*

Adviser:

$$\begin{cases} a = \phi & \text{if } q_{CA} = q_H \\ a = s_A & \text{if } q_{CA} = q_L \end{cases}$$

[1] The chooser's prediction of his reputation by the adviser after choosing a alternative is derived as

$$\begin{aligned}
b_A(0,0) = b_A(1,1) &= \frac{\sigma_H[(1 \quad \varepsilon_C)q_{AH} + \varepsilon_C]}{\sigma_H[(1 \quad \varepsilon_C)q_{AH} + \varepsilon_C] + \sigma_L[\varepsilon_C q_{AL} + (1 \quad \varepsilon_C)]} \\
b_A(0,1) = b_A(1,0) &= \frac{\sigma_H(1 \quad \varepsilon_C)q'_{AH}}{\sigma_H(1 \quad \varepsilon_C)q'_{AH} + \sigma_L \varepsilon_C q'_{AL}}.
\end{aligned}$$

Chooser:

$$\begin{cases} c = s_C & \text{if } a = \phi \\ c = a & \text{if } a \neq \phi \text{ and } q_{CC} = q_L \\ c \neq a & \text{if } a \neq \phi \text{ and } q_{CC} = q_H \end{cases}$$

Proof. At first, we consider the case in which v is strictly increasing. The most part of this model are identical to that in the previous section except that conditions for the chooser not to deviate. Thus we show that he deviates for this valuation function. When $a = 0$, $s_C = 0$ and $q_{CC} = q_H$, choosing $c = 1$ is better for the chooser than $c = 0$. For this, the following must hold

$$\begin{aligned} &\Pr[x = 1 \mid a = 0, s_C = 0, q_{CC} = q_H] + \beta v(b_A(0,1)) \\ &\geqslant \Pr[x = 0 \mid a = 0, s_C = 0, q_{CC} = q_H] + \beta v(b_A(0,0)) \qquad (4.23) \\ \iff &\beta \geqslant \frac{2\Pr[x = 0 \mid a = 0, s_C = 0, q_{CC} = q_H] - 1}{v(b_A(0,1)) - v(b_A(0,0))}. \end{aligned}$$

When $a = 0$, $s_C = 0$ and $q_{CC} = q_L$, the condition for the chooser to chooses $c = 0$ is

$$\begin{aligned} &\Pr[x = 0 \mid a = 0, s_C = 0, q_{CC} = q_L] + \beta v(b_A(0,0)) \\ &\geqslant \Pr[x = 1 \mid a = 0, s_C = 0, q_{CC} = q_L] + \beta v(b_A(0,1)) \\ \iff &\beta \leqslant \frac{2\Pr[x = 0 \mid a = 0, s_C = 0, q_{CC} = q_L] - 1}{v(b_A(0,1)) - v(b_A(0,0))} \end{aligned}$$

Note that $b_A(0,1) > b_A(0,0)$ holds by lemma 4.2.4, so $v(b_A(0,1)) > v(b_A(0,0))$ since v is strictly increasing. For the existence of β that satisfies these conditions, it must hold that

$$\begin{aligned} &\frac{2\Pr[x = 0 \mid a = 0, s_C = 0, q_{CC} = q_H] - 1}{v(b_A(0,1)) - v(b_A(0,0))} \leqslant \frac{2\Pr[x = 0 \mid a = 0, s_C = 0, q_{CC} = q_L] - 1}{v(b_A(0,1)) - v(b_A(0,0))} \\ \iff &\Pr[x = 0 \mid a = 0, s_C = 0, q_{CC} = q_H] \leqslant \Pr[x = 0 \mid a = 0, s_C = 0, q_{CC} = q_L]. \end{aligned}$$

Substituting each probability, we have

$$\begin{aligned} &\frac{\sigma_H(1-\varepsilon_C)q_A q_H + \sigma_L \varepsilon_C q_A q_L}{\sigma_H(1-\varepsilon_C)q_{AH} + \sigma_L \varepsilon_C q_{AL}} \leqslant \frac{\sigma_H \varepsilon_C q_A q_H + \sigma_L(1-\varepsilon_C)q_A q_L}{\sigma_H \varepsilon_C q_{AH} + \sigma_L(1-\varepsilon_C)q_{AL}} \\ \iff &\sigma_H^2 \varepsilon_C(1-\varepsilon_C)q_{AH}q_A q_H + \sigma_L \sigma_H \varepsilon_C^2 q_{AH} q_A q_L \\ &\quad + \sigma_H \sigma_L (1-\varepsilon_C)^2 q_{AL} q_A q_H + \sigma_L^2 \varepsilon_C (1-\varepsilon_C) q_{AL} q_A q_L \\ &\leqslant \sigma_H^2 \varepsilon_C(1-\varepsilon_C)q_{AH}q_A q_H + \sigma_L \sigma_H (1-\varepsilon_C)^2 q_{AH} q_A q_L \\ &\quad + \sigma_H \sigma_L \varepsilon_C^2 q_{AL} q_A q_H + \sigma_L^2 \varepsilon_C (1-\varepsilon_C) q_{AL} q_A q_L \\ \iff &\{1 - 2\varepsilon_C\} q_{AL} q_H \leqslant \{1 - 2\varepsilon_C\} q_{AH} q_L \\ \iff &\{q_A q_L + (1-q_A)(1-q_L)\} q_H \leqslant \{q_A q_H + (1-q_A)(1-q_H)\} q_L \\ \iff &q_H \leqslant q_L. \end{aligned}$$

However, this contradicts $q_H > q_L$.

Next we consider the case in which v is weakly increasing. What is important is v such that $v(b_A(0,1)) = v(b_A(0,0))$. In this case, by (4.23),

$$\begin{aligned} &\Pr[x = 1 \mid a = 0, s_C = 0, q_{CC} = q_H] \geqslant \Pr[x = 0 \mid a = 0, s_C = 0, q_{CC} = q_H] \\ \iff &\Pr[x = 0 \mid a = 0, s_C = 0, q_{CC} = q_H] \leqslant \frac{1}{2} \\ \iff &\frac{\sigma_H(1-\varepsilon_C)q_A q_H + \sigma_L \varepsilon_C q_A q_L}{\sigma_H(1-\varepsilon_C)q_{AH} + \sigma_L \varepsilon_C q_{AL}} \leqslant \frac{1}{2} \\ \iff &\sigma_H(1-\varepsilon_C)(q_A + q_H - 1) + \sigma_L \varepsilon_C (q_A + q_L - 1) \leqslant 0. \end{aligned}$$

This is a contradiction since $q_H > q_A > q_L \geqslant 1/2$. □

4.4 Conclusion

We have investigated psychological reactance from a view point of economics. Though the reason of the phenomenon has been regarded as a restoration of threatened freedom for a long time, we proposed other explanation for it. Our key assumption is that the chooser cares his reputation by the adviser. Especially, what is important is the value of reputation depends on his self-evaluation. Our contribution is the presentation of a formal model that describes the psychological reactance as the equilibrium of a game. Though there may be other equilibria and preferences that realize a reactive behavior, we think that this finding is a big step.

There are still many problems remaining. The task that we want to deal with is the extension of the precision space of the chooser, which have been assumed to be discrete. If the precision is extended to continuous one as Levy (2004) does, we will be able to get a better sense of relation between psychological reactance and the effect of reputation concern.

Appendices

Appendix A

Appendices for Chapter 3

A.1 The probability of right choice for strategy (4.2)

In this appendix, we describe the adviser's subjective probability with which the chooser chooses a right alternative, which is discussed in section 4.2.1. These probabilities have some kind of symmetric property. For example, the probability under $(q_{CA}, s_A, a) = (q_H, 0, 1)$ is the same as that under $(q_{CA}, s_A, a) = (q_H, 1, 0)$.

$$
\begin{aligned}
&\Pr[c = x \mid q_{CA} = q_H, s_A = 0, a = 0] = \Pr[c = x \mid q_{CA} = q_H, s_A = 1, a = 1] \\
&= \frac{\alpha_L \varepsilon_A \delta + \alpha_H (1 - \varepsilon_A)(1 - \delta)}{\alpha_L \varepsilon_A + \alpha_H (1 - \varepsilon_A)} \\
&\Pr[c = x \mid q_{CA} = q_H, s_A = 0, a = 1] = \Pr[c = x \mid q_{CA} = q_H, s_A = 1, a = 0] \\
&= \frac{\alpha_L \varepsilon_A (1 - \delta) + \alpha_H (1 - \varepsilon_A)\delta}{\alpha_L \varepsilon_A + \alpha_H (1 - \varepsilon_A)} \\
&\Pr[c = x \mid q_{CA} = q_H, s_A = 0, a = \phi] = \Pr[c = x \mid q_{CA} = q_H, s_A = 1, a = \phi] \\
&= \frac{\alpha_L \varepsilon_A q_L + \alpha_H (1 - \varepsilon_A) q_H}{\alpha_L \varepsilon_A + \alpha_H (1 - \varepsilon_A)} \\
&\Pr[c = x \mid q_{CA} = q_L, s_A = 0, a = 0] = \Pr[c = x \mid q_{CA} = q_L, s_A = 1, a = 1] \\
&= \frac{\alpha_L (1 - \varepsilon_A)\delta + \alpha_H \varepsilon_A (1 - \delta)}{\alpha_L (1 - \varepsilon_A) + \alpha_H \varepsilon_A} \\
&\Pr[c = x \mid q_{CA} = q_L, s_A = 0, a = 1] = \Pr[c = x \mid q_{CA} = q_L, s_A = 1, a = 0] \\
&= \frac{\alpha_L (1 - \varepsilon_A)(1 - \delta) + \alpha_H \varepsilon_A \delta}{\alpha_L (1 - \varepsilon_A) + \alpha_H \varepsilon_A} \\
&\Pr[c = x \mid q_{CA} = q_L, s_A = 0, a = \phi] = \Pr[c = x \mid q_{CA} = q_L, s_A = 1, a = \phi] \\
&= \frac{\alpha_L (1 - \varepsilon_A) q_L + \alpha_H \varepsilon_A q_H}{\alpha_L (1 - \varepsilon_A) + \alpha_H \varepsilon_A}
\end{aligned}
$$

A.2 The chooser's predictions of right alternative for strategy (4.1)

In this appendix, we present the chooser's predictions of right alternative after observing a, s_C and q_{CC}. Again, the symmetric property in the previous appendix holds. For example, if $q_{CC} = q_H$, the probability of $x = 0$ when $(a, s_C) = (1, 0)$ is the same as that of $x = 1$ when $(a, s_C) = (0, 1)$.

$$\Pr[x=0 \mid a=0, s_C=0, q_{CC}=q_H] = \Pr[x=1 \mid a=1, s_C=1, q_{CC}=q_H]$$
$$= \frac{\sigma_H(1-\varepsilon_C)q_A q_H + \sigma_L \varepsilon_C q_A q_L}{\sigma_H(1-\varepsilon_C)q_{AH} + \sigma_L \varepsilon_C q_{AL}}$$
$$\Pr[x=1 \mid a=0, s_C=0, q_{CC}=q_H] = \Pr[x=0 \mid a=1, s_C=1, q_{CC}=q_H]$$
$$= \frac{\sigma_H(1-\varepsilon_C)(1-q_A)(1-q_H) + \sigma_L \varepsilon_C (1-q_A)(1-q_L)}{\sigma_H(1-\varepsilon_C)q_{AH} + \sigma_L \varepsilon_C q_{AL}}$$
$$\Pr[x=0 \mid a=0, s_C=0, q_{CC}=q_L] = \Pr[x=1 \mid a=1, s_C=1, q_{CC}=q_L]$$
$$= \frac{\sigma_H \varepsilon_C q_A q_H + \sigma_L(1-\varepsilon_C)q_A q_L}{\sigma_H \varepsilon_C q_{AH} + \sigma_L(1-\varepsilon_C)q_{AL}}$$
$$\Pr[x=1 \mid a=0, s_C=0, q_{CC}=q_L] = \Pr[x=0 \mid a=1, s_C=1, q_{CC}=q_L]$$
$$= \frac{\sigma_H \varepsilon_C (1-q_A)(1-q_H) + \sigma_L(1-\varepsilon_C)(1-q_A)(1-q_L)}{\sigma_H \varepsilon_C q_{AH} + \sigma_L(1-\varepsilon_C)q_{AL}}$$

$$\Pr[x=0 \mid a=0, s_C=1, q_{CC}=q_H] = \Pr[x=1 \mid a=1, s_C=0, q_{CC}=q_H]$$
$$= \frac{\sigma_H(1-\varepsilon_C)q_A(1-q_H) + \sigma_L \varepsilon_C q_A(1-q_L)}{\sigma_H(1-\varepsilon_C)q'_{AH} + \sigma_L \varepsilon_C q'_{AL}}$$
$$\Pr[x=1 \mid a=0, s_C=1, q_{CC}=q_H] = \Pr[x=0 \mid a=1, s_C=0, q_{CC}=q_H]$$
$$= \frac{\sigma_H(1-\varepsilon_C)(1-q_A)q_H + \sigma_L \varepsilon_C (1-q_A)q_L}{\sigma_H(1-\varepsilon_C)q'_{AH} + \sigma_L \varepsilon_C q'_{AL}}$$
$$\Pr[x=0 \mid a=0, s_C=1, q_{CC}=q_L] = \Pr[x=1 \mid a=1, s_C=0, q_{CC}=q_L]$$
$$= \frac{\sigma_H \varepsilon_C q_A(1-q_H) + \sigma_L(1-\varepsilon_C)q_A(1-q_L)}{\sigma_H \varepsilon_C q'_{AH} + \sigma_L(1-\varepsilon_C)q'_{AL}}$$
$$\Pr[x=1 \mid a=0, s_C=1, q_{CC}=q_L] = \Pr[x=0 \mid a=1, s_C=0, q_{CC}=q_L]$$
$$= \frac{\sigma_H \varepsilon_C (1-q_A)q_H + \sigma_L(1-\varepsilon_C)(1-q_A)q_L}{\sigma_H \varepsilon_C q'_{AH} + \sigma_L(1-\varepsilon_C)q'_{AL}}$$

$$\Pr[x=0 \mid a=\phi, s_C=0, q_{CC}=q_H] = \Pr[x=1 \mid a=\phi, s_C=1, q_{CC}=q_H]$$
$$= \frac{\alpha_H(1-\varepsilon_A)(1-\varepsilon_C)q_H + \alpha_L \varepsilon_A \varepsilon_C q_L}{\alpha_H(1-\varepsilon_A)(1-\varepsilon_C) + \alpha_L \varepsilon_A \varepsilon_C}$$
$$\Pr[x=1 \mid a=\phi, s_C=0, q_{CC}=q_H] = \Pr[x=0 \mid a=\phi, s_C=1, q_{CC}=q_H]$$
$$= \frac{\alpha_H(1-\varepsilon_A)(1-\varepsilon_C)(1-q_H) + \alpha_L \varepsilon_A \varepsilon_C (1-q_L)}{\alpha_H(1-\varepsilon_A)(1-\varepsilon_C) + \alpha_L \varepsilon_A \varepsilon_C}$$
$$\Pr[x=0 \mid a=\phi, s_C=0, q_{CC}=q_L] = \Pr[x=1 \mid a=\phi, s_C=1, q_{CC}=q_L]$$
$$= \frac{\alpha_H(1-\varepsilon_A)\varepsilon_C q_H + \alpha_L \varepsilon_A (1-\varepsilon_C)q_L}{\alpha_H(1-\varepsilon_A)\varepsilon_C + \alpha_L \varepsilon_A (1-\varepsilon_C)}$$
$$\Pr[x=1 \mid a=\phi, s_C=0, q_{CC}=q_L] = \Pr[x=0 \mid a=\phi, s_C=1, q_{CC}=q_L]$$
$$= \frac{\alpha_H(1-\varepsilon_A)\varepsilon_C (1-q_H) + \alpha_L \varepsilon_A (1-\varepsilon_C)(1-q_L)}{\alpha_H(1-\varepsilon_A)\varepsilon_C + \alpha_L \varepsilon_A (1-\varepsilon_C)}$$

A.3 The posterior of q_C by the chooser for strategy (4.1)

We collect up the posterior of q_C by the chooser. These are the probabilities of $q_C = q_H$ with observation (a, s_C, q_{CC}).

$$
\begin{aligned}
b_C(0,0,H) = b_C(1,1,H) &= \frac{\sigma_H(1-\varepsilon_C)q_{AH}}{\sigma_H(1-\varepsilon_C)q_{AH} + \sigma_L\varepsilon_C q_{AL}} \\
b_C(0,0,L) = b_C(1,1,L) &= \frac{\sigma_H\varepsilon_C q_{AH}}{\sigma_H\varepsilon_C q_{AH} + \sigma_L(1-\varepsilon_C)q_{AL}} \\
b_C(0,1,H) = b_C(1,0,H) &= \frac{\sigma_H(1-\varepsilon_C)q'_{AH}}{\sigma_H(1-\varepsilon_C)q'_{AH} + \sigma_L\varepsilon_C q'_{AL}} \\
b_C(0,1,L) = b_C(1,0,L) &= \frac{\sigma_H\varepsilon_C q'_{AH}}{\sigma_H\varepsilon_C q'_{AH} + \sigma_L(1-\varepsilon_C)q'_{AL}} \\
b_C(\phi,0,H) = b_C(\phi,1,H) &= \frac{\alpha_H(1-\varepsilon_A)(1-\varepsilon_C)}{\alpha_H(1-\varepsilon_A)(1-\varepsilon_C) + \alpha_L\varepsilon_A\varepsilon_C} \\
b_C(\phi,0,L) = b_C(\phi,1,L) &= \frac{\alpha_H(1-\varepsilon_A)\varepsilon_C}{\alpha_H(1-\varepsilon_A)\varepsilon_C + \alpha_L\varepsilon_A(1-\varepsilon_C)}
\end{aligned}
$$

A.4 The probability of right choice for strategy (4.22)

In this appendix, we describe the adviser's subjective probability with which the chooser chooses a right alternative when he follows 4.22, which is discussed in section 4.2.2, where $\delta_L \equiv \varepsilon_C(1-q_A)q_L + (1-\varepsilon_C)q_A$ and $\delta_H \equiv \varepsilon_C(1-q_A) + (1-\varepsilon_C)q_A q_H$.

$$
\begin{aligned}
\Pr[c = x \mid q_{CA} = q_H, s_A = 0, a = 0] &= \frac{\alpha_L\varepsilon_A\delta_L + \alpha_H(1-\varepsilon_A)(1-\delta_H)}{\alpha_L\varepsilon_A + \alpha_H(1-\varepsilon_A)} \\
\Pr[c = x \mid q_{CA} = q_H, s_A = 0, a = 1] &= \frac{\alpha_L\varepsilon_A(1-\delta_L) + \alpha_H(1-\varepsilon_A)\delta_H}{\alpha_L\varepsilon_A + \alpha_H(1-\varepsilon_A)} \\
\Pr[c = x \mid q_{CA} = q_H, s_A = 0, a = \phi] &= \frac{\alpha_L\varepsilon_A q_L + \alpha_H(1-\varepsilon_A)q_H}{\alpha_L\varepsilon_A + \alpha_H(1-\varepsilon_A)} \\
\Pr[c = x \mid q_{CA} = q_H, s_A = 1, a = 0] &= \frac{\alpha_L\varepsilon_A(1-\delta_L) + \alpha_H(1-\varepsilon_A)\delta_H}{\alpha_L\varepsilon_A + \alpha_H(1-\varepsilon_A)} \\
\Pr[c = x \mid q_{CA} = q_H, s_A = 1, a = 1] &= \frac{\alpha_L\varepsilon_A\delta_L + \alpha_H(1-\varepsilon_A)(1-\delta_H)}{\alpha_L\varepsilon_A + \alpha_H(1-\varepsilon_A)} \\
\Pr[c = x \mid q_{CA} = q_H, s_A = 1, a = \phi] &= \frac{\alpha_L\varepsilon_A q_L + \alpha_H(1-\varepsilon_A)q_H}{\alpha_L\varepsilon_A + \alpha_H(1-\varepsilon_A)}
\end{aligned}
$$

$$\Pr[c = x \mid q_{CA} = q_L, s_A = 0, a = 0] = \frac{\alpha_L(1-\varepsilon_A)\delta_L + \alpha_H\varepsilon_A(1-\delta_H)}{\alpha_L(1-\varepsilon_A) + \alpha_H\varepsilon_A}$$

$$\Pr[c = x \mid q_{CA} = q_L, s_A = 0, a = 1] = \frac{\alpha_L(1-\varepsilon_A)(1-\delta_L) + \alpha_H\varepsilon_A\delta_H}{\alpha_L(1-\varepsilon_A) + \alpha_H\varepsilon_A}$$

$$\Pr[c = x \mid q_{CA} = q_L, s_A = 0, a = \phi] = \frac{\alpha_L(1-\varepsilon_A)q_L + \alpha_H\varepsilon_A q_H}{\alpha_L(1-\varepsilon_A) + \alpha_H\varepsilon_A}$$

$$\Pr[c = x \mid q_{CA} = q_L, s_A = 1, a = 0] = \frac{\alpha_L(1-\varepsilon_A)(1-\delta_L) + \alpha_H\varepsilon_A\delta_H}{\alpha_L(1-\varepsilon_A) + \alpha_H\varepsilon_A}$$

$$\Pr[c = x \mid q_{CA} = q_L, s_A = 1, a = 1] = \frac{\alpha_L(1-\varepsilon_A)\delta_L + \alpha_H\varepsilon_A(1-\delta_H)}{\alpha_L(1-\varepsilon_A) + \alpha_H\varepsilon_A}$$

$$\Pr[c = x \mid q_{CA} = q_L, s_A = 1, a = \phi] = \frac{\alpha_L(1-\varepsilon_A)q_L + \alpha_H\varepsilon_A q_H}{\alpha_L(1-\varepsilon_A) + \alpha_H\varepsilon_A}.$$

References

M Armstrong, J Vickers, and J Zhou. Consumer protection and the incentive to become informed. *Journal of the European Economic Association*, 7(2-3):399–410, 2009.

J W Brehm. *A theory of psychological reactance*. New York: Academic Press, 1966.

Peter A Diamond. A framework for social security analysis. *Journal of Public Economics*, 8(3):275–298, 1977.

Simone Galperti. Commitment, Flexibility, and Optimal Screening of Time Inconsistency. *Econometrica*, 83(4):1425–1465, 2015.

Faruk Gul and Wolfgang Pesendorfer. Temptation and Self-Control. *Econometrica*, 69(6): 1403–1435, 2001.

Faruk Gul and Wolfgang Pesendorfer. Self-control, revealed preference and consumption choice. *Review of Economic Dynamics*, 7(2):243–264, 2004.

Daniel Kahneman and Amos Tversky. Choices , Values , and Frames. *American Psycologist*, 39(4):341–350, 1984.

Per Krusell, Burhanettin Kuruscu, and Jr. Smith, Anthony A. Temptation and Taxation. *Econometrica*, 78(6):2063–2084, 2010.

Çağri S Kumru and Athanasios C Thanopoulos. Social security and self control preferences. *Journal of Economic Dynamics and Control*, 32(3):757–778, 2008.

David Laibson. Golden Eggs and Hyperbolic Discounting. *Quarterly Journal of Economics*, 112(2):443–477, 1997.

Gilat Levy. Anti-herding and strategic consultation. *European Economic Review*, 48(3): 503–525, 2004.

Yusufcan Masatlioglu and Efe A Ok. Rational choice with status quo bias. *Journal of Economic Theory*, 121(1):1–29, 2005.

Kerstin Roeder. Optimal taxes and pensions with myopic agents. *Social Choice and Welfare*, 42(3):597–618, 2014.

Paul. A Samuelson. Optimum social security in a life-cycle growth model. *International Economic Review*, 16(3):539–544, 1975.

William Samuelson and Richard Zeckhauser. Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1):7–59, 1988.

Amos Tversky and Daniel Kahneman. Loss Aversion in Riskless Choice : A Reference-Dependent Model. *Quarterly Journal of Economics*, 106(4):1039–1061, 1991.