Contracting with Heterogenous Quasi-Hyperbolic Agents: Investment Good Pricing under Asymmetric Information *

Jianye YAN†

*This is the first chapter of the thesis. I am deeply indebted to my advisor, Professor Thomas Mariotti, for his indispensable instructions on my research and long-run substantial encouragement to me especially during both physically and psychologically recovering procedure after I suffered from a car accident. I also thank Professor Jean Tirole for his advices and discussions with me on the research topic of quasi-hyperbolic discounting. And I am grateful to the participants in the 6th PhD Presentation Meeting of the Royal Economic Society at City University London and 2010 Annual Meeting of the Association of Southern European Economic Theorists at University of Alicante, too. The conversations with Professor François Salanié, Professor Claude Crampes, and Sanxi Li also benefit me. Of course all the possible faults and defects inside the paper are my own.

†Toulouse School of Economics. Address: MH 208, TSE, Manufacture de Tabacs, 21 allées de Brienne, 31000, Toulouse, France. Phone: 0033-6-62608675. Email: jianye.yan@tse-fr.eu.

Abstract

Growing evidence from experiments and field studies supporting the existence of time-inconsistent discounting behavior in decision making encourages economists to investigate its impact on standard economic theories. We extend DellaVigna and Malmendier (2004)'s monopolistic pricing under quasi-hyperbolic discounting to the asymmetric information case where the consumer may hold private information on different dimensions: heterogeneous consumption benefit, heterogeneous short-run impatience, and heterogeneous naivety. Some important properties established in its complete information pricing scenario do not hold any more. Interestingly, the degree of short-run impatience exhibits a non-monotonic impact on the seller's profit when the consumption benefit is the consumer's private information. As long as the principal is imprecise about the time preference of her agents, an over-consumption of the investment good for the sophisticated consumer will be involved. At last, in the case of diverse naivety, when the principal can screen her agents, more naive agent gets more rent in the sense of fictitious utility, but in fact the consumer's real surplus decreases in the degree of naivety.

Key words: *Time-inconsistency*; *Quasi-hyperbolic discounting*; *Investment Good*; *Screening Menu*

JEL classification: D03; D82; D91

1 Introduction

Growing evidences from experiments and field studies (for example: experimental evidences from Larwood and Whittaker (1977), Kirby and Herrnstein (1995); field studies by Madrian and Shea (2001), Ariely and Wertenbroch (2002), DellaVigna and Malmendier (2006), etc.) support the existence of a lack of self-control on the part of individual decision makers, which may lead to time-inconsistent behavior. This leads to the following question: how a rational profit-maximizing firm can take advantage from this irrational aspect of consumers' preference and tailor its contracts or pricing schemes in response to it? This is particularly relevant in *investment good* industries. When the benefit (utility) from consuming comes later than the purchasing (consuming) date: a typical example emphasized by DellaVigna and Malmendier (2004) (DM, henceforth), is the attendance to a health club which involves current exercising effort cost and only delivers future health benefits.

DM study the optimal pricing scheme by a time-consistent firm facing a time-inconsistent consumer, with a quasi-hyperbolic discount function. Their conclusions are strikingly different from the standard finding in industrial organization, according to which an optimal two-part tariff should feature marginal cost pricing plus a fixed fee which extracts the entire consumer surplus. Instead, they show that a monopolist in the investment good industry facing a time-inconsistent consumer will charge a price below marginal cost, along with a higher flat fee, which is consistent with the empirical evidence about health club pricing: a small per entrance fee coupled with a large registration fee. Thanks to this non-standard skewed pricing, the firm's profit is not hurt by the consumer's short-run impatience as long as the latter is fully aware of his time-inconsistency problem, i.e., he is sophisticated.

To derive these results, DM assume that there is no asymmetric information between the firm and the consumer; nevertheless, in the reality, we also sometimes encounter the private information on the consumer side appearing in the market. In this paper, we suppose instead that the consumer may hold private information on different dimensions. Three kinds of asymmetric information are considered: (1) first, the intrinsic benefit from consuming the investment good may be the consumer's private information; (2) second, the consumer may hold private information about his degree of short-run impatience; (3) third, different consumers may differ in their self-awareness about their future short-run impatience, i.e., his degree of naivety may be the consumer's private information, where we adopt the definition of *naivety* as in O'Donoghue and Rabin (2001).

The main results of our paper are as follows. When the consumer holds private information about his intrinsic benefit of consuming the investment good, we find that under the second-best pricing, the first-best below-marginal-cost per-usage pricing property established in DM does not hold for the whole population; specifically, the per-usage price for some relatively inefficient types (low benefit consumer) may be higher than the marginal cost. More importantly, unlike in the first-best, the existence of time-inconsistency of the consumer hurts the firm's profit even if the consumer is sophisticated. When time-inconsistency is combined with asymmetric information, the firm cannot get a full reimbursement of its per-usage loss by charging an entrance fee because of the existence of the informational rent. And interestingly, this rent and the firm's expected profit is non-monotonic in the degree of short-run impatience: when the consumer is close to a fully patient one, more short-run impatience hurts the firm; while when the consumer's short-run patience level is extremely low, more short-run impatience improves the firm's profit, where the intuition is that the firm can take advantage of the significant conflict between the consumer's long-run and short-run self in the sense of charging an extremely huge entrance fee, since the relatively patient and sophisticated long-run self is urgent to pay and participate the club in order to commit his future impulsive self to consume under a low per-usage price.

When the consumer's short-run impatience becomes his private information, the below-marginal-cost per-usage pricing property is still preserved and the shape of the second-best pricing scheme depends on the distribution of the consumer's cost of consuming the investment good. For example, if the consuming cost is uniformly distributed, then the only feasible pricing scheme is a pooling one; otherwise, separating pricing schemes may exist. No matter which kind of second-best pricing scheme that emerges at the optimum, when the consumer has private information about his short-run time preference parameter, the first-best property that the firm's profit is invariant as long as the consumer is sophisticated does not hold.

At last, we model the private information on the consumer's naivety. In the quasi-hyperbolic (β, δ)-discounting framework, following O'Donoghue and Rabin (2001), we parameterize the consumer's naivety by his subjective prediction of the future short-run impatience, which is denoted by $\widehat{\beta} \in [\beta, 1]$. The firm knows the short-run impatience parameter of the consumer, but does not know to which extent the consumer himself knows it: the firm cannot observe the cognitive belief of the consumer, $\widehat{\beta}$, while the consumer sticks to his belief $\widehat{\beta}$, and doesn't know the true β or the fact that the firm knows β when contracting. In one word, the two contracting parties hold different beliefs about the consumer's future short-run impatience, and since the consumer is naive, he does not update his belief after observing the pricing scheme offered by the firm. When the firm can only offer a pooling price in the second-best, the firm's first-best benefit from the consumer's naivety disappears, i.e., the firm can only obtain a sophisticated consumer's profit when trading with a naive consumer. When the firm can screen the degree of naivety, we find that the sophisticated type over-consumes the investment good, while the consumption allocation for a fully naive consumer coincides with its complete information level. Though the more naive consumer gets more fictitious rent when contracting, in fact, the firm extracts more from him in the sense of real consumption surplus.

In the literature, traditionally there are mainly two approaches to model individual's self control/present bias problem (see Amador, Werning and Angeletos, 2006). On top of quasi-hyperbolic discounting, another one is Gul-Pesendorfer temptation preference formulation (Gul and Pesendorfer, 2001). Esteban and Miyagawa (2006), and Esteban, Miyagawa and Shum (2007) are a set of papers on firm(s)' pricing strategy against a population of diversified consumers who hold Gul-Pesendorfer temptation preferences. Thus this paper contributes to the literature on the incomplete information contracting with a pool of quasi-hyperbolic agents, as both an extension to the complete information first-best of DM and a parallel comparison with the literature on the second-best pricing within Gul-Pesendorfer framework; and we also look forward to shedding some light on the broader interdiscipline among contract theory, industrial organization and behavioral economics.

Eliaz and Spiegler (2006) is also a paper concerning second-best contracting between a time-consistent principal and a time-inconsistent agent, where the private information is the agent's degree of naivety. They use a general formulation depicting time-inconsistency, which is compatible to both the quasi-hyperbolic approach and Gul-Pesendorfer temptation preference. They do not restrict or parameterize the functional form, while just considered a pair of function (u, v), which represent the preferences of today and tomorrow. The agent naively and privately holds a belief in probability θ that his tomorrow's preference would remain u with probability θ, and that with probability $1 - \theta$, it will become v. The main distinguishing feature with us is: we follow O'Donoghue and Rabin (2001) and DM, measuring the agent's naivety by a parameter $\widehat{\beta} \in [\beta, 1]$, hence the imaginary preference of tomorrow is inside a continuum and not linear in $\widehat{\beta}$, unlike the linear combination between two extremes u and v in Eliaz and Spiegler (2006), where the utility is linear in the naivety type θ. But some findings are similar to us, e.g., their agent's net utility of today decreases with his naivety type after the participation to the contract.

The remainder of the paper is organized as follows. Section 2 briefly reviews the concepts of quasi-hyperbolic discounting and naivety, and the investment good pricing in DM. In Section 3 we introduce in turn the three dimensions of the consumer's private information to see the impact of asymmetric information structure on contracting. Section 4 discusses some other kinds of screening prices based on non-uniform distribution of exercising cost for sophisticated and naive consumer, respectively. Section 5 concludes.

2 Quasi-hyperbolic Discounting, Naivety and Investment Good

In this section we discuss the concepts of quasi-hyperbolic discounting and naivety in the literature; and later Proposition 0 represents the first-best investment good pricing properties worked out by DM.

Investment Good If the benefit and utility of consuming a good comes one period later than the date of consuming, then we call this kind of good an *investment good*. For example, health club participation, continuation school program, and rehabilitation center[1] are some typical investment goods: exercising/studying/curing today brings contemporaneous disutility but results in fitness/knowledge/health tomorrow. The time lag between the consumption and the benefit of the investment good influences the choice of a time-inconsistent consumer who suffers from short-run impatience due to a lack of self-control. DellaVigna and Malmendier (2004, 2006) propose a 3-date framework (refer to Figure 1) to analyze health club pricing. At date 0, a firm wants to attract a consumer to buy an investment good, where the consumer incurs a cost c at date 1 if consuming and will receive a late benefit b at date 2, which can be interpreted as the consumer's intrinsic willingness to pay for the good at this date. The parameter c can be interpreted as the physical and psychological disutility of exercising. The firm proposes a two-part tariff (L, p) at date 0. L represents the entrance fee, while p is the per-usage price for the participants of the health club. At date 0 the consumer has to decide whether or not to accept this tariff scheme, that is, whether or not to join the club. If he does not accept, he obtains a zero reservation utility. If he accepts, he is committed to pay L to the firm at time $t = 1$.[2] Moreover, at

[1] Normally a typical advertisement appearing at the gate of a rehabilitation center is: "Building Brighter Futures For People With Miserable Addictions".

[2] DM assume that L is incurred at date 1 in order to circumvent the short-run impatient participation to the health club due to time-inconsistency. Their main highlight is the decision discrepancy between the consumer's self 0 and self 1 about whether or not to exercise at date 1. So L and p are incurred in the same period and represent the point of view of long-run and short-run self, respectively. This assumption is also critical to obtain the first-best property under complete information that the firm's profit is not affected by the consumer's time-inconsistency as long as he is sophisticated. We show below that this

this date, the consumer who joined the club at date 0 decides whether to consume or not: if he does not consume, he is exempted from paying the price p or incurring the cost c, and he will get zero benefit at date 2; if he consumes, he bears a total cost $c + p$ at date 1 and will obtain a benefit b at date 2. The cost c is unknown to both the firm and the consumer when contracting at date 0, but the distribution of c is common knowledge between the firm and the consumer. Denote by F the corresponding c.d.f., and by f the corresponding p.d.f.

Figure 1 is about here

Time-inconsistent Preference The consumer has a self-control problem that takes the form of short-run impatience. Following DM, we model this self-control problem by using quasi-hyperbolic discounting. In this 3-date framework, the present value of future utilities at date 0 of a quasi-hyperbolic consumer is $u_0 = v_0 + \beta \sum_{t=1}^{2} \delta^t v_t$, where v_t is the instantaneous utility flow. The corresponding discount function is represented by the long dotted line in Figure 2. The parameter $\delta < 1$ is the usual exponential discounting factor; while the parameter $\beta < 1$, which corresponds to the spread between long dotted line and the solid line in Figure 2, measures the short-run impatience, or myopia, of the quasi-hyperbolic consumer. The firm is assumed to be fully patient and time-consistent,[3] with a discount factor δ. The short dotted line in Figure 2 measures the intertemporal evaluation of a quasi-hyperbolic consumer with $\beta < 1$ starting from date 1. The difference between the long and the short dotted lines reflects the consumer's time-inconsistency: the

latter property is violated in the second-best.

[3]Here we regard the firm as an organization whose corporate intertemporal decision making does not suffer from individual self-control problems.

preferences toward the future (date 2) are different from the points of view between today (date 0) and tomorrow (date 1).

Figure 2 is about here

Sophistication vs. Naivety Under (β, δ)-discounting, *naivety*, a notion introduced by O'Donoghue and Rabin (2001), indicates the extent to which a quasi-hyperbolic consumer is aware of his future self-control problem. For example, a *sophisticated* consumer perfectly predicts that he will suffer from a self-control problem tomorrow, and thus accurately predicts his tomorrow's time preference, as represented by the short dotted line in Figure 2. By contrast, a *fully naive* consumer is totally unaware today of his short-run impatience tomorrow, i.e., he believes falsely that he will be time-consistent tomorrow: on Figure 2, a fully naive consumer falsely believes that the discounting from date 1 to date 2 remains along the long dotted line. In an intermediate case, labeled *partial naivety,* the consumer knows that he will lack self-control tomorrow, but does not predict this problem to its full extent, i.e., the perceived discounting from date 1 to date 2 is located between the long dotted line and the short dotted line in Figure 2.

In the 3-date setting, the prediction of the consumer's self 0 about his tomorrow's utility is $\widehat{u}_1 = v_1 + \widehat{\beta}\delta v_2$, where $\widehat{\beta} \in [\beta, 1]$; while the consumer at date 1 will evaluate present and future utilities according to $u_1 = v_1 + \beta\delta v_2$. The cases, $\widehat{\beta} = \beta$, $\beta < \widehat{\beta} < 1$ and $\widehat{\beta} = 1$, represent "sophisticate", "partial naivety" and "full naivety", respectively. An imperfectly self-aware consumer ($\widehat{\beta} > \beta$) exhibits not only time-inconsistent preferences, but also a decision inconsistency since his self tomorrow will typically deviate from the naive prediction made by his self today.

The DM Pricing Coming back to the 3-date investment good consumption framework, it should be noted that the gross benefit b comes at date 2, but that the price p and the cost c are incurred at date 1. Hence, on behalf of the consumer's self 1, the decision whether to exercise at date 1 depends on the sign of his net benefit $\beta\delta b - p - c$; the down payment $-L$ does not enter into consideration since it is a sunk cost. However, from the point of view of the consumer's self 0, his prediction about self 1's decision whether to exercise depends on the sign of $\widehat{\beta}\delta b - p - c$, which is less demanding than the real consumption criteria since $\widehat{\beta} \geq \beta$, reflecting the consumer's possible naivety. At date 0, taking into account the predicted self 1's behavior, the consumer decides whether to join the club, i.e., to accept the two-part tariff (L, p) or not, according to the following inequality:

$$\beta\delta\left(-L + \int_{-\infty}^{\widehat{\beta}\delta b - p} (\delta b - p - c)\, dF(c)\right) \geq 0.$$

The multiplicative term $\beta\delta$ can be directly cancelled so the consumer's long-run decision rule becomes:

$$-L + \int_{-\infty}^{\widehat{\beta}\delta b - p} (\delta b - p - c)\, dF(c) \geq 0. \tag{1}$$

Equation (1) constitutes the participation constraint of the consumer. Conditional on the consumer's participating at date 0 and consuming at date 1, the firm will incur a cost[4] a at date 1, which can be interpreted as the per-usage cost of equipments. Under DM's complete information structure, there is no asymmetry of information between the firm and the consumer, i.e., the firm knows the true value of β and the consumer's self image $\widehat{\beta}$

[4]DM also introduces an initial setup cost K for the firm. Since this cost plays a role neither in the consumer's decision making, nor in the firm's pricing, we disregard it, ignoring the firm's participation constraint.

when contracting with the consumer. Thus the firm's pricing behavior becomes a first-best contract design between a time-consistent principal and a time-inconsistent agent:

$$\max_{L,p} \quad \delta \left[L + F\left(\beta\delta b - p\right)\left(p - a\right)\right], \text{ s.t., (1).} \tag{2}$$

Notice that the short-run impatience in the probability of exercising is β, not $\widehat{\beta}$, which reflects that the final consumption decision will be made by the consumer's self 1. At the optimum, the participation constraint (1) binds, as the firm extracts the entire consumer surplus. The optimal pricing rule is then as follows:

$$p^{FB} - a = -\left(1 - \widehat{\beta}\right)\delta b \frac{f\left(\widehat{\beta}\delta b - p^{FB}\right)}{f\left(\beta\delta b - p^{FB}\right)} - \frac{F\left(\widehat{\beta}\delta b - p^{FB}\right) - F\left(\beta\delta b - p^{FB}\right)}{f\left(\beta\delta b - p^{FB}\right)} \tag{3}$$

The superscript "FB" stands for "first-best"; a detailed computation can be found in the proof of Proposition 0 provided in the Appendix.

If the consumer is fully patient and time-consistent, i.e., $\widehat{\beta} = \beta = 1$, then Equation (3) boils down to $p = a$, which reflects the standard two-part tariff pricing property, i.e., marginal cost per-usage pricing plus a lump-sum fee which extracts all the consumer's surplus. By contrast, if the consumer is time-inconsistent, i.e., $\beta \leq \widehat{\beta} \leq 1$ with at least one strict inequality, the right-hand side of Equation (3) is always negative. Hence, the marginal cost per-usage pricing property is violated when the firm faces a quasi-hyperbolic consumer, no matter whether he is sophisticated or naive. We separate the difference between p and a into two parts by defining $A \equiv -\left(1 - \widehat{\beta}\right)\delta b \frac{f\left(\widehat{\beta}\delta b - p^{FB}\right)}{f\left(\beta\delta b - p^{FB}\right)}$ and $B \equiv -\frac{F\left(\widehat{\beta}\delta b - p^{FB}\right) - F\left(\beta\delta b - p^{FB}\right)}{f\left(\beta\delta b - p^{FB}\right)}$. Term A is a per-usage price discount, whose goal is to secure a short-run impatient consumer's participation at date 0, based on his naive prediction of his future behavior. Term B, more specifically, the difference between two probabilities in the numerator, reflects the firm's ability to extract surplus from the consumer by taking advantage of his naivety.[5] Notice that both A and B are negative if $\beta < \widehat{\beta} < 1$. Furthermore, $A = 0$ if the consumer is fully naive; $B = 0$ if the consumer is sophisticated. The following proposition collects the main insights of DM about the investment good pricing for a time-inconsistent and naive consumer under complete information.

Proposition 0 (The first-best pricing in DM)

(i) For a quasi-hyperbolic consumer, whether sophisticated or naive, the per-usage price is lower than the marginal cost: when $\beta < 1$,

$$p^{FB} < a = p^*,$$

where p^* stands for the standard marginal cost per-usage pricing.

(ii) For a sophisticated quasi-hyperbolic consumer, the real consumption probability and the firm's profit are both unaffected by the consumer's short-run impatience β.

(iii) For a naive quasi-hyperbolic consumer, the firm's profit is increasing in $\widehat{\beta}$ *when* β *is given.*

We provide an illustrative proof of this result in the Appendix.[6]

[5] "the consumer will take advantage of the discount less often than he anticipates, and the firm will make higher profits", Page 364, DM.

[6] In the proof of Proposition 0 in the appendix, we only present the intuitive necessary condition of the solution. For the sufficiency of the first-order condition and the existence of p^{FB}, we need a technical ABP (*asymptotically bounded peaks*) assumption on the density f of c, which rules out the possibility of unbounded peaks on the tails of $f(c)$. The fully rigorous proof of the proposition and the details of ABP can be seen from DM, Page 362-363.

The main insight of Proposition 0 is that the consumer's time-inconsistency has an impact on the pricing scheme, in contrast with traditional industrial organization theory. Specifically, the optimal marginal cost pricing in a two-part tariff is violated. The below-marginal-cost per-usage pricing brings a loss to the firm at date 1, but by binding the participation constraint (1) of the consumer, the firm gets a full reimbursement of the loss in per-usage price by inflating the entrance fee L. This reflects a fact that the firm takes into account the consumer's myopia and adjusts the payoff structure across periods accordingly. This skewed pricing scheme[7] helps reduce the market distortion and the profit loss due to the consumer's self-control problem.

To discuss the impact of time-inconsistency to the firm's profits, it is useful to distinguish two cases.

I. Sophistication When the consumer is sophisticated, part (ii) of Proposition 0 tells us that the pricing structure change due to time-inconsistency leads to no loss for the firm. Indeed, it follows from the proof of Proposition 0 in the Appendix, that $\beta\delta b - p^{FB}_{\widehat{\beta}=\beta} = \delta b - a$, i.e., for a sophisticated consumer, the consumption probability is independent from his short-run impatience and coincides with a time-consistent consumer's one. Thus there is no distortion on the investment good consumption when both players are sophisticated even if one party, the consumer, suffers from a self-control problem. Notice also that L is a committed pre-payment, through which the firm can contract with the long-run patient self of the consumer about the participation in the club beforehand. In addition, the consumer's long-un self also benefits from this skewed pricing scheme. The combination of a long-run

[7] We call $\left(L^{FB}, p^{FB}\right)$ a skewed pricing scheme, in contrast with the standard balanced two-part tariff pricing (L^*, p^*), where p^* equals marginal cost and L^* equals the consumer surplus.

committed L and a short-run discount, that is, $p^{FB} < a$, avoids the impulsive and inefficient decision of the short-run self 1, which helps the sophisticated time-inconsistent consumer overcome his own self-control problem. At last, it is worth emphasizing that the firm cannot take advantage from the sophisticated consumer's self-control problem, unlike when she deals with a naive consumer.

II. Naivety By contrast, if the consumer is naive, the firm will exploit the consumer's naivety to some extent and extract more surplus by taking advantage of the wedge between the consumer's expectations and his actual behavior. In that case, the ignorance of one party in the contract benefits another party. And this exploitation behavior becomes more severe as the extent of naivety, as measured by $\widehat{\beta} - \beta$, increases.

DM only work out the first-best contract between a time-consistent firm and a time-inconsistent consumer; although there is uncertainty at the date of contracting, there is no private information on the consumer's side. In the next two sections, we extend this 3-date investment good pricing model to the case of asymmetric information. We will see that some results in Proposition 0 no longer hold under the asymmetric information structure.

3 Second-best Pricing for an Investment Good under Uniform Exercising Cost

In this section, we derive the second-best pricing rule in the investment good framework of DM when the consumer suffers from a self-control problem and, moreover, may hold private information in different dimensions: (i) consumers may differ with respect to the benefit b of consuming the investment good, while the seller does not know the value of b, which is the standard dimension of adverse selection on the efficiency parameter of the consumer; (ii) When the consumer is sophisticated, the firm may not be aware of the exact extent of the consumer's short-run impatience, i.e., β is the consumer's private information; (iii) for a naive consumer, the firm may not be aware of the exact degree of the consumer's naivety, i.e., $\widehat{\beta}$ is the consumer's private information, where this last scenario is like a non-common priors contracting: the firm knows the true β and the consumer privately holds a naive belief $\widehat{\beta}$. In the following, we treat these three cases in turn to assess the impact of asymmetric information on the 3-date investment good model.[8]

For the sake of tractability, throughout this section we restrict ourselves to the case of an uniform distribution of exercising cost c.[9] In the next section, we extend some of our results to the case of a non-uniformly distributed exercising cost.

First, we have the following lemma which simplifies the first-best pricing under uniform distribution of c.

Lemma 1 *If c is drawn from an uniform distribution on a support $[\underline{c}, \overline{c}]$ that satisfies the boundary condition $\underline{c} < \delta b - a < \overline{c}$, then, for any $\beta \leq \widehat{\beta} \leq 1$, $p_S^{FB} = p_N^{FB} = a - (1 - \beta)\, \delta b$, where the subscripts "S" and "N" stand for sophisticated and naive consumer, respectively. Thus, $\frac{\partial p^{FB}}{\partial \widehat{\beta}} = 0$. Besides, L^{FB} is increasing in $\widehat{\beta}$ for a given β.*

Proof of Lemma 1. c is uniformly distributed, Equation (3) simplifies

[8] This paper prevents the complicated multi-dimensional screening problem, which may constitute a direction of future work based on the latest development and progress in the techniques of adverse selection theory.

[9] The uniform distribution is an extreme case satisfying the ABP assumption (see DM and footnote 7), since there is no peak in the whole density on the support of c.

to $p^{FB} - a = -\left(1-\widehat{\beta}\right)\delta b - \left(\widehat{\beta}\delta b - \beta\delta b\right) = -\left(1-\beta\right)\delta b$, which is independent of $\widehat{\beta}$. The boundary condition $\underline{c} < \delta b - a < \overline{c}$ ensures that for any $\beta \leq \widehat{\beta} \leq 1$, the consumption probability belongs to the interval $(0,1)$, since $\Pr\left(c \leq \beta\delta b - p^{FB}\right) = \Pr\left(c \leq \delta b - a\right) = \frac{\delta b - a - \underline{c}}{\overline{c} - \underline{c}}$. Using the fact that (1) is binding, we obtain that

$$L^{FB} = \int_{-\infty}^{\widehat{\beta}\delta b - p^{FB}} \left(\delta b - p^{FB} - c\right) dF(c). \tag{4}$$

Finally $\frac{\partial p^{FB}}{\partial \widehat{\beta}} = 0$ implies that $\frac{\partial L^{FB}}{\partial \widehat{\beta}} = \left(1-\widehat{\beta}\right)(\delta b)^2 f(\widehat{\beta}\delta b - p^{FB}) \geq 0$, where the equality holds only if $\widehat{\beta} = 1$. ■

This lemma tells us that under uniform exercising cost, there is no distortion in the consumption probability of the investment good even for a naive time-inconsistent consumer because the per-usage price is insensitive to the naivety and coincides with that of a sophisticated consumer. The down payment L is the only instrument through which the firm can exploit a time-inconsistent consumer when the latter is naive.

We now investigate whether the main features of investment good pricing for a quasi-hyperbolic consumer emphasized in Proposition 0 under symmetric information still hold or not under the three kinds of incomplete information structures outlined above.

3.1 Private Information on b

Suppose first that the intrinsic efficiency parameter, i.e., the consumption benefit b of the investment good, is the consumer's private information. To consider one effect at a time, we assume that the time preference parameters of the consumer, β and $\widehat{\beta}$, are publicly observed by the firm. This allows to concentrate on the adverse selection on b and to avoid a multi-dimensional private information problem. The distribution of b satisfies the following assumption.

Assumption A1 *$b \in \left[\underline{b}, \overline{b}\right]$ is drawn from a c.d.f. H with p.d.f. h and is independent from c, but subject to the boundary condition*

$$\underline{c} < \delta\underline{b} - a - \frac{\delta}{h(\underline{b})} < (2-\beta)\,\delta\overline{b} - a < \overline{c}. \tag{5}$$

Moreover, the distribution of b satisfies the monotone hazard rate property, i.e., $\frac{1-H(b)}{h(b)}$ is decreasing in b.

The boundary conditions (5) ensure that the decisions of consumers will be interior, i.e., for any type b of the consumer and any $\beta \leq \widehat{\beta} \leq 1$, the resulting predicted and real consumption probabilities at date 1 strictly belong to $(0,1)$.[10] The purpose of a large enough support for c is to provide enough variability in c so that no one is excluded from the market, e.g., $\underline{c}$ is not so large that the consumption probability of a type $\underline{b}$ consumer is 0, and no pooling at the top arises, e.g., $\overline{c}$ is not so small that there exists $b^{\circ} \in \left[\underline{b}, \overline{b}\right)$, such that for all $b \in \left[b^{\circ}, \overline{b}\right]$, type-$b$ consumers are 100% sure to exercise at date 1.

In a two-part tariff, the firm has two instruments for pricing: the entrance fee L and the per-usage price p. Under asymmetric information, the firm attempts at finding an incentive compatible and individually rational menu of prices $(L\,(b),p\,(b))$, which maximizes its expected profit. Formally, the firm's program is:

$$\max_{L,p} \int_{\underline{b}}^{\overline{b}} \delta\left(L\,(b) + F(\beta\delta b - p\,(b))\,(p\,(b) - a)\right) h(b)db \tag{6}$$

[10]See the proof of Proposition 1.

subject to the following incentive compatibility (IC) and individual rationality (IR) constraints:

$$\text{ICs: } \forall b \in \left[\underline{b}, \overline{b}\right], b \in \arg\max_{b' \in \left[\underline{b}, \overline{b}\right]} u\left(b, b'\right), \tag{7}$$

$$\text{IRs: } \forall b \in \left[\underline{b}, \overline{b}\right], u\left(b, b\right) \geq 0, \tag{8}$$

where

$$u\left(b, b'\right) \equiv -L\left(b'\right) + \int_{-\infty}^{\widehat{\beta}\delta b - p(b')} \left(\delta b - p\left(b'\right) - c\right) dF(c)$$

is the left-hand side of (1) after a choice $\left(L\left(b'\right), p\left(b'\right)\right)$ from the two-part tariff menu. Define

$$v\left(b, \beta, \widehat{\beta}\right) \equiv \int_{-\infty}^{\widehat{\beta}\delta b - p} \left(\delta b - p - c\right) dF(c). \tag{9}$$

Spence-Mirrlees' single crossing condition for the program (6) subject to (7) and (8) requires that $\frac{\partial v}{\partial b}$ be monotone in p, i.e., $\frac{\partial^2 v}{\partial b \partial p}$ is globally positive or negative. In that case, the program can be transformed into a tractable optimal control problem. Here

$$\frac{\partial^2 v}{\partial b \partial p} = -\widehat{\beta}(1 - \widehat{\beta})\delta^2 b f'(\widehat{\beta}\delta b - p) - \delta f(\widehat{\beta}\delta b - p) \tag{10}$$

When c is uniformly distributed, $\frac{\partial^2 v}{\partial b \partial p} < 0$ globally since then $f' = 0$.

Discussion on the Distribution of c Before turning to the main results of second-best pricing with respect to b, it will be useful to discuss the trade-off between the tractability of the model and the generality of the exercising cost distribution. First, we observe that the Spence-Mirrlees condition is satisfied as long as $f'(c) > 0$ for all c. For robustness concern, we have checked what happens in this case. It turns out that the analysis only involves more complication like bunching, while not leading to any new qualitative insights compared to the case of a uniform c. This is why we keep the assumption

$f' = 0$ and its associated simpler and explicit solution to illustrate the main second-best insights. Second, observe from (10) that if the distribution of c is "close to uniform", i.e., the density does not change too sharply, then the Spence-Mirrlees condition works as well, from which we can anticipate that the results in this subsection are not only restricted to a uniform c, but also valid to a relatively broad family of exercising cost distributions. Third, allowing a more general distribution for c which would violate the Spence-Mirrlees' single crossing condition may be an interesting starting point for future work based on the latest development of literature on principal-agent models under adverse selection (see for instance Araujo and Moreira (2010)).

Under Assumption A1, we find that when the consumer holds private information, the second-best pricing no longer satisfies the below-marginal-cost per-usage pricing property for some types. Moreover, the short-run impatience of the consumer now involves some loss on the firm's profit under incomplete and asymmetric information, unlike in the first-best setting of DM. The details are stated in the following two propositions and their associated comments and corollaries.

Proposition 1 *Under Assumption A1, the optimal price menu ($L^{SB}(b)$,$p^{SB}(b)$), where the superscript "SB" stands for "second-best", is separating, with*

$$p^{SB}(b) = a - (1-\beta)\,\delta b + \delta \frac{1 - H(b)}{h(b)}.$$

$p^{SB}(b)$ is decreasing in b and $L^{SB}(b)$ is increasing in b. Moreover, $p^{SB}(\bar{b}) = p^{FB}(\bar{b})$, while $p^{SB}(b) > p^{FB}(b)$ for $b < \bar{b}$.

The proof of Proposition 1 is in the Appendix.

When we introduce private information on the traditional intrinsic efficiency parameter b, the second-best per-usage price is upward distorted with respect to the first-best described in Lemma 1. First, $p^{SB}(b)$ coincides with the first-best value $p^{FB}(b)$ only for the most efficient consumer with type $\overline{b}$, which is the usual "no distortion at the top" result. So the efficient consumption probability of the investment good is only preserved for the short-run impatient consumer who holds the highest willingness to pay. By contrast, if $b < \overline{b}$, $p^{SB} > p^{FB}$ so $\beta\delta b - p^{SB} < \delta b - a$, i.e., the real consumption probability is lower than the first-best thus there is a loss in efficiency even if the consumer is sophisticated, in contrast with part (ii) of Proposition 0.

Secondly, and more importantly, $p^{SB}(b)$, unlike p^{FB} for $\beta < 1$, is no longer necessarily lower than the marginal cost a. Whether p is higher or lower than a depends on the relative size of $(1-\beta)\,b$ and $\frac{1-H(b)}{h(b)}$. When b is close to $\overline{b}$, the below-marginal-cost per-usage pricing is preserved, but it is attenuated by the asymmetric information on b. As a result, when b becomes smaller, $p^{SB}(b)$ may become larger than a. Indeed, under the following assumption about the distribution and the support of b, the DM's below-marginal-cost per-usage pricing property no longer holds globally.

Assumption A1' *The benefit b from consuming is drawn as under A1, and in addition* $(1-\beta)\,\underline{b}h(\underline{b}) < 1$.

Observe that the additional boundary condition, $(1-\beta)\,\underline{b}h(\underline{b}) < 1$, is satisfied by most kinds of distributions of b when β is close to 1.

Under Assumption A1', we have the following corollary.

Corollary 1 *Under Assumption A1 and A1', and the same parametric/non-parametric assumptions as in Proposition 1, there exists a threshold type $\widetilde{b} \in \left(\underline{b}, \overline{b}\right)$ such that for $b < \widetilde{b}$, $p^{SB}(b) > a$.*

Proof of Corollary 1. That $(1-\beta)\,\underline{b}h(\underline{b}) < 1$ implies that $-(1-\beta)\,\delta\underline{b} + \delta\frac{1}{h(\underline{b})} > 0$, and thus that $p^{SB}(\underline{b}) > a$. Besides, we know that $p^{SB}\left(\overline{b}\right) =$

$p^{FB}\left(\overline{b}\right) < a$ and $p^{SB}(\cdot)$ is continuous and strictly decreasing in b, so there is a unique $\widetilde{b} \in \left(\underline{b}, \overline{b}\right)$ such that $p^{SB}\left(\widetilde{b}\right) = a$, and $p^{SB}(b) > a$ for all $b < \widetilde{b}$. ■

This corollary tells us that the firm asks for a higher-than-marginal-cost per-usage price for consumers with a low benefit of exercising. Figure 3 illustrates this result.

Figure 3 is about here

The failure of the below-marginal-cost per-usage pricing property comes from the upward second-best distortion of p. Would this upward distortion constitute a barrier to the first-best downward discount on p, thus hinder the full refund for the firm's profit when the consumer is sophisticated and time-inconsistent? The answer from the following Proposition 2 is "yes".

Proposition 2 *Under the parametric/non-parametric assumptions of Proposition 1, if the consumer is sophisticated ($\widehat{\beta} = \beta$) and privately knows b, then the firm's second-best profit is non-monotone in β, specifically: $\pi_S^{SB,b}$, where the subscript "S" stands for "sophisticated consumer", is strictly convex in the interval $\beta \in (0, 1]$ and reaches its minimum at $\beta = \frac{1}{2}$. In addition, $\pi_S^{SB,b}$ of β is symmetric around $\frac{1}{2}$, i.e., $\pi_S^{SB,b}(\beta) = \pi_S^{SB,b}(1-\beta)$.*

The proof of Proposition 2 is in the Appendix.

From this proposition, first, we see that the property in part (ii) of Proposition 0, that under symmetric information, the firm's profit is unaffected by the consumer's short-run impatience as long as the latter is sophisticated, does not hold any more under asymmetric information.

Under symmetric information, when facing a short-run impatient consumer, the strategy of the firm is to discount downward the per-usage price p, and instead enhance the flat fee L to take advantage of the consumer's long-run patience, so that the loss in the per-usage price can be fully recovered. But when the firm does not know the consumer's benefit of exercising at the contracting stage, optimal pricing calls for an upward distortion in the per-usage price. Indeed, this involves a negative impact on the firm's profit of a sophisticated consumer's time-inconsistency. From Proposition 2 and its proof, $\pi_S^{SB,b}$ is increasing in β in the region $\left[\frac{1}{2}, 1\right]$, i.e., the loss in the profit is more severe when β becomes farther from 1.

But when β is less than $\frac{1}{2}$, the firm's profit becomes increasing in the consumer's short-run impatience level $1-\beta$. Why does this happen? Notice that $1-\beta$ does not only represent the short-run impatience of the consumer, but also the difference in decision making between his long-run and short-run selves, i.e., the extent of his time-inconsistency. Deciding to participate to the club in advance is an exogenous commitment device that allows him to overcome the future short-run impatience on behalf of his long-run self; so when the consumer suffers from a severe self-control problem, i.e., when $\beta < \frac{1}{2}$, his rational long-run self is urgently willing to participate to the club and sacrifice more surplus when contracting, which allows the firm to ask him a higher entrance fee L than it would, if it were facing a relatively more patient consumer. This partially alleviates the loss in profit due to short-run impatience.

Finally, the second-best profit locus is symmetric with respect to $\beta = \frac{1}{2}$ in the interval $\beta \in (0, 1)$. This means that when the consumer is extremely impatient in the short-run (β is close to 0), the firm's expected profit goes back to the profit it would earn when facing a time-consistent consumer. An intuitive explanation for this phenomenon is as follows: when β approaches 0, the sophisticated long-run self anticipates that at date 1, it is likely that he will not exercise, so at date 0 he is willing to give much more rent L to the firm to help pre-commit his future self. When $\beta = 0$, in theory, the consumer is indifferent between accepting any (L, p) at date 0 since all the payments, cost, and benefit occur tomorrow and the day after are totally irrelevant from self 0's perspective; to maintain continuity, the second-best pricing scheme at $\beta = 0$ leads to the full reimbursement for the firm's loss in profit; and the firm's profit comes back to its level when the firm faces a time-consistent consumer. Figure 4 depicts the evolution of the firm's profit when β varies from 0 to 1 for a sophisticated consumer.

Figure 4 is about here

In Figure 4, the convex curve is the firm's expected profit when facing a sophisticated consumer with known β but unknown b, while the higher flat solid line represents the complete information profit in DM, which is unaffected by the sophisticated consumer's short-run impatience. The wedge between this line and the curve reflects the loss in profit due to asymmetric information on b.

For comparison purposes, it is useful to compute the expected consumer's informational rent as well as the utilitarian social welfare. We define the informational rent $U(b)$ by the left-hand side of (1), i.e., the participation constraint. As for the measure of social welfare, we follow the paternalistic definition of O'Donoghue and Rabin (2001), where $SW = \pi + \delta E(U(b))$. Paternalism means in this context that the consumer's discount factor in the social welfare function is his long-run discount factor δ, not his short-run discount factor $\beta\delta$. In this sense, a paternalistic social planner corrects for the consumer's short-run impatience in evaluating social welfare, but takes into account the consumer's equilibrium consumption, which leads to a rent $U(b)$ in our context. The following result holds.

Corollary 2 *Suppose that the information structure is as in Proposition 2 and that the consumer is sophisticated. Then,*

(i) The expected consumer informational rent $E(U(b))$ *is non-monotone and strictly concave in* β*, and reaches its maximum at* $\beta = \frac{1}{2}$*; in addition, it is symmetric around* $\beta = \frac{1}{2}$*.*

(ii) The social welfare is invariant in β*.*

The proof of Corollary 2 is in the Appendix.

The important information conveyed by this corollary is that though the firm's profit is affected by the consumer's short-run impatience in a non-monotonic way, paternalistic social welfare remains unaffected. The difference between symmetric and asymmetric information on b therefore lies only in the allocation of surplus between the firm and the consumer. From the expression for $p^{SB}(b)$ in Proposition 1, we obtain that the consumption probability at date 1 is:

$$F(\beta\delta b - p^{SB}(b)) = \delta b - a - \delta\frac{1 - H(b)}{h(b)},$$

which is independent of β. Hence, the short-run impatience of a sophisticated consumer has no impact on the social efficiency, while only the second-best information structure on b has an impact, i.e., the adjusted term $\delta\frac{1-H(b)}{h(b)}$, which is invariant in β. The evolution of β from 0 to 1 influences the bargaining power between the firm and the consumer, whereby we have introduced the intuition of non-monotonicity of informational rent with respect to β in the analysis after Proposition 2.

3.2 Sophisticated Consumer's Private Information on Short-run Patience Level β

In this subsection, we consider the following alternative information structure: the firm is perfectly aware of b and faces a sophisticated consumer, but the latter's short-run patience parameter β is his private information. The firm's prior about β is represented by a distribution over an interval $\left[\underline{\beta}, \overline{\beta}\right]$ with c.d.f. $G(\cdot)$ and p.d.f. $g(\cdot)$, where $0 < \underline{\beta} < \overline{\beta} \leq 1$. The consumer's exercising cost c and his β are independently drawn.

We first examine whether the Spence-Mirrlees' single crossing condition holds in this context, which requires that $\frac{\partial^2 v}{\partial \beta \partial p}$ to be globally positive or negative, where the function v follows from the definition in (9). We only additionally impose $\widehat{\beta} = \beta$ since the consumer is assumed to be sophisticated here. Thus

$$\frac{\partial^2 v}{\partial \beta \partial p} = -(1-\beta)(\delta b)^2 f'(\beta \delta b - p) = 0 \qquad (11)$$

when c is uniformly distributed. As a result, the firm cannot screen the consumer's β using any kind of pricing menu, since the consumer evaluates β and p separately: there exist no separating two-part tariffs $(L(\beta), p(\beta))$ which can induce the consumer to reveal his type β. The following lemma formalizes this idea.

Lemma 2 *When the sophisticated consumer privately observes $\beta \in \left[\underline{\beta}, \overline{\beta}\right]$ and the exercising cost c is uniformly distributed, there exist no separating price menu=$(L(\beta'), p(\beta'))_{\beta' \in \left[\underline{\beta}, \overline{\beta}\right]}$ which screen the consumer's type β.*

Proof of Lemma 2. Given a price scheme $(L(\beta'), p(\beta'))_{\beta' \in \left[\underline{\beta}, \overline{\beta}\right]}$, the IC constraints of the consumer are

$$\forall \beta \in \left[\underline{\beta}, \overline{\beta}\right], \beta = \arg\max_{\beta' \in \left[\underline{\beta}, \overline{\beta}\right]} \quad u(\beta, \beta'),$$

where $u(\beta,\beta') = -L(\beta') + \int_{\underline{c}}^{\beta\delta b - p(\beta')} (\delta b - p(\beta') - c)\, dF(c)$. When c is uniformly distributed,

$$u(\beta,\beta') = -L(\beta') + \frac{1}{\overline{c}-\underline{c}} q(p(\beta')) + \frac{1}{\overline{c}-\underline{c}} x(\beta) - \frac{\underline{c}}{\overline{c}-\underline{c}} \left(\delta b - \frac{1}{2}\underline{c}\right),$$

where $q(p) = \frac{1}{2}p^2 - (\delta b - \underline{c})\, p$ and $x(\beta) = \frac{1}{2}\beta(\delta b)^2$. Thus $u(\beta,\beta')$ is additively separable in (β,β'), which implies that all the types will choose the same element in the menu $(L(\beta'),p(\beta'))_{\beta'\in[\underline{\beta},\overline{\beta}]}$. Therefore there exists no separating pricing equilibrium. ■

Given this lemma, we search for an optimal second-best pooling pricing scheme. The following proposition holds.

Proposition 3 *When the consumer is sophisticated and β is his private information, the optimal second-best pooling price is*

$$p^{SB,\beta} = a - (1 - E(\beta))\,\delta b < a.$$

Moreover, the second-best profit of the firm when trading with a type β consumer, $\pi_S^{SB,\beta}(\beta)$, is decreasing in β.

Proof of Proposition 3. According to the prior of β, the firm's pooling program is

$$\max_{L,p} \int_{\underline{\beta}}^{\overline{\beta}} \delta\left(L + F(\beta\delta b - p)(p - a)\right) dG(\beta), \tag{12}$$

$$\text{s.t. IRs: } \forall\beta \in [\underline{\beta},\overline{\beta}],\ u(\beta) = -L + \int_{\underline{c}}^{\beta\delta b - p} (\delta b - p - c)\, dF(c) \geq 0 \tag{13}$$

From (13) we obtain that

$$u'(\beta) = (1-\beta)(\delta b)^2 f(\beta\delta b - p) > 0. \tag{14}$$

Therefore, letting $L = \int_{\underline{c}}^{\underline{\beta}\delta b - p} (\delta b - p - c)\, dF(c)$, all the individual rationality constraints are satisfied. Substitute into (12) yields:

$$\begin{aligned} & \max_p \int_{\underline{\beta}}^{\overline{\beta}} \delta \left(\int_{\underline{c}}^{\underline{\beta}\delta b - p} (\delta b - p - c)\, dF(c) + F(\beta\delta b - p)\,(p - a) \right) dG(\beta) \\ = & \max_p \quad \delta \left(\int_{\underline{c}}^{\underline{\beta}\delta b - p} (\delta b - p - c)\, dF(c) + \int_{\underline{\beta}}^{\overline{\beta}} F(\beta\delta b - p)\,(p - a)\, dG(\beta) \right). \end{aligned}$$

The first-order condition with respect to p is

$$\begin{aligned} & \int_{\underline{\beta}}^{\overline{\beta}} \left(F(\beta\delta b - p) - f(\beta\delta b - p)\,(p - a) \right) dG(\beta) \\ = & \ F(\underline{\beta}\delta b - p) + \left(1 - \underline{\beta}\right) \delta b f(\underline{\beta}\delta b - p) \end{aligned} \qquad (15)$$

The second-order condition with respect to p is:

$$\begin{aligned} \frac{\partial^2 \pi_S^{SB,\beta}}{\partial p^2} &= \delta \left(f\left(\underline{\beta}\delta b - p\right) + \left(1 - \underline{\beta}\right) \delta b f'\left(\underline{\beta}\delta b - p\right) \right) \\ &\quad + \delta \left(\int_{\underline{\beta}}^{\overline{\beta}} \left(f'\left(\beta\delta b - p\right)(p - a) - 2f\left(\beta\delta b - p\right) \right) dG\left(\beta\right) \right) \\ &\leq 0. \end{aligned}$$

where $\pi_S^{SB,\beta}$ denotes the expected profit. When c is uniformly distributed, $\frac{\partial^2 \pi_S^{SB,\beta}}{\partial p^2} = -\frac{\delta}{\bar{c} - \underline{c}} < 0$ globally. So the first-order condition is sufficient for a maximum. By imposing $f(\cdot) = \frac{1}{\bar{c} - \underline{c}}$ into (15), we obtain

$$\int_{\underline{\beta}}^{\overline{\beta}} \left(\beta\delta b - p - \underline{c} - (p - a) \right) dG\left(\beta\right) = \underline{\beta}\delta b - p - \underline{c} + \left(1 - \underline{\beta}\right) \delta b.$$

By rearranging,

$$p^{SB,\beta} = a - \left(1 - E\left(\beta\right)\right) \delta b < a. \qquad (16)$$

Finally, by (12), $\pi_S^{SB,\beta}\left(\beta\right) = \delta\left(L^{SB,\beta} + F\left(\beta\delta b - p^{SB,\beta}\right)\left(p^{SB,\beta} - a\right)\right)$, so that

$$\frac{\partial \pi_S^{SB,\beta}\left(\beta\right)}{\partial \beta} = \delta^2 b f\left(\beta\delta b - p^{SB,\beta}\right)\left(p^{SB,\beta} - a\right) < 0.$$

The result follows. ■

From (16), we see that the below-marginal-cost per-usage pricing property is preserved under this information structure. By contrast, part (ii) of Proposition 0, i.e., the invariance of the firm's profit with respect to β when the consumer is sophisticated no longer holds when we introduce asymmetric information on β: indeed, a more impatient consumer brings more profits to the firm. The reason is that the pricing is pooling. In fact, from (16), $p^{SB,\beta}$ is the average of p^{FB}s, representing the firm's second-best strategy when facing a heterogeneous and imprecise market. It is more likely that a higher β type consumer will exercise at date 1, which involves a higher per-usage loss to the firm due to the below-marginal-cost per-usage pricing property.

3.3 Naive Consumer's Private Information on $\widehat{\beta}$

In this subsection the firm knows the consumer's short-run patience level β and his benefit from exercising b, but does not know to what extent the consumer himself is self-aware of his short-run impatience; that is, the consumer's perception of β, $\widehat{\beta}$, is his private information. Thus, while the firm knows the actual short-run impatience of the consumer β, it does not observe the self image $\widehat{\beta}$ of the consumer. The firm only knows a prior of $\widehat{\beta}$ on $[\beta, 1]$ with a c.d.f. $J(\cdot)$ and density $j(\cdot)$. We assume that $\widehat{\beta}$ and c are independently distributed. The consumer privately holds a belief $\widehat{\beta}$ about his short-term impatience, which may differ from the true β. Because he is naive, he does not update his belief after observing the pricing scheme offered by the firm i.e., the pricing scheme does not constitute a signal to the cognition of the consumer. The both parties contract based on their non-common priors.

As in Section 3.2, Spence-Mirrlees' single crossing condition is not satisfied under uniform distribution of exercising cost c since

$$\frac{\partial^2 v}{\partial \widehat{\beta} \partial p} = -\left(1-\widehat{\beta}\right)(\delta b)^2 f'\left(\widehat{\beta}\delta b - p\right) = 0, \tag{17}$$

because $f'(\cdot) = 0$, where v is defined by (9). The following lemma parallels Lemma 2.

Lemma 3 *When the firm is not aware of the consumer's degree of naivety $\widehat{\beta} \in [\beta, 1]$, and the exercising cost c is uniformly distributed, there exist no separating price schemes $(L\left(\widehat{\beta}\right), p\left(\widehat{\beta}\right))$ which screen the consumer's type $\widehat{\beta}$.*

The proof is exactly similar to that of Lemma 2, and therefore we omit it here.

Given this lemma, we search for an optimal second-best pooling pricing scheme. One has the following result.

Proposition 4 *When the firm does not know to which extent the consumer is naive, the optimal second-best pooling price is $p^{SB,\widehat{\beta}} = a - (1-\beta)\delta b = p^{FB}$, and the firm's profit is the same as that it would obtain if it was to face a sophisticated consumer.*

Proof of Proposition 4. Based on the firm's prior on $\widehat{\beta}$, its pooling program is

$$\max_{L,p} \int_{\beta}^{1} \delta\left(L + F(\beta\delta b - p)(p-a)\right) dJ\left(\widehat{\beta}\right), \tag{18}$$

$$\text{s.t. IRs: } \forall \widehat{\beta} \in [\beta, 1], u\left(\widehat{\beta}\right) = -L + \int_{\underline{c}}^{\widehat{\beta}\delta b - p} (\delta b - p - c)\, dF(c) \geq 0. \tag{19}$$

From (19) we obtain that

$$u'\left(\widehat{\beta}\right) = \left(1-\widehat{\beta}\right)(\delta b)^2 f\left(\widehat{\beta}\delta b - p\right) \geq 0, \tag{20}$$

where the equality holds only if $\widehat{\beta}=1$. Letting $L=\int_{\underline{c}}^{\beta\delta b-p}(\delta b-p-c)\,dF(c)$, all the individual rationality constraints are satisfied. Substituting into (18) yields

$$\begin{aligned} &\max_{p}\int_{\beta}^{1}\delta\left(\int_{\underline{c}}^{\beta\delta b-p}(\delta b-p-c)\,dF(c)+F(\beta\delta b-p)(p-a)\right)dJ\left(\widehat{\beta}\right)\\ =\;&\max_{p}\quad\delta\left(\int_{\underline{c}}^{\beta\delta b-p}(\delta b-p-c)\,dF(c)+F(\beta\delta b-p)(p-a)\right)\end{aligned}$$

which is just the first-best program when the firm faces a sophisticated consumer, as can be seen from (2) with (1) binding, and $\widehat{\beta}=\beta$: $p=a-(1-\beta)\,\delta b=p^{FB}$. ■

The distinctive feature of the present setting, in which there is private information on naivety, is that the different types only differ at date 0; till date 1 all the types coincide since the consumption decision by the consumer is made by the common and real β. So when we involve asymmetric information in the dimension of naivety, the firm gives up the cheating intention in contracting toward naivety since she is not aware of the consumer's self-awareness, i.e., "the source to be cheated is imprecise". While the firm offers a sophisticated consumer's price to the whole naive population and keeps the first-best profit by a sophisticated consumer. In the second-best, (iii) of Proposition 0 may not hold.

4 Second-best Screening under Non-uniform Exercising Cost

Section 3 documented several violations of the first-best pricing properties for an investment good when different forms of asymmetric information were introduced. These results were obtained under the assumption of a uniformly distributed exercising cost. In the case where the private benefit b from consuming is the consumer's private information, the qualitative results of second-best pricing are robust for some reasonable perturbation of the distribution of the exercising cost, i.e., when the density of exercising cost does not change too much, so that the Spence-Mirrlees condition remains satisfied. An undesirable feature of assuming a uniform cost distribution was that, when the consumer's short-run time preference and his self-awareness become his private information, separating price schedules are ruled out. In this section, we allow some dispersion in the density of c in order to investigate separating pricing equilibria when β or $\widehat{\beta}$ are the consumer's private information.

According to (11) and (17), we cannot allow any peaks in the density of c on its support for otherwise single crossing may fail. A typical example of a non-uniform distribution satisfying the *no peak density* restriction is an exponential distribution with density function $f(c) = \lambda e^{-\lambda c}$ on $[0, +\infty)$, for $\lambda > 0$. So $f'(c) < 0$ globally, i.e., single crossing is satisfied. We consider a more general family of cost distribution:

Assumption A2 *At date 0, both the firm and the consumer know that the realization of cost c at date 1 follows a distribution on $[\underline{c}, +\infty)$ such that $f'(c) < 0$ globally and $\underline{c} < \delta b - a$. Moreover, the distribution of c satisfies the monotone hazard rate property, i.e., $\frac{1-F(c)}{f(c)}$ is non-increasing in c.*

The boundary condition $\underline{c} < \delta b - a$ ensures that the predicted and the real consumption probabilities are both strictly positive for all the types of β and $\widehat{\beta}$ thus ruling out corner solutions.

4.1 Private β of a Sophisticated Consumer

Suppose as in Section 3.2 that the firm faces a sophisticated consumer, whose short-run impatience parameter it does not know. We denote as before by $G(\cdot)$ and $g(\cdot)$ the c.d.f. and p.d.f. of the firm's prior about β.

Under Assumption A2, the single crossing condition holds globally:

$$\frac{\partial^2 v}{\partial \beta \partial p} = -(1-\beta)(\delta b)^2 f'(\beta\delta b - p) > 0.$$

Hence the prerequisite for a screening menu is satisfied.[11] The firm's program for a price menu is:

$$\max_{L(\beta),p(\beta)} \int_{\underline{\beta}}^{\overline{\beta}} \delta \left(L(\beta) + F(\beta\delta b - p(\beta))(p(\beta) - a)\right) dG(b) \tag{21}$$

subject to:

$$\text{ICs: } \forall\beta \in \left[\underline{\beta}, \overline{\beta}\right], -L(\beta) + v(\beta, p(\beta)) = \max_{\beta' \in \left[\underline{\beta},\overline{\beta}\right]} \left(-L(\beta') + v(\beta, p(\beta'))\right) \tag{22}$$

$$\text{IRs: } \forall\beta \in \left[\underline{\beta}, \overline{\beta}\right], -L(\beta) + v(\beta, p(\beta)) \geq 0 \tag{23}$$

The following proposition depicts the second-best screening prices feature under this information structure and the distribution assumptions.

Proposition 5 *If the consumer is sophisticated and β is his private information, and if the distribution of c satisfies A2, then an optimal separated screening menu $(L^{SB}(\beta)$,$p^{SB}(\beta))$ exists:*

[11] For any kind of distribution of c, the firm can always implement a pooling equilibrium, which is depicted by Section 3.2. But when the distribution of c converges from a decreasing density to a uniform, the feasible set of the firm's pricing scheme (pooling or separating) shrinks suddenly and discontinuously, i.e., when c is uniformly distributed, then only a pooling scheme is feasible. This parallels the relationship between Section 4.2 and 3.3.

(i) $p^{SB}(\beta)$ *is continuous in* β*.*

(ii) The per-usage price for type $\overline{\beta}$ *coincides with its first-best level (no distortion at the top):* $p^{SB}\left(\overline{\beta}\right) = p_S^{FB}\left(\overline{\beta}\right) = a - \left(1 - \overline{\beta}\right)\delta b$*.*

(iii) The per-usage price for type $\underline{\beta}$ *is distorted downward with respect to its first-best level:* $p^{SB}\left(\underline{\beta}\right) < p_S^{FB}\left(\underline{\beta}\right)$*.*

(iv) For the types between $\underline{\beta}$ *and* $\overline{\beta}$*, there may exist some bunching interval (see Line AC in Figure 5), i.e, pooling per-usage price. But for the types outside the bunching interval,* $p^{SB}(\beta)$ *is strictly increasing in* β*. For all* $\beta \in \left(\underline{\beta}, \overline{\beta}\right)$*, the per-usage price is distorted downward with respect to its first-best level:* $p^{SB}(\beta) < p_S^{FB}(\beta)$*.*

(v) $L^{SB}(\beta)$ *is strictly decreasing in* β *for the screened types (outside the bunching intervals), while is invariant in* β *in each bunching interval.*

The proof of Proposition 5 is in the Appendix.

An illustrative shape of $p^{SB}(\beta)$ and its relationship with the first-best are shown in Figure 5. Form Part (ii)(iii)(iv) of Proposition 5, the below-marginal-cost per-usage pricing property is still preserved for the whole population when the firm faces a pool of diversely short-run impatient but sophisticated consumers.

Figure 5 is about here

Proposition 5 implies an interesting phenomenon: when a sophisticated consumer holds private information about β and the firm can screen among different types, the asymmetric information leads to over-consumption of the investment good except for the most short-run patient type ($\overline{\beta}$). From (3), we can compute the first-best consumption probability of a sophisticatedly time-inconsistent consumer, and obtain that it achieves the efficient time-consistent consumer's level:

$$F\left(\beta\delta b - p_S^{FB}(\beta)\right) = F\left(\beta\delta b - a + (1-\beta)\delta b\right) = F\left(\delta b - a\right). \qquad (24)$$

While here the second-best consumption probability is $F\left(\beta\delta b - p^{SB}(\beta)\right) \geq F(\delta b - a)$ since $p^{SB}(\beta) \leq p_S^{FB}(\beta)$, where the equality only holds for $\beta = \overline{\beta}$. This over-consumption can be also illustrated geometrically from Figure 5. The critical fact is that the first-best per-usage price is increasing in β; on the other hand, the set of incentive compatibility constraints on the interval $\left[\underline{\beta}, \overline{\beta}\right]$ requires that the second-best per-usage price is non-decreasing in β (Region AC represents some "bunching" types): these lead to a downward wedge from p_S^{FB} to p^{SB} for all βs.

But the over-consumption in the second-best is not necessarily true globally for all types in the next subsection when $\widehat{\beta}$ is the consumer's private information since in general we do not have a monotonic first-best p across $\widehat{\beta}$ on $[\beta, 1]$ (see Figure 6 and the subsequent analysis).

4.2 Private $\widehat{\beta}$ of a Naive Consumer

Now, as in Section 3.3, the firm is aware of the imperfect self-awareness of the consumer (naivety) but is imprecise about the extent of his naivety ($\widehat{\beta}$). The firm's prior about $\widehat{\beta}$ is just a distribution over the interval $[\beta, 1]$ with c.d.f. $J(\cdot)$ and density $j(\cdot)$.

Since we adopt the same preassumption as in Section 4.1 (Assumption A2), and according to (17), single crossing condition is globally satisfied. Therefore it may be possible to screen the consumer's naivety type through a menu of two-part tariffs $(L^{SB}\left(\widehat{\beta}\right), p^{SB}\left(\widehat{\beta}\right))$. The firm's program for an optimal pricing menu in this information structure is:

$$\max_{L(\widehat{\beta}), p(\widehat{\beta})} \int_{\beta}^{1} \delta\left(L\left(\widehat{\beta}\right) + F\left(\beta\delta b - p\left(\widehat{\beta}\right)\right)\left(p\left(\widehat{\beta}\right) - a\right)\right) dJ\left(\widehat{\beta}\right) \qquad (25)$$

subject to:

$$\text{ICs: } \forall\widehat{\beta} \in [\beta, 1], -L\left(\widehat{\beta}\right)+v\left(\widehat{\beta}, p\left(\widehat{\beta}\right)\right) = \max_{\widehat{\beta}' \in [\beta,1]} \left(-L\left(\widehat{\beta}'\right) + v\left(\widehat{\beta}, p\left(\widehat{\beta}'\right)\right)\right) \tag{26}$$

$$\text{IRs: } \forall\widehat{\beta} \in [\beta, 1], -L\left(\widehat{\beta}\right) + v\left(\widehat{\beta}, p\left(\widehat{\beta}\right)\right) \geq 0 \tag{27}$$

The following proposition depicts the second-best screening prices feature under Assumption A2.

Proposition 6 *If the consumer is naive and $\widehat{\beta}$ is his private information, and if the distribution of c satisfies Assumption A2 and b is large enough, then an optimal separated screening menu exists:*

(i) $p^{SB}\left(\widehat{\beta}\right)$ is continuous in $\widehat{\beta}$.

(ii) The per-usage price for $\widehat{\beta} = 1$ coincides with its first-best (no distortion at the "top" (fully naive)).

(iii) The per-usage price for $\widehat{\beta} = \beta$ is distorted downward with respect to its first-best level.

(iv) For the types between $\widehat{\beta} = \beta$ and $\widehat{\beta} = 1$, there may exist some bunching interval (see Figure 6), i.e, pooling per-usage price. But for the types outside the bunching interval, $p^{SB}\left(\widehat{\beta}\right)$ is strictly increasing in $\widehat{\beta}$.

(v) $L^{SB}\left(\widehat{\beta}\right)$ is strictly decreasing in $\widehat{\beta}$ for the screened types (outside the bunching intervals), while is invariant in $\widehat{\beta}$ in each bunching interval.

The proof of this proposition is in the Appendix.

An illustrative shape of $p^{SB}\left(\widehat{\beta}\right)$ and its comparison with the first-best for the naive consumer are shown in Figure 6. Also from part (ii)(iii)(iv) of Proposition 6,

$$\frac{\partial p^{SB}\left(\widehat{\beta}\right)}{\partial\widehat{\beta}} \geq 0 \Rightarrow \forall\widehat{\beta}, p^{SB}\left(\widehat{\beta}\right) \leq p^{SB}\left(\widehat{\beta} = 1\right) = p_N^{FB}\left(\widehat{\beta} = 1\right) < a.$$

Thus the below-marginal-cost per-usage pricing property is still preserved when the firm faces a pool of diversely naive consumers.

Figure 6 is about here

Second, part (iii) of Proposition 6 tells us that the asymmetric information on the consumer's naivety extent involves definitely an over-consumption of the investment good for the sophisticated type consumer since $p^{SB} < p_S^{FB}$ when $\widehat{\beta} = \beta$. According to (24), this involves an over-consumption at date 1 with respect to the socially efficient consumption probability since when $\widehat{\beta} = \beta$, $F\left(\beta\delta b - p^{SB}\right) > F\left(\beta\delta b - p_S^{FB}\right) = F\left(\delta b - a\right)$.

But from the illustration in Figure 6, the over-consumption with respect to the first-best complete information scenario does not necessarily appear for naive consumers. For example, $p^{SB}\left(\widehat{\beta}\right) \geq p_N^{FB}\left(\widehat{\beta}\right)$ for the bunching types $\left[\widehat{\beta}_1, \widehat{\beta}_2\right]$, which means,

$$F\left(\beta\delta b - p^{SB}\left(\widehat{\beta}\right)\right) \leq F\left(\beta\delta b - p_N^{FB}\left(\widehat{\beta}\right)\right), \forall\widehat{\beta} \in \left[\widehat{\beta}_1, \widehat{\beta}_2\right],$$

where the equality holds only at the two boundaries $\widehat{\beta}_1$ and $\widehat{\beta}_2$. This is different from the global over-consumption in the Subsection 4.1, whereby the reason is that the per-usage price under complete information is not necessarily non-decreasing in $\widehat{\beta}$ (see the locus of the curve $p_N^{FB}\left(\widehat{\beta}\right)$ in Figure 6 and the analytical expression in (3)).

Third, for a fully naive consumer, his consumption probability is not influenced by the firm's asymmetric information ($p^{SB}\left(\widehat{\beta} = 1\right) = p_N^{FB}\left(\widehat{\beta} = 1\right)$).

More importantly and interestingly, though the consumer's informational rent $U\left(\widehat{\beta}\right)$ ($= -L^{SB}\left(\widehat{\beta}\right) + v\left(\widehat{\beta}, p^{SB}\left(\widehat{\beta}\right)\right)$) is increasing in $\widehat{\beta}$, this is just the fictitious utility/rent based on naive expectation at date 0. The following corollary states the wedge between the fictitious and the real rent/consumption surplus of a naive consumer, and describes both their evolutions across the degree of naivety, $\widehat{\beta}-\beta$.

Corollary 3 *Under a given and public β, diverse and private $\widehat{\beta}$, and A2, the consumer's fictitious rent $U\left(\widehat{\beta}\right)$ is increasing in $\widehat{\beta}$; while the real consumption surplus is $U\left(\widehat{\beta}\right)-\int_{\beta\delta b-p^{SB}\left(\widehat{\beta}\right)}^{\widehat{\beta}\delta b-p^{SB}\left(\widehat{\beta}\right)}\left(\delta b-p^{SB}\left(\widehat{\beta}\right)-c\right)dF\left(c\right)$, which is decreasing in $\widehat{\beta}$.*

The proof of Corollary 3 is in the Appendix.

For a consumer with naive expectation $\widehat{\beta}$, the wedge between his fictitious and real surplus is $\int_{\beta\delta b-p^{SB}\left(\widehat{\beta}\right)}^{\widehat{\beta}\delta b-p^{SB}\left(\widehat{\beta}\right)}\left(\delta b-p^{SB}\left(\widehat{\beta}\right)-c\right)dF\left(c\right)$, which reflects the different cut-off points of exercising cost in the naive and sophisticated expectations. For a sophisticated type consumer, the fictitious and real surplus coincide because the wedge cancels; and from the principle of screening equilibrium, $U\left(\widehat{\beta}=\beta\right)=0$. So from Corollary 3, the increasing (Res. decreasing) fictitious (Res. real) rent in $\widehat{\beta}$ tell us an interesting finding: a naive participant thought that he extracts positive rent from the firm, but in fact he is exploited (negative real rent) after he learns the true β at date 1. Figure 7 also illustrates this.

Figure 7 is about here

5 Concluding Remarks

We introduce asymmetric information structure in DM's monopolistic first-best investment good pricing framework. Different dimensions of the consumer's private information, e.g., on the short-run patience level (β), on the degree of naivety ($\widehat{\beta}$), and on the intrinsic willingness to pay (b), are considered, respectively.

When the consumer privately holds the future benefit of consuming the investment good (b), the known DM below-marginal-cost per-usage pricing property in the complete information benchmark is not true for the whole population in the second-best. The firm will distort the per-usage price upward due to the incomplete information and thus may exert a higher-than-marginal-cost per-usage price for the relatively inefficient types (low willingness to pay).

More importantly, under the asymmetric information on b, the firm's profit is no longer unaffected by the consumer's short-run impatience even when the consumer is sophisticated, which is different from DM's main finding in the first-best. The upward distortion on the per-usage price obstructs the first-best discount to attract a short-run impatient consumer, so a lower β will involve a negative impact on the firm's second-best profit even if the consumer is sophisticated when $\beta \in \left[\frac{1}{2}, 1\right]$.

But when $\beta < \frac{1}{2}$ and the consumer is sophisticated, i.e., perfectly self-aware of his future short-run impatience to consume a healthy investment good, he can regard the firm/club as an exogenous correction instrument to overcome his own self-control problem thanks to the existence of a pre-committed payment L before consumption and the lower per-usage price during consumption, thus the consumer becomes more urgent to participate the health club. Therefore the firm can ask for a higher entrance fee L to reimburse the per-usage loss. As a result, the firm's second-best expected profit is decreasing in β when $\beta \in \left(0, \frac{1}{2}\right]$ (lower β generates higher profit). All these creates a non-monotonic, strictly convex and symmetric second-best profit locus with respect to β when β varies from 0 to 1.

In addition, the social welfare (the sum of profit and consumer surplus) is invariant in β as long as the consumer is sophisticated. The firm's imprecise knowledge on b only influences the surplus allocation between the firm and the consumer, while not the total volume in the economy.

When the consumer holds private information on the time preference (β) and his belief toward tomorrow's value ($\widehat{\beta}$), the second-best pricing feature depends crucially on the distribution type of the exercising cost (c). When c is uniformly distributed, the firm can only conduct a pooling pricing. When the firm does not know the short-run patience level of a sophisticated consumer, under the second-best pooling pricing, the firm prefers a short-run impatient consumer since higher β realizes more consumption which involves more loss to the firm in the per-usage stage. When the firm does not know the naivety level of a naive consumer, the benefit to the profit due to the consumer's naivety in the first-best disappears under a second-best pooling pricing. The firm can only offer a sophistication price scheme and reserve a sophisticated consumer's profit level for the whole population.

When the distribution of the exercising cost is non-uniform, the separated screening pricing menus exist in the information structure of private β and $\widehat{\beta}$, respectively. The second-best equilibrium involves an over-consumption of the investment good with respect to its first-best (complete information pricing) for all the sophisticated types. Facing the diverse naivety, the sophisticated consumer's participation constraint is binding at 0; while the more naive consumer gets more fictitious rent when contracting, but in fact he is more heavily exploited in the sense of negative real rent.

At last, in the future we incline to conduct some field or laboratory studies to support or test the theoretical findings in this paper, and to compare with the known DM first-best properties and their evidences in the investment good industry as well.

Appendix: Proofs

Proof of Proposition 0.

(i)

The firm's problem is $\max_{L,p}(2)$, s.t. (1). Substituting for L in (2) by binding (1) yields

$$\max_{p}\pi(p) = \delta\left(\int_{-\infty}^{\widehat{\beta}\delta b-p}(\delta b-p-c)\,dF(c)-\overline{u}-K+F(\beta\delta b-p)(p-a)\right)$$

The first order condition with respect to p is:

$$-\left(1-\widehat{\beta}\right)\delta bf\left(\widehat{\beta}\delta b-p\right)-\int_{-\infty}^{\widehat{\beta}\delta b-p}dF(c)-f(\beta\delta b-p)(p-a)+F(\beta\delta b-p)=0$$

By rearranging this, we get Equation (3).

If $\beta=\widehat{\beta}=1$, $p^{FB}_{\beta=\widehat{\beta}=1}=a$.

If $\beta<1$ and the consumer is sophisticated ($\widehat{\beta}=\beta$):

$$p^{FB}_{\beta=\widehat{\beta}<1}-a=-(1-\beta)\,\delta b<0 \tag{28}$$

If $\beta<1$ and the consumer is naive ($\beta<\widehat{\beta}\leq 1$):

$$p^{FB}_{\beta<\widehat{\beta}\leq 1}-a\leq-\frac{F\left(\widehat{\beta}\delta b-p^{FB}_{\beta<\widehat{\beta}\leq 1}\right)-F\left(\beta\delta b-p^{FB}_{\beta<\widehat{\beta}\leq 1}\right)}{f\left(\beta\delta b-p^{FB}_{\beta<\widehat{\beta}\leq 1}\right)}<0$$

So the "below-marginal-cost per-usage pricing" property for $\beta<1$ is shown.

(ii)

For sophisticated quasi-hyperbolic consumer, the real consumption probability is $F\left(\beta\delta b-p^{FB}_{\beta=\widehat{\beta}<1}\right)\overset{\text{by (28)}}{=}F(\beta\delta b-a+(1-\beta)\,\delta b)=F(\delta b-a)$, where the last term is exactly the resulting consumption probability of the time-consistent consumer since $p^{FB}_{\beta=\widehat{\beta}=1}=a$.

When $\widehat{\beta}=\beta$, by imposing the fact that (1) is binding and Equation (28),

$$\begin{aligned}
&\pi\left(p^{FB}_{\beta=\widehat{\beta}<1}\right)\\
=\;&\delta\left(\int_{-\infty}^{\beta\delta b-p^{FB}_{\beta=\widehat{\beta}<1}}\left(\delta b-p^{FB}_{\beta=\widehat{\beta}<1}-c\right)dF\left(c\right)+F\left(\beta\delta b-p^{FB}_{\beta=\widehat{\beta}<1}\right)\left(p^{FB}_{\beta=\widehat{\beta}<1}-a\right)\right)\\
=\;&\delta\left(\left(\delta b-a+\left(1-\beta\right)\delta b\right)\int_{-\infty}^{\delta b-a}dF\left(c\right)-\int_{-\infty}^{\delta b-a}cdF\left(c\right)-F\left(\delta b-a\right)\left(1-\beta\right)\delta b\right)\\
=\;&\delta\left(F\left(\delta b-a\right)\left(\delta b-a\right)-\int_{-\infty}^{\delta b-a}cdF\left(c\right)\right)
\end{aligned}$$

which is independent of β.

(iii)

We use the envelope theorem:

$$\frac{\partial\pi\left(p^{FB}\right)}{\partial\widehat{\beta}}=\left(1-\widehat{\beta}\right)\delta^3b^2f\left(\widehat{\beta}\delta b-p^{FB}\right)\geq 0$$

where the equality holds only if $\widehat{\beta}=1$.

So if given β, the firm's profit strictly increases with the extent of naivety in the entire region where $\widehat{\beta}$ varies from β to 1. ■

Proof of Proposition 1. Define the informational rent $U\left(b\right)=u\left(b,b\right)=-L\left(b\right)+v\left(b,p\left(b\right)\right)$ of type b, so from ICs in (7) and the envelope theorem, $U'\left(b\right)=\frac{\partial v}{\partial b}=\widehat{\beta}\left(1-\widehat{\beta}\right)\delta^2bf\left(\widehat{\beta}\delta b-p\right)+\delta F\left(\widehat{\beta}\delta b-p\right)>0$. Thus $\forall b\in\left(\underline{b},\overline{b}\right]$, $u\left(b,b\right)>u\left(\underline{b},\underline{b}\right)$; so the IRs in (8) can just reduce to the IR of the bottom type $\underline{b}$, i.e., $u\left(\underline{b},\underline{b}\right)\geq 0$. Since leaving rents to the agent is costly for the firm, this bottom IR has to be binding.

And ICs in (7) tell us that $\forall b,b'\in\left[\underline{b},\overline{b}\right]$,

$$-L\left(b'\right)+v\left(b',p\left(b'\right)\right)\geq -L\left(b\right)+v\left(b',p\left(b\right)\right)\tag{29}$$

$$-L\left(b\right)+v\left(b,p\left(b\right)\right)\geq -L\left(b'\right)+v\left(b,p\left(b'\right)\right)\tag{30}$$

Summing (29) and (30), we obtain that:

$$\begin{aligned} & v\left(b', p\left(b'\right)\right) - v\left(b', p\left(b\right)\right) + v\left(b, p\left(b\right)\right) - v\left(b, p\left(b'\right)\right) \\ = & \int_b^{b'} \int_{p(b)}^{p(b')} \frac{\partial^2 v}{\partial \widehat{b} \partial p} \left(\widehat{b}, p\right) dp d\widehat{b} \geq 0 \end{aligned} \tag{31}$$

From (10) under the uniform c, we know that $\frac{\partial^2 v}{\partial \widehat{b} \partial p} < 0$, so the inequality in (31) restricts that $b' \geq b \Leftrightarrow p\left(b'\right) \leq p\left(b\right)$, which gives a monotone constraint on the optimal $p\left(\cdot\right)$: $p'\left(b\right) \leq 0$ for $\forall b \in \left[\underline{b}, \overline{b}\right]$.

Substituting $L\left(b\right) = -U\left(b\right) + v\left(b, p\left(b\right)\right)$, we can transform the original program (6) subject to (7)(8) into the following optimal control problem ($p\left(b\right)$ as control variable, $U\left(b\right)$ as state variable):

$$\max_{U(b),p(b)} \int_{\underline{b}}^{\overline{b}} \delta\left(v\left(b, p\left(b\right)\right) - U\left(b\right) + F\left(\beta\delta b - p\left(b\right)\right)\left(p\left(b\right) - a\right)\right) dH\left(b\right) \tag{32}$$

subject to:

$$U'\left(b\right) = \widehat{\beta}\left(1 - \widehat{\beta}\right)\delta^2 b f\left(\widehat{\beta}\delta b - p\left(b\right)\right) + \delta F\left(\widehat{\beta}\delta b - p\left(b\right)\right) \tag{33}$$

$$p'\left(b\right) \leq 0 \tag{34}$$

$$U\left(\underline{b}\right) = 0 \tag{35}$$

Imposing the uniform distribution of c ($F\left(c\right) = \frac{c - \underline{c}}{\overline{c} - \underline{c}}$ and $f\left(c\right) = \frac{1}{\overline{c} - \underline{c}}$),

$$\begin{aligned} v\left(b, p\left(b\right)\right) &= \int_{\underline{c}}^{\widehat{\beta}\delta b - p(b)} \frac{\delta b - p\left(b\right) - c}{\overline{c} - \underline{c}} dc \\ &= \left(\left(1 - \frac{\widehat{\beta}}{2}\right)\delta b - \frac{1}{2} p\left(b\right) - \frac{\underline{c}}{2}\right) \frac{\widehat{\beta}\delta b - p\left(b\right) - \underline{c}}{\overline{c} - \underline{c}} \end{aligned} \tag{36}$$

Then we can rewrite (32) and (33) as the following (37) and (38), respectively:

$$\max_{U(b),p(b)} \int_{\underline{b}}^{\overline{b}} \delta\left(v\left(b, p\left(b\right)\right) - U\left(b\right) + \frac{\beta\delta b - p\left(b\right) - \underline{c}}{\overline{c} - \underline{c}}\left(p\left(b\right) - a\right)\right) h\left(b\right) db \tag{37}$$

$$U'(b)=\widehat{\beta}\left(1-\widehat{\beta}\right)\delta^2 b\frac{1}{\overline{c}-\underline{c}}+\delta\frac{\widehat{\beta}\delta b-p(b)-\underline{c}}{\overline{c}-\underline{c}} \tag{38}$$

By integration by part:

$$\begin{aligned}
&\int_{\underline{b}}^{\overline{b}} U(b)\,h(b)\,db &(39)\\
=\;& [U(b)(H(b)-1)]_{\underline{b}}^{\overline{b}}+\int_{\underline{b}}^{\overline{b}} U'(b)(1-H(b))\,db\\
=\;& \int_{\underline{b}}^{\overline{b}} U'(b)(1-H(b))\,db
\end{aligned}$$

Insert (39) into (37) and then substitute $U'(b)$ with its expression in (38), we can compact the program (37) and the constraints (38)(35) into:

$$\begin{aligned}
\max_{p(b)} \;:\; & \int_{\underline{b}}^{\overline{b}} \delta\left(\frac{\beta\delta b-p(b)-\underline{c}}{\overline{c}-\underline{c}}(p(b)-a)\right)h(b)\,db &(40)\\
&+\int_{\underline{b}}^{\overline{b}} \delta\left(\left(\left(1-\frac{\widehat{\beta}}{2}\right)\delta b-\frac{1}{2}p(b)-\frac{\underline{c}}{2}\right)\frac{\widehat{\beta}\delta b-p(b)-\underline{c}}{\overline{c}-\underline{c}}\right)h(b)\,db\\
&-\int_{\underline{b}}^{\overline{b}} \delta^2\left(\frac{\widehat{\beta}\delta b-p(b)-\underline{c}}{\overline{c}-\underline{c}}+\frac{\widehat{\beta}\left(1-\widehat{\beta}\right)\delta b}{\overline{c}-\underline{c}}\right)(1-H(b))\,db
\end{aligned}$$

First of all we ignore the monotonicity constraint $p'(b)\le 0$, so that (40) becomes a point-wise optimization problem $\max\limits_{p}\delta\int_{\underline{b}}^{\overline{b}}\pi(p,b)\,db$, where

$$\begin{aligned}
\pi(p,b) \;=\;& h(b)\left(\left(1-\frac{\widehat{\beta}}{2}\right)\delta b-\frac{1}{2}p-\frac{\underline{c}}{2}\right)\frac{\widehat{\beta}\delta b-p-\underline{c}}{\overline{c}-\underline{c}} &(41)\\
&+\frac{\beta\delta b-p-\underline{c}}{\overline{c}-\underline{c}}(p-a)\,h(b)\\
&-\delta\left(\frac{\widehat{\beta}\delta b-p-\underline{c}}{\overline{c}-\underline{c}}+\frac{\widehat{\beta}\left(1-\widehat{\beta}\right)\delta b}{\overline{c}-\underline{c}}\right)(1-H(b))
\end{aligned}$$

Then the first order condition with respect to p of (41) is:

$$\begin{aligned}\frac{\partial \pi}{\partial p} &= \left(-\frac{1}{2}\frac{\widehat{\beta}\delta b - p - \underline{c}}{\overline{c}-\underline{c}} - \frac{\left(\left(1-\frac{\widehat{\beta}}{2}\right)\delta b - \frac{1}{2}p - \frac{\underline{c}}{2}\right)}{\overline{c}-\underline{c}}\right) h(b) \quad (42)\\ &\quad -\left(\frac{p-a}{\overline{c}-\underline{c}} - \frac{\beta\delta b - p - \underline{c}}{\overline{c}-\underline{c}}\right) h(b)\\ &\quad +\frac{\delta}{\overline{c}-\underline{c}}\left(1-H(b)\right)\\ &= 0\end{aligned}$$

By solving (42), we get:

$$p(b) = a - (1-\beta)\delta b + \delta\frac{1-H(b)}{h(b)} \quad (43)$$

Under A1, $\frac{1-H(b)}{h(b)}$ is decreasing in b, so the relaxed solution in (43) satisfies automatically the constraint $p'(b) \leq 0$.

Next we need check that under the boundary conditions for b and c in (5), the predicted (sophisticatedly or naively, at date 0) and real (at date 1) consumption probability from the interior solution in (43) belongs to $(0,1)$ for any type b so that no corner solution happens, i.e., (43) is exactly the final solution of the original program (6) subject to (7)(8). It suffices to show that $\forall b$, $\beta\delta b - p(b) - \underline{c} > 0$ and $\widehat{\beta}\delta b - p(b) - \overline{c} < 1$, where $p(b)$ is defined by (43).

Under (43) and (5),

$$\forall b, \beta\delta b - p(b) = \delta b - a - \delta\frac{1-H(b)}{h(b)} \geq \delta\underline{b} - a - \frac{\delta}{h(\underline{b})} > \underline{c} \quad (44)$$

where the inequality "$\geq$" in (44) comes from the assumption that $\frac{1-H(b)}{h(b)}$ is decreasing in b and the fact that $H(\underline{b}) = 0$.

On the other side,

$$\begin{aligned}\forall b, \widehat{\beta}\delta b - p(b) &\leq \delta b - p(b) = (2-\beta)\delta b - a - \delta\frac{1-H(b)}{h(b)}\\ &\leq (2-\beta)\delta\overline{b} - a < \underline{c} \quad (45)\end{aligned}$$

since $H\left(\overline{b}\right)=1$.

Given (43), $L(b)=-U(b)+v(b,p(b))$. By the envelope theorem in ICs, $L'(b)=\frac{\partial v}{\partial p}\cdot p'(b)$, where $\frac{\partial v}{\partial p}=-\left(1-\widehat{\beta}\right)\delta bf\left(\widehat{\beta}\delta b-p\right)-F\left(\widehat{\beta}\delta b-p\right)<0$, so $L'(b)>0$.

Compared with the results in Lemma 1, we obviously see that $p\left(\overline{b}\right)=p^{FB}\left(\overline{b}\right)$ and $p(b)>p^{FB}(b)$ for $\forall b<\overline{b}$ since the distortion $\delta\frac{1-H(b)}{h(b)}>0$ when $b<\overline{b}$. ■

Proof of Proposition 2. By (6) and substituting $L(b)=-U(b)+v(b,p(b))$, the resulting expected profit of the firm is:

$$\pi^{SB,b}=\int_{\underline{b}}^{\overline{b}}\delta\left(v(b,p(b))-U(b)+F(\beta\delta b-p(b))(p(b)-a)\right)dH(b) \quad (46)$$

where $p(b)$ is the second-best solution in (43).

From (35), we have $U(b)=\int_{\underline{b}}^{b}U'\left(\widehat{b}\right)d\widehat{b}$, where $U'(\cdot)$ is given by (33). After imposing the expression of $U(b)$, of $v(b,p(b))$ (from (36)), and the uniform distribution function of c into (46), $\pi^{SB,b}=\frac{\delta}{\overline{c}-\underline{c}}\int_{\underline{b}}^{\overline{b}}\widetilde{\pi}(b)\,dH(b)$, where

$$\begin{aligned}\widetilde{\pi}(b) &= \left(\left(1-\frac{\widehat{\beta}}{2}\right)\delta b-\frac{1}{2}p(b)-\frac{\underline{c}}{2}\right)\left(\widehat{\beta}\delta b-p(b)-\underline{c}\right)\\ &\quad -\int_{\underline{b}}^{b}\left(\widehat{\beta}\left(1-\widehat{\beta}\right)\delta^2\widehat{b}+\delta\left(\widehat{\beta}\delta\widehat{b}-p\left(\widehat{b}\right)-\underline{c}\right)\right)d\widehat{b}\\ &\quad +(p(b)-a)(\beta\delta b-p(b)-\underline{c}) \quad (47)\end{aligned}$$

Thus the profit for a sophisticated ($\widehat{\beta}=\beta$) b is:

$$\begin{aligned}\widetilde{\pi}_{\widehat{\beta}=\beta}(b) &= \frac{1}{2}(\beta\delta b-p(b)-\underline{c})((2-\beta)\delta b-p(b)-\underline{c}) \quad (48)\\ &\quad -\int_{\underline{b}}^{b}\left(\beta(2-\beta)\delta^2\widehat{b}-\delta p\left(\widehat{b}\right)-\delta\underline{c}\right)d\widehat{b}\\ &\quad +(p(b)-a)(\beta\delta b-p(b)-\underline{c})\\ &= \frac{1}{2}(\beta\delta b-p(b)-\underline{c})(2\delta b-2a-\underline{c}-(\beta\delta b-p(b)))\\ &\quad -\int_{\underline{b}}^{b}\left(\beta(2-\beta)\delta^2\widehat{b}-\delta p\left(\widehat{b}\right)-\delta\underline{c}\right)d\widehat{b}\end{aligned}$$

By imposing (43) for $p(b)$, we obtain:

$$\begin{aligned}\widetilde{\pi}_{\widehat{\beta}=\beta}(b) &= \frac{1}{2}\left(\delta b - a - \delta m(b) - \underline{c}\right)\left(\delta b - a - \underline{c} + \delta m(b)\right) \\ &\quad - \left(1+\beta-\beta^2\right)\delta^2 \int_{\underline{b}}^{b} \widehat{b}d\widehat{b} + \delta\left(a+\underline{c}\right)\left(b-\underline{b}\right) + \delta^2 \int_{\underline{b}}^{b} m\left(\widehat{b}\right)d\widehat{b}\end{aligned}$$

where $m(b) = \frac{1-H(b)}{h(b)}$.

We can simply write $\widetilde{\pi}_{\widehat{\beta}=\beta}(b) = C(b) - (1+\beta-\beta^2)\,\delta^2 \int_{\underline{b}}^{b} \widehat{b}d\widehat{b}$. The function $1+\beta-\beta^2$ is symmetric with respect to $\beta = \frac{1}{2}$ (by the fact that $1+(1-\beta)-(1-\beta)^2 = 1+\beta-\beta^2$), so $\widetilde{\pi}_{\widehat{\beta}=\beta}(b;\beta) = \widetilde{\pi}_{\widehat{\beta}=\beta}(b;1-\beta)$. Therefore,

$$\begin{aligned}\pi^{SB,b}_{\widehat{\beta}=\beta}(\beta) &= \frac{\delta}{\overline{c}-\underline{c}}\int_{\underline{b}}^{\overline{b}} \widetilde{\pi}_{\widehat{\beta}=\beta}(b;\beta)\,dH(b) \\ &= \frac{\delta}{\overline{c}-\underline{c}}\int_{\underline{b}}^{\overline{b}} \widetilde{\pi}_{\widehat{\beta}=\beta}(b;1-\beta)\,dH(b) \\ &= \pi^{SB,b}_{\widehat{\beta}=\beta}(1-\beta)\end{aligned} \tag{49}$$

By taking the derivative of $\pi^{SB,b}_{\widehat{\beta}=\beta}$ w.r.t. β, we get:

$$\begin{aligned}\frac{\partial \pi^{SB,b}_{\widehat{\beta}=\beta}}{\partial\beta} &= \frac{\delta}{\overline{c}-\underline{c}}\int_{\underline{b}}^{\overline{b}} \frac{\partial \widetilde{\pi}_{\widehat{\beta}=\beta}(b;\beta)}{\partial\beta}dH(b) \\ &= \frac{\delta^3(2\beta-1)}{\overline{c}-\underline{c}}\int_{\underline{b}}^{\overline{b}}\int_{\underline{b}}^{b} \widehat{b}d\widehat{b}dH(b) \\ &= \frac{\delta^3(2\beta-1)}{\overline{c}-\underline{c}}\int_{\underline{b}}^{\overline{b}} \frac{1}{2}\left(b^2-\underline{b}^2\right)dH(b)\end{aligned} \tag{50}$$

$b^2-\underline{b}^2 \geq 0$ and the equality holds only if $b=\underline{b}$, so $\int_{\underline{b}}^{\overline{b}} \frac{1}{2}\left(b^2-\underline{b}^2\right)dH(b) > 0$. Thus

$$\frac{\partial^2 \pi^{SB,b}_{\widehat{\beta}=\beta}}{\partial\beta^2} = \frac{2\delta^3}{\overline{c}-\underline{c}}\int_{\underline{b}}^{\overline{b}} \frac{1}{2}\left(b^2-\underline{b}^2\right)dH(b) > 0$$

Therefore $\pi^{SB,b}_{\widehat{\beta}=\beta}$ is globally and strictly convex in β. From (50),

$$\frac{\partial \pi^{SB,b}_{\widehat{\beta}=\beta}}{\partial \beta} \begin{cases} > 0, \text{ if } \beta > \frac{1}{2} \\ = 0, \text{ if } \beta = \frac{1}{2} \\ < 0, \text{ if } \beta < \frac{1}{2} \end{cases} \tag{51}$$

So $\beta = \frac{1}{2}$ reaches the minimum of $\pi^{SB,b}_{\widehat{\beta}=\beta}$ when β varies from 0 to 1. ■

Proof of Corollary 2.

(i)

By (39), $E(U(b)) = \int_{\underline{b}}^{\overline{b}} U(b) h(b) db = \int_{\underline{b}}^{\overline{b}} U'(b)(1 - H(b)) db$. Then impose the expression of $U'(b)$ in (38) and the fact of sophistication ($\widehat{\beta} = \beta$), we obtain:

$$E(U(b)) = \delta \int_{\underline{b}}^{\overline{b}} \left(\frac{\beta\delta b - p(b) - \underline{c}}{\overline{c} - \underline{c}} + \frac{\beta(1-\beta)\delta b}{\overline{c} - \underline{c}} \right) (1 - H(b)) db \tag{52}$$

From Proposition 1, we know that $p(b) = a - (1-\beta)\delta b + \delta \frac{1-H(b)}{h(b)}$. So $\beta\delta b - p(b)$ is independent of β. Thus we can reduce (52) to:

$$E(U(b)) = \frac{\beta(1-\beta)\delta^2}{\overline{c} - \underline{c}} \int_{\underline{b}}^{\overline{b}} b(1 - H(b)) db + X \tag{53}$$

where X is not a function of β. So it's obvious from (53) that $E(U(b))$ is symmetric around $\beta = \frac{1}{2}$.

$$\frac{\partial E(U(b))}{\partial \beta} = \frac{(1-2\beta)\delta^2}{\overline{c} - \underline{c}} \int_{\underline{b}}^{\overline{b}} b(1 - H(b)) db \tag{54}$$

Since $1 - H(b) \geq 0$ for all $b \in [\underline{b}, \overline{b}]$, where the equality holds only if $b = \overline{b}$, we have $\int_{\underline{b}}^{\overline{b}} b(1 - H(b)) db > 0$. Hence,

$$\frac{\partial E(U(b))}{\partial \beta} \begin{cases} < 0, \text{ if } \beta > \frac{1}{2} \\ = 0, \text{ if } \beta = \frac{1}{2} \\ > 0, \text{ if } \beta < \frac{1}{2} \end{cases} \tag{55}$$

(55) is the non-monotonicity of $E(U(b))$ in β.

$$\frac{\partial^2 E(U(b))}{\partial \beta^2} = \frac{-2\delta^2}{\overline{c}-\underline{c}}\int_{\underline{b}}^{\overline{b}} b(1-H(b))\,db < 0 \tag{56}$$

(56) is the strict concavity of $E(U(b))$. So from (55), $\beta = \frac{1}{2}$ is the unique maximum.

(ii)

$SW = \pi^{SB,b}_{\widehat{\beta}=\beta} + \int_{\underline{b}}^{\overline{b}} \delta U(b)\,dH(b)$. Then replace $\pi^{SB,b}_{\widehat{\beta}=\beta}$ by the expressions in (36) and (37), and impose $\widehat{\beta} = \beta$, we obtain:

$$SW = \frac{\delta}{\overline{c}-\underline{c}}\int_{\underline{b}}^{\overline{b}} (\beta\delta b - p(b) - \underline{c})\left(\left(1-\frac{\beta}{2}\right)\delta b + \frac{1}{2}p(b) - \frac{\underline{c}}{2} - a\right) dH(b) \tag{57}$$

Since we know that $\beta\delta b - p(b)$ is independent of β, so (57) is invariant in β. ■

Proof of Proposition 5.

(i)

Define the informational rent $U(\beta) = -L(\beta) + v(\beta, p(\beta))$ of type β, so from ICs in (22) and the envelope theorem,

$$U'(\beta) = \frac{\partial v}{\partial \beta} = (1-\beta)(\delta b)^2 f(\beta\delta b - p(\beta)) > 0$$

Thus $\forall \beta \in \left(\underline{\beta}, \overline{\beta}\right]$, $U(\beta) > U\left(\underline{\beta}\right)$; so the IRs in (23) can just reduce to the IR of the bottom type $\underline{\beta}$, i.e., $U\left(\underline{\beta}\right) \geq 0$, which has to be binding: $U\left(\underline{\beta}\right) = 0$.

And ICs in (22) tell us that $\forall \beta, \beta' \in \left[\underline{\beta}, \overline{\beta}\right]$,

$$-L(\beta') + v(\beta', p(\beta')) \geq -L(\beta) + v(\beta', p(\beta)) \tag{58}$$

$$-L(\beta) + v(\beta, p(\beta)) \geq -L(\beta') + v(\beta, p(\beta')) \tag{59}$$

Summing (58) and (59), we obtain that:

$$\begin{aligned} & v\left(\beta', p\left(\beta'\right)\right)-v\left(\beta', p\left(\beta\right)\right)+v\left(\beta, p\left(\beta\right)\right)-v\left(\beta, p\left(\beta'\right)\right) \\ = & \int_{\beta}^{\beta'} \int_{p(\beta)}^{p(\beta')} \frac{\partial^2 v}{\partial \widetilde{\beta} \partial p}\left(\widetilde{\beta}, p\right) d p d \widetilde{\beta} \geq 0 \end{aligned} \quad (60)$$

$\frac{\partial^2 v}{\partial \widetilde{\beta} \partial p}=-\left(1-\widetilde{\beta}\right)(\delta b)^2 f'\left(\widetilde{\beta} \delta b-p\right)>0$ under A2. So the inequality in (60) restricts that $\beta' \geq \beta \Leftrightarrow p\left(\beta'\right) \geq p(\beta)$, which gives a monotone constraint on the optimal $p(\cdot)$: $p'(\beta) \geq 0$ for $\forall \beta \in\left[\underline{\beta}, \overline{\beta}\right]$.

Substituting $L(\beta)=-U(\beta)+v(\beta, p(\beta))$, we can transform the original program (21) subject to (22)(23) into the following optimal control problem ($p(\beta)$ as control variable, $U(\beta)$ as state variable):

$$\max_{U(\beta), p(\beta)} \int_{\underline{\beta}}^{\overline{\beta}} \delta(v(\beta, p(\beta))-U(\beta)+F(\beta \delta b-p(\beta))(p(\beta)-a)) d G(\beta) \quad (61)$$

subject to:

$$U'(\beta)=(1-\beta)(\delta b)^2 f(\beta \delta b-p(\beta)) \quad (62)$$

$$p'(\beta) \geq 0 \quad (63)$$

$$U\left(\underline{\beta}\right)=0 \quad (64)$$

By integration by part:

$$\begin{aligned} & \int_{\underline{\beta}}^{\overline{\beta}} U(\beta) d G(\beta) \quad (65) \\ = & [U(\beta)(G(\beta)-1)]_{\underline{\beta}}^{\overline{\beta}}+\int_{\underline{\beta}}^{\overline{\beta}} U'(\beta)(1-G(\beta)) d \beta \\ = & \int_{\underline{\beta}}^{\overline{\beta}} U'(\beta)(1-G(\beta)) d \beta \end{aligned}$$

We insert (65) and the expression of v into (61), then we can compact the program (61) and the constraints (62)(64) into:

$$\max_{p(\beta)} \delta \int_{\underline{\beta}}^{\overline{\beta}}\left(g(\beta) \kappa(\beta, p(\beta))-U'(\beta)(1-G(\beta))\right) d \beta \quad (66)$$

where $\kappa\left(\beta, p\left(\beta\right)\right)=\int_{\underline{c}}^{\beta\delta b-p(\beta)}\left(\delta b-p\left(\beta\right)-c\right)dF\left(c\right)+F\left(\beta\delta b-p\left(\beta\right)\right)\left(p\left(\beta\right)-a\right)$.

First we ignore the monotonicity constraint $p'\left(\beta\right)\geq 0$, so that (66) becomes a point-wise optimization problem $\max_{p}\delta\int_{\underline{\beta}}^{\overline{\beta}}\pi\left(p,\beta\right)d\beta$, where $\pi\left(p,\beta\right)=g\left(\beta\right)\kappa\left(\beta,p\right)-U'\left(\beta\right)\left(1-G\left(\beta\right)\right)$.

Then the first order condition with respect to p is:

$$\begin{aligned}\frac{\partial\pi}{\partial p} &= (1-\beta)(\delta b)^{2}f'(\beta\delta b-p)(1-G(\beta)) \\ &\quad -((p-a+(1-\beta))\delta b f(\beta\delta b-p))g(\beta) \\ &= 0\end{aligned} \tag{67}$$

By solving (67), we obtain:

$$\begin{aligned}p &= a-(1-\beta)\delta b+(1-\beta)(\delta b)^{2}\frac{1-G(\beta)}{g(\beta)}\cdot\frac{f'(\beta\delta b-p)}{f(\beta\delta b-p)} \\ &= p^{FB}_{\beta=\widehat{\beta}}+(1-\beta)(\delta b)^{2}\frac{1-G(\beta)}{g(\beta)}\cdot\frac{f'(\beta\delta b-p)}{f(\beta\delta b-p)}\end{aligned} \tag{68}$$

Now we consider the monotonicity constraint $p'\left(\beta\right)\geq 0$. $\frac{\partial p^{FB}_{\beta=\widehat{\beta}}}{\partial\beta}=\delta b>0$, but we do not know the sign of $\frac{\partial}{\partial\beta}\left((1-\beta)\frac{1-G(\beta)}{g(\beta)}\cdot\frac{f'(\beta\delta b-p)}{f(\beta\delta b-p)}\right)$. The most important thing is that we check whether the relaxed control solution in (68) satisfies $p\left(\overline{\beta}\right)>p\left(\underline{\beta}\right)$; if so, we can always realize a non-decreasing control trajectory starting from $p\left(\underline{\beta}\right)$ and ending at $p\left(\overline{\beta}\right)$ even if p in (68) decreases in β in some sub-intervals inside $\left(\underline{\beta},\overline{\beta}\right)$[12]. By (68), $p\left(\overline{\beta}\right)=a-\left(1-\overline{\beta}\right)\delta b>a-\left(1-\underline{\beta}\right)\delta b=p^{FB}_{\beta=\widehat{\beta}}\left(\underline{\beta}\right)>p\left(\underline{\beta}\right)$ since $G\left(\overline{\beta}\right)=1$ and $f'\left(\cdot\right)<0$. So there

[12]In fact we can circumvent "bunching" by appending a parametric specification on the distribution of c and the monotone hazard rate property on β. For example, c is exponentially distributed on $[0,+\infty)$, i.e., $f\left(c\right)=\lambda e^{-\lambda c}$, so $\frac{f'(\cdot)}{f(\cdot)}$ is constant globally; plus the fact that $\frac{1-G(\beta)}{g(\beta)}$ is decreasing in β and $f'\left(\cdot\right)<0$: $\frac{\partial}{\partial\beta}\left((1-\beta)\frac{1-G(\beta)}{g(\beta)}\cdot\frac{f'(\beta\delta b-p)}{f(\beta\delta b-p)}\right)>0$ so that the relaxed p in (68) is increasing in β (the monotone constraint is satisfied automatically).

exists a non-decreasing optimal control on $\left[\underline{\beta},\overline{\beta}\right]$, $p^{SB}\left(\beta\right)$, satisfying

$$\begin{aligned} p^{SB}\left(\overline{\beta}\right) &= a-\left(1-\overline{\beta}\right)\delta b = p^{FB}_{\beta=\widehat{\beta}}\left(\overline{\beta}\right); \qquad (69)\\ p^{SB}\left(\underline{\beta}\right) &= a-\left(1-\underline{\beta}\right)\delta b+\left(1-\underline{\beta}\right)\left(\delta b\right)^2\frac{1}{g\left(\underline{\beta}\right)}\cdot\frac{f'\left(\underline{\beta}\delta b-p^{SB}\left(\underline{\beta}\right)\right)}{f\left(\underline{\beta}\delta b-p^{SB}\left(\underline{\beta}\right)\right)};\\ \forall\beta &\in \left[\underline{\beta},\overline{\beta}\right], \frac{\partial p^{SB}\left(\beta\right)}{\partial\beta}\geq 0. \end{aligned}$$

An illustration of $p^{SB}\left(\beta\right)$ with compared to the relaxed solution in (68) and $p^{FB}_{\beta=\widehat{\beta}}$ are shown in Figure 5.

We know that $L\left(\beta\right)=-U\left(\beta\right)+v\left(\beta,p\left(\beta\right)\right)$, so by envelope theorem,

$$\begin{aligned} \frac{\partial L^{SB}\left(\beta\right)}{\partial\beta} &= \frac{\partial v}{\partial p}\cdot\frac{\partial p^{SB}\left(\beta\right)}{\partial\beta} \qquad (70)\\ &= -\left(\left(1-\beta\right)\delta bf\left(\beta\delta b-p^{SB}\left(\beta\right)\right)+F\left(\beta\delta b-p^{SB}\left(\beta\right)\right)\right)\frac{\partial p^{SB}\left(\beta\right)}{\partial\beta}\\ &\leq 0 \end{aligned}$$

(ii)

It's already shown in (69).

(iii)

In the relaxed solution from (68), the distortion part $\left(1-\beta\right)\left(\delta b\right)^2\frac{1-G(\beta)}{g(\beta)}\cdot\frac{f'(\beta\delta b-p)}{f(\beta\delta b-p)}<0$. So for those β whose $p^{SB}\left(\beta\right)$ coincides with the relaxed solution $p\left(\beta\right)$ in (68), we have $p^{SB}\left(\beta\right)\leq p^{FB}_{\beta=\widehat{\beta}}\left(\beta\right)$, where the equality holds only if $\beta=\overline{\beta}$.

For those β in the bunching interval, we use the "reduction to absurdity". Suppose $\exists\check{\beta}\in$ a bunching interval $\left(\beta_1,\beta_2\right)$, where $p^{SB}\left(\check{\beta}\right)\geq p^{FB}_{\beta=\widehat{\beta}}\left(\check{\beta}\right)$. Because of bunching, $p^{SB}\left(\beta_1\right)=p^{SB}\left(\check{\beta}\right)$. And $\check{\beta}>\beta_1\Rightarrow p^{FB}_{\beta=\widehat{\beta}}\left(\check{\beta}\right)>p^{FB}_{\beta=\widehat{\beta}}\left(\beta_1\right)$ since $p^{FB}_{\beta=\widehat{\beta}}\left(\beta\right)$ is strictly increasing in β. So $p^{SB}\left(\beta_1\right)>p^{FB}_{\beta=\widehat{\beta}}\left(\beta_1\right)$. While β_1 is the starting point of this bunching interval, that is, $p^{SB}\left(\beta_1\right)$ coincides with the relaxed $p\left(\beta_1\right)$ of (68). According to the preceding paragraph,

$p^{SB}(\beta_1) \leq p^{FB}_{\beta=\widehat{\beta}}(\beta_1)$, where a contradiction appears. So for those β in the bunching interval, we also have $p^{SB}(\beta) < p^{FB}_{\beta=\widehat{\beta}}(\beta)$.

At last, since we do not impose corner solution/boundary constraints in Program (21) subject to (22)(23), we need to check that under the boundary condition $\underline{c} < \delta b - a$ and the solution in (69), the resulting consumption probability of any type (β) strictly belongs to $(0,1)$, i.e., interior solution. The maximum cost encountered by a type β to exercise is $\beta\delta b - p^{SB}(\beta) \geq \beta\delta b - p^{FB}_{\beta=\widehat{\beta}}(\beta)$, where the equality holds only if $\beta = \overline{\beta}$. While $\beta\delta b - p^{FB}_{\beta=\widehat{\beta}}(\beta) = \delta b - a > \underline{c}$, so $\forall \beta \in \left[\underline{\beta}, \overline{\beta}\right]$, $\beta\delta b - p^{SB}(\beta) > \underline{c}$, the consumption probability is strictly positive. And since the upper bound of c is $+\infty$, so the consumption probability is always lower than 1. Till here the expressions in (69) are indeed the optimal second-best pricing to Program (21) subject to (22)(23) (no corner solution conflict) and all the associated shown properties work. ■

Proof of Proposition 6.

(i)

Define the informational rent $U\left(\widehat{\beta}\right) = -L\left(\widehat{\beta}\right) + v\left(\widehat{\beta}, p\left(\widehat{\beta}\right)\right)$ of type $\widehat{\beta}$, so from ICs in (26) and the envelope theorem,

$$U'\left(\widehat{\beta}\right) = \frac{\partial v}{\partial \widehat{\beta}} = \left(1-\widehat{\beta}\right)(\delta b)^2 f\left(\widehat{\beta}\delta b - p\left(\widehat{\beta}\right)\right) \geq 0 \tag{71}$$

where the equality holds only if $\widehat{\beta} = 1$. So $\forall \widehat{\beta} \in (\beta, 1]$, $U\left(\widehat{\beta}\right) > U(\beta)$; so the IRs in (27) can just reduce to the IR of the bottom type β, i.e., $U(\beta) \geq 0$, which has to be binding: $U(\beta) = 0$.

And ICs in (26) tell us that $\forall \widehat{\beta}, \widehat{\beta}' \in [\beta, 1]$,

$$-L\left(\widehat{\beta}'\right) + v\left(\widehat{\beta}', p\left(\widehat{\beta}'\right)\right) \geq -L\left(\widehat{\beta}\right) + v\left(\widehat{\beta}', p\left(\widehat{\beta}\right)\right) \tag{72}$$

$$-L\left(\widehat{\beta}\right) + v\left(\widehat{\beta}, p\left(\widehat{\beta}\right)\right) \geq -L\left(\widehat{\beta}'\right) + v\left(\widehat{\beta}, p\left(\widehat{\beta}'\right)\right) \tag{73}$$

Summing (72) and (73), we obtain that:

$$\begin{aligned} & v\left(\widehat{\beta}',p\left(\widehat{\beta}'\right)\right)-v\left(\widehat{\beta}',p\left(\widehat{\beta}\right)\right)+v\left(\widehat{\beta},p\left(\widehat{\beta}\right)\right)-v\left(\widehat{\beta},p\left(\widehat{\beta}'\right)\right) \\ = & \int_{\widehat{\beta}}^{\widehat{\beta}'}\int_{p\left(\widehat{\beta}\right)}^{p\left(\widehat{\beta}'\right)}\frac{\partial^2 v}{\partial\widetilde{\widehat{\beta}}\partial p}\left(\widetilde{\widehat{\beta}},p\right)dpd\widetilde{\widehat{\beta}}\geq 0 \end{aligned} \quad (74)$$

$\frac{\partial^2 v}{\partial\widetilde{\widehat{\beta}}\partial p}=-\left(1-\widetilde{\widehat{\beta}}\right)(\delta b)^2 f'\left(\widetilde{\widehat{\beta}}\delta b-p\right)>0$ under A2. So the inequality in (74) restricts that $\widehat{\beta}'\geq\widehat{\beta}\Leftrightarrow p\left(\widehat{\beta}'\right)\geq p\left(\widehat{\beta}\right)$, which gives a monotone constraint on the optimal $p(\cdot)$: $p'\left(\widehat{\beta}\right)\geq 0$ for $\forall\widehat{\beta}\in[\beta,1]$.

Substituting $L\left(\widehat{\beta}\right)=-U\left(\widehat{\beta}\right)+v\left(\widehat{\beta},p\left(\widehat{\beta}\right)\right)$, we can transform the original program (25) subject to (26)(27) into the following optimal control problem ($p\left(\widehat{\beta}\right)$ as control variable, $U\left(\widehat{\beta}\right)$ as state variable):

$$\max_{U\left(\widehat{\beta}\right),p\left(\widehat{\beta}\right)}\int_{\beta}^{1}\delta\left(v\left(\widehat{\beta},p\left(\widehat{\beta}\right)\right)-U\left(\widehat{\beta}\right)+F\left(\beta\delta b-p\left(\widehat{\beta}\right)\right)\left(p\left(\widehat{\beta}\right)-a\right)\right)dJ\left(\widehat{\beta}\right) \quad (75)$$

subject to:

$$U'\left(\widehat{\beta}\right)=\left(1-\widehat{\beta}\right)(\delta b)^2 f\left(\widehat{\beta}\delta b-p\left(\widehat{\beta}\right)\right) \quad (76)$$

$$p'\left(\widehat{\beta}\right)\geq 0$$

$$U(\beta)=0 \quad (77)$$

By integration by part:

$$\begin{aligned} & \int_{\beta}^{1}U\left(\widehat{\beta}\right)dJ\left(\widehat{\beta}\right) \\ = & \left[U\left(\widehat{\beta}\right)\left(J\left(\widehat{\beta}\right)-1\right)\right]_{\beta}^{1}+\int_{\beta}^{1}U'\left(\widehat{\beta}\right)\left(1-J\left(\widehat{\beta}\right)\right)d\widehat{\beta} \\ = & \int_{\beta}^{1}U'\left(\widehat{\beta}\right)\left(1-J\left(\widehat{\beta}\right)\right)d\widehat{\beta} \end{aligned} \quad (78)$$

We insert (78) and the expression of v into (75), then we can compact the program (75) and the constraints (76)(77) into:

$$\max_{p\left(\widehat{\beta}\right)}\delta\int_{\beta}^{1}\left(j\left(\widehat{\beta}\right)\kappa\left(\beta,\widehat{\beta},p\left(\widehat{\beta}\right)\right)-U'\left(\widehat{\beta}\right)\left(1-J\left(\widehat{\beta}\right)\right)\right)d\widehat{\beta} \quad (79)$$

where

$$\kappa\left(\beta,\widehat{\beta},p\left(\widehat{\beta}\right)\right)=\int_{\underline{c}}^{\widehat{\beta}\delta b-p\left(\widehat{\beta}\right)}\left(\delta b-p\left(\widehat{\beta}\right)-c\right)dF\left(c\right)+F\left(\beta\delta b-p\left(\widehat{\beta}\right)\right)\left(p\left(\widehat{\beta}\right)-a\right)$$

First we ignore the monotonicity constraint $p'\left(\widehat{\beta}\right)\geq 0$, so that (79) becomes a point-wise optimization problem:

$$\max_{p} j\left(\widehat{\beta}\right)\kappa\left(\beta,\widehat{\beta},p\left(\widehat{\beta}\right)\right)-U'\left(\widehat{\beta}\right)\left(1-J\left(\widehat{\beta}\right)\right) \tag{80}$$

Then the solution of (80) is:

$$\begin{aligned} p &= a-\left(1-\widehat{\beta}\right)\delta b\frac{f\left(\widehat{\beta}\delta b-p\right)}{f\left(\beta\delta b-p\right)} \\ &\quad -\frac{F\left(\widehat{\beta}\delta b-p\right)-F\left(\beta\delta b-p\right)}{f\left(\beta\delta b-p\right)} \\ &\quad +\left(1-\widehat{\beta}\right)\left(\delta b\right)^{2}\frac{1-J\left(\widehat{\beta}\right)}{j\left(\widehat{\beta}\right)}\cdot\frac{f'\left(\widehat{\beta}\delta b-p\right)}{f\left(\beta\delta b-p\right)} \end{aligned} \tag{81}$$

Now the most important thing for the existence of a separated screening is that under the relaxed solution in (81), $p\left(\widehat{\beta}=1\right)>p\left(\widehat{\beta}=\beta\right)$, i.e., the starting and ending point of the relaxed trajectory satisfies the monotonicity constraint $p'\left(\widehat{\beta}\right)\geq 0$, then even if $p\left(\widehat{\beta}\right)$ in (81) is not non-decreasing in $\widehat{\beta}$, we can do "bunching" in between; otherwise ($p\left(\widehat{\beta}=1\right)\leq p\left(\widehat{\beta}=\beta\right)$) from the Pontryagin principle, we cannot have a non-constant control $p\left(\widehat{\beta}\right)$ under $p'\left(\widehat{\beta}\right)\geq 0$, then we go back to the result of pooling in Section 3.3. By (81), the condition $p\left(\widehat{\beta}=1\right)>p\left(\widehat{\beta}=\beta\right)$ is:

$$\begin{aligned} &\frac{f'\left(\beta\delta b-p\left(\widehat{\beta}=\beta\right)\right)}{f\left(\beta\delta b-p\left(\widehat{\beta}=\beta\right)\right)}\cdot\frac{\delta b}{j\left(\beta\right)} \\ &< 1-\frac{F\left(\delta b-p\left(\widehat{\beta}=1\right)\right)-F\left(\beta\delta b-p\left(\widehat{\beta}=1\right)\right)}{\left(1-\beta\right)\delta bf\left(\beta\delta b-p\left(\widehat{\beta}=1\right)\right)} \end{aligned} \tag{82}$$

We can find a sufficient condition for the validity of (82) is that b is large enough[13]; more rigorously: $\exists \widehat{b} > 0$, s.t., $\forall b \geq \widehat{b}$, (82) is satisfied.

We rearrange (82) as:

$$\frac{f'\left(\beta\delta b - p\left(\widehat{\beta} = \beta\right)\right)}{f\left(\beta\delta b - p\left(\widehat{\beta} = \beta\right)\right)} \cdot \frac{\delta b}{j(\beta)} - \frac{1 - F\left(\delta b - p\left(\widehat{\beta} = 1\right)\right)}{(1-\beta)\,\delta b f\left(\beta\delta b - p\left(\widehat{\beta} = 1\right)\right)} \tag{83}$$
$$< 1 - \frac{1 - F\left(\beta\delta b - p\left(\widehat{\beta} = 1\right)\right)}{(1-\beta)\,\delta b f\left(\beta\delta b - p\left(\widehat{\beta} = 1\right)\right)}$$

LHS of (83) is negative, so it's sufficient to make (82) hold by showing $\exists \widehat{b} > 0$, s.t., $\forall b \geq \widehat{b}$, $\frac{1-F\left(\beta\delta b - p\left(\widehat{\beta}=1\right)\right)}{(1-\beta)\delta b f\left(\beta\delta b - p\left(\widehat{\beta}=1\right)\right)} < 1$. By A2, $\frac{1-F\left(\beta\delta b - p\left(\widehat{\beta}=1\right)\right)}{f\left(\beta\delta b - p\left(\widehat{\beta}=1\right)\right)}$ is non-increasing in $[\underline{c}, +\infty)$, so the maximum hazard rate of c is $\frac{1}{f(\underline{c})}$, i.e., $\frac{1-F\left(\beta\delta b - p\left(\widehat{\beta}=1\right)\right)}{f\left(\beta\delta b - p\left(\widehat{\beta}=1\right)\right)}$ is bounded above and always positive. So

$$\lim_{b\to+\infty} \frac{1}{(1-\beta)\,\delta b} = 0^+ \Rightarrow \lim_{b\to+\infty} \frac{1 - F\left(\beta\delta b - p\left(\widehat{\beta} = 1\right)\right)}{(1-\beta)\,\delta b f\left(\beta\delta b - p\left(\widehat{\beta} = 1\right)\right)} = 0^+$$

Therefore, $\exists \widehat{b} > 0$, s.t., $\forall b \geq \widehat{b}$, $\frac{1-F\left(\beta\delta b - p\left(\widehat{\beta}=1\right)\right)}{(1-\beta)\delta b f\left(\beta\delta b - p\left(\widehat{\beta}=1\right)\right)} < 1$, i.e., (82) holds and it's possible to conduct a separated screening menu. Thanks to "bunching" technique if necessary, we can obtain an optimal control in the interval $[\beta, 1]$,

[13]if we impose a parametric application of A2, e.g., to assume that c is exponentially distributed on $[0, +\infty)$, i.e., $f(c) = \lambda e^{-\lambda c}$, then (82) is satisfied for $\forall b > 0$. LHS of (82) is negative, so if $\frac{F\left(\delta b - p\left(\widehat{\beta}=1\right)\right) - F\left(\beta\delta b - p\left(\widehat{\beta}=1\right)\right)}{(1-\beta)\delta b f\left(\beta\delta b - p\left(\widehat{\beta}=1\right)\right)} < 1$, then (82) holds. With the functional form $f(c) = \lambda e^{-\lambda c}$, $\frac{F\left(\delta b - p\left(\widehat{\beta}=1\right)\right) - F\left(\beta\delta b - p\left(\widehat{\beta}=1\right)\right)}{(1-\beta)\delta b f\left(\beta\delta b - p\left(\widehat{\beta}=1\right)\right)} = \frac{1-e^{-\lambda(1-\beta)\delta b}}{\lambda(1-\beta)\delta b} < 1$ since $e^{-x} > 1 - x$ for $\forall x > 0$. In addition, if with the exponential c, the relaxed solution in (81) is automatically increasing in $\widehat{\beta}$, i.e., no need of "bunching".

$p^{SB}\left(\widehat{\beta}\right)$, of the program (75) as:

$$\begin{aligned}
p^{SB}\left(\widehat{\beta}=1\right) &= a-\frac{F\left(\delta b-p^{SB}\left(\widehat{\beta}=1\right)\right)-F\left(\beta\delta b-p^{SB}\left(\widehat{\beta}=1\right)\right)}{f\left(\beta\delta b-p^{SB}\left(\widehat{\beta}=1\right)\right)}; \qquad (84)\\
p^{SB}\left(\widehat{\beta}=\beta\right) &= a-(1-\beta)\,\delta b+(1-\beta)\,(\delta b)^2\,\frac{1}{j\,(\beta)}\cdot\frac{f'\left(\beta\delta b-p^{SB}\left(\widehat{\beta}=\beta\right)\right)}{f\left(\beta\delta b-p^{SB}\left(\widehat{\beta}=\beta\right)\right)};\\
\forall\widehat{\beta} &\in [\beta,1]\,,\frac{\partial p^{SB}\left(\widehat{\beta}\right)}{\partial\widehat{\beta}}\geq 0.
\end{aligned}$$

An illustration of $p^{SB}\left(\widehat{\beta}\right)$ with compared to the relaxed solution in (81) and $p^{FB}_{\widehat{\beta}>\beta}$ are shown in Figure 6.

We know that $L\left(\widehat{\beta}\right)=-U\left(\widehat{\beta}\right)+v\left(\widehat{\beta},p\left(\widehat{\beta}\right)\right)$, so by envelope theorem,

$$\begin{aligned}
\frac{\partial L^{SB}\left(\widehat{\beta}\right)}{\partial\widehat{\beta}} &= \frac{\partial v}{\partial p}\cdot\frac{\partial p^{SB}\left(\widehat{\beta}\right)}{\partial\widehat{\beta}} \qquad (85)\\
&= -\left(\left(1-\widehat{\beta}\right)\delta b f\left(\widehat{\beta}\delta b-p\right)+F\left(\widehat{\beta}\delta b-p\right)\right)\frac{\partial p^{SB}\left(\widehat{\beta}\right)}{\partial\widehat{\beta}}\\
&\leq 0
\end{aligned}$$

(ii)

By (3) and imposing $\widehat{\beta}=1$, we have

$$p^{FB}_{\widehat{\beta}=1>\beta}=a-\frac{F\left(\delta b-p^{FB}_{\widehat{\beta}=1>\beta}\right)-F\left(\beta\delta b-p^{FB}_{\widehat{\beta}=1>\beta}\right)}{f\left(\beta\delta b-p^{FB}_{\widehat{\beta}=1>\beta}\right)}$$

which is the same equation as in (84) for $\widehat{\beta}=1$, so $p^{SB}\left(\widehat{\beta}=1\right)=p^{FB}_{\widehat{\beta}=1>\beta}$.

From (28) and $\widehat{\beta}=\beta$ in (84), we can obtain

$$\begin{aligned}
& p^{SB}\left(\widehat{\beta}=\beta\right)-p^{FB}_{\widehat{\beta}=\beta<1}\\
&= (1-\beta)\,(\delta b)^2\,\frac{1}{j\,(\beta)}\cdot\frac{f'\left(\beta\delta b-p^{SB}\left(\widehat{\beta}=\beta\right)\right)}{f\left(\beta\delta b-p^{SB}\left(\widehat{\beta}=\beta\right)\right)}\\
&< 0
\end{aligned}$$

since $f'(\cdot) < 0$. ■

Proof of Corollary 3. The informational rent of a $\widehat{\beta}$-type consumer is $U\left(\widehat{\beta}\right) = -L^{SB}\left(\widehat{\beta}\right) + v\left(\widehat{\beta}, p^{SB}\left(\widehat{\beta}\right)\right)$. From part (i) of the proof of Proposition 6, $U'\left(\widehat{\beta}\right) \geq 0$, where the equality holds only if $\widehat{\beta} = 1$.

The real rent/consumption surplus of the investment good for a $\widehat{\beta}$-type consumer is $\widetilde{U}\left(\widehat{\beta}\right) = -L^{SB}\left(\widehat{\beta}\right) + v\left(\beta, p^{SB}\left(\widehat{\beta}\right)\right)$. Impose the definition of v at Page 17, the wedge between fictitious and real rent is:

$$\widetilde{U}\left(\widehat{\beta}\right) = U\left(\widehat{\beta}\right) - \int_{\beta\delta b - p^{SB}\left(\widehat{\beta}\right)}^{\widehat{\beta}\delta b - p^{SB}\left(\widehat{\beta}\right)} \left(\delta b - p^{SB}\left(\widehat{\beta}\right) - c\right) dF(c)$$

and

$$\begin{aligned}
\widetilde{U}'\left(\widehat{\beta}\right) &= U'\left(\widehat{\beta}\right) - \left(\delta b - p^{SB\prime}\left(\widehat{\beta}\right)\right)\left(1-\widehat{\beta}\right)\delta b f\left(\widehat{\beta}\delta b - p^{SB}\left(\widehat{\beta}\right)\right) \qquad (86)\\
&\quad - p^{SB\prime}\left(\widehat{\beta}\right)\left((1-\beta)\,\delta b f\left(\beta\delta b - p^{SB}\left(\widehat{\beta}\right)\right) - \int_{\beta\delta b - p^{SB}\left(\widehat{\beta}\right)}^{\widehat{\beta}\delta b - p^{SB}\left(\widehat{\beta}\right)} f(c)\, dc\right)
\end{aligned}$$

By imposing (71), (86) can be rearranged as:

$$\begin{aligned}
\widetilde{U}'\left(\widehat{\beta}\right) &= p^{SB\prime}\left(\widehat{\beta}\right)\delta b\left(1-\widehat{\beta}\right) f\left(\widehat{\beta}\delta b - p^{SB}\left(\widehat{\beta}\right)\right) \qquad (87)\\
&\quad - p^{SB\prime}\left(\widehat{\beta}\right)\delta b\,(1-\beta)\, f\left(\beta\delta b - p^{SB}\left(\widehat{\beta}\right)\right)\\
&\quad + p^{SB\prime}\left(\widehat{\beta}\right)\int_{\beta\delta b - p^{SB}\left(\widehat{\beta}\right)}^{\widehat{\beta}\delta b - p^{SB}\left(\widehat{\beta}\right)} f(c)\, dc\\
&= p^{SB\prime}\left(\widehat{\beta}\right)\delta b\left(1-\widehat{\beta}\right)\left(f\left(\widehat{\beta}\delta b - p^{SB}\left(\widehat{\beta}\right)\right) - f\left(\beta\delta b - p^{SB}\left(\widehat{\beta}\right)\right)\right)\\
&\quad + p^{SB\prime}\left(\widehat{\beta}\right)\left(\int_{\beta\delta b - p^{SB}\left(\widehat{\beta}\right)}^{\widehat{\beta}\delta b - p^{SB}\left(\widehat{\beta}\right)} f(c)\, dc - \left(\widehat{\beta} - \beta\right)\delta b f\left(\beta\delta b - p^{SB}\left(\widehat{\beta}\right)\right)\right)\\
&= p^{SB\prime}\left(\widehat{\beta}\right)\left(\delta b\left(1-\widehat{\beta}\right)\cdot Y + Z\right)
\end{aligned}$$

where $Y = f\left(\widehat{\beta}\delta b - p^{SB}\left(\widehat{\beta}\right)\right) - f\left(\beta\delta b - p^{SB}\left(\widehat{\beta}\right)\right) < 0$ for $\widehat{\beta} > \beta$ since

$f'(\cdot) < 0$ under A2; and,

$$\begin{aligned} Z &= \int_{\beta\delta b - p^{SB}\left(\widehat{\beta}\right)}^{\widehat{\beta}\delta b - p^{SB}\left(\widehat{\beta}\right)} f(c)\,dc - \left(\widehat{\beta} - \beta\right)\delta b f\left(\beta\delta b - p^{SB}\left(\widehat{\beta}\right)\right) \\ &= \int_{\beta\delta b - p^{SB}\left(\widehat{\beta}\right)}^{\widehat{\beta}\delta b - p^{SB}\left(\widehat{\beta}\right)} \left(f(c) - f\left(\beta\delta b - p^{SB}\left(\widehat{\beta}\right)\right)\right) dc < 0 \end{aligned}$$

since $f(c) - f\left(\beta\delta b - p^{SB}\left(\widehat{\beta}\right)\right) \leq 0$, $\forall c \in \left[\beta\delta b - p^{SB}\left(\widehat{\beta}\right), \widehat{\beta}\delta b - p^{SB}\left(\widehat{\beta}\right)\right]$, where the equality holds only when $c = \beta\delta b - p^{SB}\left(\widehat{\beta}\right)$.

So $\delta b\left(1 - \widehat{\beta}\right) \cdot Y + Z \leq 0$, where the equality holds only if $\widehat{\beta} = \beta$.

From part (i) of Proposition 6, $p^{SB\prime}\left(\widehat{\beta}\right) \geq 0$, where the equality holds for the "bunching" types.

Hence, return back to (87), $\widetilde{U}'\left(\widehat{\beta}\right) \leq 0$, where the equality holds for $\widehat{\beta} = \beta$ and bunching types. ■

Figure 1. The Timing Setting of an Investment Good Pricing Model

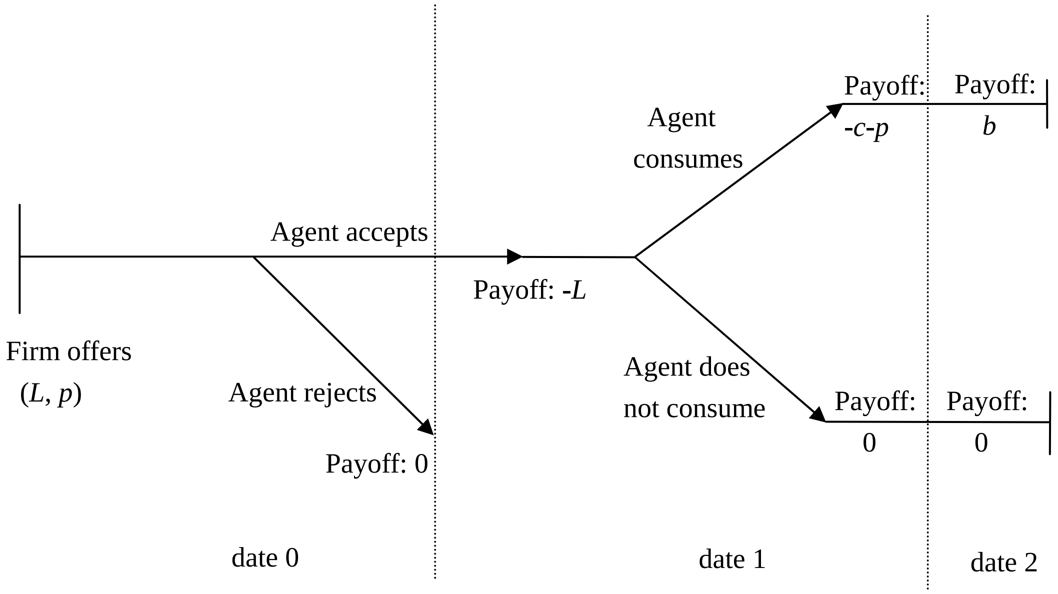


Figure 2.
Illustration of Quasi-hyperbolic Discounting

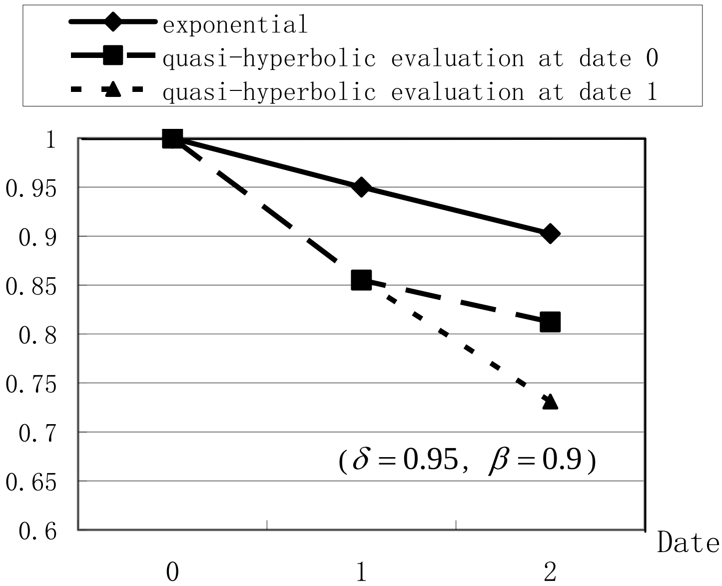


Figure 3.

The Violation of Below-Marginal-Cost Per-usage Pricing Property

when *b* is Consumer's Private Information

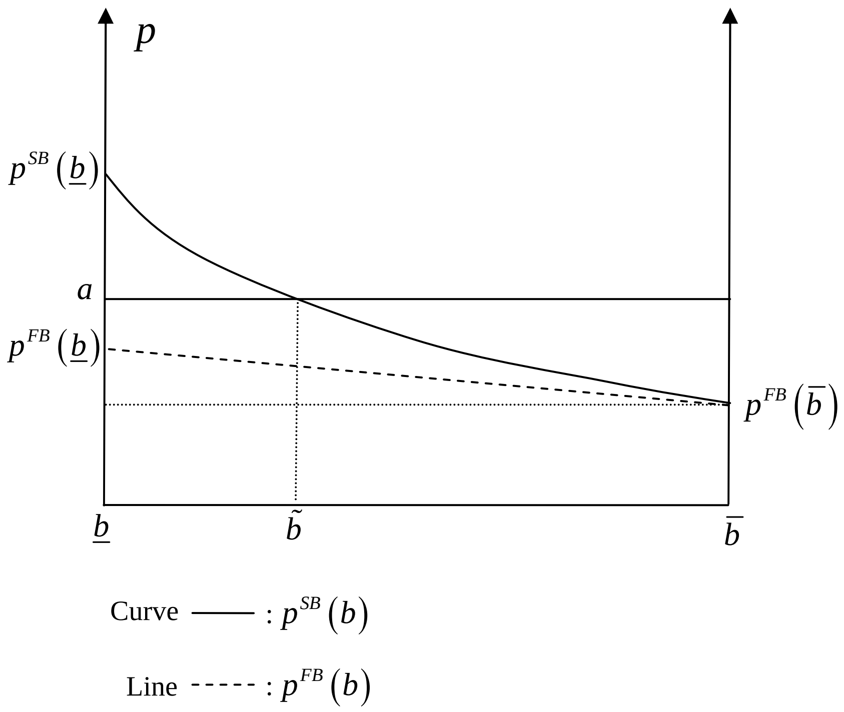


Figure 4.

The Impact of β on Firm's Profit Facing a Sophisticated Consumer when b is Private Information

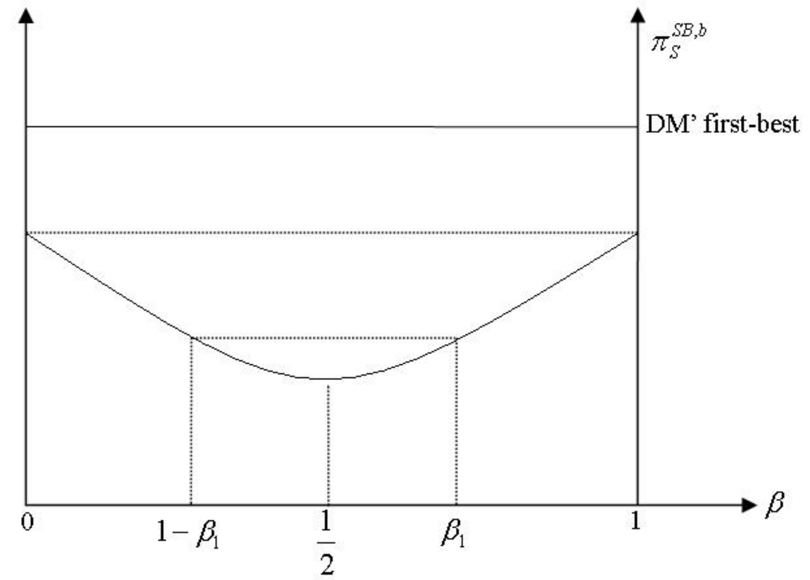
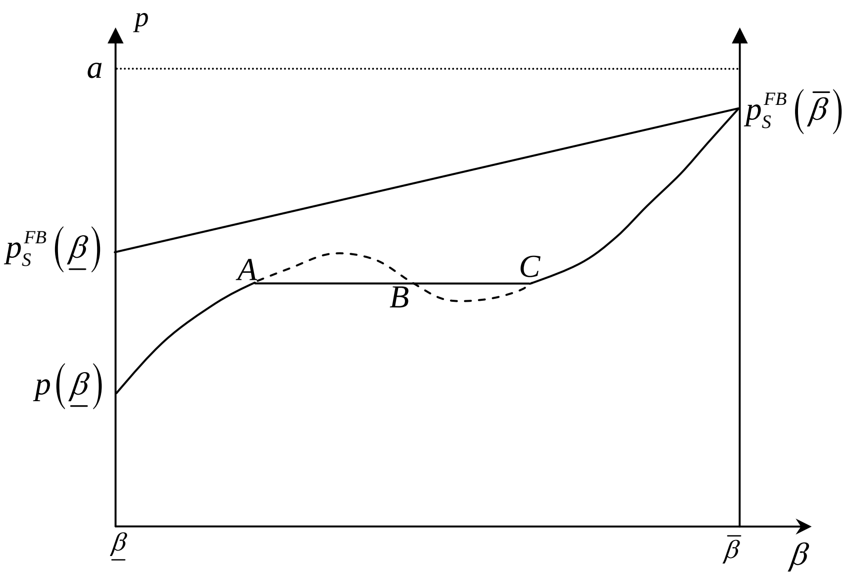


Figure 5. $p^{SB}(\beta)$**, relaxed** $p(\beta)$ **and** $p_S^{FB}(\beta)$



$p(\beta)$: Curve $p(\underline{\beta})ABC\, p_S^{FB}(\bar{\beta})$

$p^{SB}(\beta)$: Curve $p(\underline{\beta})A$ + Line AC + Curve $C\, p_S^{FB}(\bar{\beta})$

Figure 6. $p^{SB}(\hat{\beta})$**, relaxed** $p(\hat{\beta})$ **and** $p_N^{FB}(\hat{\beta})$

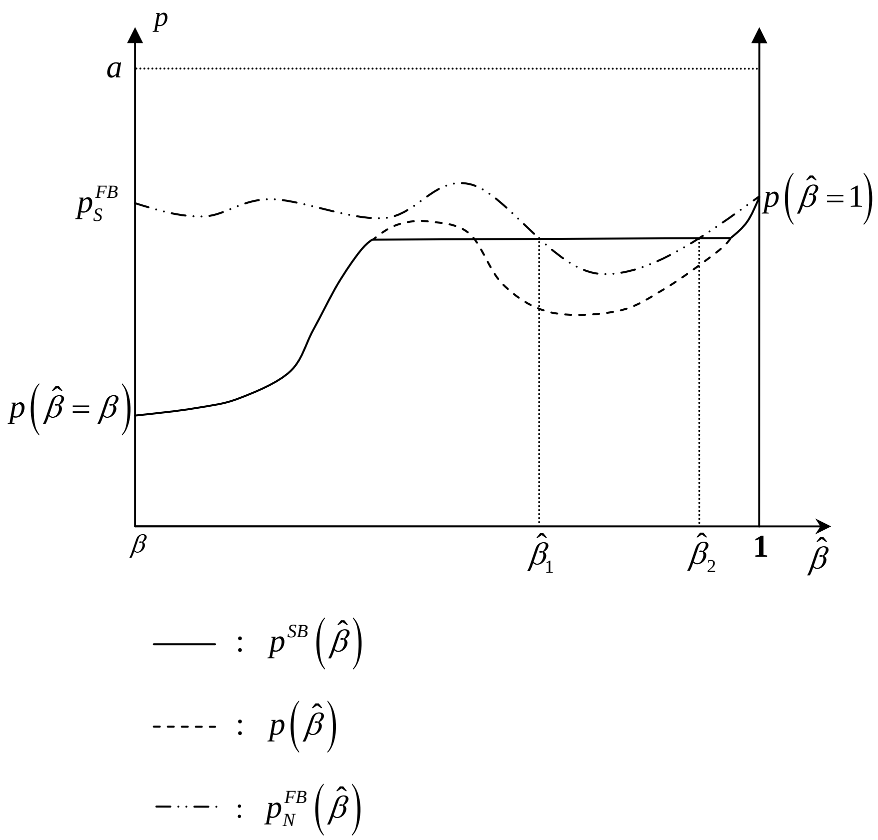
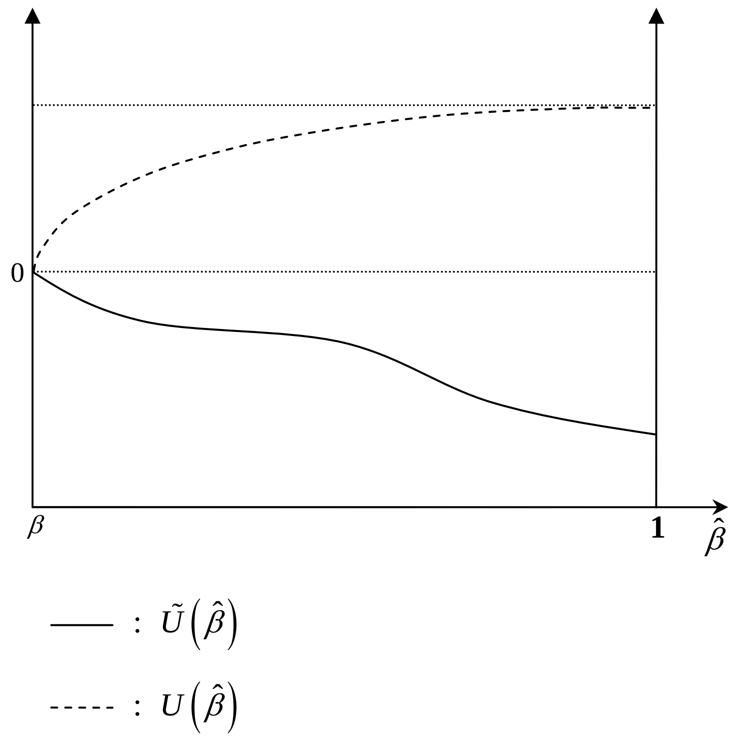


Figure 7. fictitious rent $U\left(\hat{\beta}\right)$ and real rent $\tilde{U}\left(\hat{\beta}\right)$



References

[1] Amador, Manuel, Iván Werning and George-Marios Angeletos, 2006, "Commitment vs. Flexibility", *Econometrica*, 74: 365-396.

[2] Ariely, Dan and Klaus Wertenbroch, 2002, "Procrastination, Deadlines, and Performance: Self-Control by Precommitment," *Psychological Science*, 13: 219–224.

[3] Araujo, Aloisio and Humberto Moreira, 2010, "Adverse selection problems without the Spence–Mirrlees condition", *Journal of Economic Theory*, 145: 1113-1141.

[4] Armstrong, Mark and Jean-charles Rochet, 1999, "Multi-dimensional screening: A user's guide", *European Economic Review*, 43: 959-979.

[5] Benabou, Roland and Jean Tirole, 2004, "Willpower and Personal Rules", *Journal of Political Economy*, 112: 848-886.

[6] Chung, Shin-Ho and Richard J. Herrnstein, 1961, "Relative and Absolute Strengths of Response as a Function of Frequency of Reinforcement", *Journal of the Experimental Analysis of Animal Behavior*, 4: 267-72.

[7] DellaVigna, S. and U. Malmendier, 2004, "Contract Design and Self-Control: Theory and Evidence", *Quarterly Journal of Economics*, 119: 353-402.

[8] DellaVigna, S. and U. Malmendier, 2006, "Paying Not to Go to the Gym", *American Economic Review*, 96: 694-719.

[9] Eliaz, Kfir and Ran Spiegler, 2006, "Contracting with Diversely Naive Agents", *Review of Economic Studies*, 73: 689–714.

[10] Esteban, Susanna and Eiichi Miyagawa, 2006, "Temptation, self-control, and competitive nonlinear pricing", *Economics Letters*, 90: 348-355.

[11] Esteban, Susanna, Eiichi Miyagawa and Matthew Shum, 2007, "Nonlinear pricing with self-control preferences", *Journal of Economic Theory*, 135: 306-338.

[12] Fang, Hanming and Yang Wang, 2010, "Estimating Dynamic Discrete Choice Models with Hyperbolic Discounting, with an Application to Mammography Decisions", Penn Institute for Economic Research Working Paper 10-033.

[13] Gilpatric, Scott M., 2008, "Present-biased preferences, self-awareness and shirking", *Journal of Economic Behavior & Organization*, 67: 735-754.

[14] Gong, Liutang, 2002, *Methods of Dynamic Economics* (in Chinese), Peking University Press.

[15] Guesnerie, R., and J. J. Laffont, 1984, "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm", *Journal of Public Economics*, 25: 329-369.

[16] Gul, F. and Pesendorfer, W., 2001, "Temptation and Self Control", *Econometrica*, 69: 1403-1436.

[17] Kirby, K.N., and R.J. Herrnstein, 1995, "Preference Reversals Due to Myopic Discounting of Delayed Rewards", *Psychological Science*, 6: 83-89.

[18] Laffont, Jean-Jacques and David Martimort, 2002, *The Theory of Incentives*, Princeton University Press.

[19] Laffont, Jean-Jacques and Jean Tirole, 1993, *A Theory of Incentives in Procurement and Regulation*, MIT Press.

[20] Laibson, D., 1997, "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112: 443–477.

[21] Larwood, Laurie and William Whittaker, 1977, "Managerial Myopia: Self-Serving Biases in Organizational Planning," *Journal of Applied Psychology*, 62: 194–198.

[22] Luttmer, Erzo G.J. and Thomas Mariotti, 2003, "Subjective Discounting in an Exchange Economy", *Journal of Political Economy*, 111: 959-989.

[23] Madrian, Brigitte C. and Dennis Shea, 2001, "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior," *Quarterly Journal of Economics*, 116: 1149–1187.

[24] Maskin, Eric and John Riley, 1984, "Monopoly with Incomplete Information", *Rand Journal of Economics*, 15: 171-196.

[25] O'Donoghue, Ted D. and Matthew Rabin, 2001, "Choice and Procrastination", *Quarterly Journal of Economics*, 116: 121-160.

[26] O'Donoghue, Ted and Matthew Rabin, 2005, "Optimal taxes for sin goods", *Swedish Economic Policy Review*, 12: 7-40.

[27] Samuelson, P., 1937, "A Note on Measurement of Utility", *Review of Economic Studies*, 4: 155-161.

[28] Spence, M., 1977, "Nonlinear Prices and Welfare", *Journal of Public Economics*, 8: 1-18.

[29] Strotz, R., 1956, "Myopia and Inconsistency in Dynamic Utility Maximization", *Review of Economic Studies*, 23: 165–180.