Test design under voluntary participation
Supplementary material

Frank Rosar
University of Bonn

March 18, 2017

1. Parametrization of HARA utility

In the paper, I impose the assumptions $\hat{u}' > 0$, $\hat{u}'' < 0$ and $\hat{u}''' \geq 0$ on the agent's reduced-form utility function. I argue here that these assumptions are satisfied for a broad class of HARA utility functions.

The parameterized class of utility functions

$$\hat{u}(\mu) = \begin{cases} -\frac{1}{1-c_1}(c_1\mu + c_2)^{-\frac{1}{c_1}+1} & \text{if } c_1 \in \mathbb{R}\backslash\{0,1\} \text{ and } c_2 \in [\max\{-c_1, 0\}, \infty) \\ -c_2 \exp(-\frac{1}{c_2}\mu) & \text{if } c_1 = 0 \text{ and } c_2 \in (0, \infty) \\ \ln(\mu + c_2) & \text{if } c_1 = 1 \text{ and } c_2 \in (0, \infty) \end{cases}$$

exhibits hyperbolic absolute risk aversion (HARA). The Arrow-Pratt measure of absolute risk-aversion is given by $r_A(\mu) \equiv -\hat{u}''(\mu)/\hat{u}'(\mu) = 1/(c_1\mu + c_2)$. As special cases, this class includes quadratic utility ($c_1 = -1$), cubic utility ($c_1 = -1/2$), exponential/CARA utility ($c_1 = 0$), logarithmic utility ($c_1 = 1$), and CRRA utility ($c_2 = 0$). The parameter space is illustrated by the grey areas in Figure S-1; the special cases are indicated by the colored line segments. $\hat{u}' > 0$ and $\hat{u}'' < 0$ is true for the entire parameter space; $\hat{u}''' \geq 0$ is true for $c_1 \geq -1$. That is, the assumptions in my reduced model are satisfied for any HARA utility function with parameters in the dark grey area. This includes all mentioned special cases.

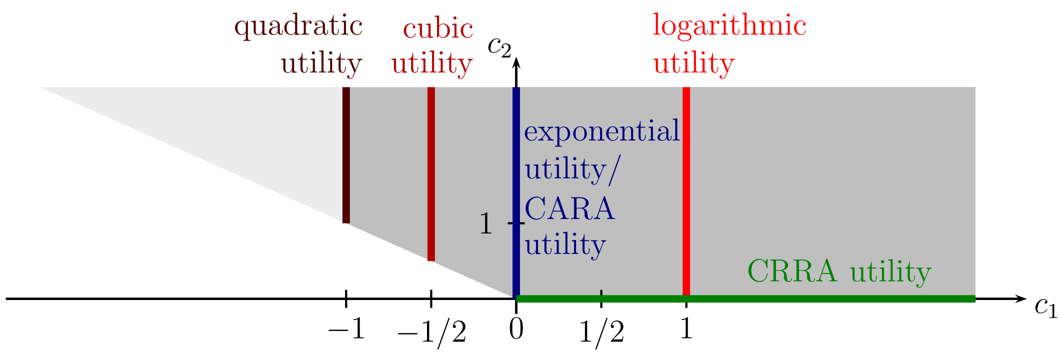

Figure S-1: Parameter space and special cases of HARA utility

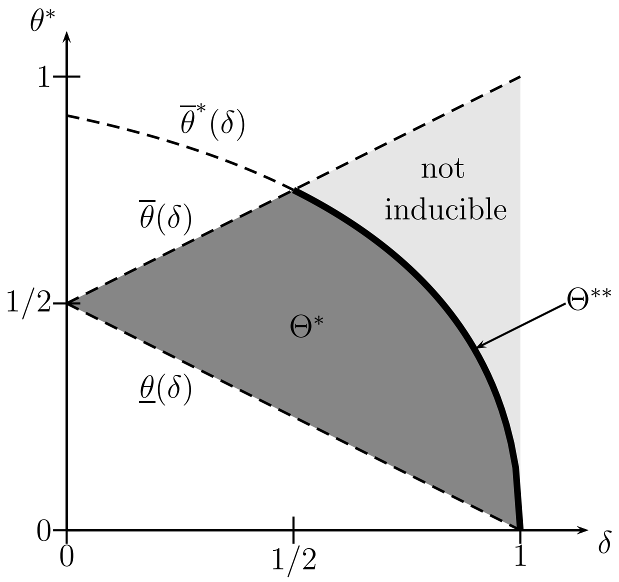

Figure S-2: Inducible participation thresholds $[\hat{u}(\mu) = -(1-\mu)^2,\ \theta \sim \mathrm{U}[\underline{\theta}(\delta), \overline{\theta}(\delta)]]$

2. Discussion of the optimal participation behavior in the uniform-quadratic case

It follows from the analysis in the paper that the participation threshold that the principal does optimally induce solves the following program:

$$\begin{aligned} \max_{\theta^* \in \Theta^*} \quad & (1 - F(\theta^*))\mathbb{E}_\tau[\hat{v}(\mu)] + F(\theta^*)\hat{v}(\mathbb{E}_F[\theta|\theta \leq \theta^*]) \\ \text{s.t.} \quad & \pi^{\text{NFP}}_{\rho(\theta^*)} \text{ and } \chi_{\theta^*} \text{ induce } \tau \\ & \rho(\theta^*) \text{ solves } \mathbb{E}_{\tau_{\theta^*}}[\hat{u}(\mu)] = \hat{u}(\mathbb{E}_F[\theta|\theta \leq \theta^*]) \text{ if } \theta^* \in \Theta^* \backslash \Theta^{**} \\ & \rho(\theta^*) = 1 \text{ if } \theta^* \in \Theta^{**} \end{aligned} \tag{S-1}$$

In this supplementary material, I discuss the choice of the optimal participation threshold in the uniform-quadratic case:

Assumption S-1 $\hat{u}(\mu) = -(1-\mu)^2$, $\hat{v}(\mu) = (\mu - \mathbb{E}_F[\theta])^2$, *and* θ *is uniformly distributed on* $[\underline{\theta}(\delta), \overline{\theta}(\delta)] \equiv [(1-\delta)/2, (1+\delta)/2]$ *with* $\delta \in (0, 1)$.

Under this assumption, a higher δ is associated with more informative private information. I obtain from Lemma A2 in the paper that $\Theta^* = \{\theta^* \in [\underline{\theta}, \overline{\theta}) | \theta^* \leq \overline{\theta}^*(\delta)\}$ and $\Theta^{**} = \{\theta^* \in [\underline{\theta}, \overline{\theta}) | \theta^* = \overline{\theta}^*(\delta)\}$ with $\overline{\theta}^*(\delta) \equiv 2\sqrt{\underline{\theta}(\delta)} - \underline{\theta}(\delta)$. Figure S-2 illustrates how the sets Θ^* and Θ^{**} depend on δ. A rough intuition is the following: Inducing a given participation threshold θ^* consists of two parts, the motivation of participation for private signals $\theta > \theta^*$ and the deterrence of participation for private signals $\theta < \theta^*$. By making the test sufficiently uninformative, the motivation part can always be satisfied. Thus, crucial for inducibility is the deterrence part. If $\delta \leq 1/2$, the agent's signaling motive is weak for any possible private signal. By making the test sufficiently informative, he can for any possible private signal be deterred from participating. As a consequence, any threshold $\theta^* \in [\underline{\theta}(\delta), \overline{\theta}(\delta))$ is inducible. If $\delta > 1/2$, there exist signals for that the agent is sufficiently certain to be good such that he cannot be deterred from participating by increasing the informativeness of the test. Every threshold $\theta^* \in (\overline{\theta}^*(\delta), \overline{\theta}(\delta))$ is then not inducible. Due to a partial unraveling effect that becomes stronger as δ increases, the set of private signals for that the agent cannot be deterred from participating becomes larger as δ increases. If δ is close to one, there is almost full unraveling for any informative test.

In Section 3 of the paper, I have argued that an accurate NFP test induces less participation than any sufficiently inaccurate NFP test. Although the switch from a very inaccurate NFP test to an accurate NFP test causes clearly less participation, participation may locally increase as the accuracy of the test increases. Responsible for this is that an increase in the participation threshold improves the pool of participants as well as the pool of non-participants. Depending on the relative strength of these improvements, participation may get more or less attractive in response to an increase in the supposed participation threshold. This implies that it is a priori not clear whether the equilibrium participation threshold increases or decreases in response to a small change in the test accuracy. The subsequent proposition establishes that there is under Assumption S-1 also locally for any intermediate participation threshold a trade-off between accuracy and participation.

Proposition S-1 (Trade-off between accuracy and participation) *Suppose that Assumption S-1 holds and consider for each $\theta^* \in \Theta^*$ the test $\pi^{NFP}_{\rho(\theta^*)}$ that induces θ^*. Then, $\rho(\theta^*)$ is increasing; that is, more participation is induced by a less accurate test.*

Proof. It follows from Lemma A2 in the proof of Proposition 2 in the paper that Θ^* is under Assumption S-1 a convex set. Let $\rho(\theta^*)$ be as defined in (S-1). $\rho(\theta^*)$ is the unique solution to $\mathbb{E}_{\tau_{\theta^*}}[\hat{u}(\mu)] = \hat{u}(\mathbb{E}_F[\theta|\theta \leq \theta^*])$ if $\theta^* \in \Theta^* \backslash \Theta^{**}$. By Lemma A2 (b) in the proof of Proposition 2 in the paper, the same equation holds if $\theta^* \in \Theta^{**}$. Under Assumption S-1, the solution is also in this case unique. Thus, for any $\theta^* \in \Theta^*$, I can write the equation that uniquely determines the test accuracy ρ as

$$g(\rho, \theta^*) \equiv (1 - \rho\theta^*)\hat{u}(\mu_1(\rho, \theta^*)) + \rho\theta^*\hat{u}(1) - \hat{u}(\mathbb{E}_F[\theta|\theta \leq \theta^*]) = 0$$

with

$$\mu_1(\rho, \theta^*) \equiv \frac{\mathbb{E}_F[\theta|\theta \geq \theta^*] - \mathbb{E}_F[\theta|\theta \geq \theta^*]\rho}{1 - \mathbb{E}_F[\theta|\theta \geq \theta^*]\rho}.$$

By applying the Implicit Function Theorem, I obtain that $\rho(\theta^*)$ is a continuously differentiable function with

$$\rho'(\theta^*) = -\frac{\partial g(\rho, \theta^*)/\partial \theta^*}{\partial g(\rho, \theta^*)/\partial \rho}.$$

I will conclude this proof by showing that $\partial g(\rho, \theta^*)/\partial \theta^* > 0$ (Step 1) and that $\partial g(\rho, \theta^*)/\partial \rho < 0$ (Step 2).

Step 1. I have,

$$
\begin{aligned}
\frac{\partial g(\rho,\theta^*)}{\partial\theta^*} &= \rho(\hat{u}(1)-\hat{u}(\mu_1(\rho,\theta^*))))\\
&\quad +(1-\rho\theta^*)\frac{1}{2}\frac{1-\rho}{(1-\mathbb{E}_F[\theta|\theta\geq\theta^*]\rho)^2}\hat{u}'(\mu_1(\rho,\theta^*))-\frac{1}{2}\hat{u}'(\mathbb{E}_F[\theta|\theta\leq\theta^*])\\
&> \rho(\hat{u}(1)-\hat{u}(\mu_1(\rho,\theta^*)))\\
&\quad +\frac{1}{2}\left(\frac{1-\rho\theta^*}{1-\mathbb{E}_F[\theta|\theta\geq\theta^*]\rho}\frac{1-\rho}{1-\mathbb{E}_F[\theta|\theta\geq\theta^*]\rho}-1\right)\hat{u}'(\mu_1(\rho,\theta^*))\\
&> \rho(\hat{u}(1)-\hat{u}(\mu_1(\rho,\theta^*)))\\
&\quad +\frac{1}{2}\left(\frac{1-\rho}{1-\mathbb{E}_F[\theta|\theta\geq\theta^*]\rho}-1\right)\hat{u}'(\mu_1(\rho,\theta^*))\\
&= \rho(\hat{u}(1)-\hat{u}(\mu_1(\rho,\theta^*)))-\frac{1}{2}\rho(1-\mu_1(\rho,\theta^*))\hat{u}'(\mu_1(\rho,\theta^*))\\
&= 0.
\end{aligned}
$$

The transformations arise as follows: The first equality follows from using that

$$
\frac{\mathrm{d}\mathbb{E}_F[\theta|\theta\geq\theta^*]}{\mathrm{d}\theta^*}=\frac{\mathrm{d}\mathbb{E}_F[\theta|\theta\leq\theta^*]}{\mathrm{d}\theta^*}=\frac{1}{2}
$$

under Assumption S-1, and that

$$
\frac{\partial\mu_1(\rho,\theta^*)}{\rho\theta^*}=\frac{1}{2}\frac{1-\rho}{(1-\mathbb{E}_F[\theta|\theta\geq\theta^*]\rho)^2}.
$$

To see why the first inequality holds, notice first that necessary for $g(\rho,\theta^*)=0$ is that $\mu_1(\rho,\theta^*)<\mathbb{E}_F[\theta|\theta\leq\theta^*]$. The inequality follows because $\hat{u}''<0$ implies then that $\hat{u}'(\mu_1(\rho,\theta^*))>\hat{u}'(\mathbb{E}_F[\theta|\theta\leq\theta^*])$. The second inequality follows from using that $\theta^*<\mathbb{E}_F[\theta|\theta\geq\theta^*]$ implies that $(1-\rho\theta^*)/(1-\mathbb{E}_F[\theta|\theta\geq\theta^*]\rho)>1$. The second equality follows from using that I can rewrite the expression in parentheses as $-\rho(1-\mu_1(\rho,\theta^*))$. The third equality follows from using that the structure of $\hat{u}$ in Assumption S-1 implies that $\hat{u}(1)-\hat{u}(\mu_1)=1/2\cdot(1-\mu_1)\hat{u}'(\mu_1)$.

Step 2. I have,

$$
\begin{aligned}
\frac{\partial g(\rho,\theta^*)}{\partial\rho} &= \theta^*(\hat{u}(1)-\hat{u}(\mu_1(\rho,\theta^*))))\\
&\quad -(1-\rho\theta^*)\frac{\mathbb{E}_F[\theta|\theta\geq\theta^*]}{1-\mathbb{E}_F[\theta|\theta\geq\theta^*]\rho}(1-\mu_1(\rho,\theta^*))\hat{u}'(\mu_1(\rho,\theta^*))\\
&= \theta^*(1-\mu_1(\rho,\theta^*))\\
&\quad \cdot\Big[\frac{\hat{u}(1)-\hat{u}(\mu_1(\rho,\theta^*))}{1-\mu_1(\rho,\theta^*)}-\frac{1-\rho\theta^*}{1-\mathbb{E}_F[\theta|\theta\geq\theta^*]\rho}\frac{\mathbb{E}_F[\theta|\theta\geq\theta^*]}{\theta^*}\hat{u}'(\mu_1(\rho,\theta^*))\Big]\\
&> \theta^*(1-\mu_1(\rho,\theta^*))\cdot\Big[\frac{\hat{u}(1)-\hat{u}(\mu_1(\rho,\theta^*))}{1-\mu_1(\rho,\theta^*)}-\hat{u}'(\mu_1(\rho,\theta^*))\Big]\\
&> 0.
\end{aligned}
$$

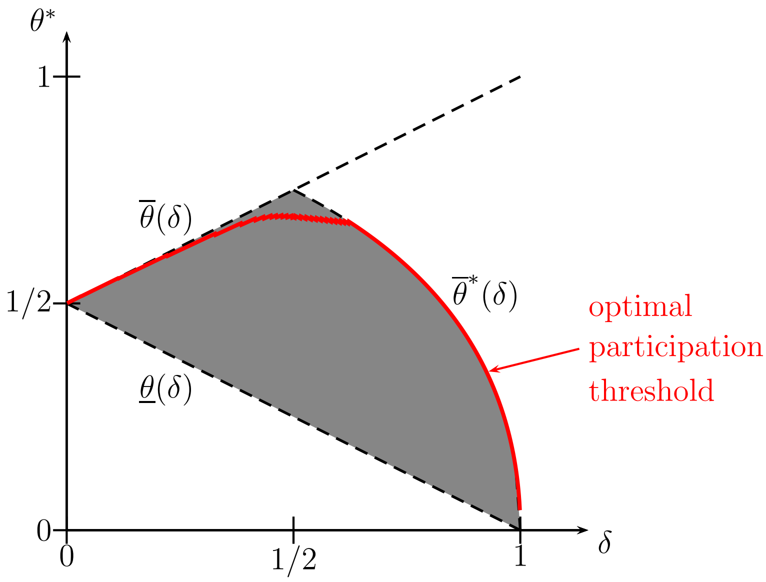

Figure S-3: Optimal participation threshold under Assumption S-1

The transformations arise as follows: The first equality follows from using that

$$\begin{aligned}\frac{\partial \mu_1(\rho,\theta^*)}{\partial \rho} &= -\frac{\mathbb{E}_F[\theta|\theta \geq \theta^*](1-\mathbb{E}_F[\theta|\theta \geq \theta^*])}{(1-\mathbb{E}_F[\theta|\theta \geq \theta^*]\rho)^2} \\ &= -\frac{\mathbb{E}_F[\theta|\theta \geq \theta^*]}{1-\mathbb{E}_F[\theta|\theta \geq \theta^*]\rho}(1-\mu_1(\rho,\theta^*)).\end{aligned}$$

The second equality follows from rearranging. The first inequality follows from $\mathbb{E}_F[\theta|\theta \geq \theta^*] > \theta^*$. The second inequality follows from $\hat{u}'' < 0$. q.e.d.

I can interpret any binary test as a pass-fail test where the negative result $s = 1$ corresponds to "fail" and the positive result $s = 2$ corresponds to "pass". If the test is completely informative, the agent passes if, and only if, he is good; if the test is subject to false positives, the agent does sometimes pass even when he is bad; if the test is subject to false negatives, the agent does sometimes fail although he is good. If the test is modified such that the implied quality perception associated to failing increases, I can interpret this as a reduction of the "stigma of failure". On the other hand, if the test is modified such that passing gets more likely for any θ (i.e., such that the induced $\tau_\theta(\mu_2)$ increases for any θ), I can interpret this as "grade inflation". Decreasing the accuracy of a NFP test makes passing less likely and increases the quality perception associated with failing. This allows for the interpretation that more participation is optimally induced by reducing the stigma of failure and not by inflating grades (even though this would be possible as well).

I come now to the question of how the informativeness of the agent's private information δ affects the optimal participation threshold. The grey area in Figure S-3 indicates which participation thresholds are inducible; the red solid line illustrates the optimal participation threshold.[1] If the informativeness of the agent's private signal is low, any participation threshold $\theta^* \in [\underline{\theta}(\delta), \overline{\theta}(\delta))$ is inducible. As signals are spread out in a small interval around 1/2, the worst possible quality perception associated to non-participation is quite high and the agent's signaling motive is weak. Fostering participation is costly for the principal in terms of test accuracy. The principal prefers little participation in a relatively accurate test to a higher participation in a much less accurate test. Thus, the optimal participation threshold lies close to $\overline{\theta}(\delta)$. If the informativeness of the agent's private signal is high, the agent's signaling motive is strong as the quality perception associated to non-participation is low. An unravelling effect kicks in causing relatively high participation in an accurate test. As inducing participation for private signals close to $\underline{\theta}(\delta)$ requires a very inaccurate test, fostering even more participation is, again, very costly in terms of test accuracy. An accurate test and therewith the highest inducible participation threshold is optimal. As inducing full participation is never optimal, there is always a role for indirect learning through the agent's participation decision.

3. Relation to problems with non-binary quality

Suppose the agent's quality ω can assume three values, say $\Omega = \{\omega_1, \omega_2, \omega_3\}$ with $0 = \omega_1 < \omega_2 < \omega_3 = 1$. In an important class of economic environments, the principal's and the agent's payoff depend only on the expected quality from the principal's perspective.[2] Let $\mu \in \Delta(\Omega)$ denote now the principal's posterior belief. $\check{\mu}(\mu) \equiv \mathbb{E}_\mu[\omega]$ describes then the expected quality associated to the principal's belief. I can write reduced-form payoffs as $\hat{v}(\check{\mu}(\mu))$ and $\hat{u}(\check{\mu}(\mu))$, and I can impose on $\hat{v}$ and on $\hat{u}$ the same assumptions that I imposed in my model with binary quality. Suppose further that the agent's private signal is a prior belief $\theta \in \Theta$ with $\Theta \subset \Delta(\Omega)$. Thus, when the agent has to take his participation decision, the expected quality from his perspective is $\check{\theta}(\theta) \equiv \mathbb{E}_\theta[\omega]$. When I impose assumptions that imply that the agent cares only through $\check{\theta}(\theta)$ about θ and that $\check{\theta}(\theta)$ allows only for inferences about $\check{\mu}(\mu)$, my binary problem is strongly related to the non-binary problem.

The analysis of the non-binary problem is, however, somewhat more involved. A test and a participation strategy induce a distribution of expected qualities $\check{\mu}$ from the principal's perspective, say $\check{\tau}$, and from the agent's interim perspective, say $\check{\tau}_\theta$. The principal's test design problem can then be transformed into the problem of designing a distribution of expected qualities $\check{\tau}$. Yet besides the participation constraint and the Bayesian plausibility constraint that arise also in the binary model, a "technical feasibility constraint" arises additionally in the non-binary model. To see this, suppose that quality is $\omega_1 = 0$, $\omega_2 = 1/2$ and $\omega_3 = 1$ with probability 1/3 each. Any induced distribution of expected qualities $\check{\tau}$ has then expected value 1/2 (by Bayesian plausibility) and a support that is a subset of $[0, 1]$ (because the expected

[1] I used Maple to derive the optimal participation threshold numerically.

[2] Payoffs with such a structure arise, for example, when the principal takes a decision to minimize the quadratic distance between her decision and the actual quality and the agent cares only about the principal's decision.

quality associated to each test result is a convex combination of ω_1, ω_2 and ω_3). However, not every distribution $\check{\tau}$ with an expected value of 1/2 and such a support can be induced. As the convex combination associated with at least one test result has to put positive weight on ω_2, it is for instance not possible to induce a distribution where $\check{\mu}_1 = 0$ and $\check{\mu}_2 = 1$ arise both with probability 1/2.

Besides the additional technical issue that has to be handled, imposing an assumption on the independence of information and test results conditional on quality may be more restrictive in the non-binary model than in the binary model. In reasonable problems with non-binary quality, this assumption is violated. Therefore, allowing for more general forms of private information would be interesting as well. This would give rise to additional issues that make the analysis less tractable (e.g., the existence of tests that induce participation behavior that is non-monotonic in $\check{\theta}(\theta)$).