Selling Training Data

Jingmin Huang, Wei Zhao and Renjie Zhong

Renmin University of China

2024

This slide deck:

Selling Training Data

- data (or training features) is trained to make **predictions** to improve the quality of **decision making under uncertainty**
- data buyer seeks to augment the private data and refine the initial prediction by purchasing more data
- data seller versions the training data and designs the associated tariff to screen buyers with different private training data
- our question: what is the **optimal data selling mechanism**?
- two screening toolkits: **prediction accuracy of data** and **price**

Input Data and Training Data

- in Bergemann et al. (2018), seller designs *input data*, like cookie and purchase history, to buyers with **relevant knowledge**
  1. lenders with **independent knowledge of a borrower** purchase data to make predictions about whether to lend
  2. health care providers **with access to a patient's family** make predictions to enhances health care delivery
- in ours, buyers with private dataset purchase *training data* to **append her private one**, in order to **train algorithms** to predict outcomes and recognize patterns
  1. the purchased medical data is used for **training supervised learning models** to predict disease diagnosis or drug efficacy
  2. the purchased financial data is used for the **stock price prediction/credit scoring**.

Differences between Data and Normal Goods

- **two key attributes of data** makes it different from normal goods, complicating this design problem
- **first**, data generates **predictions across different states**, and its predictive accuracy in different states may be different
- the buyers share different demands for refined accuracy in different states, i.e. the **preference** is **multi-dimensional**
- the seller can allocate different predictive accuracy in different states in a data, i.e. the **allocation** is **multi-dimensional**
- the seller can utilize the interactions between different dimensions to extract the information rent

Differences between Data and Normal Goods

- **second**, the value of data is **intrinsically combinatorial**, i.e., different datasets have **correlations in signals**, leading to sub-additive or super-additive valuations.
- the **complementarity** and **substitute** between data differs, implying that merging training data $\neq$ naive combination of predictions
- the buyer make **endogenous combination** of predictions to **minimize the dis-utility from statistical error**
- the seller can utilize the complementarity between the selling data and the private data to relax the **incentive compatibility** of buyers with different private data

Trade-off

basic trade-off: **information rent extraction** versus **efficient value extraction**

1. efficient value extraction $\Rightarrow$ provide data with perfect prediction to all types
2. rent extraction $\Rightarrow$ decrease the prediction accuracy for low types

remind that data is

1. **multi-dimensional** $\Rightarrow$ decrease prediction accuracy in different states for different types
2. **intrinsically combinatorial** $\Rightarrow$ utilize the endogenous combination to avoid data distortion

Main Result

binary situation: binary type, binary action, binary state

training data screening: **four** different selling schemes

1. **exclusive policy**
2. **no** discrimination
3. **partial** discrimination
4. **perfect** discrimination

normal goods screening: 1 & 2 (Riley and Zeckhauser, 1983)

input data (information) screening: 1 & 2 & 4 (Bergemann et al., 2018) in the general case, we fix the predictive accuracy in one dimension

in this case, the optimal mechanism is **two-tiered pricing**:

1. in the first tier, a partially informative data, perfectly predicting one state
2. in the second tier, data with complete prediction accuracy
3. the threshold is determined similar to the monopolist pricing

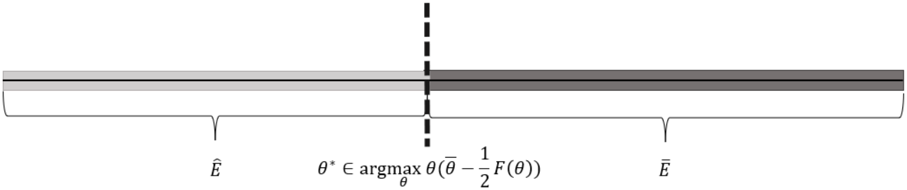


Figure: Two-tierd Pricing Mechanism

Literature Review

1. the design and price of information: Admati and Pfleiderer (1986); Admati and Pfleiderer (1990); Babaioff et al. (2012); Bergemann and Bonatti (2015); **Bergemann et al. (2018)**, Li (2022)
2. multi-dimensional screening: Adams and Yellen (1976); McAfee et al. (1989); Armstrong and Rochet (1999); Carroll (2017); Haghpanah and Hartline (2021); Yang (2021); Yang (2022)
3. applications of the infinite-dimensional extension of Carathéodory's theorem in economic design: Fuchs and Skrzypacz (2015); Bergemann et al. (2018); Kang (2023); Loertscher and Muir (2023); Dworczak and Muir (2024); Le Treust and Tomala (2019); Doval and Skreta (2024)

This slide deck:

Timeline

1. the seller posts a mechanism $\mathcal{M} = \{\mathcal{E}, t\}$
   1. a collection of experiments ϵ
   2. associated tariff $t : \mathcal{E} \to \mathbb{R}_+$
2. the buyer chooses an experiment $E \in \epsilon$ and pays price $t(E)$
3. the true state ω is realized
4. the buyer receive **two signals** to update her belief, one from her **private experiment**, another from **the experiment E she purchased**, and she chooses an action a to maximize her expected utility
5. payoffs are realized

Individual Decision Problem

- finite states $\Omega = \{\omega_1, ..., \omega_I\}$ and common prior $\mu \in \Delta(\Omega)$, $\mu_i \equiv \mu(\omega_i)$
- the buyer chooses an action to maximize his expected payoff based on his information, from the finite action set $A = \{a_1, ..., a_J\}$
- utility function $u : A \times \Omega \to \mathbb{R}$, $u_{ij} \equiv u(a_j, \omega_i)$

u	a_1	$\cdots$	a_J
ω_1	u_{11}	$\cdots$	u_{1J}
$\vdots$	$\vdots$		$\vdots$
ω_I	u_{I1}	$\cdots$	u_{IJ}

- hereafter we assume matching utility if $u(a_j, \omega_i) = 1_{i=j}$ and $I = J$ to simplify harmful algebra

Buyer's Private Data

- type $\theta \sim F(\theta)$ buyer, with private data E'_θ, decides to purchase data E_θ
- $E'_\theta = \{S, \pi'_\theta\}$ consists of signals $S' = \{s'_1, ..., s'_K\}$, with $\pi : \omega \to \Delta S$, $\pi'_{\theta ik} \equiv Pr[s'_k|\omega_i]$

E'_θ	s'_1	$\cdots$	s'_K
ω_1	$\pi'_{\theta 11}$	$\cdots$	$\pi'_{\theta 1K}$
$\vdots$	$\vdots$		$\vdots$
ω_I	$\pi'_{\theta I1}$	$\cdots$	$\pi'_{\theta IK}$

- each signal induces a posterior $\Rightarrow$ the agent's type is the **distribution of posteriors**, i.e. $\{\mu_{\theta 1}, ..., \mu_{\theta K}\}$ with $\Pr(\mu_{\theta k}) = \sum_{i=1}^{I} \mu_i \pi'_{\theta ik}$ for all $k \in 1, ..., K$, where $\mu_{\theta k} \in \Delta(\Omega)$

Outside Option without Purchasing Data

- without seller's data, conditioning on the prediction (signal) from the private data he accepted, the buyer chooses an optimal action
- the payoff is his expected value (in signals), constituting the **outside option** in this mechanism
- optimal action and payoff conditional on accepting s'_k for agent θ:

$$a\left(s'_k \mid E_\theta\right) \in \arg\max_{a_j \in A}\{\textstyle\sum_{i=1}^{I} \mu_{\theta k i} u_{ij}\} \text{ and } u(s'_k \mid E_\theta) \triangleq \max_j\{\textstyle\sum_{i=1}^{I} \mu_{\theta k i} u_{ij}\}$$

- expected payoff for agent θ:

$$u_\theta \triangleq \textstyle\sum_{k=1}^{K} \Pr(\mu_{\theta k}) u(s'_k \mid E_n) = \sum_{k=1}^{K} \max_j \left\{\sum_{i=1}^{I} \mu_i \pi'_{\theta i k} u_{ij}\right\}$$

Value of Data in Individual Decision Problem

- suppose that the buyer combine the prediction from the purchased data with the one from his private data w.l.o.g (order invariance of Bayesian Updating)
- the optimal action and payoff conditional on accepting s_r and s'_k for agent n:

$$a\left(s_r \mid s'_k\right) \in \underset{a_j \in A}{\arg\max}\left\{\sum_{i=1}^{I}\left(\frac{\mu_{\theta k i}\pi_{ir}}{\sum_{i'=1}^{I}\mu_{\theta k i'}\pi_{i' r}}\right)u_{ij}\right\}$$

$$u\left(s_r \mid s'_k\right) \triangleq \max_{j}\left\{\sum_{i=1}^{I}\left(\frac{\mu_{\theta k i}\pi_{ir}}{\sum_{i'=1}^{I}\mu_{\theta k i'}\pi_{i' r}}\right)u_{ij}\right\}$$

- expected payoff for agent n:

$$u(\mathrm{E}, \theta) \triangleq \sum_{r=1}^{R}\sum_{k=1}^{K}\max_{j}\left\{\sum_{i=1}^{I}\mu_i \pi'_{nik}\pi'_{ir}u_{ij}\right\}$$

- the value of data: $V(\mathrm{E}, \theta) \triangleq u(\mathrm{E}, \theta) - u_\theta$

Designer's Problem

- the seller posts a menu $\mathcal{M} = \{\mathcal{E}, t\}$ to maximize his profits
- we can restrict to the direct menu $\mathcal{M} = \{E_\theta, t_\theta\}_{\theta \in \Theta}$ by the revelation principle

Designer's Problem

$$\max_{\mathcal{M}} \int_\Theta t_\theta dF(\theta)$$

$$V(E_\theta, \theta) - t_\theta \geq 0, \ \forall \theta \in \Theta \qquad \text{(IR)}$$

$$V(E_\theta, \theta) - t_\theta \geq V(E_{\theta'}, \theta) - t_{\theta'}, \ \forall \theta, \theta' \in \Theta \qquad \text{(IC)}$$

This slide deck:

Motivation

we need some structural properties of the experiments in the optimal menu to

1. drop the **maximizer operator** $\Rightarrow$ "revelation principle" (recommendation)
2. reduce the **dimensions of screening** $\Rightarrow$ structural properties
3. tackle the **interactions between obedience in information design and mutual IC analysis**

"revelation principle" in information design

1. signal set $S \Rightarrow$ action profile $\times_{k=1}^{K} A$ for all possible posteriors
2. signal realization $s \Rightarrow$ recommendation profile $a^r = (a_{r1}, ..., a_{rK})$ for all possible posteriors

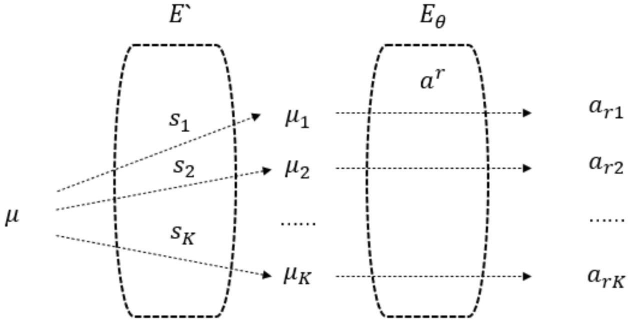


Figure: Direct Recommendation Mechanism

binary state $\{\omega_1, \omega_2\}$, binary action $\{a_1, a_2\}$

binary signal induces binary posterior $\{\mu_{\theta 1}, \mu_{\theta 2}\}$, and reorder $\frac{\mu_{\theta 1}(\omega_1)}{\mu_{\theta 1}(\omega_2)} \geq \frac{\mu_{\theta 2}(\omega_1)}{\mu_{\theta 2}(\omega_2)}$

only need to design recommendation schemes with a^1, a^2, a^3 as below:

E_θ	$a^1 = (a_1, a_1)$	$a^2 = (a_1, a_2)$	$a^3 = (a_2, a_2)$	$a^4 = (a_2, a_1)$
ω_1	$1 - \pi_2$	π_2	0	0
ω_2	0	π_1	$1 - \pi_1$	0

π_i: probability inducing **statistical error** from choosing a_{-i} in ω_i

$a^1 = (a_1, a_1)$, $a^3 = (a_2, a_2)$ always be obeyed (for any θ')

obedience of $a^2 = (a_1, a_2)$ for θ matters

in binary type, $a^2 = (a_1, a_2)$ should be obeyed (for any θ') in the optimal case

This slide deck:

Binary Situation

- binary state $\{\omega_1, \omega_2\}$, action $\{a_1, a_2\}$, type $\{\theta, \theta'\}$, and posterior
- common prior $\mu = (\frac{1}{2}, \frac{1}{2})$, uniform type distribution $\Pr(\text{type } \theta) = \frac{1}{2}$
- to simplify the notation, denote the two type as:

$$
\begin{aligned}
\mu &= \left(\frac{1}{2}, \frac{1}{2}\right) \\
&= \frac{\mu_2 - \frac{1}{2}}{\mu_1 + \mu_2 - 1}(\mu_1, 1 - \mu_1) + \frac{\mu_1 - \frac{1}{2}}{\mu_1 + \mu_2 - 1}(1 - \mu_2, \mu_2) \\
&= \frac{\mu_2' - \frac{1}{2}}{\mu_1' + \mu_2' - 1}(\mu_1', 1 - \mu_1') + \frac{\mu_1' - \frac{1}{2}}{\mu_1' + \mu_2' - 1}(1 - \mu_2', \mu_2')
\end{aligned}
$$

- suppose $\mu_1, \mu_1', \mu_2, \mu_2' > \frac{1}{2}$ and $V(\bar{\mathsf{E}}, \theta) \geqslant V(\bar{\mathsf{E}}, \theta')$ w.l.o.g

Valuation for the Fully Informative Experiment

the valuation of the fully informative experiment $\bar{\mathsf{E}}$ for type θ and θ'

$$V(\bar{\mathsf{E}},\theta) = 1 - \frac{\left(\mu_2-\frac{1}{2}\right)\mu_1}{\mu_1+\mu_2-1} - \frac{\left(\mu_1-\frac{1}{2}\right)\mu_2}{\mu_1+\mu_2-1} = \underbrace{\frac{\left(\mu_2-\frac{1}{2}\right)(1-\mu_1)}{\mu_1+\mu_2-1}}_{\theta_1} + \underbrace{\frac{\left(\mu_1-\frac{1}{2}\right)(1-\mu_2)}{\mu_1+\mu_2-1}}_{\theta_2}$$

$$V(\bar{\mathsf{E}},\theta') = \underbrace{\frac{\left(\mu_2'-\frac{1}{2}\right)(1-\mu_1')}{\mu_1'+\mu_2'-1}}_{\theta_1'} + \underbrace{\frac{\left(\mu_1'-\frac{1}{2}\right)(1-\mu_2')}{\mu_1'+\mu_2'-1}}_{\theta_2'}$$

$$V(\bar{\mathsf{E}},\theta) \geqslant V(\bar{\mathsf{E}},\theta') \Rightarrow \theta_1+\theta_2 \geqslant \theta_1'+\theta_2'$$

Economic Meaning for θ_i

- now the private type $\theta = (\theta_1, \theta_2)$ and $\theta' = (\theta'_1, \theta'_2)$
- θ_i and θ'_i represent **preference** for information in state i, reflecting the **prediction accuracy** of private data
- **vertical preference**: overall prediction accuracy $\theta_1 + \theta_2$
- **horizontal preference**: prediction accuracy across different states θ_1 and θ_2
- take θ_i as example, the components of θ_i:

$$\theta_1 = \underbrace{\frac{\mu_2 - \frac{1}{2}}{\mu_1 + \mu_2 - 1}}_{\text{market share}} \underbrace{(1 - \mu_1)}_{V(\bar{E},(\mu_1,1-\mu_1))}, \ \theta_2 = \underbrace{\frac{\left(\mu_1 - \frac{1}{2}\right)}{\mu_1 + \mu_2 - 1}}_{\text{market share}} \underbrace{(1 - \mu_2)}_{V(\bar{E},(\mu_2,1-\mu_2))}$$

- the higher **prediction accuracy** for state i, the lower **preference**

Designer's Problem

u	$a^1 = (a_1, a_1)$	$a^2 = (a_1, a_2)$	$a^3 = (a_2, a_2)$
ω_1	$1 - \pi_2$	π_2	0
ω_2	0	π_1	$1 - \pi_1$

by the shared responsiveness, the obedience constraint is:

$$\underbrace{\max\{\theta_1\pi_1 + \theta_2\pi_2, \theta_1'\pi_1 + \theta_2'\pi_2\}}_{\textbf{disutility when obeying } (a_1,a_2)} \leqslant \underbrace{\min\left\{\frac{1}{2}\pi_1, \frac{1}{2}\pi_2\right\}}_{\textbf{disutility when choosing } (a_1,a_1) \textbf{ or } (a_2,a_2)}$$

$$k_1 \equiv \max\left\{\frac{\theta_2}{\frac{1}{2}-\theta_1} \frac{\theta_2'}{\frac{1}{2}-\theta_1'}\right\} \leqslant \frac{\pi_1}{\pi_2} \leqslant \min\left\{\frac{\frac{1}{2}-\theta_2}{\theta_1}, \frac{\frac{1}{2}-\theta_2'}{\theta_1'}\right\} \equiv k_2$$

valuation for this experiment

$$V(\mathrm{E}, \theta) = \theta_1 + \theta_2 - \theta_1\pi_1 - \theta_2\pi_2 \quad V(\mathrm{E}, \theta') = \theta_1' + \theta_2' - \theta_1'\pi_1 - \theta_2'\pi_2$$

Existence of Fully Informative Experiment

Lemma

The fully informative experiment $\overline{E}$ always lies in the optimal menu

- if not, replace the experiment selling to the one charging the highest fee as and charge her a higher fee
- by the existence of the fully informative experiment, the designer only designs the one for another type (two parameters)

$E_{\theta'}$	$a^1 = (a_1, a_1)$	$a^2 = (a_1, a_2)$	$a^3 = (a_2, a_2)$
ω_1	$1 - \pi_2$	π_2	0
ω_2	0	π_1	$1 - \pi_1$

- suppose allocate the fully informative one Ē to the high value type w.l.o.g

Designer's Problem

$$\max_{E, t_\theta, t_{\theta'}} \frac{1}{2}\left(t_\theta + t_{\theta'}\right)$$

s.t.

$$\begin{aligned}
&V(\overline{E}, \theta) - t_H \geqslant 0 && \text{(IR-}\theta\text{)} \\
&V(E, \theta') - t_{\theta'} \geqslant 0 && \text{(IR-}\theta'\text{)} \\
&V(\overline{E}, \theta) - t_\theta \geqslant V(E, \theta) - t_{\theta'} && \text{(IC-}\theta\text{)} \\
&V(E, \theta') - t_{\theta'} \geqslant V(\overline{E}, \theta') - t_\theta && \text{(IC-}\theta'\text{)} \\
&k_1 \leqslant \frac{\pi_1}{\pi_2} \leqslant k_2 && \text{(Responsiveness)}
\end{aligned}$$

This slide deck:

Four Selling Schemes

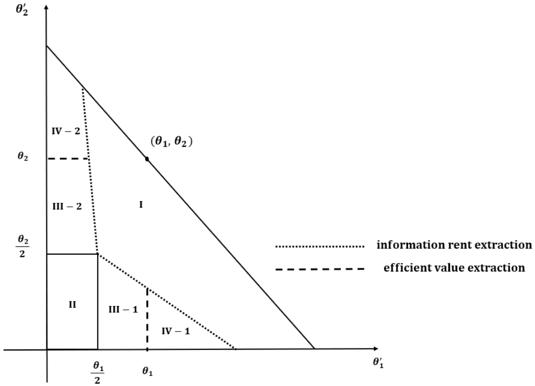


Figure: Optimal Selling Schemes

the high type always gets $\bar{E}$

the designer implements four selling schemes to low type

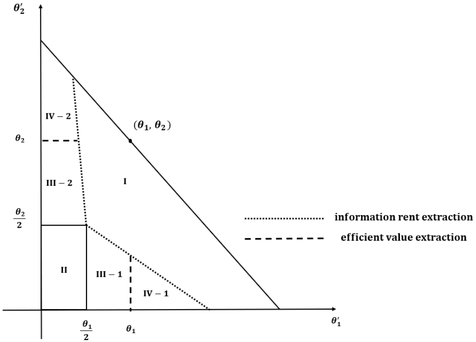
1. zone I: no discrimination
2. zone II: exclusive policy
3. zone III: partial discrimination
4. zone IV: perfect discrimination

Zone I and II

both horizontal preference are much smaller or not much smaller

$\Rightarrow$ approximately **one-dimensional** preference

Riley and Zeckhauser (1983)‘s classic **no-haggling** result applies



- zone I: no discrimination

  selling $\overline{E}$ to both types.
- zone II: exclusive policy

  only selling $\overline{E}$ to type-H

Zone III

when one much smaller while another not much smaller, trade-off emerges

both smaller horizontal values preference ⇒ info rent¿0

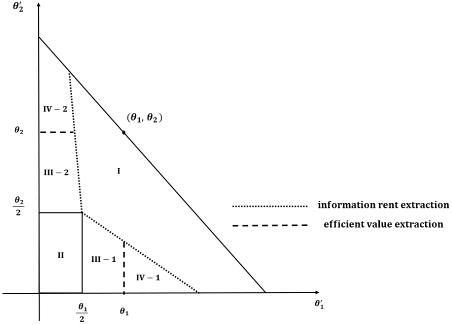


Figure: Optimal Selling Schemes

zone III: partial discrimination

selling $\bar{E}$ to type-H, and E^* to type-L

E^* is the same in this zone

(IR-L),(IC-H),(Responsiveness) is binding

- zone III-(1) (partial discrimination): $\frac{\theta_1}{2} < \theta_1' < \theta_1$ and $\theta_2' < \frac{\theta_2}{2}$

	$a^1 = (a_1, a_1)$	$a^2 = (a_1, a_2)$	$a^3 = (a_2, a_2)$
ω_1	0	1	0
ω_2	0	π_1^*	$1 - \pi_1^*$

- which is equivalent to sell the two "marginal" experiments to the two realizations

θ_1'	a_1	a_2
ω_1	1	0
ω_2	π_1^*	$1 - \pi_1^*$

θ_2'	a_1	a_2
ω_1	0	1
ω_2	0	1

- the dimension with relatively weak predictive power of type-L is sold null, and the strong one gets a partial informative experiment fully revealing one state the designer can partially differentiate the low type one to (i) fully extract the valuation of the low type and (ii) (incompletely) reduce the info rent of high type

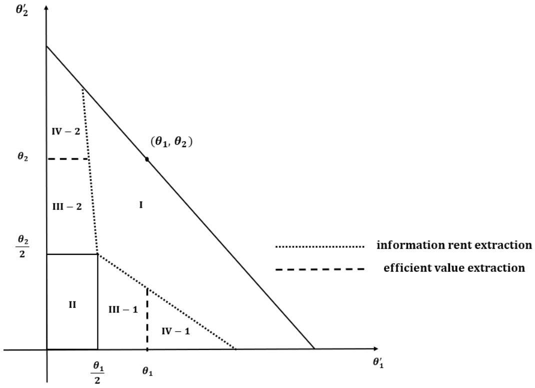


Figure: Optimal Selling Schemes

- the extraction is constrained because of the responsiveness
- E is a fixed experiment

  null to the dim with high accuracy

  a partially informative prediction fully revealing one state to another
- why fixed?

  θ-responsiveness is more stringent when one of the horizontal value is smaller while another larger

$\Rightarrow$ the information rent can be eliminated, but the elimination may cause huge loss in the extraction of low type valuation

multi-dimensions account for this result, different from incongruent preference order (Bergemann et al.,2018)

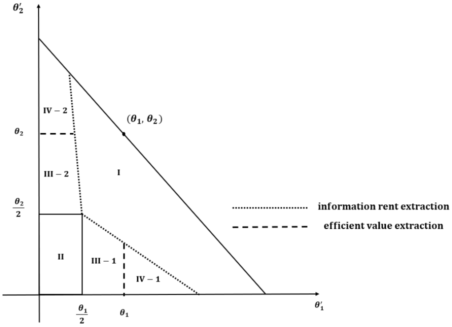


Figure: Optimal Selling Schemes

zone IV: perfect discrimination

selling $\bar{\mathrm{E}}$ to type-H, and E to type-L, E smoothly changes in this zone

(IR-L),(IC-H),(IR-H),(Responsiveness) is binding

- zone IV (perfect discrimination)

	$a^1 = (a_1, a_1)$	$a^2 = (a_1, a_2)$	$a^3 = (a_2, a_2)$
ω_1	$1 - \pi_2^*$	π_2^*	0
ω_2	0	π_1^*	$1 - \pi_1^*$

- which is equivalent to sell the two "marginal" experiments to the two realizations

θ_1'	a_1	a_2
ω_1	1	0
ω_2	π_1^*	$1 - \pi_1^*$

θ_2'	a_1	a_2
ω_1	$1 - \pi_2^*$	π_2^*
ω_2	0	1

- both dimesions of type-L get a partial informative experiment fully revealing one state

Comparisons of Zone III and IV

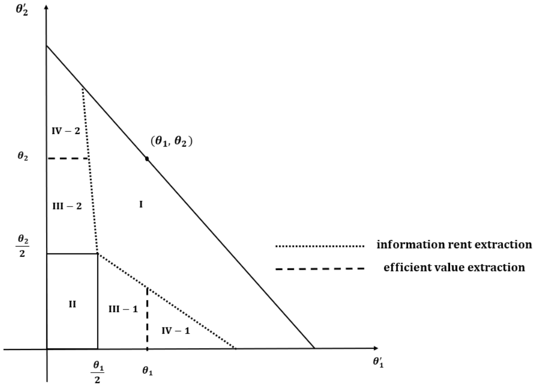


Figure: Optimal Selling Schemes

1. zone III

   indeed only one realization is sold
   a partial informative experiment
   fully revealing one state

   the experiment is the same

2. zone IV

   both realizations get a partial informative experiment fully revealing one state

   the experiment changes smoothly

Takeaway

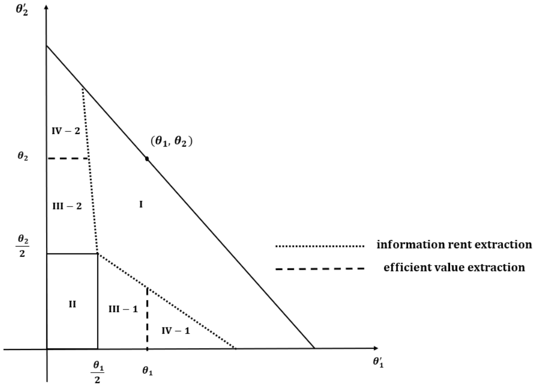


Figure: Optimal Selling Schemes

1. zone I: no discrimination

   (IR-L) is binding

2. zone II: exclusive policy

   (IR-H)

3. zone III: partial discrimination

   (IR-L),(IC-H),(Responsiveness)

4. zone IV: perfect discrimination

   (IR-L),(IC-H),(IR-H),(Responsiveness)

This slide deck:

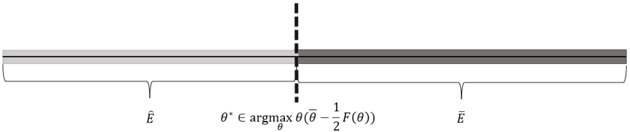
Main Result

one-dimensional private type (θ_1, θ_2) with $\theta_1 = Constant$, $\theta_2 \sim F(\theta)$ with support $[\underline{\theta}, \overline{\theta}] = [0, \frac{1}{2} - Constant]$

Theorem (The Optimality of the Cutoff Mechanism)

The optimal selling mechanism is

1. $(E_\theta, t_\theta) = (\overline{E}, t_{\theta^*})$ *for all for* $\theta \in \left[\theta^*, \overline{\theta}\right]$
2. $(E_\theta, t_\theta) = (\hat{E}, \hat{t})$ *for* $\theta \in [\underline{\theta}, \theta^*)$, *where* $\pi_1 = \frac{\theta^*}{\overline{\theta}}$, $\pi_2 = 1$
3. $\theta^* \in \arg\max_\theta \theta \left(\overline{\theta} - \frac{1}{2}F(\theta)\right)$



the optimal mechanism takes a simple and economically interpretable structure

1. the types are partitioned into two tiers according to their predictive power
2. in the first tier, a partially informative experiment $\hat{E}$, where the θ_2 dimension are sold a null while the θ_1 dimension is sold a partially informative experiment
3. in the second tier, a fully informative experiment $\overline{E}$
4. the threshold of the two tiers is determined similar to the monopolist pricing

The First Tier

$\hat{E}$	(a_1, a_1)	(a_1, a_2)	(a_2, a_2)
ω_1	0	1	0
ω_2	0	$\frac{\theta^*}{\overline{\theta}}$	$1 - \frac{\theta^*}{\overline{\theta}}$

The Second Tier

$\overline{E}$	(a_1, a_1)	(a_1, a_2)	(a_2, a_2)
ω_1	1	0	0
ω_2	0	0	1

Step 1: Analyzing IC and Responsiveness

- now the responsive zone $[\frac{\theta}{\overline{\theta}}, \frac{\frac{1}{2}-\theta}{\frac{1}{2}-\overline{\theta}}]$ is decreasing with θ
- in the optimal menu, $[\frac{\theta}{\overline{\theta}}, 1]$ for all θ

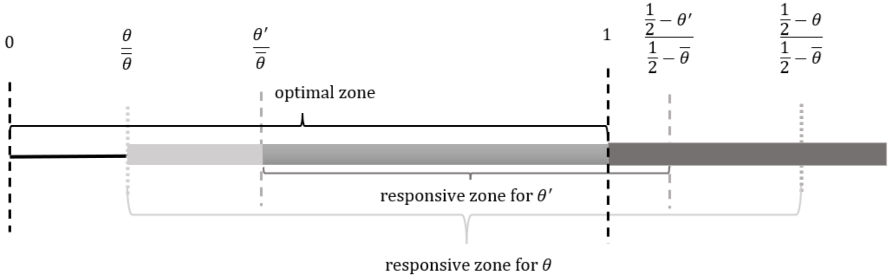


Figure: Responsive Zone

E_θ	(a_1, a_1)	(a_1, a_2)	(a_2, a_2)
ω_1	$1 - \pi_2$	π_2	0
ω_2	0	π_1	$1 - \pi_1$

E_θ	(a_1, a_1)	(a_2, a_2)
ω_1	1	0
ω_2	π_1	$1 - \pi_1$

E_θ	(a_1, a_1)	(a_1, a_2)	(a_2, a_2)
ω_1	$1 - \pi_2$	π_2	0
ω_2	0	π_1	$1 - \pi_1$

E_θ	(a_1, a_1)	(a_2, a_2)
ω_1	$1 - \pi_2$	π_2
ω_2	0	1

the responsive form of E_θ for θ'

$$V(E_\theta, \theta') - t_\theta = \max\left\{1 - \frac{1}{2}\pi_1, 1 - \theta'\pi_2 - m\pi_1, 1 - \frac{1}{2}\pi_2\right\} - u(\theta') - t_\theta$$

$$= 1 - u(\theta') - t_\theta - \frac{1}{2}\pi_1 - \min\{0, \theta'\pi_2 - \bar{\theta}\pi_1\}$$

$$= 1 - u(\theta') - \underbrace{t_\theta - \frac{1}{2}\pi_1}_{\text{transfer of } E_\theta} - \min\{0, \underbrace{\pi_2(\theta' - \lambda(\theta))}_{\pi_2}\}$$

Figure: Responsiveness and IC

Trade-off between IC and Responsiveness

In the optimal mechanism, there always exists $\theta^* \in \Theta$,

1. $E_\theta = \overline{E}$ if and only if $\theta \geq \theta^*$
2. for all $\theta < \theta^*$,
   1. the responsiveness of θ is not binding and
   2. there exists θ', E_θ is non-responsive for θ', and $\mathrm{IC}[\theta' \to \theta]$ is binding

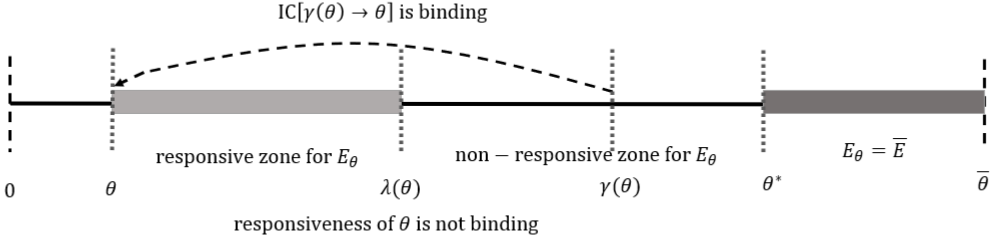


Figure: Trade-off between IC and Responsiveness

Step 2: Characterizing the Structure of the Optimal Menu

- define the function $\lambda(\theta)$ as below
  1. $\lambda(\theta) = \overline{\theta}\frac{\pi_1(\theta)}{\pi_2(\theta)}$ if $\pi_2(\theta) \neq 0$
  2. $\lambda(\theta) = \overline{\theta}$ if $\pi_2(\theta) = 0$. $\lambda(\theta) \in \left[\theta, \overline{\theta}\right]$
- the responsiveness of θ is binding if and only if $\lambda(\theta) = \theta$ or $\overline{\theta}$.
- the experiment E_θ is responsive for $\theta' \in [\theta, \lambda(\theta)]$, and pools the recommendation profile (a_1, a_2) with (a_1, a_1) for $\theta' \in \left[\lambda(\theta), \overline{\theta}\right]$
- define the single-valued correspondence $\gamma(\theta)$ as below:
  1. $\gamma(\theta) = a$ if $\theta = \lambda(\theta)$
  2. $\gamma(\theta) \in \{\theta' \mid \text{IC}\,[\theta' \to \theta] \text{ is binding}\}$ if $\theta < \lambda(\theta)$

Properties of $\lambda(\theta)$ and $\gamma(\theta)$

In the optimal menu,

1. $\lambda(\theta) \leq \lambda(\hat{\theta}) \leq \gamma(\theta)$ for $\hat{\theta} \in [\theta, \lambda(\theta)]$
2. $\pi_2(\theta) : \Theta \to [0, 1]$ is non-increasing
3. $\lambda(\theta) : \Theta \to \Theta$ is non-decreasing

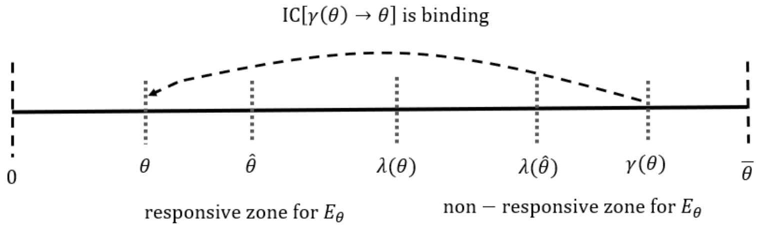


Figure: Properties of $\lambda(\theta)$ and $\gamma(\theta)$

Selling Tiers

$(\pi_1(\theta), \pi_2(\theta), t_\theta) = (\pi_1(\hat{\theta}), \pi_2(\hat{\theta}), t_{\hat{\theta}})$ iff $\lambda(\theta) = \lambda\left(\hat{\theta}\right)$

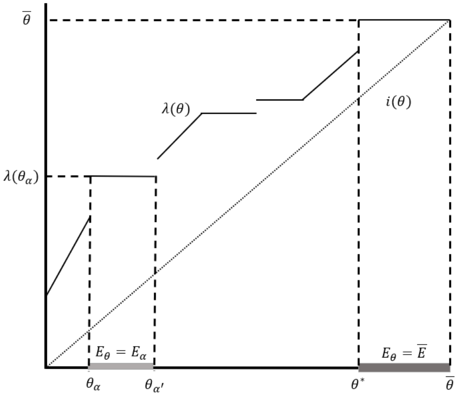


Figure: Tiered Pricing Mechanism

Step 3: Solving the Designer's Problem

Denote $V(\theta) = V(E_\theta, \theta) - t_\theta$ as the net value of type θ.

Equivalent Transformation of Constraints

In the optimal mechanism, the IC,IR and Responsiveness conditions are equivalent to

1. $\frac{1}{2}\pi_1(\theta) + t_\theta = t^*$ for all $\theta \in \Theta$, t^* is the associated tariff for all $\theta \in [\theta^*, \overline{\theta}]$
2. $V(\theta) = \int_0^\theta (1 - \pi_2(t))dt + V(\underline{\theta})$
3. $\mathrm{IR}[\underline{\theta}]$ holds
4. $\pi_2(\theta) : \Theta \to [0, 1]$ is non-increasing

Designers‘ Problem of Choosing Optimal π_2

$$\max_{pi_2(\theta)} \int_{\underline{\theta}}^{\overline{\theta}} \Phi(\theta) d\pi_2(\theta) \quad \text{s.t.} \quad \pi_2(\theta) : \Theta \to [0,1] \text{ is non-increasing}$$

where $\Phi(\theta) : \Theta \to \mathbb{R}$.

- designer‘s problem: maximizing a linear functional subject to monotonicity.
- by the infinite-dimensional extension of **Carathéodory’s theorem**, it follows that the optimal π_2 is an **extreme point of the set of non-increasing allocation rules**
- therefore π_2 is a step function with $\operatorname{im} \pi_2 \subseteq \{0, 1\}$

- with the form of π_2 and the tiered-pricing structure, we can further deduce the optimality of two-tiered pricing of tiers
- we can transform the designer's problem as the choice of the optimal threshold θ^*.

Optimal Threshold

$\theta^* \in \arg\max_\theta \theta \left(\overline{\theta} - \frac{1}{2}F(\theta)\right)$ is the optimal threshold of the tiers.

Discussion of the Main Result

- $2 \neq 1+1$, even with one fixed
  1. in the dimension with random predictive power, like π_2 in this case, the classic no-haggling result applies
  2. in the dimension with the same predictive power, there also exists differentiation in the allocation, which results from the interaction between the horizontal preference
- compared to one-dimensional screening, the multi-dimensional preference broadens the seller‘s scope of differentiation.
- the seller can focus on the extraction of other dimensions of the type when the valuation is low in some dimension