A Rank-Based Reward between a Principal and a Field of Agents: Application to Energy Savings

Clémence Alasseur, Erhan Bayraktar, Roxana Dumitrescu, Quentin Jacquet

A Rank-Based Reward between a Principal and a Field of Agents: Application to Energy Savings

Clémence Alasseur[1] Erhan Bayraktar[2] Roxana Dumitrescu[3] Quentin Jacquet[1,4]

[1] EDF Lab Saclay, Palaiseau, France

{clemence.alasseur,quentin.jacquet}@edf.fr

[2]Department of Mathematics, University of Michigan, USA

erhan@umich.edu

[3]Department of Mathematics, King's College London, United Kingdom

roxana.dumitrescu@kcl.ac.uk

[4]INRIA, CMAP, Ecole Polytechnique, Palaiseau, France

Abstract

We consider a problem where a Principal aims to design a reward function to a field of heterogeneous agents. In our setting, the agents compete with each other through their rank within the population in order to obtain the best reward. We first explicit the equilibrium for the mean-field game played by the agents, and then characterize the optimal reward in the homogeneous setting. For the general case of a heterogeneous population, we develop a numerical approach, which is then applied to the specific case study of the market of Energy Saving Certificates.

Keywords: Ranking games, Principal-Agent problem, Mean-field games, Energy savings

1 Introduction

1.1 Motivation

Energy retailer has incentives to generate energy consumption saving at the scale of its customer portfolio. For example in France, since 2006, power retailers – called *Obligés* – have a target of a certain amount of Energy Saving Certificates[1] to hold at a predetermined future date (usually 3 or 4 years). If they fail to obtain this number of certificates, then they face financial penalties. Certificates can be acquired either by certifying energy savings at the customer or by buying certificates on the market. If a retailer holds more certificates than its target at the end of the period, the surplus can be sold on the Energy Saving Certificates market. The pluri-annual energy savings goal is determined by the government, and is function of the cumulative discounted amount of energy saved (thanks to thermal renovation for instance).

E. Bayraktar is partially supported by the National Science Foundation under grant DMS-2106556 and by the Susan M. Smith chair.

[1] https://www.powernext.com/french-energy-saving-certificates

Similar mechanisms – called *White certificates* – have been implemented in several countries in Europe (Great Britain, Italy or Denmark). We refer to [1] for a recent report on the French case.

There is evidence from behavioral economy that energy consumption reductions can be motivated by providing a financial reward and/or information on social norms or comparison to customers, see e.g. see Alcott and Todd [2] or Dolan and Metcafe [3]. Especially, in [3], the authors find that social norms reduce consumption by around 6% (0.2 standard deviations). Secondly, they obtain that large financial rewards for targeted consumption reductions work very well in reducing consumption, with a 8% reduction (0.35 standard deviations) in energy consumption. For recent years, electricity providers are aware of this lever to make energy savings, and contracts offering bonus/rewards in compensation of reduction efforts appear, see e.g. the offers of "SimplyEnergy"[2], "Plüm énergie"[3] or "OhmConnect"[4]. The interest of this kind of solutions is reinforced in the current situation of gas and power shortage where many countries intend to diminish their global energy consumption[5] .

1.2 Related Works and Contributions

In this paper, we analyze games with large populations of agents (consumers) interacting through their empirical distribution. We focus on interactions based on the *rank* of each players: in our context, the rank measures the reduction effort of a consumer compared with the rest of the population, and is computed as the cumulative probability associated with his consumption. A rank $r \in [0, 1]$ indicates that the consumer is among the r percent of the population with the highest consumption reduction. The retailer will then design a rank-based reward function to incentive agents to reduce their consumption.

We directly study the *mean-field* limit of the model, considering a continuum of consumers. Mean-field games have been introduced simultaneously by Lasry and Lions [4, 5, 6] and Huang, Caines and Malhamé [7, 8]. In particular, they provide efficient ways to compute an approximation of the equilibrium for games with large number of players, which are rarely tractable. In the specific case of rank-based interactions, Bayraktar and Zhang [9] provide results of existence and uniqueness of the equilibrium for a general class of reward function, which enables the study of complex games [10, 11]. The design of reward function is then modeled as a *Principal-Agent* problem, see e.g. the work of Sannikov [12] in continuous-time settings. In such problems, the Principal (retailer) aims to design a monetary reward that is offered to the agent, depending on the quantity of work achieved by the latter. The additional difficulty in this context is the presence of a continuum of agents, and the interaction between each of them through the mean-field game. Such extension of the Principal-Agent problem have been considered by Carmona and Wang [13] and Elie, Mastrolia and Possamaï [14]. Shrivats, Firoozi and Jaimungal [15] apply this theory to the market of Renewable Energy Certificate (REC). As we do here, the overall population is clustered into a finite number of (infinite-size) independent sub-populations, each of them composed of indistinguishable agents. This heterogeneous context has been less regarded as it increases further the difficulty, both on analytic and numerical aspects. In contrast to our work, Shrivats et al. consider the interaction between a regulator and a field of providers, whereas we focus here on the the interaction between a provider and a field of consumers.

[2] `https://www.simplyenergy.com.au/residential/energy-efficiency/reduce-and-reward`
[3] `https://plum.fr/cagnotte/`
[4] `https://www.ohmconnect.com/`
[5] `https://www.politico.eu/article/eu-countries-save-energy-winter/`

For purely rank-based reward function and homogeneous population, the optimal behavior of the principal has been studied by Bayraktar and Zhang [11]. In particular, they provide analytic characterization of the (unique) equilibrium for the mean-field game at the lower level, and look at the Principal-Agent problem – also called in this context *tournament design* – for several principal objective functions.

In this paper, we first extend the theoretical results provided in [11] to rewards that not only depend on the rank, but also linearly depends on the process value (consumption). Figure 1 represents the different assumptions that are considered in [9], [11] and in this paper. We then propose a formulation of the Principal-Agent problem in the context of Energy Savings. We show that the problem can be converted to a problem of targeting an equilibrium distribution. We further provide an analytic solution for the homogeneous case and, in the more general setting, we develop an algorithm to optimize the shape of the reward. This algorithm is based on a black-box solver and we use here the solver CMA-ES [16]. We show the efficiency of the approach on homogeneous population, and present results on the general case. We then provide a detailed interpretation of the numerical results. In particular, getting back to the N-players game setting, we show that the use of rank-based reward function can induce a substantial consumption reduction for the agent. Finally, we consider several extensions suitable to our context. In particular, we focus on time-dependent costs of effort for the agents, reflecting the collective awareness of agents on the energy reduction's necessity. We are also able to provide some invariance results, which show that the use of more sophisticated reward (a function that jointly depends on the rank and the consumption of the agent) is, at the equilibrium, equivalent to a reward that belongs to the class we are studying.

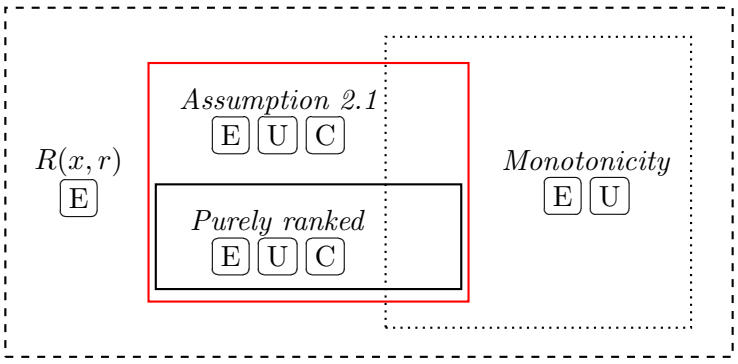


Figure 1: Classification of the results on mean-field equilibrium
The labels "E" (resp. "U", "C") stands for "Existence" (resp. "Uniqueness", "Characterization")

The rest of the paper is organized as follows: in Section 2, we first define the model and characterize the solution to the agents' problem. In Section 3 and 4, we propose a numerical approach to solve the problem in the heterogeneous setting, where no analytic solution is known. Finally, we tackle some extensions that naturally raise when considering applications to Energy Savings. The proofs of the main results are given in the appendix.

2 Model

We consider a *heterogeneous* population which can be divided beforehand into K clusters of indistinguishable consumers. Each cluster $k \in [K] : \{1, \dots, K\}$ represents a proportion ρ_k of the overall population.

Let $(\Omega, \mathbb{F}, \mathbb{P})$ be a complete filtered probability space, which supports a family of K independent Brownian motions $\{W_k\}_{1\leq k\leq K}$. Let $\mathbb{A}$ be the set of progressively measurable processes a satisfying the integrability condition $\mathbb{E}\int_0^T |a(s)|ds < \infty$. For $a_k \in \mathbb{A}$, we denote by $X_k(t)$ the forecasted energy consumption of a customer of cluster k (typically an household), forecast made at time t for consumption a time $T > t$. Assume that X_k^a follows the dynamics:

$$X_k^a(t) = X_k(0) + \int_0^t a_k(s)ds + \sigma_k \int_0^t dW_k(s), \quad X_k(0) = x_k^{\text{nom}}. \tag{1}$$

Here, the process a_k is then viewed as the consumer's effort to reduce his electricity consumption. Without any effort, customers are expected to have a *nominal* consumption of x_k^{nom}. In this model, we suppose that there is no common noise between the different clusters. Note that we do not explicitly impose bounds on the process X_k – typically non-negativity assumption, but this is naturally enforced by the cost of effort and the volatility parameter σ_k.

Let $\mathbb{R}\times[0,1] \ni (x,r) \mapsto R(x,r) \in \mathbb{R}$ be a continuous bounded function that is non-increasing in both arguments. The set of such reward functions is denoted by $\mathcal{R}_b$. For any probability measure μ on $\mathbb{R}$, we write $R_\mu(x) = R(x, F_\mu(x))$ where F_μ denotes the cumulative distribution function on μ. When $R(x,r)$ is independent of x, we say that the reward is *purely ranked-based*. In the sequel, we will consider the following decomposition assumption:

Assumption 2.1. *The reward R has the form*

$$R\,(x,r) = B(r) - px \ , \tag{2}$$

where $p \in \mathbb{R}$ and $B \in \mathcal{R}_b^r$ with $\mathcal{R}_b^r$ the set of purely ranked-based (decreasing) bounded functions. We then call R the total reward *and its rank-dependent part B the* additional reward*.*

In the energy context, the second member $-px$ represents the natural incentive to reduce the consumption, coming from the price p to consume one unit of energy, whereas the first member is the additional financial reward offered to consumers based on their rank. In the N-players game setting, each cluster k contains N_k players, and the ranking of a player i from this cluster, consuming $X_k^i(T)$, is measured by the fraction $\frac{1}{N_k}\sum_{j=1}^{N_k} \mathbb{1}_{X_k^j(T)\leq X_k^i(T)}$ of players having equal or best performance (so that the worst performer (the highest consumption) has rank one and the top performer has rank $1/N_k$).

Assumption 2.2. *(i) Each cluster is* independent*: the rank of an agent of cluster $k \in [K]$ is only determined by the distribution of the cluster k.*

(ii) The same total reward is proposed to each cluster, i.e., the price of electricity p and the additional reward B are common to all the clusters.

Assumption 2.2 means that the clusters evolve separately, but are linked through a common reward function. We finally define f_k^{nom} as

$$f_k^{\text{nom}}(x) := \varphi\left(x\,;\, x_k^{\text{nom}}, \sigma_k\sqrt{T}\right) \ , \tag{3}$$

where $\varphi(\,\cdot\,;\mu,\sigma)$ is the pdf for $\mathcal{N}(\mu,\sigma)$. The function $f_k^{\rm nom}$ corresponds to p.d.f. of $X_k^a(T)$ under a zero effort (a_k is a constant process equals to 0), and is called *nominal* pdf.

2.1 Mean-field game between consumers

In all this section, we suppose the reward $R(x,r)$ known, and Assumption 2.2 verified. Let us fix a cluster $k \in [K]$, as there is no interaction between clusters.

An agent of k is able to produces the effort a_k to reduce its consumption, but has to pay the quadratic cost $c_k\, a_k^2(t)$ with $c_k > 0$ a given positive constant. In our context, this cost corresponds to the purchase of new equipment, more efficient than old one (new heating installation, isolation, ...). In exchange, the consumer receives the monetary reward $B(r)$, depending on his rank $r = F_{\mu_k}(x)$ within the sub-population, where μ_k is the k-subpopulation's distribution. Its objective is then:

$$V_k(R,\mu_k) := \sup_a \mathbb{E}\left[R_{\mu_k}(X_k^a(T)) - \int_0^T c_k a_k^2(t)dt\right] \ . \qquad (P^{\rm cons})$$

The quantity $V_k(R,\mu_k)$ is then the *optimal utility* of an agent of class k, knowing the provider's reward and the population distribution.

2.1.1 Previous results

We recall here some results which will be used throughout the paper. The first theorem gives the explicit solution of the agent's best response to a population distribution μ_k:

Theorem 2.1 ([11],Proposition 2.1)**.** *Given $R \in \mathcal{R}$ and $\tilde{\mu}_k \in \mathcal{P}(\mathbb{R})$, let*

$$\beta_k(\tilde{\mu}) = \int_{\mathbb{R}} f_k^{\rm nom}(x) \exp\left(\frac{R_{\tilde{\mu}}(x)}{2c_k\sigma_k^2}\right) dx \quad (<\infty) \ . \qquad (4)$$

Then, the optimal terminal distribution μ_k^* of the player of cluster k has p.d.f.

$$f_{\mu_k^*}(x) = \frac{1}{\beta(\tilde{\mu}_k)} f_k^{\rm nom}(x) \exp\left(\frac{R_{\tilde{\mu}_k}(x)}{2c_k\sigma_k^2}\right) \ , \qquad (5)$$

and the optimal value is then $V_k(R,\tilde{\mu}_k) = 2c_k\sigma_k^2 \ln \beta_k(\tilde{\mu}_k)$.

This result is obtained using the Schrödinger bridge approach, see [17] for connections with optimal transport theory. We give below the definition of an equilibrium.

Definition 2.2. *We say that $\mu_k \in \mathcal{P}(R)$ is an* equilibrium *(terminal distribution) if it is a fixed-point of the mapping $\Phi_k : \mu_k \mapsto \mu_k^*$, with μ_k^* given by the solution of the equation* (5)*.*

This equilibrium has been first studied using abstract tools to obtain existence in a very general setting. We recall here the result:

Theorem 2.3 ([9], Theorem 3.2)**.** *The mapping Φ_k has a fixed-point, i.e., there exists at least one equilibrium.*

We give below a characterization of an equilibrium distribution. This characterization gives analytic expression of the quantile when the reward is purely ranked-based:

Theorem 2.4 ([11], Theorem 3.2)**.** *Given $R \in \mathcal{R}_b$, the distribution $\mu_k \in \mathcal{P}(\mathbb{R})$ is an equilibrium terminal distribution for cluster k if and only if its quantile function q_{μ_k} satisfies*

$$N\left(\frac{q_{\mu_k}(r) - x_k^{\text{nom}}}{\sigma_k\sqrt{T}}\right) = \frac{\int_0^r \exp\left(-\frac{R_{\mu_k}(q_{\mu_k}(z))}{2c_k\sigma_k^2}\right) dz}{\int_0^1 \exp\left(-\frac{R_{\mu_k}(q_{\mu_k}(z))}{2c_k\sigma_k^2}\right) dz} \ ,$$

where N is the standard normal c.d.f. In the specific case of a purely ranked-based reward, we obtain that the equilibrium ν_k is unique and the quantile is given by

$$q_{\nu_k}(r) = x_k^{\text{nom}} + \sigma_k\sqrt{T}N^{-1}\left(\frac{\int_0^r \exp\left(-\frac{B(z)}{2c_k\sigma_k^2}\right) dz}{\int_0^1 \exp\left(-\frac{B(z)}{2c_k\sigma_k^2}\right) dz}\right) \ . \tag{6}$$

The mean consumption at the equilibrium is then $m_{\mu_k} = \int_0^1 q_{\mu_k}(r)dr$.

2.1.2 New results

We provide here an extension of Theorem 2.4 to the case when the reward map R can also depend on x (note that all results given in [11] are provided in the case of purely ranked rewards). The next theorem makes explicit the equilibrium for this more general form of reward R, that naturally arises in our case study.

Theorem 2.5. *Suppose the reward is of the form defined in Assumption 2.1. Then, the equilibrium μ_k is unique, and it satisfies*

$$q_{\mu_k}(r) = q_{\nu_k}(r) - \frac{pT}{2c_k} \ , \tag{7}$$

where ν_k is the (unique) equilibrium distribution for the specific case $p = 0$ (purely ranked-based reward), defined in (6)*.*

Theorem 2.5 shows that the add of a linear part in "x" acts as a shift on the probability density function. Our uniqueness result of the equilibrium μ generalizes the one obtained in [11], under the additional assumptions that the map $r \mapsto R(x, r)$ is convex and $r \mapsto R_x(x, r)$ is non decreasing. We prove that, with the special structure of the reward (linear part in x), no convexity condition is required for purely-ranked part B.

Corollary 2.6. *In the particular case when the provider does not offer additional reward B, i.e. the total reward R only has a linear dependence on x, we have the following characterization: the equilibrium follows the normal distribution $\mathcal{N}\left(x_k^{\text{pi}}, \sigma_k\sqrt{T}\right)$, where $x_k^{\text{pi}} = x_k^{\text{nom}} - \frac{pT}{2c_k}$ is the consumption under the natural incentive associated with the price p. Moreover, the optimal consumer's utility is*

$$V_k(R, \mu_k) = V_k^{\text{pi}} := -px_k^{\text{nom}} + \frac{p^2T}{4c_k} \ . \tag{8}$$

Corollary 2.6 shows that the electricity price induces a natural incentive to reduce the consumption. Therefore, without supplementary reward, the consumer already makes an effort.

2.2 Retailer's problem

In this section, we suppose that Assumption 2.1 is satisfied. Therefore, the equilibrium distribution is unique and is defined by (6)*. For a mean-field equilibrium* $(\mu_k)_{k\in[K]}$*, the mean consumption of the overall population is then* $m = \sum_{k\in[K]} \rho_k m_{\mu_k}$.

We denote the mapping from the pure-ranked part of the reward functions and the corresponding equilibrium distribution by

$$\epsilon_k : \mathcal{R}_b^r \mapsto \mathcal{P}(\mathbb{R}) \ .$$

We also define $\zeta_{k,\mu} := f_\mu / f_k^{\text{nom}}$ the normalized distribution. The problem of the retailer can then be written as

$$\max_{B\in\mathcal{R}_b^r} \left\{ s\left(\sum_{k\in[K]} \rho_k m_{\mu_k}\right) + (p - c_r) \sum_{k\in[K]} \rho_k m_{\mu_k} - \int_0^1 B(r)dr \;\middle|\; \begin{array}{l} \mu_k = \epsilon_k(B) \\ V_k(B) \geq V_k^{\text{pi}} \end{array} \right\} \qquad (P^{\text{ret}})$$

where

- $\diamond$ $s(\cdot)$ denotes the valuation of the energy savings,
- $\diamond$ c_r denotes the cost of production of energy,
- $\diamond$ and m_μ is the mean consumption at the equilibrium μ.

The constraint on the utility ensures that consumers play the game, as it procures a better utility than without additional reward. In the sequel, we denote by $g(\cdot)$ the function defined as

$$g : m \mapsto s(m) - c_r m \ .$$

Note that in [11] the results are obtained in the case of linear dependence of the gain functional with respect to the equilibrium distribution. We extend here the results to the case of concave nonlinear dependence. We make the following assumption on the function s.

Assumption 2.3. *The function* $s : \mathbb{R} \to \mathbb{R}$ *is supposed to be decreasing, concave and differentiable and such that* $s' : \mathbb{R} \to \mathbb{R}$ *is bounded.*

2.2.1 Homogeneous population

We consider in this section the specific case where there is a unique cluster of customers. Therefore, we omit the dependence in k. Using Theorem A.2 (developed in Appendix), the problem (P^{ret}) can be

reformulated as a constrained maximization problem on the distribution space:

$$(P^{\text{ret}}) \iff \begin{cases} \max\limits_{\mu} & g\left(\int_{-\infty}^{+\infty} y f_\mu(y)dy\right) - V^{\text{pi}} - 2c\sigma^2 \int_{-\infty}^{+\infty} \ln\left(\frac{f_\mu(y)}{f^{\text{nom}}(y)}\right) f_\mu(y)dy \\ \text{s.t.} & \int_{-\infty}^{+\infty} f_\mu(y)dy = 1 \\ & y \mapsto \ln\left(\frac{f_\mu(y)}{f^{\text{nom}}(y)}\right) + \frac{p}{2c\sigma^2} y \text{ bounded and decreasing} \end{cases} \tag{9}$$

Knowing the optimal equilibrium distribution μ^* of (9), the corresponding reward is

$$B^*(r) = V^{\text{pi}} + 2c\sigma^2 \ln\left(\zeta_{\mu^*}(q_{\mu^*}(r))\right) + p q_{\mu^*}(r) \ .$$

One can verify that (μ^*, B^*) satisfies the characterization of an equilibrium given in Theorem 2.4. Note also that, by construction, the utility condition $V(B) \geq V^{\text{pi}}$ is automatically verified.

Consider the following *relaxed optimization problem*

$$\begin{array}{ll} \max\limits_{\mu} & g\left(\int_{-\infty}^{+\infty} y f_\mu(y)dy\right) - V^{\text{pi}} - 2c\sigma^2 \int_{-\infty}^{+\infty} \ln\left(\frac{f_\mu(y)}{f^{\text{nom}}(y)}\right) f_\mu(y)dy \\ \text{s.t.} & \int_{-\infty}^{+\infty} f_\mu(y)dy = 1 \end{array} \tag{$\widetilde{P}^{\text{ret}}$}$$

The discussion about the relation between the initial problem (9) and the relaxed one ($\widetilde{P}^{\text{ret}}$) is provided further. The optimal solution of this relaxed problem is then characterized by the following lemma:

Lemma 2.7. *Let Assumption 2.3 holds and let $\delta : \mathbb{R} \to \mathbb{R}$ be a function given by*

$$\delta(m) = p - c_r + s'(m) \ .$$

The optimal distribution μ^* satisfies the following equation:

$$\begin{aligned} f_\mu(y) &= \frac{1}{\alpha(\mu)} f^{\text{nom}}(y) \exp\left(y \frac{g'(m_\mu)}{2c\sigma^2}\right) \\ \alpha(\mu) &= \int_{-\infty}^{+\infty} f^{\text{nom}}(y) \exp\left(y \frac{g'(m_\mu)}{2c\sigma^2}\right) dy \end{aligned} \tag{10}$$

This result is obtained using Karush-Kuhn-Tucker conditions, which are sufficient for this convex problem, see Appendix A. In contrast with [11], the optimal distribution is not explicit anymore due to the more general gain function $g(\cdot)$. The optimal distribution is implicitly known through a fixed-point equation, which is given in the following theorem.

Theorem 2.8. *Let Assumption 2.3 holds. Then, the distribution $\mu^* \hookrightarrow \mathcal{N}(m^*, \sigma\sqrt{T})$ where m^* satisfies the fixed-point equation*

$$m - x^{\text{pi}} = \frac{T}{2c}\delta(m) \ , \tag{11}$$

is optimal for the problem $(\widetilde{P}^{\mathrm{ret}})$*. Moreover, the associated reward* B^* *is*

$$B^*(r) = \frac{c}{T}\left[(x^{\mathrm{pi}})^2 - (m^*)^2\right] + q_{\mu^*}(r)\delta(m^*) \ , \tag{12}$$

and the associated retailer gain is

$$\pi = s(m^*) - m^* s'(m^*) + \left(\frac{m^* + x^{\mathrm{pi}}}{2}\right)\delta(m^*) \ . \tag{13}$$

The function $\delta(\cdot)$ could be interpreted as the *reduction desire* of the provider. The decreasing condition on B is then equivalent to $\delta(m) < 0$. It shows that the additional reward is decreasing iff the saving function s must be sufficiently important compared to the retailer marginal benefit $p - c_r$. In this case, an optimal solution for the relaxed optimization problem $(\widetilde{P}^{\mathrm{ret}})$ is an ε-optimal solution for the initial optimization problem (9). The proof follows by similar arguments as the ones provided in Theorem 5.4. in [11], but the ε is different, also depending on the bounds of g'.

Remark. For quadratic function $s : m \mapsto \alpha_2 m^2 + \alpha_1 m + \alpha_0$, the fixed point of (11) is analytically known:

$$m^* = \frac{x^{\mathrm{nom}} + \frac{(\alpha_1 - c_r)T}{2c}}{1 - \frac{\alpha_2 T}{c}} \ .$$

2.3 Heterogeneous population

We now consider that there exists a finite number of clusters $k > 1$. The transformation which leads to (9) still applies, but there is an additional constraint given by the equality of the reward for all the clusters. The constrained optimization problem reads as follows

$$\begin{aligned}
\max_{\mu_1,\ldots,\mu_K} \quad & g\left(\sum_{k\in[K]} \rho_k \int_{-\infty}^{+\infty} y f_{\mu_k}(y)dy\right) - \sum_{k\in[K]} \rho_k \left[V_k^{\mathrm{pi}} + 2c_k\sigma_k^2 \int_{-\infty}^{+\infty} \ln\left(\frac{f_{\mu_k}(y)}{f_k^{\mathrm{nom}}(y)}\right) f_{\mu_k}(y)dy\right] \\
\text{s.t.} \quad & \int_{-\infty}^{+\infty} f_{\mu_k}(y)dy = 1, \quad \forall k \in [K] \\
& y \mapsto \ln\left(\frac{f_{\mu_k}(y)}{f_k^{\mathrm{nom}}(y)}\right) + \frac{p}{2c_k\sigma_k^2} y \quad \text{bounded and decreasing} \\
& V_1^{\mathrm{pi}} + 2c_1\sigma_1^2 \ln\left(\frac{f_{\mu_1}(y)}{f_1^{\mathrm{nom}}(y)}\right) = \ldots = V_K^{\mathrm{pi}} + 2c_K\sigma_K^2 \ln\left(\frac{f_{\mu_K}(y)}{f_K^{\mathrm{nom}}(y)}\right), \quad \forall y \in \mathbb{R}
\end{aligned} \tag{14}$$

Even if we relax the problem as in the homogeneous case by removing the constraint of boundedness and monotonicity, the KKT conditions are still not sufficient to ensure optimality, due to the nonlinearity of the last set of equality constraints. In fact, this set of constraints represents the coupling condition of having the same reward function for all the cluster (Assumption 2.2). The next section is dedicated to numerical algorithm which can be used for heterogeneous population.

3 Reward optimization

3.1 Restriction to piecewise linear reward

For a given $N \in \mathbb{N}$, we denote by Σ_N a range of rank values such that $\Sigma_N := \{0 = \eta_1 < \eta_2 < \ldots < \eta_N = 1\}$. Let $M \in \mathbb{R}_+$, then we define the class of bounded piecewise linear rewards adapted to Σ_N as

$$\widehat{\mathcal{R}}_M^N := \left\{ r \in [0,1] \mapsto \sum_{i=1}^{N-1} \mathbb{1}_{r \in [\eta_i, \eta_{i+1}[} \left[b_i + \frac{b_{i+1} - b_i}{\eta_{i+1} - \eta_i}(r - \eta_i) \right] \quad \middle| \quad \begin{array}{l} b \in [-M, M]^N \\ b_1 \geq \ldots \geq b_N \end{array} \right\} .$$

The reward function obtained as a linear interpolation of a non-increasing vector b is denoted by $R_M^N(b)$. For this special class of reward, the computation of some integrals can be simplified. The integral that appears in the equilibrium characterization (6) becomes

$$\int_0^1 \exp\left(-\frac{B(z)}{2c_k\sigma_k^2}\right) dz = 2c\sigma^2 \sum_{i=1}^{N-1} \frac{\eta_{i+1} - \eta_i}{b_{i+1} - b_i} \left[\exp\left(-\frac{b_{i+1}}{2c\sigma^2}\right) - \exp\left(-\frac{b_i}{2c\sigma^2}\right) \right]$$

and the integral of the reward simplifies into

$$\int_0^r B(z)dz = \sum_{i=1}^{N-1} (\eta_{i+1} - \eta_i) \left(\frac{b_{i+1} + b_i}{2} \right) .$$

We define the following transformation:

$$\begin{array}{rccc} \phi_M^N : &]\text{-}1,1]^N & \to \widehat{\mathcal{R}}_M^N & \\ & z & \mapsto R_M^N(b) & \end{array} \quad \text{where} \quad \begin{cases} b_1 = Mz_1 \\ b_i = \frac{1}{2}(b_{i-1} - M) + \frac{1}{2}(b_{i-1} + M)z_i, \ i > 1 \end{cases} \tag{15}$$

Note that, for any $M \in \mathbb{R}_+$ and $N \in \mathbb{N}$, the function ϕ_M^N is invertible and $\left(\phi_M^N\right)^{-1}$ is defined as:

$$\left(\phi_M^N\right)^{-1}(b) = \begin{cases} z_1 = \dfrac{1}{M} b_1 \\ z_i = \dfrac{2b_i - b_{i-1} + M}{b_{i-1} + M}, \ i > 1 \end{cases}$$

As the map ϕ_M^N is invertible, exploring the space $[-1,1]^N$ is sufficient to cover the space $\widehat{\mathcal{R}}_M^N$. Optimizing on $\widehat{\mathcal{R}}_M^N$ is then equivalent to optimize on $[-1,1]^N$ via the transformation ϕ_M^N. As an example, Figure 2 displays a reward function in both spaces.

3.2 Description of the algorithm

We denote by $\pi_\lambda : \mathcal{R}_b^r \to \mathbb{R}$ the Lagrangian function of (P^{ret}), defined as

$$\pi_\lambda(B) := g\left(\sum_{k \in [K]} \rho_k m_{\mu_k} \right) + p \sum_{k \in [K]} \rho_k m_{\mu_k} - \int_0^1 B(r)dr + \lambda \sum_{k \in [K]} \rho_k \max(V_k^{\text{pi}} - V_k(B)), \tag{16}$$

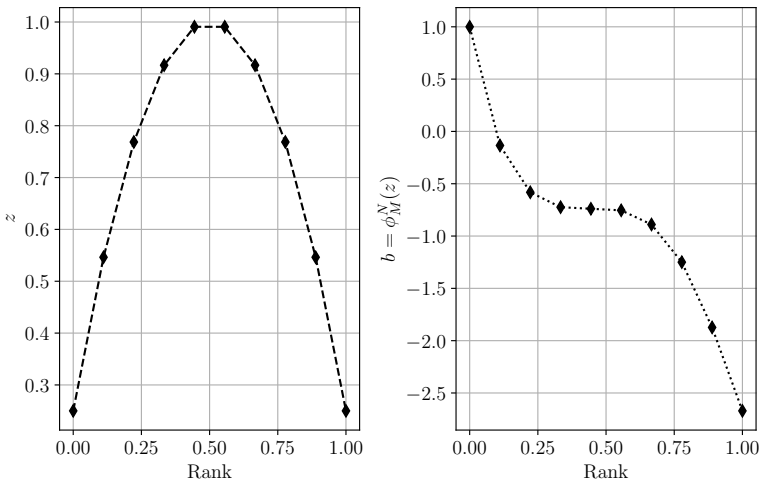


Figure 2: Example of transformation using function ϕ_M for $M = 4$ and $N = 10$

where the equilibrium distribution $\mu_k = \epsilon_k(B)$ is computed with closed-form formula given in Theorem 2.5. For fixed Lagrangian multiplier λ, this function constitutes a relaxed version of the initial problem (P^{ret}). Algorithm 1 aims to maximize the function π_λ. To this end, we do not directly search the optimal reward but, as described previously, we use the invertible map ϕ_N^M to search in the space $[-1, 1]^N$. The search is then achieved by an off-the-shelf solver (it can be evolutionary methods for instance as the evaluation of π_λ is relatively cheap).

Algorithm 1 Optimization of the reward

Require:

- M, N, λ, Σ_N,
- solver Π,
- initial point z^0,

Construct Θ as

$$\Theta : z \in]\text{-}1, 1]^N \mapsto (\pi_\lambda \circ R_M^N)(z) \tag{17}$$

Apply Π to maximize Θ (starting from z^0) and get the final state z^Π.
return $B^\Pi = (R_M^N)(z^\Pi)$.

Remark. The reward function found by Algorithm 1 is bounded and decreasing, but might violate the utility constraint $V_k(B) \geq V_k^{\text{pi}}$ for small penalization values of λ. Note that if the optimizer for the discrete problem on a sufficiently precise grid is a global optimizer, then we get an ε-solution of the initial problem.

4 Numerical results

We implement Algorithm 1, using CMA-ES [16] as optimization solver through the C++ interface [18]. We use 20 discretization points / optimization variables. We always take $z^0 \equiv 1$ as initial guess. This initial guess has the main advantage to satisfy the utility constraint (otherwise the problem is infeasible). The σ parameter of CMA is fixed to 0.05. The numerical results were obtained on a laptop i7-1065G7

CPU@1.30GHz.

We use in this toy model the following parameters:

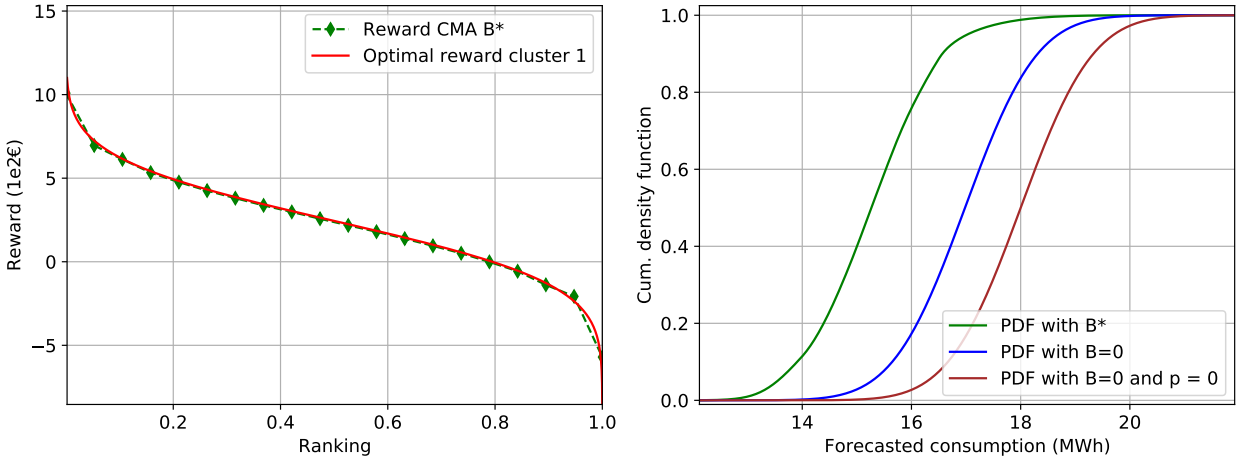
Parameter	Segment 1	Segment 2	Unit
T	3		years
p	0.17		€/kWh
c_r	0.15		€/kWh
$X(0)$	18	12	MWh
σ	0.6	0.3	MWh
c	2.5	5	€ $[\text{MWh}]^{-2}$ $[\text{years}]^2$
s	$m \mapsto 0.1m^2$		€
ρ	0.5	0.5	-

Table 1: Parameters of the instance

The price which is considered here corresponds to the price of electricity in French regulated offers, and the initial forecasted consumption is around the french mean consumption over 3 years[6] Note that this duration corresponds to the canonical duration of a period in the Energy Saving Certificates market.

The two segments we design differ by their consumption, their volatility and their cost of effort. Indeed, as the second cluster already consumes less than the first one, he has more difficulty to reduce the energy consumed, as it may be already reserved to necessary usages.

4.1 Homogeneous population



(a) Analytic optimal reward in red, compared to the reward function found by CMA

(b) Comparison of the three CDF: nominal, price incentive and with the optimal reward

Figure 3: Optimization in the homogeneous case

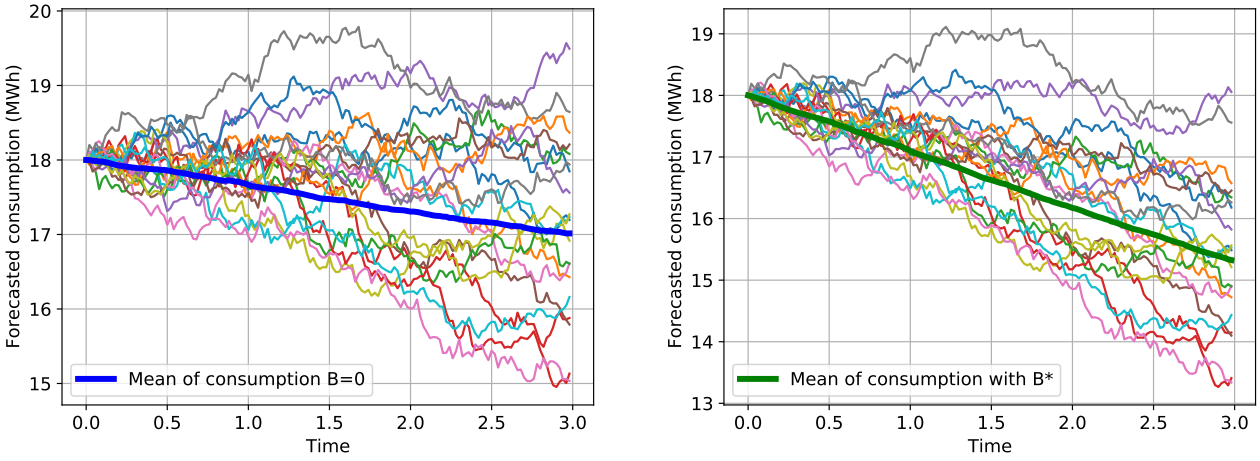
Figure 3a shows the reward found by Algorithm 1. As a comparison, the optimal reward (computed with (12)) is also drawn. The two reward are very close, meaning that the algorithm has converged to the global optimum. Figure 3b depicts three cumulative distribution function. The nominal one has a mean value of 18, the one with the price as unique incentive has a mean value around 17, and the cdf

[6] `https://www.cre.fr/en/Electricity/retail-electricity-market`

obtained with the optimal reward has a mean value around 15. As expected, the additional reward has induced a higher effort in the population, and so a higher energy reduction.

Note that the consumers has an average bill of $px^{\text{pi}} = 2\,890$€ (for $T = 3$ years) and the additional reward takes values in $[-500, 1\,000]$€. Therefore, the additional reward is, in this toy model, important in comparison to the original bill.

The N-players game. We now illustrate numerically the behavior of several agents at the ε-Nash equilibrium, with the optimal reward, which is obtained from the limit equilibrium (see Theorem 3.8 in [9]). The simulation of the trajectories is done using a Euler-Maruyama scheme. We refer to [19] for details on the discretization, as for convergence rates.



(a) Trajectories without additional reward

(b) Trajectories with optimal control from mean-field approximation

Figure 4: Trajectories for 20 consumers (homogeneous case)

Figure 4 displays the evolution of the forecasted consumption. At time $t = T = 3$, we then have the true consumption on the horizon of each agent. In the graph, the values goes from 15 to 19.5. In both subfigures, the solid lines represents the mean consumption over the large (finite) set of consumers (typically, we average on 1000 curves). Note that the hazard realization is the same in the two graphs, so that each realization can be compared. Once again, we observe that the consumption reduction is greater with the additional reward.

For completeness, Figure 5 shows the optimal cost of effort of each agent, i.e., $t \mapsto \int_0^t ca^2(s)ds$. We observe that each agent is facing almost the same cost, but three curves seems to diverge. In fact, they corresponds to three agents that have undergone the most extreme Brownian realization. The pink agent has to maintain a high effort to keep the best ranking whereas for the grey and purple agents, the effort they need to do be better ranked is so important (as they are already badly ranked) that they prefer to make a small effort and stay in the same position.

4.2 Heterogeneous population

Figure 6 displays the optimal reward found bu the algorithm. In addition, we also draw the optimal reward that we would have obtained if the population was only composed of the first (resp. second)

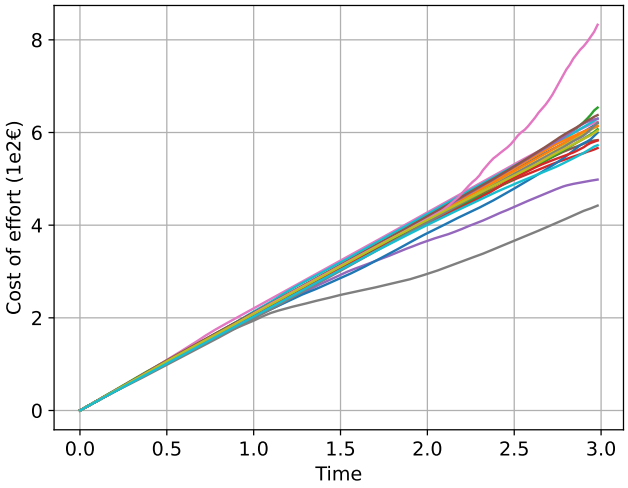


Figure 5: Optimal cost of effort for 20 consumers (homogeneous case)

cluster. We cannot guarantee the optimality of this reward, but the solution seems to be very reasonable, as it is a compromise of the two (homogeneous) optima. One can observe that the consumption reduction is less important with this reward (Figure 6b) than in the previous case (Figure 3b). In Fig. 6c we display the retailer's objective value for each reward (B_1 resp. B_2 denotes the optimal reward associated with the objective criteria π_1 resp. π_2, for the subpopulation 1 resp. 2). On both cluster, we can observe the gap coming from the heterogeneity of the population: the retailer cannot perfectly design a reward to each cluster. We also show the value of the utility constraint. We observe that the inequality constraint in not saturated when using B^*. This means that the first cluster has a strictly greater utility with this reward compared to the case without additional reward.

The N-players game. As for the homogeneous case, we illustrate numerically the behaviour of the consumption of several agents in Figure 7.

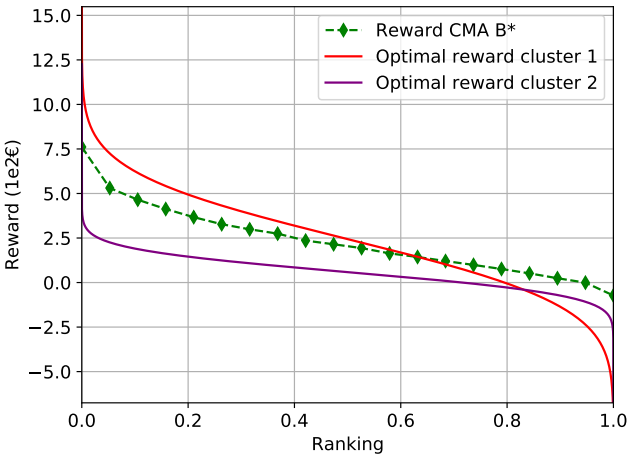
5 Extensions

In this section, we propose several extensions to more general settings. In this context, the equilibrium is not analytically known anymore. Therefore, we propose a numerical algorithm based on a fixed point method to compute an equilibrium.

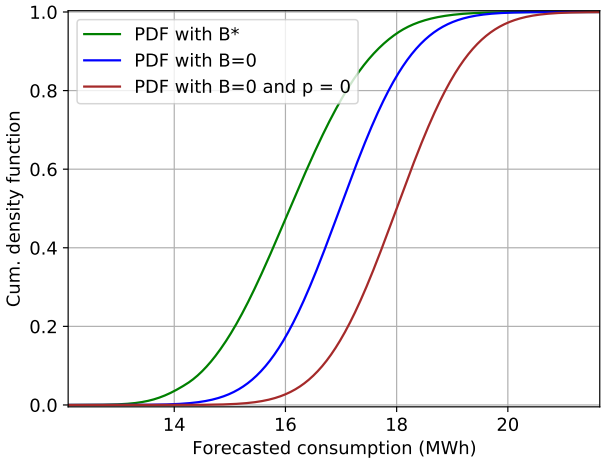
We denote by $W_1(f_1, f_2)$ the 1-Wasserstein metric for distribution $f_1, f_2 \in \mathcal{P}_1(\mathbb{R}) = \{\mu \in \mathcal{P}(\mathbb{R}) : \int_{\mathbb{R}} |x| d\mu(x) < \infty\}$. We provide an algorithm based on the *best response* characterization given by equation (5). The algorithm is detailed in Algorithm 2. Several sequences of damping coefficients have been tested on this problem:

- ◇ Iteration-independent damping $l_i = 1/2$,
- ◇ Decreasing damping $l_i = \left(\frac{1}{i+1}\right)^p$, $p \in \mathcal{N}$.

The convergence with the damping $l_i = 1/(i+1)$ is slow but guaranteed (as the sequence converges to zero) whereas the other damping may not converge.



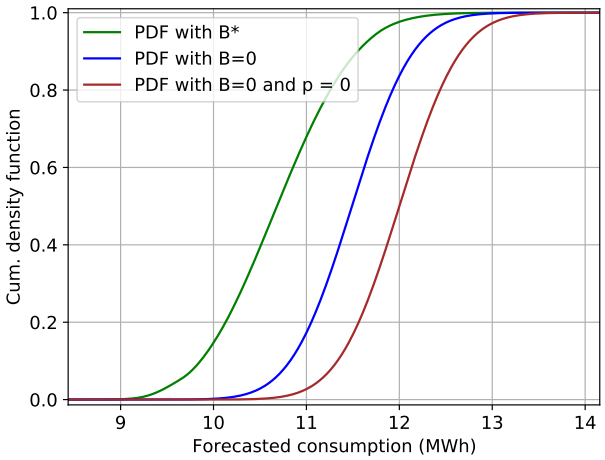
(a) Red and purple rewards are the optimal reward in the homogeneous case. The reward function found by CMA is displayed in green.



(b) Comparison of the three CDF (first cluster): nominal, price incentive and with the optimal reward

		B^*		
		$k=1$	$k=2$	$k=1,2$
π	$k=1$	9.76	-	7.42
	$k=2$	-	2.95	2.82
$V_k - V_k^{\text{pi}}$	$k=1$	0	-	1.67
	$k=2$	-	0	0

(c) Objectives and utility constraints for the two homogeneous cases and the heterogeneous one



(d) Comparison of the three CDF (second cluster): nominal, price incentive and with the optimal reward

Figure 6: Optimization in the heterogeneous case

General reward $R(x,r)$. We consider first general form of reward coupling between x and r, such that Assumption 2.1 does not hold anymore.

Theorem 5.1 (Invariance)**.** *Let $R^*(x,r)$ be an optimal bounded and decreasing reward function for the following problem*

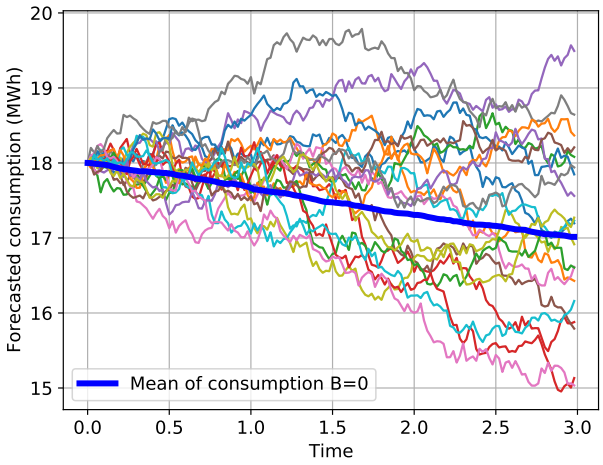
$$\max_{R(x,r)} \left\{ g\left(m_\mu\right) - \int_{-\infty}^{+\infty} R_\mu(x)dx \;\middle|\; \begin{array}{l} \mu = \epsilon(B) \\ V(B) \geq V^{\text{pi}} \end{array} \right\} \tag{18}$$

This equilibrium distribution obtained with R^* is denoted by μ^*. Then,

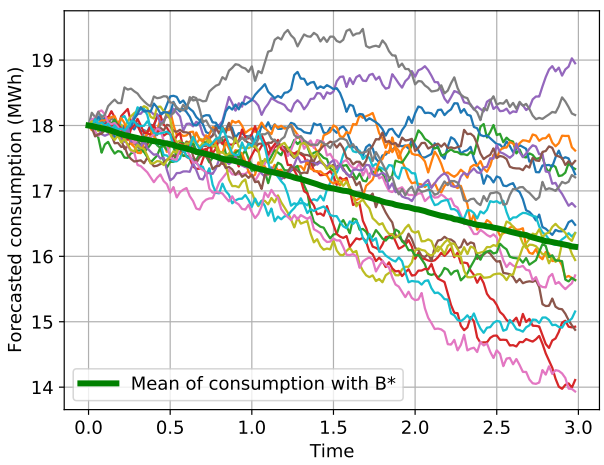
(i) the purely ranked reward function $\hat{B} : r \mapsto R^*(q_{\mu^*}(r), r)$ is also an optimal reward,

(ii) the reward function $\hat{R} : x \mapsto R^*(x, F_{\mu^*}(x))$ is also an optimal reward.

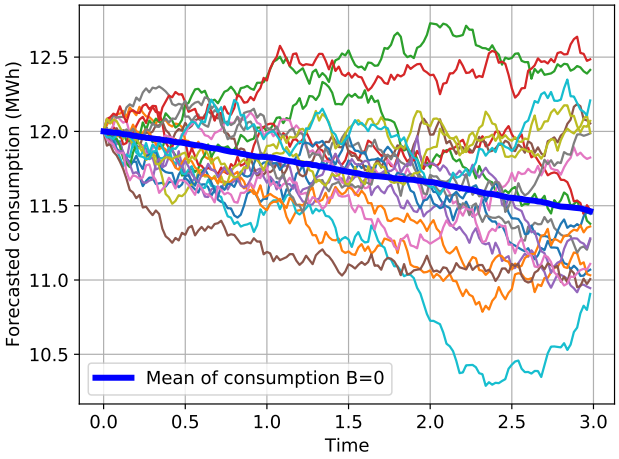
In practice, Theorem 5.1 has very useful implications. It states that complicated reward policies



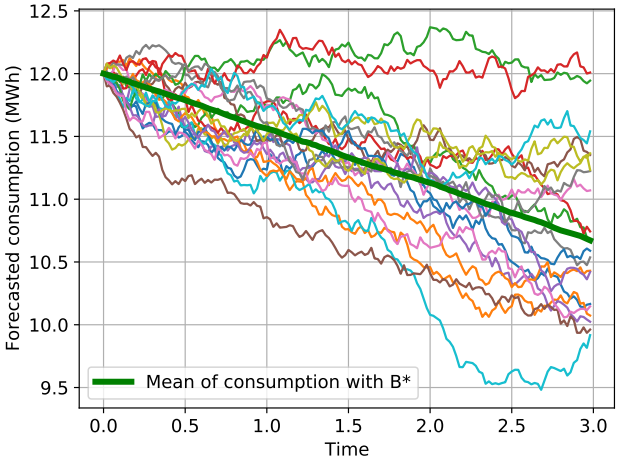
(a) Trajectories without additional reward, first cluster



(b) Trajectories with optimal control from mean-field approximation, first cluster



(c) Trajectories without additional reward, second cluster



(d) Trajectories with optimal control from mean-field approximation, second cluster

Figure 7: Trajectories for 20 consumers (heterogeneous case)

Algorithm 2 Fixed-point Resolution

Require:

- initial p.d.f. $f_{\mu_k^{(0)}}$ of cluster k,
- error tolerance ε,
- iteration maximum n_{max},
- sequence of damping coefficients $\{l_i\}_{i\in\mathbb{N}}$.

$d,\ i \leftarrow 2\varepsilon,\ 0$

while $d \geq \varepsilon$ or $n \leq n_{max}$ **do**

$f_{\mu_k^{(i+1/2)}} \leftarrow \Phi_k(f_{\mu_k^{(i)}})$

$f_{\mu_k^{(i+1)}} \leftarrow l_i f_{\mu_k^{(i+1/2)}} + (1-l_i) f_{\mu_k^{(i)}}$ ▷ damping l_i

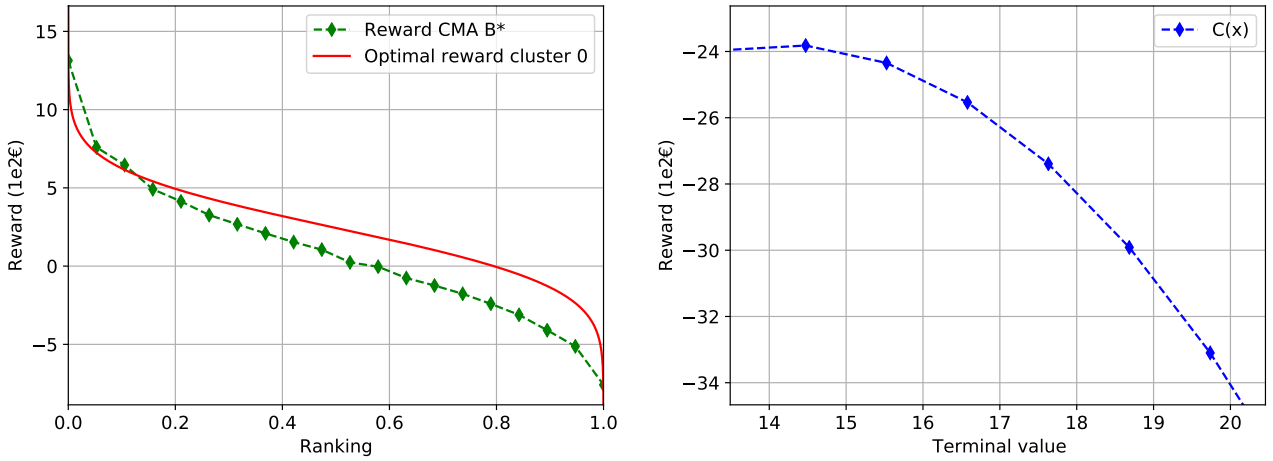
$d \leftarrow W_1\left(f_{\mu_k^{(i)}}, f_{\mu_k^{(i+1)}}\right)$ ▷ distance between two iterates

$i \leftarrow i+1$

end while

simplify into simple rules. The first item shows that we can construct a purely *competitive* game in the sense that the consumers receives incentives only through their rank. The second item shows that we can construct a *decentralized* reward since the incentive of each customer only depends on their own consumption. Note that this notion of invariance applies at the equilibrium, and the equivalence of the reward is no longer true outside the equilibrium.

Numerical experiments. In order to illustrate the invariance theorem, we show an example of optimization with $R(x, r) = B(r) + C(x)$, where $C(x)$ is taken as a quadratic function. During the optimization, we use Algorithm 2 with a linear decreasing damping $(1/(i+1)$. The optimal reward we obtained is depicted in Figure 8.



(a) Rank-based part of the optimal reward

(b) Value-based part of the optimal reward

Figure 8: Example of equivalent optimal reward with non-linear function of the terminal value

Note that the optimal distribution induced by this reward is not depicted. In fact, as shown in Theorem 5.1, the same optimal equilibrium as in Figure 3b.

Time-dependent effort cost In the context of ecological transition, the consumers are more willing to contribute to the energy reduction, and therefore the effort cost c can be viewed as a time dependent parameter, modeling the change of customers' behaviour.

In this case, with a cost profile $c_k(t)$, $t \in [0, T]$ for each cluster k, the consumer's problem becomes

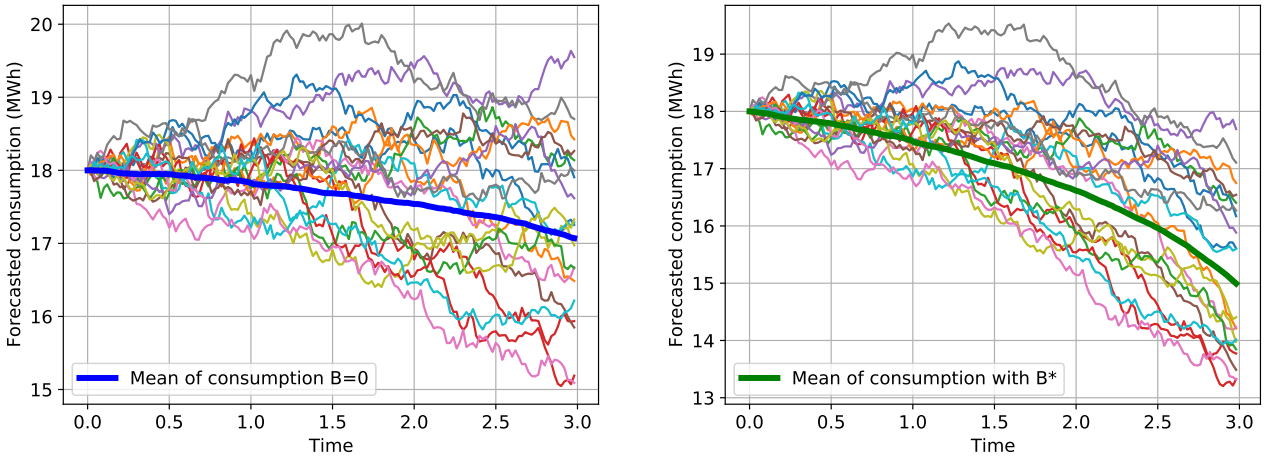
$$V_k(R, \mu_k) := \sup_a \mathbb{E}\left[R_{\mu_k}(X_k^a(T)) - \int_0^T c_k(t) a_k^2(t) dt\right] \quad . \tag{19}$$

Theorem 5.2. *Assume that the cost profiles are bounded such that there exist* $(\underline{c}_k, \overline{c}_k)$ *verifying for all* $t \leq T$

$$0 < \underline{c}_k \leq c_k(t) \leq \overline{c}_k \quad .$$

Then, Theorem 2.3 still applies, i.e., for any reward function $R(x, r)$*, there exists at least one equilibrium distribution.*

Numerical experiments. We now consider a cost decreasing with the time: $c(t) = 5.5 - 1.5t$. Figure 9 draws the same 20 consumers as in the previous cases for the optimal reward obtained in Figure 3a.



(a) Trajectories without additional reward, first cluster

(b) Trajectories with optimal control from mean-field approximation, first cluster

Figure 9: Trajectories with decreasing cost of effort

6 Conclusion

In this work, we study a specific type of incentive – based on the rank of each agent – in order to reduce the global consumption of a population. This reward introduces a competition between the agents, and we are able to give the equilibrium distribution in the mean-field context. The optimal design of the retailer is then analyzed, both theoretically and analytically. We then study the behavior of the model on example inspired by real application case. We numerically observe that ranked-based reward function can be efficient mechanisms to make substantial energy reduction.

References

[1] Ministère de la Transition Energétique. *Bilan de la 4eme période des CEE 2018-2021.* URL: `https://www.ecologie.gouv.fr/dispositif-des-certificats-deconomies-denergie`.

[2] Hunt Allcott and Rogers Todd. "The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation". In: *American Economic Review* 104.10 (2014), pp. 3003–37.

[3] Paul Dolan and Robert Metcalfe. "Neighbors, knowledge, and nuggets: two natural field experiments on the role of incentives on energy conservation". In: *Becker Friedman Institute for Research in Economics Working Paper* (2015).

[4] Jean-Michel Lasry and Pierre-Louis Lions. "Jeux à champ moyen. I-Le cas stationnaire". In: *Comptes Rendus Mathématique* 343.9 (2006), pp. 619–625.

[5] Jean-Michel Lasry and Pierre-Louis Lions. "Jeux à champ moyen. II–Horizon fini et contrôle optimal". In: *Comptes Rendus Mathématique* 343.10 (2006), pp. 679–684.

[6] Jean-Michel Lasry and Pierre-Louis Lions. "Mean field games". In: *Japanese Journal of Mathematics* 2.1 (2007), pp. 229–260.

[7] Minyi Huang, Roland P. Malhame, and Peter E. Caines. "Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle". In: *Commun. Inf. Syst.* 6.3 (2006), pp. 221–252.

[8] M. Huang, P. E. Caines, and R. P. Malhame. "Large-Population Cost-Coupled LQG Problems With Nonuniform Agents: Individual-Mass Behavior and Decentralized Nash Equilibria". In: *IEEE Transactions on Automatic Control* 52.9 (2007), pp. 1560–1571.

[9] Erhan Bayraktar and Yuchong Zhang. "A rank-based mean field game in the strong formulation". In: *Electronic Communications in Probability* 21 (2016), pp. 1–12.

[10] Erhan Bayraktar, Jakša Cvitanić, and Yuchong Zhang. "Large tournament games". In: *The Annals of Applied Probability* 29.6 (Dec. 2019). URL: https://doi.org/10.1214%2F19-aap1490.

[11] Erhan Bayraktar and Yuchong Zhang. "Terminal Ranking Games". In: *Mathematics of Operations Research* 46.4 (Nov. 2021), pp. 1349–1365. URL: https://doi.org/10.1287%2Fmoor.2020.1107.

[12] Yuliy Sannikov. "A Continuous-Time Version of the Principal–Agent Problem". In: *Review of Economic Studies* 75.3 (July 2008), pp. 957–984. URL: https://doi.org/10.1111/j.1467-937x.2008.00486.x.

[13] René Carmona and Peiqi Wang. "Finite-State Contract Theory with a Principal and a Field of Agents". In: *Management Science* 67.8 (Aug. 2021), pp. 4725–4741. URL: https://doi.org/10.1287/mnsc.2020.3760.

[14] Romuald Elie, Thibaut Mastrolia, and Dylan Possamaï. "A Tale of a Principal and Many, Many Agents". In: *Mathematics of Operations Research* 44.2 (May 2019), pp. 440–467. URL: https://doi.org/10.1287/moor.2018.0931.

[15] Arvind Shrivats, Dena Firoozi, and Sebastian Jaimungal. *Principal agent mean field games in REC markets*. 2021. URL: https://arxiv.org/abs/2112.11963.

[16] Nikolaus Hansen. "The CMA Evolution Strategy: A Comparing Review". In: *Towards a New Evolutionary Computation*. Springer Berlin Heidelberg, pp. 75–102. URL: https://doi.org/10.1007/3-540-32494-1_4.

[17] Yongxin Chen, Tryphon T. Georgiou, and Michele Pavon. "On the Relation Between Optimal Transport and Schrödinger Bridges: A Stochastic Control Viewpoint". In: *Journal of Optimization Theory and Applications* 169.2 (Sept. 2015), pp. 671–691. URL: https://doi.org/10.1007/s10957-015-0803-z.

[18] Alexander Fabisch. *CMA-ESpp*. https://github.com/AlexanderFabisch/CMA-ESpp. 2013.

[19] Hoang-Long Ngo and Dai Taguchi. "Strong rate of convergence for the Euler-Maruyama approximation of stochastic differential equations with irregular coefficients". In: *Mathematics of Computation* 85.300 (Oct. 2015), pp. 1793–1819. URL: https://doi.org/10.1090%2Fmcom3042.

A Appendix

In this section, we collect several results and proofs.

Theorem A.1.

$$f_k^{\text{nom}}(x)\exp(\tau x) = \exp\left(\tau x_k^{\text{nom}} + \frac{1}{2}\tau^2\sigma_k^2 T\right)\varphi\left(x; x_k^{\text{nom}} + \tau\sigma_k^2 T, \sigma_k\sqrt{T}\right) \quad . \tag{20}$$

Proof.

$$f^{\text{nom}}(x)\exp(\tau x) = \frac{1}{\sigma\sqrt{T}\sqrt{2\pi}}\exp\left(-\frac{(x-x^{\text{nom}})^2 - 2\tau\sigma^2 Tx}{2\sigma^2 T}\right)$$
$$= \frac{1}{\sigma\sqrt{T}\sqrt{2\pi}}\exp\left(-\frac{(x-[x^{\text{nom}}+\tau\sigma^2 T])^2}{2\sigma^2 T} + \tau x^{\text{nom}} + \frac{1}{2}\tau^2\sigma^2 T\right)$$

□

Theorem A.2. *(i) For a given cluster k, the set of equilibria attainable by a reward function B is given by*

$$\epsilon_k(\mathcal{R}_b^r) = \{\mu \in \mathcal{P}^+(\mathbb{R}) : 2c_k\sigma_k^2 \ln\zeta_{k,\mu_k}(q_{\mu_k}(r)) + pq_{\mu_k}(r) \text{ is bounded and decreasing}\}$$

(ii) If $\mu_k \in \epsilon_k(\mathcal{R}_b^r)$, then

$$\epsilon_k^{-1}(\mu_k) = \left\{2c_k\sigma_k^2 \ln\zeta_{k,\mu_k}(q_{\mu_k}(r)) + pq_{\mu_k}(r) + C : C \in \mathbb{R}\right\}$$

(iii) Suppose that additional reservation "utility" constraint $V_k(B) \geq V_k^{\text{pi}}$ and budget constraint $\int_0^1 B(r)dr \leq K$, then the constant C_k in (ii) is restricted to

$$V_k^{\text{pi}} \leq C_k \leq K - 2c_k\sigma_k^2\int_0^1 \ln\zeta_{k,\mu_k}(q_{\mu_k}(r))dr - pm_{\mu_k} \ .$$

In particular, such a C_k exists if and only if

$$2c_k\sigma_k^2\int_0^1 \ln\zeta_{k,\mu_k}(q_{\mu_k}(r))dr - pm_{\mu_k} \leq K - V_k^{\text{pi}} \ .$$

Proof. (ii) The condition of Theorem 2.4 is verified:

$$\int_0^r \exp\left(-\frac{R_\mu(q_\mu(z))}{2c\sigma^2}\right)dz = \int_0^r (\zeta_\mu(q_\mu(r)))^{-1}\,dz = \int_{-\infty}^{q_\mu(r)} f^{\text{nom}}(z)dz \ .$$

As the uniqueness is concerned, suppose that B and B' lead to the same distribution μ with $p \neq 0$. Then, B and B' lead to the same distribution ν with $p = 0$, see Theorem 2.5. Therefore, as shown in [11], B and B' are equal up to a constant. □

Proof of Theorem 2.5

We give here the proof for a given class and, for simplicity, we omit the dependence in k.

Characterization of an equilibrium. First, suppose that ν is an equilibrium distribution for the case

$p = 0$. Let $\gamma \in \mathbb{R}$ whose value will be determined later. By definition of f_ν (see (5)), we get

$$\int_0^r \exp\left(-\frac{B(z) - p(q_\nu(z) + \gamma)}{2c\sigma^2}\right) dz = \int_{-\infty}^{q_\nu(r)} \exp\left(-\frac{B(F_\nu(x))}{2c\sigma^2} + \frac{p}{2c\sigma^2}(x + \gamma)\right) f_\nu(x) dx$$
$$= \frac{e^{\frac{p}{2c\sigma^2}\gamma}}{\beta(\nu)} \int_{-\infty}^{q_\nu(r)} \exp\left(-\frac{B(F_\nu(x))}{2c\sigma^2} + \frac{p}{2c\sigma^2}x\right) f^{\text{nom}}(x) \exp\left(\frac{B(F_\nu(x))}{2c\sigma^2}\right) dx.$$

Using (20) with $\tau = \frac{p}{2c\sigma^2}$ and the change of variables $u = \frac{x - (x^{\text{nom}} + \frac{pT}{2c})}{\sigma\sqrt{T}}$, we deduce

$$\int_0^r \exp\left(-\frac{B(z) - p(q_\nu(z) + \gamma)}{2c\sigma^2}\right) dz = \frac{1}{\beta(\nu)} e^{\frac{1}{2c\sigma^2}\left(\gamma + px^{\text{nom}} + \frac{Tp^2}{4c}\right)} \int_{-\infty}^{q_\nu(r)} \varphi\left(x; x^{\text{nom}} + \frac{pT}{2c}, \sigma\sqrt{T}\right) dx$$
$$= \frac{1}{\beta(\nu)\sqrt{2\pi}} e^{\frac{1}{2c\sigma^2}\left(\gamma + px^{\text{nom}} + \frac{Tp^2}{4c}\right)} \int_{-\infty}^{\frac{q_\nu(r) - (x^{\text{nom}} + \frac{pT}{2c})}{\sigma\sqrt{T}}} \exp\left(-\frac{u^2}{2}\right) du$$
$$= \frac{1}{\beta(\nu)} e^{\frac{1}{2c\sigma^2}\left(\gamma + px^{\text{nom}} + \frac{Tp^2}{4c}\right)} N\left(\frac{q_\nu(r) - (x^{\text{nom}} + \frac{pT}{2c})}{\sigma\sqrt{T}}\right) .$$

Therefore, taking $\gamma = -\frac{pT}{2c}$, we end up with

$$N\left(\frac{\left[q_\nu(r) - \frac{pT}{2c}\right] - x^{\text{nom}}}{\sigma\sqrt{T}}\right) = \frac{\int_0^r \exp\left(-\frac{B(z) - p\left[q_\nu(z) - \frac{pT}{2c}\right]}{2c\sigma^2}\right) dz}{\int_0^1 \exp\left(-\frac{B(z) - p\left[q_\nu(z) - \frac{pT}{2c}\right]}{2c\sigma^2}\right) dz} .$$

By setting $q_\mu(r) = q_\nu(r) - \frac{pT}{2c}$, we recover the characterization of an equilibrium (see Theorem 2.4).

Conversely, suppose now that μ is the equilibrium for $p \in \mathbb{R}$. Then, following the same steps,

$$N\left(\frac{\left[q_\mu(r) + \frac{pT}{2c}\right] - x^{\text{nom}}}{\sigma\sqrt{T}}\right) = \frac{\int_0^r \exp\left(-\frac{B(z)}{2c\sigma^2}\right) dz}{\int_0^1 \exp\left(-\frac{B(z)}{2c\sigma^2}\right) dz} .$$

The distribution ν defined as $q_\nu(r) = q_\mu(r) + \frac{pT}{2c}$ is a valid equilibrium.

Uniqueness of the equilibrium. Suppose that there exist two distinct equilibrium distributions μ and μ' such that $q_\mu \neq q_{\mu'}$. Then by the above proof, we derive the existence of two distinct equilibrium distributions ν and ν' for the case $p = 0$ satisfying $q_\nu \neq q_{\nu'}$. We get a contradiction by the uniqueness of the equilibrium for purely ranked-based rewards.

Proof of Lemma 2.7

We apply the KKT conditions on $(\widetilde{P}^{\text{ret}})$:

$$\begin{cases} 0 = yg'(m_\mu) - 2c\sigma^2 \ln\left(\dfrac{f_\mu(y)}{f^{\text{nom}}(y)}\right) + \lambda,\ \forall y \in \mathbb{R} \\ \displaystyle\int_{-\infty}^{+\infty} f_\mu(y)dy = 1 \end{cases}, \lambda \in \mathbb{R}$$

From which we can deduce:

$$\begin{aligned} \forall y \in \mathbb{R},\ f_\mu(y) &= \frac{1}{\alpha(\mu)} f^{\text{nom}}(y) \exp\left(y\frac{g'(m_\mu)}{2c\sigma^2}\right) \\ \alpha(\mu) &= \int_{-\infty}^{+\infty} f^{\text{nom}}(y) \exp\left(y\frac{g'(m_\mu)}{2c\sigma^2}\right) dy \end{aligned}$$

Then, as the objective function is concave in f_μ (from 2.3), and the equality constraint is affine, the optimality condition is sufficient.

Proof of Lemma 2.7

Integrating (10) gives us

$$\begin{aligned} m_\mu = \int_{-\infty}^{+\infty} y f_\mu(y)dy &= \frac{1}{\alpha(\mu)} \int_{-\infty}^{+\infty} f^{\text{nom}}(y) \exp\left(y\frac{g'(m_\mu)}{2c\sigma^2}\right) dy \\ &= \int_{-\infty}^{+\infty} y\phi\left(y; x^{\text{nom}} + \frac{Tg'(m_\mu)}{2c}, \sigma\sqrt{T}\right) dy \\ &= x^{\text{nom}} + \frac{Tg'(m_\mu)}{2c} = x^{\text{pi}} + \frac{T}{2c}\delta(m_\mu)\ . \end{aligned}$$

We can now recover the reward:

$$\begin{aligned} B^*(r) &= V^{\text{pi}} + 2c\sigma^2 \ln\left(\zeta_{\mu^*}(q_{\mu^*}(r))\right) + pq_{\mu^*}(r) \\ &= V^{\text{pi}} + q_{\mu^*}(r)\left[p - c_r + s'(m_{\mu^*})\right] - 2c\sigma^2 \ln\left(\int_{-\infty}^{+\infty} f^{\text{nom}}(y) \exp\left(y\frac{g'(m_{\mu^*})}{2c\sigma^2}\right) dy\right) \\ &= V^{\text{pi}} + \frac{c}{T}\left[(x^{\text{nom}})^2 - m^2\right] + q_{\mu^*}(r)\delta(m_{\mu^*}) \\ &= \frac{c}{T}\left[(x^{\text{pi}})^2 - m^2\right] + q_{\mu^*}(r)\delta(m_{\mu^*})\ . \end{aligned}$$

From the definition of the provider objective,

$$\begin{aligned} \pi &= g(m) + pm - \int_0^1 B^*(r)dr \\ &= s(m) - c_r m + pm - \frac{c}{T}\left[(x^{\text{pi}})^2 - m^2\right] - m\left[p - c_r + s'(m)\right] \\ &= s(m) - ms'(m) + \left(\frac{x^{\text{pi}} + m}{2}\right)\frac{2c}{T}\left(m - x^{\text{pi}}\right) \\ &= s(m) - ms'(m) + \left(\frac{x^{\text{pi}} + m}{2}\right)\delta(m)\ . \end{aligned}$$

Proof of Theorem 5.1

(i) By construction, the reward $\tilde{B}$ is also bounded and decreasing. Then, the cost induced by the additional reward is the same with R^* and $\hat{B}$:

$$\int_{-\infty}^{+\infty} R^*_{\mu^*}(x) f_{\mu^*}(x) dx = \int_0^1 \hat{B}(r) dr \ .$$

Finaly, μ^* is also an equilibrium for the reward $\hat{B}$:

$$\frac{1}{\hat{\beta}(\mu^*)} f^{\text{nom}}(x) \exp\left(\frac{\hat{B}(F_{\mu^*}(x))}{2c\sigma^2}\right) = \frac{1}{\beta^*(\mu^*)} f^{\text{nom}}(x) \exp\left(\frac{R^*_{\mu^*}(x)}{2c\sigma^2}\right) = f_{\mu^*} \ ,$$

where $\hat{\beta}$ and β^* are computed respectively with $\hat{B}$ and R^*. The last equality comes from the characterization of an equilibrium. Therefore, the reward function $\hat{B}$ satisfies the constraints and produces the same objective value as R^*. It is also optimal.

(ii) The proof follows the same ideas as at the previous item.