2012

Sellers with Misspecified Models

Kristof Madarasz, *London School of Economics and Political Science*
Andrea Prat

Sellers with Misspecified Models *

Kristóf Madarász (LSE) †and Andrea Prat (Columbia)‡

Abstract

Principals often operate on the basis of misspecified models of their agents' preferences. We show that even slight misspecification can lead to large and non-vanishing losses. Instead we propose a two-step scheme, whereby: (i) the principal identifies the model-optimal menu; (ii) modifies prices by offering to share with the agent a fixed proportion of the profit she would receive if this item was sold at the model-optimal price. We show that her loss is bounded and vanishes smoothly as the model converges to the truth. Finally, two-step mechanisms without a sharing rule like (ii) will not yield a valid approximation.

1 Introduction

As George Box famously put it, "Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful." In agency theory, a principal is assumed to operate on the basis of the agent's preferences. Her model will however, at best, be an approximation of the truth. Hence she may not be able to design the truly optimal contract. How should a principal who knows that her model is potentially misspecified act in such a circumstance?

In such a context - following March and Simon's (1958) classic approach to organizational decision making - one can ask two related questions. Can the principal find a solution that achieves an acceptable payoff even if her model turns out to be wrong? How will such a contract differ from the contract she would offer if her model was exactly true?

*An earlier version of this paper was circulated under the title "Screening with an Approximate Type Space". We thank Ken Arrow, Mark Armstrong, Sylvain Chassang, Vincent Conitzer, Philippe Jehiel, Sergiu Hart, Thomas Mariotti, John Moore, Jim Minifie, Ilya Segal, Ennio Stachetti, Tomasz Sadzik, Balazs Szentes, Miriam Sinn, Joel Sobel, and seminar participants at Berkeley, Brown, Cambridge, Caltech, Chicago, Columbia, Gerzensee, Harvard-MIT, Hebrew, Helsinki, LSE, NYU, Stanford, UCL, for useful suggestions.

†k.p.madarasz@lse.ac.uk, London School of Economics, Houghton Street, WC2A 2AE, London UK

‡ap3116@columbia.edu, Department of Economics, and Graduate School of Business, Columbia University

Our paper attempts to answer these questions in the context of one of the classic problems of all of microeconomics: single-agent mechanism design with quasi-linear preferences. This model – commonly referred to as the 'screening problem' – has found various key economic applications from regulation and taxation and to labor markets, insurance and incentive design. In its classic interpretation of nonlinear pricing, a multi-product monopolist offers a menu of product-price specifications to a buyer or a continuum of buyers (e.g. Wilson 1993).

In the standard formulation, the principal knows the true distribution of the agent's preferences. In this paper, we re-visit the general screening problem, but instead assume that the principal (seller) does not know the true distribution of the agent's (buyers') preferences. The principal faces model uncertainty and has access only to a model that is potentially misspecified. The seller knows that her model is potentially wrong - for example because it is typically simpler than reality - and she has a sense – to be formalized shortly – of how much her model varies from the buyers' true preferences. Can such a principal guarantee herself an outcome that is not much worse than what she could expect had she access to the true preferences of the agent?

For instance, the agent's preferences may depend on his physical location. While geography may affect preferences continuously, data often comes in a discretized form: the seller may know roughly how many people live in a certain zip code area, but not their exact location or the exact way location affects preferences. Of course, she could assume an arbitrary continuous distribution, specific to each zip code area and a specific utility function. However, the seller could instead be interested in finding a contract that achieves robust performance given any possible within-area distributions.

As another example of model uncertainty, consider a situation where the agent's willingness-to-pay for an object depends on a list of attributes: income, age, location, etc. This list may be long though, and the principal may operate on the basis of a model that explicitly included the effect of the most important factors, but left some characteristics unmodeled or modeled only in an approximate way. In this instance, she wishes to design a mechanism using her model that is robust to misspecifications of the minor attributes.

Our paper shows two results, a negative and a positive one. On the negative side, we show that designing a contract as if the principal's misspecified model was correct leads to potentially large losses, and these losses do not vanish even as the model gets arbitrarily close to the truth. As a consequence, even a small degree of misspecification can cause largely suboptimal outcomes and cause a principal who ignores model uncertainty to be surprised by how badly her contract performs relative to what she could expect if her model was true. The positive result instead

identifies a simple two-step procedure that departs from the above naive solution in a systematic way and produces a valid approximation for a very large class of situations. In fact we show that any contract that will be robust to preference misspecifications must be similar to the contract identified by this procedure. The rest of this section summarizes these two results in informal terms.

Outside of the problem of model uncertainty, even if the principal had access to a correct model of the agent's behavior, when taking all factors into account, the correct solution of the screening problem might be prohibitively difficult. Indeed Conitzer and Sandholm (Theorem 1, 2004) show that finding an exact solution to the single-agent mechanism design problem we consider here is NP-complete. Our method offers a smooth trade-off between adopting a simpler representation and facing a tolerable loss relative to the optimal revenue, even in the large class of domains where naively relying on the simpler representation will lead to a great loss.

To introduce our results, we first have to discuss model misspecification, and a way of measuring the quality of the principal's model. In our setup, the model is a finite approximation of the agent's true preferences, hence every model type represents an uncountable set of possible true types. We define the approximation index of a model to be the maximal distance – in terms of willingness to pay for any product – between any model type and any of the true types it represents, For any model and any value of the approximation index, there is a set of true preference profiles whose maximal distance from the model is weakly less than the value we have chosen - where we will also allow for a probabilistic or local interpretation of this statement. In the geographical example a natural approximation index is given by the maximal preference distance between two types in any district and the set of possible agent preference profiles includes all possible geographical distributions given the partition of agents into districts.

In March and Simon's spirit, the near-optimal contract we are seeking can rely only on the information available to the principal. The menu offered to the agents will thus depend only on the principal's model and the approximation index, and no other information about the true types. For any true type space satisfying the approximation index bound, the approximation loss is given by the difference between the profit that the principal would get if she optimized over the true type space and the profit she gets with the menu computed by the algorithm. A near-optimal solution puts a bound on this approximation loss and guarantees that the bound vanishes as the approximation index goes to zero.

To introduce the negative result (Theorem 2), note that finding a near-optimal solution given model uncertainty in our strategic setting poses a challenge that is absent in non-strategic environments. Even when all primitives are well-behaved,

the fact that the agent best-responds to the menu offered to him by the principal creates room for discontinuity: a small change in the menu might lead to a large change in the principal's expected payoff. The discontinuity is heightened by two elements. First, in the exact solution of the screening problem the principal's payoff function is discontinuous exactly at the equilibrium allocation: this is because profit maximization implies a system of binding incentive-compatibility and participation constraints. Second, outside the narrow monotonic one-dimensional case, binding constraints may well be non-local (Wilson 1993, Armstrong 1996, Rochet and Choné 1998, Armstrong and Rochet 1999). This makes dealing with misspecification challenging: a small perturbation of a payoff type might lead to large changes in equilibrium choice behavior and affect the principal's payoff discontinuously.

Our negative result examines the profit loss of a principal who behaves as if her model type space was correct. She simply computes the optimal menu given her model and offers it to the agent. We show that, already when preference heterogeneity is two dimensional, the upper bound to the profit loss remains finite even as the approximate index goes to zero. This means that a naive principal experiences a discontinuous loss when moving from the case where her model is exactly correct to the case where it is only almost correct. The result is proven by showing that there exists a broad class of screening problems in which the optimal solution involves binding nonlocal constraints for a positive measure of types. Here, as argued above, a small perturbation of payoff types creates large changes in equilibrium behavior and a large loss for the principal. We also provide a simple example of this situation in the paper (Section 3).

One might try to address our negative result by finding sufficiently restrictive conditions that guarantee that only local incentive constraints bind. As argued above, this works well only when preference heterogeneity is one dimensional (e.g., Mussa and Rosen 1978), but not when preferences vary along multiple dimensions, and arguably "In most cases that we can think of, a multidimensional preference parameterization seems critical to capturing the basic economics of the environment" (Rochet and Stole, 2003).[1] Our goal is not to identify a contract or a mechanism that works very well in one specific environment, but rather to find one that produces an acceptable outcome for a large class of screening problems given the model uncertainty faced by the principal – one that uses generic preferences, cost functions, and (potentially multi-dimensional) type spaces. This will then also help ensure that the contract will be robust to violations of exact preferences restrictions.

The positive result (Theorem 1) identifies an approximation scheme that works

[1]Battaglini and Lamba (2012) consider a dynamic version of the one-dimensional screening model with imperfect type persistence. They show that even in a twice repeated setting, non-local incentive constraint bind and that solutions based on the first-order approach considering only adjacent constraints are not valid.

in any smooth type space. We call our solution concept a *profit-participation mechanism.* Given an approximate type space and its corresponding approximation index, we define the profit-participation mechanism, based on two steps:

(i) The principal solves for the optimal menu, a vector of product-price pairs, based on the set of all feasible products as if the model type space was true.

(ii) The principal then takes the menu obtained in the first step, keeps the product vector unchanged and modifies the price vector. In particular, our principal willingly offers a discount on each product proportional to the profit she would get if that product was sold at the model-optimal price. The size of the absolute discount, which is determined by the mechanism, depends only on the approximation index, i.e., our measure of model uncertainty.

Theorem 1 studies the difference between the expected profit (over the true type space) generated by the menu obtained by our profit-participation mechanism and the expected profit (over the true type space) generated by the menu that would be optimal given the true type space – as the agent's preferences are potentially misspecified, both these expected profits are unknown. We are nevertheless able to prove the existence of an upper bound to this difference and show that this upper bound is a smooth function of the approximation index. Furthermore, for any screening problem, the upper bound vanishes smoothly as the approximation index goes to zero.

Profit participation yields a near-optimal solution in the presence of model uncertainty because it addresses the violation of optimally binding non-local incentive constraints. By willingly offering a profit-related discount, the principal makes the agent a shared residual claimants of her model profit. This guarantees that allocations that yield more profit in the menu that is optimal given the model types become relatively more attractive for the agent. Now, a true type that is close to a model type may still not choose the product that is meant for that model type. Even if he chooses a different product from the modified menu, this must now be one that would have yielded an approximately higher profit in the original menu – the difference in the principal's profit is bounded by a constant that is strictly decreasing in the discount.

While a profit-related discount is beneficial because it puts an upper bound to the profit loss due to the discrepancy between the choice of a true type and a model type, it also has a cost in terms of lower sale prices. The discount rate used in the profit-participation mechanism strikes an optimal balance between this cost and the above described benefit. We show that as the approximation index decreases, a given upper bound to the profit loss can be achieved with a lower discount, and as the model tends to the truth, the optimal discount goes to zero as well.

One may wonder whether there are other ways of achieving a valid approximation besides the one we propose. We show that profit-sharing is a necessary feature of any valid approximation scheme within a large class of mechanisms. Our negative result (Theorem 2) does not just cover the naive mechanism mentioned earlier. It applies to any model-based mechanism, namely any scheme that begins with step (i) of the profit-participation mechanism. In other words it applies to all algorithms that can modify model-optimal prices according to *any* rule. Such modification includes the naive mechanism discussed above – in which the model-optimal price vector is left unchanged – as well as any scheme based on price manipulation.

Theorem 2 proves that, if preference heterogeneity is multi-dimensional, then any model-based mechanism which violates a profit-participation condition cannot be a valid approximation: the upper bound to the profit loss does not vanish as the approximation index goes to zero. This means that if there are model-based mechanisms that do at least as well as the profit-participation mechanism, they must be similar in spirit to the one we propose, in that they contain an element of profit participation.

The economic insight from our result is that approximate models can play a useful role in general contracting environments as long as the risk of misspecification is dealt with in an appropriate manner. A principal who has only an imperfect model of the agent's behavior can start by taking her simpler model at face value and find its optimal solution. However, the resulting allocation is not robust to model misspecification. To make sure that small errors in the model do not lead to serious profit losses, the principal must act 'magnanimously'. She needs to make the agent the joint residual claimant on part of her profit that she would make if her model was true. Such apparent generosity takes the form of a discount that should be greater for more lucrative products. The comparative statics implication of our results (Corollary 1) is that, as the principal's model of the agent preferences becomes less accurate, the principal needs to price products in a more magnanimous way.

The paper is structured as follows. Section 2 introduces the screening problem and defines the notion of an approximate type space. Section 3 provides a simple example, where the type of the agent corresponds to his geographical location, to illustrate the logic behind our two results. For exposition reasons, we report first the positive result and then the negative result. This is because the latter not only includes the 'naive' mechanism, but all mechanisms that do not use profit sharing, and hence is best discussed after introducing our positive result. Therefore, Section 4 discusses the positive result: we develop Profit-Participation Pricing and establish an approximation bound (Lemma 1), we show the positive result of the paper, namely that the profit-participation mechanism is a valid approximation scheme

(Theorem 1) given model uncertainty, and we discuss comparative statics (Corollary 1). Section 5 shows that model-based mechanisms are valid approximation schemes only if they contain an element of profit participation (Theorem 2). Section 6 concludes.

1.1 Literature

To the best of our knowledge, this is the first paper to discuss near-optimal screening when the principal faces model uncertainty and uses a misspecified type space.

There is of course a large body of work on approximation in single-agent problems in many disciplines including economics. However, as mentioned earlier, our set-up presents a form of discontinuity that is due to the strategic interaction between the principal and the agent. The agent's payoff in our model is always continuous in the agent's type, as the Theorem of the Maximum predicts (Ausubel and Deneckere 1993). It is the principal's payoff that may be discontinuous because of the misspecified response of the agent. As our example in Section 4.3 illustrates, this discontinuity exists even when the agent's utility is continuous in allocation and type.[2]

Near optimal nonlinear pricing was first analyzed by Wilson (1993, section 8.3), who discusses the approximate optimality of multi-part tariffs (with a one-dimensional type). The closest work in terms of approximation in mechanism design is Armstrong (1999), who studies near-optimal nonlinear tariffs for a monopolist as the number of products goes to infinity, under the assumption that the agent's utility is additively separable across products. He shows that the optimal mechanism can be approximated by a simple menu of two-part tariffs, in each of which prices are proportional to marginal costs (if agent's preferences are uncorrelated across products, the mechanism is even simpler: a single cost-based two-part tariff). There are a number of key differences between our approach and Armstrong's. Perhaps, the most important one is that his approximation moves from a simplification of the contract space while we operate on the type space.[3]

Independent work by Chassang (2011) studies approximately optimal contracts in a dynamic delegated investment problem with moral hazard, adverse selection, and a limited liability constraint on both principal and agent. The paper identifies a class of calibrated contracts that perform approximately as well as a linear bench-

[2]For an analysis of strategic approximation in games with symmetric information, see Reny (2012).

[3]See also Chu, Leslie, and Sorensen (2010) for a theoretical and empirical analysis of this problem. A growing literature at the intersection of computer science and economics studies near-optimal mechanisms under computational complexity (e.g., Conitzer and Sandholm 2004) or communication complexity (e.g., Nisan and Segal 2006). However, the focus is on designers who face restrictions on the space of mechanisms rather than model uncertainty.

mark contract with a number of attractive properties. The performance bound is independent of the underlying process for returns. While our two papers apply to different environment and use different approaches, they both include the feature that the principal shares part of her profit with the agent in a manner that is designed to correct for small but potentially damaging failures of benchmark incentive compatibility constraints. We suspect that this may be a general property of mechanisms that are near-optimal in a variety of environments.

The goal of our paper is also related to Gabaix (2010). In that decision-theoretic framework, the agent operates on the basis of a "sparse" representation of the world. The objective is to find the optimal sparse model compatible with the limited cognitive skills of the decision-maker. Our paper does not endogenize the approximate type space; instead we focus on identifying and tackling the challenges to approximation that arise in strategic environments.

2 Setup

We begin by introducing the standard single-agent quasilinear mechanism design problem. Let Y be a compact set of available alternatives. The principal selects a subset of the set Y and assigns transfer prices $p \in \mathbb{R}$ to each element of this subset. The resulting menu is thus a set of items, a set of alternative-price pairs. Let's denote a menu by M and assume that it always contains the outside option y_0 whose price p_0 is normalized to zero. Once a menu is offered, the agent is asked to choose exactly one item from this menu.

The principal's profit is the transfer price net the cost of producing the object:

$$\pi(t, y, p) = p - c(y)$$

where the above specification follows Rochet and Chone (1998), and much of the literature on non-linear pricing, in that the principal's payoff does not directly depend on the agent's type.

For any set of alternatives Y, the agent's preferences are described by two independently defined objects: the *truth* and the *model*. These two will be linked by an *approximation index* and they will give rise to the *approximation problem* we analyze.

Truth. The agent's *true* preferences depend on his private type t from set $T \subseteq \Re^n$ - a compact and connected measurable set within a finite dimensional Euclidean space - drawn from a probability distribution which has a continuous density

function $f(t)$.[4] In particular, the agent's payoff is his type-dependent valuation of the object y net the transfer price to the principal:

$$v(t, y) - p$$

Thus the principal faces a single-agent mechanism design problem which - given a fixed (Y, c) - can be summarized by (T, v, f).We refer to T and v (T_v henceforth) as the true type space.

We assume that there is a finite upper bound on the principal's profit

$$\Pi_{\max} = \sup_{y \in Y, t \in T} v(t, y) - c(y)$$

and denote the supremum of the principal's expected payoff over all feasible menus by $\Pi^*(T_v, f)$. The principal's expected profit is then bounded from below by zero and above by $\Pi_{\max}$. We scale v, c such that $\Pi_{\max}$ is normalized to be 1.

Model. Our key point of departure is that the principal facing the above screening problem, does not have access to the truth. Instead, she is constrained to operate on the basis of a *model* that might systematically differ from the truth. The principal's model is a possibly incorrect representation of the agent's preferences. The model uses a discrete type set S where the preferences of a model type $s \in S$ of the agent, are given by

$$u(s, y) - p$$

with associated probability distribution function $g \in \Delta S$.

We refer to S and u (S_u henceforth) as the model set or equivalently as the approximate type space: the principal's full model is denoted by (S_u, g). Note that the model can generically differ from the truth both in terms of type distribution and also in terms of the preferences. The preferences of the model types considered by the principal might not correspond to the preferences of any actual type from T.

Approximation Index. We now introduce a measure of the quality of the model, namely a distance between the model and the truth. This index describing how misspecified the model is will satisfy two important conditions. First, it will be a simple scalar that reflects some 'distance' between the model and the truth that goes to zero as the model tends to the truth. This corresponds to the minimal information that the principal can have about how misspecified her model potentially is. Second, it will have a worst-case element, which will allow us to find upper

[4]The existence of such a density implies that there are no atoms. The set-up can be extended to encompass discrete type spaces.

bounds to the profit loss. This pessimistic nature of the measure guarantees that as our measure goes to zero any other non-worst case measure would go to zero too.

Given any truth (T_v, f) and model (S_u, g) the corresponding true approximation index $\varepsilon_{\text{true}}$ is defined as follows:

1. An approximation partition $\mathcal{P}$ is a finite measurable partition of T with $\#S$ (possibly non-connected) cells J_s, such that the probability mass of true types belonging to cell J_s (computed according to density f) equals the probability (according to g) of model type s. Let Γ be the (obviously non-empty) set of all approximation partitions.[5]

2. For each cell J_s in an approximation partition $\mathcal{P}$ in Γ, define the upper bound to the maximal utility distance between any type and its associated model type as $d_s(\mathcal{P}) = \sup_y \sup_{t \in J_s} |u(s,y) - v(t,y)|$ and define the upper bound for the whole partition as $d(\mathcal{P}) = \max_{s \in S} d_s(\mathcal{P})$.

3. The *true approximation index* of $(T_v, f; S_u, g)$ is $\varepsilon_{\text{true}} = \inf_{\mathcal{P} \in \Gamma} d(\mathcal{P})$. Note that $\varepsilon_{\text{true}}$ exists and, given that S is finite, it is strictly positive.

4. Let the *approximation index* ε be any number strictly greater than $\varepsilon_{\text{true}}$. Also, define an *$\varepsilon$-approximation partition* as any approximation partition $\mathcal{P}$ with $d(\mathcal{P}) \leq \varepsilon$. There must exist at least one ε-approximation partition.[6]

Approximation Problem. In this setting, the principal knows the model and an approximation index ε that reflects model uncertainty. She does not know anything about the truth except that the distance between the truth and her model is not greater than ε. The paper considers solutions that are based only on the cost function, the model, and the approximation index, namely $(S_u, g, Y, c, \varepsilon)$ but do not require any further information about the truth. The goal then is to put a bound on the payoff loss that the principal experiences because she uses the model rather than the truth. Namely, our goal is to obtain a solution that guarantees that, for *any* (T_v, f) such that $\varepsilon_{\text{true}} \leq \varepsilon$, the difference between the principal would receive under the truly optimal mechanism for (T_v, f) and the payoff that the principal receives using a model-based algorithm in an environment in which the agent's true

[5] Since f is absolutely continuous and T is Lebesgue measurable, it can be proven based on Lyapunov's original convexity theorem that such a partition always exists.

To sketch a proof, index the elements of S by i where $i \in \{1, ..., \#S\}$. Take $i = 1$, since f is absolutely continuous and for all measurable $J(s_1) \subset T$ it follows that $f(J(s_1)) \leq f(T)$, it follows that we can always find a set $J(s_1) \subset T$ such $f(J(s_1)) = f_S(s_1)$. See Theorem 2 of Ross (2005) for a proof. Since S is finite, we can repeat this procedure and find $J(s_2)$ in the domain $T \backslash J(s_1)$ and thus proceed inductively.

[6] If there exists a partition $\mathcal{P}^*$, that attains $\inf_{\mathcal{P} \in \Gamma} d(\mathcal{P})$, then we can also allow $\varepsilon = \varepsilon_{\text{true}}$ and $\mathcal{P}^*$ is an ε-partition.

preferences are described by (T_v, f) is bounded above by some function of ε, with the property that the bound goes to zero as $\varepsilon \to 0$. While this problem can be posed under specific distributional or functional assumptions, we are interested in finding a solution that is informationally minimal and hence works for any environment defined above.

The analysis can be easily extended to situations where the principal is not fully certain that $\varepsilon_{\text{true}} \leq \varepsilon$. If the principal thinks that there is a small probability δ that $\varepsilon_{\text{true}} > \varepsilon$, one can simply modify the upper bound to the loss by adding a worst-case scenario (a profit of zero) that occurs with probability δ. In the same spirit, one can extend the analysis to cases where the principal faces more local model uncertainty and hence her confidence varies locally, allowing for different indices to apply to different regions of the preference space.

2.1 Discussion

Our principal is constrained to operate on a model rather than the truth, but she is willing to take a stand on the maximal payoff distance between her model and the truth. The principal does not know the truth (T_v, f), but she knows that her representation, (S_u, g) is such that $\varepsilon_{\text{true}} < \varepsilon$. The approximation index imposes an informationally minimal restrictions on what the true type space might be relative to the principal's model. For any given model, there is a potentially very large class of preference specification and distributions over these (T_v, f) such that they satisfy the approximation index, their $\varepsilon_{\text{true}} < \varepsilon$.

Unlike the rest of the literature on approximation in mechanism design, our paper does not in any way restrict the set of contracts that the principal may use. What prevents the principal to achieve the optimal profit is purely model uncertainty.

Why does our principal use an approximate model rather than the truth? In many situations, it is unrealistic to assume that the principal can fully specify the underlying true distribution of preferences. Rather she operates under weaker epistemic conditions, like some non-probabilistic uncertainty (Walley 1991) or some coarser understanding of the agent's behavior.[7] She knows that her model may be wrong, but she does not know in which direction. She knows, however, an upper bound on how misspecified her model might be, e.g., she knows an approximation index ε .

In other situations, the principal can improve the quality of her model but that comes at a cost. For instance, in the geographical model she could obtain the

[7]Related concepts are applied to macroeconomics (Hansen and Sargent 2010), but there it is the agent who faces model uncertainty, not the designer, i.e., the principal/ government.

exact location of agents. Our result can help the principal decide whether to incur the additional cost: the upper bound on the profit loss provides a measure of the potential benefit of learning the truth.

3 Example

We now introduce a class of examples to illustrate the potential problems that can arise when a principal utilizes a (slightly) mis-specified model.

The agent's true type is $T \in [0,1]^m$ where m is a positive integer. Assuem for simplicity that it is uniformly distributed. His true payoff is

$$v(t,y) = \max\left(w(t,y), \bar{w}(t)\right)$$

where $w(y,t)$ is single-peaked in $y \in [0,1]^m$. Let

$$y^*(t) = \arg\max w(t,y)$$

and $y^*(t) \neq y^*(t')$ if $t \neq t'$. Assume that $w(t, y^*(t)) \geq \bar{w}(t)$ for all t. However, of course, it could be that $w(t,y) < \bar{w}(t)$ for other values of y.

A practical example of such a product might be a car. The consumer may need a car, any car, for a practical reason, captured by $\bar{w}(t)$. However, he may also have a taste for a particular model, maybe sports cars or 4x4s, so he is willing to pay $v(y,t) > w(t)$ for certain subsets of y. The particular value $y^*(t)$ represents the ideal car for consumer t.

The model type S is a discretized version of T. For simplicity assume that S is a regular grid and each model type s corresponds to the midpoint of a cell in the grid. Let $\varepsilon_{\text{true}}$ be the maximal difference $v(t,y) - v(s,y)$ whenever t belongs to the grid of s over all $s \in S$.

We make an additional assumption: For every t and t'

$$v(y^*(t), t) - w(t) \geq v(y^*(t), t') - w(t'), \tag{1}$$

which has an intuitive interpretation: type t is willing to pay more that type t' to move from a product with no aesthetic value to the one preferred by t. This is reasonable as this is the product preferred by t, not t'.

For instance, this is satisfied if

$$v(y^*(t), t) = -a\|(y-t)\| \text{ and } w(t) = b\|t\|$$

where $\|\cdot\|$ is the Euclidean distance, and we make the asumption that $a \geq b$. That

is because

$$v\left(y^{*}\left(t\right),t\right)-w\left(t\right)=-a\left\|0\right\|-b\left\|t\right\|\geq-a\left\|t-t'\right\|-b\left\|t'\right\|=v\left(y^{*}\left(t\right),t'\right)-w\left(t'\right),$$

due to the triangle inequality ($\left\|t-t'\right\|+\left\|t'-0\right\|\geq\left\|0-t\right\|$) and the fact that $a\geq b$.

It is also satisfied if we assume that

$$v\left(y^{*}\left(t\right),t\right)=h\left(y-t\right)+w\left(t\right)$$

where $h\left(y-t\right)$ is single-peaked and has maximal value $h\left(0\right)$. That is because

$$v\left(y^{*}\left(t\right),t\right)-w\left(t\right)=h\left(0\right)\geq h\left(y-t\right)=v\left(y^{*}\left(t\right),t'\right)-w\left(t'\right).$$

While the modeler knows v for all values in T, the principal only knows v for values in S and has approximation index ε (namely he knows that $\varepsilon_{\text{true}}\leq\varepsilon$).

The goal of this example is to show that if the principal solves the problem as if the model space was correct he incurs a large loss and that this loss does not vanish as $\varepsilon\rightarrow0$. We will do so by showing that the solution involves generic binding non-local constraints.

We begin by defining the lowest price the monopolist charge for any product that is offered for sale:

$$p_{\min}=\min_{s\in S,\hat{y}(t)\neq y_{0}}p\left(\hat{y}\left(s\right)\right)$$

The minimal price $p_{\min}$ divides the type set S into three subsets:

- Excluded types S_0: those for whom

$$w\left(s,y^{*}\left(s\right)\right)<p_{\min}.$$

- Served types S_1 such that

$$w\left(s,y^{*}\left(s\right)\right)\geq p_{\min}.\geq\bar{w}\left(s\right)$$

- Served types S_2 such that

$$\bar{w}\left(t\right)\geq p_{\min}$$

We show that for types in S_2 the binding constraint is nonlocal. If there exists a solution where each type in S_1 is charged $w\left(s,y^{*}\left(s\right)\right)$ and each type in S_2 is charged $w\left(s,y^{*}\left(s\right)\right)-\bar{w}(s)+p_{\min}$, that solution must be optimal because it is based on the maximal conceivable willingness-to-pay given $p_{\min}$. To see that this solution exists, assume that each type t in S_1 and S_2 is offered his ideal product $y^{*}\left(s\right)$ at

price $p(s) = w(s, y^*(s))$ in S_1 and price $p(s) = w(s, y^*(s)) - \bar{w}(s) + p_{\min}$ in S_2. Inspect first the incentive-compatibility constraints for s in S_1 and S_2 are:

$$w(s, y^*(s)) - p(s) \geq \max(w(s', y^*(s)), \bar{w}(s)) - p(s') \text{ for all } s' \in S_1 \cup S_2$$

and they can be split into two sets of inequalities. The first set has the form

$$w(s, y^*(s)) - p(s) \geq \bar{w}(s) - p(s') \geq \bar{w}(s) - p_{\min}$$

and it is satisfied by definition because $p_{\min} \geq \bar{w}(s)$ in S_1 and because of the definition of $p(s)$ in S_2. The second set of inequalities has the form:

$$w(s, y^*(s)) - p(s) \geq w(s', y^*(s)) - p(s').$$

in the case of S_1, one can see the inequalities are satisfied when $p(\cdot)$ is replaced by its definition. In the case of S_2, the second set of inequalities can be rewritten as

$$\bar{w}(s) \geq w(s', y^*(s)) - w(s', y^*(s')) + \bar{w}(s')$$

and it is thus satisfied by assumption (1).

This shows that for every $p_{\min}$ there is an optimal solution where for every $s \in S_1$ the binding constraint is the participation constraint and for every $s \in S_2$ the binding contraint is the incentive compatibilty contraint with the lowest-priced product.

Now consider what happens when the menu that is optimal for model type S is used for the true type T. Take a type t near s. If he chooses the product meant for s, namely $y^*(s)$, he receives utility $w(t, y^*(s)) - p(s)$ but this is lower than the utility that comes from some other product (if $s \in S_2$). instead of making a profit $p(s) = w(s, y^*(s)) - \bar{w}(s) + p_{\min}$, the principal makes only $p_{\min}$.

Note that simply providing a discount on all prices will not solve this problem because it will also reduce $p_{\min}$.

As we shall show in the rest of the paper, the logic of this simple example is general. An approximation scheme based on profit discounting always works (Theorem 1). No discounting or schemes that are not based on profit discounting will not work (Theorem 2).

4 Profit-Participation Mechanism

In this section, we show the positive result of our paper: we can always obtain a valid approximation if we use a profit-participation mechanism. As mentioned in

the introduction, it is expositionally simpler to begin with the positive result, as our negative result discusses mechanisms that do not contain the profit-participation property that characterizes the positive result.

The section begins with an intermediate result (Lemma 1) on profit-participation discounting. We then prove the main result (Theorem 1) and we conclude with a brief mention of comparative statics.

4.1 Profit-Participation Pricing

In this section we introduce the key component of our solution method, Profit-Participation Pricing, and present an intermediate result that bounds the profit loss for the principal.

First, we define a notion of expected profit that can be applied to both the truth and the model. Let us define it for the model first. It is the expected profit that the principal would receive if her model was the true. Hold the set of products Y and the cost function c constant and, for any menu $M = \{y_k, p_k\}_k$ and model (S_u, g), define

$$\Pi\left(S_u, g, M\right) = \sum_{s \in S} g\left(s\right)\left(p\left(y\left(s\right)\right) - c\left(y\left(s\right)\right)\right),$$

where $\left(y\left(s\right), p(y\left(s\right))\right)$ is the allocation selected by type s from M, which for any $s \in S$ is given by

$$u\left(s, y\left(s\right)\right) - p\left(y\left(s\right)\right) \geq u\left(s, y\right) - p(y) \text{ for all } \left(y, p\left(y\right)\right) \in M$$

with the proviso that, whenever the agent is indifferent between two or more allocations, he chooses one that yields the highest profit to the principal.

An identical definition holds for the truth, (allowing for the fact that f is a density rather than a function) where equivalently, $\Pi\left(T_v, f, M\right)$ denotes the expected profit that menu M will generate under the true type distribution.[8]

The definition of Profit-Participation Pricing is as follows:

Definition 1 *For any menu* $M = \left(\left(y', p(y')\right), ..., \left(y^k, p(y^k)\right)\right)$*, let the menu derived by Profit-Participation Pricing be* $\tilde{M} = \left(\left(y', \tilde{p}(y')\right), ..., \left(y^k, \tilde{p}(y^k)\right)\right)$ *where the product vector is unchanged and the new price vector* $\tilde{p}(y)$ *is given by*

$$\tilde{p}\left(y\right) = p\left(y\right) - \tau\left(p\left(y\right) - c\left(y\right)\right)$$

[8] At this stage, there are a number of equivalent ways to express the menu, the agent's choice, and the principal's expected profit. Perhaps, the most standard one is based on the use of a direct mechanism. For reasons that will become clear later, we prefer to use an indirect mechanism formulation in which allocations are indexed by the product.

such that

$$\tau = \sqrt{2\varepsilon}.$$

In words, using Profit-Participation Pricing the principal leaves the product component of a menu fixed, but she gives a specific profit-based discount on all products using a constant fraction. To highlight this, note that the above transformation can be equivalently expressed as:

$$\overbrace{\tilde{p}(y) - c(y)}^{\text{new profit}} = (1-\tau)\overbrace{(p(y) - c(y))}^{\text{old profit}}$$

In the rest of the analysis below, we fix the principal's model: S_u with associated probability distribution g, the cost function c, and an approximation index ε.

We now turn to an approximation lemma. For any truth and any model with an associated approximation index, pick any mechanism that contains the outside option. Using profit-participation pricing the principal can bound the difference between the principal's expected payoff generated by this mechanism under her model and the expected profit generated by the Profit-Participation discounted version of her mechanism given the agent's true behavior:

Lemma 1 *Consider a model (S_u, g) with an approximation index ε, and let M be any menu. Let $\tilde{M}$ be the menu derived through Profit Participation Pricing. Then for any truth (T_v, f) satisfying the approximation index:*

$$\Pi\left(T_v, f, \tilde{M}\right) - \Pi\left(S_u, g, M\right) \geq -2\sqrt{2\varepsilon}.$$

Proof. By the definition of the approximation index ε, we know that there exists a partition P of the true type space T with approximation index no greater than ε.

Take any menu M and compute the discounted menu $\tilde{M}$. Consider any model type s and any true type that belongs to the set associated with s, namely any $t \in J(s)$.

Suppose that a model type s is offered menu M and a true type t is offered menu $\tilde{M}$. There are two possibilities: (i) t and s choose the same product; (ii) t and s choose different products.

Case (i) is straightforward. Denote the allocation chosen by both agents by $(\widehat{y}, p(\widehat{y}))$. The only loss for the principal is due to the price discount determined by τ:

$$\tilde{p}(\widehat{y}) - c(\widehat{y}) = (1-\tau)\left(p(\widehat{y}) - c(\widehat{y})\right).$$

Focus now on case (ii). Suppose when $\tilde{M}$ is offered, t chooses an allocation y'

different from $\widehat{y}$. Because $t \in J(s)$, we know that

$$\begin{aligned} |v(t, \widehat{y}) - u(s, \widehat{y})| &\leq \varepsilon; \\ |v(t, y') - u(s, y')| &\leq \varepsilon. \end{aligned}$$

implying that the payoff difference between the two products cannot be much smaller for the true type than for the model type:

$$u(s, \widehat{y}) - u(s, y') + 2\varepsilon \geq v(t, \widehat{y}) - v(t, y') \geq u(s, \widehat{y}) - u(s, y') - 2\varepsilon. \tag{2}$$

This does not preclude, however, that the choices of the two types are different, as assumed in (ii).

Next, consider a revealed preference argument. With the *original* price vector p, the model type prefers $\widehat{y}$ to y':

$$u(s, \hat{y}) - p(\hat{y}) \geq u(s, y') - p(y'). \tag{3}$$

With the discounted price vector, type t prefers y' to $\widehat{y}$:

$$v(t, \hat{y}) - \tilde{p}(\hat{y}) \leq v(t, y') - \tilde{p}(y'). \tag{4}$$

By subtracting (4) from (3), we get that

$$\begin{aligned} &p(y') - \tilde{p}(y') - (p(\hat{y}) - \tilde{p}(\hat{y})); \\ &\geq v(t, \hat{y}) - v(t, y') - (u(s, \hat{y}) - u(s, y')). \end{aligned} \tag{5}$$

By (2), the right-hand side of (5) is bounded below by -2ε. Given the definition of $\tilde{p}$, the left-hand side of (5) can also be written as:

$$\overbrace{\tau(p(y') - c(y'))}^{\text{discount for } y'} - \overbrace{\tau(p(\hat{y}) - c(\hat{y}))}^{\text{discount for } \hat{y}}.$$

Summing up,

$$\tau(p(y') - c(y') - p(\hat{y}) + c(\hat{y})) \geq -2\varepsilon. \tag{6}$$

There are two potential sources of loss, one due to the deviation from $\hat{y}$ to y', the latter due to the price discount. The loss caused by the deviation given the above inequality is

$$p(y') - c(y') - p(\hat{y}) + c(\hat{y}) \geq -\frac{2\varepsilon}{\tau}. \tag{7}$$

The loss due to the price discount is (recalling that profit is bounded above by $\Pi_{\max}$,

which was normalized to 1),

$$\tilde{p}(y') - c(y') - (p(y') - c(y')) = -\tau(p(y') - c(y')) \geq -\tau. \tag{8}$$

Adding the above two inequalities together we get that

$$\tilde{p}(y') - c(y') - (p(\hat{y}) - c(\hat{y})) \geq -\tau - \frac{2\varepsilon}{\tau}. \tag{9}$$

We can now see the explicit trade-off between the two sources of loss: the direct-loss from discounting and the deviation-loss. By optimizing on this, we can bound their sum. In particular, if we set τ equal to

$$\arg\min_{\tau} \tau + \frac{2\varepsilon}{\tau} = \sqrt{2\varepsilon},$$

we get

$$\tilde{p}(y') - c(y') - (p(\hat{y}) - c(\hat{y})) \geq -2\sqrt{2\varepsilon}.$$

Taking expectations appropriately, we get the statement of the lemma .

The lemma contains the main intuition for why this type of approximation scheme works. Profit participation puts a bound on the loss that the principal suffers if the type space is not what she thought it was. By offering profit-based price discounts, the principal ensures that allocations that generate higher profit to her become relatively more attractive to the agent. Profit-Participation Pricing is in effect a system of local incentives. The agent becomes a sharing residual claimant on the principal's profit, and now types near model types are encouraged to choose similarly high-margin allocations as the model types.

A key feature of Profit-Participation Pricing is that there is no guarantee that types close to a model type will choose in the same way as their respective model types. The principal still does not know how often different allocations will be chosen by the agent. In fact, the principal cannot even guarantee that, when offered the discounted menu, model types will choose the allocation they were choosing previously. However, the principal knows that whichever allocation types choose with the discounted menu, the deviation from the allocation chosen by model types in the non-discounted menu cannot be very damaging to profit.

The existence of this bound is based on a trade-off introduced by Profit-Participation Pricing. First, offering a price discount leads to a loss to the principal proportional to τ. Second, the greater is the profit-based discount, the smaller is the potential loss that the principal might need to suffer due to a deviation. Setting $\tau = \sqrt{2\varepsilon}$ optimizes on this trade-off between the loss from lower prices and the loss from

deviations and establishes the above upper bound.[9]

4.2 Profit Participation Mechanism

In the previous section, we did not mention optimality. The set of alternatives and the prices were not chosen with expected profit in mind; rather we considered any menu. We now introduce our solution concept: we combine finding the optimal menu given the principal's model with modifying such a menu via Profit-Participation Pricing.

Definition 2 *The profit-participation mechanism (PPM) consists of the following steps:*

(i) Find an optimal menu $\hat{M}$ for the screening problem defined by S_u, g, Y,c;

(ii) Apply Profit-Participation Pricing to $\hat{M}$ to obtain a discounted menu $\tilde{M}$.

PPM takes the pricing problem described in Section 2 as its input and outputs a menu $\tilde{M}$. Our focus now is on the profit difference comparing two scenarios: the principal's expected profit given the true - but unknown - optimal solution and the principal's expected profit if she offers $\tilde{M}$ to the true type space. This comparison captures the approximation loss.

Formally, let M^* be any menu, including, if it exists, the menu that is optimal for the true type space.

Definition 3 *For any given menu M^*, let the PPM loss be $\Pi\left(T_v, f, M^*\right) - \Pi\left(T_v, f, \tilde{M}\right)$. We say that the PPM loss is bounded above by a number x if it is less than x.*

If there exists an optimal menu for the true type space – namely if (Y, c, T_v, f) are such that there exists a mechanism that maximizes the principal's expected payoff – then the definition above includes the optimal menu, and the PPM loss for any menu is bounded above by the PPM loss for the optimal menu.

Note that the theorem does not require the principal to know M^*, nor the expected profit that she could achieve given this mechanism, nor the true expected profit given menu $\tilde{M}$.

We can now state the main result of the paper in terms of the known parameters of our setup.

[9] There is an interesting connection between the proof of Lemma 1 and Theorem 21 of Balcan et al (2008). In their set-up the principal searches for the optimal mechanism on a discretized set of prices. One of the steps in the proof consists in analyzing the effect of offering to the agent a discretized price vector and putting a bound on the price loss that the principal may experience because the agent chooses a different product.

Theorem 1 *The PPM loss for any M^* is always bounded above by $4\sqrt{2\varepsilon}$.*

Proof. Step 1. Take any menu M^* that contains the outside option. The associated expected payoff is:

$$\Pi\left(T_v, f, M^*\right) = \int_{t\in T} f\left(t\right)\left(p\left(y\left(t\right)\right) - c\left(y\left(t\right)\right)\right) dt,$$

where the pair $y\left(t\right), p\left(y\left(t\right)\right)$ is the allocation selected by type t. The system of incentive-compatibility constraints (including, as usual, the participation constraints) is:

$$v\left(t, y\left(t\right)\right) - p\left(y\left(t\right)\right) \geq v\left(t, y\right) - p(y) \text{ for all } t \in T \text{ and all } \left(y, p\left(y\right)\right) \in M^*. \quad (10)$$

Both M^* and $\Pi\left(T_v, M^*\right)$ are unknown objects and they remain unknown in our approach.

Step 2. For the rest of the proof, fix P to be any ε-approximation partition defined for (S_u, g) and (T_v, f). The true approximation index is, by definition, not greater than ε. To clarify notation, given the bijection between model types and partition cells let's index the elements of the model type set S by $i \in \{1, ..., \#S\}$. Then for each cell $J(s_i)$ in P, we can find at least one type $t_i^* \in J(s_i)$ that generates a payoff to the principal which is at least as large as the average payoff for the cell:

$$p\left(y\left(t_i^*\right)\right) - c\left(y\left(t_i^*\right)\right) \geq \int_{t\in J(s_i)} \frac{f\left(t\right)}{g(s_i)}\left(p\left(y\left(t\right)\right) - c\left(y\left(t\right)\right)\right) dt. \quad (11)$$

Now, create a new model $\left(\bar{S}_v, \bar{g}\right)$ with the following properties. Let the new type space be defined by $\bar{S} \equiv \left(t_i^*\right)_{i=1,..\#S}$ and the model payoff be $v\left(t_i^*, y\right)$ for all $i \in \{1, ..., \#S\}$ and for $y \in Y$. Finally, let $\bar{g}(t_i^*)$ be such that $\bar{g}(t_i^*) = g(s_i)$.for all $i \in \{1, ..., \#S\}$

Suppose that the agent's true preferences were determined by $\left(\bar{S}_v, \bar{g}\right)$ rather than (T_v, f). Given menu M^*, the system of incentive compatibility constraints are as follows: for any $t_i^* \in \bar{S}$.

$$v\left(t_i^*, y\left(t_i^*\right)\right) - p\left(y\left(t_i^*\right)\right) \geq v\left(t_i^*, y'\right) - p(y') \text{ for all } \left(y', p\left(y'\right)\right) \in M^*$$

Since all the inequalities in the new system were already present in (10), it follows that, if $\left(y\left(t\right), p\left(y\left(t\right)\right)\right)$ was the allocation selected by type $t \in \bar{S} \subset T$ before, it still is in $\left(\bar{S}_v, \bar{g}\right)$.

The expected profit that the principal would obtain with M^* and $\left(\bar{S}_v, \bar{g}\right)$ would then not be lower than $\Pi\left(T_v, M^*\right)$:

$$\begin{aligned}\Pi\left(\bar{S}_v, \bar{g}, M^*\right) &= \sum_{i=1,\ldots \#S} \bar{g}(t_i^*)\left(p\left(y\left(t_i^*\right)\right) - c\left(y\left(t_i^*\right)\right)\right) \geq \\ &\geq \sum_{i=1,\ldots \#S} \int_{t \in J(s_i)} f\left(t\right)\left(p\left(y\left(t\right)\right) - c\left(y\left(t\right)\right)\right) dt \\ &= \Pi\left(T_v, f, M^*\right),\end{aligned}$$

where the inequality is due to (11).

Step 3. We now apply Lemma 1 for the first time. The lemma applies to any truth and any model that satisfies certain conditions. For the purpose of this step, we define the truth to be (S_u, g) and the model to be $\left(\bar{S}_v, \bar{g}\right)$. The approximation partition P used above is still obviously an approximation partition in this case and hence ε is still a valid approximation index.

Let M' be the menu derived by Profit-Participation Pricing with approximation index ε from M^*. By applying Lemma 1 given 'truth' (S_u, g) and 'model' $\left(\bar{S}_v, \bar{g}\right)$ with approximation index ε, we obtain

$$\Pi\left(S_u, g, M'\right) - \Pi\left(\bar{S}_v, \bar{g}, M^*\right) \geq -2\sqrt{2\varepsilon}.$$

Step 4. However, the menu M' is not optimal for the model (S_u, g). Pick a menu $\hat{M}$ that is optimal for that model:

$$\hat{M} \in \arg\max_M \Pi\left(S_u, g, M\right).$$

By definition,

$$\Pi\left(S_u, g, \hat{M}\right) \geq \Pi\left(S_u, g, M'\right).$$

Step 5. Let us apply Lemma 1 for the second time. Let the truth be (T_v, f) and the model (S_u, g). We discount $\hat{M}$ through Profit-Participation Pricing to become $\tilde{M}$. The lemma guarantees that:

$$\Pi\left(T_v, f, \tilde{M}\right) - \Pi\left(S_u, g, \hat{M}\right) \geq -2\sqrt{2\varepsilon}.$$

Summing up the above five steps:

$$\Pi\left(T_v, f, M^*\right) = [\text{expected profit for any } M^*]; \quad \text{(Step 1)}$$

$$\Pi\left(\bar{S}_v, \bar{g}, M^*\right) \geq \Pi\left(T_v, f, M^*\right); \quad \text{(Step 2)}$$

$$\Pi\left(S_u, g, M'\right) \geq \Pi\left(\bar{S}_v, \bar{g}, M^*\right) - 2\sqrt{2\varepsilon}; \quad \text{(Step 3)}$$

$$\Pi\left(S_u, g, \hat{M}\right) \geq \Pi\left(S_u, g, M'\right); \quad \text{(Step 4)}$$

$$\Pi\left(T_v, f, \tilde{M}\right) \geq \Pi\left(S_u, g, \hat{M}\right) - 2\sqrt{2\varepsilon}; \quad \text{(Step 5)}$$

and hence the profit-loss due to using $\tilde{M}$ instead of any menu M^* is bounded by:

$$\Pi\left(T_v, f, \tilde{M}\right) \geq \Pi\left(T_v, f, M^*\right) - 4\sqrt{2\varepsilon}.$$

.

The proof of the theorem constructs the bound to the PPM loss by applying Lemma 1 twice. In the first application the lemma bounds the difference between the principal's expected profit for any menu M^* and the model type profit given any model type set satisfying the approximation index ε. The second application bounds the difference between the maximal model type profit and the true profit given the discounted menu identified by profit-participation pricing given the principal's misspecified model. Taken together, the two steps bound the difference between the maximal profit and the profit obtained with the discounted version of the optimal model type menu.

Again, the bound is valid without requiring the principal to know anything beyond her model and the upper bound to the inaccuracy of the model: an approximation index ε

4.3 Discussion

Theorem 1 offers two novel lessons. First, even in a situation in which the principal is uncertain about the model, profit-participation offers a very simple way of arriving at a menu that guarantees an expected payoff that is demonstrably close to the maximal expected payoff. Second, the quality of the approximation depends on the quality of the model. As the approximation index ε vanishes, the PPM solution tends to the optimal solution. The more confident the principal is about the quality of her model, the more she can behave as if her model was correct.

The fact that theorem 1 puts an upper bound on the loss due to model misspecification has a worst-case feel, akin to strong ambiguity aversion or maximin expected utility, e.g., Gilboa and Schmeidler (1989). Hence one may wonder why

the principal should be preoccupied with worst outcomes rather than average outcomes. However, we are not interested in finding an optimal mechanism for an ambiguity-averse principal. What we find is a near-optimal mechanism for an expected payoff-maximizing principal. There may be better mechanisms given additional information– a topic we discuss in section 5 – but PPM has the important property that it is asymptotically optimal: for a large class of problems the loss is guaranteed to vanish as $\varepsilon \to 0$. Thus, the benefit of replacing our general-purpose near-optimal mechanism with one that is more tailored to the situation at hand goes to zero in the limit.[10]

If, as March and Simon (1958) argued, near optimal solution is a response to cognitive bounds, we should observe that more bounded decision-makers choose different arrangements. Within our set-up, there is a model uncertainty reducing sense in which we can perform comparative statics on cognitive.

Suppose that the principal adds model types to her existing model $S \subset S'$. In the context of the earlier example, this could be the case for example when the principal chooses to interview another type at an unexplored location at some sampling cost γ, and add this type's preferences to her model. It follows form the definition that the new approximation index is $\varepsilon' \leq \varepsilon$. Suppose that the principal can devote a budget of N to sample additional types. Then as N increases, the number of products on the menu ($\#\widehat{Y}$) increases and the discount τ used in PPM decreases. To sum up:

Corollary 1 *If the principal uses PPM, then, as N increases:*

(i) $\#\widehat{Y}$ increases (there are a larger number of items on the menu)

(ii) τ decreases (the principal prices alternatives with higher actual profit margins hence less magnanimously).

As model uncertainty rises, a principal who uses PPM to solve our problem, will use mechanisms that are simpler, as measured in the number of distinct items offered to the agent. Such simpler menus will be derived from mechanisms that employ rougher categorizations of true types into approximate types and offer items at greater discounts relative to the prices optimal for the approximate type space.

[10]There might appear to be a relation between our model uncertainty interpretation and the quantal response equilibrium of McKelvey and Palfrey (1995). We can imagine that the model used by our principal excludes some dimensions of the agent's type. This means that, as in a quantal-response equilibrium, if we take the model at face value, we still have some unexplained variability in the agent's choice. However, there are two crucial differences. First, while quantal-response equilibria operate by perturbing the players' strategies, our agent instead always plays best response. Second, quantal-response equilibrium postulates a particular form of perturbation, while our principal may not have such information.

Outside of the problem of model uncertainty, by imposing some minimal structure on the true type space - Lipschitz continuity of the agent's utility in his type - PPM can also help reduce the complexity of contracting. Conitzer and Sandholm (Theorem 1, 2004) show that finding an exact solution to the single-agent mechanism design problem we consider here is NP-complete. Our result suggests a simple way of finding an approximate solution. Partition the agent type space by using a regular grid: the approximation index is then given by the maximal cell size. Select the preferences of any single type from the cell to be its representative type. Then using this sparse and misspecified model, obtain a bound to the approximation loss by using Theorem 1.This way, the principal uses a much simpler albeit wrong model, which if used naively might generate discontinuous losses, but which under profit-participation allows for a simple trade-off between approximation loss and how complex the model need to be to general robust revenue. [11]

5 Alternative Mechanisms

As we have shown, PPM is a valid approximation scheme, but are there other approximation schemes that perform equally well or perhaps better? To address this question, we first have to note that the performance of any approximation scheme depends on the class of problems to which it is applied. According to the No Free Lunch Theorem of Optimization, elevated performance over one class of problems tends to be offset by performance over another class (Wolpert and Macready 1997). The more prior information the principal has, the more tailored the mechanism can be. For more restrictive classes of problems (e.g. one-dimensional problems with the standard regularity conditions), it is easy to think of mechanisms that perform better than PPM. Given model uncertainty, a more pertinent question is whether there are other valid mechanisms for the general class of problems we consider.

Since our results apply to a large class of multi-dimensional screening problems, defined only by the approximation index ε, we shall now ask whether there are other mechanisms, besides PPM, that work for this class of problems. We begin by defining the class of mechanisms that are based on the principal's model and modify prices given the optimal solution of this model:

Definition 4 *A mechanism is model-based if it can be represented as a two-step process where first one performs step (i) of the PPM and then, modifies the price vector $p(y)$ according to some function*

$$\tilde{p}(y) = \Psi\left(p(y), c(y), \varepsilon\right).$$

[11] For more detail see Madarasz and Prat (2010, Section 5).

The function Ψ obviously does not operate on the price of the outside option y_0, which is a primitive of the problem. We focus our attention on mechanisms that return minimal exact solutions, namely solutions where all alternatives offered are bought with positive probability.

The function Ψ can encompass a number of mechanisms. In the naive one, the principal takes the model seriously tout court, without modifying prices.

Example 1 *In the naive mechanism,*

$$\Psi\left(p(y), c(y), \varepsilon\right) = p(y).$$

In the flat discount mechanism, the principal acts magnanimously by discounting prices, but her generosity is not related to model profits:

Example 2 *In the flat discount mechanism*

$$\Psi\left(p(y), c(y), \varepsilon\right) = p(y) - \delta,$$

for some $\delta > 0$, which may depend on ε.

Finally, we can also represent the PPM in this notation:

Example 3 *In PPM*

$$\Psi\left(p(y), c(y), \varepsilon\right) = (1 - \tau)\, p(y) + \tau c(y),$$

for some $\tau > 0$, which may depend on ε.

The following definition is aimed at distinguishing between mechanisms depending on whether they contain an element of profit participation or not.

Definition 5 *A model based mechanism violates profit participation if, for a given $\varepsilon > 0$, there exists $\bar{p} > 0$ and $\bar{c} > 0$ such that for all $p' < p'' \leq \bar{p}$ and $c \leq \bar{c}$*

$$p'' - \Psi\left(p'', c, \varepsilon\right) \leq p' - \Psi\left(p', c, \varepsilon\right).$$

Under profit participation the principal shares her profits and hence her price discount that is strictly increasing in her profit. Profit participation is violated when there is a price/cost region including the origin where an increase in the principal's profit does not translate into a strict increase in the absolute value of the price discount.

The profit-participation condition is strong for two reasons: profit participation fails if a *weak* inequality is violated; it fails if the inequality is violated for *some*

prices and cost levels, not necessarily all prices and cost levels. As our theorem is negative, a strong condition means that we can exclude a larger class of mechanisms.

It is easy to see that both the naive price mechanism and the flat-discount mechanism violate profit participation (indeed, they violate it for all values of $p'' > p'$ and c). Instead, with PPM, we have

$$p'' - \Psi(p'', c) = \tau p'' + \tau c > \tau p' + \tau c = p' - \Psi(p', c) \text{ for all } p'' > p';$$

and by construction, PPM never violates profit participation.

We now show that a model-based mechanism that violates profit participation cannot be a valid approximation:

Theorem 2 *Suppose the agent's preferences are at least two-dimensional. The upper bound to the profit loss generated by a model-based mechanism that violates profit participation does not vanish as* $\varepsilon \to 0$.

Proof. Suppose that the mechanism is model-based but violates profit participation for some $\bar{p} > 0$ and $\bar{c} > 0$. Select

$$p' = \frac{1}{2}\bar{p}$$
$$p'' \in \left(\frac{1}{2}\bar{p}, \bar{p}\right)$$

Suppose that $c = 0$. For the p' and p'' chosen above it must be that:

$$p'' - \Psi(p'', 0, \varepsilon) \leq p' - \Psi(p', 0; \varepsilon) \tag{12}$$

Define $h = p'$ and $q = p'' - p'$. Consider the following problem:

$$\begin{aligned} T &= [0, 2] \\ f(t) &= \frac{1}{2} \text{ for all } t \in [0, 2] \\ Y &= [1, 2] \cup \{\bar{y}\} \cup y_0 \\ u(t, y) &= \begin{cases} h + q(t - 1 - 2|y - t|) & \text{if } y \in [1, 2] \\ h & \text{if } y = \bar{y} \\ 0 & \text{if } y = y_0 \end{cases} \\ c(y) &= 0 \text{ for all } y \end{aligned}$$

In this screening problem, types below $t = 1$ prefer a "generic" alternative $\bar{y}$. Types above $t = 1$ prefer a "personalized" alternative $y = t$.

It is easy to see that in the optimal solution of this screening problem types below 1 buy $\bar{y}$ at price h and each type $t > 1$ is offered a personalized alternative $\hat{y}(t) = t$ at price $h + q(t-1)$. The principal's expected profit is $h + \frac{1}{4}q$.

To show that the mechanism Ψ does not yield a valid approximation, we consider the following sequence of models (S_k, u_k, g_k) where $S_k \subseteq T$ and $u_k(s,y) = v(s,y)$ for all k: sets with associated stereotype probability distributions:

$$\begin{cases} S_0 = \{0,1,2\} \\ g_{S_0}(0) = g_{S_0}(2) = \frac{1}{4}; g_{S_0}(1) = \frac{1}{2} \end{cases}$$

$$\begin{cases} S_1 = \left\{0, \frac{1}{2}, 1, \frac{3}{2}, 2\right\} \\ g_{S_1}(0) = g_{S_1}(2) = \frac{1}{8}; g_{S_1}\left(\frac{1}{2}\right) = g_{S_1}(1) = g_{S_1}\left(\frac{3}{2}\right) = \frac{1}{4} \end{cases}$$

$$\vdots$$

$$\begin{cases} S_n = \left\{0, \frac{1}{2^n}, \ldots, 1, 1+\frac{1}{2^n}, \ldots, 2\right\} \\ g_{S_n}(0) = g_{S_n}(2) = \frac{1}{2^{n+2}}; g_{S_n}(s) = \frac{1}{2^{n+1}} \text{ for all other } s \end{cases}$$

$$\vdots$$

Given the prior f, a valid approximation index for stereotype set S_n is $\varepsilon_n = 3q(\frac{1}{2^{n+1}})$.

Hold n fixed. The exact solution of the screening problem for (S_n, u_n, g_n) involves offering $\bar{y}$ at price $p(\bar{y}) = h$ as well as a vector of alternatives identical to the vector of types $\left\{1 + \frac{1}{2^n}, \ldots, 2 - \frac{1}{2^n}, 2\right\}$, each of them priced at $p(\hat{y}(s)) = h + q(s-1)$. The minimum price is h, while the maximum price is $h + q$. Hence, by our definition of h and q, all prices are between p' and p''.

The mechanism returns the following prices:

$$\begin{aligned} \tilde{p}(\bar{y}) &= \Psi(p(\bar{y}), 0; \varepsilon_n) = \Psi(h, 0; \varepsilon_n) \\ \tilde{p}(\hat{y}(s)) &= \Psi(p(\hat{y}(s)), 0; \varepsilon_n) = \Psi(h + q(s-1), 0; \varepsilon_n) \end{aligned}$$

Now recall that by definition $\Psi(p, c)$ violates profit participation. Hence, for any $t \in [1,2]$,

$$h + q(s-1) - \Psi(h + q(s-1), 0; \varepsilon_n) \leq h - \Psi(h, 0; \varepsilon_n) \tag{13}$$

Now take any type $t \in [1,2]$ which is not a model type (a set of measure 1 for every S_n) and consider his choice between the allocation meant for any stereotype $s \in [1,2]$ modified by Ψ ($\hat{y}(s)$ at price $\Psi(h + q(s-1), 0; \varepsilon_n)$) and the allocation meant for model types below $t = 1$ ($\bar{y}$ at price $\Psi(h, 0; \varepsilon_n)$). If he buys $\hat{y}(s)$ he gets payoff

$$h + q(t - 1 - 2|s - t|) - \Psi(h + q(s-1), 0; \varepsilon_n)$$

If he buys $\bar{y}$ he gets utility

$$h - \Psi\left(h, 0; \varepsilon_n\right)$$

He chooses $\hat{y}\left(s\right)$ only if

$$q\left(t - 1 - 2\left|s - t\right|\right) - \Psi\left(h + q(s-1), 0; \varepsilon_n\right) \geq -\Psi\left(h, 0; \varepsilon_n\right)$$

which, if one subtracts (13) from it, implies:

$$q\left(t - 1 - 2\left|s - t\right|\right) - q\left(s - 1\right) \geq 0,$$

which can be re-written as

$$t - s \geq 2\left|s - t\right|,$$

which is always false. Hence, all types that are not model types choose $\bar{y}$ rather than a nearby personalized alternative. For any S_n, the expected profit of the principal if she uses Ψ is h.

Hence, the limit of the profit as $n \to \infty$ ($\varepsilon_n \to 0$) is still h, which is strictly lower than the profit with the maximal profit with the true type, which, as we saw above, is $h + \frac{q}{2}$. .

The proof of the above Theorem 2 proceeds by constructing a straightforward class of problems with binding non-local constraints. In particular, we assume that the product space includes a generic inferior product and a continuum of type-specific products. In the optimal solution, a non-zero measure of types face a binding incentive-compatibility constraint between a personalized alternative and the generic alternative. Hence, all types nearby a model type strictly prefer the generic alternative to the model type's optimal allocation. This creates a non-vanishing loss for the principal.

The intuition behind the Theorem has to do with the knife-edge nature of mechanisms that do not include profit participation. In the exact solution of the principal's model there is a binding constraint (IC or PC) for every alternative offered. The profit-participation mechanism – and analogous schemes – relax these constraints in the right direction. They add slack to constraints that ensures that the agent does not choose alternatives with a lower profit. Mechanisms without profit-participation return a price vector that still displays binding constraints – or don't even satisfy those constraints. When profit participation is violated types near a model type might choose different profit-margin allocations. If only local constraints are binding, the magnitude of such misallocations vanishes as $\varepsilon \to 0$. In multi-dimensional screening problems, however, non-local constraints will typically bind in optimum. Here, the magnitude of the misallocation does not vanish even as $\varepsilon \to 0$.

As one would expect, the theorem does not apply if the type space is one-dimensional. For instance, in the Mussa and Rosen's (1978) set-up, the naive mechanism is a valid approximation: only local IC constraints are binding and deviations are guaranteed to get smaller as the model type space becomes finer. As we argued in the introduction though, most economically relevant screening problems are likely to be multi-dimensional and hence Theorem 2 provides useful general guidance.

Given the nature of the statement, the proof is by example. One may object that the class of problems used in the proof has zero measure within the set of all possible screening problems. However, all we need is a class of problems where non-local constraints are binding in optimum for a positive measure of types. As binding non-local constraints are an endemic feature of multi-dimensional screening problems (Rochet and Choné 1998, Rochet and Stole 2003), endless other examples could be found in which the optimal solution is not robust to small perturbation of the agent's preferences.

Theorem 2 says that if a mechanism performs at least as well as PPM, it must be either very similar to PPM in that it relies on profit participation, or very different in that it is not even model based. Hence, the theorem points to three interesting questions that we leave for future research. Are there other profit-participation mechanisms that perform better than PPM? Are there non-model based mechanisms that perform better than PPM? Are there not-overly-restrictive classes of screening problems where the Naive Mechanism is guaranteed to be a valid approximation?

6 Conclusion

We consider a principal who faces a screening problem but faces model uncertainty and operates on a potentially misspecified model of the agent's behavior We characterize the upper bound to the expected loss that the principal incurs if she uses the profit-participation mechanism. We show that this loss vanishes smoothly as the model type space tends to the true one. We also prove that this is not true for similar model-based mechanisms that do not contain a profit participation element.

The economic insight of this paper is that a principal who operates on the basis of only an approximate type space cannot just ignore the misspecification error, but she can find a simple and economically intuitive way to limit the damage. It would be interesting to know whether this insight holds beyond our set-up. A strength of our approach is that it does not make specific functional or distributional assumptions and applies to settings where the potential allocation space is very large.

Our analysis has a number of limitations that future research could address. First, we assume that the principal's cost depends only on the product characteristics, but not on the type of the agent (as in insurance problems). Second, we assume that there is only one agent (or a continuum thereof). It would be interesting to extend the analysis and explore the role of profit-participation in implementing near-optimal social choice correspondences in environments with multiple agents, perhaps linking it to notions of robustness to small perturbations in such contexts (Meyer-ter-Vehn and Morris 2011). Third, one could explore environments where payoffs are not quasilinear.

References

[1] Armstrong, Mark. (1996): Multiproduct Nonlinear Pricing. *Econometrica* 64(1): 51–75.

[2] Armstrong, Mark (1999): Price Discrimination by a Many-Product Firm. *Review of Economic Studies* 66: 151–168.

[3] Armstrong, Mark and Rochet Jean-Charles. (1999): Multi-dimensional Screening: A User's Guide. *European Economic Review* 43: 959-979.

[4] Balcan, Maria-Florina, et al. (2008): Reducing mechanism design to algorithm design via machine learning. *Journal of Computer and System Sciences* 74: 1245-1270.

[5] Battaglini Marco and Rohit Lamba (2012): Optimal Dynamic Contracting. Mimeo, Princeton University.

[6] Box, George and Draper Norman (1987): Empirical Model Building and Response Surfaces. John Wiley.

[7] Chassang, Sylvain (2013): Calibrated Incentive Contracts. (forthcoming) *Econometrica*

[8] Chu, Chenghuan Sean, Phillip Leslie, and Alan Sorensen. (2011): Bundle-Size Pricing as an Approximation to Mixed Bundling. *American Economic Review* 101(1): 263-303.

[9] Conitzer, Vincent, and Thomas Sandholm.(2004): Self-interested Automated Mechanism Design and Implications for Optimal Combinatorial Auctions In Proceedings of the 5th ACM Conference on Electronic Commerce (EC-04), pp. 132-141, New York, NY, USA.

[10] Gabaix, Xavier (2010): A Sparsity-Based Model of Bounded Rationality. Mimeo, NYU.

[11] Gilboa, Itzhak, and David Schmeidler. (1989): Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* 18(2): 141-153.

[12] Hansen, Lars Peter, and Thomas J. Sargent. (2010): Wanting Robustness in Macroeconomics. Working paper, NYU.

[13] Hotelling, Harod. (1929) Stability in Competition. *Economic Journal 39:*.41-57.

[14] Madarasz, Kristof and Andrea Prat (2010): *Screening with an Approximate Type Space*. CEPR Discussion Papers 7900.

[15] McKelvey, Richard, and Thomas Palfrey (1995): Quantal Response Equilibria for Normal Form Games. *Games and Economic Behavior* 10: 6–38.

[16] March, James, and Herbert Simon.(1958): *Organizations.* New York: Wiley.

[17] Meyer-ter-Vehn, Moritz and Stephen Morris.(2011): The Robustness of Robust Implementation. *Journal of Economic Theory* 146, 2093–2104.

[18] Mussa, Michael, and Sherwin Rosen. (1978): Monopoly and Product Quality. *Journal of Economic Theory* 18: 301–317.

[19] Nisan, Noam and Ilya Segal. (2006): The Communication Requirements of Efficient Allocations and Supporting Prices. *Journal of Economic Theory* 129(1): 192-224.

[20] Reny, Philip J. (2011). Strategic Approximations of Discontinuous Games. *Economic Theory*, 48: 17-29.

[21] Rochet, Jean-Charles, and Philippe Choné. (1998): Ironing, Sweeping, and Multidimensional Screening. *Econometrica* 66(4): 783–826.

[22] Rochet, Jean-Charles, and Lars Stole. (2003): The Economics of Multidimensional Screening. in *Advances in Economics and Econometrics*, Vol 1, eds. M. Dewatripont, L.P. Hansen, and S. Turnovsky, Cambridge.

[23] Rosen, Sherwin (1974): Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy* 82(1), 34-55.

[24] Ross, David. (2005): An Elementary Proof of Lyapunov's Theorem. *The American Mathematical Monthly* 112(7): 651-653.

[25] Salop, Steven (1979): Monopolistic competition with outside goods. *The Bell Journal of Economics* 10 (1): 141–156

[26] Walley, Peter. (1991): *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.

[27] Wilson, Robert B. (1993): *Nonlinear Pricing*. Oxford University Press.

[28] Wolpert, David H. and William G. Macready. (1997): No Free Lunch Theorems for Optimization. *IEEE Transactions on Evolutionary Computation* 1(1): 67-82