Kristóf Madarász and Andrea Prat

Sellers with misspecified models

Article (Accepted version)
(Refereed)

Original citation:
Madarász, Kristóf and Prat, Andrea (2016) Sellers with misspecified models. Review of Economic Studies, 84 (2). pp. 790-815. ISSN 0034-6527
DOI: 10.1093/restud/rdw030

Sellers with Misspecified Models[*]

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

Abstract

Principals often operate on misspecified models of their agents' preferences. When preferences are such that non-local incentive constraints may bind in the optimum, even slight misspecification of the preferences can lead to large and non-vanishing losses. Instead, we propose a two-step scheme whereby the principal: (i) identifies the model-optimal menu; and (ii) modifies prices by offering to share with the agent a fixed proportion of the profit she would receive if an item were 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, be an approximation of the truth at best. 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 would such a contract differ from the contract she would offer if her model was exactly true?

Our paper attempts to answer these questions in the context of one of the classic problems in 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 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 revisit the general screening problem but assume that the principal

[*] An earlier version of this paper was circulated under the title "Screening with an Approximate Type Space." We thank the anonymous referees and the Editor, as well as 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

(seller) does not know the true distribution of the agent's (buyer's) preferences. Instead, the principal faces model uncertainty and has access only to a misspecified model. Crucially, this uncertainty is such, that her model may misspecify not only probabilities, but also preferences, in that, her model may not enlist exact preference specifications (types) that can occur in reality. The seller knows that her model is potentially wrong – for example, because it is typically simpler than reality – and has a sense – to be formalized shortly – of how much her model could deviate from the true description of the buyer's behavior. Can such a seller guarantee herself an outcome that is not much worse than what she could expect if she had access to the true description of the buyer's preferences?

For instance, consider a situation in which the agent's willingness to pay for an object depends on a list of attributes: income, age, family background, profession, etc. This list may be long, though, and the principal may operate on the basis of a coarser model that explicitly includes the effect of the most important factors, leaving some characteristics unmodeled or modeled only in an approximate way. Nevertheless, she wishes to design a mechanism using her model, which is robust to all possible misspecifications of the impact of minor attributes on the shape of these preferences.

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

Finally, outside of the problem of model uncertainty, even if the principal had access to the correct model of the agent's behavior, when taking all factors into account, identifying the optimal 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 that we consider here is NP-complete. Hence, the principal might want to operate on the basis of a simpler model that listed fewer contingencies than what can occur in reality. Our method allows to describe a trade-off between adopting such a coarser representation of the type space and facing a tolerable loss relative to operating on the full type space, even in the large class of domains where naively relying on a slightly coarser representation will lead to a much greater loss.

Our paper proceeds in three steps. First, we demonstrate that designing a contract as if the principal's misspecified model was correct leads to potentially large losses, and these losses need not vanish even as the distribution of preferences described in the model gets arbitrarily close to the true distribution of preferences. Second, we identify 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. Lastly, we show that any contract that is based on the solution of the principal's model and, given our general class of problems, will always be robust to small preference misspecifications must be similar to the contract identified by this procedure. The rest of this section summarizes these results in informal terms.

To introduce our results, we first discuss model misspecification and a way of measuring the quality of the principal's model. In our setup, the principal's model is a finite approximation of the agent's true preferences, and each model type can represent a potentially uncountable set of nearby preference profiles. To express the distance between any two types' preferences we consider the maximal difference between their respective willingnesses to pay for any given product. We then define an approximation index of a model to be a valid upper bound on the distance between the preferences of any model type and any of the true types it represents. An approximation index provides an informationally minimal restriction on what the true problem might be, but ensures that the probability that the agent's true type is no further away from a given model type than this index is at least as high as the probability that the principal's model assigns to this model type.

For any model and any value of the approximation index, there is a very large class of true preference profiles, and distributions over them, whose distance from the model is weakly less than this index - where we can also allow for a probabilistic and local interpretation of this statement. Correspondingly, any such index also allows for the presence of unforeseen contingencies, that is, the presence of true types whose exact preferences may not be listed in the model's support.

In the spirit of March and Simon (1958), the near-optimal contract we are seeking can rely only on information that is available to the principal. The menu offered to the agent will, thus, depend only on the principal's misspecified model and the approximation index, and on no other information about the true types. For any true type space, 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 in this type space from the menu computed by her model-based algorithm. A near-optimal solution ensures that, given any true environment satisfying the approximation index, the payoff that this solution generates in that environment is not much worse than what the principal could achieve if she knew the truth. It puts a bound on the approximation loss and guarantees that this loss always vanishes smoothly as the approximation index goes to zero.

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 that the principal offers him 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 the following fact. Given the exact solution of the screening problem, the principal's payoff function is discontinuous exactly at the equilibrium allocation: this is true because profit maximization implies a system of binding incentive-compatibility constraints. Importantly, outside of the case in which preference heterogeneity is such that the single-crossing property holds and only local incentive constraints bind, binding incentive constraints may well be non-local (Wilson, 1993; Armstrong, 1996; Rochet and Choné, 1998; Armstrong and Rochet, 1999). In the presence of such non-local incentive constraint, a type is indifferent between his allocation and the allocation of another type with distinctly different preferences. This fact makes dealing with preference misspecification challenging: here, even a slight perturbation of a payoff type might lead to a large change in equilibrium choice behavior and affect the principal's payoff discontinuously.

Given our setup and measure of misspecification, we first illustrate, in the context of a simple but economically relevant example, the profit loss for a principal who naively behaves as if her misspecified model was correct, (Section 3). She simply computes the model-optimal menu and offers it to the agent. Here, if it is guaranteed that only local incentive constraints bind, the principal's payoff from naively operating on the model converges to the payoff she would realize if she optimized based on the correct description of the problem. In contrast, if non-local incentive constraints can bind in the solution, the profit loss may remain strictly positive, even as such misspecification goes to zero. A naive principal, now, experiences a potentially discontinuous loss when moving from the case where her model is exactly correct to the case where it is only almost correct. In the presence of binding non-local constraints, a small perturbation of preferences, relative to those described in the model, can create large changes in equilibrium behavior and might then cause large losses for the principal.

One might try to address such non-robustness due to small preference perturbations by finding sufficiently restrictive conditions that guarantee that only local incentive constraints bind. Importantly, this works well, for example, in the classic setting of Mussa and Rosen (1978), where the single-crossing property is satisfied and the problem of non-vanishing losses does not arise. In contrast, when binding constraints are non-local, misspecification of preferences causes the naive approach to fail. Such non-local constraints are present in a great variety of intuitive and potentially key economic settings, and are characteristic of 'multi-dimensional' problems, e.g., Rochet and Stole, 2003.[1] In fact, Section 3 illustrates that an environment in which only local constraints bind can be very close to one in which binding constraints are non-local, and still yield very different results in terms of robustness to misspecifications.

Thus, our goal is not to identify a contract or a mechanism that works very well in specific environments in which certain preference restrictions hold exactly, but, rather, to find one that produces an acceptable outcome for a large class of screening problems even in the presence of preference misspecifications – one that uses general preferences, cost functions, and type spaces. This will ensure that the contract will be robust to violations of exact preference restrictions. Indeed, our second result (Theorem 1) identifies an approximation scheme that works in any smooth type space. We call our solution concept a *profit-participation mechanism.* Given a model type space and its corresponding approximation index, we define the profit-participation mechanism in 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 her model 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

[1] Relatedly, work by Jehiel, Meyer-Ter-Vehn, Moldovanu, and Zame (2006) and Gershkov, Goeree, Kushnir, Moldovanu, and Shi (2013) spot key effects of multi-dimensionality on the properties of optimal mechanisms. Battaglini and Lamba (2012) consider a dynamic version of the one-dimensional screening model with imperfect type persistence and show that non-local incentive constraints also bind there.

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 of these are unknown to the principal. We are, nevertheless, able to prove the existence of an upper bound to this difference and show that it is a smooth decreasing function of the approximation index. For any screening problem, the upper bound on the performance loss vanishes smoothly with the square-root of the approximation index, and, hence, the loss goes to zero as the model tends to the truth.

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 claimant of her model profit. This guarantees that allocations that yield more profit in the model-optimal menu become relatively more attractive to the agent. A true type close to a model type, may still not choose the product that is meant for this model type. At the same time, even if he chooses a different product from the modified menu, this must now be less damaging than before – the profit difference is bounded by an amount that is strictly decreasing in the discount.

While a profit-related discount is beneficial because it puts an upper bound on 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 the loss from lower prices and the gain from increased robustness. We show that as the approximation index decreases a given upper bound on 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.

Finally, one may wonder whether there are other ways of achieving a generally valid approximation besides the one we propose. Our final result shows that profit-sharing is a necessary feature of any valid approximation scheme within a large class of mechanisms given the general model uncertainty, allowing for small unforeseen contingencies, we consider. This result does not cover only the naive mechanism mentioned earlier. It applies to any model-based pricing mechanism – namely, any scheme that begins with step (i) of the profit-participation mechanism. In other words, it applies to all algorithms that start with the solution to the principal's model and then change the model-optimal prices according to some fixed rule that depends only on the approximation index. Theorem 2 shows that if this rule does not satisfy profit participation, then it cannot be a generally robust approximation: the profit loss need not vanish as the approximation index goes to zero. This means that if there are model-based pricing mechanisms that generally 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-based discounting.

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 appropriately. A principal who has only an imperfect model of the agent's behavior can start by taking her simpler model at face value and finding its solution. However, the resulting allocation is not robust to small preference misspecifications. To make sure that such 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 to part of the profit 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 paper is structured as follows. Section 2 introduces the screening problem and defines the notion of an approximation index. Section 3 describes a simple setting with heterogeneous tastes for different specifications of a durable good, such as a car, and demonstrates the presence of non-vanishing losses. Section 4 discusses the positive result: we develop profit-participation pricing; establish an approximation bound (Lemma 1); and show that the profit-participation mechanism is a valid approximation scheme (Theorem 1) under model uncertainty. 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 setup presents a form of discontinuity that is due to the strategic interaction between the principal and the agent. The principal's optimal payoff may be discontinuous when the agent's behavior is misspecified. As our example in Section 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 the 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]

[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

Our work is related to Chassang (2013) and Carroll (2013). Chassang 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 benchmark contract with a number of attractive properties. The performance bound is independent of the underlying process for returns. Carroll (2013) considers a moral hazard problem in which the principal knows some, but not all, of the actions that are available to the agent and shows that, under general circumstances, the only contract that is robust to all possible actions is linear. While the present paper studies a different class of problems, we also find that the robust mechanism requires the principal to share a fixed proportion of her profit with the agent. This share goes to zero as the principal's model becomes more and more accurate.

Also related is Bergemann and Schlag (2011), who study monopoly pricing of a single product when the seller faces uncertainty about the distribution of buyer valuations; they show that given a minimax regret criterion, the optimal pricing policy is stochastic. More broadly, our paper is related to Gabaix (2014), who considers a single decision maker who faces an explicit attention cost for each dimension to which she pays attention; thus, she decides to pay only a form of partial attention to each given dimension based on its importance given a quadratic utility approximation.

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 compact 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 options, i.e., a set of alternative-price pairs. We denote such a feasible menu by M and assume that it always contains the outside option y_0, whose price p_0 and cost, $c(y_0)$, are assumed to be zero. Once a menu is offered, the agent is asked to choose exactly one item from it.

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

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

where the above specification follows Rochet and Choné (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* (reality) and the *model.* These two will be linked by an *approximation index* and will give rise to the approximation problem that we analyze.

Truth (Reality)**.** The agent's true preferences depend on his private type t from a compact set $T \subseteq \mathbb{R}$, drawn according to an absolutely continuous density $f(t)$.[4] In particular, the agent's real payoff is his type-dependent valuation of the object y net of the transfer price to

than on model uncertainty.

[4] To simplify some of the exposition, we assume a continuum of true types. Our analysis, however, would remain valid in the case where T was a purely abstract set with only a *finite* number of elements.

the principal:

$$v\left(t,y\right)-p$$

We refer to T and v (T_v henceforth) as the true type space. Thus, the principal faces a single-agent mechanism design problem which - given a fixed technology (Y,c) - can be summarized by (T_v,f).[5]

We assume that there is a finite upper bound on the principal's profit $D=\Pi_{\max}$ with

$$D=\sup_{y\in Y,t\in T} v\left(t,y\right)-c\left(y\right),$$

and denote the supremum of the principal's expected profit over all feasible menus - menus containing the outside option - by $\Pi^*\left(T_v,f\right)$. The principal's expected profit is then bounded from below by zero and from above by $\Pi_{\max}$. We scale v and c such that $\Pi_{\max}$ is normalized to be 1. Let Θ denote the set of all true screening problems that satisfy the above assumptions.

Seller's Model. Our key point of departure is that the principal facing the above screening problem does not have access to the true type space. 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

$$u\left(s,y\right)-p$$

with associated probability distribution function $g\in\Delta S$. We refer to S and u (S_u henceforth) as the set of model types or, equivalently, as the approximate type space.[6] The principal's model is then denoted by (S_u,g).

Approximation Index. We now introduce our measure of the model's quality, which expresses the degree of misspecification between the model and the truth. This measure satisfies two important conditions. First, it is a simple scalar that reflects a 'distance' between the model and the truth that will go to zero as the model tends to the truth. This will corresponds to a minimal form of information that the principal can have about how misspecified her model of the agent's preferences potentially is. Second, it has a maximal-distance element, which allows us to find upper bounds on the profit loss. This worst-case aspect of the measure guarantees that as our measure goes to zero, any other non-maximal-distance measure would go to zero, too.

Given any truth (T_v,f) and model (S_u,g), the 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. Each cell is associated with a separate model type, 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 (non-empty) set

[5] All our results hold if we also assume that $v(t,y)$ is continuous in t and y.

[6] To simplify notation, we also assume that the same upper-bound D continues to apply here; that is, $\sup_{y\in Y,s\in S} u(s,y)-c(y)\le D=1$.

of all approximation partitions.[7]

2. For each cell J_s of a given $\mathcal{P}$ in Γ, define 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* is $\varepsilon_{\text{true}} = \inf_{\mathcal{P} \in \Gamma} d(\mathcal{P})$. Note that $\varepsilon_{\text{true}}$ exists and, given that S is finite, assuming that v is not locally constant, it is also strictly positive.

4. Let a valid *approximation index* ε be any number strictly greater than $\varepsilon_{\text{true}}$. Define an *$\varepsilon$-approximation partition* of the truth to be any approximation partition $\mathcal{P}$ with $d(\mathcal{P}) \leq \varepsilon$.

Our measure of misspecification is based on the distance between preferences induced by the sup-norm. This corresponds to the maximal difference between two types' respective willingnesses to pay for any given product. Given this distance, for each type we can consider the set of preference specifications that are within ε-distance of this type's preferences. If there is an assignment of true types to model types, such that, for each model type s, the (true) probability that the agent's preferences are within ε-distance of this model type's preferences is at least $g(s)$, the probability that the principal's model attaches to s, then an ε-approximation partition exists. Here, all true types are within ε-distance of model types with the appropriate probabilities. If such an ε-approximation partition exists, we say that the model is an ε-approximation of the truth, or, alternatively, that the truth is within ε-distance of the model.[8]

Our approximation index links probability distributions over preferences with non-common supports. It captures the intuitive idea of wanting robustness with respect to small vanishing misspecifications of the underlying preferences. Under this measure, even if the model fails to describe the true preferences exactly, it can still be close to the truth as long as the model preferences are close to the true preferences with the appropriate probabilities. In the next section, we then explore the robustness of the naive mechanism with respect to this measure. There are, of course, other metrics over probability distributions that penalize non-common supports more. Under those, any small perturbation of the underlying preferences may correspond to a larger deviation in the corresponding distribution, making robustness under such considerations potentially easier to satisfy. Naturally, since our solution will guarantee a limited payoff loss to the principal when using any model distribution whose deviation from the true distribution remains limited under our measure, the same limited loss applies even when such a distribution has a different distance from the truth under a different metric.

In our setting, there are a finite number of model types and a continuum of true types. This is not essential, the logic of our results would continue to hold if T was finite (even if it had

[7] Since f is absolutely continuous and T is Lebesgue measurable, it follows from Lyapunov's convexity theorem that such a partition 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 $f(J(s_1)) \leq f(T)$, and 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.

[8] There is a connection between our measure of a true approximation index expressing the distance between the 'model' and 'reality' and the Lévy-Prokhorov metric on probability distributions, Prokhorov (1956), when considering the 'model' and 'reality' as probability distributions over the space of utility functions over the set of products Y where this space is endowed with the sup norm as above.

the same cardinality as S). Instead, what is key for our approach is that, given any positive approximation index, the principal's model may still misspecify preferences and not contain in its support the exact description of all preference specifications that can occur in reality. For any positive ε the agent's true types may lie outside of the model's support. Importantly, while any positive approximation index allows for such preference misspecification, it also bounds its maximal possible extent. It also implies that any true type is within ε-distance of a model type. Furthermore, as ε goes to zero, any difference between the preferences of a model type and that of the corresponding true types vanishes.

The possibility of such small unforeseen contingencies plays a key role in our paper. Our approach, thus, differs from considering a principal who knows the true state space and is uncertain only about the prior over this state space, e.g., Gilboa and Schmeidler (1989). Holding the set of types constant, and slightly misspecifying only probabilities per type, will *not* lead to the kind of non-robustness problem we describe in the next section. Instead, it arises if the principal cannot be certain to describe exactly all preference specifications (contingencies) that can occur in reality, and, thus, her model may misspecify not only probabilities, but preferences as well. Correspondingly, the solution that we are looking for, is then such, that it will be robust to slight preference misspecifications as well.

Approximation Problem. The principal knows a misspecified model along with an approximation index ε. The goal, then, is to obtain a solution that is robust to *any* possible misspecification satisfying this approximation index. Such a solution guarantees that for *any* problem that is within ε-distance of the principal's model, the principal faces a limited loss. In particular, that given any such (T_v, f), the difference between the principal's maximal expected payoff given (T_v, f) and the expected payoff she receives using her model-based algorithm when the true environment is (T_v, f) is bounded from above by some function of ε, with the property that the bound smoothly goes to zero as $\varepsilon \to 0$.[9]

2.1 Discussion

Our principal is constrained to operate on a misspecified model. Nevertheless, she is willing to take a stand on how far her model could be from the truth. The principal does not know (T_v, f), but knows that her representation (S_u, g) thereof is such that $\varepsilon_{true} \leq \varepsilon$. She knows that her model may fail to list contingencies that can occur in reality, but cannot specify which ones. The approximation index imposes an informationally minimal restriction on what the true type space could be. For any given model, there is a potentially very large class of true preference specifications and distributions over these that satisfy the approximation index.[10]

Why does the principal use an approximate model? In many situations, her understanding of the agent's behavior might not match the agent's true behavior exactly. The principal might

[9] The analysis can be extended to situations in which the principal is not fully certain that $\varepsilon_{\text{true}} \leq \varepsilon$. If the principal thinks that there is a probability δ that $\varepsilon_{\text{true}} > \varepsilon$, one can 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 in which 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.

[10] Related uncertainty is considered by Bergemann and Morris (2005) in a multi-agent setting, but there, uncertainty concerns the agents' beliefs and higher-order beliefs of each other.

not be able to surely and correctly describe all contingencies that could occur in reality. In other words, she may face unforeseen contingencies. She may operate under weaker epistemic conditions where she cannot be certain to describe the environment exactly. Instead, her representation of the environment may misspecify types or lump different types together. At the same time, she knows an upper bound on how misspecified her model might be, that is, she knows an approximation index ε.

The complexity of the problem provides an important reason for operating on a coarser, and, hence, misspecified model. Even if the principal could correctly enlist *all* the possible contingencies that can occur in reality this may be prohibitively costly. Specifically, as Conitzer and Sandholm (2004) show, the screening problem that we study here is NP-complete. In the absence of an algorithm that can identify a solution for the screening problem that is polynomial in the size of the input, operating on richer and richer type spaces increases the computational (and communicational) burden associated with the mechanism very fast.

To illustrate the complexity of the problem, consider a finite T and Y, and suppose that $\#Y = \alpha\#T$, where $\alpha > 1$ and # refers to the cardinality of the set. Note, that if the principal had access to the truth under the Revelation Principle, she could always solve the mechanism design problem in two stages: (i) for each possible allocation of alternatives to types, see if it is implementable, and, if it is, compute the profit-maximizing price vector; (ii) given the maximized profit values in (i), choose the allocation with the highest profit. While each step in (i) is a linear program, the number of allocations that we must consider in (i) is as high as $\#Y^{\#T}$. Instead, the reduction in complexity when moving from a larger set of types $\#T$ to a smaller set of types $\#S$, as in our geographic example, $\#Y^{\#T} - \#Y^{\#S}$, can be very substantial. Hence, a principal who operates on a coarser model which lists a smaller number of distinct preference specifications (types) can greatly reduce the burden associated with finding an acceptable solution. For a more detailed formal discussion of the motivation in terms of computational complexity, see Madarász and Prat (2010).

In a variety of settings, the principal may be able to improve the quality of her model at some cost. For instance, in the geographical example, she could obtain the exact location of agents. She could also decide to expand the set of preference specifications considered in a way that reduced the extent of potential misspecification. Our result can help the principal decide whether to incur the additional cost: the bound on the profit loss that our paper identifies provides a measure of the potential benefit of improving the model. We return to these points in Section 4.

3 Example

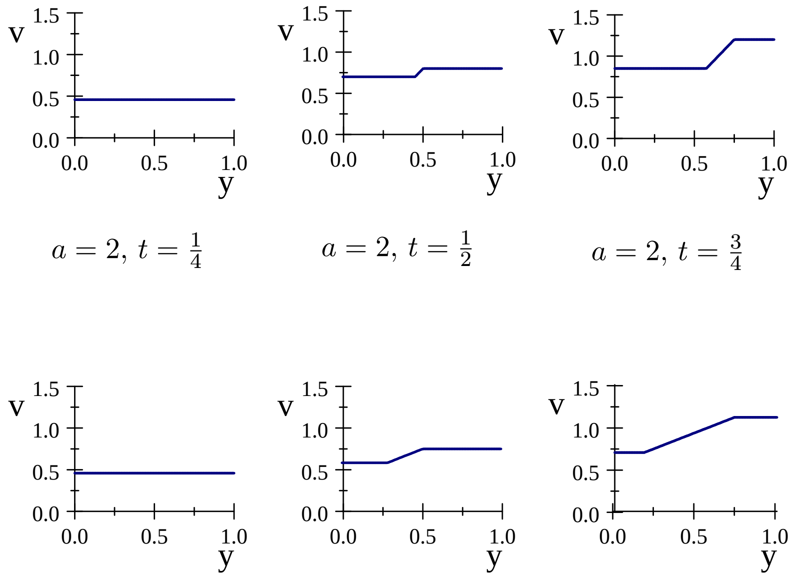
We now present an example to illustrate the problem that can arise when the principal utilizes a (slightly) misspecified model. This example also allows us to provide intuition about the source of this problem and why our solution concept of a profit-participation mechanism, formally introduced in Section 4, will address it, thus, motivating Theorems 1 and 2.

Consider a monopolist selling different specifications of a durable product - e.g., a variety of cars - to consumers with unobserved preference heterogeneity. The agent's true type t is drawn uniformly from $T = [0,1]$.[11] The agent's true valuation of a product $y \in [0,1]$, each produced by the monopolist at cost 0, is given by $v\,(t,y) = \max\{w\,(t,y)\,,\bar{w}\,(t)\}$.

To motivate these preferences, note that a consumer may need a car, *any* kind of car, for a practical reason such as commuting to work. This is captured by the 'pragmatic' type-specific value $\bar{w}\,(t)$. At the same time, he may also have a taste for a particular model y, maybe a sportscar or a 4x4, so he is willing to pay more for a subset of cars that have additional 'aesthetic' value, $w\,(t,y) > \bar{w}\,(t)$. Specifically, assume that

$$w(t,y) = (b+1)t + \min\{0, a\,(y-t)\} \text{ and } \bar{w}\,(t) = bt + k,$$

where $k > 0$, $a > 0$ and $b \in (0,1)$. The pragmatic value of a car is $bt+k$. The aesthetic value is increasing in quality y, for $y < t$, and stays constant thereafter. Both the pragmatic value and the best aesthetic value are increasing in the agent's type, though the latter increases faster.



$a = 2,\ t = \frac{1}{4}$ $a = 2,\ t = \frac{1}{2}$ $a = 2,\ t = \frac{3}{4}$

$a = \frac{3}{4},\ t = \frac{1}{4}$ $a = \frac{3}{4}, t = \frac{1}{2}$ $a = \frac{3}{4},\ t = \frac{3}{4}$

Figure 1

Figure 1 above depicts the valuation of three possible types t and two possible values of a. Lower types value the product only for pragmatic reasons. Higher types are willing to pay more for sufficiently higher qualities, but very low quality products give them only pragmatic value. (For all figures in this section, we picked $k = 1/3$ and $b = 1/2$.)

A key feature of this example is that the single-crossing condition on preferences

$$v\left(t'',y''\right) - v\left(t'',y'\right) \geq v\left(t',y''\right) - v\left(t',y'\right) \text{ for all } t'' > t' \text{ and } y'' > y'$$

[11] The fact that the set of true types is a continuum is not necessary. The same points discussed here can hold even if the cardinality of the model set and the true set are the same.

holds if $a < 1$, and is violated if $a > 1$.[12] To see this, note that for any y, the derivative $v_y(t, y)$ (where it exists) is given by

$$a \text{ if } y \in \left(\frac{(a-1)\,t + k}{a}, t\right) \text{ and } 0 \text{ otherwise.}$$

If $a < 1$, the interval $\left(\frac{(a-1)t+k}{a}, t\right)$ expands in both directions as t increases, which implies supermodularity.[13] If $a > 1$, it shifts to the right. The consequences of slight misspecification in this section hinge on whether the above single crossing is violated or satisfied.

We begin by showing that if the single-crossing condition is violated, operating on a misspecified model leads to large and non-vanishing losses. Consider the following model: let S^m be a discrete set of types $\{0, 1/m, 2/m, .., 1\}$, each with equal probability, and preferences $v(s, y)$.[14] Here, as m increases, the true approximation index decreases, and goes to zero as m goes to infinity. The next claim describes the optimal contract for the model for a set of specifications.

Claim 1 *Suppose that $a > 1$ and $\frac{2b}{1+2b} < k < b$. The optimal contract for S^m involves choosing thresholds $0 \leq s_1 \leq s_2 < 1$. Types below s_1 are excluded. Types $s \in [s_1, s_2)$ receive $\bar{y} = 0$ at price $k + bs_1$. Types $s \geq s_2$ receive $y(s) = s$ at price $p(s) = s + bs_1$.*

In the solution to S^m, low types are excluded; intermediate types receive the 'pragmatic' value of the car; high types receive products best tailored to their preferences.[15] Furthermore, (almost) all served types receive positive rent (increasing in type). In particular, each type s belonging to $[s_2, 1]$ faces a binding non-local incentive-compatibility constraint: he is indifferent between $y = s$ at price $s + bs_1$, and receiving only the pragmatic-value ($\bar{y} = 0$) at price $k + bs_1$. Here, local constraints are not binding. This is a crucial observation because it means that a slight misspecification of preferences may lead to large deviations.

What happens now when this menu, which is optimal for the model type space S^m is actually offered to the true type space T? To see this, consider two contiguous high (model) types, $s'' > s' > s_2$. Take an unmodeled type in-between, $t \in (s', s'')$. This type t receives a net payoff of $b(t - s_1)$ if he buys $\bar{y}$ providing only the pragmatic value. Hence, he will never choose $y(s')$, since

$$(1 + b)\,t + \min\{0, a\left(s' - t\right)\} - p\left(s'\right) = b\,(t - s_1) - (a - 1)\left(t - s'\right) < b\,(t - s_1)\,.$$

[12] Note that the single-crossing condition also covers the outside option y_0. However, as the set of product y is $[0, 1]$, we must assign a conventional negative value to the outside option – for instance, $y_0 = -1$. We then assume that the set of possible products is $\{-1\} \cup [0, 1]$, and we set $v(t, -1) = 0$ for all t. The condition holds as stated in the text.

[13] The single-crossing condition is assumed when solving standard non-linear pricing problems (Wilson, 1993, p. 71) in order to guarantee that only local downward incentive-compatibility constraints are binding.

[14] One might be tempted to say that the principal who knows S^m will correctly "guess" the exact shape of $v(y, t)$ correctly. Of course, given model uncertainty, that is generally not possible for an unknown $v(y, \cdot)$, and there are a vast number of different preferences specifications close to those described in the model.

[15] For m not too small, $0 < s_1$ holds.

Similarly, he will never choose $y\left(s''\right)$ since

$$(1+b)\,t+\min\{0,a\left(s''-t\right)\}-p\left(s''\right)=b\left(t-s_1\right)-\left(s''-t\right)<b\left(t-s_1\right).$$

As a consequence, all true types not equal to model types choose the pragmatic product $\bar{y}$ and pay a price of $k+bs_1$ instead of $s+bs_1$. Formally, let $\Pi_T\left(S^m\right)$ be the principal's true expected profit when she uses – on the true type space (T_v, f) – the menu that is optimal for S^m.

Claim 2 *If $a>1$ and $\frac{2b}{1+2b}<k<b$, then $\Pi^*\left(T_v,f\right)-\lim_{m\to\infty}\Pi_T\left(S^m\right)=\frac{(1-k)^2}{2}>0$.*

Figure 2 below illustrates the above claim for $a=2$. Dots represents the principal's expected profit if she uses a model with $m+1$ types, the solid line the truly optimal profit. (As $m\to\infty$, the profit loss remains roughly 40%.)

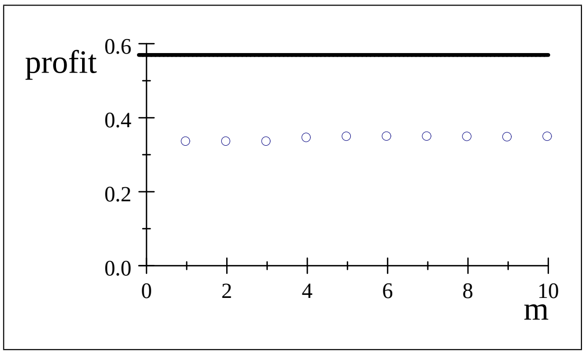


Figure 2

The above fact derives from the *joint* presence of preference misspecification and binding non-local constraints. To see this even more clearly, consider the case in which $a<1$. Here, the single-crossing condition holds and local incentive-compatibility constraints bind. Despite preference misspecification, the naive mechanism is now a valid approximation.

Claim 3 *If $a<1$, then $\Pi^*\left(T_v,f\right)-\lim_{m\to\infty}\Pi_T\left(S^m\right)=0$.*

To see the logic, consider a true type t between two adjacent model types s' and s'', which are offered different allocations (y',p') and (y'',p'') in the solution. Given the single-crossing property, this type t will choose one of these two allocations over any other allocation. Because local downward constraints bind, this then guarantees that type s'' will become indifferent between (y'',p'') and (y',p'), and types between s'' and s'' would choose (y',p') over any other allocation. The principal does experience a loss due to misspecification, but this loss is now bounded from above by the preference distance between model types and true types and, therefore, vanishes as that distance between the model and the truth goes to zero.

4 Profit-Participation Mechanism

We now turn to the second result of our paper: we can always obtain a valid approximation by employing a profit-participation mechanism (PPM). The section begins with an intermediate result (Lemma 1) on profit-participation discounting. We then prove the main result (Theorem 1) and conclude with some further observations.

4.1 Profit-Participation Pricing

We now introduce the key component of our solution concept, 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 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

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

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) and replacing (S_u, g) with (T_v, f). Equivalently, $\Pi\left(T_v, f, M\right)$ denotes the expected profit that menu M will generate given the true type distribution.[16]

Let us now define Profit-Participation Pricing.

Definition 1 *For any menu $M = \{y_k, p_k\}_k$, let the menu derived by Profit-Participation Pricing be $\tilde{M} = \{y_k, \tilde{p}_k\}_k$, where the products are unchanged and the new price vector $\tilde{p}(y_k)$ is given by*

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

for some $\tau > 0$.

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:

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

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.

[16] 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.

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 menu 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 from M through Profit Participation Pricing with $\tau = \sqrt{2\varepsilon}$. Then, for any truth (T_v, f) satisfying the approximation index:*

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

Proof. See Appendix .

The lemma contains the main intuition for why this 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 for 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 to 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 true types close to a model type will choose in the same way as their respective model types. Moreover, the principal still does not know how often different allocations will be chosen by the agent. In fact, she cannot even guarantee that, when offered the discounted menu, model types will choose the allocation they were choosing previously. Crucially, however, the principal knows that whichever allocation a true type chooses from the discounted menu, the deviation from the allocation chosen by the corresponding model type in the undiscounted menu cannot be very damaging to her 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. The maximal loss from deviation is bounded by $\frac{2\varepsilon}{\tau}$. To see this, note that, given a revealed preference argument, twice the approximation index puts a bound on the maximal price difference between the allocation chosen by a model type and the one chosen by a nearby true type. A profit-based discount links this price difference to the difference between the profits these different allocations deliver to the principal, and, at the same time, lowers a true agent type's maximal incentive to deviate to the allocation that provides the lower profit. Given the properties of an approximation index, the greater is this percentage profit-based discount, the lower is the loss in expected profits which can result from the difference between the choices of the true types and the choices of the model types. Setting $\tau = \sqrt{2\varepsilon}$ then optimizes this trade-off between the loss from lower prices and the loss from deviations and establishes the above upper bound.[17]

[17] There is an interesting connection between the proof of Lemma 1 and Theorem 21 of Balcan et al (2008).

4.2 Profit-Participation Mechanism

So far, we have not mentioned optimality. We have not chosen the set of alternatives and prices with expected profit in mind; rather, we have 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 (unknown) maximal expected profit given the true problem, and the principal's true, but also unknown, expected profit if she offers $\tilde{M}$ to the true type space. This comparison captures the approximation loss.

Formally, take any M^* containing the outside option, including, if it exists, the menu that is optimal for the true type space.

Definition 3 *For any M^*, let the PPM loss be $\Pi\left(T_v, f, M^*\right) - \Pi(T_v, f, \tilde{M})$.*

If there exists an optimal menu for the true type space – namely, if (Y, c, T_v, f) is 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 truly optimal menu.

We can now state the main result of the paper in terms of the known parameters of our setup. Note that the theorem does not require the principal to know M^*, the true expected profit that she could achieve given this mechanism, or the true expected profit given menu $\tilde{M}$.

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

Proof. See Appendix .

The proof of the theorem constructs the bound to the PPM loss by applying Lemma 1 twice. In the first application, it bounds the difference between the principal's true expected profit from any menu M and the maximal model-profit, given any model that satisfies the approximation index ε. The second application bounds the difference between the maximal model-profit and the profit in the true environment from the menu identified by PPM. Taken together, they bound the difference between the truly maximal profit and the profit obtained with the discounted version of the model-optimal menu.

The above bound is valid without requiring the principal to know anything beyond her model and the upper bound to the inaccuracy of the model: a valid approximation index

In their setup 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.

ε. Intuitively, a small known change in the model would have a small effect on the value of the maximal profit. Lemma 1 ensures that PPM yields a mechanism that guarantees a good approximation, even if this change in the model is unknown.

In sum, clearly the most robust mechanism the principal could offer is the "sell-out" mechanism, in which the agent is charged the principal's cost for any bundle he might choose. However, in this mechanism the principal can extract profit only by using a lump-sum participation fee, but this fee cannot exceed the surplus available on any type in order to ensure that all surplus-generating types participate. On the other hand, if the mechanism is designed to maximize the expected profit for some model, but the model misspecifies utility functions of the agents' types even slightly, the resulting profit could be much lower due to those types choosing very different bundles from the menu. PPM establishes a happy medium between these extremes: adding a small "profit-sharing" discount to the model-optimal mechanism ensures the robustness to small misspecifications of the model.

4.3 Discussion

Theorem 1 offers two novel lessons. First, even if the principal faces model uncertainty profit participation offers a simple and economically intuitive way of arriving at a menu that guarantees a payoff that is demonstrably close to the truly maximal payoff. Second, the quality of this approximation depends only on the quality of the model. 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 imposes an upper bound on the loss due to misspecification has a worst-case feel, akin to maxmin expected utility, e.g., Gilboa and Schmeidler (1989). Note, however, that we are not offering an optimal mechanism for such an ambiguity-averse principal. Such a principal knows the set of true states of the world and faces uncertainty only over the right distribution over them. In contrast, as mentioned, our principal faces unforeseen contingencies and cannot be sure to enlist all states of the world correctly. We, thus, offer a near-optimal mechanism for an expected payoff-maximizing principal with the property that it is robust to all small misspecifications of the states of the world. For *any* true environment that is sufficiently close to the principal's model, but may contain slightly difference preference specifications than those listed in the model, the profit guaranteed by PPM in that environment is close to what the maximal expected profit would be in that environment.

Finally, by imposing some minimal structure on the true type space - Lipschitz continuity of the agent's utility in his Euclidean type - PPM can also reduce the computational complexity of contracting. As mentioned, Conitzer and Sandholm (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's type space: the approximation index is then given by the maximal cell size and the Lipschitz constant. Select the preferences of any single type from the cell to be its representative type. Using this sparse and misspecified model, the principal can obtain a bound to the approximation loss by using Theorem 1. This way, PPM allows the principal to use a much simpler, albeit misspecified, model that operates on the basis of a potentially much smaller set of preference specifications and, thus, reduce the cost associated with computational (or communicational) complexity.[18] If the principal were to use this model naively, it might generate discontinuous losses. Under profit-participation, there is a simple smooth trade-off between the size of the approximation loss and how complex the model needs to be to generate robust revenue.

5 Alternative Mechanisms

As we have shown, PPM is a valid approximation scheme, that is, it guarantees that the losses go to zero as model uncertainty vanishes, but are there other approximation schemes that exhibit perform equally well or better? To address this question, we first must 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 may perform better than PPM. In the presence of the model uncertainty we consider, however, a more pertinent question is whether there are other mechanisms that are robust, given the general class of problems we consider.

Since our results apply to a large class of screening problems, we now ask whether other mechanisms besides PPM work for this whole class of problems. We begin by defining the class of mechanisms that are based on the principal's model and then modify prices based on some rule, given the solution to this model:

Definition 4 *A mechanism is model-based if it can be represented as a two-step process that (i) first finds an optimal menu $\hat{M}$ for the screening problem defined by S_u, g, Y,c; and (ii) then modifies the price vector $p(y)$ according to some function*

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

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 to the principal's model, that is, solutions where all alternatives offered are bought with positive probability given the principal's model. Such a 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,*

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

[18] For a treatment on the corresponding computational and communicational complexity and details on the interpretation of PPM as a polynomial time approximation scheme (PTAS), see Proposition 1 of Section 5 in Madarász and Prat (2010).

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.

In the proportional discount mechanism, the principal acts magnanimously by discounting prices proportionally.

Example 3 *In the proportional discount mechanism,*

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

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

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

Example 4 *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 model-based mechanisms depending on whether or not they satisfy a condition of profit participation. This condition, stated below, is more permissive than our specific PPM, that is, it can be satisfied by mechanisms that do not exactly match PPM. Nevertheless, it is sufficiently restrictive to rule out a large class of mechanisms that systematically differ from PPM.

A perturbation satisfies our condition of profit participation when the price discount it offers on a product-price pair is always *strictly* increasing in the associated profit. In particular, a function Ψ satisfies our condition of profit participation as long as there do not exist distinct cost levels at which a higher-priced but lower-profit allocation is always discounted weakly more than a lower-priced but higher-profit allocation.

Definition 5 *A function Ψ satisfies profit participation for a given ε, if there are no $c'' > c'$ such that for all $p''$and p',*

$$\text{if } 0 < p'' - p' < c'' - c', \text{ then } p'' - \Psi(p'', c'', \varepsilon) \geq p' - \Psi(p', c', \varepsilon). \tag{1}$$

The above definition points to the existence of only two distinct cost levels where, for a positive range of prices, the discount is weakly decreasing, instead of strictly increasing, in profit. Specifically, at these cost levels, a higher-priced allocation associated with a lower profit is always discounted weakly more than a lower-priced allocation associated with a higher profit.[19] Hence, the logic of profit participation is systematically violated; for a positive range of prices, given by the set of all prices that satisfy the condition $0 < p'' - p' < c'' - c'$, the discount is weakly decreasing in profit. If such cost levels exist for a given Ψ and ε, we say that Ψ does not satisfy profit participation at this ε.

Note that none of the above mechanisms, except for PPM, satisfy profit-participation for any given ε. Consider, for example, the naive mechanism. Here,

$$p'' - \Psi\left(p'', c'', \varepsilon\right) = 0 \geq 0 = p' - \Psi(p', c', \varepsilon)$$

always holds, since the discount on any price is zero. Thus, a higher-profit allocation is always discounted *weakly* less than a lower-profit allocation. More generally, in the first three examples, the discount is always non-decreasing in price and is independent of the cost. Thus, given any $c'' > c'$, a higher-profit allocation associated with a lower price is *always* discounted (at least weakly) less than a lower-profit allocation associated with a higher price. In contrast, this is never true under PPM. Here, the discount is always strictly increasing in profit, and, hence, PPM satisfies profit participation for any approximation index ε.[20]

We now show that a model-based mechanism that does not satisfy profit participation cannot guarantee that losses vanish as the model tends to the truth. To state this formally, we need to introduce some additional notation.

Fix a truth (T_v, f) and an approximation index ε. First, consider the collection of all models that are ε-approximations of this truth. We denote this collection of models by $\Delta(T_v, f, \varepsilon)$. Now, fix a modifier Ψ, and consider all model-based mechanisms for which it is true that, each mechanism is (i) based on an optimal menu for a model in $\Delta(T_v, f, \varepsilon)$ which is (ii) modified by Ψ. We denote the set of all resulting menus by $\Lambda(T_v, f, \varepsilon, \Psi)$. We can then express the supremum of the difference between the principal's truly maximal profit and the profit generated by such a menu. Formally, we can express this *maximal loss* as

$$L(T_v, f, \varepsilon, \Psi) = \sup_{M \in \Lambda(T_v, f, \varepsilon, \Psi).} \Pi^*(T_v, f) - \Pi(T_v, f, M).$$

We already showed in Section 3 that this maximal loss can be very significant, for any given ε, when Ψ is the naive mechanism. With this new notation, Theorem 1 stated that if Ψ is the one used in PPM, then this maximal loss always decreases in ε, and goes to zero as $\varepsilon \to 0$, for any true problem (T_v,f). In other words, if Ψ is PPM, the principal is guaranteed that this maximal loss vanishes as her model becomes more and more precise, for *any* possible true problem. We now show that if Ψ does not satisfy profit participation, this can no longer hold.

Theorem 2 *Fix any Ψ and $\varepsilon > 0$. If Ψ does not satisfy profit participation for this ε, then there exists $(T_v, f) \in \Theta$ such that $L(T_v, f, \varepsilon, \Psi) \geq 1/8$.*

[19] Note that given y' and y'', the condition $0 < p(y'') - p(y') < c(y'') - c(y')$ implies both that $p(y'') > p(y')$ and that $\pi(y'') < \pi(y')$.

[20] Formally, if $0 < p'' - p' < c'' - c'$, then $p'' - \Psi\left(p'', c'', \varepsilon\right) = \tau(p'' - c'') < \tau(p' - c') = p' - \Psi(p', c', \varepsilon)$ for any $\tau > 0$.

Proof. See Appendix .

The theorem implies that given any Ψ that does not satisfy profit participation, one can find a problem (in fact, a class of problems) for which the share of profit lost because of Ψ stays strictly bounded away from zero, even if the principal uses a model that is arbitrarily close to the truth.

The above statement is a natural counterpart to Theorem 1. While Theorem 1 shows that PPM works for *any* type space, the above result claims that model-based mechanisms that do not satisfy profit participation do not share this property. Given a discounting rule, if this rule violates profit participation, there is always a true environment where there can be significant losses, even if the principal's model is arbitrarily close to this truth. Theorem 2 implies that if a mechanism over the set of all admissible problems performs as well as PPM, it must be either very similar to PPM, in that it satisfies profit participation, or very different in that it is not even model-based.

The proof proceeds by constructing a straightforward class of problems with binding non-local constraints. We fix a true environment where the product space includes a generic product produced at a higher cost and a continuum of specific personalized products produced at a lower cost. In the solution to the truth, a non-zero measure of types face a binding non-local incentive-compatibility constraint. This IC constraint is between a lower-priced personalized alternative and the higher-priced generic alternative, with the profit margin on the former being higher than on the latter. The fact that a model-based mechanism does not satisfy the profit participation condition implies, here, that the price of the generic good is discounted more. Hence, when perturbing preferences even slightly, a perturbed type strictly prefers the generic alternative to the model type's allocation. As in Section 3, this creates a discrete profit loss for the principal for any degree of preference misspecification as long as the discounting rule violates profit participation.

The proof of Theorem 2 is based on constructing a setting where the violation of the profit participation condition, in the joint presence of preference misspecifications and binding non-local constraints, leads to non-vanishing losses. Clearly, many other settings with the same feature can be identified. Such cases could be associated with an even higher absolute profit loss than identified in the specific counter-example for the proof. The result, however, does not establish how endemic non-local incentive constraints are in screening problems. We believe that such non-local constraints are present in a wide range of serious economic settings, e.g., the 'multi-dimensional' screening problems in Rochet and Choné (1998). Establishing more precise claims about prevalence, however, requires future research, both theoretical and empirical.

The theorem uses the fact that in many settings the solution to a screening problem involves binding non-local incentive constraints. As mentioned, such a feature is not present in the classic setup of Mussa and Rosen (1978) where the single-crossing property holds and binding constraints are local. However, as we argued in Section 3, there are many economically relevant screening problems that involve binding non-local constraints. Here, Theorem 2 provides useful general guidance in the presence of model uncertainty.

The above result also points to three interesting questions. 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? We also leave these questions to future research.

6 Conclusion

We consider a principal who faces model uncertainty and unforeseen contingencies and is constrained to operate on a potentially misspecified model of the agent's preferences. We characterize the upper bound on the expected loss that such a principal incurs as long as she uses a profit-participation mechanism. We show that this loss vanishes smoothly as the model type space tends to the true one, and also prove that this is not true for similar model-based mechanisms that do not contain an element of profit participation.

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 from using an incorrect model. One strength of our approach is that it does not make specific functional or distributional assumptions, and it applies to settings in which the potential allocation space is very large. It would be interesting to know whether the same economic insight holds beyond our setup. As mentioned, future research can investigate whether there are non-model-based mechanisms that perform better than PPM, or specify conditions under which the Naive Mechanism is robust to model uncertainty.

Our analysis also has a number of important limitations that future research could address. First, the principal's cost depended only on the product characteristics, but not on the type of the agent - as in insurance problems. The current approach cannot directly be extended to type-dependent costs because the profit-participation scheme assumes that profit depends only on the product and the price. It would be interesting to extend it by allowing profit to depend on a conjecture over what types buy the product. Second, we assumed 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 in which payoffs are not quasilinear. Finally, it may be interesting to see use the local misspecification approach in other agency problems besides nonlinear pricing. In this vein, Carroll and Meng (forthcoming) extend this profit participation approach to a moral hazard problem where the principal's model may be slightly misspecified.

7 Appendix

Proof of Claim 1. Suppose that $a > 1$. Consider first the set of possible menus where all types are served. The lowest type can be charged, at most, k. Since for all $s < k$ the pragmatic value is greater than the best aesthetic value, these will buy the same car as the lowest type and pay k.

For any type $s \geq k$, the upper bound to the price the principal can charge a type s equals s: the best aesthetic value for this type, $(1+b)\,s$, minus the net utility he would get if he bought the cheapest car offered to the lowest type, $bs + k - k$. The upper bound is therefore simply s. If a menu achieves that upper bound, we know it is optimal among the class of menus where all types are served. Suppose that every $s \geq k$ is offered his ideal product $y = s$ at price $p = s$. This would achieve the bound and guarantee that type s preferred his allocation to the outside option, as well as, to a car with only pragmatic value. We then need to check that it is also incentive-compatible with respect to all other allocations assigned to types above k. Type $s \geq k$ prefers his allocation to the allocation assigned to a different type $s' \geq k$ if

$$(1+b)\,s - s \geq \max\left((1+b)\,s + \min\left(0, a\left(s'-s\right)\right), bs + k\right) - s'.$$

Note, we already know that s prefers his own allocation to a pragmatic car, hence, we can exclude that he would prefer the allocation meant for s' for its pragmatic value $bs + k$. This fact, together with other rearrangements, leads to:

$$s' - s \geq \min\left(0, a\left(s'-s\right)\right)$$

If $s' \geq s$, the inequality is satisfied. If $s' < s$, the inequality becomes $(a-1)\,(s-s') \geq 0$, which is satisfied if $a > 1$, which is what we have assumed.

Hence, we found the optimal menu in case all types are served. Its expected profit (in the model world S^m) is:

$$\frac{1}{m+1}\sum_{j=0}^{m}\max\left(k, \frac{j}{m}\right) = \frac{1}{m+1}\left(k(\text{int}\,(km)+1) + \sum_{j=\text{int}(km)+1}^{m}\frac{j}{m}\right).$$

If the principal excludes types below $\frac{m_0}{m}$, she can raise the price of the pragmatic product from k to $k + bs_{\bar{m}}$ and the price of all specific goods from s to $s + bs_{\bar{m}}$, where $s_{\bar{m}} = \frac{m_0}{m}$. The profit then, assuming that $s_{\bar{m}} \leq k$, becomes

$$\frac{1}{m+1}\left(\left(k + b\frac{m_0}{m}\right)(\text{int}\,(km) + 1 - m_0) + \sum_{j=\text{int}(km)+1}^{m}\left(\frac{j}{m} + b\frac{m_0}{m}\right)\right).$$

Disregarding integer constraints, the optimal m_0 maximizes:

$$\left(k + b\frac{m_0}{m}\right)(km + 1 - m_0) + (m - km)\,b\frac{m_0}{m},$$

with first-order condition:

$$\frac{m_0}{m} = \frac{b-k}{2b} + \frac{1}{2m}.$$

The optimal $\frac{m_0}{m}$ is positive, but lower than k if $\frac{2b}{1+2b} < k < b$, which is what we assumed. Once integer constraints are considered, the optimal m_0 is then either the integer (weakly) above or below $\frac{b-k}{2b}m + \frac{1}{2}$.

To sum up, under these parameter values, the optimal menu for the model S^m can be described as follows. Types are divided into three disjoint sets: low types S_1, medium types S_2 and high types S_3. The low types are excluded. The medium types all buy the same product $\bar{y}$ (without loss of generality assume this is the lowest product $\bar{y} = 0$) at price $k + bs_1$, where s_1 is the lowest element of S_2. Every type in S_3 is offered his ideal product $y(s) = s$ at price $p(s) = s + bs_1$.

Proof of Claim 2. Given the argument above, for any S^m, the expected profit that the principal obtains if she offers the menu that is optimal for S^m to the true type T is simply the price of the pragmatic good times the share of types served. Recall, that the optimal m_0 is given by the integer (weakly) above or the integer below $\frac{b-k}{2b}m + \frac{1}{2}$. Therefore, the first non-excluded type is:

$$s_1 = \frac{m_0}{m} \in \left\{ \frac{\text{int}\left(\frac{b-k}{2b}m + \frac{1}{2}\right)}{m}, \frac{\text{int}\left(\frac{b-k}{2b}m + \frac{1}{2}\right) + 1}{m} \right\},$$

and the price of the pragmatic product is:

$$p_1 = k + bs_1 \in \left\{ k + b\frac{\text{int}\left(\frac{b-k}{2b}m + \frac{1}{2}\right)}{m}, k + b\frac{\text{int}\left(\frac{b-k}{2b}m + \frac{1}{2}\right) + 1}{m} \right\}.$$

The principal's expected profit from the solution to the model (on the truth) is:

$$\left(1 - \frac{m_0}{m}\right) p_1 \in \left\{ \begin{array}{c} \left(1 - \frac{\text{int}\left(\frac{b-k}{2b}m+\frac{1}{2}\right)}{m}\right)\left(k + b\frac{\text{int}\left(\frac{b-k}{2b}m+\frac{1}{2}\right)}{m}\right), \\ \left(1 - \frac{\text{int}\left(\frac{b-k}{2b}m+\frac{1}{2}\right)+1}{m}\right)\left(k + b\frac{\text{int}\left(\frac{b-k}{2b}m+\frac{1}{2}\right)+1}{m}\right) \end{array} \right\}.$$

As m increases, the two possible values of the expected profit both converge to:

$$\lim_{m\to\infty} \frac{m - m_0}{m} p_1 = \left(1 - \frac{b-k}{2b}\right)\left(k + \frac{b-k}{2}\right) = \frac{(b+k)^2}{4b}.$$

If, instead, we use the menu that is optimal for T, we set:

$$\begin{aligned} t_1 &= \frac{b-k}{2b}, \\ t_2 &= k, \\ p_1 &= \frac{b+k}{2}, \\ p(t) &= t + bt_1 = t + \frac{b-k}{2} = t - k + \frac{b+k}{2}, \end{aligned}$$

and the expected profit is:

$$\Pi^* = (t_2 - t_1) p_1 + \int_{t_2}^{1} p(t)\, dt = (1 - t_1) p_1 + \int_{k}^{1} (t-k)\, dt = \frac{(b+k)^2}{4b} + \frac{(1-k)^2}{2}$$

.

Proof of Claim 3. Consider a true type $t \in (s', s'')$ not included in the principal's model, where $s' < s''$ are two neighboring model types. Let the allocations chosen in equilibrium by s' and s'' be (y', p') and (y'', p'') respectively. Hence, for any other allocation (y, p) that is offered in the menu (including the outside option $(y_0, 0)$), it must be that:

$$v\left(s', y'\right) - p' \geq v\left(s', y\right) - p, \tag{2}$$

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

However, single crossing implies that for every $y < y'$:

$$v\left(s', y'\right) - v\left(s', y\right) \leq v\left(t, y'\right) - v\left(t, y\right),$$

which, combined with the IC constraints in (2), guarantees that type t prefers (y', p') to any other allocation (y, p) with $y < y'$. A similar line of reasoning (using the IC constraints in (3), and the single-crossing property) guarantees that t prefers (y'', p'') to any other allocation (y, p) with $y > y''$. This means that t always chooses either (y', p') or (y'', p''). Hence, the principal's payoff is always at least p' (price is monotonic). As $\#S \to \infty$, the difference between p' and p'' vanishes, and the principal's expected profit tends to the optimal profit under the truth T .

Proof of Lemma 1. First note that the loss due to the price discount is (recalling that profit is bounded above by $\Pi_{\max}$, which was normalized to 1),

$$\tilde{p}\left(y\right) - c\left(y\right) - \left(p\left(y\right) - c\left(y\right)\right) = -\tau\left(p\left(y\right) - c\left(y\right)\right) \geq -\tau. \tag{4}$$

By the definition of the approximation index ε, we know that there exists a partition $\mathcal{P}$ 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 cell associated with s – namely, $t \in J(s)$. 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 types by $(\widehat{y}, p(\widehat{y}))$. The only loss for the principal is due to the price discount determined by τ:

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

Focus, now, on case (ii). Suppose that when $\tilde{M}$ is offered, t chooses an allocation y' different from $\widehat{y}$ chosen by s. Because $t \in J(s)$, we know that

$$\left|v\left(t, \widehat{y}\right) - u\left(s, \widehat{y}\right)\right| \leq \varepsilon,$$

$$\left|v\left(t, y'\right) - u\left(s, y'\right)\right| \leq \varepsilon,$$

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

$$u(s,\widehat{y}) - u(s,y') + 2\varepsilon \geq v(t,\widehat{y}) - v(t,y') \geq u(s,\widehat{y}) - u(s,y') - 2\varepsilon. \tag{5}$$

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 s prefers $\widehat{y}$ to y':

$$u(s,\hat{y}) - p(\hat{y}) \geq u(s,y') - p(y'). \tag{6}$$

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

$$v(t,\hat{y}) - \tilde{p}(\hat{y}) \leq v(t,y') - \tilde{p}(y'). \tag{7}$$

By subtracting (7) from (6), we get that:

$$\begin{aligned} & p(y') - \tilde{p}(y') - (p(\hat{y}) - \tilde{p}(\hat{y})) \\ & \geq v(t,\hat{y}) - v(t,y') - (u(s,\hat{y}) - u(s,y')). \end{aligned} \tag{8}$$

By (5), the right-hand side of (8) is bounded below by -2ε. Given the definition of $\tilde{p}$, the left-hand side of (8) 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,

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

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

$$p(y') - c(y') - p(\hat{y}) + c(\hat{y}) \geq -\frac{2\varepsilon}{\tau}. \tag{10}$$

Adding this inequality and the inequality in (4), we get that:

$$\tilde{p}(y') - c(y') - (p(\hat{y}) - c(\hat{y})) \geq -\tau - \frac{2\varepsilon}{\tau}. \tag{11}$$

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 that:

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

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

Proof of Theorem 1. Take a given menu M^* containing the outside option. The true expected profit from it is:

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

where $y\left(t\right), p\left(y\left(t\right)\right)$ is such that $v\left(t, y\left(t\right)\right) - p\left(y\left(t\right)\right) \geq v\left(t, y\right) - p(y)$ for all $t \in T$ and all $\left(y, p\left(y\right)\right) \in M^*$. Consider now $\left(S_u, g\right)$ with elements $i \in \{1, ..., \#S\}$. There exists an ε-approximation partition P of (T_v, f), such that for each cell $J(s_i)$ of P, there exist $t_i^* \in J(s_i)$ such that:

$$\left[p\left(y\left(t_i^*\right)\right) - c\left(y\left(t_i^*\right)\right)\right] g(s_i) \geq \int_{t \in J(s_i)} \left[p\left(y\left(t\right)\right) - c(y\left(t\right))\right] f(t) dt. \tag{12}$$

Step 1. Construct a model $(\bar{S}_v, \bar{g})$ with $\bar{S} \equiv \left(t_i^*\right)_{i=1,..\#S}$ and preferences $v\left(t_i^*, y\right)$, and $\overline{g}(t_i^*)$ $= g(s_i)$ for all $i \in \{1, ..., \#S\}$. Since all types in $\bar{S}_v$ are contained in T_v, it follows that they all choose the same options from M^* as they did before. It must then be that:

$$\Pi\left(\bar{S}_v, \bar{g}, M^*\right) \geq \Pi\left(T_v, f, M^*\right).$$

Step 2. We now apply Lemma 1 for the first time. The partition P used above is still an ε-approximation partition between $\left(\bar{S}_v, \bar{g}\right)$ and $\left(S_u, g\right)$.[21] Let M' be the menu derived by profit-participation pricing from M^*. By Lemma 1,

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

Step 3. Consider, now, the menu $\hat{M}$ that is optimal for $\left(S_u, g\right)$:

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

By definition,

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

Step 4. We employ Lemma 1 for the second time. When we discount $\hat{M}$ through profit-participation pricing to obtain $\tilde{M}$, we get that:

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

[21] Note that while S_u is discrete, since there is a bijection between the elements of S_u and the elements of $\overline{S}_v$, leaving probabilities unaffected, Lemma 1 applies.

Summing up the above steps:

$$
\begin{aligned}
&\Pi\left(T_v, f, M^*\right) = [\text{expected profit for any } M^*]; \\
&\Pi\left(\bar{S}_v, \bar{g}, M^*\right) \geq \Pi\left(T_v, f, M^*\right) && \text{(Step 1)} \\
&\Pi\left(S_u, g, M'\right) \geq \Pi\left(\bar{S}_v, \bar{g}, M^*\right) - 2\sqrt{2\varepsilon}; && \text{(Step 2)} \\
&\Pi\left(S_u, g, \hat{M}\right) \geq \Pi\left(S_u, g, M'\right); && \text{(Step 3)} \\
&\Pi\left(T_v, f, \tilde{M}\right) \geq \Pi\left(S_u, g, \hat{M}\right) - 2\sqrt{2\varepsilon} && \text{(Step 4)}
\end{aligned}
$$

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

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

.

Proof of Theorem 2. Suppose that Ψ does not satisfy profit participation for a given ε. By definition, there exists $c'' > c'$ such that if $0 < p'' - p' < c'' - c'$, then $p'' - \Psi(p'', c'', \varepsilon) \geq p' - \Psi(p', c', \varepsilon)$. Consider the following problem:

$$
\begin{aligned}
T &= [0,1] \ \text{ with uniform density;} \\
Y &= [0,1] \cup \{\bar{y}\} \cup y_0; \\
u\left(t, y\right) &= \begin{cases} l + q(t - 2\left|t - y\right|) & \text{if } y \in [0,1] \\ h & \text{if } y = \bar{y} \\ 0 & \text{if } y = y_0 \end{cases} \\
c\left(\bar{y}\right) &= c'' \text{ and } c(y) = c' \text{ for all } y \in [0,1] \ \text{ and } c(y_0) = 0.
\end{aligned}
$$

Define $h = 2c'' - c'$, $l = c'$ and $q = 2c'' - 2c'$. Consider type t^* such that $h - c'' = u(t^*, t^*) - c'$. We can rewrite this as $c'' - c'' = 2(c'' - c')t^*$, which implies that $t^* = 0.5$. It is easy to see, that in the optimal solution to this screening problem, types below t^* buy $\bar{y}$ at price h, and each type above t^* is offered a personalized alternative $\hat{y}\left(t\right) = t$ at price $l + qt$. The principal's expected profit is

$$0.5(c'' - c') + \int_{0.5}^{1} (l + qt - c')dt = 1.\,25(c'' - c').$$

Consider now a model $(S^n_{u^n}, g^n)$ in which $S^n \subseteq T$ and $u^n(s, y) = u(s, y)$. Furthermore, given $n \in \mathbb{N}$, let:

$$
\begin{cases}
S^n_{u^n} = \left\{0, \frac{1}{2^{n+1}}, \frac{2}{2^{n+1}} \ldots, 1\right\}; \\
g^n\left(0\right) = g^n\left(1\right) = \frac{1}{2^{n+2}}; g^n\left(s\right) = \frac{1}{2^{n+1}} \text{ for all other } s.
\end{cases}
$$

Given the prior f, a valid approximation index for model set $S^n_{u^n}$ is $\varepsilon_n = 3q(\frac{1}{2^{n+1}})$. Hence, for any given $\varepsilon > 0$, if n is sufficiently large, then $\varepsilon_n < \varepsilon$, and for any such n, $(S^n_{u^n}, g^n)$ is such an ε-approximation of the truth.

The optimal solution to $(S^n_{u^n}, g^n)$ involves offering $\bar{y}$ at price $p\left(\bar{y}\right) = h$, as well as a vector of personalized alternatives $\hat{y}(s) = s$ for types $s > t^*$ at prices $l + qs$. The model-based mechanism modified by Ψ then returns prices:

$$\tilde{p}(\bar{y}) = \Psi\left(p(\bar{y}), c'', \varepsilon_n\right) = \Psi\left(h, c'', \varepsilon_n\right),$$
$$\tilde{p}(\hat{y}(s)) = \Psi\left(p(\hat{y}(s)), c', \varepsilon_n\right) = \Psi\left(l + qs, c', \varepsilon_n\right).$$

Now recall that $\Psi(p, c, \varepsilon)$ violates profit participation given ε at $c'' - c' > p'' - p'$, and, indeed, for all $s > t^*$,

$$0 < p(\bar{y}) - p(\hat{y}(s)) = (2c'' - 2c')(1 - s) < c'' - c'$$

since $t^* = 0.5$. It then follows that for any $s > t^*$,

$$h - \Psi(h, c'', \varepsilon) \geq l + qs - \Psi(l + qs, c', \varepsilon). \tag{13}$$

Now take any $t \in T$ that is not a model type (a set of positive measure given f for any $\varepsilon_n > 0$), and consider such a t's choice between the allocation $\hat{y}(s)$, given some $s > t^*$, now at price $\Psi(l + qs, c'; \varepsilon_n)$, and the allocation $\bar{y}$, now at price $\Psi(h, c''; \varepsilon_n)$. If such a t buys $\hat{y}(s)$, he gets a payoff of:

$$l + q(t - 2|s - t|) - \Psi\left(l + qs, c', \varepsilon_n\right).$$

If he buys $\bar{y}$, he gets a payoff of:

$$h - \Psi\left(h, c'', \varepsilon_n\right).$$

Hence, such a t chooses $\hat{y}(s)$ only if:

$$l + q(t - 2|s - t|) - \Psi\left(l + qs, c', \varepsilon_n\right) \geq h - \Psi\left(h, c'', \varepsilon_n\right),$$

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

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

which can be re-written as

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

which is always false. Hence, all types that are not model types choose $\bar{y}$ rather than any of the personalized alternatives. Hence, the expected profit is $h - c'' = c'' - c'$. Since the above holds for any n, as $\varepsilon_n \to 0$, the loss is still $0.25(c'' - c')$. Given that $2(c'' - c') \leq 1$, by normalization, the bound follows .

References

[1] Armstrong, Mark. (1996): Multiproduct Nonlinear Pricing. *Econometrica* 64: 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] Bergemann, Dirk, and Steven Morris. (2005): Robust Mechanism Design. *Econometrica* 73, 1771-1813.

[7] Bergemann, Dirk, and Karl Schlag. (2011): Robust Monopoly Pricing. *Journal of Economic Theory* 146: 2527–2543.

[8] Blumrosen, Liad, Noam Nisan, and Ilya Segal. (2007): Auctions with Severely Bounded Communication. *Journal of Artificial Intelligence Research* 28: 233-266.

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

[10] Carroll, Gabriel. (2015): Robustness and Linear Contracts. *American Economic Review* 105: 536-63.

[11] Carroll, Gabriel and Delong Meng. (2015): Locally Robust Contracts for Moral Hazard, *Journal of Mathematical Economics* forthcoming.

[12] Chassang, Sylvain. (2013): Calibrated Incentive Contracts. *Econometrica* 81: 1935–1971.

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

[14] 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), 132-141, New York, NY, USA.

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

[16] Gershkov, Alex, Jacob Goeree, Alexey Kushnir, Benny Moldovanu, and Xianweh Shi. (2013): On the Equivalence of Bayesian and Dominant Strategy Implementation. *Econometrica* 81: 197–220.

[17] Gilboa, Itzhak, and David Schmeidler. (1989): Maxmin Expected Utility with Non-unique Prior. *Journal of Mathematical Economics* 18: 141-153.

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

[19] Jehiel, Philippe, Moritz Meyer-ter-Vehn, Benny Moldovanu, and William R. Zame. (2006): The Limits of Ex-post Implementation. *Econometrica* 74: 585–610.

[20] Madarász, Kristóf and Andrea Prat. (2010): *Screening with an Approximate Type Space.* CEPR Discussion Papers 7900.

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

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

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

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

[25] Prokhorov, Yuri V. (1956): Convergence of Random Processes and Limit Theorems in Probability Theory. *Theory of Probability and Its Applications* 1: 157–214.

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

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

[28] 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.

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

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

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