Statistical Mechanism Design: Robust Pricing and Reliable Projections[*]

Duarte Gonçalves[†] Bruno Furtado[‡]

August 6, 2020

Abstract

This paper studies the robustness of pricing strategies when a firm is uncertain about the distribution of consumers' willingness-to-pay. When the firm has access to data to estimate this distribution, a simple strategy is to implement the mechanism that is optimal for the estimated distribution. We find that such empirically optimal mechanism boasts strong profit and regret guarantees. Moreover, we provide a toolkit to evaluate the robustness properties of different mechanisms, showing how to consistently estimate and conduct valid inference on the profit generated by any one mechanism, which enables one to evaluate and compare their probabilistic revenue guarantees.

Keywords: Monopoly; Robust mechanism design; Pricing; Hypothesis testing; Nonparametric estimation.

JEL Classifications: D82; D42; C14; L12.

[*]We are grateful to Yeon-Koo Che, Mark Dean, Navin Kartik, Elliot Lipnowski and Michael Woodford for their comments and suggestions.

[†]Department of Economics, Columbia University and BRU-IUL; duarte.goncalves@columbia.edu.

[‡]Department of Economics, Columbia University; bruno.furtado@columbia.edu.

1. Introduction

When facing uncertainty about consumers' willingness-to-pay, pricing and projections are two central elements of a firm's business plan. As the profitability of any pricing strategy is uncertain, it may be desirable to pursue strategies that provide revenue guarantees for the firm. On the other hand, firms often depend on projections about their future revenue under different scenarios to inform budget planning, inventory management, capital investments, and more. Consequently, they need to be able not only to consistently estimate their expected profits but also to have reliable confidence bounds. While pricing and projections are conceptually intertwined, these can and are often considered separately.

In this paper, we propose a data-based approach to tackle both these issues. In our setup, the firm faces uncertainty about the true distribution of consumers' types in the otherwise canonical setup of Maskin and Riley (1984). As is standard, we allow the firm to design mechanisms – pairs of prices and quantities – though our results also hold in the more restrictive setting of uniform pricing. However, differently from what is typically assumed in the robust mechanism design literature, instead of holding precise information about features of the distribution of consumers' willingness-to-pay, we assume the firm observes a finite sample drawn from this distribution.

Our contribution is twofold. First, we study a specific pricing strategy, the *empirically optimal mechanism*, and its finite sample robustness properties. Second, we provide a toolkit to perform statistical inference on the profit obtained for any arbitrary mechanism, including the empirically optimal one, enabling a data-based approach to the evaluation and comparison of different pricing strategies.

Empirically optimal mechanisms are constructed in a simple and intuitive manner. Fixing any given distribution of consumer types, we start by obtaining a mechanism that is optimal for that distribution. This mapping from distributions to optimal mechanisms is then coupled with a consistent estimator for the true distribution. Hence, empirically optimal mechanisms maximize expected profit when the estimate is taken to be the true distribution. Importantly, this class of menus relies on a fully nonparametric, prior-free estimator of the type distribution.

We show that mechanisms constructed in this manner are asymptotically optimal, achieving the optimal profit with probability one as the sample size grows. Moreover, empirically optimal mechanisms are robust in the spirit of Bergemann and Schlag (2011), that is, small perturbations of the estimated distribution, induced by changes in the underlying data, do not affect the firm's expected profits. This follows from establishing Lipschitz continuity of the firm's value function in the distribution.

Empirically optimal mechanisms also entail strong probabilistic guarantees for both profit and regret – the difference between the optimal expected profit and the expected profit that a mechanism delivers under the true distribution – and we characterize how these relate to the sample size. For any given finite sample size, we obtain a probabilistic lower bound for expected profit and a probabilistic upper bound for regret. Crucially, these bounds are not asymptotic and depend only on known constants. These results are then used to establish how many samples the firm needs, from an *ex-ante* perspective, to obtain probabilistic bounds on expected profit and on maximal regret. Our findings are related to the growing literature on sample-based revenue guarantees (see e.g. Cole and Roughgarden 2014; Huang et al. 2018; Guo et al. 2020), as we provide a non-asymptotic lower bound on sampling requirements for profit and regret guarantees.

We then provide tools for estimating expected profit and reliably conducting inference, not only for a mechanism that is empirically optimal given an estimate of the distribution, but for any fixed mechanism. We derive an estimator for expected profit that is consistent and unbiased, and, when appropriately rescaled, asymptotically normal. Further, we note the validity of a bootstrap implementation to conducting inference.

This approach enables estimation of the expected profit that any given mechanism attains for purposes such as budgeting, regardless of the criteria that guided the choice of the mechanism. Moreover, the ability to conduct data-based inference on the expected profit also expands the criteria that can be used to select a mechanism, namely by considering probabilistic revenue guarantees given by confidence intervals, by testing which in a set of alternative mechanisms achieves a higher expected profit, or by directly analyzing the distribution of the difference in profit between any two mechanisms. As such, this methodology of estimation and inference applies generally and opens up a new, data-based approach to deriving robust revenue guarantees.

In particular, we show how empirically optimal mechanisms can be used to estimate and conduct inference on the *optimal expected profit*, that is, the maximum expected profit that could be obtained were the firm to know the true distribution. Conducting inference on the optimal expected profit is useful for two reasons. First, as regret considerations constitute a standard criterion in the literature for selecting among competing mechanisms and as regret, by its definition, depends on knowledge of the optimal expected profit, a means to consistently estimate and conduct inference on the latter will enable the same for the former. Second, the maximum achievable profit can itself be an object of interest when, for instance, comparing the expected return of alternative investment possibilities.

While consistency of our suggested estimator follows from our results on asymptotic optimality of empirically optimal mechanisms, its asymptotic normality follows from a novel envelope theorem for the firm's value function. In other words, we show that the value function is Fréchet differentiable in the distribution of consumer types and that its Fréchet derivative equals that of the profit function at the empirically optimal mechanism. With Fréchet differentiability in hand, a Delta method for statistical functionals applies, and asymptotic normality of our estimator ensues. We then again consider bootstrap implementations for conducting inference.

In order to study the finite sample properties of the proposed estimators, we conduct Monte Carlo simulations of their bootstrap implementation for the standard specification of sale of an indivisible good. We perform this exercise for the expected profit of a given mechanism as well as for the optimal profit. The evidence shows that the empirical coverage of our estimators approximates well the associated confidence intervals, even with relatively few samples.

Finally, we illustrate how our results on empirically optimal mechanisms partially extend to an auction setting. We consider the case where the firm auctions a single item to a finite number of risk-neutral bidders with independent private values drawn from the same distribution. In particular, analog versions of asymptotic optimality and profit and regret guarantees are shown to hold in this setting as well.

Empirically optimal mechanisms correspond to one of the simplest forms of statistically informed mechanism design: observing a sample, estimating a distribution, and implementing a mechanism that is optimal for the estimated distribution. As we demonstrate, this extremely simple mechanism not only has sound revenue guarantees, but also allows practitioners to estimate the maximum profit attainable. Our results on estimation of the expected profit enable designers to construct confidence intervals and perform hypothesis testing, constituting a useful tool for practitioners and empiricists. From a theoretical perspective, we hope our data-based perspective on robust mechanism design proves useful in the broader study of mechanism design under model uncertainty.

Related Literature

The most directly related literature studies robust mechanism design with a monopolist who has perfect knowledge not about the whole distribution as in the more standard models (e.g. Maskin and Riley 1984), but only about some features of this distribution. With such information, the firm can then narrow down the set of possible distributions to consider and adopt a pricing strategy that maximizes the worst-case profit or minimizes the worst-case regret. This approach tries to address the concern that the optimal mechanism is not robust to the firm having less than exact information on the distribution of consumers' willingness-to-pay, in line with the general research program of robust mechanism design.

The papers in this literature closest to ours are Bergemann and Schlag (2008; 2011) and Carrasco et al. (2018). These papers model a firm that does not know the distribution of consumer types, but has access to imperfect information that allows it to refine the set of possible distributions. Focusing on a linear specification, they assume the firm acts as if it faces an adversarial nature that chooses a distribution to maximize regret.[1] Bergemann and Schlag (2008) assume that the firm knows only an upper bound for the support; Bergemann and Schlag (2011) study the case where the firm also knows that the true distribution of consumers' willingness-to-pay is in a given neighborhood of a given target distribution; Carrasco et al. (2018) posits that the firm knows either the first moment and an upper bound for the support of the distribution,[2] or the first two or three moments of the distribution.

These papers then characterize the regret-minimizing mechanisms, whereby the firm hedges against uncertainty by randomizing over prices. In contrast, we assume that the firm does not know specific features of the distribution but instead has ac-

[1] Bergemann and Schlag (2011) also consider the case where nature minimizes profit and show the firm chooses a deterministic uniform pricing rule.

[2] This is also the case in a closely related paper, Carrasco et al. (2018), where the assumptions on consumer's utility function of linearity in quantity and on linearity of the firm's cost are relaxed.

cess to a sample drawn from the true distribution, from which these features could potentially be estimated and upon which an empirical version of these mechanisms could be implemented. Our analysis shows that, in addition to its other desirable properties, the empirically optimal mechanism generates nearly minimal maximum regret in the sense of these papers.

In another closely related paper, Madarász and Prat (2017) allow for a firm that is uncertain over both the distribution of types and the functional form of consumers' utility functions, while at the same time endowing the firm with a possibly misspecified benchmark model of consumer demand. They provide a uniform bound on regret that depends on the distance between the firm's benchmark model and the truth, where the measure of distance between models is related to the largest absolute value of the difference of willingness-to-pay across all possible types and quantities. Instead of looking at the worst-case scenario solution, the authors show that adjusting the pricing strategy that is optimal for the misspecified model in a specific manner involving this distance leads to a uniform bound on regret.

The present study takes a similar approach to Segal (2003), in that both papers determine the optimal mechanism from an estimate of the distribution function. While we focus on issues of estimation and inference, Segal (2003) is concerned with designing a mechanism that makes optimal use of the information contained in consumers' reported valuations. This optimal mechanism implements a bidding system in which consumers receive the good if their bid is above a threshold price that depends on others' bids. This induces consumers to reveal their own type while simultaneously allowing the firm to use others' bids to infer the distribution. The paper shows that as the number of consumers increases, profits converge to the optimal profit with a known type distribution, and discusses rates of convergence. However, the dependency of the optimal price on others' valuations forces the imposition of strong assumptions on the estimated distribution function, rendering non-parametric methods problematic. Our framework, on the other hand, is singularly suited to rely on non-parametric estimation of and inference on the type distribution and does not require consumers to directly communicate their types.[3] Furthermore, we believe that a key product of our paper is its novel toolkit for conducting inference and obtaining revenue guarantees for any mechanism.

Our work is also related to the literature on sample complexity. Huang et al. (2018) deals with the case where the firm observes independent samples from the unknown true type distribution and consumers have quasilinear-linear utility functions as above. The authors provide probabilistic asymptotic bounds on the number of samples from the true distribution of consumers' valuations that are necessary to achieve a share of $1-\varepsilon$ of the optimal profit. Similarly, our sample complexity results provide non-asymptotic sample upper bounds for regret and profit for specific mechanisms, but for the more general class of utility functions. Fu et al. (2020), in an auction setup, characterize the number of samples from the true distribution of bidders' types that are necessary for the designer to achieve full surplus extraction. Opposite to this paper, their type space is finite and the unknown distribution of types belongs to a finite set of joint distributions over all the bidders' values such that the types are correlated in a specific manner.

A brief outline of the paper is as follows. Section 2 introduces the main theoretical framework. In Section 3, we define our class of empirically optimal mechanisms and examine some of their main properties: asymptotic optimality and profit guarantees. After exploring this particular class of mechanisms, in Section 4 we turn to the question of providing a statistical toolkit to estimate and conduct inference on profit, including optimal expected profit. Section 5 illustrates an extension of our

[3]For instance, the sample may be obtained by implementing revealed-preference elicitation techniques. An example would be providing a menu that corresponds to a direct mechanism inducing perfect discrimination between all possible types – where the quantity is a strictly increasing function of the type – and infer consumers' types from their choices. This would then be incentive compatible if consumers are myopic, the good is durable or if it is possible to exclude them from future purchases.

results to the auction setting with independent private values. Finally, we conclude with a discussion of specific suggestions for further work in Section 6. All omitted proofs are included in the appendix.

2. Setup

Let $\Theta := [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}$, denote the set of feasible consumer types, which are distributed according to the cumulative distribution $F_0 \in \mathscr{F}$, where $\mathscr{F}$ corresponds to the set of all distributions on Θ endowed with the supremum-norm $\|\cdot\|_\infty$, i.e., $\|F\|_\infty := \sup_{t\in\mathbb{R}} |F(t)|$ for all $F \in \mathscr{F}$. Consumers' utility is given by $u(\theta, x, p) = v(\theta, x) - p$, where θ is the consumer's type, $x \in X := [0, \bar{x}]$ denotes quantity and $p \in \mathbb{R}_+$ price. We assume that v is twice continuously differentiable, concave in x, supermodular, $v(\underline{\theta}, x) = v(\theta, 0) = 0$ for all θ and x, increasing in both arguments and, wherever strictly positive, strictly so.

The firm can choose a menu or mechanism M from the set $\mathscr{M}$ of all compact menus $M' \subset X \times \mathbb{R}_+$ containing the element $(0,0)$. These comprise pairs of quantity and prices that the consumers can choose, with the option of consuming nothing being always available. We impose the further restriction that if $(x, 0) \in M'$, then $x = 0$; that is, the firm does not give away strictly positive quantities of the good for free.[4] The firm incurs a twice differentiable, convex and strictly increasing cost for quantity sold, $c : X \to \mathbb{R}$, where $c(0) = 0$. When choosing menu $M \in \mathscr{M}$ subject to a distribution $F \in \mathscr{F}$ of consumer types, the firm's expected profit $\pi(M, F)$ is given by

$$\pi(M,F) := \int p(\theta) - c(x(\theta))\, dF(\theta)$$

for some $(x(\theta), p(\theta)) \in \arg\max_{(x,p)\in M} u(\theta, x, p)$, with ties broken in favor of the firm.

[4]This restriction facilitates Section 4's inference exercise for an arbitrary fixed menu and is without loss of revenue from the firms' perspective.

Our setup encompasses many of the variations in the literature,[5] being mostly the same as that in the one buyer version of Myerson (1981) and Maskin and Riley (1984), except that there c is assumed to be linear and $\mathscr{F}$ corresponds to the distributions with a strictly positive density. Mussa and Rosen (1978), instead, assume that $v(\theta,x) = \theta \cdot x$ and specify c to be strictly convex. In Bergemann and Schlag (2011) and Carrasco et al. (2018), $v(\theta,x) = \theta \cdot x$ and $\mathscr{F}$ is a subset of distributions that satisfy some pre-specified conditions.

We consider the case where neither the firm nor the consumers know the true distribution of types, $F_0 \in \mathscr{F}$, which motivates the choice of dominant strategies as our solution concept. Instead, the firm has access to a sample $S^n = (\theta_i, i = 1,\ldots,n) \in \Theta^n$, $n \in \mathbb{N}$, where each θ_i is independently drawn from F_0. This gives rise to the problem of selecting a menu depending on the realized sample. Let $\mathscr{S}$ denote the set of all samples, $\mathscr{S} := \bigcup_{n\in\mathbb{N}} \Theta^n$. A *sample-based mechanism* is then a mapping $M_S : \mathscr{S} \to \mathscr{M}$, which selects a specific menu depending on the realized sample.

The robust mechanisms mentioned earlier can easily be adjusted to incorporate the information in the sample in order to estimate the features of the distribution that they assume to be known. For example, given a particular sample S^n, the firm can then estimate the support of the true distribution and implement the mechanism given in Bergemann and Schlag (2008). Alternatively, it can estimate the first few moments of the true distribution and implement the mechanism in Carrasco et al. (2018). A natural question is then whether, given sampling uncertainty, these sample-based mechanisms would exhibit probabilistic robustness properties akin to the deterministic ones that hold when these features are perfectly known. This question is relevant in practice, since *any* information about an unknown distribution is usually obtained from finite data. We therefore take a more direct and,

[5] It does not include, for instance, the case where the firm's cost depends directly on the consumers' type as in Example 4.1 in Toikka (2011).

arguably, simpler approach that makes full use of the sample itself to "learn" about the underlying true distribution and inform mechanism choice.

3. Empirically Optimal Mechanisms

In this section, we introduce the class of empirically optimal mechanisms. This class of mechanisms is defined by two elements: an estimator of the true distribution and a mapping that takes each estimated distribution to a menu that would be optimal were the estimate to coincide with the true distribution.[6]

Let $\widehat{\mathscr{F}}$ denote the set of estimators that are consistent for F_0, that is, the set of estimators $\hat{F}$ such that (i) $\hat{F} : \mathscr{S} \to \mathscr{F}$; and (ii) $\|\hat{F}(S^n) - F_0\|_\infty \xrightarrow{p} 0$, for any F_0 in $\mathscr{F}$. Let $M^* : \mathscr{F} \to \mathscr{M}$ be a fixed selection from the set of optimal menus for every distribution, that is, $\forall F \in \mathscr{F}$, $M^*(F) \in \mathscr{M}^*(F) := \arg\max_{M \in \mathscr{M}} \pi(M, F)$. An *empirically optimal mechanism* $\hat{M}^*$ is a sample-based mechanism that simply joins together a consistent estimator and a selection from the set of optimal menus, that is, $\hat{M}^* = M^* \circ \hat{F}$. We refer to the set of empirically optimal mechanisms as $\widehat{\mathscr{M}}^*$.

While extremely simple, nothing ensures us that such a sample-based mechanism is either well-defined in general environments or that it constitutes a reasonable approach to pricing under uncertainty. The purpose of this section is to address these issues and show that this "naive" approach to pricing delivers several desirable properties including strong probabilistic robustness guarantees.

3.1 Existence

In order to show that the set of empirically optimal mechanisms $\widehat{\mathscr{M}}^*$ is nonempty, we start by briefly noting that an optimal menu exists for any distribution $F \in \mathscr{F}$,

[6] Note that, if the seller did have a prior, $\mu \in \Delta(\Delta(\Theta))$, over the set of possible distributions, it would still be sufficient to consider only the expected distribution according to the seller's posterior, $\mathbb{E}_\mu[F|S^n] \in \mathscr{F}$, in order to determine which mechanism to choose. That is, due to linearity of the profit function on the distribution, $\max_{M \in \mathscr{M}} \mathbb{E}_\mu[\pi(M,F)|S^n] = \max_{M \in \mathscr{M}} \pi(M, \mathbb{E}_\mu[F|S^n])$.

that is, $\mathscr{M}^*(F) \neq \emptyset$ for all $F \in \mathscr{F}$. We include a formal proof of this statement in the appendix.

Since this implies that a selection $M^* : \mathscr{F} \to \mathscr{M}$ such that $M^*(F) \in \mathscr{M}^*(F)$ exists and as there are consistent estimators for any $F_0 \in \mathscr{F}$, the set of empirically optimal mechanisms $\widehat{\mathscr{M}}^*$ is nonempty. An example of a consistent (and unbiased) estimator for F_0 is the empirical cumulative distribution, defined as $\hat{F}(S^n)(\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{\theta_i \leq \theta\}}$, which, for any $F_0 \in \mathscr{F}$, has a uniform rate of convergence, that is, $\|\hat{F}(S^n) - F_0\|_\infty \xrightarrow{p} 0$ (in fact, by the Glivenko–Cantelli theorem, uniform convergence occurs almost surely).

Note that under some specific conditions, an exact characterization of the set of optimal menus is known. For instance, if $v(\theta, x) = \theta \cdot x$ and $c(x) = \overline{c} \cdot x$, it is well-known that any optimal mechanism is F-almost everywhere equal to an indicator function $\mathbf{1}_{\{p^* \geq \theta\}}$, where $p^* \in \arg\max_{p \in \text{supp}(F)} (p - \overline{c}) \cdot \int \mathbf{1}_{\{p \geq \theta\}} dF(\theta)$. Such an explicit solution simplifies the problem of characterizing empirically optimal mechanisms dramatically. When, instead, v is multiplicatively separable, and F_0 is absolutely continuous and has convex support, any optimal mechanism is almost everywhere equal to pointwise maximization of the ironed virtual value as shown in Toikka (2011). Hence, the problem of characterizing empirically optimal mechanisms can be made computationally tractable by ensuring not only consistency, but also absolutely continuity and convex support of estimates $\hat{F}(S^n)$. Although the empirical cumulative distribution is not absolutely continuous, one such estimator $\hat{F}$ can be easily obtained by adopting any smooth interpolation of the empirical cumulative distribution, for example a linear interpolation, a cubic spline or an interpolation relying on Bernstein polynomials.[7]

[7] In the appendix, we provide a simple proof for the fact that linear interpolations retain the uniform convergence (and therefore consistency) properties of the empirical distribution. See Babu et al. (2002) and Leblanc (2011) for details on interpolation of the empirical cumulative distribution using Bernstein polynomials.

3.2 Asymptotic Optimality

Having defined our class of empirically optimal mechanisms, we establish in this section that they are asymptotically optimal: the realized expected profit given the mechanism converges in probability to the optimal expected profit as the sample size grows.

Such convergence is not guaranteed for arbitrary sample-based mechanisms, even for those that have desirable robustness properties, as it requires that the sample-based mechanism makes full use of the sample. For instance, if the mechanism relies only on statistics such as estimates for a finite number of moments or the support of the distribution, then it is immediate that it will not, in general, converge in probability to the optimal expected profit. Relying on an estimator for the true distribution itself (an infinite-dimensional parameter) is then key to obtaining asymptotic optimality.

We start by making an important observation:

Lemma 1. For any $M \in \mathscr{M}$, $\pi(M,F)$ is Lipschitz continuous in $F \in \mathscr{F}$, with a Lipschitz constant L that does not depend on M.

We defer the proof to the appendix, but highlight the main steps here. The proof of Lemma 1 first makes use of the revelation principle to focus on the elements of any arbitrary menu that are payoff relevant, as these are given by a bounded non-decreasing function, and hence of bounded variation. Then, we appeal to a result that provides an upper bound on Riemann–Stieltjes integrals of functions of bounded variation to obtain the result. In particular, we find a Lipschitz constant of at most

$$L = 2\left(v(\bar{\theta},\bar{x}) + (\bar{\theta} - \underline{\theta}) \cdot \max_{\theta' \in \Theta} v_1(\theta',\bar{x}) + c(\bar{x})\right).$$

For every $F \in \mathscr{F}$, define the firm's value function as

$$\Pi(F) := \sup_{M \in \mathscr{M}} \pi(M, F).$$

Lemma 1 leads to a further result, this time regarding (Lipschitz) continuity of the value function:

Lemma 2. Π is Lipschitz continuous, with Lipschitz constant L.

Proof. For any $F, G \in \mathscr{F}$,

$$|\Pi(F) - \Pi(G)| = \left| \sup_{M \in \mathscr{M}} \pi(M,F) - \sup_{M \in \mathscr{M}} \pi(M,G) \right| \leq \sup_{M \in \mathscr{M}} |\pi(M,F) - \pi(M,G)| \leq L \cdot \|F - G\|_\infty$$

□

Finally, the desired result of asymptotic optimality of our class of empirically optimal mechanisms follows immediately:

Proposition 1. Let $\hat{M}^*$ be an empirically optimal mechanism given by $\hat{M}^* = M^* \circ \hat{F}$. Then, $|\pi(\hat{M}^*(S^n), F_0) - \Pi(F_0)| \xrightarrow{p} 0$.

Proof. By Lemmata 1 and 2,

$$\begin{aligned} |\pi(\hat{M}^*(S^n), F_0) - \Pi(F_0)| &\leq |\pi(\hat{M}^*(S^n), F_0) - \pi(\hat{M}^*(S^n), \hat{F}(S^n))| + |\pi(\hat{M}^*(S^n), \hat{F}(S^n)) - \Pi(F_0)| \\ &\leq L \cdot \|\hat{F}(S^n) - F_0\|_\infty + |\Pi(\hat{F}(S^n)) - \Pi(F_0)| \leq 2L \cdot \|\hat{F}(S^n) - F_0\|_\infty \xrightarrow{p} 0. \end{aligned}$$

□

Proposition 1 provides a simple justification for using an empirically optimal mechanism to guide the firm's pricing strategy: As the sample size grows large, such sample-based mechanisms deliver an expected profit close to the optimal one. We stress the minimal informational assumptions made. In particular, this result does not depend on the firm knowing the support of the true distribution F_0 ex ante, as the empirically optimal mechanism is defined making use only of the estimated cumulative distribution and there are consistent estimators that require no assumptions on (and in fact, asymptotically learn) the support of F_0. Moreover, as $\Pi(F_0) - \pi(\hat{M}^*(S), F_0) \leq 2L\|\hat{F}(S) - F_0\|_\infty$, we conclude that empirically optimal mechanisms are robust in a sense akin to Bergemann and Schlag (2011), since for any $\varepsilon > 0$, samples inducing $\|\hat{F}(S) - F_0\|_\infty < \varepsilon$ imply that $\Pi(F_0) - \pi(\hat{M}^*(S^n), F_0) \leq 2L\varepsilon$.

3.3 Robustness Properties

While some sample-based implementations of existing robust mechanisms would not be asymptotically optimal, it is possible that others would. Thus, an obvious question is: Do empirically optimal mechanisms provide robustness guarantees with finite samples that render them especially appealing? In this section, we argue that this is indeed the case.

Robustness properties of mechanisms regard the worst-case scenarios. The existing literature has focused on two main properties. One corresponds to the worst-case profit that the firm can expect given that the true distribution lies in a specific set $A \subseteq \mathscr{F}$. This is in part motivated by appealing to the characterization of preferences exhibiting ambiguity aversion by Gilboa and Schmeidler (1989), which entails a maxmin representation, whereby the decision-maker (here, the firm) evaluates each act (mechanism) by assuming the worst-case payoff. The robustness of a specific mechanism according to this criterion is then given by the lower bound on expected profit it can attain, $\min_{F \in A} \pi(M, F)$. The second robustness criterion that has been considered in the literature depends on the notion of regret: How much profit the firm may be forgoing by committing to mechanism M when the true distribution is F_0, that is, $R(M, F_0) := \Pi(F_0) - \pi(M, F_0)$.

A simple implication of Lemmata 1 and 2 is that we can immediately obtain probabilistic bounds on regret and on how far the realized expected profit may be from the profit the firm expects to obtain given its estimated distribution.

Proposition 2. Let $\hat{M}^* = M^* \circ \hat{F}$ be an empirically optimal mechanism. Suppose that $\hat{F} \in \widehat{\mathscr{F}}$ is such that $\forall S^n \in \mathscr{S}$, $\mathbb{P}(\|\hat{F}(S^n) - F_0\|_\infty > \delta) \leq p(n,\delta)$ for some function $p : \mathbb{N} \times \mathbb{R}_+ \to [0,1]$. Then,

(i) $\mathbb{P}\left(|\pi(\hat{M}^*(S^n), \hat{F}(S^n)) - \pi(\hat{M}^*(S^n), F_0)| > \delta\right) \leq p(n, \delta/L)$; and

(ii) $\mathbb{P}\left(R(\hat{M}^*(S^n), F_0) > 2\delta\right) \leq p(n, \delta/L)$,

where L denotes the Lipschitz constant from Lemma 1.

While Proposition 2 is a trivial observation, it enables the firm to obtain strong, non-asymptotic probabilistic bounds on both profit and regret whenever basing an empirically optimal mechanism on an estimator $\hat{F}$ with specific properties. Whenever the firm knows an upper bound $\bar{\theta}$ for the support of the true distribution, L can be obtained in a way that depends exclusively on known constants and these probabilistic profit and regret guarantees can be computed explicitly. The next two examples illustrate how this result can be applied by focusing on estimators $\hat{F}$ with well-known properties, such as the empirical cumulative distribution and smooth interpolations of it.

Example 1. Let $\hat{M}^* = M^* \circ \hat{F}$ be any empirically optimal mechanism such that $\hat{F}$ denotes the cumulative distribution estimator. By the Dvoretzky–Kiefer–Wolfowitz inequality (Dvoretzky et al. 1956) with Massart's (1990) constant, we have that

$$\mathbb{P}(\|\hat{F}_n - F_0\|_\infty > \delta) \leq 2\exp\left(-2n\delta^2\right)$$

and hence $p(n,\delta) = 2\exp\left(-2n\delta^2\right)$. Then, Proposition 2 applies and we can obtain regret lower than 2δ with probability of at least $1 - p(n,\delta)$ and a confidence bound with range 2δ such that the true expected profit differs from $\pi(\hat{M}^*(S^n), \hat{F}(S^n))$ by less than δ also with probability greater than $1 - p(n,\delta)$. □

Example 2. Suppose that $\mathscr{F}$ is restricted to the set of absolutely continuous distributions on Θ, of which F_0 is known to be an element of, and that v is multiplicatively separable in $\theta \in \Theta$ and $x \in X$. If we were to constrain $\hat{F}(S)$ to also be absolutely continuous, admitting a strictly positive density and having convex support, an analytic characterization of $M^*(\hat{F}(S))$ is known, given by pointwise maximization of the ironed virtual value given $\hat{F}(S)$.[8] This would then simplify the computational cost of finding the optimal mechanism. Take $\hat{F}$ to be any interpolation of the empirical cumulative distribution that results in a valid distribution function that is absolutely continuous and has convex support, such as the linear interpolation (see Lemma 10 in the appendix). Note that given that F_0 is atomless by assumption, $\mathbb{P}(\exists k, \ell \in \{0,1,\ldots,n-1\} : s_k = s_\ell) = 0 \ \forall n \in \mathbb{N}$, and thus the linear interpolation is well-defined with probability 1. Furthermore, for any such interpolation $\hat{F}$, with probability 1, the estimate given by $\hat{F}(S^n)$ differs from the empirical cumulative distribution $\hat{F}^E(S^n)$ by at most $1/n$ at any given point. Hence,

$$\begin{aligned}
\mathbb{P}(\|\hat{F}(S^n) - F_0\|_\infty > \delta) &\le \mathbb{P}(\|\hat{F}^E(S^n) - F_0\|_\infty + \|\hat{F}(S^n) - \hat{F}^E(S^n)\|_\infty > \delta) \\
&\le \mathbb{P}(\|\hat{F}^E(S^n) - F_0\|_\infty > \delta - 1/n) \\
&\le 2\exp\left(-2n(\delta - 1/n)^2\right),
\end{aligned}$$

where, again, the last inequality follows from the Dvoretzky–Kiefer–Wolfowitz inequality with Massart's constant. It follows that $p(n,\delta) = 2\exp\left(-2n(\delta - 1/n)^2\right)$. As $\hat{F} \in \widehat{\mathscr{F}}$, Proposition 2 applies and the regret and confidence bounds obtain. □

[8] Toikka (2011) has shown that for any absolutely continuous distribution F on Θ with a strictly positive density, the set of maximizers $\mathscr{M}^*(F)$ pointwise maximize $\bar{J}(\theta)v(\theta,x) - c(x)$, where $\bar{J}(\theta)$ is the ironed version of $J(\theta) := \theta - \frac{1-F(\theta)}{f(\theta)}$.

As this next example shows, under some assumptions on the true distribution F_0 one can even obtain not only non-asymptotic, but also non-probabilistic regret and confidence bounds.

Example 3. Suppose that it is known that the true distribution F_0 admits a density f_0 with total variation bounded by $B < \infty$ and has support contained in $[0,1]$.[9] Let V_B denote the set of all densities on $[0,1]$ with total variation bounded by B. Take any kernel density estimator, given by

$$\hat{f}(S^n)(u) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{u-\theta_i}{h}\right),$$

where h is a smoothing parameter such that $h \to 0$ as $n \to \infty$, θ_i denotes the value of observation i, $K \geq 0$ and $\int K(u)du = 1$. From theorem 3 in Datta (1992), we know that, for any such kernel density estimator $\hat{f}_n$, one has

$$\sup_{f \in V_B} \|\hat{f}(S^n) - f\|_1 \leq (2B+1)k_1 h + \left(\frac{k_2}{nh}\right)^{1/2},$$

where $k_1 := \int |u|K(u)du$ and $k_2 := \int K^2(u)du$. Define $q(n,h) := (2B+1)k_1 h + \left(\frac{k_2}{nh}\right)^{1/2}$ and let $\hat{F}(S^n)(\theta) := \int \mathbf{1}_{u\leq\theta}\hat{f}(S^n)(u)du$ denote the estimated cumulative distribution. Note that there are several kernel density estimators available such that

$$k_1 = \int |u|K(u)du < \infty \text{ and } k_2 = \int K^2(u)du < \infty,$$

e.g. if K is the uniform kernel, triangle, Epanechnikov, among others.
In order for $\hat{F}$ to belong to $\widehat{\mathscr{F}}$ we need for it to (1) be uniformly consistent and (2) to have compact support. There are several ways to achieve this, namely by relying on a kernel such that $\int_{|u|>\delta} K(u)du = 0$ for some finite $\delta > 0$ – which is satisfied by most of the standard kernels, namely the ones cited above. For any such kernel, $k_1, k_2 < \infty$ and $\forall S^n \in \mathscr{S}$, supp $\hat{F}(S^n) \subseteq \Theta$ is compact, by choosing lower and upper

[9]This can be generalized to some closed interval in $\mathbb{R}$.

bounds $\underline{\theta}$, $\bar{\theta}$ appropriately.

Therefore,

$$\begin{aligned}\|\hat{F}(S^n)-F_0\|_\infty &= \sup_{\theta\in\Theta}\left|\hat{F}(S^n)(\theta)-F_0(\theta)\right| = \sup_{\theta\in\Theta}\left|\int \mathbf{1}_{u\le\theta}\left(\hat{f}(S^n)(u)-f_0(u)ds\right)\right| \\ &\le \sup_{\theta\in\Theta}\int \mathbf{1}_{u\le\theta}|\hat{f}(S^n)(u)-f_0(u)|ds \le \|\hat{f}(S^n)-f_0\|_1 \le q(n,h),\end{aligned}$$

where $\|\cdot\|_1$ denotes the L^1 norm, i.e. $\|f\|_1 := \int|f(u)|du$. It follows immediately that if $h\to 0$ and $n\cdot h\to\infty$, $\|\hat{F}(S^n)-F_0\|_\infty\to 0$ and, consequently, $\hat{F}\in\widehat{\mathscr{F}}$. The argument above implies that

$$\begin{aligned}&R(\hat{M}^*(S^n),F_0)\le q(n,h)/2L;\\ &|\pi(\hat{M}^*(S^n),\hat{F}(S^n))-\pi(\hat{M}^*(S^n),F_0)|\le q(n,h)/L.\end{aligned}$$

As such, if f_0 has total variation bounded by B, we can obtain non-probabilistic bounds on regret and expected profit when implementing an empirically optimal mechanism based on kernel density estimation. □

Other sample-based mechanisms could potentially yield even stronger robustness properties. However, for any sample-based mechanism M_S, one has

$$\begin{aligned}R(M_S(S^n),F_0)-R(\hat{M}^*(S^n),F_0) &= \pi(M_S(S^n),F_0)-\pi(\hat{M}^*(S^n),F_0)\\ &\le \Pi(F_0)-\pi(\hat{M}^*(S^n),F_0) = R(\hat{M}^*(S^n),F_0).\end{aligned}$$

As the examples above show, choosing $\hat{F}$ appropriately ensures that $\mathbb{P}(R(\hat{M}^*(S^n),F_0)>\delta)$ declines exponentially with n. Hence, the gains in profit and regret guarantees from implementing alternative pricing policies are modest at best, in a formal sense.

To conclude this section, we note that Proposition 2 can instead be used to determine how many samples the firm requires in order to obtain specific robustness guarantees when relying on empirically optimal mechanisms. That is, another reading of Proposition 2 is that the firm only needs at most N samples – where N is the smallest integer such that $\alpha \geq p(N, \delta/L)$ – to secure at most 2δ of regret with probability $1-\alpha$. Alternatively, the same N samples provide a (conservative) confidence interval for profit with range 2δ with confidence level of $1-\alpha$. This results in a non-asymptotic sample complexity bound for a specific class of sample-based mechanisms. In contrast to the sample-complexity bounds obtained in Huang et al. (2018), which pertain to the share of the optimal profit that the firm is able to secure, $R(M, F_0)/\Pi(F_0)$, we focus on bounding regret directly and our bounds are not asymptotic, that is, they hold for finite samples.

4. Inference and Robustness

While the revenue and regret guarantees derived in the previous section are useful, upon observing a particular sample and deciding on a menu, the firm may want to be able to obtain consistent projections for the expected profit. In this section, we show how to obtain consistent and unbiased estimates and conduct inference on the expected profit.

Our results enable unbiased and consistent estimation of and inference on the expected profit not only of empirically optimal mechanisms but of any given mechanism $M \in \mathcal{M}$. Moreover, we show how empirically optimal mechanisms serve a special purpose, in that they can be used to estimate and conduct inference on the optimal expected profit. With these tools, one can provide confidence intervals for the expected profit with specific asymptotic coverage for any one mechanism, calculate probabilistic bounds for regret, or test whether one mechanism yields higher expected profit than another.

4.1 Expected Profit

An immediate consequence of Lipschitz continuity of the firm's profit function with respect to the distribution is that, for any mechanism and any consistent estimator of the distribution of types, one can consistently estimate the expected profit that such a mechanism would generate.

Proposition 3. For any true distribution $F_0 \in \mathscr{F}$, consistent estimator $\hat{F} \in \widehat{\mathscr{F}}$, and mechanism $M \in \mathscr{M}$, we have $\pi(M,\hat{F}(S^n)) \xrightarrow{p} \pi(M,F_0)$. Moreover, if $\hat{F}$ is an unbiased estimator such as the empirical distribution function, $\mathbb{E}\left[\pi(M,\hat{F}(S^n))\right] = \pi(M,F_0)$, that is, the plug-in estimator is also unbiased.

Proof. By Lemma 1, we have that, $\forall \hat{F} \in \widehat{\mathscr{F}}$ and $\forall M \in \mathscr{M}$,

$$|\pi(M,\hat{F}(S^n)) - \pi(M,F_0)| \leq L\|\hat{F}(S^n) - F_0\|_\infty \xrightarrow{p} 0$$

Moreover, if $\hat{F}$ is an unbiased estimator of F_0, by linearity of $\pi(M,F)$ in F, we have that $\mathbb{E}\left[\pi(M,\hat{F}(S^n))\right] = \pi(M,\mathbb{E}\left[\hat{F}(S^n)\right]) = \pi(M,F_0)$. □

An important aspect in estimation is the ability to conduct inference. For instance, the firm could be interested in using statistical inference in order to compare different mechanisms, that is, to test whether a specific mechanism would deliver a higher expected profit than another. Another possible application would be to obtain valid confidence intervals for the expected profit under a particular mechanism, as it is an arguably crucial tool for the development of routine business activities such as drawing up budgets under different scenarios, with varying degrees of confidence.

While the firm could potentially derive confidence intervals for the expected profit by adjusting Proposition 2 and Example 1 to the mechanism it is considering, these bounds would exhibit two drawbacks when used for this purpose. First, they require knowledge of $\bar{\theta}$, the upper bound on the type distribution. Second, and more critically, they generally do not provide the correct asymptotic coverage, that is, they would be exceedingly conservative.

In order to address these drawbacks, we suggest a simple estimation procedure that does yield asymptotically valid inference. First, we focus on the empirical distribution function as our estimator, since it admits a functional central limit theorem (by Donsker's theorem) when properly centered and rescaled. Then, because profit is linear and continuous in the type distribution, it is Fréchet differentiable, with derivative given by $\dot{\pi}_M(\cdot) = \pi(M,\cdot)$.[10] One can thus obtain the asymptotic distribution of our consistent estimator for the expected profit $\pi(M,\hat{F}(S^n))$ by appealing to a simple functional Delta method result.

Theorem 1. Let $\hat{F}$ denote the empirical distribution estimator. Then, $\forall M \in \mathscr{M}$, $\forall F_0 \in \mathscr{F}$,

$$\sqrt{n}\left(\pi(M,\hat{F}(S^n)) - \pi(M,F_0)\right) \xrightarrow{d} N(0,\sigma^2_{M,F_0}),$$

where $\sigma^2_{M,F_0} := \mathbb{E}\left[(\dot{\pi}_M(\delta_\theta - F_0))^2\right] = \mathbb{E}\left[(\pi(M,\delta_\theta - F_0))^2\right]$ and $\theta \sim F_0$.

The function δ_θ denotes the cumulative distribution associated with a Dirac measure at θ, that is, $\delta_\theta(x) = \mathbf{1}_{\{x\geq\theta\}}$.

Theorem 1 states that the distribution of the empirical process,

$$G_n := \sqrt{n}\left(\pi(M,\hat{F}(S^n)) - \pi(M,F_0)\right),$$

converges weakly to $N(0,\sigma^2_{M,F_0})$. A question then arises of how to estimate, in practice, the asymptotic distribution in a consistent manner, as it depends on the unknown distribution F_0. We provide two alternatives. One option is the use of a plug-in estimator for σ^2_{M,F_0}. This can be done directly – as the functional depen-

[10]The Fréchet derivative $\dot{\pi}_M$ is defined on the space of functions on Θ of bounded variation endowed with the supremum-norm. We refer to the appendix for details.

dence of σ^2_{M,F_0} on F_0 is known and a consistent estimate for F_0 is readily available –, or by following other plug-in methods as those in Shao (1993). Another option is to rely on the classical bootstrap to approximate the distribution of G_n by the distribution of $\hat{G}_n := \sqrt{n}\left(\pi(M,\hat{F}(S^n_B)) - \pi(M,\hat{F}(S^n))\right)$, conditional on S^n, where S^n_B denotes the resampling of n observations from S^n with uniform weights. That this approach does in fact consistently estimate the limiting distribution was shown by Parr (1985a, Theorem 4).

We note that other bootstrap methods would also yield consistent estimates, such as subsampling (bootstrap without replacement) (Politis and Romano 1994) or Jackknife procedures (Parr 1985b). Moreover, given that the Fréchet derivative of profit is sufficiently well-behaved, under some smoothness assumptions on the true distribution F_0, smoothed versions of the bootstrap (Cuevas and Romo 1997) can also be considered.

4.2 Optimal Profit and Regret

We now extend our statistical inference results, highlighting how empirically optimal mechanisms can be used to provide a consistent and asymptotically normal estimator for optimal profit.

Given that the firm's value function is Lipschitz continuous in the distribution and that empirically optimal mechanisms attain the optimal profit for a consistent estimate of the true distribution, it is easy to see that one can then use them as a tool to consistently estimate the optimal profit that the firm would obtain, were it to know F_0. We formalize this observation as follows:

Proposition 4. For any true distribution $F_0 \in \mathscr{F}$ and empirically optimal mechanism $\hat{M}^*$ given by $\hat{M}^* = M^* \circ \hat{F}$, $\pi(\hat{M}^*(S^n), \hat{F}(S^n)) \xrightarrow{p} \Pi(F_0)$.

Proof. Similarly to Proposition 3, we again have that, by Lemma 2, Π is Lipschitz continuous and therefore,

$$|\pi(\hat{M}^*(S^n), \hat{F}(S^n)) - \Pi(F_0)| = |\Pi(\hat{F}(S^n)) - \Pi(F_0)| \le L\|\hat{F}(S^n) - F_0\|_\infty \xrightarrow{p} 0.$$

□

It is less straightforward that one could take an approach to conducting inference on the optimal profit similar to that derived for fixed mechanisms. Specifically, this would require proving that the firm's value function Π is also Fréchet differentiable.[11] We confirm that indeed such an approach is valid by proving an interesting technical result in this generalized Maskin-Riley setup: an envelope theorem for the firm's value function. In other words, our next result shows that the value function is Fréchet differentiable at any distribution $F \in \mathscr{F}$ and that its Fréchet derivative coincides with that of the expected profit at F with the optimal menu for F.

Theorem 2 (Envelope Theorem)**.** Π is Fréchet differentiable at all $F \in \mathscr{F}$. Moreover, its Fréchet derivative at F is given by $\dot{\Pi}_F = \dot{\pi}_{M_F}$, $\forall M_F \in \mathscr{M}^*(F)$.

Then, defining the empirical process $\hat{G}_n := \sqrt{n}\left(\Pi(\hat{F}(S_B^n)) - \Pi(\hat{F}(S^n))\right)$, conditional on S^n, an adapted version of Theorem 1 ensues:

Theorem 3. Let $\hat{F}$ denote the empirical distribution estimator. Then, $\forall F_0 \in \mathscr{F}$,

$$\sqrt{n}\left(\Pi(\hat{F}(S^n)) - \Pi(F_0)\right) \xrightarrow{d} N(0, \sigma_{F_0}^2),$$

where $\sigma_{F_0}^2 = \mathbb{E}\left[\left(\dot{\Pi}_{F_0}(\delta_\theta - F_0)\right)^2\right] = \mathbb{E}\left[(\pi(M_F, \delta_\theta - F_0))^2\right]$ and $\theta \sim F_0$. Moreover, $\hat{G}_n \xrightarrow{d} N(0, \sigma_{F_0}^2)$.

[11]In fact, the now standard functional Delta method requires only the weaker notion of Hadamard differentiability; see, e.g., van der Vaart and Wellner (1996, ch. 3.9). However, the stronger notion of Fréchet differentiability has the benefit of allowing us to bypass the measurability complications that arise when using weaker notions.

In this case, opposite to the case of inference under a fixed mechanism, we do not have a valid (consistent) plug-in estimator for $\sigma^2_{F_0}$, which depends on F_0 and, more problematically, also on an optimal mechanism under the distribution F_0, $M_0 \in \mathscr{M}^*(F_0)$. An important argument in favor of a bootstrap approach to estimating the asymptotic distribution in this case is that it bypasses this issue.

Under some conditions, it may make sense to rely on different estimators. For instance, as discussed in Example 2, when $v(\theta, x)$ is multiplicatively separable and F_0 is known to be absolutely continuous and with compact and convex support, the functional form of the solution is exactly known. This allows for a drastic simplification of the problem from a computational point of view, since it dispenses with the hurdle of finding the optimal mechanism for a given distribution. Especially in the context of implementing a bootstrap approach, the gains can be substantial. However, an estimate of the ironed virtual value depends on a suitable estimate of the density f. Therefore, we find it especially relevant that, when F_0 is known to be absolutely continuous, one can use as an estimator the simple linear interpolation of the empirical distribution discussed earlier to obtain a bootstrap estimator for the asymptotic distribution. Further, we note that this extremely simple approach is not only consistent for the true distribution F_0, but also for its density.

Proposition 5. Let $\hat{F}$ denote the linear interpolation of the empirical distribution estimator. Then, for any absolutely continuous $F_0 \in \mathscr{F}$,

(1) $\sqrt{n}\left(\Pi(\hat{F}(S^n)) - \Pi(F_0)\right) \xrightarrow{d} N(0, \sigma^2_{F_0})$, where $\sigma^2_{F_0} = \mathbb{E}\left[\left(\dot{\Pi}_{F_0}(\delta_\theta - F_0)\right)^2\right]$;

(2) $\hat{G}_n \xrightarrow{d} N(0, \sigma^2_{F_0})$; and

(3) $\|\hat{f}(S^n) - f_0\|_1 \xrightarrow{p} 0$, where $\hat{f}(S^n)$ and f_0 denote the Radon-Nikodym derivatives of $\hat{F}(S^n)$ and F_0, respectively.

While estimating the optimal expected profit may be relevant for investment decisions, as it provides an upper bound on the return of a given investment, these results can also be used to estimate regret. As regret, $R(M, F)$, is given by $R(M, F) =$

$\Pi(F) - \pi(M,F)$, it is Fréchet differentiable at any distribution $F \in \mathscr{F}$, for any fixed $M \in \mathscr{M}$, as the sum of Fréchet differentiable functionals is itself Fréchet differentiable. Then, using similar arguments as those in Proposition 4 and Theorem 3, one can conduct inference on regret. Consequently, for any mechanism $M \in \mathscr{M}$, one can not only obtain asymptotically valid probabilistic bounds for expected profit, but also for regret.

4.3 Simulation Evidence

To conclude this section, we present empirical evidence on the finite sample properties of our estimators.

We conduct Monte Carlo simulations on the empirical coverage of the confidence intervals for expected profit under a fixed mechanism – uniform pricing, with the price set at 1/2 – and for the optimal expected profit, using empirically optimal mechanisms for the empirical distribution. We use the approximation obtained by classic bootstrapping (N out of N), which we have showed to be asymptotically valid. We show simulation results for confidence levels $\alpha \in \{.1, .05, .01\}$, with varying sample size N. For each sample size, we draw 1,000 samples and, for each sample, we estimate the confidence interval by drawing 1,000 bootstrap samples from the original sample.

We focus on the case where consumers have quasilinear-linear utility and the unit cost is normalized to zero, as in Bergemann and Schlag (2011) and Carrasco et al. (2018). We show results for three different parameterizations of F_0 relying on the Beta distribution: Beta(1/4,1/4), Uniform(0,1) and Beta(4,4).[12]

In Table 1 we present evidence for the empirical coverage frequency at sample sizes of 500, 1,000 and 2,500. As is immediate upon inspection of the table, our estimators have extremely good finite sample properties, with the empirical cover-

[12]The empirical coverage results were consistent across other parameterizations and using Beta mixtures or mixtures with degenerate distributions.

Table 1. **Empirical Coverage Frequencies**

(a) **Profit with Fixed Mechanism**

		Beta(1/4,1/4)			Unif(0,1)			Beta(4,4)		
	$1-\alpha$	.90	.95	.99	.90	.95	.99	.90	.95	.99
	500	.891	.944	.983	.892	.946	.987	.895	.943	.988
N	1,000	.901	.934	.985	.901	.946	.986	.896	.953	.985
	2,500	.909	.954	.989	.903	.953	.989	.889	.948	.985

(b) **Optimal Profit**

		Beta(1/4,1/4)			Unif(0,1)			Beta(4,4)		
	$1-\alpha$	.90	.95	.99	.90	.95	.99	.90	.95	.99
	500	.881	.933	.989	.894	.945	.983	.895	.950	.987
N	1,000	.886	.945	.985	.888	.943	.981	.894	.941	.984
	2,500	.910	.952	.988	.890	.941	.988	.884	.933	.984

Note: This table shows the frequency with which the estimated confidence interval with asymptotic coverage of $1-\alpha$ contained the true expected profit, $\pi(M,F_0)$ in the case of a fixed mechanism and $\Pi(F_0)$ in the case of the optimal expected profit. The fixed mechanism corresponds to uniform pricing at 1/2. The estimated confidence interval followed a centered bootstrap procedure with 1,000 samples redrawn with replacement from the original sample with 1,000 iterations, for each sample size N.

age frequencies being very close to the theoretical asymptotic coverage probability, regardless of which of the three distributions is considered. We also investigated the behavior of our estimators under small samples. As Figure 1 shows, they fare reasonably well for sample sizes between 50 and 300.

We also considered the regret incurred by adopting an empirically optimal mechanism that depends on the empirical distribution with finite samples. As illustrated in Figure 2, we empirically study the average regret as a share of the optimal expected profit, that is, $\big(\Pi(F_0)-\pi(\hat{M}^*(S^n),F_0)\big)/\Pi(F_0)$. The average is taken across 1,000 samples of varying size in increments of 10 observations. Even with just 50 observations, the empirically optimal mechanism on average attains regret that is under 4% of the optimal expected profit. For the purpose of comparison, the robust mechanism in Carrasco et al. (2018, Section 5.1), relying on an estimate of the mean and assuming knowledge of the upper bound of the distribution, exhibits

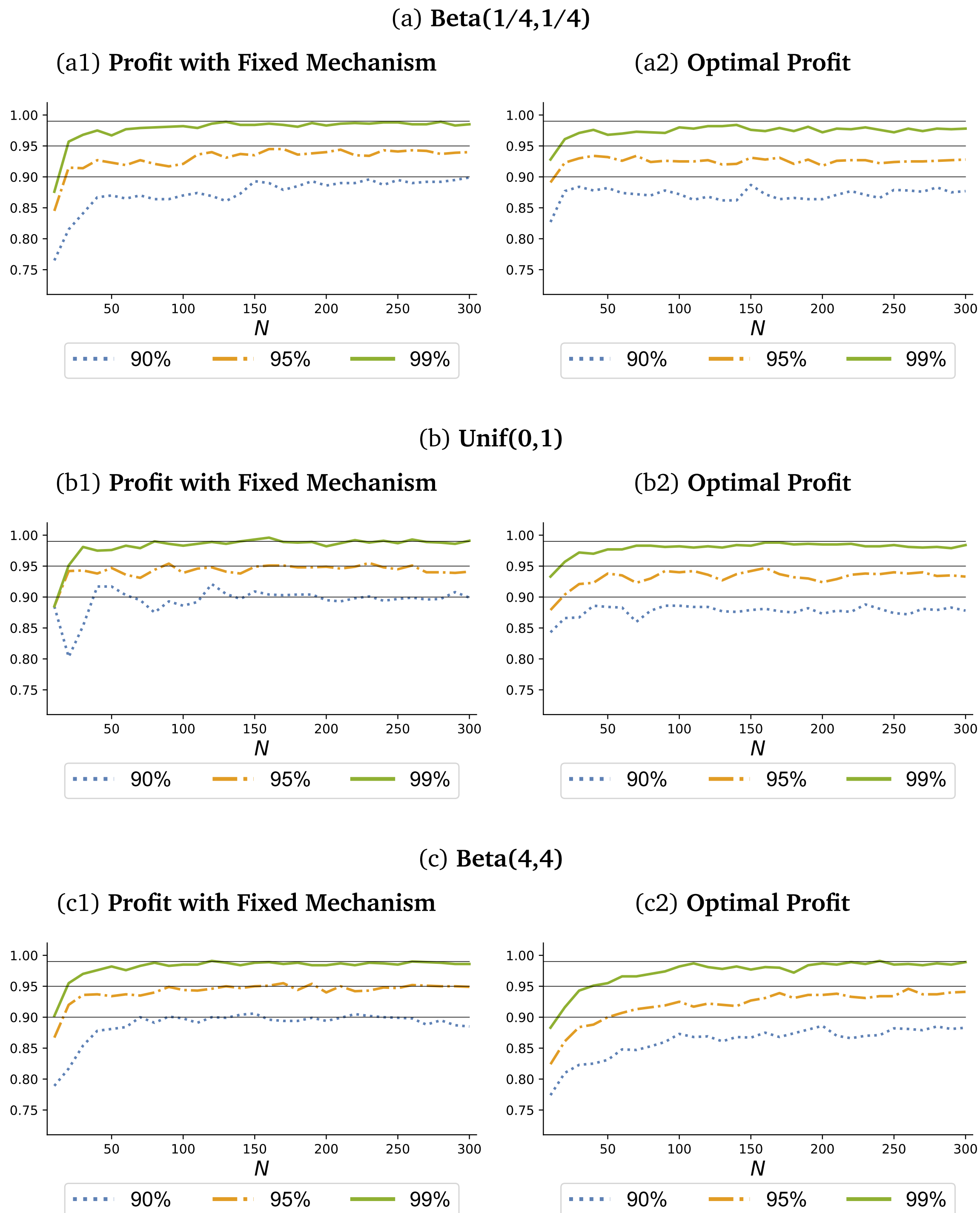

Figure 1. **Empirical Coverage Frequencies**

Note: This figure shows the frequency with which the estimated confidence interval with asymptotic coverage of $1-\alpha = .9, .95, .99$ contained the true expected profit, $\pi(M,F_0)$ in the case of a fixed mechanism and $\Pi(F_0)$ in the case of the optimal expected profit. The procedure is as described in the note to Table 1. Sample size N varies between 10 and 300 with increments of 10 observations. The fixed mechanism corresponds to uniform pricing at 1/2.

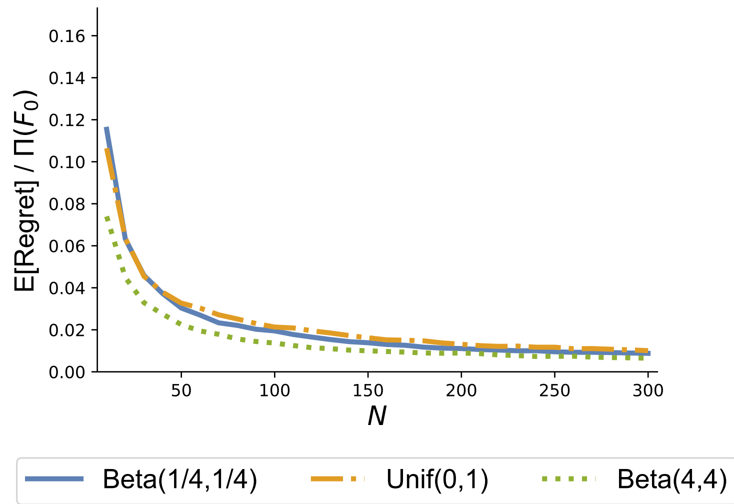

Figure 2. **Regret of the Empirically Optimal Mechanism as a share of Optimal Profit**

Note: This figure shows the average regret of the empirically optimal mechanism as a fraction of the optimal expected profit under the true distribution F_0. The average is taken over 1,000 samples for each sample size N between 10 and 300 with increments of 10 observations.

average regret no lower than 20% of the optimal expected profit under any of the three distributions we consider.[13]

5. Extension to Single-Item Auctions

Before concluding, we discuss how to apply some of the insights developed in this paper to the related setting of single-unit auctions. In particular, we show how simple empirically optimal mechanisms in this context exhibit some of the desirable robustness features shown in Section 3.

Suppose that the firm has a single item to auction to $M \geq 2$ bidders. The firm values the item at $c > 0$, and each bidder $i = 1, \ldots, M$ is risk-neutral and values the item at θ_i, drawn independently from distribution F_0. To make matters simple, we assume that F_0 is absolutely continuous with convex and compact support. In

[13]We observe that for the minimax regret distribution derived in Carrasco et al. (2018), by construction, the empirically optimal mechanism will attain the optimal profit with probability one and regardless of the number of samples, and, therefore, also attain the minimal regret.

such case, it is well-known that revenue equivalence holds and that a second-price auction with a reserve price is optimal for the firm, with bidders disclosing their types. Then, the optimal reserve price when the type distribution F satisfies the same assumptions solves

$$\max_{r \in \Theta} \pi(r, F),$$

where $\pi(r,F) = \int_r^{\bar{\theta}} dF_{(2;M)}$ and $F_{(2;M)}$ denotes the distribution of the second-highest willingness-to-pay, given a distribution of types F and M bidders. That is,

$$F_{(2;M)}(\theta) = M \cdot F(\theta)^{M-1}(1 - F(\theta)) + F(\theta)^M.$$

Consider the case where the firm has access to a sample of n observations drawn from F_0 and the reserve price is set before bids are submitted.[14] Similar to before, denote an empirically optimal reserve price $\hat{r}^*$ as the composition of a consistent estimator, $\hat{F}$, of the true distribution F_0, based on the realized sample S^n, and a selection from the set of reserve prices that are optimal for F, r^*.

The next proposition provides an analogue of Propositions 1 and 2 to this specific setting:

Proposition 6. Let $\hat{r}^*$ be an empirically optimal reserve price given by $\hat{r}^* = r^* \circ \hat{F}$. Then, $|\pi(\hat{r}^*(S^n), F_0) - \Pi(F_0)| \xrightarrow{p} 0$. Moreover, if $\hat{F} \in \widehat{\mathscr{F}}$ is such that $\forall S^n \in \mathscr{S}$, $\mathbb{P}(\|\hat{F}(S^n) - F_0\|_\infty > \delta) \leq p(n, \delta)$ for some function $p : \mathbb{N} \times \mathbb{R}_+ \to [0,1]$, then,

(i) $\mathbb{P}\left(|\pi(\hat{r}^*(S^n), \hat{F}(S^n)) - \pi(\hat{r}^*(S^n), F_0)| > \delta\right) \leq p(n, \delta/L)$; and

(ii) $\mathbb{P}(R(\hat{r}^*(S^n), F_0) > 2\delta) \leq p(n, \delta/L)$,

where $L = 2M(M-1)$ and $\Pi(F) := \sup_{r \in \Theta} \pi(r, F)$.

[14] The same arguments apply when the reserve price is secret and takes into account the bids submitted, as these would just translate into a larger sample of $n + M$ observations.

The key insight is that the expected profit is linear in distribution of the second-order statistic and that this in turn is Lipschitz continuous in the distribution of types, $\|F_{(2;M)} - G_{(2;M)}\|_\infty \le 2M(M-1)\|F - G\|_\infty$. For $\hat{r}^*$ to be empirically optimal, though, we must have that $\hat{F}(S^n)$ is absolutely continuous and has convex and compact support. Similarly to Example 2, when $\hat{F}$ is the linearly interpolated empirical distribution we have that $p(n,\delta) = 2\exp(-2n(\delta - 1/n)^2)$, delivering regret and confidence bounds.[15]

Our results on the properties on the robustness of empirically optimal mechanisms extend naturally to auction settings as this application illustrates. There is, however, a natural limitation in extending our results on inference: expected profit is linear in the distribution of the second-order statistic, not in the distribution of types themselves.

6. Concluding Remarks

This paper has studied two separate but related questions. The first is how a firm should price when uncertain about the distribution of consumers' willingness-to-pay. The second is how to conduct inference regarding the expected profit, both under any fixed pricing strategy and for the optimal profit. When the firm has access to a sample of consumers' valuations, we showed that adopting an extremely simple approach – estimating the distribution using the sample and then pricing optimally for the estimated distribution – yields attractive robustness properties, in particular obtaining probabilistic lower bounds both for the expected profit and for regret. On the other hand, we provided a toolkit to conduct inference for the expected profit. This enables practitioners to obtain confidence intervals not only for expected profit but also, for example, for the difference in profit that two different mechanisms

[15] Cole and Roughgarden (2014) provide alternative sample-complexity bounds for this problem, characterizing the asymptotic number of samples needed to achieve $(1-\epsilon)$ share of the optimal profit.

induce. More generally, this allows for a data-based approach to robust mechanism design, where robustness properties are inferred from available data.

Two important concerns that we have not discussed are how to obtain such a sample and whether a mechanism that trades-off experimentation and exploitation performs better when sampling comes at a cost. Under the assumption of myopic consumer behavior, such a sample can be procured by means of a survey. If consumers are assumed to be forward-looking, they may gain from misrepresenting their type, but incentive compatibility can be restored if the firm can preclude surveyed customers from purchasing the item in the future (or supply it to them for free in the case of unit demand). While under some conditions, optimal experimentation asymptotically attains the optimal profit (see e.g. Aghion et al. 1991), our results show that empirically optimal mechanisms do so as well, insofar as the sample acquired grows. Therefore, the firm's overall loss of revenue will depend solely on the convergence rate of its estimator.

A different issue is how to optimality elicit types (i.e., generate a sample) through revealed choices. If consumers' types are not knowable but through their choices – for instance, consumers may not be aware of their types – the firm could conduct market research and elicit such a sample by means of a mechanism that induces each type to self-select to a different pair of quantity and price. This market research can then lead to the original sample or expand an existing sample – in which case a cost-benefit analysis on the net value of acquiring additional observations may come into play. Although existence and feasibility of such mechanisms is immediate, exploring how to optimally conduct such sample elicitation endeavors may be an interesting avenue for future research.

References

Aghion, P., P. Bolton, C. Harris, and B. Jullien (1991). Optimal learning by experimentation. *Review of Economic Studies 58*(4), 621.

Babu, G. J., A. J. Canty, and Y. P. Chaubey (2002). Application of bernstein polynomials for smooth estimation of a distribution and density function. *Journal of Statistical Planning and Inference 105*(2), 377–392.

Battauz, A., M. D. Donno, and F. Ortu (2015). Envelope theorems in banach lattices and asset pricing. *Mathematics and Financial Economics 9*(4), 303–323.

Beesack, P. R. (1975). Bounds for riemann-stieltjes integrals. *The Rocky Mountain Journal of Mathematics 5*(1), 75–78.

Bergemann, D. and K. H. Schlag (2008). Pricing without priors. *Journal of the European Economic Association 6*(2-3), 560–569.

Bergemann, D. and K. H. Schlag (2011). Robust monopoly pricing. *Journal of Economic Theory 146*(6), 2527–2543.

Bickel, P. J. and D. A. Freedman (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics 9*(6), 1196–1217.

Carrasco, V., V. F. Luz, N. Kos, M. Messner, P. Monteiro, and H. Moreira (2018). Optimal selling mechanisms under moment conditions. *Journal of Economic Theory 177*, 245–279.

Carrasco, V., V. F. Luz, P. K. Monteiro, and H. Moreira (2018). Robust mechanisms: the curvature case. *Economic Theory 68*(1), 203–222.

Cole, R. and T. Roughgarden (2014). The sample complexity of revenue maximization. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing - STOC '14*. ACM Press.

Cuevas, A. and J. Romo (1997). Differentiable functionals and smoothed bootstrap. *Annals of the Institute of Statistical Mathematics 49*(2), 355–370.

Darst, R. and H. Pollard (1970). An inequality for the riemann-stieltjes integral. *Proceedings of the American Mathematical Society 25*(4), 912–912.

Datta, S. (1992). Some nonasymptotic bounds for l_1 density estimation using kernels. *The Annals of Statistics 20*(3), 1658–1667.

Dvoretzky, A., J. Kiefer, and J. Wolfowitz (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics 27*(3), 642–669.

Fu, H., N. Haghpanah, and J. Hartline (2020). Full surplus extraction from samples. *Working Paper*, 33.

Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics 18*(2), 141–153.

Guo, C., Z. Huang, and X. Zhang (2020). Sample complexity of single-parameter revenue maximization. *ACM SIGecom Exchanges 17*(2), 62–70.

Huang, Z., Y. Mansour, and T. Roughgarden (2018). Making the most of your samples. *SIAM Journal on Computing 47*(3), 651–674.

Leblanc, A. (2011). On estimating distribution functions using bernstein polynomials. *Annals of the Institute of Statistical Mathematics 64*(5), 919–943.

Madarász, K. and A. Prat (2017). Sellers with misspecified models. *Review of Economic Studies 84*(2), 790–815.

Maskin, E. and J. Riley (1984). Monopoly with incomplete information. *The RAND Journal of Economics 15*(2), 171.

Massart, P. (1990). The tight constant in the dvoretzky-kiefer-wolfowitz inequality. *The Annals of Probability 18*(3), 1269–1283.

Milgrom, P. and I. Segal (2002). Envelope theorems for arbitrary choice sets. *Econometrica 70*(2), 583–601.

Mirrlees, J. A. (1971). An Exploration in the Theory of Optimum Income Taxation. *The Review of Economic Studies 38*(2), 175–208.

Mussa, M. and S. Rosen (1978). Monopoly and product quality. *Journal of Economic Theory 18*(2), 301–317.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research 6*(1), 58–73.

Niculescu, C. P. and L.-E. Persson (2018). *Convex Functions and Their Applications*. Springer International Publishing.

Parr, W. C. (1985a). The bootstrap: Some large sample theory and connections with robustness. *Statistics & Probability Letters 3*(2), 97–100.

Parr, W. C. (1985b). Jackknifing differentiable statistical functionals. *Journal of the Royal Statistical Society: Series B (Methodological) 47*(1), 56–66.

Politis, D. N. and J. P. Romano (1994). Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics 22*(4), 2031–2050.

Segal, I. (2003). Optimal pricing mechanisms with unknown demand. *American Economic Review 93*(3), 509–529.

Shao, J. (1993). Differentiability of statistical functionals and consistency of the jackknife. *The Annals of Statistics 21*(1), 61–75.

Toikka, J. (2011). Ironing without control. *Journal of Economic Theory 146*(6), 2510–2526.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer New York.

7. Omitted Proofs

Existence of optimal mechanisms for arbitrary type distributions

Lemma 3. $\mathscr{M}^*(F) \neq \emptyset$ for all $F \in \mathscr{F}$.

Proof. Define $\bar{p} := \max_{(\theta,x)\in\Theta\times X} v(\theta,x)$ and endow the power set of $X \times [0,\bar{p}]$ with the Hausdorff metric. It is well-known that with the topology induced by the Hausdorff metric, the set of all compact subsets of $X \times [0,\bar{p}]$, denoted by $\mathscr{K}(X \times [0,\bar{p}])$, is itself compact. Define the set of mechanisms $\mathscr{M}_{IR}$ as

$$\mathscr{M}_{IR} := \left\{ M \in \mathscr{M} : \max_{(x,p)\in M} p \leq \bar{p} \right\} \subseteq \mathscr{K}(X \times [0,\bar{p}]).$$

Clearly, $\mathscr{M}_{IR}$ is closed, and therefore also compact. Since $v(\cdot,\cdot)$ is continuous and clearly so is the mapping $\Theta \times \mathscr{M}_{IR} \ni (\theta, M) \mapsto M \cup (0,0)$, we have by the Maximum Theorem that the correspondence $\mathscr{M}_{IC} : \Theta \times \mathscr{M}_{IR} \rightrightarrows X \times [0,\bar{p}]$ such that

$$(\theta, M) \mapsto \mathscr{M}_{IC}(\theta, M) = \operatorname*{argmax}_{(x,p)\in M\cup(0,0)} v(\theta,x) - p$$

is non-empty, compact valued and upper hemicontinuous. Applying the Maximum Theorem once again, we obtain that the value function $\gamma : \Theta \times \mathscr{M}_{IR} \to \mathbb{R}$ where

$$(\theta, M) \mapsto \gamma(\theta, M) = \max_{(x,p)\in\mathscr{M}_{IC}(\theta,M)} p - c(x)$$

is upper semicontinuous, i.e., for any $(\theta_0, M_0) \in \Theta \times \mathscr{M}_{IR}$, $\limsup_{(\theta,M)\to(\theta_0,M_0)} \gamma(\theta, M) \leq \gamma(\theta_0, M_0)$. Now, since γ is bounded we can apply (reverse) Fatou's Lemma to obtain that for any $M_0 \in \mathscr{M}_{IR}$,

$$\limsup_{M_n\to M_0} \int \gamma(\theta, M_n) dF(\theta) \leq \int \limsup_{M_n\to M_0} \gamma(\theta, M_n) dF(\theta) \leq \int \gamma(\theta, M_0) dF(\theta).$$

This proves that the mapping $(F,M)\mapsto\int\gamma(\cdot,M)dF$ is upper semicontinuous in M for every $F\in\mathscr{F}$. Therefore, since $\mathscr{M}_{IR}$ is compact, the extreme value theorem guarantees the existence of $M^*\in\operatorname{argmax}_{M\in\mathscr{M}_{IR}}\pi(M,F)=\operatorname{argmax}_{M\in\mathscr{M}}\pi(M,F)$ for all $F\in\mathscr{F}$, where the latter equality comes from the fact that $(x,p)\in X\times(\bar{p},\infty)$ will never be chosen over $(0,0)$. □

Proof of Lemma 1

We first state two results that will prove useful in determining the Lipschitz continuity of $\pi(M,F)$ in $F\in\mathscr{F}$.

Let $\mathscr{P}:=\{P=\theta_0,\ldots,\theta_{n_P}:\underline{\theta}=\theta_0\leq\theta_1\leq\cdots\leq\theta_{n_P}=\bar{\theta}\}$ and $V(f):=\sup_{P\in\mathscr{P}}\sum_{i=0}^{n_P-1}|f(\theta_{i+1})-f(\theta_i)|$, where $V(f)$ denotes the total variation of a function $f:\Theta\to\mathbb{R}$. We have:

Lemma 4 (Beesack-Darst-Pollard Inequality (Darst and Pollard 1970; Beesack 1975))**.** Let f,g,h be real-valued functions on a compact interval $I=[a,b]\subset\mathbb{R}$, where h is of bounded variation with total variation $V(h)$ on I and such that $\int_a^b fdg$ and $\int_a^b hdg$ both exist. Then,

$$m\int_a^b fdg+V(h)\sup_{a\leq a'\leq b'\leq b}\int_{a'}^{b'}fdg\geq\int_a^b hfdg\geq m\int_a^b fdg+V(h)\inf_{a\leq a'\leq b'\leq b}\int_{a'}^{b'}fdg,$$

where $m=\inf\{h(x):x\in I\}$.

The following is a standard result in the mechanism design literature based on Mirrlees (1971) and Milgrom and Segal (2002).

Lemma 5. The choices from a menu $M\in\mathscr{M}$ satisfy

$$(x(\theta),p(\theta))\in\arg\max_{(x,p)\in M\cup\{(0,0)\}}u(\theta,x,p)$$

if, and only if, $x(\cdot)$ is nondecreasing, $x(\underline{\theta})=0$ and $p(\theta)=v(\theta,x(\theta))-\int_{\underline{\theta}}^{\theta}v_1(s,x(s))ds$.

For any $F \in \mathscr{F}$, we can thus rewrite the problem – with some abuse of notation – by choosing directly a function x from the set $\mathscr{X}$, where

$$\mathscr{X} := \left\{x : [\underline{\theta}, \bar{\theta}] \to X, x \text{ is nondecreasing and } x(\underline{\theta}) = 0\right\},$$

in order to maximize profit, given by

$$\pi(x, F) := \int_{\Theta} \left(v(\theta, x(\theta)) - \int_{\underline{\theta}}^{\theta} v_1(s, x(s)) ds - c(x(\theta)) \right) dF(\theta).$$

Consider the normed vector space $(BV(\Theta), \|\cdot\|_{\infty})$, where $BV(\Theta) := \{g : \Theta \to \mathbb{R} \mid V(g) < \infty\}$. For any fixed $M \in \mathscr{M}$, consider its corresponding allocation function $x \in \mathscr{X}$ and extend the functional $\pi(M, \cdot)$ to $BV(\Theta)$ by defining

$$\bar{\pi}(M, H) = \int_{\Theta} \left(v(\theta, x(\theta)) - \int_{\underline{\theta}}^{\theta} v_1(s, x(s)) ds - c(x(\theta)) \right) dH(\theta), \quad \forall H \in BV(\Theta).$$

Clearly, $\bar{\pi}(M, F) = \pi(M, F)$ for all $F \in \mathscr{F}$. Moreover, note that for all $F, G \in BV(\Theta)$,

$$\begin{aligned} |\bar{\pi}(M, F) - \bar{\pi}(M, G)| &= \left| \int_{\Theta} \left(v(\theta, x(\theta)) - \int_{\underline{\theta}}^{\theta} v_1(s, x(s)) ds - c(x(\theta)) \right) d(F - G)(\theta) \right| \\ &\leq \left| \int_{\Theta} h_1 d(F - G)(\theta) \right| + \left| \int_{\Theta} h_2(\theta) d(F - G)(\theta) \right| \end{aligned}$$

where

$$\begin{aligned} h_1(\theta) &:= v(\theta, x(\theta)) \\ h_2(\theta) &:= \int_{\underline{\theta}}^{\theta} v_1(s, x(s)) ds + c(x(\theta)) \end{aligned}$$

As both v and x are nondecreasing, we have that for any $x(\cdot)$,

$$\begin{aligned} V(h_1) &= v(\bar{\theta}, x(\bar{\theta})) - v(\underline{\theta}, x(\underline{\theta})) \\ &\leq v(\bar{\theta}, \bar{x}) =: L_1 < \infty \end{aligned}$$

As v is supermodular and nondecreasing in θ and c is increasing and convex,

$$\begin{aligned} V(h_2) &= \int_{\underline{\theta}}^{\bar{\theta}} v_1(s, x(s))ds + c\left(x(\bar{\theta})\right) - c\left(x(\underline{\theta})\right) \\ &\leq (\bar{\theta} - \underline{\theta}) \cdot \max_{\theta' \in \Theta} v_1(\theta', \bar{x}) + c(\bar{x}) =: L_2 < \infty \end{aligned}$$

Moreover, we note that $\inf_{\theta \in \Theta} h_1(\theta) = v(\underline{\theta}, x(\underline{\theta})) = 0$ and $\inf_{\theta \in \Theta} h_2(\theta) = \int_{\underline{\theta}}^{\underline{\theta}} v_1(s, x(s))ds + c\left(x(\underline{\theta})\right) = c(0) = 0$. Hence, for h_i, with $m_i := \inf_{\theta \in \Theta} h_i(\theta)$, $i = 1, 2$, and letting $H(\theta) = F(\theta) - G(\theta)$,

$$\begin{aligned} \left| \int_{\underline{\theta}}^{\bar{\theta}} h_i(\theta) dF(\theta) - \int_{\underline{\theta}}^{\bar{\theta}} h_i(\theta) dG(\theta) \right| &= \left| \int_{\underline{\theta}}^{\bar{\theta}} h_i(\theta) d[F(\theta) - G(\theta)] \right| \\ &\leq \max \left\{ m_i[H(\bar{\theta}) - H(\underline{\theta})] + V(h_i) \sup_{\underline{\theta} \leq \alpha < \beta \leq \bar{\theta}} \int_{\alpha}^{\beta} dH, \right. \\ &\qquad \left. -m_i[H(\bar{\theta}) - H(\underline{\theta})] - V(h_i) \inf_{\underline{\theta} \leq \alpha < \beta \leq \bar{\theta}} \int_{\alpha}^{\beta} dH \right\} \\ &= V(h_i) \cdot \max \left\{ \sup_{\underline{\theta} \leq \alpha < \beta \leq \bar{\theta}} \int_{\alpha}^{\beta} dH, \sup_{\underline{\theta} \leq \alpha < \beta \leq \bar{\theta}} -\int_{\alpha}^{\beta} dH \right\} \\ &= V(h_i) \cdot \max \left\{ \sup_{\underline{\theta} \leq \alpha < \beta \leq \bar{\theta}} H(\beta) - H(\alpha), \sup_{\underline{\theta} \leq \alpha < \beta \leq \bar{\theta}} H(\alpha) - H(\beta) \right\} \\ &\leq V(h_i) \cdot \sup_{\underline{\theta} \leq \alpha < \beta \leq \bar{\theta}} |H(\alpha)| + |H(\beta)| \\ &= 2 \cdot V(h_i) \cdot \|F - G\|_\infty \end{aligned}$$

Combining the preceding inequalities results in

$$|\bar{\pi}(M, F) - \bar{\pi}(M, G)| \leq 2(L_1 + L_2)\|F - G\|_\infty \quad \forall F, G \in BV(\Theta),\ M \in \mathscr{M}.$$

The result is obtained by restricting the domain to $\mathscr{F}$.

Proof of Theorem 2

Let $\mathscr{Y}$ be a normed vector space with an open subset $\mathscr{A} \subseteq \mathscr{Y}$. As is standard, we denote the dual space of $\mathscr{Y}$ by $\mathscr{Y}^*$. Before we prove the main result, let us recall two different notions of differentiability:

Definition 1 (Gâteaux Differentiability)**.** A functional $T : \mathscr{Y} \to \mathbb{R}$ is Gâteaux differentiable at $F \in \mathscr{Y}$ if there exists a linear functional $\dot{T}_F \in \mathscr{Y}^*$ such that for all $H \in \mathscr{Y}$,

$$\dot{T}_F(H) = \lim_{t\to\infty} \frac{T(F+tH) - T(F)}{t}.$$

Definition 2 (Fréchet Differentiability)**.** A functional $T : \mathscr{Y} \to \mathbb{R}$ is Fréchet differentiable at $F \in \mathscr{Y}$ if there is a linear continuous functional $\dot{T}_F$ defined on $\mathscr{Y}$ such that

$$\lim_{\|H\|\to 0} \frac{|T(F+H) - T(F) - \dot{T}_F(H)|}{\|H\|} = 0.$$

Note that if T is Gâteaux differentiable at $F \in \mathscr{Y}$, then its derivative is unique. Moreover, if it is Fréchet differentiable at $F \in \mathscr{Y}$, it is Gâteaux differentiable and its derivatives agree.

An important generalization of Gâteaux differential for the case of convex and continuous functionals is that of a subdifferential.

Definition 3. The subdifferential of $T : \mathscr{Y} \to \mathbb{R}$ at $F \in \mathscr{Y}$ is the set

$$\partial T(F) := \left\{ D \in \mathscr{Y}^* : T(G) \geq T(F) + D(G-F) \text{ for each } G \in \mathscr{Y} \right\}.$$

Since the normed linear space $(BV(\Theta), \|\cdot\|_\infty)$ is a metric space, and thus $BV(\Theta)$ is open, the following lemma guarantees that ∂T is nonempty:

Lemma 6 (Theorem 3.3.1. Niculescu and Persson (2018)). If $T : \mathscr{A} \to \mathbb{R}$ is a continuous and convex functional, then $\partial T(a) \neq \emptyset$, for all $a \in \mathscr{A}$.

The next result shows why subdifferentials can be thought of as a generalization of Gâteaux differentials.

Lemma 7 (Proposition 3.6.9. Niculescu and Persson (2018)). Let $T : \mathscr{A} \to \mathbb{R}$ be a continuous and convex functional. T is Gâteaux differentiable at $a \in \mathscr{A}$ if and only if $\partial T(a)$ is a singleton.

In what follows, we borrow from the proof strategy of Theorem 2 in Battauz et al. (2015), although their conditions do not directly apply to our problem.

First fix $M \in \mathscr{M}$ and notice that $\bar{\pi}(M,\cdot) : BV(\Theta) \to \mathbb{R}$ is a linear functional. Therefore, it has a Fréchet derivative. In fact, for all $F, H \in BV(\Theta)$,

$$\lim_{\|H\|_\infty \to 0} \frac{|\bar{\pi}(M,F+H) - \bar{\pi}(M,F) - \bar{\pi}(M,H)|}{\|H\|_\infty}$$
$$= \lim_{\|H\|_\infty \to 0} \frac{|\bar{\pi}(M,F) + \bar{\pi}(M,H) - \bar{\pi}(M,F) - \bar{\pi}(M,H)|}{\|H\|_\infty} = 0.$$

This implies that the Fréchet derivative of $\bar{\pi}(M,F)$ is independent of F and given by $\dot{\pi}_M(\cdot) = \bar{\pi}(M,\cdot)$. Therefore, $\bar{\pi}(M,\cdot)$ is also Gâteaux differentiable, with Gâteaux derivative at F given by $\bar{\pi}(M,\cdot)$.

Define the functional $\bar{\Pi} : BV(\Theta) \to \mathbb{R}$ by

$$\bar{\Pi}(H) = \sup_{M \in \mathscr{M}} \bar{\pi}(M,H), \quad \forall H \in BV(\Theta).$$

Since the set $\{\bar{\pi}(M,H) : M \in \mathscr{M}\}$ is bounded for any $H \in BV(\Theta)$, it is clear that $\bar{\Pi}(H) < \infty$. Furthermore, it is immediate that $\bar{\Pi}(F) = \Pi(F)$ for all $F \in \mathscr{F}$.

As $\bar{\pi}(M,H)$ is linear in H for any M, one immediately has that $\bar{\Pi}$ is convex in $H \in BV(\Theta)$, as the supremum of a family of linear functionals. Moreover, from the proof of Lemma 2, it is easy to see that $\bar{\Pi}$ remains Lipschitz continuous. Therefore, by Lemma 6, $\partial\bar{\Pi}(H) \neq \emptyset$ for any $H \in BV(\Theta)$.

By Lemma 3, $M^*(F) \neq \emptyset$ for any $F \in \mathscr{F}$. Fix any $F \in \mathscr{F}$. Take any $D \in \partial\bar{\Pi}(F)$ and any $M_F \in \mathscr{M}^*(F)$. Note, for any $G \in BV(\Theta)$, that $\bar{\pi}(M_F, F) - \bar{\pi}(M_F, G) \geq \bar{\Pi}(F) - \bar{\Pi}(G) \geq D(F-G)$, hence $\partial\bar{\Pi}(F) \subseteq \partial\bar{\pi}(M_F, F)$ and, as $\bar{\pi}(M_F, \cdot)$ is continuous and linear, by Lemma 7, $\partial\bar{\pi}(M_F, F) = \{\dot{\pi}_{M_F}\}$. Then, for any $G \in BV(\Theta)$,

$$
\begin{aligned}
& \bar{\pi}(M_F,F) - \bar{\pi}(M_F,G) \geq \bar{\Pi}(F) - \bar{\Pi}(G) \geq \dot{\pi}_{M_F,F}(F-G) \\
\Leftrightarrow \quad & \bar{\pi}(M_F,F) - \bar{\pi}(M_F,G) - \dot{\pi}_{M_F}(F-G) \geq \bar{\Pi}(F) - \bar{\Pi}(G) - \dot{\pi}_{M_F}(F-G) \geq 0 \\
\Rightarrow \quad & |\bar{\pi}(M_F,F) - \bar{\pi}(M_F,G) - \dot{\pi}_{M_F}(F-G)| \geq |\bar{\Pi}(F) - \bar{\Pi}(G) - \dot{\pi}_{M_F}(F-G)| \geq 0.
\end{aligned}
$$

By Fréchet differentiability of $\bar{\pi}(M_F, \cdot)$, we then have that $\forall \{G_n\}_n$ such that $\|G_n - F\|_\infty \to 0$,

$$
0 \leq \frac{|\bar{\Pi}(G_n) - \bar{\Pi}(F) - \dot{\pi}_{M_F}(G_n - F)|}{\|G_n - F\|_\infty} \leq \frac{|\bar{\pi}(M_F, G_n) - \bar{\pi}(M_F, F) - \dot{\pi}_{M_F}(G_n - F)|}{\|G_n - F\|_\infty} \to 0
$$

and, consequently, $\bar{\Pi}$ is Fréchet differentiable at $F \in \mathscr{F}$. As F was arbitrary, we have that $\bar{\Pi}$ is Fréchet differentiable at any $F \in \mathscr{F}$.

Proof of Theorems 1 and 3

We will prove Theorem 3. The proof for Theorem 1 is virtually the same.

By Theorem 2, $\Pi(F)$ is Fréchet differentiable at any $F \in \mathscr{F}$ and, thus, it can be written as $\Pi(F) = \Pi(F_0) + \dot{\Pi}_{F_0}(F - F_0) + o(\|F - F_0\|_\infty)$. Furthermore, we note that $\hat{F}$ is an unbiased estimator of F_0 and then, by linearity,

$$
\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \pi(M^*(F_0), \delta_{\theta_i})\right] = \mathbb{E}\left[\pi(M^*(F_0), \hat{F}(S^n))\right] = \pi\left(M^*(F_0), \mathbb{E}\left[\hat{F}(S^n)\right]\right) = \pi(M^*(F_0), F_0),
$$

where δ_θ denotes the cumulative distribution function associated with a Dirac delta measure at θ and θ_i is the i-th observation in the sample S^n. Thus,

$$
\begin{aligned}
\sqrt{n}\left(\Pi(\hat{F}(S^n)) - \Pi(F_0)\right) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\pi(M^*(F_0),\delta_{\theta_i}) - \pi(M^*(F_0),F_0)\right) + \sqrt{n}\cdot o(\|\hat{F}(S^n)-F_0\|_\infty) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\pi(M^*(F_0),\delta_{\theta_i}) - \pi(M^*(F_0),F_0)\right) \\
&\quad + \sqrt{n}\cdot\|\hat{F}(S^n)-F_0\|_\infty\cdot\frac{o(\|\hat{F}(S^n)-F_0\|_\infty)}{\|\hat{F}(S^n)-F_0\|_\infty} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\pi(M^*(F_0),\delta_{\theta_i}) - \pi(M^*(F_0),F_0)\right) + O_p(1)\cdot o(1) \\
&\xrightarrow{d} N(0,\sigma^2_{F_0}),
\end{aligned}
$$

where $\sigma^2_{F_0} = \mathbb{E}\left[\left(\dot{\Pi}_{F_0}(\delta_\theta - F_0)\right)^2\right]$.

Finally, following Parr (1985a), we have that

$$
\begin{aligned}
\hat{G}_n &= \sqrt{n}\left(\Pi(\hat{F}(S^n_B)) - \Pi((\hat{F}(S^n)))\right) \\
&= \sqrt{n}\left(\left\{\Pi(F_0) + \dot{\Pi}_{F_0}(\hat{F}(S^n_B)-F_0) + o(\|\hat{F}(S^n_B)-F_0\|_\infty)\right\}\right. \\
&\qquad \left. - \left\{\Pi(F_0) + \dot{\Pi}_{F_0}(\hat{F}(S^n)-F_0) + o(\|\hat{F}(S^n)-F_0\|_\infty)\right\}\right) \\
&= \sqrt{n}\left(\dot{\Pi}_{F_0}(\hat{F}(S^n_B)-\hat{F}(S^n)) + o(\|\hat{F}(S^n_B)-F_0\|_\infty) + o(\|\hat{F}(S^n)-F_0\|_\infty)\right) \\
&= \sqrt{n}\left(\dot{\Pi}_{F_0}(\hat{F}(S^n_B)-\hat{F}(S^n)) + o(\|\hat{F}(S^n_B)-\hat{F}(S^n)\|_\infty) + o(\|\hat{F}(S^n)-F_0\|_\infty)\right),
\end{aligned}
$$

where we used Fréchet differentiability to obtain a first-order von Mises expansion of Π in the second equality, linearity of $\dot{\Pi}_{F_0}$ in the third and the triangle inequality in the last. As, $\hat{F}(S^n_B)$ and $\hat{F}(S^n)$ denote empirical distributions of $\hat{F}(S^n)$ and F_0, by the Dvoretzky et al.'s (1956) inequality, we have that

$$
\hat{G}_n = \sqrt{n}\dot{\Pi}_{F_0}(\hat{F}(S^n_B)-\hat{F}(S^n)) + o_p(1) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\pi(M^*(F_0),\delta_{\theta^B_i}) - \pi(M^*(F_0),\hat{F}(S^n))\right) + o_p(1)
$$

As $\mathbb{E}\left[\pi(M^*(F_0),\delta_{\theta_i^B})\,|\,S^n\right] = \pi(M^*(F_0),\hat{F}(S^n))$ and $\mathbb{E}\left[\left(\pi(M,\hat{F}(S_B^n))\right)^2\right] = \sigma_{F_0}^2 < \infty$, by the central limit theorem in Bickel and Freedman (1981), we have that $\hat{G}_n \xrightarrow{d} G_0$. Finally, as the limiting distribution is continuous, convergence is uniform.

Proof of Proposition 5

Let $\hat{F}^E$ denote the empirical distribution estimator. By Lemmata 2 and 10, $\Pi(\hat{F}(S^n)) = \Pi(\hat{F}^E(S^n)) + O_p(n^{-1})$. As such, (1) follows from the observation that $\sqrt{n}\{\Pi(\hat{F}(S^n)) - \Pi(F_0)\} = \sqrt{n}\{\Pi(\hat{F}^E(S^n)) - \Pi(F_0)\} + O_p(n^{-1/2})$ which, together with Slutsky's theorem and Theorem 3 implies $\sqrt{n}\{\Pi(\hat{F}(S^n)) - \Pi(F_0)\} \xrightarrow{d} N(0,\sigma_{F_0}^2)$.
(2) results from the analogous observation that:

$$\hat{G}_n = \sqrt{n}\left(\Pi(\hat{F}(S_B^n)) - \Pi((\hat{F}(S^n))\right) = \sqrt{n}\left(\Pi(\hat{F}^E(S_B^n)) - \Pi((\hat{F}^E(S^n))\right) + O_p(n^{-1/2}) \xrightarrow{d} N(0,\sigma_{F_0}^2).$$

For (3), we first prove the following lemmata:

Lemma 8. Consider distribution functions F, F_1, $F_2,\ldots$ such that $\|F_n(t) - F(t)\|_\infty \xrightarrow{a.s.} 0$ for all $t \in \mathbb{R}$. For all $n \in \mathbb{N}$, let μ_n be the measure on $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ induced by F_n and μ be the measure induced by F. Then $\mu_n(A) \to \mu(A)$ for all $A \in \mathscr{B}(\mathbb{R})$ almost surely.

Proof. Let $\{I_j := (a_j,b_j] \subseteq \mathbb{R} \mid j \in \mathbb{N}\}$ be a collection of disjoint intervals, and denote $I = \bigcup_{j=1}^\infty I_j$. Then, since for any $n \in \mathbb{N}$, μ_n and μ are finite measures,

$$\begin{aligned}
\left|\mu_n(I) - \mu(I)\right| &= \left|\sum_{j=1}^\infty \mu_n(I_j) - \sum_{j=1}^\infty \mu(I_j)\right| = \left|\sum_{j=1}^\infty F_n(b_j) - F_n(a_j) - \sum_{j=1}^\infty F(b_j) - F(a_j)\right| \\
&= \left|\sum_{j=1}^\infty \left(F_n(b_j) - F(b_j)\right) - \left(F_n(a_j) - F(a_j)\right)\right| \\
&\le \sum_{j=1}^\infty \left|F_n(b_j) - F(b_j)\right| + \left|F_n(a_j) - F(a_j)\right| \\
&\le \sum_{j=1}^\infty 2\sup_{t\in\mathbb{R}} |F_n(t) - F(t)|
\end{aligned}$$

Let $c_n = \sup_{t\in\mathbb{R}} |F_n(t) - F| \to 0$. There exists a monotone convergent subsequence $c_{n_k} = 2\sup_{t\in\mathbb{R}} \left|F_{n_k}(t) - F(t)\right| \to 0$ as $k \to \infty$. Therefore, by the monotone convergence theorem,

$$\lim_{k\to\infty} \left|\mu_{n_k}(I) - \mu(I)\right| \leq \lim_{k\to\infty} \sum_{j=1} 2\sup_{t\in\mathbb{R}} \left|F_{n_k}(t) - F(t)\right| = \sum_{j=1} \lim_{k\to\infty} 2\sup_{t\in\mathbb{R}} \left|F_{n_k}(t) - F(t)\right| = 0.$$

This implies there exists a convergent subsequence $\mu_{n_k}(I) \to \mu(I)$.

Now take any convergent subsequence $\mu_{n_m}(I)$ of $\mu_n(I)$, which we will denote by $\mu_m(I)$. We have

$$\left|\mu_m(I) - \mu(I)\right| \leq \sum_{j=1}^{\infty} 2\sup_{t\in\mathbb{R}} |F_m(t) - F(t)|.$$

Then $c_m = 2\sup_{t\in\mathbb{R}} |F_m(t) - F(t)| \to 0$ has a monotone convergent subsubsequence c_{m_r}. We can thus apply the monotone convergence theorem once again to conclude that

$$\left|\mu_{m_r}(I) - \mu(I)\right| \to 0 \quad \text{as } r \to \infty.$$

Since the subsequence $\mu_m(I)$ is convergent by assumption, it must converge to the limit of each of its subsequences, and we have

$$\lim_{m\to\infty} \mu_m(I) = \mu(I).$$

Therefore, every convergent subsequence of $\mu_n(I)$ converges to $\mu(I)$. Since $\left(\mu_n(I)\right)_{n\in\mathbb{N}}$ is bounded, this implies that $\lim_{n\to\infty} \mu_n(I) = \mu(I)$, and thus $\lim_{n\to\infty} \mu_n$ is a pre-measure that agrees with μ in the ring formed by disjoint unions of intervals of the type $(a,b]$, $b > a$. Therefore, since μ is σ-finite, Carathéodory's extension theorem implies that $\lim_{n\to\infty} \mu_n$ must agree with μ on $\mathscr{B}(\mathbb{R})$ almost surely. □

Lemma 9. Let $\{F_n : \Theta \to [0,1] \mid n \in \mathbb{N}\}$ be a sequence of absolutely continuous distribution functions with Radon-Nikodym derivatives given by f_n. If there exists a distribution function F such that $\|F_n - F\|_\infty \to 0$, then it is absolutely continuous with Radon-Nikodym derivative f and $\|f_n - f\|_1 \to 0$.

Proof. As $\forall n \in \mathbb{N}$, F_n is absolutely continuous and Θ has finite Lebesgue measure, then the Radon-Nikodym derivatives $\{f_n\}_{n\in\mathbb{N}}$ are uniformly integrable with respect to the Lebesgue measure. Let μ be the measure associated with F. By the Vitali-Hahn-Saks theorem and Lemma 8, we have that

$$\lim_{n\to\infty} \int_A f_n(\theta)d\theta = \mu(A)$$

for all $A \in \mathscr{B}(\Theta)$. Since $\{f_n\}_{n\in\mathbb{N}}$ is uniformly integrable, the Dunford-Pettis theorem implies that every subsequence of $\{f_n\}_{n\in\mathbb{N}}$ has a convergence subsubsequence converging to g in $L^1(\Theta)$. Denote such a subsubsequence by $\{f_{n_k}\}_{k\in\mathbb{N}}$. Then, for every $A \in \mathscr{B}(\mathbb{R})$,

$$\int_A g(\theta)d\theta = \lim_{k\to\infty} \int_A f_{n_k} d\theta = \lim_{k\to\infty} \mu_{n_k}(A) = \int_A f(\theta)d\theta.$$

This implies that $f = g$ almost surely and $\|f_n - f\|_1 \to 0$. □

Given that, $\{\hat{F}(S^n)\}_{n\in\mathbb{N}}$ is absolutely continuous with probability 1 and $\|\hat{F}(S^n) - F_0\|_\infty \xrightarrow{a.s.} 0$, the previous Lemmata imply that $\|f_n - f_0\|_1 \xrightarrow{p} 0$, which concludes the proof for (3).

Proof of Proposition 6

First, we note that $\pi(r,F)$ is Lipschitz continuous in F, with a Lipschitz constant that is independent of r. Note that

$$\begin{aligned}|\pi(r,F) - \pi(r,G)| &= \left| \int_\Theta \mathbf{1}_{\theta\geq r} d(F_{(2;M)} - G_{(2;M)}) \right| \\ &\leq \|F_{(2;M)} - G_{(2;M)}\|_\infty\end{aligned}$$

$$
\begin{aligned}
&= \sup_{\theta\in\Theta} |M\cdot(F(\theta)^{M-1} - G(\theta)^{M-1}) + (M-1)\cdot(G(\theta)^M - F(\theta)^M)| \\
&\leq M\cdot\sup_{\theta\in\Theta}|F(\theta)^{M-1} - G(\theta)^{M-1}| + (M-1)\cdot\sup_{\theta\in\Theta}|F(\theta)^M - G(\theta)^M)| \\
&\leq 2M(M-1)\|F-G\|_\infty.
\end{aligned}
$$

where the first inequality uses the Beesack-Darst-Pollard inequality – see Lemma 4. By the same arguments as in Lemma 2, we have that $\Pi(F) := \sup_{r\in\Theta}\pi(r,F)$ is also Lipschitz continuous in F and, by those made in Propositions 1 and 2, the result follows.

Other Proofs

Let the linearly interpolated empirical cumulative distribution be given by

$$
\hat{F}(S^n)(\theta) = \sum_{k=0}^{n-1} \mathbf{1}_{\{\theta_{(k)}\leq\theta<\theta_{(k+1)}\}} \frac{1}{n}\frac{\theta-\theta_{(k)}}{\theta_{(k+1)}-\theta_{(k)}} + \mathbf{1}_{\{\theta_{(n)}\leq\theta\}},
$$

where $\theta_{(k)}$ denotes the k-th smallest observation in the sample S^n and $\theta_{(0)} = \underline{\theta}$. The following holds:

Lemma 10. For any absolutely continuous $F_0 \in \mathscr{F}$, (1) $\|\hat{F}(S^n) - F_0\|_\infty \xrightarrow{a.s.} 0$ and (2) with probability 1, $\hat{F}(S^n) \in \mathscr{F}$, $\hat{F}(S^n)$ has convex support and is absolutely continuous.

Proof. Note that, with probability 1, any sampled value $\theta_i > \underline{\theta}$ as F_0 is absolutely continuous and so $\hat{F}(S^n)$ is well defined. By construction, $\hat{F}(S^n)$ has convex support and is absolutely continuous. As the probability that any two sampled observations have the same value is null, we have that, with probability 1, $\|\hat{F}(S^n) - \hat{F}^E(S^n)\|_\infty = 1/n$ where $\hat{F}^E$ denotes the empirical cumulative distribution and, with probability 1, $\|\hat{F}(S^n) - F_0\|_\infty = \|\hat{F}^E(S^n) - F_0\|_\infty + 1/n$. Consequently, as $\|\hat{F}^E(S^n) - F_0\|_\infty \xrightarrow{a.s.} 0$ and as F_0 is absolutely continuous, then the result follows. □