

Distribution-Free Inference on Willingness-to-Pay in Binary Choice Experiments

Ian Pitman* Alexander Torgovitsky†

October 30, 2026

Abstract

Choice experiments are widely used in a large number of fields, including a growing subfield of labor economics that measures the valuation of non-pecuniary workplace amenities. In contrast to many other types of randomized experiments, the evaluation of a choice experiment typically relies on parametric assumptions about the distribution of willingness-to-pay (WTP) for the choices, most commonly some flavor of logit or probit model. We develop distribution-free methods for inference in binary (dichotomous) choice experiments in which subjects choose from two options. Our methods are based on sharp characterizations of the identified set that are valid regardless of the variation in the data. We use our methods to re-evaluate the influential study by [Mas and Pallais \(2017\)](#) on the valuation for workplace amenities. Most of their conclusions survive without the logistic distributional assumption. The main exceptions are the valuation of overtime work, for which the experiment offered only two prices, as well as some gender differences, which our distribution-free methods suggest may be artifacts of the logit functional form.

*Kenneth C. Griffin Department of Economics, University of Chicago.

†Kenneth C. Griffin Department of Economics, University of Chicago.

1 Introduction

Experimental methods are prized in the social sciences in large part because they can be evaluated nonparametrically, without auxiliary functional form assumptions (LaLonde, 1986; Burtless, 1995; Athey and Imbens, 2017). A choice experiment is a popular experimental approach in which individuals are asked to make a choice from a set of options with randomly assigned prices or other characteristics. In contrast to most other experiments, choice experiments are usually evaluated under parametric distributional assumptions that translate choice frequencies into preferences (valuations). In this paper, we develop distribution-free methods for evaluating choice experiments in which the choice decision is binary (dichotomous).

Our work is motivated by a recent resurgence of interest in choice experiments in labor economics (for example, Mas and Pallais, 2017; Wiswall and Zafar, 2018; Mahmud et al., 2021; Maestas et al., 2023; Drake et al., 2025). These authors have argued that designing natural field experiments in which individuals make (or believe they make) real labor supply choices helps address well-known critiques of stated valuation approaches (McFadden, 1994; Hausman, 2012). In addition to labor economics, choice experiments have also long been common in environmental economics (e.g. Hoyos et al., 2009) and health economics (de Bekker-Grob et al., 2012). They are used to study a dizzying array of topics beyond economics, from dentistry (e.g. Boyers et al., 2021), to organic food (e.g. Güney and Giraldo, 2019), to the location of heroin injection sites (Berrigan and Zucchelli, 2022), to name just a few.

The evaluation of binary choice experiments was originally developed by Bishop and Heberlein (1979), Hanemann (1984), and Cameron (1988). Focusing on binary decisions is common due to its simplicity and perceived incentive compatibility (Carson et al., 2014; Johnston et al., 2017, Recommendation 8). The typical approach is to estimate a binary outcome logit model in which one of the features of a choice is its randomly varied price. The latent logistically distributed random variable is interpreted as willingness-to-pay (WTP). The model and estimated coefficients are then used to estimate the mean, standard deviation, and quantiles of WTP based on the assumed logistic distribution. Practitioners routinely express concern over the role of these distributional assumptions and recommend estimating distribution-free models (see Recommendation 16 and pp. 360–361 in Johnston et al., 2017; see also Freedman, 2008). A major obstacle to doing so is the fact that features of the WTP distribution are generally only partially identified without parametric distributional assumptions.

It is known that the distribution function of WTP is bounded between two step functions determined by the choice shares at the observed prices (Boman et al., 1999;

Manski and Tamer, 2002; Bhattacharya, 2021). The implied bounds on the mean were derived by Green et al. (1998) and Boman et al. (1999), and shown to be sharp by Magnac and Maurin (2008). We show that every distribution function between the two step functions is consistent with the data, which implies that the bounds are *uniformly* sharp in the sense of Firpo and Ridder (2019). We use this result to derive sharp bounds on the quantiles using a set-valued definition of the quantile that is invariant to relabeling the choice. We also derive sharp bounds on the variance, drawing on Stoye (2010) and the literature on imprecise probabilities (Langewisch and Choobineh, 2004; Kreinovich et al., 2006).

We develop approaches to estimation and inference under two different frameworks.

First, we consider estimation and inference approaches in which the *choice shares* are parameterized, but the distribution of WTP is left unspecified. The motivation for this approach is a separation between the choice shares, which are nonparametrically point identified, and features of the WTP distribution, which are not. Using a parametric model, such as a logit, then operates transparently as a means of dimension reduction, while playing no additional role in isolating the distribution of WTP. The inference problem is still a nonstandard one due to partial identification, but is simplified by the fact that typical parametric models make it credible to assume that the choice shares are “regular,” in that they are strictly monotone and bounded away from zero and one.

Then, we consider fully nonparametric approaches that also estimate choice shares nonparametrically. These estimators can lead to estimated choice shares that are not increasing (even if the model is correctly specified), straining the regularity assumption for the choice shares, and leading to sample-analog bounds that cross or are undefined. We develop confidence sets that accommodate in-sample violations of the shape restrictions and remain valid as long as the model is correctly specified in the population.

We use our methods to re-evaluate the influential choice experiment used by Mas and Pallais (2017) to estimate the value for alternative work arrangements. We are able to confirm most of their qualitative conclusions while removing their logistic distributional assumption. One exception is with overtime work, which only had two randomized prices in the experiment. We find that the authors’ estimates on the value of overtime work are heavily driven by extrapolation manufactured by the logistic distribution; without it, we find that the data is consistent with mean WTP to avoid overtime that is anywhere from moderately negative to strongly positive. The other exception is gender differences in the WTP to work from home or to avoid employer discretion, both of which have little evidence without the logit distribution.

Our paper is related to a methodological literature on distribution-free methods for

discrete choice analysis. Most relevant to our setting are methods that focus on the distribution of latent valuations for binary choice experiments, many of which have been developed in the contingent valuation literature.¹ [Kriström \(1990\)](#) proposes linearly interpolating the distribution function of WTP between observed prices, which effectively assumes that WTP is uniformly distributed between them; see also [Duffield and Patterson \(1991\)](#) and [Hanemann and Kanninen \(2001\)](#). [Cosslett \(1983\)](#) and [Watanabe \(2010\)](#) show that mean WTP is nonparametrically point identified when prices are drawn from a continuous distribution whose support covers that of WTP, and [Lewbel et al. \(2011\)](#) estimate moments and quantiles under a related asymptotic large-support assumption on prices. All of these papers make sufficient assumptions to ensure point identification.

We focus on partial identification, so our work is more closely related to the previously mentioned papers by [Green et al. \(1998\)](#), [Boman et al. \(1999\)](#), [Magnac and Maurin \(2008\)](#), and [Bhattacharya \(2021\)](#). We develop inference methods for bounds that have analytical characterizations, in contrast to a literature that develops computational approaches, such as [Manski \(2008\)](#), [Chesher and Rosen \(2013\)](#), [Tebaldi et al. \(2023\)](#), and [Kamat and Norris \(2026\)](#). The benefit of our narrower identification focus is that the statistical inference problem is more tractable.

Recent work by [Thakral and Tô \(2026\)](#) also derives distribution-free bounds in binary choice experiments. Their focus is on extending the standard model to accommodate choice errors, whereas our focus is on statistical inference in the standard model without choice errors. The authors do not consider inference or sampling uncertainty for any of their bounds. Without choice errors, the bounds they derive for the mean coincide with those of [Green et al. \(1998\)](#). Their bounds on quantiles coincide with ours except when a choice share equals the quantile level, a difference that arises because they use the conventional single-valued definition of quantiles, which is not invariant to relabeling. They do not show that their bounds are sharp. They apply their results to [Mas and Pallais \(2019\)](#), which is a different (but related) application from [Mas and Pallais \(2017\)](#), the application that we consider.

The organization of the paper is as follows. In [Section 2](#), we review standard evaluation approaches for binary choice experiments. [Section 3](#) derives bounds on features of the WTP distribution. [Section 4](#) discusses different estimators of the choice shares and introduces the idea of “regular” choice shares. [Section 5](#) develops inference procedures for regular choice shares, while [Section 6](#) extends these procedures to non-regular choice shares. [Section 7](#) uses our results to reanalyze [Mas and Pallais \(2017\)](#). [Section](#)

¹This contrasts with distribution-free methods that focus on identification of index coefficients, such as [Manski \(1975, 1985\)](#), [Horowitz \(1992\)](#), or [Komarova \(2013\)](#).

8 contains some brief concluding remarks focused on recommendations for empirical researchers evaluating binary choice experiments or contingent valuation studies.

2 Binary Choice Experiments

In this section, we define the primitive objects of a binary choice experiment and review parametric identification of the willingness-to-pay (WTP) distribution. The approach is originally due to [Hanemann \(1984\)](#) and [Cameron \(1988\)](#).

2.1 Choices

Individual i is asked to make a binary decision that depends on a price, P_i , that has support $\mathcal{P}$. Let $Y_i(p)$ on $\{0, 1\}$ denote the choice they would make if presented with price p . Their actual choice is $Y_i \equiv Y_i(P_i)$. In the application of [Mas and Pallais \(2017\)](#), the binary decision is whether to accept a job $Y_i = 1$ with a particular amenity when the wage is P_i . We reanalyze this application in Section 7.

The setting is experimental and prices are assumed to be randomly assigned.

Assumption RA. (Random assignment) P_i is independent of $\{Y_i(p) : p \in \mathcal{P}\}$.

Random assignment implies that the causal effect of P_i on the binary decision Y_i is point identified as a difference in means:

$$\mathbb{P}[Y_i(p') = 1] - \mathbb{P}[Y_i(p) = 1] = \mathbb{P}[Y_i = 1 \mid P_i = p'] - \mathbb{P}[Y_i = 1 \mid P_i = p].$$

This is an extensive-margin measure of how much price affects choosing the amenity. The object of a choice experiment is typically to measure the intensity of preference for an amenity, as reflected through the distribution of WTP. A simple potential outcomes model is not sufficient for this purpose because it does not describe how choices are made.

2.2 Threshold-crossing model of valuations

Let V_i denote individual i 's willingness-to-pay (or valuation) for Y_i . Let $B_i \in \{0, 1\}$ denote i 's *tie-breaker*: the choice they would make in the event of indifference. We assume that potential choices are governed by a threshold-crossing model.

Assumption TC. (Threshold crossing) For a pair of random variables (V_i, B_i) on $\overline{\mathbb{R}} \times \{0, 1\}$, potential choices are given by

$$Y_i(p) = \mathbb{1}[V_i > p] + \mathbb{1}[V_i = p]B_i. \tag{1}$$

As a model of preferences, (1) posits that all individuals are weakly more likely to choose the amenity when faced with a lower price. This seems like an unobjectionable assumption when P_i truly reflects a price, but could be more controversial if the randomly assigned incentive P_i represents some other adjustment to the choice environment.² In Appendix A, we modify model (1) to account for the possibility that an unknown proportion of individuals are inattentive to P_i , a common concern in choice experiments (see, for example, Johnston et al., 2017, Section 5.4).

Denote the WTP distribution by $F(p) \equiv \mathbb{P}[V_i \leq p]$. Together with Assumption RA, Assumption TC implies that

$$\begin{aligned} \mathbb{P}[Y_i = 0 \mid P_i = p] &= \mathbb{P}[Y_i(p) = 0] \\ &= \mathbb{P}[\mathbb{1}[V_i > p] + \mathbb{1}[V_i = p]B_i = 0] \\ &= F(p) - \mathbb{P}[V_i = p, B_i = 1]. \end{aligned} \tag{2}$$

If F is continuous, so that the second term has probability zero, then $F(p)$ is nonparametrically point identified at all prices $p \in \mathcal{P}$. In order to point identify a summary measure of F (such as its mean), the researcher must decide how to extrapolate from prices $p \in \mathcal{P}$ to all other prices.

The extrapolation problem is a consequence of limited price support. If prices are continuously distributed with a support that contains the support of V_i , then (2) point identifies F everywhere, so that any feature of F is also point identified. Cosslett (1983), Watanabe and Asano (2009), and Watanabe (2010) develop estimators premised on continuous variation. Our motivation is a choice experiment in which the researcher is constrained to offer a small, finite grid of prices, so that continuous variation is unavailable.

2.3 Parametric identification

The standard way to extrapolate with limited price variation is by imposing a parametric assumption. In this section, we consider the case in which V_i is assumed to be logistically distributed, as in Mas and Pallais (2017), following Hanemann (1984), Cameron and James (1987), and Cameron (1988). Similar arguments apply to other parameterizations.

Let U_i be a standard logistic random variable with location parameter zero and

²Bhattacharya (2015) shows that this structure follows from utility maximization whenever utility is continuous and strictly increasing in the numeraire, without any restrictions on preference heterogeneity or income effects, so that V_i can be interpreted as an equivalent variation.

scale parameter one. Suppose valuations can be expressed as

$$V_i = \frac{U_i - \alpha}{\beta}, \quad (3)$$

for some $\alpha \in \mathbb{R}$ and $\beta > 0$. Then V_i is a logistic random variable with location parameter $-\alpha/\beta$ and scale parameter $1/\beta$. In particular, by (2) we have that for $p \in \mathcal{P}$,

$$\mathbb{P}[Y_i = 0 \mid P_i = p] = F(p) = \Lambda(\alpha + \beta p),$$

where $\Lambda(u) \equiv 1/(1 + \exp(-u))$ is the standard logistic function. In this example, α and β are point identified so long as $\mathcal{P}$ contains two distinct prices.

The parametric assumption in (3) ensures that point identification of α and β suffices to point identify the entire WTP distribution. Many features of F can be expressed as smooth functions of α and β , including mean WTP,

$$\mathbb{E}[V_i] = -\frac{\alpha}{\beta},$$

the standard deviation of WTP,

$$S[V_i] = \frac{\pi}{\beta\sqrt{3}} \approx \frac{1}{.55\beta},$$

and the r th quantile of WTP for $r \in (0, 1)$,

$$F^{-1}(r) = \frac{\Lambda^{-1}(r) - \alpha}{\beta} = \frac{1}{\beta} \left(\log \left(\frac{r}{1-r} \right) - \alpha \right).$$

Estimation and inference through maximum likelihood and the delta method is straightforward. However, the validity of this approach depends on the assumed logistic distribution.

3 Nonparametric Bounds on Willingness-to-Pay

In this section, we derive uniformly sharp bounds on the distribution of WTP without restricting that distribution to be in any specific parametric family. We use these bounds to derive sharp bounds on the mean, quantiles, and variance of the distribution of WTP. Reporting all three measures of WTP has been argued to be important for contingent valuation studies (Johnston et al., 2017, Recommendation 18).

3.1 Notation

We assume that the support of P_i is discrete and finite with $J \geq 1$ distinct prices

$$\mathcal{P} = \{p_1, p_2, \dots, p_J\}$$

labeled in increasing order, so that $-\infty < p_1 < \dots < p_J < +\infty$. For $j \in \{1, \dots, J\}$, denote the share of individuals who would not choose the amenity at price p_j by

$$s_j \equiv \mathbb{P}[Y_i(p_j) = 0] = \mathbb{P}[Y_i = 0 \mid P_i = p_j], \quad (4)$$

where the last equality follows from Assumption [RA](#). Our analysis employs two extremal “choke prices” p_0 and p_{J+1} under which all individuals would either choose the amenity or not choose the amenity, respectively. We state this as the following assumption for reference, although the condition is sufficiently flexible that it is not necessarily an assumption.

Assumption CP. (Choke prices) There are known choke prices $p_0 < p_1$ and $p_{J+1} > p_J$ such that

$$\begin{aligned} s_0 &\equiv \mathbb{P}[Y_i(p_0) = 0] = 0, \\ \text{and } s_{J+1} &\equiv \mathbb{P}[Y_i(p_{J+1}) = 0] = 1. \end{aligned}$$

These choke prices could be $p_0 = -\infty$ and $p_{J+1} = +\infty$.

Both choke prices are assumed to be known to the researcher. All of our analysis allows the researcher to set one or both of the choke prices to be infinite, in which case Assumption [CP](#) is not a substantive assumption. The restrictiveness of Assumption [CP](#) is therefore entirely at the discretion of the researcher.

Let $\mathcal{F}$ denote the set of all univariate distribution functions on the extended real line, $\overline{\mathbb{R}} \equiv \mathbb{R} \cup \{-\infty, +\infty\}$. A function $G : \overline{\mathbb{R}} \rightarrow [0, 1]$ belongs to $\mathcal{F}$ if and only if it is weakly increasing, continuous from the right, and satisfies $G(+\infty) = 1$.³ The identified set for F , denoted by $\mathcal{F}^*$, is the subset of $\mathcal{F}$ consistent with Assumptions [RA](#), [TC](#), [CP](#), and the observed conditional distribution of Y_i given P_i .

³Because $G \in \mathcal{F}$ is a distribution function on the *extended* real line, we allow it to have point masses at $\pm\infty$ to represent the case in which there are individuals who would not choose (or would always choose) the amenity at any finite price. This means that it is possible that $\lim_{v \rightarrow -\infty} G(v) > 0$ and $\lim_{v \rightarrow +\infty} G(v) < 1$, unlike a distribution defined only on the real line.

3.2 The distribution function

The following proposition provides a sharp characterization of $\mathcal{F}^*$, which is illustrated in Figure 1.

Proposition 1. Suppose that Assumptions [RA](#), [TC](#), and [CP](#) are satisfied. Let

$$F_L(p) \equiv \mathbb{1}[p_{J+1} \leq p] + \sum_{j=0}^J \mathbb{1}[p_j \leq p < p_{j+1}]s_j$$

and $F_U(p) \equiv \mathbb{1}[p_{J+1} \leq p] + \sum_{j=0}^J \mathbb{1}[p_j \leq p < p_{j+1}]s_{j+1}.$

Then the identified set for F is given by

$$\mathcal{F}^* = \{G \in \mathcal{F} : F_L(p) \leq G(p) \leq F_U(p) \text{ for all } p \in \overline{\mathbb{R}}\}.$$

The bounds $F_L, F_U \in \mathcal{F}^*$ are themselves distribution functions in the identified set, so they are uniformly sharp in the sense of [Firpo and Ridder \(2019, Definition 2\)](#).⁴

Proposition 1 forms the basis of the subsequent analysis. The tie-breaker B_i ensures that both F_L and F_U are members of $\mathcal{F}^*$. If we were to assume that $B_i = 0$ for all i , then (2) would imply $F(p_j) = s_j$, whereas $F_U(p_j) = s_{j+1}$. The identified set would contain distributions arbitrarily close to F_U , but not F_U itself. Assuming $B_i = 1$ for all i would exclude F_L in the same way. The bounds on the features of F that we consider below would be unchanged, but some would no longer be attained; for example, the identified set for the mean would be half-open.

3.3 The mean

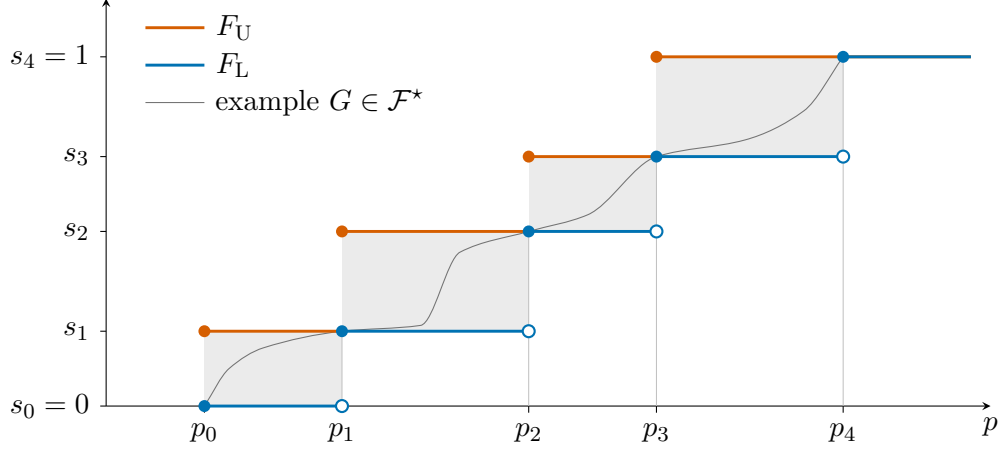
For any $G \in \mathcal{F}$, let $m(G)$ denote the (possibly infinite) mean of G , if it exists. The identified set for $m(F)$ is defined as

$$\mathcal{M}^* \equiv \{m(G) : G \in \mathcal{F}^* \text{ such that } m(G) \text{ exists}\}.$$

Since m is increasing in the first-order stochastic dominance order, Proposition 1 immediately provides sharp bounds on mean WTP as the means of F_U and F_L , respectively ([Manski, 2003](#)).

⁴With finite choke prices, F_L and F_U place no mass at $\pm\infty$, so they are uniformly sharp even among distributions on $\mathbb{R}$. With an infinite choke price, they can place mass at $\pm\infty$; for example, if $p_0 = -\infty$, then $\lim_{p \downarrow -\infty} F_U(p) = s_1$. Among distributions on $\mathbb{R}$, they are then pointwise but not uniformly sharp, which shows that the limit conditions in [Firpo and Ridder \(2019, Theorem 2\)](#) need not hold.

Figure 1: The bounding distributions in Proposition 1



Notes: Illustration with $J = 3$ observed prices and finite choke prices. The shaded boxes show the region $F_L(p) \leq G(p) \leq F_U(p)$ that contains every $G \in \mathcal{F}^*$.

Proposition 2. Suppose that Assumptions [RA](#), [TC](#), and [CP](#) are satisfied. Then

$$\mathcal{M}^* = [m(F_U), m(F_L)] = \left[\sum_{j=0}^J (s_{j+1} - s_j) p_j, \sum_{j=0}^J (s_{j+1} - s_j) p_{j+1} \right],$$

where the above expressions use the convention that $0 \times \pm\infty \equiv 0$. The length of the identified set is

$$m(F_L) - m(F_U) = \sum_{j=0}^J (s_{j+1} - s_j) (p_{j+1} - p_j). \quad (5)$$

All components of the lower bound are determined by the observed data, except potentially for p_0 . If $s_1 = 0$, so that the lowest observed price p_1 is itself a choke price, then p_0 drops out of the expression. Otherwise, the lower bound of $\mathcal{M}^*$ will be an affine function of p_0 , with slope given by the share s_1 of individuals who do not choose the amenity at the lowest observed price. Setting $p_0 = 0$, which is reasonable if the amenity is an uncontroversial “good,” produces the [Turnbull \(1976\)](#) lower bound, which is commonly reported in the contingent valuation literature ([Haab and McConnell, 1997, 2002](#)). If the researcher is unwilling to specify a lower choke price, so that $p_0 = -\infty$, then the lower bound is completely uninformative whenever $s_1 > 0$. Symmetric statements apply to the upper bound with respect to the choke price p_{J+1} .⁵

⁵If both choke prices are finite then the midpoint of $\mathcal{M}^*$ is the mean of the distribution function that interpolates linearly between the choice shares at adjacent prices. This is the basis of the estimator proposed

An important implication of Proposition 2 is that in the absence of an assumed or observed choke price, a binary choice experiment contains no information about the mean WTP (Haab and McConnell, 1997, p. 259). Estimates of the mean are determined entirely by the tail behavior of an assumed parametric functional form, in the sense that any hypothetical value of mean WTP can always be justified by a distribution with appropriate tail behavior (Manski, 1988).⁶ Such a distribution might appear bizarre, placing small amounts of mass far out in the tails. If the researcher thinks that this possibility is unreasonable, then they should be willing to impose a priori choke prices that preclude it.

3.4 The quantiles

For any $r \in [0, 1]$ and $G \in \mathcal{F}$, let $q(r, G)$ denote the r th quantile of G . We view the standard definition of a quantile as problematic for the current purposes because its identified set is not invariant (equivariant) to whether the amenity is labeled as $Y_i = 1$ or $Y_i = 0$. We explain this point in more detail in Appendix B. Instead, we employ a set-valued definition of quantiles on the extended real line (Ehm et al., 2016). This makes our arguments differ from prior quantile bounding arguments for distribution-free binary choice experiments (Haab and McConnell, 2002; Crooker and Herriges, 2004; Thakral and Tô, 2026).⁷

The set-valued definition takes $q(r, G) \equiv [\underline{q}(r, G), \bar{q}(r, G)]$, where

$$\underline{q}(r, G) \equiv \inf\{p \in \overline{\mathbb{R}} : G(p) \geq r\}$$

and $\bar{q}(r, G) \equiv \sup\{p \in \overline{\mathbb{R}} : G(p) \leq r \text{ or } p = -\infty\}.$

The identified set for $q(r, F)$ is given by

$$\mathcal{Q}^*(r) \equiv \bigcup_{G \in \mathcal{F}^*} q(r, G).$$

First-order stochastic dominance arguments imply that $\mathcal{Q}^*(r)$ only depends on the r th quantiles of the bounding distributions F_U and F_L from Proposition 1.

by Kriström (1990); see also Duffield and Patterson (1991) and Hanemann and Kanninen (2001). Proposition 2 shows that this estimator selects one point from the identified set based on the implicit distributional assumption that WTP is uniformly distributed between adjacent prices.

⁶In the contingent valuation literature this is called the “fat tail problem” (for example, Ready and Hu, 1995).

⁷A consequence of this is that the bounds derived by Thakral and Tô (2026, Theorem 2) are narrower than those ahead in Proposition 3, because they are bounds for $\underline{q}(r, F)$, rather than for the quantile set $q(r, F) = [\underline{q}(r, F), \bar{q}(r, F)]$ defined below. A smaller difference is that their quantile bounds do not allow for finite choke prices.

Proposition 3. Suppose that Assumptions [RA](#), [TC](#), and [CP](#) are satisfied. Then for $r \in [0, 1]$,

$$\begin{aligned}\mathcal{Q}^*(r) &= [\underline{q}(r, F_U), \bar{q}(r, F_L)] \\ &= \left[\mathbb{1}[r = 0](-\infty) + \sum_{j=0}^J \mathbb{1}[s_j < r \leq s_{j+1}]p_j, \right. \\ &\quad \left. \mathbb{1}[r = 1](+\infty) + \sum_{j=0}^J \mathbb{1}[s_j \leq r < s_{j+1}]p_{j+1} \right],\end{aligned}$$

where the above expressions use the convention that $0 \times \pm\infty \equiv 0$.

Like the mean bounds, the lower quantile bound depends on the lower choke price p_0 , but not on the upper choke price p_{J+1} , with the opposite being true for the upper quantile bound. However, unlike the mean bounds, quantile bounds will still be informative for some quantiles r even if choke prices are infinite. In particular, the lower bound will be informative for $r > s_1$, while the upper bound will be informative for $r < s_J$. For ease of exposition, suppose that $s_j \neq r$ for all $j \in \{0, \dots, J+1\}$. Then the length of $\mathcal{Q}^*(r)$ is given by $p_{k+1} - p_k$, where $k \in \{0, \dots, J\}$ is the unique index satisfying $s_k < r < s_{k+1}$. Whenever $k \notin \{0, J\}$, the length of the identified set for the r th quantile will not depend on choke prices at all.

3.5 The variance

For any $G \in \mathcal{F}$, let $v(G)$ denote the (possibly infinite) variance of G , if it exists. The identified set for $v(F)$ is defined as

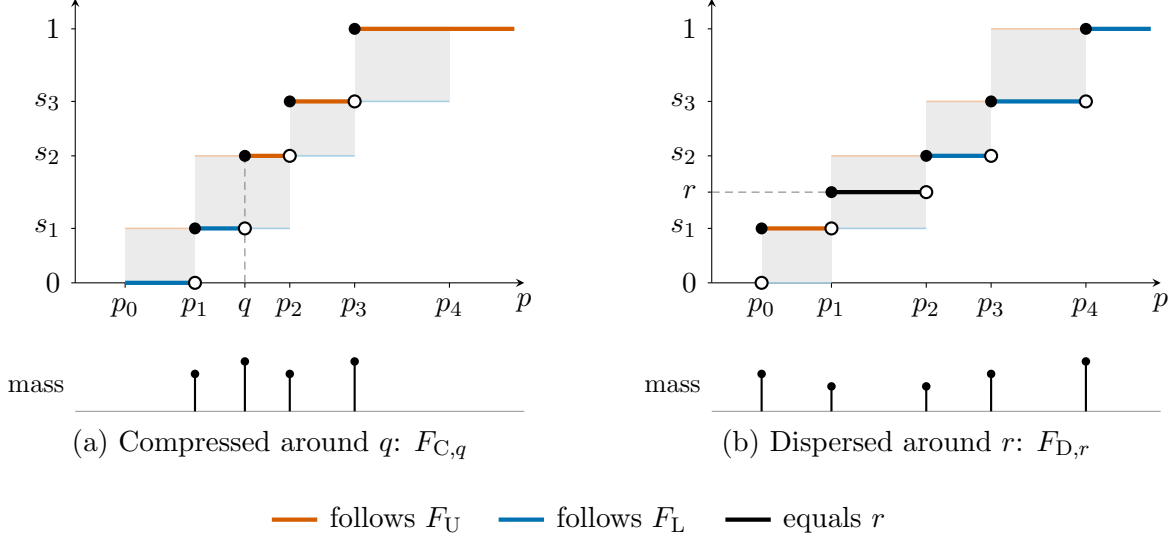
$$\mathcal{V}^* \equiv \{v(G) : G \in \mathcal{F}^* \text{ such that } v(G) \text{ exists}\}.$$

[Stoye \(2010\)](#) showed that $\mathcal{V}^*$ can be characterized in terms of “compressed” and “dispersed” distributions derived from F_U and F_L .⁸ For $q \in [p_0, p_{J+1}]$, define the compressed distribution around q as

$$F_{C,q}(p) \equiv \begin{cases} F_L(p), & p < q, \\ F_U(p), & q \leq p. \end{cases}$$

⁸Note that [Stoye \(2010\)](#) uses the opposite labeling scheme from us, so that his F_U is our F_L and his F_L is our F_U . Our scheme is motivated by pointwise comparisons, while Stoye’s is motivated by first-order stochastic dominance comparisons.

Figure 2: Compressed and dispersed distributions



Notes: Compressed and dispersed distributions for the same observed prices and choice shares as in Figure 1. Each distribution function is drawn in the color of the bound it coincides with. The faint lines are F_L and F_U . The lower plots show the probability mass that each distribution places at each point.

Similarly, define the dispersed distribution around $r \in [0, 1]$ as

$$F_{D,r}(p) \equiv \begin{cases} F_U(p), & p < \underline{q}(r, F_U), \\ r, & \underline{q}(r, F_U) \leq p < \bar{q}(r, F_L), \\ F_L(p), & \bar{q}(r, F_L) \leq p. \end{cases}$$

These constructions are motivated by the observation that each $G \in \mathcal{F}^*$ places mass $s_{j+1} - s_j$ in each price bin $[p_j, p_{j+1}]$; different $G \in \mathcal{F}^*$ only differ in how they allocate mass within these bins. The compressed distribution around q allocates it as close to q as possible, pushing mass towards the upper endpoints p_{j+1} for bins with $p_{j+1} \leq q$ and towards the lower endpoints p_j for bins with $p_j \geq q$. The dispersed distribution around r does the opposite, pushing mass towards the lower endpoint of bins with $s_{j+1} \leq r$, towards the upper endpoint of bins with $s_j \geq r$, and splitting the mass of the bin with $s_j < r < s_{j+1}$ between its two endpoints. Figure 2 illustrates.

If the choke prices p_0 and p_{J+1} are finite, so that all the means and variances of the compressed and dispersed distributions exist and are finite, then it follows from [Stoye \(2010, Theorem 2\(i\)\)](#) that

$$\mathcal{V}^* = \left[\inf_{q \in [p_0, p_{J+1}]} v(F_{C,q}), \sup_{r \in [0, 1]} v(F_{D,r}) \right].$$

That is, the variance lower bound is the smallest variance among compressed distributions, while the variance upper bound is the largest variance among dispersed distributions. In general, these optimization problems are not convex in q or r ; [Stoye \(2010, p. 335\)](#) suggests solving them through a one-dimensional grid search.

In the current case, however, the distributions in $\mathcal{F}^*$ have a known structure that has been studied in the literature on imprecise probabilities. [Kreinovich et al. \(2006, Section 4, pp. 220–224\)](#) show that this allows the variance bounds to be characterized more explicitly: the lower bound places the mass in each bin at the point of the bin closest to the mean, the upper bound places it at the endpoint farther from the mean, and the mass is split when the mean is equal to the bin's midpoint. The following proposition restates that characterization in terms of compressed and dispersed distributions, then extends it to infinite choke prices.

Proposition 4. Suppose that Assumptions [RA](#), [TC](#), and [CP](#) are satisfied. Without loss of generality, if $s_1 = 0$, let p_0 be finite, while if $s_J = 1$, let p_{J+1} be finite. Let the “midpoint” distribution function be given by

$$F_M(p) \equiv \mathbb{1} \left[\frac{p_J + p_{J+1}}{2} \leq p \right] + \sum_{j=1}^J \mathbb{1} \left[\frac{p_{j-1} + p_j}{2} \leq p < \frac{p_j + p_{j+1}}{2} \right] s_j.$$

Then

$$\mathcal{V}^* = \begin{cases} [v(F_{C,q^*}), +\infty], & \text{if } p_0 = -\infty \text{ or } p_{J+1} = +\infty, \\ [v(F_{C,q^*}), v(F_{D,r^*})], & \text{otherwise,} \end{cases}$$

where

- $q^* \in [p_1, p_J]$ is a solution to $m(F_{C,q}) = q$; a solution exists, and it is unique unless $s_k = 0$ and $s_{k+1} = 1$ for some $k \in \{1, \dots, J-1\}$; and
- $r^* \in (0, 1)$ is a solution to either $m(F_{D,r}) = \bar{q}(r, F_M)$ or $F_M(m(F_{D,r})) = r$; a solution exists and is unique.

Proposition 4 shows that the lower bound of the variance of WTP is attained by the compressed distribution F_{C,q^*} whose mean is equal to q^* . A consequence is that this lower bound does not depend on choke prices, because $q^* \in [p_1, p_J]$. On the other hand, the variance upper bound does depend on choke prices. It is infinite if $p_0 = -\infty$ and $s_1 > 0$, or if $p_{J+1} = +\infty$ and $s_J < 1$. When both choke prices are finite, the optimal dispersion parameter is either an interior solution $m(F_{D,r^*}) = \bar{q}(r^*, F_M)$, or a corner solution $F_M(m(F_{D,r^*})) = r^*$. In the case of an interior solution, the mean of the optimal dispersed distribution lies at the midpoint of some price bin $[p_j, p_{j+1}]$; this

pushes as much of the probability mass away from the mean as possible. In the case of a corner solution, the variance upper bound is attained by simply setting $r^* = s_j$ for some $j \in \{1, \dots, J\}$.

4 Choice Share Estimation

In the next two sections, we develop distribution-free methods for conducting inference on the mean, variance, and quantiles of WTP. The methods are centered around estimation of the choice shares, s_j for $j \in \{1, \dots, J\}$, which are point identified by the distribution of Y_i given $P_i = p_j$. We assume throughout that we observe an i.i.d. sample $\{Y_i, P_i\}_{i=1}^n$. Estimation of the choice shares is premised on the following assumption.

Assumption FF. (Functional form) For all $j \in \{1, \dots, J\}$,

$$s_j = L(h(p_j; \theta_0)), \quad (6)$$

where $L : \mathbb{R} \rightarrow (0, 1)$ is a known, continuously differentiable distribution function with strictly positive derivative ℓ , $\theta_0 \in \mathbb{R}^K$ is an unknown parameter, and $h : \mathbb{R} \times \mathbb{R}^K \rightarrow \mathbb{R}$ is a known index function that is continuously differentiable in its second argument. The map $\theta \mapsto (h(p_1; \theta), \dots, h(p_J; \theta))$ is injective and its Jacobian has full column rank at θ_0 .

The leading example of Assumption FF is the linear logit model used by [Mas and Pallais \(2017\)](#), in which $L = \Lambda$ is the standard logistic function from Section 2.3, $\theta = (\alpha, \beta)$, and $h(p; \theta) = \alpha + \beta p$. A probit model instead takes L to be the standard normal distribution function. At the other extreme, taking $h(p_j; \theta) = \theta_j$ with $K = J$ produces a fully nonparametric specification that makes Assumption FF without loss of generality, as long as the choice shares are strictly between zero and one. For parameterized cases, the other conditions in Assumption FF require the parameterization to be smooth and non-redundant, so that θ is identified.

Even in parameterized cases, Assumption FF is not at odds with our agenda of distribution-free inference. Whereas the standard approach in equation (3) of Section 2.3 assumes that $F(p)$ follows a parameterized logistic distribution for *all* $p \in \mathbb{R}$, Assumption FF only imposes a parametric form at the *observed* prices $p_1, \dots, p_J$. The crucial distinction is that choice shares at the observed prices are point identified, while those at unobserved prices are not. Assumption FF explicitly avoids extrapolating from the grid of observed prices to hypothetical prices that were not part of the experiment. The role of the distributional assumption in extrapolation can be isolated by comparing point estimates from the standard approach with bounds that use the same estimated

choice shares, but which avoid using the distributional assumption to extrapolate to other prices. Imposing a functional form on the observed shares also comes with the practical benefit of increasing power through dimension reduction, something that will be evident in the application in Section 7.

We consider estimation and inference with and without the assumption that the true choice shares at the observed prices—which are unknown but point identified—are strictly monotonic and interior to the unit interval.

Assumption RS. (Regular shares) $0 < s_1 < s_2 < \dots < s_J < 1$.

We say that choice shares are *regular* when Assumption RS holds. As a notational convention, we let $\mathcal{S} \equiv \{s \in \mathbb{R}^J : 0 \leq s_1 \leq \dots \leq s_J \leq 1\}$ denote the order polytope, so that choice shares are regular if and only if $(s_1, \dots, s_J)$ lies in $\text{int}(\mathcal{S})$, the interior of $\mathcal{S}$. Under Assumption FF, Assumption RS is equivalent to $\theta_0 \in \Theta$, where

$$\Theta \equiv \{\theta \in \mathbb{R}^K : h(p_j; \theta) < h(p_{j+1}; \theta) \text{ for all } j \in \{1, \dots, J-1\}\}$$

is an open set. If $h(p; \theta) = \alpha + \beta p$, then $\Theta = \mathbb{R} \times \mathbb{R}_{++}$, and Assumption RS is satisfied if $\beta_0 > 0$. As this case suggests, Assumption RS will tend to be more suitable in guiding finite-sample behavior when θ is more tightly parameterized; more flexible specifications will more frequently produce estimates on the boundary of $\mathcal{S}$, where the constraint binds and the normal approximation is unreliable, especially when there are many observed prices.

Under Assumptions FF and RS, we can estimate θ_0 by maximum likelihood. Let $\hat{\theta}_n$ denote the maximum likelihood estimator constrained to the closure of Θ . Under standard conditions—which include Assumptions FF and RS—we have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma_0).$$

The asymptotic variance, Σ_0 , can be estimated by the usual sandwich estimator $\hat{\Sigma}_n \equiv \hat{H}_n^{-1} \hat{S}_n \hat{H}_n^{-1}$, where $\hat{H}_n$ and $\hat{S}_n$ are the sample averages of the Hessian and of the outer product of the score of the log-likelihood, evaluated at $\hat{\theta}_n$.

Estimation and inference for choice shares then follows from the delta method. Collect the observed prices and population choice shares into vectors $\mathbf{p} \equiv [p_1, \dots, p_J]'$ and $\mathbf{s} \equiv [s_1, \dots, s_J]'$. Let $\hat{\mathbf{s}}_n \equiv [L(h(p_1; \hat{\theta}_n)), \dots, L(h(p_J; \hat{\theta}_n))]'$ denote the vector of estimated choice shares. Applying the delta method gives

$$\sqrt{n}(\hat{\mathbf{s}}_n - \mathbf{s}) \xrightarrow{d} \text{diag}\{\ell(h(\mathbf{p}; \theta_0))\} \nabla h(\theta_0) \mathcal{N}(0, \Sigma_0) \equiv \mathcal{N}(0, \Gamma_0),$$

where $\text{diag}\{\cdot\}$ constructs a square diagonal matrix from a vector, $\nabla h(\theta)$ denotes the $J \times K$ Jacobian of $\theta \mapsto h(\mathbf{p}; \theta)$, and Γ_0 denotes the resulting asymptotic variance matrix. The usual estimator of Γ_0 is

$$\hat{\Gamma}_n \equiv \text{diag}\{\ell(h(\mathbf{p}; \hat{\theta}_n))\} \nabla h(\hat{\theta}_n) \hat{\Sigma}_n \nabla h(\hat{\theta}_n)' \text{diag}\{\ell(h(\mathbf{p}; \hat{\theta}_n))\},$$

which is consistent under standard assumptions.

In the saturated specification with $h(p_j; \theta) = \theta_j$, unconstrained maximum likelihood estimation of θ reduces to the binning estimator of the choice shares

$$\hat{s}_{j,n} \equiv \frac{\sum_{i=1}^n \mathbb{1}[P_i = p_j](1 - Y_i)}{\sum_{i=1}^n \mathbb{1}[P_i = p_j]}.$$

The estimated choice shares are asymptotically normal by the central limit theorem, whether or not Assumption [RS](#) is satisfied, and Γ_0 is the diagonal matrix

$$\Gamma_0 = \text{diag} \left\{ \left[\frac{s_1(1-s_1)}{\mathbb{P}[P_i=p_1]} \quad \dots \quad \frac{s_J(1-s_J)}{\mathbb{P}[P_i=p_J]} \right]' \right\}.$$

The estimated choice shares from the binning estimator may be non-monotonic, so that $\hat{s}_{j,n} > \hat{s}_{k,n}$ even though $p_j < p_k$, and they could be equal to zero or one. Non-monotone binned estimates suggest that Assumption [RS](#) may be close to failing relative to sampling error, which motivates the methods in Section [6](#).⁹

5 Inference with Regular Shares

This section develops inference on features of WTP under Assumption [RS](#), using the constrained maximum likelihood estimator from Section [4](#).

5.1 The mean

Suppose that both choke prices in Assumption [CP](#) are finite, so that both the lower and upper bounds on mean WTP are not infinite. Proposition [2](#) showed that the bounds

⁹Nonparametric isotonic regression ([Ayer et al., 1955](#)) is another possible way to estimate the shares; see [Kriström \(1990\)](#) for its use in contingent valuation. The isotonic regression would always be weakly increasing, but its distribution is non-standard when some adjacent choice shares are equal or close relative to sampling error (see, for example, [Chetverikov et al., 2018](#); [Fang and Santos, 2019](#)). Isotonic regression would therefore further compound the inference problems created by partial identification that we try to solve in the following sections. We do use isotonic regression in two places: as the point around which we linearize the variance bounds in Section [6.2](#), and in our estimate of Γ_0 for the nonparametric confidence intervals, where we replace each s_j with its isotonic regression estimate.

are a known affine function of choice shares,

$$\begin{bmatrix} m(F_U) \\ m(F_L) \end{bmatrix} = A\mathbf{s} + b, \quad \text{for } A \equiv \begin{bmatrix} p_0 - p_1 & \cdots & p_{J-1} - p_J \\ p_1 - p_2 & \cdots & p_J - p_{J+1} \end{bmatrix}, \quad b \equiv \begin{bmatrix} p_J \\ p_{J+1} \end{bmatrix}.$$

An analog estimator of the bounds is $[\hat{m}_{l,n}, \hat{m}_{u,n}]' \equiv A\hat{\mathbf{s}}_n + b$. Applying the continuous mapping theorem, we conclude that

$$\sqrt{n} \left(\begin{bmatrix} \hat{m}_{l,n} \\ \hat{m}_{u,n} \end{bmatrix} - \begin{bmatrix} m(F_U) \\ m(F_L) \end{bmatrix} \right) \xrightarrow{d} \mathcal{N}(0, A\Gamma_0 A'). \quad (7)$$

The asymptotic variance can be consistently estimated with $A\hat{\Gamma}_n A'$.

With asymptotically normal endpoints, a confidence interval for mean WTP can be constructed using arguments from [Imbens and Manski \(2004\)](#) and [Stoye \(2009, 2021\)](#). As observed by [Stoye \(2009\)](#), different inference procedures may be appropriate depending on whether or not the estimated bounds may cross with positive probability. In the current setting, $\hat{\theta}_n$ is constrained to the closure of Θ , so $\hat{\mathbf{s}}_n$ is weakly increasing in price, which ensures that $\mathbb{P}[\hat{m}_{u,n} \geq \hat{m}_{l,n}] = 1$. By Lemma 3 of [Stoye \(2009\)](#), the corresponding estimator $\hat{m}_{u,n} - \hat{m}_{l,n}$ of the length of the identified set is superefficient at zero. A $1 - \alpha$ confidence interval for mean WTP is given by

$$C_n \equiv \left[\hat{m}_{l,n} - \frac{\hat{c}_{l,n} \hat{\sigma}_{l,n}}{\sqrt{n}}, \hat{m}_{u,n} + \frac{\hat{c}_{u,n} \hat{\sigma}_{u,n}}{\sqrt{n}} \right],$$

where $\hat{\sigma}_{l,n}$ and $\hat{\sigma}_{u,n}$ are the estimated asymptotic standard deviations of $\hat{m}_{l,n}$ and $\hat{m}_{u,n}$, respectively, and the critical values $\hat{c}_{l,n}$ and $\hat{c}_{u,n}$ are computed following [Stoye \(2009, Equations \(4\) and \(5\)\)](#). As [Stoye \(2009\)](#) emphasizes, the uniform validity of C_n is predicated on (7) holding uniformly, which in this context requires the delta method to be reliable.

5.2 The variance

The form of the variance WTP bounds reported in Proposition 4 depends on the implicitly defined points q^* and r^* . In Lemma 1, we show that the bounds can also be represented as the optimal values of two quadratic programs, each parameterized by the vector of choice shares, $\mathbf{s}$. We use this representation to show that the bounds are differentiable in $\mathbf{s}$.

Lemma 1. Suppose that both choke prices p_0 and p_{J+1} are finite, and let $\nu = (\nu_l, \nu_u) : \mathcal{S} \rightarrow \mathbb{R}_+^2$ map choice shares to the lower and upper bounds on the variance of WTP,

as in Proposition 4. Then

- (i) ν is continuously differentiable on $\text{int}(\mathcal{S})$;
- (ii) for every $\mathbf{s} \in \mathcal{S} \setminus \text{vert}(\mathcal{S})$, ν is Hadamard directionally differentiable at $\mathbf{s}$ tangentially to the tangent cone of $\mathcal{S}$ at $\mathbf{s}$, and its directional derivative is a linear map $\nabla\nu(\mathbf{s})$;
- (iii) $\nabla\nu$ is continuous on $\mathcal{S} \setminus \text{vert}(\mathcal{S})$.

Explicit expressions for $\nabla\nu_l$ and $\nabla\nu_u$ are given in the proof.

The proof of Lemma 1 is in Supplemental Appendix SA.1.

Lemma 1 allows us to apply the delta method, followed by the same Imbens–Manski–Stoye approach as for the mean of WTP. Given Assumption RS, we have that

$$\sqrt{n}(\nu(\hat{\mathbf{s}}_n) - \nu(\mathbf{s})) \xrightarrow{d} \mathcal{N}(0, \nabla\nu(\mathbf{s})\Gamma_0\nabla\nu(\mathbf{s})'),$$

where an expression for the Jacobian $\nabla\nu(\mathbf{s})$ is given in the proof of Lemma 1. The corresponding plug-in estimator $\nabla\nu(\hat{\mathbf{s}}_n)$ is well-defined whenever the maximum likelihood estimator exists and lies in Θ with probability approaching one under Assumption RS. By the continuous mapping theorem, the asymptotic covariance matrix is consistently estimable by $\nabla\nu(\hat{\mathbf{s}}_n)\hat{\Gamma}_n\nabla\nu(\hat{\mathbf{s}}_n)'$, since $\nabla\nu(\mathbf{s})$ is continuous on $\text{int}(\mathcal{S})$. As in Section 5.1, inference for the true parameter $v(F)$ follows from Stoye (2009), because the estimated bounds $\nu(\hat{\mathbf{s}}_n)$ cannot cross by construction.

5.3 The quantiles

While bounds on the mean and variance of WTP are given by differentiable functions of choice shares, bounds on the quantiles of WTP are not, necessitating a different approach to inference.

Consider the problem of testing the null hypothesis

$$H_{0j}(r) : [p_j, p_{j+1}] \subseteq \mathcal{Q}^*(r),$$

for any $j \in \{0, \dots, J\}$. The following corollary uses the expression for $\mathcal{Q}^*(r)$ given in Proposition 3 to reformulate this null hypothesis as two inequalities involving choice shares.

Corollary 1. Suppose that Assumptions RA, TC, and CP are satisfied. Then for $r \in [0, 1]$ and $j \in \{0, \dots, J\}$, $[p_j, p_{j+1}] \subseteq \mathcal{Q}^*(r)$ if and only if $s_j \leq r \leq s_{j+1}$.

Corollary 1 justifies writing

$$H_{0j}(r) : s_j - r \leq 0 \quad \text{and} \quad r - s_{j+1} \leq 0.$$

We can therefore construct confidence sets for WTP quantiles by inverting moment inequality tests. These confidence intervals have the form

$$C_n(r) \equiv \bigcup_{j=0}^J \{[p_j, p_{j+1}] : \text{do not reject } H_{0j}(r) \text{ at level } \alpha\}.$$

The coverage that we want to guarantee is for an arbitrary element of the set of quantiles $q(r, F)$, where F is the true distribution of WTP.¹⁰ In particular, this guarantees coverage of the canonical point-valued definition of the r th quantile $\underline{q}(r, F) \in q(r, F)$. To see how this works with a confidence set like $C_n(r)$ constructed from test inversion, let $\tilde{q} \in \mathcal{Q}^*(r)$. Then there exists a $k \in \{0, \dots, J\}$ such that $\tilde{q} \in [p_k, p_{k+1}] \subseteq \mathcal{Q}^*(r)$, so that the null hypothesis $H_{0k}(r)$ holds. If $H_{0k}(r)$ is not rejected, then $[p_k, p_{k+1}] \subseteq C_n(r)$, and so $\tilde{q} \in C_n(r)$. This means that

$$\mathbb{P}[\tilde{q} \notin C_n(r)] \leq \mathbb{P}[\text{reject } H_{0k}(r) \text{ at level } \alpha],$$

which is asymptotically at most α if the test of $H_{0k}(r)$ controls size.

Constructing such a test under Assumption FF requires first constructing an asymptotically normal estimator of the moments. Write the null hypothesis $H_{0j}(r)$ in terms of the parameter θ_0 for $j \in \{1, \dots, J-1\}$ as

$$H_{0j}(r) : \phi_j(\theta_0) \leq 0, \quad \text{for} \quad \phi_j(\theta) \equiv \begin{bmatrix} h(p_j; \theta) - L^{-1}(r) \\ L^{-1}(r) - h(p_{j+1}; \theta) \end{bmatrix},$$

where we suppress the dependence of H_{0j} and ϕ_j on r in the following.¹¹ Let $A_j \equiv \nabla \phi_j(\theta_0)$ denote the $2 \times K$ Jacobian of ϕ_j at θ_0 , and let $\hat{A}_j \equiv \nabla \phi_j(\hat{\theta}_n)$. By the delta method, we have that

$$\sqrt{n}(\phi_j(\hat{\theta}_n) - \phi_j(\theta_0)) \xrightarrow{d} \mathcal{N}(0, A_j \Sigma_0 A_j'),$$

¹⁰An alternative is to cover the entire set $q(r, F)$. This would be equivalent to covering the entire identified set $\mathcal{Q}^*(r)$ itself, because there exists $G \in \mathcal{F}^*$ with $q(r, G) = \mathcal{Q}^*(r)$; see the proof of Proposition 3 for additional details. Focusing on an arbitrary element of $q(r, F)$ is therefore more consistent with our approaches for the mean and variance, which are more standard in partial identification inference.

¹¹For $j = 0$ and $j = J$, one of the two inequalities is vacuous, since $s_0 = 0$ and $s_{J+1} = 1$, so ϕ_j consists only of the remaining component, and the maxima and minima below are taken over that single component.

recalling that Σ_0 is the asymptotic covariance matrix of $\hat{\theta}_n$.¹²

We implement the method proposed by Romano et al. (2014), which we briefly outline for the test of $H_{0j}(r)$ and a fixed $j \in \{0, \dots, J\}$. Define the test statistic

$$T_{nj} \equiv \max \left\{ \text{diag}\{\hat{\sigma}_{nj}\}^{-1} \sqrt{n} \phi_j(\hat{\theta}_n) \right\},$$

where $\hat{\sigma}_{nj} \in \mathbb{R}^2$ denotes the vector of estimated asymptotic standard deviations derived from $\hat{A}_j \hat{\Sigma}_n \hat{A}_j'$, and the maximum is taken over the two components of the vector. We reject $H_{0j}(r)$ when $T_{nj} > \hat{c}_{nj}$ for some suitable critical value $\hat{c}_{nj}$. The critical value is constructed by first constructing a level $1 - \delta$ upper confidence rectangle for $\phi_j(\theta_0)$,

$$M_{nj} \equiv \left\{ \mu \in \mathbb{R}^2 : \mu \leq \phi_j(\hat{\theta}_n) + \hat{e}_{nj} \hat{\sigma}_{nj} / \sqrt{n} \right\},$$

where $\delta \ll \alpha$ and $\hat{e}_{nj}$ is the $1 - \delta$ quantile of $\max\{\mathcal{N}(0, \hat{\Omega}_{nj})\}$, with

$$\hat{\Omega}_{nj} \equiv \text{diag}\{\hat{\sigma}_{nj}\}^{-1} \hat{A}_j \hat{\Sigma}_n \hat{A}_j' \text{diag}\{\hat{\sigma}_{nj}\}^{-1}.$$

Then the least favorable value of $\mu \in M_{nj}$ consistent with $H_{0j}(r)$ is given by

$$\mu_{nj}^* \equiv \min \left\{ 0, \phi_j(\hat{\theta}_n) + \hat{e}_{nj} \hat{\sigma}_{nj} / \sqrt{n} \right\},$$

where $0 \in \mathbb{R}^2$ and the minimum is taken componentwise. The critical value $\hat{c}_{nj}$ is the $1 - \alpha + \delta$ quantile of $\max\{\mathcal{N}(\text{diag}\{\hat{\sigma}_{nj}\}^{-1} \sqrt{n} \mu_{nj}^*, \hat{\Omega}_{nj})\}$. The test rejects $H_{0j}(r)$ if some component of μ_{nj}^* is equal to zero and $T_{nj} > \hat{c}_{nj}$. Checking if some component of μ_{nj}^* is equal to zero ensures that we do not reject when every point in M_{nj} is consistent with the null hypothesis.¹³

6 Inference Without Regular Shares

This section develops inference methods that are robust to failures of Assumption RS, based on the binning estimator of the choice shares. A challenge that this creates for inference is that the estimated choice shares can imply bounds for WTP features that

¹²Note that $\phi_j(\theta) = A_j \theta + b_j$ is affine in the common case that the index $h(p; \theta) = \alpha + \beta p$ is linear. In this case, the delta method is not necessary.

¹³The first-step confidence set is intended to increase power against alternatives in which one inequality is violated while the other is clearly slack. Typically, this costs power against alternatives in which both inequalities are violated. However, notice that in our setting, these alternatives can only happen if $s_j > s_{j+1}$, contrary to the model. This suggests that the two-step procedure of Romano et al. (2014) should perform quite well in our setting, a finding that is supported by the Monte Carlo simulations in Supplemental Appendix SA.2.

cross or are not even well-defined.

6.1 The mean

With non-monotonic choice shares, the confidence interval C_n from Section 5.1 implicitly embeds a specification test, meaning it will sometimes be empty or spuriously precise (Stoye, 2021). To avoid this undesirable behavior, we use the method developed by Stoye (2021) and report a $1 - \alpha$ confidence interval for mean WTP of the form

$$\tilde{C}_n \equiv \left[\hat{m}_{l,n} - \frac{\hat{c}_n \hat{\sigma}_{l,n}}{\sqrt{n}}, \hat{m}_{u,n} + \frac{\hat{c}_n \hat{\sigma}_{u,n}}{\sqrt{n}} \right] \cup C_n^*,$$

where C_n^* is a standard $1 - \alpha$ confidence interval for the pseudo-true parameter

$$m^* \equiv \left[\frac{\sigma_u}{\sigma_l + \sigma_u} \quad \frac{\sigma_l}{\sigma_l + \sigma_u} \right]' (A\mathbf{s} + b),$$

with all of the quantities defined analogously to Section 5.1. The adjusted confidence interval covers the pseudo-true parameter under misspecification, but it also prevents the confidence interval from being empty or misleadingly short under correct specification. The critical value $\hat{c}_n$ is calibrated to ensure correct coverage while reducing the conservative bias that arises from taking the union of two intervals, as in Stoye (2021, Equation (2.2)). As long as the estimated correlation between $\hat{m}_{l,n}$ and $\hat{m}_{u,n}$ is not too high ($\hat{\rho}_n \leq 0.8$), we can simply use $\hat{c}_n = \Phi^{-1}(1 - \alpha)$.

6.2 The variance

The analog estimator of the variance bounds is not even well-defined when the estimated choice shares are non-monotonic. To address this, we construct an adjusted estimator of the variance bounds that is linearized around the isotonic regression estimator $\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n)$ that projects $\hat{\mathbf{s}}_n$ onto the order polytope $\mathcal{S}$ according to some weighted L^2 norm. The adjusted estimator is

$$\hat{\nu}_n(\hat{\mathbf{s}}_n) \equiv \nu(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n)) + \nabla \nu(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n)) (\hat{\mathbf{s}}_n - \Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n)), \quad (8)$$

where $\nabla \nu(\mathbf{s})$ is the Hadamard derivative found in Lemma 1. As we showed in that lemma, this function is continuous on $\mathcal{S} \setminus \text{vert}(\mathcal{S})$, including for $\mathbf{s} \in \text{bd}(\mathcal{S})$, just not at the vertices. The following proposition uses these results to derive the limiting distribution of the adjusted estimator.

Proposition 5. Suppose that both choke prices p_0 and p_{J+1} are finite. Let $\hat{\mathbf{s}}_n$ be the binning estimator of Section 4 and let $\hat{\nu}_n(\hat{\mathbf{s}}_n)$ be the adjusted estimator (8) of the

variance bounds. If $\mathbf{s} \in \mathcal{S} \setminus \text{vert}(\mathcal{S})$, then

$$\sqrt{n}(\hat{\nu}_n(\hat{\mathbf{s}}_n) - \nu(\mathbf{s})) \xrightarrow{d} \mathcal{N}(0, \nabla \nu(\mathbf{s}) \Gamma_0 \nabla \nu(\mathbf{s})'),$$

and $\nabla \nu(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n)) \hat{\Gamma}_n \nabla \nu(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n))'$ is a consistent estimator of the asymptotic covariance matrix.

As in Section 6.1, we combine Proposition 5 with the confidence interval of Stoye (2021), because the estimated bounds can still cross with positive probability.

Proposition 5 excludes the possibility that the true choice shares $\mathbf{s}$ are a vertex of $\mathcal{S}$. This case seems empirically unlikely; the researcher would have to have designed their experiment so badly that the true choice shares form a step vector $\mathbf{s} = (0, \dots, 0, 1, \dots, 1)$, in which case the estimated choice shares $\hat{\mathbf{s}}_n$ will also be this step vector. In this case, the data is only informative about the support of the WTP distribution, which must lie entirely in the step region.

6.3 The quantiles

The method developed in Section 5.3 is designed for cases in which the population choice shares are regular, but it does not require it, and it could be implemented even if the estimated choice shares are not regular, as commonly happens with the binning estimator. While doing so would not threaten size control, it has notable drawbacks, because the approach tests $H_{0j}(r)$ by only using estimated shares at the nearby prices p_j and p_{j+1} . If a tight parameterization is maintained via Assumption FF, then this is sensible, because the estimated $\hat{\theta}_n$ already uses information from multiple prices. Without such a parameterization, the test is not utilizing information at other prices, even though their population shares are related through monotonicity. As a more practical matter, using the method in Section 5.3 with shares that are not monotone often leads to confidence sets that are the union of several disjoint intervals, which is intuitively unappealing.¹⁴

A method that uses information from all of the shares together can be developed by rewriting the null hypothesis as

$$H_{0j}(r) : s_k - r \leq 0 \ \forall k \in \{0, \dots, j\} \quad \text{and} \quad r - s_k \leq 0 \ \forall k \in \{j+1, \dots, J+1\}.$$

This formulation is equivalent to the one in Section 5.3 because population choice shares

¹⁴Applying the method in Section 5.3 with the binning estimator produces disjoint confidence intervals in roughly 75% of the cases we consider. The convex hull of these intervals tends to be considerably longer than the intervals that use the method developed in this section.

are assumed to be monotonically increasing. A test statistic based on this formulation incorporates information from all of the shares through the full set of inequalities, rather than restricting attention to just two components. The test statistic we use is

$$T_{nj} \equiv \max\{\text{diag}\{d_j\} \text{diag}\{\hat{\varsigma}_n\}^{-1} \sqrt{n}(\hat{\mathbf{s}}_n - r)\},$$

where $\hat{\varsigma}_n$ denotes the vector of estimated asymptotic standard deviations derived from $\hat{\Gamma}_n$, and $d_j \in \mathbb{R}^J$ is the vector whose k th component is 1 for $k \leq j$ and -1 for $k > j$.

Lemma 3 in Appendix C shows that T_{nj} is unimodal as a function of the index j , meaning that it is weakly decreasing up to a certain point, then weakly increasing afterwards. This implies that the sublevel sets $\{j \in \{0, \dots, J\} : T_{nj} \leq c\}$ consist of consecutive indices, so that confidence sets formed through test inversion will be intervals.¹⁵ This property also suggests a natural estimator of the quantile bounds because the set $\bigcup_{j \in \hat{\mathcal{J}}_n} [p_j, p_{j+1}]$, where $\hat{\mathcal{J}}_n \equiv \arg \min_{0 \leq j \leq J} T_{nj}$, is a sample analog of $\mathcal{Q}^*(r)$ that is guaranteed to be a non-empty interval.

The two-step approach of Romano et al. (2014) proceeds in much the same way as in Section 5.3. First, we construct a level $1 - \delta$ upper confidence rectangle for $\text{diag}\{d_j\}(\mathbf{s} - r)$,

$$M_{nj} \equiv \{\mu \in \mathbb{R}^J : \mu \leq \text{diag}\{d_j\}(\hat{\mathbf{s}}_n - r) + e\hat{\varsigma}_n/\sqrt{n}\},$$

where $\delta \ll \alpha$ and e is the $1 - \delta$ quantile of $\max\{\mathcal{N}(0, \mathbf{I})\}$, with $\mathbf{I}$ denoting the $J \times J$ -dimensional identity matrix (since Γ_0 is a diagonal matrix for binning estimators of the choice shares). The challenge now is to find the least favorable value of $\mu \in M_{nj}$ consistent with both the null hypothesis $H_{0j}(r)$ and the model restriction that $\mathbf{s} \in \mathcal{S}$.¹⁶ The following proposition shows that, as in Romano et al. (2014), there is a single value of μ that is least favorable.

Proposition 6. Let $r \in [0, 1]$ and $j \in \{0, \dots, J\}$, and define $\mu_{nj}^* \in \mathbb{R}^J$ by its compo-

¹⁵This reasoning can in principle be jeopardized by using the two-step approach of Romano et al. (2014), which leads to critical values $\hat{c}_{nj}$ that depend on j . In practice, we find that the critical value does not vary enough with j to produce disconnected confidence sets.

¹⁶This restriction was not needed for the test in Section 5.3, because that test involved only s_j and s_{j+1} , and $H_{0j}(r)$ already implies $s_j \leq s_{j+1}$. Here it matters, because $H_{0j}(r)$ does not restrict the order of $s_1, \dots, s_j$ or of $s_{j+1}, \dots, s_J$.

nents

$$\begin{aligned} [\mu_{nj}^*]_k &= \min_{l \in \{k, \dots, j\}} \{\min\{r, \hat{s}_{l,n} + e\hat{\varsigma}_{l,n}/\sqrt{n}\}\} - r \text{ for all } k \in \{1, \dots, j\}, \quad \text{and} \\ [\mu_{nj}^*]_k &= r - \max_{l \in \{j+1, \dots, k\}} \{\max\{r, \hat{s}_{l,n} - e\hat{\varsigma}_{l,n}/\sqrt{n}\}\} \text{ for all } k \in \{j+1, \dots, J\}. \end{aligned}$$

Then $\mu_{nj}^* \in M_{nj}$ is consistent with both $H_{0j}(r)$ and $\mathbf{s} \in \mathcal{S}$, and $\mu \leq \mu_{nj}^*$ for every $\mu \in M_{nj}$ that is consistent with both.

Since the critical value is weakly increasing in each component of μ , μ_{nj}^* yields the largest critical value among all such μ . We compute $\hat{c}_{nj}$ as the $1 - \alpha + \delta$ quantile of $\max\{\mathcal{N}(\text{diag}\{\hat{\varsigma}_n\}^{-1}\sqrt{n}\mu_{nj}^*, \mathbf{I})\}$, and reject $H_{0j}(r)$ if some component of μ_{nj}^* is equal to zero and $T_{nj} > \hat{c}_{nj}$.

7 Re-evaluating the Value of Alternative Work Arrangements

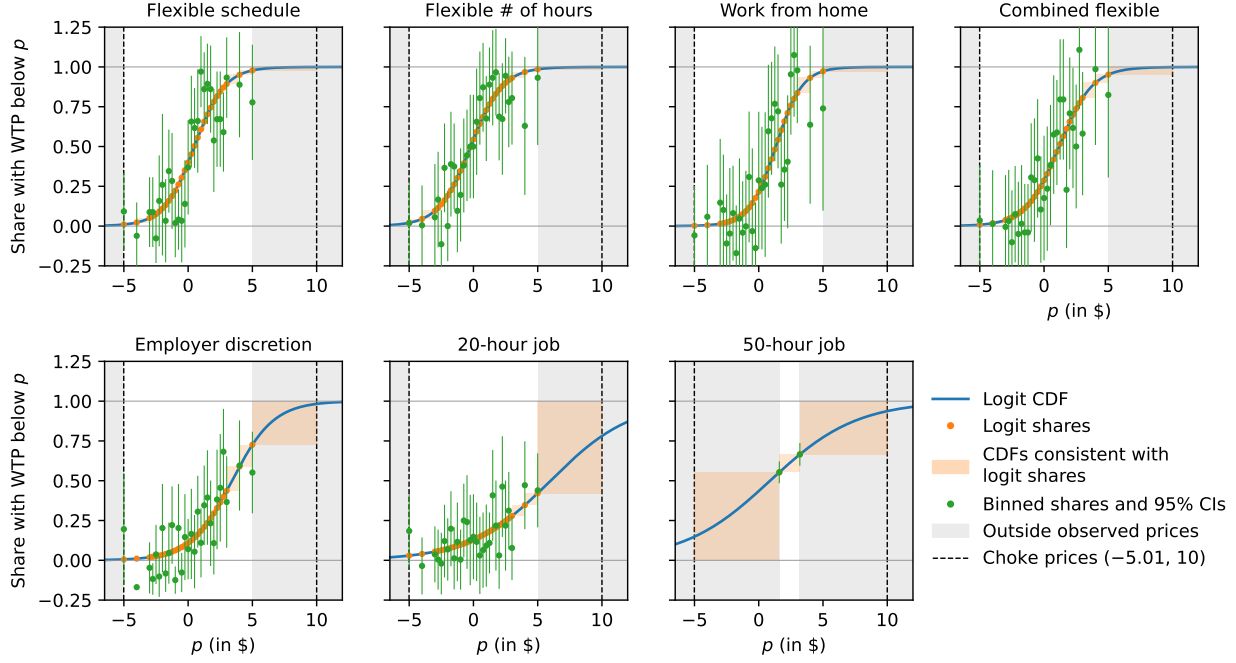
In this section, we apply our methods to the binary choice experiment of [Mas and Pallais \(2017\)](#), who measure how much workers value alternative work arrangements. Supplemental Appendix [SA.2](#) reports the results of a Monte Carlo simulation calibrated to this experiment. We find that our confidence intervals are close to size-correct for the mean and variance (standard deviation), while being conservative for quantiles.

7.1 Choice shares

Applicants to a national call center chose between a baseline position (40 hours, 9 A.M.–5 P.M., Monday–Friday, on site) and a randomly selected alternative, with a randomly assigned wage difference. We let $Y_i = 1$ denote choosing the position with the amenity, and P_i the hourly wage given up by doing so. In the four flexibility arms (a flexible schedule, a flexible number of hours, working from home, and the three combined), the amenity is the alternative position ($Y_i = 1$). For employer discretion over scheduling, part-time work (a 20-hour job), and overtime (a 50-hour job), which [Mas and Pallais \(2017\)](#) treat as disamenities, it is the baseline position ($Y_i = 0$), so WTP is the willingness to pay to avoid the alternative. In every arm except overtime, the observed prices run from $-\$5$ to $\$5$, in steps of $\$0.25$ between $-\$3$ and $\$3$ and steps of $\$1$ outside that range.

Following [Mas and Pallais \(2017\)](#), we allow for applicants who did not read the descriptions (were inattentive). We use their estimated inattention rates, which are between 25% and 49%, depending on the arm, together with the adjustment in Ap-

Figure 3: Estimated choice shares



Notes: Binning estimates of inattention-corrected choice shares are shown in green together with 95% confidence intervals. The blue line is the estimated logit distribution function, while the orange dots show shares at the observed prices implied by the logit fit and the shaded orange region shows the distribution functions consistent with those shares. The region in gray lies outside of the support of observed prices. Dashed lines mark the choke prices $p_0 = -\$5.01$ and $p_{J+1} = \$10$. The 50-hour arm has only two prices, so the logit and binning estimates of choice shares are the same.

pendix A.¹⁷ Their inattention estimates assume that $-\$5$ is a choke price, so that no attentive applicant takes the dominated position when it pays $\$5$ less. Following their assumption, we set $p_0 = -\$5.01$, so that $-\$5$ can be included in the grid of observed prices. The upper choke price is $p_{J+1} = \$10$ in our main specification, and we also report results when p_{J+1} is set to $\$6$ and $\$12$. Figure 3 plots binning estimates of the inattention-corrected shares together with the logit fit to these shares.

In the following, we compare three estimates of each feature of the WTP distribution: the logit point estimate of Mas and Pallais (2017); bounds under Assumption FF with a logistic link and a linear index, which use the logit's fitted shares but not its extrapolation between and beyond the observed prices; and nonparametric bounds, which use the binned shares. The first comparison shows how much the logit relies on extrapolation, and the second how much it relies on the logistic fit at the observed prices. We

¹⁷See Section II of Mas and Pallais (2017) for further details. We do not account for uncertainty in the estimated inattention rates when performing inference.

construct confidence intervals for the bounds that use the logit shares with the methods in Section 5, and for the nonparametric bounds with the methods in Section 6. Figure 4 shows bounds on the mean and standard deviation for each amenity, while Figure 5 shows bounds on quantiles. Tables SA.2–SA.7 in Supplemental Appendix SA.3 report all of our estimates numerically.

7.2 Flexibility

For three of the four flexibility arms, the lower nonparametric bound on mean WTP lies above the logit estimate (Figure 4): [\$0.65, \$1.82] against \$0.48 for a flexible schedule, [\$0.13, \$0.82] against $-\$0.22$ for flexible hours, and [\$1.53, \$3.00] against \$1.33 for working from home. These lower bounds do not depend on the assumed upper choke price. Extrapolation does not explain the gap, because the estimated bounds when using the logit shares contain the logit point estimate in every arm. Instead, the gap comes from the logistic fit at the observed prices, which has difficulty matching both the jump near zero and the mass above \$5 (see Figure 3) that is present in each arm.

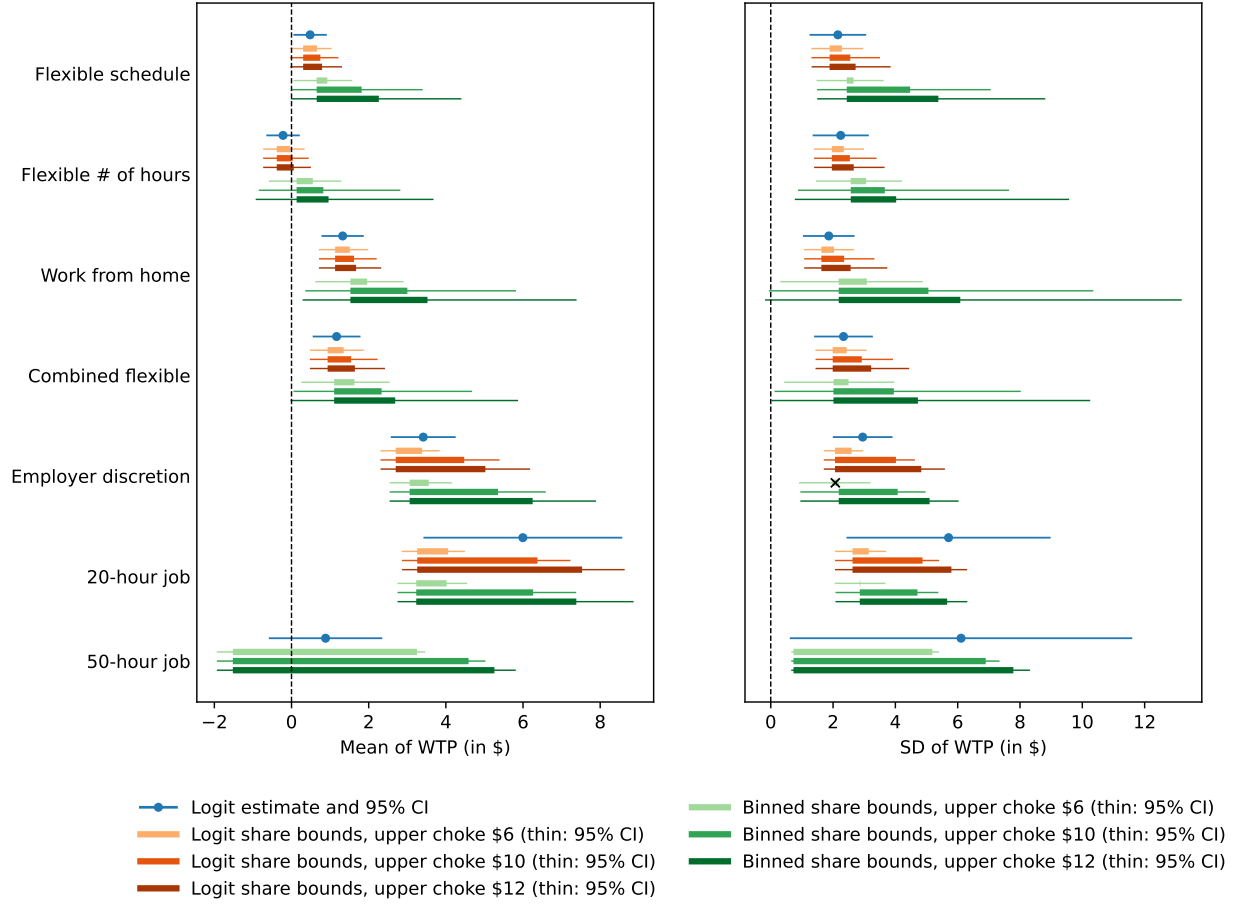
Mas and Pallais (2017, p. 3734) anticipate this problem, noting that “the logistic CDF does not capture the extreme nonlinearity.” As a solution, they also estimate a breakpoint model with a mass point at zero, which results in estimates of mean WTP (their Online Appendix Table 4) that lie inside our bounds for a flexible schedule (\$1.75), working from home (\$1.72), and the combined arm (\$1.42), and just below them for flexible hours ($-\$0.01$). Our bounds reach the same conclusion without needing to assume a specific distribution of valuations.

The quantile bounds in Figure 5 tell a similar story. The median valuation for a flexible schedule is estimated to be between \$0 and \$0.25 with a tight confidence interval, supporting the conclusion of Mas and Pallais (2017) that most workers do not value scheduling flexibility. However, larger quantiles lead to significantly wider bounds, reflecting a long tail of individuals who do value flexibility. The logit distribution is unable to capture these individuals. Bounds on the standard deviation in Figure 4 show that the logit point estimates tend to understate dispersion for the flexibility arms.

7.3 Employer discretion

Mas and Pallais (2017) find that the average applicant would give up 20% of wages to avoid employer discretion, and that almost 40% would turn the job down even if it paid 25% more (Mas and Pallais, 2017, p. 3726). Figure 4 shows that their first conclusion holds without the logit distributional assumption. The nonparametric lower bound on

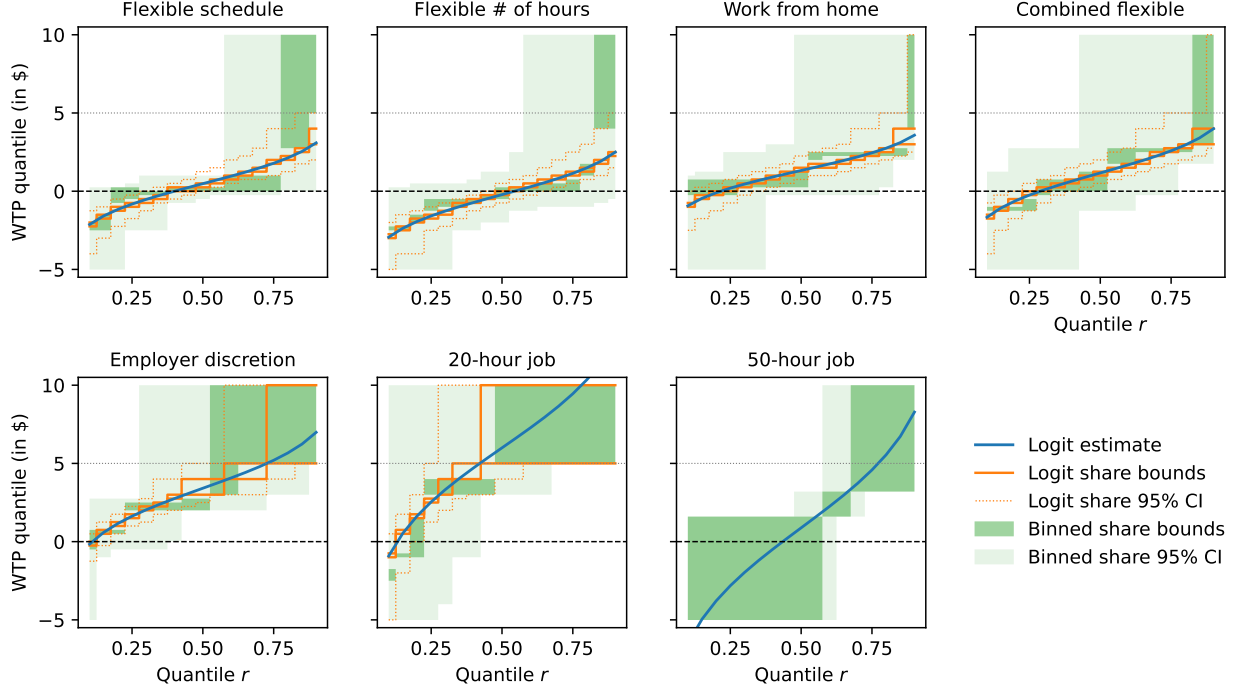
Figure 4: The mean and standard deviation of WTP



Notes: Blue shows logit point estimates and 95% confidence intervals from [Mas and Pallais \(2017\)](#). Orange shows distribution-free bounds and 95% confidence intervals when choice shares are estimated with a logit model with a linear index in price, while green shows the same with binning estimates of the choice shares. An $\times$ marks bound estimates that cross (lower above upper), which can happen when using binning estimates of the choice shares due to sampling error.

mean WTP is \$3.06, or 19% of wages, for any upper choke price, and its confidence interval starts at \$2.56. Figure 4 also shows that the logit distributional assumption may lead them to understate how much the average applicant values avoiding employer discretion. Because both the logit and the binning estimates allow for significant mass above the highest observed price of \$5, the upper bound on mean WTP depends heavily on the upper choke price, rising from \$3.56 to \$6.25 as the upper choke price rises from \$6 to \$12.

Figure 5: Quantiles of WTP



Notes: Blue shows the logit point estimate. Orange shows distribution-free bounds and 95% confidence intervals when choice shares are estimated with a logit model with a linear index in price, while green shows the same with binning estimates of the choice shares. The upper choke price for these bounds is set to $p_{J+1} = \$10$. The dotted gray line marks \$5, the highest observed price in every arm except the 50-hour arm. For the 50-hour arm, the bounds under Assumption FF equal the nonparametric bounds and are not shown.

7.4 Hours and overtime

Mas and Pallais (2017) estimate that the median applicant would give up \$6.00 an hour to avoid a 20-hour job, which implies a median value of time between 20 and 40 hours of \$4, far below the \$16 market wage (Mas and Pallais, 2017, p. 3741).¹⁸ Figure 3 shows that both logit and binning estimates have over 50% of applicants with WTP above the highest observed price of \$5. The implication is that the estimate of WTP to avoid a 20-hour job depends heavily on how one extrapolates beyond \$5. In particular, the median WTP must be at least \$5, but is otherwise restricted only by the upper choke price. The distribution-free bounds strongly support the conclusion of Mas and Pallais (2017) that most workers have a median value of time that is well below the market wage.

Because they use a logit model, Mas and Pallais (2017) also estimate that the *mean*

¹⁸Following Mas and Pallais (2017, p. 3741, footnote 35), the value of time between 20 and 40 hours is $16 - 2 \text{ WTP}$. Analogously, the value of time between 40 and 50 hours is $16 + 5 \text{ WTP}$.

Table 1: Gender differences in mean WTP

	p_{J+1}	Logit	Distribution-free bounds	
			Logit shares	Binned shares
Employer discretion	\$6	2.16	[1.06, 2.44]	[1.49, 2.43]
		[0.50, 3.81]	[0.33, 3.19]	[0.49, 3.35]
	\$10		[0.27, 4.04]	[0.61, 4.91]
			[−0.76, 5.39]	[−0.95, 6.56]
	\$12		[−0.13, 4.83]	[0.16, 6.15]
			[−1.34, 6.53]	[−1.79, 8.30]
Work from home	\$6	0.91	[0.55, 1.28]	[0.00, 1.27]
		[−0.11, 1.93]	[−0.32, 2.12]	[−1.98, 3.21]
	\$10		[0.48, 1.35]	[0.53, 3.27]
			[−0.56, 2.23]	[−3.16, 7.79]
	\$12		[0.44, 1.39]	[0.79, 4.27]
			[−0.70, 2.30]	[−4.04, 10.36]

Notes: Bounds on the difference in mean WTP (women minus men). For each arm and upper choke price, the first row reports the logit point estimate or bound estimates, and the second row 95% CIs. The logit estimates do not depend on the upper choke price.

WTP to avoid a 20-hour job is \$6.00. Figure 4 shows that the distribution-free evidence on the mean provides a somewhat different picture. It shows that the data is consistent with a mean WTP that is as low as \$3.23 for any upper choke price, which implies a value of time of at most \$9.54, closer to the \$16 market wage.

The overtime arm offered only two prices, \$1.60 and \$3.20. Both jobs paid the same \$16 base wage, with the two prices reflecting overtime premiums from 1.5 or 2 times the base wage, expressed per hour of a 50-hour week. This leaves estimates of the mean, standard deviation, and most quantiles highly dependent on functional form. The logit point estimate of mean WTP reported by Mas and Pallais (2017, Table 6) is \$0.88, which Figure 4 shows is only one of a range of possible mean WTPs consistent with the data, between −\$1.52 and \$4.59. The implied value of time between 40 and 50 hours is \$8.40 to \$38.95, compared to the point estimate of \$20.41 reported by Mas and Pallais (2017, Table 6).

7.5 Gender differences

Mas and Pallais (2017) find that women are willing to pay twice as much as men to avoid employer discretion, and more than twice as much to work from home, although the latter difference is not statistically significant. Table 1 shows that both conclusions are potentially driven by the functional form of the logit. The distribution-free lower bounds on the difference in mean WTP to avoid employer discretion are near zero

when the upper choke price is \$10 or higher. Conversely, distribution-free bounds on the difference in WTP to work from home contain a wide range of possibilities that are driven by the choke price. Table SA.8 in Supplemental Appendix SA.3 shows that there is also more evidence of differences by gender in WTP for other amenities when using distribution-free bounds.

8 Conclusion

In this paper we have provided distribution-free methods for inference on features of the distribution of WTP in binary choice experiments, such as contingent valuation studies. The methods are relatively straightforward to implement and provide confidence intervals that seem to be not overly conservative. They allow researchers to assess the extent to which distributional assumptions drive their conclusions about WTP. One clear message is that distributional assumptions will tend to have a bigger impact when there are fewer prices in the experiment, or when these prices do not adequately cover the support of WTP. In a reanalysis of Mas and Pallais (2017), we found that some of their headline results on the value for flexibility are supported by distribution-free estimates, while other results on the value of overtime work and gender differences seem to be driven by the functional form of the logit.

A Identification with Inattention

Identifying the distribution of valuations from a discrete choice experiment is complicated by the possibility that participants' decisions are an imperfect proxy for their actual market behavior. Recent experimental designs in labor economics, such as those in [Mas and Pallais \(2017\)](#) and [He et al. \(2021\)](#), try to mitigate this concern by embedding the field experiment in a real market setting, such as a job application or advertisement. Nevertheless, there is still often a concern that some proportion of participants are inattentive ([Mas and Pallais, 2017](#)). The observed choices of these participants in the experiment might not coincide with what their actual market choices would be. In this appendix, we show how to adjust the framework to accommodate inattention.

We write potential choice outcomes for individual i as $Y_i(a, p)$ where $a \in \{0, 1\}$ represents attention and $p \in \mathcal{P}$ is a price, as before. Then $Y_i(1, p)$ represents their (attentive) market choice, while $Y_i(0, p)$ represents their inattentive choice. Observed choices are given by $Y_i \equiv Y_i(A_i, P_i)$, where A_i is an unobserved indicator for whether individual i is attentive or not.

We assume that market choices follow a threshold-crossing model, as in [Section 2.2](#).

Assumption TCI. (Threshold crossing with inattention) For a pair of random variables (V_i, B_i) on $\overline{\mathbb{R}} \times \{0, 1\}$, potential market choices are given by

$$Y_i(1, p) = \mathbb{1}[V_i > p] + \mathbb{1}[V_i = p]B_i.$$

By contrast, we leave inattentive choices unmodeled, requiring only that they satisfy the following exclusion restriction.

Assumption EXI. (Exclusion restriction with inattention) For all $p \in \mathcal{P}$, $\mathbb{P}[Y_i(0, p) = Y_i(0, \cdot)] = 1$ for some random variable $Y_i(0, \cdot)$ that does not depend on p .

We also modify [Assumption RA](#) accordingly for the new potential outcomes.

Assumption RAI. (Random assignment with inattention) P_i is independent of $\{A_i, Y_i(0, \cdot), Y_i(1, p) : p \in \mathcal{P}\}$.

Under [Assumptions EXI](#) and [RAI](#), prices are randomly assigned and have no causal effect on attentiveness, inattentive choices, or valuations. Combined with [Assumption](#)

TCI, it follows that

$$\begin{aligned}
& \mathbb{P}[Y_i = 0 \mid P_i = p] \\
&= \mathbb{P}[(1 - A_i)Y_i(0, \cdot) + A_iY_i(1, p) = 0 \mid P_i = p] \\
&= \mathbb{P}[(1 - A_i)Y_i(0, \cdot) + A_i(\mathbb{1}[V_i > p] + \mathbb{1}[V_i = p]B_i) = 0] \\
&= \mathbb{P}[A_i = 0] \mathbb{P}[Y_i(0, \cdot) = 0 \mid A_i = 0] \\
&\quad + (1 - \mathbb{P}[A_i = 0])(\mathbb{P}[V_i \leq p \mid A_i = 1] - \mathbb{P}[V_i = p, B_i = 1 \mid A_i = 1]).
\end{aligned}$$

Without further assumptions, the population distribution of valuations $\mathbb{P}[V_i \leq p]$ is not point identified at $p \in \mathcal{P}$, even if we put aside tie-breaking issues (for instance, by assuming $\mathbb{P}[B_i = 0] = 1$). This is simply because we never observe market choices for inattentive individuals, who constitute a non-random subset of the population.

A natural response is to shift the focus to identifying the distribution of valuations among the attentive subpopulation. Let $F(p) \equiv \mathbb{P}[V_i \leq p \mid A_i = 1]$ denote the distribution of valuations among attentive individuals. For $j \in \{1, \dots, J\}$, define the following inattention-adjusted choice shares

$$\begin{aligned}
s_j &\equiv \mathbb{P}[Y_i(1, p_j) = 0 \mid A_i = 1] \\
&= \frac{\mathbb{P}[Y_i = 0 \mid P_i = p_j] - \mathbb{P}[A_i = 0] \mathbb{P}[Y_i(0, \cdot) = 0 \mid A_i = 0]}{1 - \mathbb{P}[A_i = 0]}.
\end{aligned} \tag{9}$$

All of our identification results from the main text immediately apply to this attentive subpopulation by simply using the inattention-adjusted choice shares. The only change is in interpretation: WTP now refers to the attentive subpopulation.

The inattention-adjusted choice shares need further assumptions to be point identified. The following two parameters need to be pinned down:

- $\mathbb{P}[A_i = 0]$, the proportion of individuals who are inattentive, and
- $\mathbb{P}[Y_i(0, \cdot) = 0 \mid A_i = 0]$, the proportion of inattentive individuals who do not choose the amenity.

These two parameters do not depend on price, so they can be identified from two restrictions on the inattention-adjusted choice shares (9). There are three ways to obtain these restrictions. The experiment can include choke prices at both ends of the price grid, which is a finite-grid version of the approach in Hausman et al. (1998). It can include a choke price at one end, and the researcher can assume a value for the inattentive choice proportion $\mathbb{P}[Y_i(0, \cdot) = 0 \mid A_i = 0]$. Or, the parameters can be taken from an auxiliary experiment. Mas and Pallais (2017) use the second route in most of

their treatments and the third in the rest.

The first route involves strengthening Assumption [CP](#) so that the binary choice experiment includes both lower and upper choke prices.

Assumption CPI. (Choke prices with inattention) The extremal observed prices $p_1, p_J \in \mathcal{P}$ are also choke prices, meaning

$$s_1 \equiv \mathbb{P}[Y_i(1, p_1) = 0 \mid A_i = 1] = 0,$$

$$\text{and } s_J \equiv \mathbb{P}[Y_i(1, p_J) = 0 \mid A_i = 1] = 1.$$

Under Assumptions [RAI](#), [EXI](#), [TCI](#), and [CPI](#), we have that

$$\mathbb{P}[A_i = 0] = 1 - (\mathbb{P}[Y_i = 0 \mid P_i = p_J] - \mathbb{P}[Y_i = 0 \mid P_i = p_1])$$

$$\mathbb{P}[Y_i(0, \cdot) = 0 \mid A_i = 0] = \frac{\mathbb{P}[Y_i = 0 \mid P_i = p_1]}{1 - (\mathbb{P}[Y_i = 0 \mid P_i = p_J] - \mathbb{P}[Y_i = 0 \mid P_i = p_1])}.$$

The inattention-adjusted choice shares s_j are then point identified as

$$s_j = \frac{\mathbb{P}[Y_i = 0 \mid P_i = p_j] - \mathbb{P}[Y_i = 0 \mid P_i = p_1]}{\mathbb{P}[Y_i = 0 \mid P_i = p_J] - \mathbb{P}[Y_i = 0 \mid P_i = p_1]}.$$

Note that this route requires more than simply assuming that some hypothetical and potentially infinite choke prices p_0, p_{J+1} exist, as in Assumption [CP](#). The choke prices actually need to have been included in the experiment.

[Mas and Pallais \(2017\)](#) take the second route. They assume that inattentive individuals make choices as if by flipping a fair coin, so $\mathbb{P}[Y_i(0, \cdot) = 0 \mid A_i = 0] = 1/2$. In this case, only a single choke price is required to achieve point identification of the inattention-adjusted choice shares. They assume that $p_1 = -\$5$ is a choke price: no attentive applicant chooses the position without the amenity when it pays \$5 less. Then $\mathbb{P}[Y_i = 0 \mid P_i = p_1] = \mathbb{P}[A_i = 0]/2$, so the inattention rate is twice the share of applicants who choose the dominated position. In effect, the coin-flip assumption takes the place of an upper choke price. [Mas and Pallais \(2017\)](#) explicitly decline to treat the upper end of their price grid as a choke price, “because there might be applicants who have a strong preference for flexibility” ([Mas and Pallais, 2017](#), p. 3738).

The 20- and 50-hour treatments in [Mas and Pallais \(2017\)](#) have no dominated position, so neither the first nor the second route is available there. The authors instead take the third route by imposing the average inattention rate from their other treatments. This amounts to deducing $\mathbb{P}[A_i = 0]$ from an auxiliary experiment, under an appropriate form of structural invariance across experiments. They continue to

maintain the coin-flip assumption on the choices of the inattentive.

B Quantile Definitions and Binary Labels

Researchers make a seemingly innocuous practical decision of whether to code the indicator of interest as $Y_i = 1$ or $Y_i = 0$ or, equivalently, whether to use Y_i or $1 - Y_i$. The convention used by [Mas and Pallais \(2017\)](#) is that the amenity $Y_i = 1$ reflects the researcher’s prior understanding of what constitutes a “good.” This assessment may be based on economic theory if, for example, the amenity expands individuals’ choice sets. However, if neither of the presented alternatives clearly dominates the other, the definition of the choice indicator is less clear-cut.

This arbitrary decision does not affect identification of means or variances. Let $W_i = -V_i$ denote WTP under the alternative definition of the amenity, $1 - Y_i = 1$, and let F_W and F_V be the respective distributions of W_i and V_i . Then $m(F_W) = -m(F_V)$ and $v(F_W) = v(F_V)$ when these quantities exist, as one would expect.

However, for quantiles, the labeling choice does make a difference. For example, consider the conventional single-valued definition $\underline{q}(r, G) \equiv \inf\{p \in \overline{\mathbb{R}} : G(p) \geq r\}$, as in [Parzen \(1979\)](#). Let $a, b > 0$, and suppose V_i equals $-a$ or b with equal probability. Then we have

$$\underline{q}(0.5, F_V) = -a, \quad \text{and} \quad \underline{q}(0.5, F_W) = -b,$$

meaning both median WTP for the amenity and median WTP to avoid the amenity are negative. The reason that $\underline{q}$ leads to this conclusion is that relabeling swaps lower and upper quantiles. Since $F_W(p) = 1 - \lim_{p' \uparrow -p} F_V(p')$, the definitions of $\underline{q}$ and $\bar{q}$ in [Section 3.4](#) imply that

$$\underline{q}(r, F_W) = -\bar{q}(1 - r, F_V) \quad \text{and} \quad \bar{q}(r, F_W) = -\underline{q}(1 - r, F_V)$$

for all $r \in [0, 1]$.

Using the set-valued definition avoids this type of nonsensical conclusion, resulting instead in

$$q(0.5, F_V) = [-a, b], \quad \text{and} \quad q(0.5, F_W) = [-b, a],$$

as one would intuitively expect. More generally,

$$q(r, F_W) = -q(1 - r, F_V) \equiv \{-p : p \in q(1 - r, F_V)\}$$

for all $r \in [0, 1]$, so that identified sets for quantiles under one labeling translate directly into identified sets under the other.

C Proofs

Proof of Proposition 1. Let $j \in \{0, \dots, J+1\}$. By Assumptions [RA](#) and [CP](#), we have $s_j = \mathbb{P}[Y_i(p_j) = 0]$, where choice shares s_j are defined as in Section [3.1](#). By Assumption [TC](#), we have

$$\begin{aligned} \mathbb{P}[Y_i(p_j) = 0] &= \mathbb{P}[\mathbb{1}[V_i > p_j] + \mathbb{1}[V_i = p_j]B_i = 0] \\ &= \mathbb{P}[V_i \leq p_j] - \mathbb{P}[V_i = p_j, B_i = 1] = \mathbb{P}[V_i < p_j] + \mathbb{P}[V_i = p_j, B_i = 0]. \end{aligned}$$

Examining the last two equalities, we see that

$$\mathbb{P}[V_i < p_j] \leq s_j \leq \mathbb{P}[V_i \leq p_j].$$

We will use the above inequalities to develop bounds on $F(p) = \mathbb{P}[V_i \leq p]$. We will then check that the proposed bounds are valid distribution functions on $\overline{\mathbb{R}}$. Finally, we will verify that the proposed bounds are sharp.

First, consider the lower bound. For $p \geq p_j$, we have that

$$s_j \leq \mathbb{P}[V_i \leq p_j] \leq \mathbb{P}[V_i \leq p],$$

since $F(p)$ is weakly increasing. Recalling that $s_{J+1} = 1$, it follows that the function

$$F_L(p) \equiv \mathbb{1}[p_{J+1} \leq p] + \sum_{j=0}^J \mathbb{1}[p_j \leq p < p_{j+1}]s_j$$

provides a lower bound for $F(p)$, since $F(p) \geq 0$.

Second, consider the upper bound. For $p < p_j$, we have that

$$\mathbb{P}[V_i \leq p] \leq \mathbb{P}[V_i < p_j] \leq s_j,$$

since $\mathbb{P}[V_i < p_j]$ may be expressed as the left limit

$$\lim_{p \uparrow p_j} \mathbb{P}[V_i \leq p] = \mathbb{P}[V_i < p_j].$$

Recalling that $s_0 = 0$, it follows that the function

$$F_U(p) \equiv \mathbb{1}[p_{J+1} \leq p] + \sum_{j=0}^J \mathbb{1}[p_j \leq p < p_{j+1}]s_{j+1}$$

provides an upper bound for $F(p)$, since $F(p) \leq 1$.

We will now check that $F_L, F_U \in \mathcal{F}$. First, notice that both bounds are right-continuous by construction. In addition, $F_L(+\infty) = F_U(+\infty) = 1$, with $F_L(-\infty) \geq 0$ and $F_U(-\infty) \geq 0$ as required. Finally, both bounds are weakly increasing if $s_j \leq s_{j+1}$ for all $j \in \{0, \dots, J\}$. But this follows immediately from our previous results, since $p_j < p_{j+1}$ by assumption, so we have

$$s_j \leq \mathbb{P}[V_i \leq p_j] \leq \mathbb{P}[V_i < p_{j+1}] \leq s_{j+1},$$

as required.

All that remains is to verify that the proposed bounds are sharp. Consider $G \in \mathcal{F}$ such that $F_L(p) \leq G(p) \leq F_U(p)$ for all $p \in \overline{\mathbb{R}}$. We would like to show that there exists a joint distribution of (V_i, B_i) on $\overline{\mathbb{R}} \times \{0, 1\}$ consistent with Assumptions [RA](#), [TC](#), [CP](#), and the observed distribution of Y_i given P_i such that V_i has marginal distribution G .

Suppose that V_i has marginal distribution G . To ensure consistency with our assumptions and observed choice shares, we must construct a conditional distribution of B_i given V_i such that for $j \in \{0, \dots, J+1\}$, we have

$$s_j = \mathbb{P}[V_i \leq p_j] - \mathbb{P}[V_i = p_j, B_i = 1].$$

Substituting in the assumed distribution of V_i and rearranging, we see that we need

$$\mathbb{P}[V_i = p_j, B_i = 1] = G(p_j) - s_j.$$

So, to arrive at a valid distribution of B_i given $V_i = p_j$, we must check that $G(p_j) - s_j$ is bounded between 0 and $\mathbb{P}[V_i = p_j]$. By assumption, $G(p_j) \geq F_L(p_j) = s_j$, which implies that $G(p_j) - s_j \geq 0$ as desired. In addition,

$$\mathbb{P}[V_i = p_j] = \mathbb{P}[V_i \leq p_j] - \mathbb{P}[V_i < p_j] = G(p_j) - s_j + s_j - \mathbb{P}[V_i < p_j].$$

By assumption,

$$\mathbb{P}[V_i < p_j] = \lim_{p \uparrow p_j} G(p) \leq \lim_{p \uparrow p_j} F_U(p) = s_j,$$

which implies that $s_j - \mathbb{P}[V_i < p_j] \geq 0$, and therefore that $G(p_j) - s_j \leq \mathbb{P}[V_i = p_j]$ as desired. So, for $j \in \{0, \dots, J+1\}$ such that $\mathbb{P}[V_i = p_j] > 0$, let

$$\mathbb{P}[B_i = 1 \mid V_i = p_j] = \frac{G(p_j) - s_j}{\mathbb{P}[V_i = p_j]},$$

and let the conditional distribution of B_i given V_i be otherwise unrestricted. Then the joint distribution of (V_i, B_i) is consistent with our assumptions and observed choice shares, and V_i has marginal distribution G . It follows that the proposed bounds are sharp, that is,

$$\mathcal{F}^* = \{G \in \mathcal{F} : F_L(p) \leq G(p) \leq F_U(p) \text{ for all } p \in \overline{\mathbb{R}}\},$$

as desired. Q.E.D.

Lemma 2. Suppose that Assumptions [RA](#), [TC](#), and [CP](#) are satisfied. For $q \in [p_0, p_{J+1}]$ and $r \in [0, 1]$, let $F_{C,q}$ and $F_{D,r}$ be the compressed and dispersed distributions introduced in Section [3.5](#),

$$F_{C,q}(p) \equiv \begin{cases} F_L(p), & p < q, \\ F_U(p), & q \leq p, \end{cases} \quad \text{and} \quad F_{D,r}(p) \equiv \begin{cases} F_U(p), & p < \underline{q}(r, F_U), \\ r, & \underline{q}(r, F_U) \leq p < \overline{q}(r, F_L), \\ F_L(p), & \overline{q}(r, F_L) \leq p. \end{cases}$$

Then $F_{C,q} \in \mathcal{F}^*$ for every $q \in [p_0, p_{J+1}]$, and $F_{D,r} \in \mathcal{F}^*$ for every $r \in [0, 1]$. Moreover, $q \mapsto m(F_{C,q})$ is continuous, weakly increasing, and piecewise linear on $[p_0, p_{J+1}]$, with $m(F_{C,p_0}) = m(F_U)$ and $m(F_{C,p_{J+1}}) = m(F_L)$.

Proof of Lemma 2. By Proposition [1](#), it suffices to show that $F_{C,q}$ and $F_{D,r}$ are elements of $\mathcal{F}$ that lie between F_L and F_U . The latter property is clear for $F_{C,q}$, which equals F_L or F_U at every p , and for the outer pieces of $F_{D,r}$. On the middle piece of $F_{D,r}$, it follows because $p \geq \underline{q}(r, F_U)$ and right-continuity of F_U give $F_U(p) \geq r$, while $p < \overline{q}(r, F_L)$ gives $F_L(p) \leq r$. So it only remains to show that these two functions are indeed distribution functions (elements of $\mathcal{F}$).

Both $F_{C,q}$ and $F_{D,r}$ are right-continuous, because each piece in their definition is right-continuous and each piece is defined on an interval that is closed on the left. Both distributions are also equal to one at $+\infty$, because $F_L(+\infty) = F_U(+\infty) = 1$. The monotonicity of the compressed distribution, $F_{C,q}$, follows from the monotonicity of F_L and F_U , because $F_L(p) \leq F_U(p) \leq F_U(p')$ whenever $p < q \leq p'$. The monotonicity of the dispersed distribution, $F_{D,r}$, follows for a similar reason, noting that $F_U(p) < r$ for $p < \underline{q}(r, F_U)$ by the definition of $\underline{q}(r, F_U)$, and that $F_L(p) \geq r$ for $p \geq \overline{q}(r, F_L)$, by the definition of $\overline{q}(r, F_L)$ and right-continuity of F_L .

For the mean, let k_q denote the index of the left-inclusive price “bin” in which q

lies,

$$k_q \equiv \mathbb{1}[q = p_{J+1}]J + \sum_{j=1}^J \mathbb{1}[p_j \leq q < p_{j+1}]j.$$

Since $F_{C,q}$ places the mass $s_{j+1} - s_j$ of bin j at p_{j+1} for $j < k_q$, at q for $j = k_q$, and at p_j for $j > k_q$,

$$m(F_{C,q}) = \sum_{j=0}^{k_q-1} (s_{j+1} - s_j)p_{j+1} + (s_{k_q+1} - s_{k_q})q + \sum_{j=k_q+1}^J (s_{j+1} - s_j)p_j.$$

This is linear in q within each bin with slope $s_{k_q+1} - s_{k_q} \geq 0$, continuous across bins, and equal to $m(F_U)$ and $m(F_L)$ at $q = p_0$ and $q = p_{J+1}$; if a choke price is infinite, continuity at that endpoint uses the convention $0 \times \pm\infty \equiv 0$. When both choke prices are finite, these properties also follow from [Stoye \(2010, proof of Theorem 2\(i\)\)](#). Q.E.D.

Proof of Proposition 2. Given the characterization of $\mathcal{F}^*$ from Proposition 1, $m(F_U)$ and $m(F_L)$ are the smallest and largest elements of $\mathcal{M}^*$ by [Manski \(2003, Proposition 1.4\)](#). Lemma 2 implies that $\mathcal{M}^*$ also contains every value between the two bounds because each compressed distribution $F_{C,q}$ lies in $\mathcal{F}^*$ and its mean moves continuously from $m(F_U)$ to $m(F_L)$ as q ranges over $[p_0, p_{J+1}]$.

The proposed expression for $m(F_U)$ follows because a random variable on $\overline{\mathbb{R}}$ distributed according to F_U is discrete and takes value p_j with probability $s_{j+1} - s_j$ for $j \in \{0, \dots, J\}$. Therefore,

$$m(F_U) = \sum_{j=0}^J (s_{j+1} - s_j)p_j.$$

If $p_0 = -\infty$ and $s_1 = 0$, the first summand equals $0 \times -\infty \equiv 0$, by convention. For $m(F_L)$, a random variable on $\overline{\mathbb{R}}$ distributed according to F_L is discrete and takes value p_j with probability $s_j - s_{j-1}$ for $j \in \{1, \dots, J+1\}$, so has

$$m(F_L) = \sum_{j=1}^{J+1} (s_j - s_{j-1})p_j = \sum_{j=0}^J (s_{j+1} - s_j)p_{j+1}.$$

If $p_{J+1} = +\infty$ and $s_J = 1$, then the last summand equals $0 \times +\infty \equiv 0$. Q.E.D.

Proof of Proposition 3. Let $G \in \mathcal{F}^*$ and $r \in [0, 1]$. First, consider the lower bound.

By Proposition 1,

$$\begin{aligned} \{p \in \overline{\mathbb{R}} : G(p) \geq r\} &\subseteq \{p \in \overline{\mathbb{R}} : F_U(p) \geq r\}, \\ \text{so } \underline{q}(r, F_U) &= \inf\{p \in \overline{\mathbb{R}} : F_U(p) \geq r\} \leq \inf\{p \in \overline{\mathbb{R}} : G(p) \geq r\} = \underline{q}(r, G). \end{aligned}$$

Second, consider the upper bound. By Proposition 1,

$$\begin{aligned} \{p \in \overline{\mathbb{R}} : G(p) \leq r \text{ or } p = -\infty\} &\subseteq \{p \in \overline{\mathbb{R}} : F_L(p) \leq r \text{ or } p = -\infty\}, \\ \text{so } \bar{q}(r, G) &= \sup\{p \in \overline{\mathbb{R}} : G(p) \leq r \text{ or } p = -\infty\} \\ &\leq \sup\{p \in \overline{\mathbb{R}} : F_L(p) \leq r \text{ or } p = -\infty\} = \bar{q}(r, F_L). \end{aligned}$$

It follows that the proposed bounds are valid, that is,

$$q(r, G) = [\underline{q}(r, G), \bar{q}(r, G)] \subseteq [\underline{q}(r, F_U), \bar{q}(r, F_L)]$$

for any $G \in \mathcal{F}^*$.

To verify that the proposed bounds are sharp, consider the dispersed distribution $F_{D,r}$ defined in Lemma 2. That result shows that $F_{D,r} \in \mathcal{F}^*$. By construction, we also know that $q(r, F_{D,r}) = [\underline{q}(r, F_U), \bar{q}(r, F_L)]$. It follows that the proposed bounds are sharp, that is,

$$\mathcal{Q}^*(r) = [\underline{q}(r, F_U), \bar{q}(r, F_L)],$$

as claimed. Note that a single distribution $F_{D,r}$ attains both bounds on the r th quantile simultaneously. The proposed expressions for each bound follow from

$$\begin{aligned} \underline{q}(r, F_U) &= \inf\{p \in \overline{\mathbb{R}} : F_U(p) \geq r\} \\ &= \mathbb{1}[r = 0](-\infty) + \mathbb{1}[r > 0] \min\{p_j \in \{p_0, \dots, p_J\} : s_{j+1} \geq r\} \\ &= \mathbb{1}[r = 0](-\infty) + \sum_{j=0}^J \mathbb{1}[s_j < r \leq s_{j+1}] p_j, \\ \text{and } \bar{q}(r, F_L) &= \sup\{p \in \overline{\mathbb{R}} : F_L(p) \leq r \text{ or } p = -\infty\} \\ &= \mathbb{1}[r = 1](+\infty) + \mathbb{1}[r < 1] \max\{p_j \in \{p_1, \dots, p_{J+1}\} : s_{j-1} \leq r\} \\ &= \mathbb{1}[r = 1](+\infty) + \sum_{j=1}^{J+1} \mathbb{1}[s_{j-1} \leq r < s_j] p_j \\ &= \mathbb{1}[r = 1](+\infty) + \sum_{j=0}^J \mathbb{1}[s_j \leq r < s_{j+1}] p_{j+1}, \end{aligned}$$

using the convention that $0 \times \pm\infty \equiv 0$.

Q.E.D.

Proof of Proposition 4. The structure is different for the upper and lower bounds.

(Lower bound) Lemma 2 shows that $F_{C,q} \in \mathcal{F}^*$ for every $q \in [p_0, p_{J+1}]$, and that $m(F_{C,q})$ is a continuous, weakly increasing, and piecewise linear function of q , with $m(F_{C,p_0}) = m(F_U)$ and $m(F_{C,p_{J+1}}) = m(F_L)$. With k_q and the expression for $m(F_{C,q})$ as in the proof of that lemma, the variance of the compressed distribution around q is

$$v(F_{C,q}) = \sum_{j=0}^{k_q-1} (s_{j+1} - s_j) p_{j+1}^2 + (s_{k_q+1} - s_{k_q}) q^2 + \sum_{j=k_q+1}^J (s_{j+1} - s_j) p_j^2 - m(F_{C,q})^2.$$

Furthermore, every $G \in \mathcal{F}^*$ with finite variance satisfies $v(G) \geq v(F_{C,c})$ for $c = m(G)$, so the variance lower bound is the smallest variance among compressed distributions. To see this, let $U \sim \text{Unif}(0, 1)$ and $V \equiv \underline{q}(U, G)$, so that $V \sim G$. Proposition 1 implies that $\lim_{p \uparrow p_j} G(p) \leq s_j \leq G(p_j)$ for each j , so $V \in [p_j, p_{j+1}]$ whenever $U \in (s_j, s_{j+1}]$. Similarly, $V_c \equiv \underline{q}(U, F_{C,c})$ is distributed according to $F_{C,c}$ and equals the projection of c onto $[p_j, p_{j+1}]$ whenever $U \in (s_j, s_{j+1}]$, so $|V - c| \geq |V_c - c|$. Since $c = m(G)$ is finite and the variance is the smallest mean squared deviation about any point,¹⁹

$$v(G) = \mathbb{E}[(V - c)^2] \geq \mathbb{E}[(V_c - c)^2] \geq v(F_{C,c}).$$

Define the function $\varphi : [m(F_U), m(F_L)] \rightarrow \overline{\mathbb{R}}_+$ that maps the means of compressed distributions to their variances,

$$\varphi(m(F_{C,q})) = v(F_{C,q}) \text{ for all } q \in [p_0, p_{J+1}] \text{ such that } v(F_{C,q}) \text{ exists,}$$

with $\varphi(\pm\infty) \equiv +\infty$ whenever $\pm\infty \in \{m(F_U), m(F_L)\}$. Then φ is not only well-defined and continuous, but also convex on its domain. By contrast, $v(F_{C,q})$ is continuous and piecewise quadratic, but not necessarily convex as a function of $q \in [p_0, p_{J+1}]$.

To verify convexity of φ , it suffices to show that its right derivative $\partial_+ \varphi$ exists and is weakly increasing on $[m(F_U), m(F_L)) \cap \mathbb{R}$; convexity then extends to the closed domain because φ is continuous with $\varphi(\pm\infty) = +\infty$. To this end, for $\mu \in [m(F_U), m(F_L))$ let l_μ denote the index of the left-inclusive mean “bin” in which μ lies,

$$l_\mu \equiv \sum_{j=1}^J \mathbb{1}[m(F_{C,p_j}) \leq \mu < m(F_{C,p_{j+1}})] j.$$

¹⁹When both choke prices are finite, this also follows from [Stoye \(2010, Theorem 2\(i\)\)](#); the direct argument is needed because Stoye assumes that the bounding distributions have finite means, which fails when a choke price is infinite.

Then there is a unique $q \in [p_{l_\mu}, p_{l_\mu+1})$ satisfying $m(F_{C,q}) = \mu$, namely

$$q_\mu \equiv \frac{\mu - \sum_{j=0}^{l_\mu-1} (s_{j+1} - s_j)p_{j+1} - \sum_{j=l_\mu+1}^J (s_{j+1} - s_j)p_j}{s_{l_\mu+1} - s_{l_\mu}}.$$

Substituting q_μ into our expression for $v(F_{C,q})$ and taking the right derivative with respect to μ , we get that

$$\partial_+ \varphi(\mu) = 2(q_\mu - \mu).$$

Consider $j \in \{0, \dots, J\}$ with $m(F_{C,p_j}) < m(F_{C,p_{j+1}})$, or equivalently with $s_{j+1} - s_j > 0$. Then $\partial_+ \varphi$ is weakly increasing on $[m(F_{C,p_j}), m(F_{C,p_{j+1}}))$, since the coefficient on μ in the above expression is non-negative,

$$s_{j+1} - s_j \in (0, 1] \implies 2 \left(\frac{1}{s_{j+1} - s_j} - 1 \right) \geq 0.$$

In addition, suppose $m(F_{C,p_j}) > m(F_U)$. Then

$$\lim_{\mu \uparrow m(F_{C,p_j})} \partial_+ \varphi(\mu) = 2(p_{j+1} - m(F_{C,p_j}))$$

where $\underline{j} \equiv \max\{k \in \{0, \dots, j-1\} : s_{k+1} - s_k > 0\}$.

It follows that $\partial_+ \varphi$ may jump up (but not down) at $m(F_{C,p_j})$, that is,

$$\lim_{\mu \uparrow m(F_{C,p_j})} \partial_+ \varphi(\mu) \leq \partial_+ \varphi(m(F_{C,p_j})) = 2(p_j - m(F_{C,p_j})).$$

As a result, $\partial_+ \varphi$ must be weakly increasing on $[m(F_U), m(F_L))$, so φ is convex on $[m(F_U), m(F_L)]$, as desired.

It remains to show that φ is minimized at $m(F_{C,q^*})$ whenever $q^* \in [p_1, p_J]$ satisfies $m(F_{C,q^*}) = q^*$. Such a q^* must exist because $q - m(F_{C,q})$ is continuous and weakly increasing on $[p_1, p_J]$ with $p_1 - m(F_{C,p_1}) \leq 0$ and $p_J - m(F_{C,p_J}) \geq 0$. First, suppose that $m(F_{C,q^*}) \notin \{m(F_{C,p_j})\}_{j=0}^{J+1}$. Then φ is differentiable at $m(F_{C,q^*})$, with

$$\varphi'(m(F_{C,q^*})) = 2(q^* - m(F_{C,q^*})) = 0,$$

so we conclude that q^* is optimal in this case. Now, suppose that $m(F_{C,q^*}) = m(F_{C,p_j})$ for some $j \in \{0, \dots, J+1\}$; if there are multiple such indices j , pick the largest one. Then if $m(F_{C,q^*}) < m(F_L)$, we have that

$$\partial_+ \varphi(m(F_{C,q^*})) = 2(p_j - m(F_{C,q^*})) \geq 2(q^* - m(F_{C,q^*})) = 0,$$

while if $m(F_{C,q^*}) > m(F_U)$, we have that

$$\partial_- \varphi(m(F_{C,q^*})) = \lim_{\mu \uparrow m(F_{C,q^*})} \partial_+ \varphi(\mu) = 2(p_{j+1} - m(F_{C,q^*})) \leq 2(q^* - m(F_{C,q^*})) = 0,$$

where j is defined as before. Both of the above inequalities follow from the fact that $m(F_{C,q^*}) = q^* \implies q^* \in [p_{j+1}, p_J]$. But then q^* must be optimal, because if $m(F_{C,q^*}) \in \{m(F_U), m(F_L)\}$ the relevant Karush–Kuhn–Tucker conditions are satisfied, while if $m(F_{C,q^*}) \in (m(F_U), m(F_L))$ the subdifferential of φ at $m(F_{C,q^*})$ contains zero, as required.

The final claim is that q^* is unique unless $s_k = 0$ and $s_{k+1} = 1$ for some $k \in \{1, \dots, J-1\}$. For $q \in [p_k, p_{k+1})$, the slope of $q - m(F_{C,q})$ in q is $1 - (s_{k+1} - s_k)$, which is strictly positive unless $s_k = 0$ and $s_{k+1} = 1$. If no $k \in \{1, \dots, J-1\}$ has $s_k = 0$ and $s_{k+1} = 1$, then $q - m(F_{C,q})$ is strictly increasing on $[p_1, p_J]$ and has exactly one root there. If there is a k such that $s_k = 0$ and $s_{k+1} = 1$, then all of the mass lies in $[p_k, p_{k+1}]$, every q in that interval solves $m(F_{C,q}) = q$, and the variance lower bound is zero.

(Upper bound) First, we will establish that the upper variance bound is infinite when either choke price is infinite. Suppose $p_{J+1} = +\infty$ and $s_J < 1$ (otherwise we may set p_{J+1} to be finite without loss of generality). Define the following distribution:

$$\bar{F}(p) \equiv \begin{cases} F_L(p), & p < p_J, \\ 1 - (1 - s_J)/(1 + p - p_J)^2, & p_J \leq p. \end{cases}$$

Then $\bar{F} \in \mathcal{F}^*$ with $v(\bar{F}) = +\infty$. To see this, suppose $\bar{V}$ is distributed according to $\bar{F}$. Then conditional on $\bar{V} > p_J$, the random variable $1 + \bar{V} - p_J$ follows a Pareto distribution with scale parameter 1 and shape parameter 2, which has finite mean but infinite variance. Since $\bar{V}$ is bounded conditional on $\bar{V} \leq p_J$ and $\mathbb{P}[\bar{V} > p_J] = 1 - s_J > 0$, it follows that $v(\bar{F}) = +\infty$, as desired. Note that it does not suffice to simply consider F_L instead of $\bar{F}$, since $v(F_L)$ is undefined. Every value in $[v(F_{C,q^*}), +\infty)$ can be attained by a distribution that coincides with F_{C,q^*} except that half of the mass $1 - s_J$ that it places on p_J is moved to $p_J + t$ for an appropriate $t \geq 0$. A symmetric construction establishes that the upper variance bound is also infinite when $p_0 = -\infty$ and $s_1 > 0$.

Now, consider the case in which both choke prices p_0, p_{J+1} are finite. By [Stoye \(2010, Theorem 2\(i\)\)](#), which applies in the case that both choke prices are finite, the variance upper bound is $\sup_{r \in [0,1]} v(F_{D,r})$. By [Lemma 2](#), $F_{D,r} \in \mathcal{F}^*$ for every $r \in [0, 1]$. We will express $m(F_{D,r})$ and $v(F_{D,r})$ as functions of $r \in [0, 1]$ that depend on prices

and choice shares. To this end, let k_r denote the index of the right-inclusive choice share “bin” in which r lies,

$$k_r \equiv \sum_{j=1}^J \mathbb{1}[s_j < r \leq s_{j+1}]j.$$

From the definition of the dispersed distribution around r , we see that

$$\begin{aligned} m(F_{D,r}) &= \sum_{j=0}^{k_r-1} (s_{j+1} - s_j)p_j + (r - s_{k_r})p_{k_r} + (s_{k_r+1} - r)p_{k_r+1} \\ &\quad + \sum_{j=k_r+1}^J (s_{j+1} - s_j)p_{j+1}, \\ v(F_{D,r}) &= \sum_{j=0}^{k_r-1} (s_{j+1} - s_j)p_j^2 + (r - s_{k_r})p_{k_r}^2 + (s_{k_r+1} - r)p_{k_r+1}^2 \\ &\quad + \sum_{j=k_r+1}^J (s_{j+1} - s_j)p_{j+1}^2 - m(F_{D,r})^2. \end{aligned}$$

Observe that $m(F_{D,r})$ is continuous, strictly decreasing (since $p_{k_r} - p_{k_r+1} < 0$ whenever $s_{k_r+1} - s_{k_r} > 0$), and piecewise linear in r , with $m(F_{D,0}) = m(F_L)$ and $m(F_{D,1}) = m(F_U)$. Define the function $\psi : [m(F_U), m(F_L)] \rightarrow \mathbb{R}_+$ that maps means in the identified set to the corresponding variance upper bound,

$$\psi(m(F_{D,r})) = v(F_{D,r}) \text{ for all } r \in [0, 1].$$

We will establish that ψ is strictly concave on its domain by showing that its right derivative $\partial_+ \psi$ exists and is strictly decreasing on $[m(F_U), m(F_L)]$. To this end, for $\mu \in [m(F_U), m(F_L)]$ let l_μ denote the index of the left-inclusive mean “bin” in which μ lies,

$$l_\mu \equiv \sum_{j=1}^J \mathbb{1}[m(F_{D,s_{j+1}}) \leq \mu < m(F_{D,s_j})]j.$$

Then there is a unique $r \in (s_{l_\mu}, s_{l_\mu+1}]$ satisfying $m(F_{D,r}) = \mu$, namely

$$r_\mu \equiv \frac{s_{l_\mu}p_{l_\mu} - s_{l_\mu+1}p_{l_\mu+1} + \mu - \sum_{j=0}^{l_\mu-1} (s_{j+1} - s_j)p_j - \sum_{j=l_\mu+1}^J (s_{j+1} - s_j)p_{j+1}}{p_{l_\mu} - p_{l_\mu+1}}.$$

Substituting r_μ into our expression for $v(F_{D,r})$ and taking the right derivative with

respect to μ , we get that

$$\partial_+ \psi(\mu) = \frac{p_{l_\mu}^2 - p_{l_\mu+1}^2}{p_{l_\mu} - p_{l_\mu+1}} - 2\mu = p_{l_\mu} + p_{l_\mu+1} - 2\mu.$$

Consider $j \in \{0, \dots, J\}$ with $m(F_{D,s_{j+1}}) < m(F_{D,s_j})$. Then $\partial_+ \psi$ is strictly decreasing on $[m(F_{D,s_{j+1}}), m(F_{D,s_j}))$ since the coefficient on μ in the above expression is $-2 < 0$. In addition, suppose $m(F_{D,s_{j+1}}) > m(F_U)$. Then

$$\lim_{\mu \uparrow m(F_{D,s_{j+1}})} \partial_+ \psi(\mu) = p_{\bar{j}} + p_{\bar{j}+1} - 2m(F_{D,s_{j+1}})$$

where $\bar{j} \equiv \min\{k \in \{j+1, \dots, J\} : s_{k+1} - s_k > 0\}$.

It follows that $\partial_+ \psi$ jumps down at $m(F_{D,s_{j+1}})$, that is,

$$\lim_{\mu \uparrow m(F_{D,s_{j+1}})} \partial_+ \psi(\mu) > \partial_+ \psi(m(F_{D,s_{j+1}})) = p_j + p_{j+1} - 2m(F_{D,s_{j+1}}).$$

As a result, $\partial_+ \psi$ must be strictly decreasing on $[m(F_U), m(F_L))$, so ψ is strictly concave on $[m(F_U), m(F_L)]$, as desired.

It remains to show that ψ is maximized at $m(F_{D,r^*})$ whenever $r^* \in (0, 1)$ satisfies either $m(F_{D,r^*}) = \bar{q}(r^*, F_M)$ or $F_M(m(F_{D,r^*})) = r^*$. Such an r^* must exist since $m(F_{D,r})$ is continuous and strictly decreasing on $[0, 1]$, with $m(F_{D,0}) > \bar{q}(0, F_M)$ and $m(F_{D,1}) < \underline{q}(1, F_M)$. First, suppose that $m(F_{D,r^*}) = \bar{q}(r^*, F_M)$. Following the same argument as in Proposition 3, we may write

$$\bar{q}(r, F_M) = \mathbb{1}[r = 1](+\infty) + \sum_{j=0}^J \mathbb{1}[s_j \leq r < s_{j+1}] \frac{p_j + p_{j+1}}{2}.$$

So, $r^* \in (0, 1)$ satisfies $p_j + p_{j+1} - 2m(F_{D,r^*}) = 0$, where $j \in \{0, \dots, J\}$ is the index of the left-inclusive choice share “bin” in which r^* lies. This implies that the left derivative of ψ at $m(F_{D,r^*})$ is equal to zero,

$$\partial_- \psi(m(F_{D,r^*})) = \lim_{\mu \uparrow m(F_{D,r^*})} \partial_+ \psi(\mu) = 0,$$

so we may conclude that r^* is optimal in this case. Now, suppose instead that $F_M(m(F_{D,r^*})) = r^*$. Then $r^* = s_{j+1} \in (0, 1)$ for some $j \in \{0, \dots, J-1\}$; if there

are multiple such indices j , pick the smallest one. Then

$$\begin{aligned} \partial_+ \psi(m(F_{D,r^*})) &= p_j + p_{j+1} - 2m(F_{D,r^*}) \leq 0 \\ \text{and } \partial_- \psi(m(F_{D,r^*})) &= p_{\bar{j}} + p_{\bar{j}+1} - 2m(F_{D,r^*}) > 0 \end{aligned}$$

where $\bar{j}$ is defined as before, and the inequalities follow from the definition of F_M . So r^* is optimal, since the subdifferential of $-\psi$ at $m(F_{D,r^*})$ contains zero, as required.

The claim that r^* is unique follows because ψ is strictly concave, so it has a unique maximizer μ^* . Because $m(F_{D,r})$ is strictly decreasing in r , exactly one $r \in [0, 1]$ satisfies $m(F_{D,r}) = \mu^*$. We have shown that every r^* satisfying either condition in the statement maximizes ψ at $m(F_{D,r^*})$, so there is only one such r^* . Q.E.D.

Proof of Corollary 1. The claim is that

$$[p_j, p_{j+1}] \subseteq \mathcal{Q}^*(r) \iff s_j - r \leq 0 \quad \text{and} \quad r - s_{j+1} \leq 0.$$

By Proposition 3, we have that

$$[p_j, p_{j+1}] \subseteq \mathcal{Q}^*(r) \iff \underline{q}(r, F_U) \leq p_j \quad \text{and} \quad p_{j+1} \leq \bar{q}(r, F_L).$$

We first establish that $\underline{q}(r, F_U) \leq p_j \iff s_{j+1} \geq r$. Recall from the proof of Proposition 3 that

$$\underline{q}(r, F_U) = \mathbb{1}[r = 0](-\infty) + \mathbb{1}[r > 0] \min\{p_k \in \{p_0, \dots, p_J\} : s_{k+1} \geq r\},$$

which shows that the $\Leftarrow$ direction follows immediately. Conversely, suppose that $\underline{q}(r, F_U) \leq p_j$. Then either $r = 0$, in which case $s_{j+1} \geq r$ holds trivially, or $r > 0$ so there exists $k \in \{0, \dots, J\}$ with $p_k \leq p_j$ and $s_{k+1} \geq r$. It follows that $s_{j+1} \geq r$, because choice shares must be monotonically increasing.

The same argument, applied to the expression for $\bar{q}(r, F_L)$ in the proof of Proposition 3, shows that $p_{j+1} \leq \bar{q}(r, F_L) \iff s_j \leq r$. Q.E.D.

Proof of Proposition 5. We first record some properties of the isotonic regression estimator that are used in the proof. Let $T_{\mathcal{S}}(\mathbf{s})$ denote the tangent cone of $\mathcal{S}$ at $\mathbf{s}$. Because $\mathcal{S}$ is closed and convex, $\Pi_{\mathcal{S}}$ is Hadamard directionally differentiable at every $\mathbf{s} \in \mathcal{S}$, and its directional derivative is the projection $\Pi_{T_{\mathcal{S}}(\mathbf{s})}$ onto $T_{\mathcal{S}}(\mathbf{s})$ (see, for example, Fang and Santos, 2019, Section 4). Since $\Pi_{\mathcal{S}}(\mathbf{s}) = \mathbf{s}$ and $\sqrt{n}(\hat{\mathbf{s}}_n - \mathbf{s}) \rightarrow_d$

$\mathcal{N}(0, \Gamma_0)$, Fang and Santos (2019, Theorem 2.1) implies that

$$\sqrt{n}(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n) - \mathbf{s}) \xrightarrow{d} \Pi_{T_{\mathcal{S}}(\mathbf{s})}(\mathcal{N}(0, \Gamma_0)),$$

a limit supported on $T_{\mathcal{S}}(\mathbf{s})$. In particular, $\sqrt{n}(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n) - \mathbf{s}) = O_p(1)$ and $\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n) \rightarrow_p \mathbf{s}$. Since $\mathbf{s} \notin \text{vert}(\mathcal{S})$, it follows that $\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n) \notin \text{vert}(\mathcal{S})$ with probability approaching one, so that $\hat{\nu}_n(\hat{\mathbf{s}}_n)$ is well-defined with probability approaching one.

Now we proceed to analyze the estimator (8) by first centering it around $\nu(\mathbf{s})$, adding and subtracting $\mathbf{s}$, then scaling by $\sqrt{n}$ to obtain

$$\begin{aligned} \sqrt{n}(\hat{\nu}_n(\hat{\mathbf{s}}_n) - \nu(\mathbf{s})) &= \nabla \nu(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n)) \sqrt{n}(\hat{\mathbf{s}}_n - \mathbf{s}) + r_n \\ \text{where } r_n &\equiv \sqrt{n}(\nu(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n)) - \nu(\mathbf{s}) - \nabla \nu(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n))(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n) - \mathbf{s})). \end{aligned}$$

Since $\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n) \rightarrow_p \mathbf{s}$ and $\nabla \nu$ is continuous at $\mathbf{s} \in \mathcal{S} \setminus \text{vert}(\mathcal{S})$, we have that $\nabla \nu(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n))$ is consistent for $\nabla \nu(\mathbf{s})$ by the continuous mapping theorem. By Slutsky's theorem, we conclude that the first term converges to the desired limiting distribution,

$$\nabla \nu(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n)) \sqrt{n}(\hat{\mathbf{s}}_n - \mathbf{s}) \xrightarrow{d} \mathcal{N}(0, \nabla \nu(\mathbf{s}) \Gamma_0 \nabla \nu(\mathbf{s})').$$

It remains to show that $r_n \rightarrow_p 0$, which delivers the conclusion from Slutsky's theorem. From Lemma 1, we know that because $\mathbf{s} \in \mathcal{S} \setminus \text{vert}(\mathcal{S})$, ν is Hadamard directionally differentiable at $\mathbf{s}$ tangentially to $T_{\mathcal{S}}(\mathbf{s})$. That result also shows that its directional derivative is given by the linear map $\nabla \nu(\mathbf{s})$. Applying Fang and Santos (2019, Theorem 2.1) again, now to ν , we have that

$$\sqrt{n}(\nu(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n)) - \nu(\mathbf{s})) - \nabla \nu(\mathbf{s}) \sqrt{n}(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n) - \mathbf{s}) \xrightarrow{p} 0,$$

since $\sqrt{n}(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n) - \mathbf{s})$ converges in distribution to a limit supported on $T_{\mathcal{S}}(\mathbf{s})$, as shown above. Adding and subtracting to r_n then applying this observation gives

$$r_n = (\nabla \nu(\mathbf{s}) - \nabla \nu(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n))) \sqrt{n}(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n) - \mathbf{s}) + o_p(1) = o_p(1) O_p(1) + o_p(1) = o_p(1),$$

because $\nabla \nu(\mathbf{s}) - \nabla \nu(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n))$ is $o_p(1)$, as previously discussed, and $\sqrt{n}(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n) - \mathbf{s})$ is $O_p(1)$.

That $\nabla \nu(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n)) \hat{\Gamma}_n \nabla \nu(\Pi_{\mathcal{S}}(\hat{\mathbf{s}}_n))'$ is a consistent estimator of the asymptotic covariance matrix follows from the continuous mapping theorem and the consistency of $\hat{\Gamma}_n$ (Section 4). Q.E.D.

Lemma 3. For any $\mathbf{z} \in \mathbb{R}^J$, let $T(j) \equiv \max\{z_1, \dots, z_j, -z_{j+1}, \dots, -z_J\}$ and $k \in$

$\arg \min_{0 \leq j \leq J} T(j)$. Then T is weakly decreasing on $\{0, \dots, k\}$ and weakly increasing on $\{k, \dots, J\}$. In particular, this holds for $T(j) = T_{nj}$, the test statistic in Section 6.3.

Proof of Lemma 3. For $j \in \{0, \dots, J\}$, let

$$T(j) \equiv \max\{z_1, \dots, z_j, -z_{j+1}, \dots, -z_J\},$$

where $z \in \mathbb{R}^J$ is an arbitrary vector, and let $k \in \arg \min_{0 \leq j \leq J} T(j)$. We will show that T is weakly increasing on $\{k, \dots, J\}$.

Let $l \in \{k, \dots, J\}$ be the first index such that $T(l) \neq T(k)$; if no such l exists, we are done. Then it must be the case that $T(l) > T(k)$, since $T(k)$ is minimal by assumption. It follows that $T(l) = z_l$, since all other components also appear in the expression for $T(l-1) = T(k)$. Now, let $m \in \{l, \dots, J\}$ be the first index such that $T(m) \neq T(l)$; if no such m exists, we are done. Then it must be the case that $T(m) > T(l)$, since the expression for $T(m)$ includes the component $z_l = T(l)$. It follows that $T(m) = z_m$, by the same reasoning as before. The desired conclusion follows by induction.

A symmetric argument establishes that T is weakly decreasing on $\{0, \dots, k\}$. Setting $z = \text{diag}\{\hat{\varsigma}_n\}^{-1} \sqrt{n}(\hat{\mathbf{s}}_n - r)$ gives $T(j) = T_{nj}$, so the above argument applies to the test statistic proposed in Section 6.3. Q.E.D.

Proof of Proposition 6. The first-step upper confidence rectangle for $\text{diag}\{d_j\}(\mathbf{s} - r)$ is given by

$$M_{nj} \equiv \{\mu \in \mathbb{R}^J : \mu \leq \text{diag}\{d_j\}(\hat{\mathbf{s}}_n - r) + e\hat{\varsigma}_n/\sqrt{n}\}.$$

Let the “feasible set” denote the set of $\mu \in M_{nj}$ consistent with both the null hypothesis $H_{0j}(r)$ and the restriction that $\mathbf{s} \in \mathcal{S}$. Since $\text{diag}\{d_j\}$ is its own inverse, a vector μ is feasible if and only if $\mu = \text{diag}\{d_j\}(\tilde{\mathbf{s}} - r)$ for some $\tilde{\mathbf{s}} \in \mathcal{S}$ such that

$$\tilde{s}_k \leq \min\{r, \hat{s}_{k,n} + e\hat{\varsigma}_{k,n}/\sqrt{n}\} \text{ for all } k \in \{1, \dots, j\}, \quad \text{and} \quad (10)$$

$$\tilde{s}_k \geq \max\{r, \hat{s}_{k,n} - e\hat{\varsigma}_{k,n}/\sqrt{n}\} \text{ for all } k \in \{j+1, \dots, J\}. \quad (11)$$

Romano et al. (2014) exploit the fact that the critical value $\hat{c}_{nj}$ is weakly increasing in each component of μ to arrive at a simple closed-form expression for the least favorable value μ_{nj}^* . This requires a feasible μ_{nj}^* such that $\mu \leq \mu_{nj}^*$ for all μ in the feasible set.

We claim that such a μ_{nj}^* exists in the current setting as well, and is given by

$$\begin{aligned} [\mu_{nj}^*]_k &= \min_{l \in \{k, \dots, j\}} \{\min\{r, \hat{s}_{l,n} + e\hat{\varsigma}_{l,n}/\sqrt{n}\}\} - r \text{ for all } k \in \{1, \dots, j\}, \quad \text{and} \\ [\mu_{nj}^*]_k &= r - \max_{l \in \{j+1, \dots, k\}} \{\max\{r, \hat{s}_{l,n} - e\hat{\varsigma}_{l,n}/\sqrt{n}\}\} \text{ for all } k \in \{j+1, \dots, J\}. \end{aligned}$$

Consider the corresponding $\mathbf{s}_{nj}^* \equiv \text{diag}\{d_j\}\mu_{nj}^* + r$. Then $\mathbf{s}_{nj}^*$ satisfies (10) and (11) by construction. We also have that $\mathbf{s}_{nj}^* \in \mathcal{S}$, because the cumulative minimum over fewer and fewer elements must weakly increase, the cumulative maximum over more and more elements must also weakly increase, and $[\mathbf{s}_{nj}^*]_j \leq r \leq [\mathbf{s}_{nj}^*]_{j+1}$, with all components in $[0, 1]$ since r and each $\hat{s}_{l,n}$ lie in $[0, 1]$. So, the proposed μ_{nj}^* is indeed feasible.

Now, suppose for the sake of contradiction that there exists a feasible $\mu \in \mathbb{R}^J$ with $\mu_k > [\mu_{nj}^*]_k$ for some $k \in \{1, \dots, J\}$, and let $\tilde{\mathbf{s}} \equiv \text{diag}\{d_j\}\mu + r$. Suppose $k \in \{1, \dots, j\}$. Then we must have $\tilde{s}_k > \min\{r, \hat{s}_{l,n} + e\hat{\varsigma}_{l,n}/\sqrt{n}\}$ for some $l \in \{k, \dots, j\}$. The right-hand side of this inequality must be an upper bound for $\tilde{s}_l$ by (10), but $\tilde{s}_l \geq \tilde{s}_k$ since $\tilde{\mathbf{s}} \in \mathcal{S}$, so we have reached a contradiction. Suppose instead that $k \in \{j+1, \dots, J\}$. Then we must have $\tilde{s}_k < \max\{r, \hat{s}_{l,n} - e\hat{\varsigma}_{l,n}/\sqrt{n}\}$ for some $l \in \{j+1, \dots, k\}$. The right-hand side of this inequality must be a lower bound for $\tilde{s}_l$ by (11), but $\tilde{s}_l \leq \tilde{s}_k$ since $\tilde{\mathbf{s}} \in \mathcal{S}$, so we have reached a contradiction. We conclude that $\mu \leq \mu_{nj}^*$ for all μ in the feasible set. Arguing as in Romano et al. (2014), then, μ_{nj}^* is least favorable in the sense that it yields the largest possible critical value $\hat{c}_{nj}$ among all values μ in the feasible set.

Q.E.D.

References

- ATHEY, S. AND G. W. IMBENS (2017): “The Econometrics of Randomized Experiments,” in *Handbook of Economic Field Experiments*, ed. by A. V. Banerjee and E. Duflo, Elsevier, vol. 1, chap. 3, 73–140.
- AYER, M., H. D. BRUNK, G. M. EWING, W. T. REID, AND E. SILVERMAN (1955): “An empirical distribution function for sampling with incomplete information,” *The Annals of Mathematical Statistics*, 26, 641–647.
- BERRIGAN, P. AND E. ZUCHELLI (2022): “Public Preferences for Safe Consumption Sites for Opioid Use: A Discrete Choice Experiment,” *Drug and Alcohol Dependence*, 238, 109578.
- BHATTACHARYA, D. (2015): “Nonparametric Welfare Analysis for Discrete Choice,” *Econometrica*, 83, 617–649.
- (2021): “The Empirical Content of Binary Choice Models,” *Econometrica*, 89, 457–474.
- BISHOP, R. C. AND T. A. HEBERLEIN (1979): “Measuring Values of Extramarket Goods: Are Indirect Measures Biased?” *American Journal of Agricultural Economics*, 61, 926–930.
- BOMAN, M., G. BOSTEDT, AND B. KRISTRÖM (1999): “Obtaining Welfare Bounds in Discrete-Response Valuation Studies: A Non-Parametric Approach,” *Land Economics*, 75, 284–294.
- BOYERS, D., M. VAN DER POL, V. WATSON, T. LAMONT, B. GOULAO, C. RAMSAY, A. DUNCAN, L. MACPHERSON, AND J. CLARKSON (2021): “The Value of Preventative Dental Care: A Discrete-Choice Experiment,” *Journal of Dental Research*, 100, 723–730.
- BURTLESS, G. (1995): “The Case for Randomized Field Trials in Economic and Policy Research,” *Journal of Economic Perspectives*, 9, 63–84.
- CAMERON, T. A. (1988): “A new paradigm for valuing non-market goods using referendum data: Maximum likelihood estimation by censored logistic regression,” *Journal of Environmental Economics and Management*, 15, 355–379.
- CAMERON, T. A. AND M. D. JAMES (1987): “Efficient Estimation Methods for “Closed-Ended” Contingent Valuation Surveys,” *The Review of Economics and Statistics*, 69, 269–276.
- CARSON, R. T., T. GROVES, AND J. A. LIST (2014): “Consequentiality: A Theoretical and Experimental Exploration of a Single Binary Choice,” *Journal of the Association of Environmental and Resource Economists*, 1, 171–207.
- CHESHER, A. AND A. M. ROSEN (2013): “What Do Instrumental Variable Models Deliver with Discrete Dependent Variables?” *American Economic Review*, 103, 557–62.
- CHETVERIKOV, D., A. SANTOS, AND A. M. SHAIKH (2018): “The Econometrics of Shape Restrictions,” *Annual Review of Economics*, 10, 31–63.
- COSSLETT, S. R. (1983): “Distribution-Free Maximum Likelihood Estimator of the Binary Choice Model,” *Econometrica*, 51, 765–782.
- CROOKER, J. R. AND J. A. HERRIGES (2004): “Parametric and Semi-Nonparametric Estimation of Willingness-to-Pay in the Dichotomous Choice Contingent Valuation Framework,” *Environmental and Resource Economics*, 27, 451–480.
- DE BEKKER-GROB, E. W., M. RYAN, AND K. GERARD (2012): “Discrete choice experiments in health economics: a review of the literature,” *Health Economics*, 21, 145–172.
- DRAKE, M., N. THAKRAL, AND L. T. TÔ (2025): “Wage Differentials and the Price of Workplace Flexibility,” January 2025 version.
- DUFFIELD, J. W. AND D. A. PATTERSON (1991): “Inference and Optimal Design for a Welfare Measure in Dichotomous Choice Contingent Valuation,” *Land Economics*, 67, 225–239.
- EHM, W., T. GNEITING, A. JORDAN, AND F. KRÜGER (2016): “Of Quantiles and Expectiles: Consistent Scoring Functions, Choquet Representations and Forecast Rankings,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78, 505–562.
- FANG, Z. AND A. SANTOS (2019): “Inference on Directionally Differentiable Functions,” *The*

- Review of Economic Studies*, 86, 377–412.
- FIRPO, S. AND G. RIDDER (2019): “Partial identification of the treatment effect distribution and its functionals,” *Journal of Econometrics*, 213, 210–234.
- FREEDMAN, D. A. (2008): “Randomization Does Not Justify Logistic Regression,” *Statistical Science*, 23, 237–249.
- GREEN, D., K. E. JACOWITZ, D. KAHNEMAN, AND D. MCFADDEN (1998): “Referendum contingent valuation, anchoring, and willingness to pay for public goods,” *Resource and Energy Economics*, 20, 85–116.
- GÜNEY, O. I. AND L. GIRALDO (2019): “Consumers’ attitudes and willingness to pay for organic eggs: A discrete choice experiment study in Turkey,” *British Food Journal*, 122, 678–692.
- HAAB, T. C. AND K. E. MCCONNELL (1997): “Referendum Models and Negative Willingness to Pay: Alternative Solutions,” *Journal of Environmental Economics and Management*, 32, 251–270.
- (2002): *Valuing Environmental and Natural Resources: The Econometrics of Non-Market Valuation*, New Horizons in Environmental Economics, Cheltenham, UK and Northampton, MA: Edward Elgar.
- HANEMANN, M. AND B. KANNINEN (2001): “The Statistical Analysis of Discrete-Response CV Data,” in *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries*, ed. by I. J. Bateman and K. G. Willis, Oxford University Press, 302–441.
- HANEMANN, W. M. (1984): “Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses,” *American Journal of Agricultural Economics*, 66, 332–341.
- HAUSMAN, J. (2012): “Contingent Valuation: From Dubious to Hopeless,” *Journal of Economic Perspectives*, 26, 43–56.
- HAUSMAN, J., J. ABREVAYA, AND F. SCOTT-MORTON (1998): “Misclassification of the dependent variable in a discrete-response setting,” *Journal of Econometrics*, 87, 239–269.
- HE, H., D. NEUMARK, AND Q. WENG (2021): “Do Workers Value Flexible Jobs? A Field Experiment,” *Journal of Labor Economics*, 39, 709–738.
- HOROWITZ, J. L. (1992): “A Smoothed Maximum Score Estimator for the Binary Response Model,” *Econometrica*, 60, 505–531.
- HOYOS, D., P. MARIEL, AND J. FERNÁNDEZ-MACHO (2009): “The influence of cultural identity on the WTP to protect natural resources: Some empirical evidence,” *Ecological Economics*, 68, 2372–2381.
- IMBENS, G. W. AND C. F. MANSKI (2004): “Confidence Intervals for Partially Identified Parameters,” *Econometrica*, 72, 1845–1857.
- JOHNSTON, R. J., K. J. BOYLE, W. V. ADAMOWICZ, J. BENNETT, R. BROUWER, T. A. CAMERON, W. M. HANEMANN, N. HANLEY, M. RYAN, R. SCARPA, R. TOURANGEAU, AND C. A. VOSSLER (2017): “Contemporary Guidance for Stated Preference Studies,” *Journal of the Association of Environmental and Resource Economists*, 4, 319–405.
- KAMAT, V. AND S. NORRIS (2026): “Estimating Welfare Effects in a Non-Parametric Choice Model: The Case of School Vouchers,” *Review of Economic Studies*, 93, 1963–1994.
- KOMAROVA, T. (2013): “Binary choice models with discrete regressors: Identification and misspecification,” *Journal of Econometrics*, 177, 14–33.
- KREINOVICH, V., G. XIANG, AND S. FERSON (2006): “Computing mean and variance under Dempster–Shafer uncertainty: Towards faster algorithms,” *International Journal of Approximate Reasoning*, 42, 212–227.
- KRISTRÖM, B. (1990): “A Non-Parametric Approach to the Estimation of Welfare Measures in Discrete Response Valuation Studies,” *Land Economics*, 66, 135–139.
- LALONDE, R. J. (1986): “Evaluating the Econometric Evaluations of Training Programs with Experimental Data,” *The American Economic Review*, 76, 604–620.

- LANGEWISCH, A. AND F. CHOUBINEH (2004): “Mean and Variance Bounds and Propagation for Ill-Specified Random Variables,” *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 34, 494–506.
- LEWBEL, A., D. MCFADDEN, AND O. LINTON (2011): “Estimating features of a distribution from binomial data,” *Journal of Econometrics*, 162, 170–188.
- MAESTAS, N., K. J. MULLEN, D. POWELL, T. VON WACHTER, AND J. B. WENGER (2023): “The Value of Working Conditions in the United States and Implications for the Structure of Wages,” *American Economic Review*, 113, 2007–2047.
- MAGNAC, T. AND É. MAURIN (2008): “Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data,” *Review of Economic Studies*, 75, 835–864.
- MAHMUD, M., I. A. GUTIERREZ, K. B. KUMAR, AND S. NATARAJ (2021): “What Aspects of Formality Do Workers Value? Evidence from a Choice Experiment in Bangladesh,” *The World Bank Economic Review*, 35, 303–327.
- MANSKI, C. F. (1975): “Maximum score estimation of the stochastic utility model of choice,” *Journal of Econometrics*, 3, 205–228.
- (1985): “Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator,” *Journal of Econometrics*, 27, 313–333.
- (1988): “Identification of Binary Response Models,” *Journal of the American Statistical Association*, 83, 729–738.
- (2003): *Partial identification of probability distributions*, Springer.
- (2008): “Partial Identification in Econometrics,” in *The New Palgrave Dictionary of Economics*, Palgrave Macmillan UK, 1–9.
- MANSKI, C. F. AND E. TAMER (2002): “Inference on Regressions with Interval Data on a Regressor or Outcome,” *Econometrica*, 70, 519–546.
- MAS, A. AND A. PALLAIS (2017): “Valuing Alternative Work Arrangements,” *American Economic Review*, 107, 3722–3759.
- (2019): “Labor Supply and the Value of Non-Work Time: Experimental Estimates from the Field,” *American Economic Review: Insights*, 1, 111–126.
- MCFADDEN, D. (1994): “Contingent Valuation and Social Choice,” *American Journal of Agricultural Economics*, 76, 689–708.
- OYAMA, D. AND T. TAKENAWA (2018): “On the (Non-)Differentiability of the Optimal Value Function When the Optimal Solution Is Unique,” *Journal of Mathematical Economics*, 76, 21–32.
- PARZEN, E. (1979): “Nonparametric Statistical Data Modeling,” *Journal of the American Statistical Association*, 74, 105–121.
- READY, R. C. AND D. HU (1995): “Statistical Approaches to the Fat Tail Problem for Dichotomous Choice Contingent Valuation,” *Land Economics*, 71, 491–499.
- ROMANO, J. P., A. M. SHAIKH, AND M. WOLF (2014): “A Practical Two-Step Method for Testing Moment Inequalities,” *Econometrica*, 82, 1979–2002.
- STOYE, J. (2009): “More on Confidence Intervals for Partially Identified Parameters,” *Econometrica*, 77, 1299–1315.
- (2010): “Partial identification of spread parameters,” *Quantitative Economics*, 1, 323–357.
- (2021): “A simple, short, but never-empty confidence interval for partially identified parameters,” Tech. rep., Cemmap.
- TEBALDI, P., A. TORGOVITSKY, AND H. YANG (2023): “Nonparametric Estimates of Demand in the California Health Insurance Exchange,” *Econometrica*, 91, 107–146.
- THAKRAL, N. AND L. T. TÔ (2026): “Pooling Noisy Choices,” .
- TURNBULL, B. W. (1976): “The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 38, 290–295.

- WATANABE, M. (2010): “Nonparametric Estimation of Mean Willingness to Pay from Discrete Response Valuation Data,” *American Journal of Agricultural Economics*, 92, 1114–1135.
- WATANABE, M. AND K. ASANO (2009): “Distribution Free Consistent Estimation of Mean WTP in Dichotomous Choice Contingent Valuation,” *Environmental and Resource Economics*, 44, 1–10.
- WISWALL, M. AND B. ZAFAR (2018): “Preference for the Workplace, Investment in Human Capital, and Gender,” *The Quarterly Journal of Economics*, 133, 457–507.

Supplemental Appendix

Contents

SA.1 Proof of Lemma 1	S2
SA.2 Monte Carlo Simulations	S9
SA.3 Additional Empirical Results	S10

SA.1 Proof of Lemma 1

Proof of Lemma 1. Suppose that both choke prices are finite, so that the population bounds are non-trivial. Denote the function that maps choice shares to bounds on the variance of WTP by $\nu : \mathcal{S} \rightarrow \mathbb{R}_+^2$, so that

$$\nu(\mathbf{s}) = \begin{bmatrix} v(F_{C,q^*}) \\ v(F_{D,r^*}) \end{bmatrix}$$

where $v(F_{C,q^*})$ and $v(F_{D,r^*})$ are defined as in Proposition 4. The proof of Lemma 1 is somewhat different for the lower and upper bounds.

(Lower bound) To prove that the lower variance bound is continuously differentiable in $\mathbf{s}$, we will provide an alternative characterization of the bound as the optimal value of a weakly convex quadratic program. Define the $(J+1) \times J$ -dimensional matrix

$$D \equiv \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix}',$$

along with the following vectors in $\mathbb{R}^{J+1}$,

$$\underline{p} \equiv [p_0 \quad \cdots \quad p_J]', \quad \bar{p} \equiv [p_1 \quad \cdots \quad p_{J+1}]', \quad e \equiv [0 \quad \cdots \quad 0 \quad 1]', \quad f \equiv D\mathbf{s} + e.$$

Observe that $f \geq 0$ with $\mathbf{1}'f = 1$, so f lies in the J -simplex

$$\Delta^J \equiv \{f \in \mathbb{R}^{J+1} : f \geq 0, \mathbf{1}'f = 1\}.$$

It will be convenient to parameterize the program in terms of f , rather than $\mathbf{s}$. In particular, the variance lower bound can be expressed as

$$\begin{aligned} v(F_{C,q^*}) &= \min_{\eta \in [0,1]^{J+1}} p'(\text{diag}\{f\} - ff')p \\ \text{s.t.} \quad p &= \underline{p} + \text{diag}\{\bar{p} - \underline{p}\}\eta. \end{aligned} \tag{12}$$

To see this, notice that the objective of the above program corresponds to the variance of a discrete random variable supported on p with probability mass function f , where $\underline{p} \leq p \leq \bar{p}$. By construction, the set of all such distributions is a subset of $\mathcal{F}^*$, but a superset of the set of all compressed distributions $\{F_{C,q} : q \in [p_0, p_{J+1}]\}$. So by [Stoye](#)

(2010), the optimized objective must equal $v(F_{C,q^*})$; that is, the optimal compressed distribution F_{C,q^*} corresponds to a discrete random variable supported on p^* with probability mass function f , where p^* is optimal for the above program.

Suppose that $f > 0$. In this case, we will show that the solution p^* is unique. First, notice that the matrix $\text{diag}\{f\} - ff'$ is positive semidefinite, and that its null space is spanned by the constant vector $\mathbf{1} \in \mathbb{R}^{J+1}$. It follows that p^* must be unique up to a constant. Now, suppose for the sake of contradiction that $p^* + c$ is feasible for some $c > 0$. Then we must have $p^* < \bar{p}$. Consider the gradient of the objective with respect to p^* ,

$$2p^{*'}(\text{diag}\{f\} - ff') = 2(p^* - f'p^*)' \text{diag}\{f\}.$$

Since p^* is optimal by assumption, it must be the case that the above gradient is non-negative; otherwise, we could decrease the objective by increasing some component of p^* , which is feasible since $p^* < \bar{p}$. But this implies that $p^* \geq f'p^*$, and therefore that p^* is a constant vector, because $f > 0$. Note that this is infeasible if $J \geq 2$, since $p_1 < p_2$ by assumption. If $J = 1$, then we must have $p^* = [p_1, p_1]'$, which contradicts the fact that $p^* < \bar{p}$. A symmetric argument applies for $c < 0$, so we have reached a contradiction, as desired. We conclude that p^* is unique whenever $f > 0$, and f has at least two components.

Following Oyama and Takenawa (2018), we can now apply Danskin's theorem to conclude that the value function of the program given by (12), call it $\nu_l(f)$, is in fact differentiable whenever $f > 0$, with derivative given by the usual envelope formula. To verify the conditions of the theorem, observe that the set $[0, 1]^{J+1}$ is compact, the objective is differentiable in η and f , and the partial derivative of the objective with respect to f is continuous in η and f . Since p^* is unique, we have that

$$\nabla \nu_l(f) = p^{*'} \text{diag}\{p^* - 2f'p^*\}.$$

So, we have shown that ν_l is differentiable on $\text{int}(\Delta^J)$. We must now make precise the sense in which ν_l is also differentiable on $\text{bd}(\Delta^J) \setminus \text{vert}(\Delta^J)$, where $\text{vert}(\Delta^J)$ denotes the vertices of the simplex.

Suppose some subvector of f , call it $f_{\mathcal{J}}$, is equal to zero, while the remaining components $f_{-\mathcal{J}}$ are strictly positive. Then so long as $f_{-\mathcal{J}}$ contains at least two components, we may apply our previous argument to conclude that $p_{-\mathcal{J}}^*$ is unique. On the other hand, the subvector $p_{\mathcal{J}}^*$ must satisfy $\underline{p}_{\mathcal{J}} \leq p_{\mathcal{J}}^* \leq \bar{p}_{\mathcal{J}}$, but is otherwise unrestricted. By Danskin's theorem, the Hadamard directional derivative of ν_l with

respect to f in the direction h is given by

$$\begin{aligned}\nabla_h \nu_l(f) &= \min_p p' \operatorname{diag}\{p - 2f'p\}h \\ \text{s.t. } \underline{p}_{\mathcal{J}} &\leq p_{\mathcal{J}} \leq \bar{p}_{\mathcal{J}} \quad \text{and} \quad p_{-\mathcal{J}} = p_{-\mathcal{J}}^*.\end{aligned}$$

Note that it is important that we restrict attention to directions h that lie in the tangent cone of Δ^J at f . This is because the theoretical object that ν_l represents is not well-defined outside of Δ^J . As a result, the behavior of ν_l outside of Δ^J depends on the particular representation of the variance lower bound that we have chosen in (12), which is analytically convenient but otherwise arbitrary. It suffices to consider directions h with $h_{\mathcal{J}} \geq 0$, in which case the above optimization problem can be solved coordinatewise for each $j \in \mathcal{J}$ to get

$$p_{\mathcal{J}}^* \equiv \max\{\underline{p}_{\mathcal{J}}, \min\{\bar{p}_{\mathcal{J}}, f'p^*\}\}, \quad (13)$$

keeping in mind that the quantity $f'p^* = f'_{-\mathcal{J}}p_{-\mathcal{J}}^*$ does not depend on $p_{\mathcal{J}}^*$. Notice that the above solution does not depend on the choice of direction h . So, $\nu_l(f)$ is differentiable on the boundary of the simplex Δ^J in the sense that the Hadamard directional derivative of ν_l at f is given by a linear map when restricted to directions h that lie in the tangent cone of Δ^J at f .

All that remains is to consider the case in which $f_{-\mathcal{J}}$ only contains a single component, that is, f is a vertex of Δ^J . Then $p_{-\mathcal{J}}^*$ is not unique, and ν_l need not be differentiable at f in the sense given in the previous paragraph.²⁰ Since $\mathbf{s} \mapsto D\mathbf{s} + e$ is an affine bijection from $\mathcal{S}$ onto Δ^J that maps $\operatorname{vert}(\mathcal{S})$ onto $\operatorname{vert}(\Delta^J)$, these are exactly the points excluded in the statement of Lemma 1.

²⁰For example, let $J = 2$ and $f = [0, 1, 0]'$, so that $\nu_l(f) = 0$. Every direction in the tangent cone of Δ^J at f has the form $h = [h_1, -(h_1 + h_3), h_3]'$ with $h_1, h_3 \geq 0$. Suppose they are not both zero. For small $t > 0$, $\nu_l(f + th)$ is the smallest variance of a random variable that places mass th_1 , $1 - t(h_1 + h_3)$, and th_3 on points $x_1 \in [p_0, p_1]$, $x_2 \in [p_1, p_2]$, and $x_3 \in [p_2, p_3]$, respectively. Taking $x_1 = p_1$, $x_3 = p_2$, and $x_2 = m \equiv (h_1 p_1 + h_3 p_2)/(h_1 + h_3)$, and bounding the variance by the second moment about m , shows that $\nu_l(f + th) \leq t h_1 (p_1 - m)^2 + t h_3 (p_2 - m)^2$. Conversely, for any feasible x with mean μ , the variance is at least

$$t [h_1 (x_1 - \mu)^2 + h_3 (x_3 - \mu)^2] \geq t \frac{h_1 h_3}{h_1 + h_3} (x_3 - x_1)^2 \geq t \frac{h_1 h_3}{h_1 + h_3} (p_2 - p_1)^2,$$

where the first inequality minimizes over μ and the second uses $x_1 \leq p_1 < p_2 \leq x_3$. The upper bound equals the same quantity, so

$$\nu_l(f + th) = t \frac{h_1 h_3}{h_1 + h_3} (p_2 - p_1)^2.$$

The directional derivative $h \mapsto \frac{h_1 h_3}{h_1 + h_3} (p_2 - p_1)^2$ is therefore not linear: it is zero in the directions $[1, -1, 0]'$ and $[0, -1, 1]'$, but is equal to $(p_2 - p_1)^2/2$ in the direction $[1, -2, 1]'$ given by their sum.

We conclude by restating our results in terms of choice shares. Recall that $\mathcal{S} \equiv \{s \in \mathbb{R}^J : 0 \leq s_1 \leq \dots \leq s_J \leq 1\}$. Then the variance lower bound is differentiable when viewed as a function of choice shares $\nu_l : \mathcal{S} \rightarrow \mathbb{R}_+$, except perhaps at the vertices of $\mathcal{S}$. The derivative is given by

$$\nabla \nu_l(s) = p^* \text{diag}\{p^* - 2(Ds + e)'p^*\}D,$$

where p^* solves the quadratic program given by (12). Let

$$\mathcal{J} \equiv \{j \in \{1, \dots, J+1\} : (Ds + e)_j = 0\}.$$

Then the corresponding subvector $p_{\mathcal{J}}^*$ of p^* is given by (13).

Finally, we verify that the above derivative is continuous on $\mathcal{S} \setminus \text{vert}(\mathcal{S})$. Let $s \in \mathcal{S} \setminus \text{vert}(\mathcal{S})$, and denote the corresponding value of p^* by $p^*(s)$ in order to emphasize its dependence on s . Then from our previous results, we can conclude that $p^*(s)$ is the unique vector in $\mathbb{R}^{J+1}$ satisfying the optimality condition

$$p^*(s) = \max\{\underline{p}, \min\{\bar{p}, (Ds + e)'p^*(s)\}\}.$$

This follows from the definition of $p_{\mathcal{J}}^*(s)$ and the fact that $p_{-\mathcal{J}}^*(s)$ is optimal and unique. Since $p^*(s)$ must lie in the compact set $[p_0, p_{J+1}]^{J+1}$, Berge's theorem of the maximum ensures that the set of fixed points of the above equation is upper hemicontinuous at s .²¹ By uniqueness of $p^*(s)$, it follows that $p^*(s)$ is continuous at s , and therefore so is $\nabla \nu_l(s)$, as required.

(Upper bound) The proof for the upper variance bound proceeds in much the same way as the proof for the lower bound. Let both choke prices p_0 and p_{J+1} be finite so that the variance upper bound is finite. Let $\tilde{D}$ and $\tilde{e}$ be defined analogously to D and e , but as a $(J+2) \times (J+1)$ -dimensional matrix and $(J+2)$ -dimensional vector, respectively. In addition, define the $(J+2)$ -dimensional vector

$$\tilde{p} \equiv \begin{bmatrix} p_0 & \dots & p_{J+1} \end{bmatrix}',$$

and recall that $f \equiv Ds + e$. Note that by definition, all dispersed distributions $F_{D,r}$ are discrete distributions supported on $\tilde{p}$. In addition, the probability mass function

²¹Notice that it is not enough to simply apply the maximum theorem to the program given by (12), since this does not pin down $p_{\mathcal{J}}^*(s)$; we must verify continuity of $p^*(s)$ at $s \in \text{bd}(\mathcal{S}) \setminus \text{vert}(\mathcal{S})$ as well.

of any discrete distribution in the identified set $\mathcal{F}^*$ supported on $\tilde{p}$ may be written as

$$\tilde{f} \equiv \tilde{D} (\text{diag}\{1 - \eta\} \begin{bmatrix} 0 \\ \mathbf{s} \end{bmatrix} + \text{diag}\{\eta\} \begin{bmatrix} \mathbf{s} \\ 1 \end{bmatrix}) + \tilde{e}$$

for some $\eta \in [0, 1]^{J+1}$. It will be convenient to express $\tilde{f}$ in terms of f rather than $\mathbf{s}$,

$$\tilde{f} = \begin{bmatrix} 0 \\ f \end{bmatrix} + \tilde{D} \text{diag}\{f\} \eta.$$

Then the variance upper bound may be expressed as

$$\begin{aligned} v(F_{D,r^*}) &= \max_{\eta \in [0,1]^{J+1}} \tilde{p}' (\text{diag}\{\tilde{f}\} - \tilde{f} \tilde{f}') \tilde{p} \\ \text{s.t. } \tilde{f} &= \begin{bmatrix} 0 \\ f \end{bmatrix} + \tilde{D} \text{diag}\{f\} \eta. \end{aligned} \quad (14)$$

The reasoning here is analogous to that given in the case of the lower bound: the feasible set of distributions is a subset of $\mathcal{F}^*$, but a superset of the set of all dispersed distributions $\{F_{D,r} : r \in [0, 1]\}$. By [Stoye \(2010\)](#), the optimized objective must equal $v(F_{D,r^*})$; that is, the optimal dispersed distribution F_{D,r^*} corresponds to a discrete random variable supported on $\tilde{p}$ with probability mass function $\tilde{f}^*$, where $\tilde{f}^*$ is optimal for the above program.

We will now establish that the solution η^* is unique whenever $f > 0$. To show this, it will be helpful to simplify the objective of the above program:

$$\begin{aligned} \tilde{p}' (\text{diag}\{\tilde{f}\} - \tilde{f} \tilde{f}') \tilde{p} &= \tilde{p}' \text{diag}\{\tilde{p}\} \begin{bmatrix} 0 \\ f \end{bmatrix} - \begin{bmatrix} 0 & f' \end{bmatrix} \tilde{p} \tilde{p}' \begin{bmatrix} 0 \\ f \end{bmatrix} \\ &+ (\tilde{p}' \text{diag}\{\tilde{p}\} - 2 \begin{bmatrix} 0 & f' \end{bmatrix} \tilde{p} \tilde{p}') \tilde{D} \text{diag}\{f\} \eta \\ &- \eta' \text{diag}\{f\} \tilde{D}' \tilde{p} \tilde{p}' \tilde{D} \text{diag}\{f\} \eta. \end{aligned}$$

Notice that $\tilde{D}' \tilde{p} = -(\bar{p} - \underline{p})$ and $\tilde{D}' \text{diag}\{\tilde{p}\} \tilde{p} = -(\bar{p} - \underline{p})(\bar{p} + \underline{p})$. For ease of notation, let $g \equiv \text{diag}\{f\}(\bar{p} - \underline{p})$; since $f > 0$ by assumption, we have $g > 0$ as well. Then we may express the objective as

$$-\eta' g g' \eta - g' \text{diag}\{\bar{p} + \underline{p} - 2f' \bar{p}\} \eta + \bar{p}' (\text{diag}\{f\} - f f') \bar{p}.$$

The gradient of the objective at η^* is therefore given by

$$2g' \text{diag}\{f' \bar{p} - g' \eta^* - (\bar{p} + \underline{p})/2\}.$$

Now, suppose for the sake of contradiction that there exists another solution $\tilde{\eta}^* \neq \eta^*$. Then the difference between $\tilde{\eta}^*$ and η^* must lie in the null space of gg' , meaning we must have $g'\tilde{\eta}^* = g'\eta^*$. As a result, the gradient of the objective at $\tilde{\eta}^*$ is identical to the above expression. In particular, because both η^* and $\tilde{\eta}^*$ are assumed to be optimal, this implies that all components of η^* and $\tilde{\eta}^*$ at which the above gradient is non-zero must coincide: if the j th component of the gradient is positive, then $\eta_j^* = \tilde{\eta}_j^* = 1$, while if it is negative, then $\eta_j^* = \tilde{\eta}_j^* = 0$.

Since $g > 0$, there can be at most one component of the gradient that is exactly equal to zero, because the vector $(\bar{p} + \underline{p})/2$ is strictly increasing. If no such component exists, then we have reached a contradiction, because $\tilde{\eta}^* = \eta^*$. If such a component does exist, denote it by j , and the rest by $-j$. Then $\tilde{\eta}_j^*$ must satisfy

$$\begin{aligned} 0 &= 2g_j(f'\bar{p} - g_j\tilde{\eta}_j^* - g'_{-j}\tilde{\eta}_{-j}^* - (\bar{p}_j + \underline{p}_j)/2) \\ \iff \tilde{\eta}_j^* &= \frac{f'\bar{p} - g'_{-j}\tilde{\eta}_{-j}^* - (\bar{p}_j + \underline{p}_j)/2}{g_j}. \end{aligned}$$

But then we must have $\tilde{\eta}_j^* = \eta_j^*$, since we know that $g'_{-j}\tilde{\eta}_{-j}^* = g'_{-j}\eta_{-j}^*$, so we have reached a contradiction, because $\tilde{\eta}^* = \eta^*$. We conclude that η^* is unique whenever $f > 0$.²²

We can now apply Danskin's theorem to conclude that the value function of the program given by (14), call it $\nu_u(f)$, is differentiable whenever $f > 0$. To verify the conditions of the theorem, observe that the set $[0, 1]^{J+1}$ is compact, the objective is differentiable in η and f , and the partial derivative of the objective with respect to f is continuous in η and f . Since η^* is unique, we have that

$$\nabla \nu_u(f) = \tilde{p}' \text{diag}\{\tilde{p} - 2\tilde{f}^{*'}\tilde{p}\} \left(\begin{bmatrix} 0' \\ \mathbf{I} \end{bmatrix} + \tilde{D} \text{diag}\{\eta^*\} \right),$$

where $\mathbf{0} \in \mathbb{R}^{J+1}$ and $\mathbf{I}$ denotes the $(J+1) \times (J+1)$ -dimensional identity matrix.

Suppose some subvector of f , call it $f_{\mathcal{J}}$, is equal to zero, while the remaining components $f_{-\mathcal{J}}$ are strictly positive. Then we can apply our previous argument to conclude that $\eta_{-\mathcal{J}}^*$ is unique. On the other hand, the subvector $\eta_{\mathcal{J}}^*$ must satisfy $0 \leq \eta_{\mathcal{J}}^* \leq 1$, but is otherwise unrestricted. By Danskin's theorem, the Hadamard

²²Notice that unlike in the proof for the lower bound, we have not used the fact that f must have at least two components (since $J \geq 1$). Indeed, when f is a scalar (and therefore must equal 1), the unique solution is given by $\eta^* = \frac{1}{2}$.

directional derivative of ν_u with respect to f in the direction h is given by

$$\begin{aligned} \nabla_h \nu_u(f) &= \max_{\eta} \tilde{p}' \text{diag}\{\tilde{p} - 2\tilde{f}'\tilde{p}\} \left(\begin{bmatrix} \mathbf{0}' \\ I \end{bmatrix} + \tilde{D} \text{diag}\{\eta\} \right) h \\ \text{s.t. } & 0 \leq \eta_{\mathcal{J}} \leq 1, \quad \eta_{-\mathcal{J}} = \eta_{-\mathcal{J}}^*, \quad \text{and} \quad \tilde{f} = \begin{bmatrix} 0 \\ f \end{bmatrix} + \tilde{D} \text{diag}\{f\}\eta. \end{aligned}$$

Let h be such that $h_{\mathcal{J}} \geq 0$, in which case the above optimization problem can be solved coordinatewise for each $j \in \mathcal{J}$ to get

$$\eta_{\mathcal{J}}^* \equiv \mathbb{1}[(\bar{p}_{\mathcal{J}} + \underline{p}_{\mathcal{J}})/2 \leq \tilde{f}'\tilde{p}], \quad (15)$$

keeping in mind that $\tilde{f}^*$ only depends on $\eta_{-\mathcal{J}}^*$ and not $\eta_{\mathcal{J}}^*$. Notice that just as in the case of the lower bound, the solution does not depend on the choice of direction $h_{\mathcal{J}}$, so ν_u is differentiable at f in the sense previously discussed.

We conclude by restating our results in terms of choice shares. The variance upper bound is differentiable when viewed as a function of choice shares $\nu_u : \mathcal{S} \rightarrow \mathbb{R}_+$. The derivative is given by

$$\nabla \nu_u(\mathbf{s}) = \tilde{p}' \text{diag}\{\tilde{p} - 2\tilde{f}'\tilde{p}\} \left(\begin{bmatrix} \mathbf{0}' \\ I \end{bmatrix} + \tilde{D} \text{diag}\{\eta^*\} \right) D,$$

where η^* and $\tilde{f}^*$ solve the quadratic program given by (14). Let

$$\mathcal{J} \equiv \{j \in \{1, \dots, J+1\} : (D\mathbf{s} + e)_j = 0\}.$$

Then the corresponding subvector $\eta_{\mathcal{J}}^*$ of η^* is given by (15).

Finally, we verify that the above derivative is continuous on $\mathcal{S}$. Let $\mathbf{s} \in \mathcal{S}$, and denote the corresponding values of η^* and $\tilde{f}^*$ by $\eta^*(\mathbf{s})$ and $\tilde{f}^*(\mathbf{s})$, respectively, in order to emphasize their dependence on $\mathbf{s}$. We cannot apply the same proof technique used in the case of the lower bound, because $\eta^*(\mathbf{s})$ cannot be expressed as the unique fixed point of a continuous map on a compact set. However, we can still apply Berge's theorem of the maximum to conclude that the set of maximizers of (14) is upper hemicontinuous at $\mathbf{s}$. Since (14) has a unique maximizer for $\mathbf{s} \in \text{int}(\mathcal{S})$, it follows that $\eta^*(\mathbf{s})$, and consequently $\nabla \nu_u(\mathbf{s})$, is continuous on $\text{int}(\mathcal{S})$.

Now, let $\bar{\mathbf{s}} \in \text{bd}(\mathcal{S})$, let $\bar{\mathcal{J}}$ denote the set of indices at which $\bar{f} = D\bar{\mathbf{s}} + e$ is equal to zero, and let $-\bar{\mathcal{J}}$ denote the remaining indices. Note that neither $\bar{\mathcal{J}}$ nor $-\bar{\mathcal{J}}$ is empty. We would like to show that for any sequence $\{\mathbf{s}_n\}_{n=1}^{\infty}$ on $\mathcal{S}$ with $\mathbf{s}_n \rightarrow \bar{\mathbf{s}}$, we have $\nabla \nu_u(\mathbf{s}_n) \rightarrow \nabla \nu_u(\bar{\mathbf{s}})$, where for $\mathbf{s}_n \in \text{bd}(\mathcal{S})$ the components of $\eta^*(\mathbf{s}_n)$ at which $D\mathbf{s}_n + e$ is zero are given by (15), so the argument below applies to them unchanged.

First, we will establish that $\eta_{-\bar{\mathcal{J}}}^*(\mathbf{s}_n) \rightarrow \eta_{-\bar{\mathcal{J}}}^*(\bar{\mathbf{s}})$. Suppose not. Then there exists

an $\epsilon > 0$ and a subsequence n_k such that $\|\eta_{-\bar{\mathcal{J}}}^*(\mathbf{s}_{n_k}) - \eta_{-\bar{\mathcal{J}}}^*(\bar{\mathbf{s}})\| \geq \epsilon$ for all k . But then $\mathbf{s}_{n_k} \rightarrow \bar{\mathbf{s}}$ as $k \rightarrow \infty$, so by upper hemicontinuity there must exist a further subsequence n_{k_l} such that $\eta^*(\mathbf{s}_{n_{k_l}})$ converges to a maximizer of (14) at $\bar{\mathbf{s}}$ as $l \rightarrow \infty$. But the $-\bar{\mathcal{J}}$ subvector of any maximizer of (14) at $\bar{\mathbf{s}}$ must coincide with $\eta_{-\bar{\mathcal{J}}}^*(\bar{\mathbf{s}})$, providing a contradiction. It follows that $\tilde{f}^{*'}(\mathbf{s}_n)\tilde{p} \rightarrow \tilde{f}^{*'}(\bar{\mathbf{s}})\tilde{p}$, since

$$\tilde{f}^{*'}(\mathbf{s}_n)\tilde{p} = f_n'\bar{p} - (\bar{p} - \underline{p})' \text{diag}\{f_n\}\eta^*(\mathbf{s}_n),$$

where $f_n = D\mathbf{s}_n + e$, $f_{\bar{\mathcal{J}},n} \rightarrow \bar{f}_{\bar{\mathcal{J}}} = 0$, and $\eta_{\bar{\mathcal{J}}}^*(\mathbf{s}_n)$ is bounded.

Now, consider $j \in \bar{\mathcal{J}}$, and suppose that $(\bar{p}_j + \underline{p}_j)/2 \neq \tilde{f}^{*'}(\bar{\mathbf{s}})\tilde{p}$. Then for n large enough, we will have $|\tilde{f}^{*'}(\mathbf{s}_n)\tilde{p} - \tilde{f}^{*'}(\bar{\mathbf{s}})\tilde{p}| < |(\bar{p}_j + \underline{p}_j)/2 - \tilde{f}^{*'}(\bar{\mathbf{s}})\tilde{p}|$, meaning

$$\begin{aligned} (\bar{p}_j + \underline{p}_j)/2 < \tilde{f}^{*'}(\bar{\mathbf{s}})\tilde{p} &\implies (\bar{p}_j + \underline{p}_j)/2 < \tilde{f}^{*'}(\mathbf{s}_n)\tilde{p} \implies \eta_j^*(\mathbf{s}_n) = 1, \quad \text{and} \\ (\bar{p}_j + \underline{p}_j)/2 > \tilde{f}^{*'}(\bar{\mathbf{s}})\tilde{p} &\implies (\bar{p}_j + \underline{p}_j)/2 > \tilde{f}^{*'}(\mathbf{s}_n)\tilde{p} \implies \eta_j^*(\mathbf{s}_n) = 0. \end{aligned}$$

It follows that $\eta_j^*(\mathbf{s}_n) \rightarrow \eta_j^*(\bar{\mathbf{s}})$. Suppose instead that $(\bar{p}_j + \underline{p}_j)/2 = \tilde{f}^{*'}(\bar{\mathbf{s}})\tilde{p}$. Then it need not be the case that $\eta_j^*(\mathbf{s}_n) \rightarrow \eta_j^*(\bar{\mathbf{s}})$, but this is irrelevant because $\nabla \nu_u(\bar{\mathbf{s}})$ does not depend on $\eta_j^*(\bar{\mathbf{s}})$. To see this, observe that the coefficient on $\eta_j^*(\bar{\mathbf{s}})$ in the expression for $\nabla \nu_u(\bar{f})$ is given by

$$(\bar{p}_j - \underline{p}_j)(2\tilde{f}^{*'}(\bar{\mathbf{s}})\tilde{p} - (\bar{p}_j + \underline{p}_j)) = 0.$$

Since $\tilde{f}^{*'}(\mathbf{s}_n)\tilde{p} \rightarrow \tilde{f}^{*'}(\bar{\mathbf{s}})\tilde{p}$ and $\eta_j^*(\mathbf{s}_n)$ is bounded, we conclude that $\nabla \nu_u(\mathbf{s}_n) \rightarrow \nabla \nu_u(\bar{\mathbf{s}})$. So, $\nabla \nu_u(\mathbf{s})$ is continuous on $\mathcal{S}$, as claimed. Q.E.D.

SA.2 Monte Carlo Simulations

Table SA.1 reports the coverage of our confidence intervals for two simulations calibrated to the Mas and Pallais (2017) data. In Panel A, the distribution of valuations in the data-generating process is given by the maximum likelihood estimate in the actual data, which we then sample from together with the empirical distribution of prices. In Panel B, we set the distribution of valuations to be the nonparametric isotonic regression estimate. In both cases, we correct for inattention by assuming that 29% of individuals are inattentive across all treatment arms, which corresponds to the average inattention rate (2×0.145) reported by Mas and Pallais (2017) and imposed in their hours treatments (see Appendix A for details). Choke prices are given by $p_0 = -\$5.01$ and $p_{J+1} = \$10$ for all treatment arms.

The table shows that all of our methods produce coverage close to the nominal

Table SA.1: Coverage in simulations, $1 - \alpha = 0.95$

	Mean	SD	Q_1	Q_2	Q_3
<i>Panel A. Logistic simulations</i>					
Flexible schedule	94.2	92.7	98.6	97.8	98.9
Flexible # of hours	94.1	93.2	98.8	98.8	97.1
Work from home	93.0	94.1	96.9	99.1	99.5
Combined flexible	93.3	94.5	97.9	98.2	99.2
Employer discretion	92.7	95.7	97.5	99.9	97.5
20-hour job	94.4	93.9	97.6	99.6	100.0
50-hour job	94.6	94.9	100.0	99.9	100.0
<i>Panel B. Nonparametric simulations</i>					
Flexible schedule	93.3	96.0	99.7	100.0	97.8
Flexible # of hours	94.2	96.1	99.5	99.5	98.6
Work from home	92.8	94.5	99.8	99.3	99.6
Combined flexible	92.7	95.9	99.8	99.8	99.4
Employer discretion	94.3	95.1	99.4	99.5	100.0
20-hour job	94.7	94.9	99.5	99.7	100.0
50-hour job	94.6	97.0	100.0	99.9	100.0

Notes: Coverage in percent. Q_1 , Q_2 , and Q_3 are the 0.25, 0.5, and 0.75 quantiles of WTP. Panel A reports the coverage of confidence intervals constructed from logit shares (Section 5), and Panel B the coverage of confidence intervals constructed from binned shares (Section 6).

level of 95% for both the mean and standard deviation across every outcome. For the quantiles, the coverage is often well above the nominal level. This is not surprising given the discrete nature of the inference problem. For example, if $s_j < r < s_{j+1}$ for some $j \in \{0, \dots, J\}$, then $\mathbb{P}[C_n(r) = [p_j, p_{j+1}]]$ will approach one as the sample size increases, and therefore so will the coverage probability. We would only expect coverage to be close to the nominal level in particularly difficult situations, such as when $r = s_j$ for some $j \in \{1, \dots, J\}$, or even for all $j \in \{1, \dots, J\}$.

SA.3 Additional Empirical Results

This appendix collects additional results for Section 7.

Table SA.2: Estimates and 95% CIs, flexibility arms, $p_{J+1} = \$6$

	Mean	SD	Q_1	Q_2	Q_3
<i>Flexible schedule</i>					
Logit	0.48	2.15	-0.82	0.48	1.79
	[0.07, 0.89]	[1.27, 3.03]	[-1.34, -0.31]	[0.07, 0.89]	[0.99, 2.59]
Logit shares	[0.30, 0.66]	[1.89, 2.28]	[-1.00, -0.75]	[0.25, 0.50]	[1.75, 2.00]
	[-0.01, 1.02]	[1.33, 2.94]	[-1.75, -0.25]	[0.00, 1.25]	[1.00, 4.00]
Binned shares	[0.65, 0.93]	[2.44, 2.66]	[-0.50, 0.25]	[0.00, 0.25]	[0.00, 1.00]
	[0.07, 1.56]	[1.50, 3.60]	[-2.50, 0.50]	[-0.50, 1.00]	[-0.25, 6.00]
<i>Flexible # of hours</i>					
Logit	-0.22	2.24	-1.58	-0.22	1.14
	[-0.63, 0.19]	[1.37, 3.12]	[-2.18, -0.97]	[-0.63, 0.19]	[0.41, 1.87]
Logit shares	[-0.38, -0.03]	[1.96, 2.35]	[-1.75, -1.50]	[-0.25, 0.00]	[1.00, 1.25]
	[-0.72, 0.32]	[1.41, 2.97]	[-2.50, -1.00]	[-0.75, 0.50]	[0.50, 2.50]
Binned shares	[0.13, 0.55]	[2.57, 3.06]	[-1.25, -0.50]	[-0.50, 0.25]	[0.00, 0.75]
	[-0.57, 1.27]	[1.48, 4.19]	[-5.01, 0.50]	[-2.00, 1.25]	[-1.00, 6.00]
<i>Work from home</i>					
Logit	1.33	1.86	0.20	1.33	2.45
	[0.80, 1.85]	[1.06, 2.66]	[-0.38, 0.77]	[0.80, 1.85]	[1.62, 3.28]
Logit shares	[1.13, 1.51]	[1.62, 2.03]	[0.00, 0.25]	[1.25, 1.50]	[2.25, 2.50]
	[0.73, 1.97]	[1.09, 2.65]	[-0.75, 0.75]	[0.75, 2.25]	[1.75, 4.00]
Binned shares	[1.53, 1.96]	[2.18, 3.08]	[-0.25, 0.75]	[0.25, 1.25]	[2.25, 2.50]
	[0.63, 2.89]	[0.33, 4.86]	[-5.01, 2.50]	[-0.25, 6.00]	[-0.25, 6.00]
<i>Combined flexible</i>					
Logit	1.17	2.34	-0.25	1.17	2.58
	[0.57, 1.76]	[1.42, 3.25]	[-0.86, 0.37]	[0.57, 1.76]	[1.61, 3.55]
Logit shares	[0.94, 1.35]	[1.99, 2.44]	[-0.25, 0.00]	[1.00, 1.25]	[2.50, 2.75]
	[0.50, 1.85]	[1.45, 3.05]	[-1.25, 0.50]	[0.50, 2.00]	[1.75, 5.00]
Binned shares	[1.11, 1.63]	[2.01, 2.50]	[-1.25, -0.50]	[0.00, 1.25]	[2.50, 2.75]
	[0.27, 2.53]	[0.46, 3.94]	[-5.01, 2.75]	[-1.25, 6.00]	[-0.25, 6.00]

Notes: For each arm and estimator, the first row reports the logit point estimate or bound estimates, and the second row 95% CIs. Logit shares: bounds under Assumption [FF](#) with a logistic link and a linear index. For the 50-hour arm, these equal the nonparametric bounds and are not shown. The logit estimates do not depend on the upper choke price. Q_1 , Q_2 , and Q_3 are the 0.25, 0.5, and 0.75 quantiles of WTP. Bound estimates from the binned shares can cross (lower above upper) because of sampling error.

Table SA.3: Estimates and 95% CIs, employer discretion and hours, $p_{J+1} = \$6$

	Mean	SD	Q_1	Q_2	Q_3
<i>Employer discretion</i>					
Logit	3.41	2.95	1.63	3.41	5.20
	[2.60, 4.23]	[2.02, 3.88]	[1.06, 2.19]	[2.60, 4.23]	[3.92, 6.49]
Logit shares	[2.70, 3.38]	[2.06, 2.59]	[1.50, 1.75]	[3.00, 4.00]	[5.00, 6.00]
	[2.33, 3.82]	[1.72, 2.94]	[1.00, 2.25]	[2.50, 5.00]	[4.00, 6.00]
Binned shares	[3.06, 3.56]	[2.18, 1.96]	[2.00, 2.50]	[2.00, 2.75]	[5.00, 6.00]
	[2.56, 4.14]	[0.93, 3.18]	[-0.50, 2.75]	[2.00, 6.00]	[3.00, 6.00]
<i>20-hour job</i>					
Logit	6.00	5.71	2.54	6.00	9.45
	[3.44, 8.55]	[2.46, 8.95]	[1.40, 3.68]	[3.44, 8.55]	[5.04, 13.86]
Logit shares	[3.26, 4.06]	[2.62, 3.15]	[2.50, 2.75]	[5.00, 6.00]	[5.00, 6.00]
	[2.88, 4.47]	[2.08, 3.68]	[1.25, 5.00]	[4.00, 6.00]	[5.00, 6.00]
Binned shares	[3.23, 4.02]	[2.86, 2.88]	[3.00, 4.00]	[5.00, 6.00]	[5.00, 6.00]
	[2.77, 4.53]	[2.08, 3.66]	[-5.01, 6.00]	[3.00, 6.00]	[5.00, 6.00]
<i>50-hour job</i>					
Logit	0.88	6.11	-2.82	0.88	4.58
	[-0.56, 2.33]	[0.64, 11.57]	[-7.44, 1.81]	[-0.56, 2.33]	[2.41, 6.75]
Binned shares	[-1.52, 3.25]	[0.73, 5.19]	[-5.01, 1.60]	[-5.01, 1.60]	[3.20, 6.00]
	[-1.92, 3.44]	[0.68, 5.38]	[-5.01, 1.60]	[-5.01, 3.20]	[3.20, 6.00]

Notes: See Table [SA.2](#).

Table SA.4: Estimates and 95% CIs, flexibility arms, $p_{J+1} = \$10$

	Mean	SD	Q_1	Q_2	Q_3
<i>Flexible schedule</i>					
Logit	0.48	2.15	-0.82	0.48	1.79
	[0.07, 0.89]	[1.27, 3.03]	[-1.34, -0.31]	[0.07, 0.89]	[0.99, 2.59]
Logit shares	[0.30, 0.75]	[1.89, 2.55]	[-1.00, -0.75]	[0.25, 0.50]	[1.75, 2.00]
	[-0.01, 1.20]	[1.33, 3.48]	[-1.75, -0.25]	[0.00, 1.25]	[1.00, 4.00]
Binned shares	[0.65, 1.82]	[2.44, 4.47]	[-0.50, 0.25]	[0.00, 0.25]	[0.00, 1.00]
	[0.03, 3.38]	[1.51, 7.04]	[-2.50, 0.50]	[-0.50, 1.00]	[-0.25, 10.00]
<i>Flexible # of hours</i>					
Logit	-0.22	2.24	-1.58	-0.22	1.14
	[-0.63, 0.19]	[1.37, 3.12]	[-2.18, -0.97]	[-0.63, 0.19]	[0.41, 1.87]
Logit shares	[-0.38, 0.03]	[1.96, 2.54]	[-1.75, -1.50]	[-0.25, 0.00]	[1.00, 1.25]
	[-0.72, 0.43]	[1.41, 3.37]	[-2.50, -1.00]	[-0.75, 0.50]	[0.50, 2.50]
Binned shares	[0.13, 0.82]	[2.57, 3.66]	[-1.25, -0.50]	[-0.50, 0.25]	[0.00, 0.75]
	[-0.83, 2.80]	[0.90, 7.63]	[-5.01, 0.50]	[-2.00, 1.25]	[-1.00, 10.00]
<i>Work from home</i>					
Logit	1.33	1.86	0.20	1.33	2.45
	[0.80, 1.85]	[1.06, 2.66]	[-0.38, 0.77]	[0.80, 1.85]	[1.62, 3.28]
Logit shares	[1.13, 1.62]	[1.62, 2.35]	[0.00, 0.25]	[1.25, 1.50]	[2.25, 2.50]
	[0.73, 2.19]	[1.10, 3.31]	[-0.75, 0.75]	[0.75, 2.25]	[1.75, 4.00]
Binned shares	[1.53, 3.00]	[2.18, 5.06]	[-0.25, 0.75]	[0.25, 1.25]	[2.25, 2.50]
	[0.37, 5.80]	[-0.04, 10.33]	[-5.01, 2.50]	[-0.25, 10.00]	[-0.25, 10.00]
<i>Combined flexible</i>					
Logit	1.17	2.34	-0.25	1.17	2.58
	[0.57, 1.76]	[1.42, 3.25]	[-0.86, 0.37]	[0.57, 1.76]	[1.61, 3.55]
Logit shares	[0.94, 1.55]	[1.99, 2.92]	[-0.25, 0.00]	[1.00, 1.25]	[2.50, 2.75]
	[0.50, 2.21]	[1.46, 3.90]	[-1.25, 0.50]	[0.50, 2.00]	[1.75, 5.00]
Binned shares	[1.11, 2.33]	[2.01, 3.95]	[-1.25, -0.50]	[0.00, 1.25]	[2.50, 2.75]
	[0.07, 4.66]	[0.15, 8.00]	[-5.01, 2.75]	[-1.25, 10.00]	[-0.25, 10.00]

Notes: See Table SA.2.

Table SA.5: Estimates and 95% CIs, employer discretion and hours, $p_{J+1} = \$10$

	Mean	SD	Q_1	Q_2	Q_3
<i>Employer discretion</i>					
Logit	3.41	2.95	1.63	3.41	5.20
	[2.60, 4.23]	[2.02, 3.88]	[1.06, 2.19]	[2.60, 4.23]	[3.92, 6.49]
Logit shares	[2.70, 4.48]	[2.06, 4.02]	[1.50, 1.75]	[3.00, 4.00]	[5.00, 10.00]
	[2.33, 5.37]	[1.72, 4.60]	[1.00, 2.25]	[2.50, 5.00]	[4.00, 10.00]
Binned shares	[3.06, 5.35]	[2.18, 4.07]	[2.00, 2.50]	[2.00, 2.75]	[5.00, 10.00]
	[2.56, 6.57]	[0.97, 4.95]	[-0.50, 2.75]	[2.00, 10.00]	[3.00, 10.00]
<i>20-hour job</i>					
Logit	6.00	5.71	2.54	6.00	9.45
	[3.44, 8.55]	[2.46, 8.95]	[1.40, 3.68]	[3.44, 8.55]	[5.04, 13.86]
Logit shares	[3.26, 6.37]	[2.62, 4.87]	[2.50, 2.75]	[5.00, 10.00]	[5.00, 10.00]
	[2.88, 7.21]	[2.08, 5.38]	[1.25, 5.00]	[4.00, 10.00]	[5.00, 10.00]
Binned shares	[3.23, 6.26]	[2.86, 4.71]	[3.00, 4.00]	[5.00, 10.00]	[5.00, 10.00]
	[2.77, 7.36]	[2.10, 5.35]	[-5.01, 10.00]	[3.00, 10.00]	[5.00, 10.00]
<i>50-hour job</i>					
Logit	0.88	6.11	-2.82	0.88	4.58
	[-0.56, 2.33]	[0.64, 11.57]	[-7.44, 1.81]	[-0.56, 2.33]	[2.41, 6.75]
Binned shares	[-1.52, 4.59]	[0.73, 6.90]	[-5.01, 1.60]	[-5.01, 1.60]	[3.20, 10.00]
	[-1.92, 5.01]	[0.68, 7.32]	[-5.01, 1.60]	[-5.01, 3.20]	[3.20, 10.00]

Notes: See Table SA.2.

Table SA.6: Estimates and 95% CIs, flexibility arms, $p_{J+1} = \$12$

	Mean	SD	Q_1	Q_2	Q_3
<i>Flexible schedule</i>					
Logit	0.48	2.15	-0.82	0.48	1.79
	[0.07, 0.89]	[1.27, 3.03]	[-1.34, -0.31]	[0.07, 0.89]	[0.99, 2.59]
Logit shares	[0.30, 0.79]	[1.89, 2.72]	[-1.00, -0.75]	[0.25, 0.50]	[1.75, 2.00]
	[-0.01, 1.29]	[1.33, 3.83]	[-1.75, -0.25]	[0.00, 1.25]	[1.00, 4.00]
Binned shares	[0.65, 2.26]	[2.44, 5.38]	[-0.50, 0.25]	[0.00, 0.25]	[0.00, 1.00]
	[0.01, 4.38]	[1.51, 8.79]	[-2.50, 0.50]	[-0.50, 1.00]	[-0.25, 12.00]
<i>Flexible # of hours</i>					
Logit	-0.22	2.24	-1.58	-0.22	1.14
	[-0.63, 0.19]	[1.37, 3.12]	[-2.18, -0.97]	[-0.63, 0.19]	[0.41, 1.87]
Logit shares	[-0.38, 0.06]	[1.96, 2.66]	[-1.75, -1.50]	[-0.25, 0.00]	[1.00, 1.25]
	[-0.72, 0.48]	[1.42, 3.63]	[-2.50, -1.00]	[-0.75, 0.50]	[0.50, 2.50]
Binned shares	[0.13, 0.96]	[2.57, 4.03]	[-1.25, -0.50]	[-0.50, 0.25]	[0.00, 0.75]
	[-0.91, 3.66]	[0.80, 9.55]	[-5.01, 0.50]	[-2.00, 1.25]	[-1.00, 12.00]
<i>Work from home</i>					
Logit	1.33	1.86	0.20	1.33	2.45
	[0.80, 1.85]	[1.06, 2.66]	[-0.38, 0.77]	[0.80, 1.85]	[1.62, 3.28]
Logit shares	[1.13, 1.67]	[1.62, 2.56]	[0.00, 0.25]	[1.25, 1.50]	[2.25, 2.50]
	[0.73, 2.30]	[1.10, 3.72]	[-0.75, 0.75]	[0.75, 2.25]	[1.75, 4.00]
Binned shares	[1.53, 3.52]	[2.18, 6.08]	[-0.25, 0.75]	[0.25, 1.25]	[2.25, 2.50]
	[0.30, 7.37]	[-0.16, 13.17]	[-5.01, 2.50]	[-0.25, 12.00]	[-0.25, 12.00]
<i>Combined flexible</i>					
Logit	1.17	2.34	-0.25	1.17	2.58
	[0.57, 1.76]	[1.42, 3.25]	[-0.86, 0.37]	[0.57, 1.76]	[1.61, 3.55]
Logit shares	[0.94, 1.64]	[1.99, 3.22]	[-0.25, 0.00]	[1.00, 1.25]	[2.50, 2.75]
	[0.50, 2.40]	[1.46, 4.42]	[-1.25, 0.50]	[0.50, 2.00]	[1.75, 5.00]
Binned shares	[1.11, 2.69]	[2.01, 4.72]	[-1.25, -0.50]	[0.00, 1.25]	[2.50, 2.75]
	[-0.01, 5.85]	[0.05, 10.23]	[-5.01, 2.75]	[-1.25, 12.00]	[-0.25, 12.00]

Notes: See Table SA.2.

Table SA.7: Estimates and 95% CIs, employer discretion and hours, $p_{J+1} = \$12$

	Mean	SD	Q_1	Q_2	Q_3
<i>Employer discretion</i>					
Logit	3.41	2.95	1.63	3.41	5.20
	[2.60, 4.23]	[2.02, 3.88]	[1.06, 2.19]	[2.60, 4.23]	[3.92, 6.49]
Logit shares	[2.70, 5.02]	[2.06, 4.83]	[1.50, 1.75]	[3.00, 4.00]	[5.00, 12.00]
	[2.33, 6.17]	[1.72, 5.57]	[1.00, 2.25]	[2.50, 5.00]	[4.00, 12.00]
Binned shares	[3.06, 6.25]	[2.18, 5.10]	[2.00, 2.50]	[2.00, 2.75]	[5.00, 12.00]
	[2.56, 7.87]	[0.97, 6.01]	[−0.50, 2.75]	[2.00, 12.00]	[3.00, 12.00]
<i>20-hour job</i>					
Logit	6.00	5.71	2.54	6.00	9.45
	[3.44, 8.55]	[2.46, 8.95]	[1.40, 3.68]	[3.44, 8.55]	[5.04, 13.86]
Logit shares	[3.26, 7.53]	[2.62, 5.80]	[2.50, 2.75]	[5.00, 12.00]	[5.00, 12.00]
	[2.88, 8.62]	[2.08, 6.28]	[1.25, 5.00]	[4.00, 12.00]	[5.00, 12.00]
Binned shares	[3.23, 7.38]	[2.86, 5.66]	[3.00, 4.00]	[5.00, 12.00]	[5.00, 12.00]
	[2.77, 8.85]	[2.10, 6.29]	[−5.01, 12.00]	[3.00, 12.00]	[5.00, 12.00]
<i>50-hour job</i>					
Logit	0.88	6.11	−2.82	0.88	4.58
	[−0.56, 2.33]	[0.64, 11.57]	[−7.44, 1.81]	[−0.56, 2.33]	[2.41, 6.75]
Binned shares	[−1.52, 5.26]	[0.73, 7.79]	[−5.01, 1.60]	[−5.01, 1.60]	[3.20, 12.00]
	[−1.92, 5.79]	[0.68, 8.30]	[−5.01, 1.60]	[−5.01, 3.20]	[3.20, 12.00]

Notes: See Table SA.2.

Table SA.8: Gender differences in WTP, all arms other than hours

	p_{J+1}	Logit	Distribution-free bounds	
			Logit shares	Binned shares
Flexible schedule	\$6	0.42 [−0.49, 1.34]	[0.03, 0.79]	[0.96, 1.59]
			[−0.67, 1.55]	[−0.22, 2.82]
	\$10	[−0.23, 0.84]	[0.39, 3.59]	
		[−1.01, 1.69]	[−1.12, 6.60]	
Flexible # of hours	\$6	0.15 [−0.71, 1.01]	[−0.16, 0.50]	[0.33, 1.03]
			[−0.87, 1.23]	[−2.28, 3.08]
	\$10	[−0.18, 0.59]	[1.43, 1.65]	
		[−0.89, 1.37]	[−4.66, 5.60]	
Work from home	\$6	0.91 [−0.11, 1.93]	[−0.18, 0.63]	[1.98, 1.96]
			[−0.91, 1.45]	[−6.00, 6.99]
	\$10	[0.55, 1.28]	[0.00, 1.27]	
		[−0.32, 2.12]	[−1.98, 3.21]	
Combined flexible	\$6	1.52 [0.40, 2.65]	[0.48, 1.35]	[0.53, 3.27]
			[−0.56, 2.23]	[−3.16, 7.79]
	\$10	[0.44, 1.39]	[0.79, 4.27]	
		[−0.70, 2.30]	[−4.04, 10.36]	
Employer discretion	\$6	2.16 [0.50, 3.81]	[1.06, 1.81]	[1.03, 1.42]
			[0.21, 2.71]	[−1.53, 3.99]
	\$10	[1.06, 2.20]	[2.76, 2.43]	
		[0.21, 3.36]	[−2.35, 6.94]	
Employer discretion	\$6	2.16 [0.50, 3.81]	[1.06, 2.40]	[3.63, 2.93]
			[0.21, 3.73]	[−3.18, 8.58]
	\$10	[1.06, 2.44]	[1.49, 2.43]	
		[0.33, 3.19]	[0.49, 3.35]	
Employer discretion	\$6	2.16 [0.50, 3.81]	[0.27, 4.04]	[0.61, 4.91]
			[−0.76, 5.39]	[−0.95, 6.56]
	\$10	[−0.13, 4.83]	[0.16, 6.15]	
		[−1.34, 6.53]	[−1.79, 8.30]	

Notes: See Table 1. Bound estimates from the binned shares can cross (lower above upper) because of sampling error.