Pooling and the Identification of Willingness to Pay

Neil Thakral
Brown University

Linh T. Tô
Boston University

April 2026[*]

Abstract

Pooling choices across heterogeneous individuals leads to a fundamental identification problem: the pooled choice curve identifies the distribution of shock-contaminated valuations instead of the distribution of underlying willingness to pay (WTP). Consequently, the price at which the pooled choice probability equals one-half can be arbitrarily far from median WTP, even if the pooled choice curve is known perfectly. When choice probabilities are observed only at finitely many prices, we derive bounds for the mean and quantiles of the WTP distribution under economically interpretable restrictions on choice shocks and the tails of valuations. In an application to experimental estimates of the value of non-work time, the bounds are tight enough to exclude several estimates from pooled logit and finite-mixture logit specifications. We then illustrate how pooled logit models can produce misleading estimates, including estimates outside the range of individual WTPs. These distortions occur even for a population with homogeneous valuations or homogeneous choice errors, and they can worsen as preferences become more homogeneous or the share of noisy types decreases. Changes in the experimental design alone can lead to arbitrary changes in pooled-logit-implied WTP estimates.

---

[*] Thakral: Department of Economics, Brown University, Box B, Providence, RI 02912 (email: neil_thakral@brown.edu). Tô: Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215 (email: linhto@bu.edu). This material is based upon work supported by the National Science Foundation under Grant Nos. 2149371 and 2149414.

1 Introduction

In binary-choice settings, such as randomized pricing experiments, stated preference studies, and discrete choice experiments (DCEs), researchers typically pool together data from heterogeneous individuals to estimate a relationship between price and choice probability. Practitioner guides (Carson, 2000, p. 1413), textbooks (Desvousges et al., 2010, p. 88), and handbook chapters (Carson and Hanemann, 2005, p. 857) commonly describe the price at which the pooled choice probability equals 0.5 as the median willingness to pay (WTP). This paper shows that such interpretations are generally not valid because pooling data across individuals to obtain the average choice probability as a function of price does not identify the distribution of underlying individual valuations.

We start by assuming the choice probability curve is perfectly known. Even in this case, the discrepancy between the 50 percent crossing point and median or mean WTP can be arbitrarily large. An individual's WTP corresponds to the inverse of their choice probability function at one-half to determine the price difference at which that individual is indifferent between two alternatives. With the full distribution of individual WTP, one can then compute averages or quantiles to summarize the data. Pooling instead averages choice probabilities at each price and then inverts the resulting curve. Because averaging and inversion do not commute in nonlinear models, the pooled curve summarizes aggregate choice behavior (equivalently, the distribution of shock-contaminated valuations) rather than the underlying distribution of individual WTP. Deterministic choice is the only case in which the pooled crossing generically coincides with median WTP. In general, no nontrivial functional of the WTP distribution is identified from the pooled curve alone. Although motivated by DCEs, these considerations extend beyond stated-preference data. The distinction persists in any setting that pools nonlinear choice probabilities and then inverts them.

Given these impossibility results, we turn to consider the case that arises in practice, where the pooled choice probability curve is only observed at a finite set of prices. In this case, we use the pooled curve to construct bounds on median and mean WTP.

Quantiles of the underlying WTP distribution can be bounded under an assumption that large shocks are unlikely, given a choice of thresholds for a large error and how rare such errors are. In the absence of any shocks, the median WTP would have to fall between the highest observed price at which the pooled choice probability exceeds one-half and the lowest observed price at which it is at most one-half. With shocks, the pooled curve instead reflects shock-contaminated WTP. To recover underlying WTP, after expanding those bounds in both directions by the largest possible non-large shock, the remaining source of distortion is the small share of observations with large errors. Since the shocks may push those observations to the opposite side of the 50 percent crossing point, the crossing point must be adjusted in either direction based on the largest fraction of responses that could have been misclassified by such large shocks.

We propose two approaches for bounding mean WTP. The first takes the step function that bounds the pooled curve from below and above, shifts those bounds to account for the largest possible non-large shock, and adds an allowance for the small share of large shocks. The second assumes mean-zero shocks, so that mean WTP coincides with the mean of shock-contaminated WTP, and then bounds mean WTP directly based on the area under the step functions bounding the pooled curve. Both approaches can be useful in practice because, depending on the shape of the pooled curve near the boundaries, either can produce tighter bounds. Although we focus on the case of a single price attribute, the same logic applies with any number of attributes.

We demonstrate the performance of these bounds in practice by revisiting data from the DCE in Mas and Pallais (2019) that estimates the marginal value of time, or the effective wages that workers are willing to accept for longer working hours. Leveraging the features of their experimental design to posit the size and frequency of large shocks, our approaches produce informative bounds on mean and median willingness to accept (WTA). In many cases, the bounds are tight enough to rule out pooled logit and mixed logit WTA estimates, and we show that this conclusion is robust to substantially relaxing our bounding assumptions. Mas and Pallais (2019) use the logit estimates to compute the average flow value of non-work time relative to the market wage for a 40-hour job, and their estimate of the value of non-work time falls below the lower bound implied by our approach. Adjusting their estimates that fall outside the bounds to the nearest endpoint increases the implied value of non-work time by 3.9 percent.

The fact that pooled logit estimates fall outside the bounds suggests that the logit model is misspecified. To illustrate this more concretely, we show that pooling respondents with different degrees of choice consistency can produce a pooled WTA estimate that is lower than the estimated WTA for each subgroup separately (see Figure 1). This also holds when estimating more flexible specifications based on a mixed logit. The effect is most stark when the wage variation induced by the experimental design does not reach the region where respondents are approximately indifferent, and thus implied valuations are identified largely through extrapolation. We then use evidence from a meta-analysis of contingent valuation studies in environmental economics by Parsons and Myers (2017) to show that such designs arise frequently in practice.

While the most fundamental issue highlighted by our results is that working directly with the choice probability curve does not identify the distribution of underlying valuations, we also document the potential for large distortions in estimating the 50 percent take-up price by fitting a single logit to heterogeneous choice data. Even when each individual follows a correctly specified logit response rule, pooling across heterogeneous individuals generally produces a mixture of logits, which is not itself logit.[1] The pooled-logit-implied WTP may fall outside the range of individual WTPs. We present a series of examples demonstrating how this bias can arise from preference heterogeneity, heterogeneity in choice consistency, and the range of prices that respondents face; and reducing heterogeneity or reducing the share of noisy types need not reduce distortion. More generally, a pooled logit estimator converges to the single-logit curve that best approximates the true pooled curve under the experimental distribution of price differences. To describe how that approximation can fail, we first show that the distortion can be arbitrarily large, even when true WTP is bounded. We then isolate the role of experimental design: changing only the distribution of experimental price differences can move the pooled-logit-implied WTP by an arbitrarily large amount, even under homogeneous true WTP.

This paper brings an identification perspective to the extensive body of work on DCEs and stated-preference methodology. A large methodological literature examines how to estimate and report WTP within parametric discrete-choice frameworks, emphasizing that WTP is a derived object (often a ratio of coefficients) and studying estimation in preference space versus WTP space, as well as heterogeneity in choice consistency (Train and Weeks, 2005; Sonnier, Ainslie and Otter, 2007; Fiebig et al., 2010; Daly, Hess and Train, 2012).[2] We contribute to this line of work by clarifying what is identified by the pooled binary-choice curve itself, before these modeling choices are imposed. We also highlight how logit models produce design-dependent summaries of heterogeneous choice behavior in the sense of White (1982).

Our work closely relates to the literature on welfare analysis from discrete-choice data. Early studies by Small and Rosen (1981) and Hanemann (1984) provide foundations for conducting welfare analysis in discrete-choice models, including binary-response models commonly used in DCEs. More recent literature considers welfare analysis under weaker structural assumptions (see Bhattacharya 2024 for a recent overview). Bhattacharya (2015) and Bhattacharya (2018) recover welfare objects from choice probabilities, and Bhattacharya (2021) derives binary-choice restrictions that deliver robust bounds on demand and welfare. Kamat and Norris (2025) bring these ideas close to the present paper by studying bounds on average WTP when the support of observed prices is finite. Our paper highlights a distinct

---

[1] Of course, model-based approaches such as mixed logit and hierarchical Bayes recover WTP moments through distributional assumptions that may also be misspecified.

[2] See also applied guidance for DCE practice, e.g., Lancsar and Louviere (2008), Hauber et al. (2016), and Johnston et al. (2017).

source of partial identification based on the fact that pooling noisy choices does not recover the distribution of underlying valuations, even if the pooled curve is known perfectly.

The paper also relates more broadly to work on welfare identification from cross-sectional response data. Hausman and Newey (2016) show that cross sections identify the distribution of demand at each observed price-income pair, but not how the same individual would respond as prices or incomes change, so average consumer surplus is not generally point identified and must instead be bounded. The comparison is useful because it highlights a broader issue that is not specific to DCEs: in both papers, the observed response object is informative but too coarse for the welfare or valuation object of interest. In Hausman and Newey (2016), the gap comes from observing each individual at only one budget; here, it comes from averaging noisy threshold choices across individuals and then inverting the pooled curve. The central difficulty is one of aggregation: the data reveal an average response relationship, while the object of interest depends on underlying individual-level variation.

The paper also connects to the literature on nonparametric and partially identified discrete choice and demand, including work on set identification, nonparametric demand analysis, and robust welfare bounds (Chesher, Rosen and Smolinski, 2013; Kitamura and Stoye, 2018; Tebaldi, Torgovitsky and Yang, 2023; Compiani, 2022; Barseghyan et al., 2021; Cosaert and Demuynck, 2018). More broadly, the paper fits within the partial-identification tradition emphasized by Manski (2003) and surveyed by Molinari (2020). Our paper brings these ideas to binary-choice settings with pooled data. Because the bounds are written directly in terms of observed choice shares and a small number of economically interpretable restrictions on shocks, they provide a transparent link between assumptions on choice errors and conclusions about latent valuation objects.[3] This also makes the bounds straightforward to implement in applications. The same ideas may also be useful in other pooled binary-choice environments, such as randomized pricing, job amenity choices, and insurance take-up, where connecting observed take-up behavior to latent valuation objects presents similar challenges.

The paper proceeds as follows. Section 2 develops the general identification results. Section 3 derives bounds on the mean and quantiles of WTP. Section 4 provides an empirical illustration of the bounds and shows that conventional estimation approaches can produce WTP estimates outside the bounds. Section 5 illustrates the distortions that arise when estimating a logit model after pooling together data from respondents who individually follow correctly specified logit choice probabilities. Section 6 concludes.

---

[3] The role of our unlikely large shocks assumption is also conceptually related to the contamination and misclassification literature. In particular, allowing a small fraction of choices to be severely distorted by large shocks is similar in spirit to bounding the share of observations for which the observed binary outcome differs from the latent one; see, for example, Molinari (2008).

2 Pooling alone does not identify willingness to pay

2.1 Setup

Let $\Delta p \in \mathbb{R}$ denote the price difference between the alternatives with and without the attribute. Let $y_{it} \in \{0, 1\}$ indicate whether individual i selects the alternative that includes the attribute in choice situation t. Suppose each individual has a WTP parameter $\beta_i \in \mathbb{R}$, and choices follow the rule

$$y_{it} = \mathbb{1}_{\{\beta_i - \Delta p_{it} + \varepsilon_{it} \geq 0\}},$$

where ε_{it} represents an idiosyncratic preference shock, independent of β_i, with continuous, strictly increasing distribution function G. We impose a normalization on ε_{it} by setting its median to zero, i.e., $G(0) = \frac{1}{2}$. This formulation implies

$$q_{\beta_i}(\Delta p) := \Pr(y_{it} = 1 \mid \beta_i, \Delta p) = \Pr(\varepsilon_{it} \geq \Delta p - \beta_i) = 1 - G(\Delta p - \beta_i).$$

Defining $F(t) := 1 - G(-t)$, we obtain

$$q_{\beta_i}(\Delta p) := \Pr(y_{it} = 1 \mid \beta_i, \Delta p) = F(\beta_i - \Delta p). \tag{1}$$

Let B represent a random draw from the population distribution of individual WTP parameters β_i. Averaging over this distribution gives the pooled choice probability

$$m(\Delta p) := \mathbb{E}[q_B(\Delta p)] = \mathbb{E}[F(B - \Delta p)]. \tag{2}$$

This setup accommodates arbitrarily many responses per person and within-person price variation. With a panel structure, one recovers not only the pooled curve $m(\cdot)$ but also the joint distribution of $(y_{i,1}, \dots, y_{i,T_i})$ given $(\Delta p_{i,1}, \dots, \Delta p_{i,T_i})$. Pooling instead discards this joint structure and retains only the marginal relationship $m(\Delta p) = \Pr(y = 1 \mid \Delta p)$.

The defining feature of the pooling approach is that identification proceeds solely through the function $m(\cdot)$. With enough data, $m(\Delta p)$ is identified at the observed price differences, and with sufficiently broad support, one recovers the entire function. The choice rule $y_{it} = \mathbb{1}_{\{\beta_i + \varepsilon_{it} \geq \Delta p_{it}\}}$ implies

$$m(\Delta p) = \Pr(B + \varepsilon \geq \Delta p),$$

so $m(\cdot)$ depends only on the distribution of the *shock-contaminated WTP* $B + \varepsilon$. Identifying the aggregate relationship between price and choice probability suffices to identify any functional $T(m)$, including price elasticities, choice probabilities at specific price differences, or the price at which the pooled choice probability attains a given value. However, even with infinite data, pooling recovers only features of the curve $m(\cdot)$ and not distributional features of true WTP B, such as the mean, median, or tail probabilities. This is an interpretive issue that arises in practice beyond DCEs.[4]

2.2 Gap between pooled and true WTP can be arbitrarily large

Let b^p denote the price difference at which $m(b^p) = \frac{1}{2}$.[5] This pooled crossing point represents the median of shock-contaminated WTP $B+\varepsilon$ since $m(\Delta p) = \Pr(B+\varepsilon \geq \Delta p)$. The question is whether this can also be interpreted as the median of B itself.

Computing the pooled crossing point involves first averaging the individual choice probabilities to obtain the pooled curve $m(\Delta p) = \mathbb{E}[F(B - \Delta p)]$ and then inverting that pooled curve at $\frac{1}{2}$ to obtain $b^p = m^{-1}(\frac{1}{2})$. Computing quantiles of WTP involves inverting before averaging, and these two operations need not commute. The next theorem shows that even with infinite data and exact knowledge of $m(\cdot)$, the pooled crossing point can be arbitrarily far from the median of B unless the shock is degenerate.

Theorem 1. *For any $M > 0$, there exists a distribution of B such that $|b^p - \mathrm{Med}[B]| > M$, and there exists a distribution of B with finite mean such that $|b^p - \mathbb{E}[B]| > M$. Furthermore, $b^p = \mathrm{Med}[B]$ for all distributions of B if and only if $\varepsilon = 0$.*

When choice is not deterministic, inverting the pooled recovers a feature of the distribution of $B + \varepsilon$ rather than B. The proof in Appendix A.1 separates the pooled crossing from the median by placing more than half of the mass at a low value so that $\mathrm{Med}[B]$ stays fixed, and moving the remaining mass arbitrarily far to the right. This shifts the pooled curve enough to move its crossing point arbitrarily far away even though the median does not move. To separate the pooled crossing from the mean, place most mass at a low value and a smaller amount of mass at a very large value; the large support point drives $\mathbb{E}[B]$ arbitrarily far out, while the pooled crossing remains uniformly bounded. Only in the case of degenerate shocks does the pooled crossing coincide with the median of B over all distributions of B.

---

[4] For example, Ashraf, Berry and Shapiro (2010) uses an estimated take-up curve to construct a "willingness-to-pay distribution" and define WTP percentiles. In their Appendix Figure A2, the note explicitly describes a "willingness-to-pay distribution based on [an] estimated demand model" used to define these percentiles. The demand curve identifies $m(p) = \Pr(y = 1 \mid p)$ and therefore functionals of $m(\cdot)$. Inferring $\Pr(B \geq p)$ requires a model of how shocks affect choice. Otherwise, percentiles derived from the pooled curve are percentiles of shock-contaminated reservation values rather than of a stable individual-level WTP distribution. To be clear, none of this undermines the usefulness of the estimated demand curve in Ashraf, Berry and Shapiro (2010) for the core questions they study on screening and use. Our point is simply that whenever the pooled demand curve (or a parametric fit to it) is described as a WTP distribution, or when percentiles computed from it are interpreted as percentiles of a stable, individual-level WTP distribution, that interpretation relies on additional assumptions that are typically not stated.

[5] When this crossing point is not unique (e.g., to cover the case $\varepsilon \equiv 0$), we use $b^p = \inf\{p : \Pr(B + \varepsilon \leq p) \geq 1/2\}$.

As a consequence, the identified sets for mean and median WTP can be made arbitrarily wide (Corollary A.1). This follows because any model with pooled crossing b^p is observationally equivalent to a model in which WTP is degenerate at b^p. Thus, whenever b^p is far from the mean or median in one admissible model, the same pooled curve is compatible with another admissible model that sets that object equal to b^p.

Theorem A.1 in Appendix A.3 strengthens the argument further by showing that the failure is not limited to the mean and median: No non-constant real-valued functional of the WTP distribution is identified from the pooled curve alone.[6] Without making additional assumptions on the joint distribution of B and ε, multiple distinct marginal distributions of B remain observationally equivalent even if the pooled curve $m(\cdot)$ is known perfectly over the entire support of Δp.

This has implications for experimental design. If the goal is to learn the distribution of WTP with minimal assumptions, expanding price support in pooled estimation is not enough. Learning about WTP requires information that separates persistent heterogeneity from transitory shocks. One route exploits within-person variation. For an individual with response curve $q_i(\Delta p) = F(\beta_i - \Delta p)$, the normalization $F(0) = \frac{1}{2}$ implies $q_i^{-1}(\frac{1}{2}) = \beta_i$. Observations are most informative about β_i when $q_i(\Delta p)$ is near $\frac{1}{2}$. This sheds light on how adaptive designs, which concentrate price variation near each individual's indifference point, provide massive gains in learning about the WTP distribution.

If we additionally assume $\mathbb{E}[\varepsilon] = 0$, then $\mathbb{E}[B + \varepsilon] = \mathbb{E}[B]$, so pooling identifies the mean of B. The conclusion of Theorem 1 regarding the median, however, would not change since it only requires ε to have unbounded support. The claim that the mean of B is point identified when $\mathbb{E}[\varepsilon] = 0$, of course, relies on the maintained assumption that $m(\Delta p)$ is known perfectly over the entire support of Δp. We now turn to the more relevant case for empirical applications, in which $m(\cdot)$ is known only at a finite number of price differences.

3 Bounds on WTP

3.1 Quantile bounds under a finite set of observed prices

Suppose the pooled curve $m(\cdot)$ is known perfectly over a finite set of observed price differences $p_1 < \cdots < p_K$. In this case, we can obtain bounds on quantiles of B. Without any shocks, the median of B is simply bounded between the largest observed price difference at which the pooled choice probability exceeds $\frac{1}{2}$ and the smallest observed price difference at which it

---

[6] Of course, if the distribution of shocks is known, then one can recover the distribution of B by deconvolution. We return to this point in Section 6.1.

does not. Adding noise relaxes these bounds, but we can still obtain informative bounds on quantiles of B under mild restrictions on the shock distribution, namely that large shocks are unlikely.

To accommodate distributions that are not strictly monotone, define the quantile function as

$$Q_B(\tau) := \inf\{x : \tau \leq \Pr(B \leq x)\}$$

for $\tau \in (0,1)$.

Theorem 2. *Suppose there exist constants* $c \geq 0$ *and* $\alpha \in [0, \frac{1}{2})$ *such that* $\Pr(\varepsilon > c) \leq \alpha$ *and* $\Pr(\varepsilon < -c) \leq \alpha$*. Then for every* $\tau \in (\alpha, 1-\alpha)$*,*

$$\underline{Q}_B(\tau) \leq Q_B(\tau) \leq \overline{Q}_B(\tau),$$

where

$$\begin{aligned} \underline{Q}_B(\tau) &:= \sup\big\{p_j : m(p_j) > 1 - \tau + \alpha\big\} - c, \\ \overline{Q}_B(\tau) &:= \inf\big\{p_j : m(p_j) \leq 1 - \tau - \alpha\big\} + c. \end{aligned}$$

In addition, if $B \perp \varepsilon$ *and* $G(0) = \frac{1}{2}$*, then for every* $\tau \in (0,1)$*, we have* $\underline{Q}_B(\tau) \leq Q_B(\tau) \leq \overline{Q}_B(\tau)$*, where*

$$\begin{aligned} \underline{Q}_B(\tau) &:= \max\Big\{\sup\big\{p_j : m(p_j) > 1 - \tau(1-\alpha)\big\} - c, \sup\Big\{p_j : m(p_j) > 1 - \frac{\tau}{2}\Big\}\Big\}, \\ \overline{Q}_B(\tau) &:= \min\Big\{\inf\big\{p_j : m(p_j) \leq (1-\tau)(1-\alpha)\big\} + c, \inf\Big\{p_j : m(p_j) \leq \frac{1-\tau}{2}\Big\}\Big\}. \end{aligned}$$

By convention, $\sup \emptyset = -\infty$ and $\inf \emptyset = +\infty$, so the bounds are nontrivial only when the observed price grid contains cells that satisfy the relevant inequalities.[7]

The proof in Appendix B.1 highlights the two sources of partial identification. First, a finite set of observed prices localizes quantiles of the observed shock-contaminated WTP to the interval in which the pooled curve crosses the relevant threshold (e.g., 0.5 for the median). Second, the restriction that large shocks are unlikely provides a way to translate those bounds to quantiles of WTP. When $\alpha = 0$, shocks are bounded by c, so the median of B lies within c of the shock-contaminated WTP bound. In general, even without assuming independence, shocks that arbitrarily distort a share α of observations can cause a share of at most α to move across the relevant threshold, and thus the bounds are adjusted accordingly.

---

[7] Furthermore, under this convention, the result without assuming independence extends to all $\tau \in (0,1)$; when $\tau \leq \alpha$, the lower endpoint is vacuous, and when $\tau > 1-\alpha$ the upper endpoint is vacuous.

Assuming independence tightens the bounds because, instead of an unconditional mass of α potentially moving across the threshold in either direction, a share of at most α among the τ fraction of observations below the τ-quantile of B can have large shocks that push them above the threshold. Similarly, among the $1-\tau$ fraction of observations above the τ-quantile of B, a share of at most α can have large shocks that push them below the threshold. The median-zero normalization further tightens the bounds because, at any cutoff t, at least half of the observations with $B \leq t$ have $\varepsilon \leq 0$, so they also satisfy $B + \varepsilon \leq t$. Similarly, at least half of the observations with $B > t$ have $\varepsilon \geq 0$, so they also satisfy $B + \varepsilon > t$. The tightened bound takes the larger of the two lower bounds and the smaller of the two upper bounds.

The result also has immediate implications for experimental design. In particular, a finer price grid can tighten the first component of the bound in Theorem 2, but it cannot eliminate the second component.

3.2 Mean bounds under a finite set of observed prices

Unlike quantiles, which we bound based on the range of price differences at which the pooled curve crosses certain thresholds, the mean depends on the entire shape of the pooled curve and in particular its tails. We consider two natural ways to bound mean WTP: (i) restrictions on the tails of the underlying WTP distribution, and (ii) restrictions on the support of shock-contaminated WTP combined with the assumption of mean-zero shocks.

Bound on latent WTP. The first route to bounding mean WTP parallels Theorem 2 by invoking the assumption of unlikely large shocks. It adds two ingredients: step-function approximations to $m(\cdot)$, and a restriction on the tails of WTP outside a central interval.

Denote the observed prices by $p_1 < \cdots < p_K$, which define intervals $I_0 = (-\infty, p_1)$, $I_j = [p_j, p_{j+1})$ for $j = 1, \dots, K-1$, and $I_K = [p_K, \infty)$. For any price p, let $j(p)$ denote the unique index such that $p \in I_{j(p)}$. Define $\underline{m}(p) = \inf_{\tilde{p} \in I_{j(p)}} m(\tilde{p})$ and $\overline{m}(p) = \sup_{\tilde{p} \in I_{j(p)}} m(\tilde{p})$, so that $\underline{m}(p)$ and $\overline{m}(p)$ are step functions that bound $m(p)$ from below and above, respectively.

Theorem 3. *Suppose there exist constants* $c \geq 0$ *and* $\alpha \in [0, \frac{1}{2})$ *such that* $\Pr(\varepsilon > c) \leq \alpha$ *and* $\Pr(\varepsilon < -c) \leq \alpha$. *Suppose further that there exist* $\zeta, \eta \in [0, \infty)$ *and* $\ell, u \in \mathbb{R}$ *with* $\ell < u$ *such that* $\mathbb{E}[(\ell - B)_+] \leq \zeta$ *and* $\mathbb{E}[(B - u)_+] \leq \eta$, *where* $x_+ := \max\{x, 0\}$. *Then* $\mathbb{E}[B] \in [B_\ell, B_u]$, *where*

$$B_\ell := \ell + \int_{\ell+c}^{u+c} \max\{0, \underline{m}(t) - \alpha\} \, \mathrm{d}t - \zeta,$$

$$B_u := \ell + \int_{\ell-c}^{u-c} \min\{1, \overline{m}(t) + \alpha\} \, \mathrm{d}t + \eta.$$

In addition, if $B \perp \varepsilon$ *and* $G(0) = \frac{1}{2}$*, then* $\mathbb{E}[B] \in [B_\ell^\perp, B_u^\perp]$*, where*

$$B_\ell^\perp := \ell + \int_\ell^u \max\left\{0, \frac{\underline{m}(t+c) - \alpha}{1-\alpha}, 2\underline{m}(t) - 1\right\} \mathrm{d}t - \zeta,$$
$$B_u^\perp := \ell + \int_\ell^u \min\left\{1, \frac{\overline{m}(t-c)}{1-\alpha}, 2\overline{m}(t)\right\} \mathrm{d}t + \eta.$$

The bound reflects three sources of slack. The finite set of observed prices replaces $m(\cdot)$ with step functions that envelope it from below and above. The unlikely large shocks assumption translates information about shock-contaminated WTP back to information about B. Finally, the terms ζ and η account for the contribution of tails of WTP outside the interval $[\ell, u]$.

As before, assuming independence leads to tighter bounds by restricting the share of observations with large shocks based on the share of observations with WTP B above or below any given level. In particular, consider a cutoff t, and let q denote the share of observations with $B \geq t$. If the observed share with $B + \varepsilon \geq t + c$ is m, then the complementary observed share with $B + \varepsilon < t + c$ must contain a fraction of at least $(1 - \alpha)$ of the share with $B < t$, so $1 - m \geq (1 - \alpha)(1 - q)$, which gives $q \geq \frac{m-\alpha}{1-\alpha}$. Similarly, if the observed share with $B + \varepsilon \geq t - c$ is m, then this observed share must contain a fraction of at least $(1 - \alpha)$ of the share with $B \geq t$, so $m \geq (1 - \alpha)q$, which gives $q \leq \frac{m}{1-\alpha}$. The median-zero normalization further tightens the bounds because, if the observed share with $B + \varepsilon \geq t$ is m, then the complementary observed share with $B + \varepsilon < t$ must contain at least half of the share with $B < t$, so $1 - m \geq \frac{1}{2}(1 - q)$, which gives $q \geq 2m - 1$. Similarly, the observed share with $B + \varepsilon \geq t$ must contain at least half of the share with $B \geq t$, so $m \geq \frac{1}{2}q$, which gives $q \leq 2m$. As before, the tightened bound takes the larger of the two lower bounds and the smaller of the two upper bounds.

Imposing bounded support for WTP leads to the following simple special cases of interest.

Corollary 1. *If* $\ell \leq B \leq u$ *in addition to the assumptions of Theorem 3, then* $\mathbb{E}[B] \in [B_\ell, B_u]$*, where*

$$B_\ell := \ell + \int_{\ell+c}^{u+c} \max\{0, \underline{m}(t) - \alpha\} \ \mathrm{d}t,$$
$$B_u := \ell + \int_{\ell-c}^{u-c} \min\{1, \overline{m}(t) + \alpha\} \ \mathrm{d}t.$$

Under the additional assumption that $|\varepsilon| \le c$*,*

$$B_\ell := \ell + \int_{\ell+c}^{u+c} \underline{m}(t)\ \mathrm{d}t,$$
$$B_u := \ell + \int_{\ell-c}^{u-c} \overline{m}(t)\ \mathrm{d}t.$$

In addition, if $B \perp \varepsilon$ *and* $G(0) = \frac{1}{2}$*, then* $\mathbb{E}[B] \in [B_\ell^\perp, B_u^\perp]$*, where*

$$B_\ell^\perp := \ell + \int_\ell^u \max\left\{0, \frac{\underline{m}(t+c) - \alpha}{1-\alpha}, 2\underline{m}(t) - 1\right\}\ \mathrm{d}t,$$
$$B_u^\perp := \ell + \int_\ell^u \min\left\{1, \frac{\overline{m}(t-c)}{1-\alpha}, 2\overline{m}(t)\right\}\ \mathrm{d}t.$$

Under the additional assumption that $|\varepsilon| \le c$*,*

$$B_\ell^\perp := \ell + \int_\ell^u \max\{\underline{m}(t+c), 2\underline{m}(t) - 1\}\ \mathrm{d}t,$$
$$B_u^\perp := \ell + \int_\ell^u \min\{\overline{m}(t-c), 2\overline{m}(t)\}\ \mathrm{d}t.$$

The last special case parallels the bound on median WTP from Theorem 2 with $\alpha = 0$. Observing a finer set of prices tightens the step-function approximation, and the bounded c on the size of the shocks determines how far information about $B + \varepsilon$ must be shifted to recover information about B.

Bound on shock-contaminated WTP with mean-zero errors. The second route for bounding WTP exploits the fact that $\mathbb{E}[B] = \mathbb{E}[B + \varepsilon]$ when $\mathbb{E}[\varepsilon] = 0$. Restricting the support of shock-contaminated WTP then delivers a straightforward bound on mean WTP:

Theorem 4. *If* $\mathbb{E}[\varepsilon] = 0$ *and* $\operatorname{supp}(B + \varepsilon) \subseteq [L, U]$*, then* $\mathbb{E}[B] \in [B_L, B_U]$*, where*

$$B_L := L + \int_L^U \underline{m}(t)\ \mathrm{d}t,$$
$$B_U := L + \int_L^U \overline{m}(t)\ \mathrm{d}t.$$

When shocks are bounded, support assumptions on B and on $B + \varepsilon$ imply one another.[8]

---

[8] In particular, suppose $|\varepsilon| \le c$. If $\operatorname{supp}(B) \subseteq [\ell, u]$ and $|\varepsilon| \le c$, then $\ell - c \le B + \varepsilon \le u + c$. Hence $\operatorname{supp}(B + \varepsilon) \subseteq [\ell - c, u + c]$. Conversely, if $\operatorname{supp}(B + \varepsilon) \subseteq [L, U]$ and $|\varepsilon| \le c$, then for every realization, $L - c \le B \le U + c$.

The next result compares the two support-based mean bounds when one support restriction is induced from the other.

Corollary 2. *Suppose* $|\varepsilon| \leq c$ *and* $\mathbb{E}[\varepsilon] = 0$*.*

1. *If* $\operatorname{supp}(B+\varepsilon) \subseteq [L, U]$*, then the bound* $[B_L, B_U]$ *from Theorem 4 is weakly tighter than the bound* $[B_\ell, B_u]$ *from Corollary 1 based on the induced restriction* $\operatorname{supp}(B) \subseteq [L-c, U+c]$*.*
2. *If* $\operatorname{supp}(B) \subseteq [\ell, u]$*, then the bound* $[B_\ell, B_u]$ *from Corollary 1 and the bound* $[B_L, B_U]$ *from Theorem 4 based on the induced restriction* $\operatorname{supp}(B+\varepsilon) \subseteq [\ell-c, u+c]$ *satisfy* $B_L - B_\ell \in [-c, c]$ *and* $B_U - B_u \in [-c, c]$*. Moreover, the bounds are not uniformly ordered.*

Part 1 compares the direct bound on $B+\varepsilon$ from Theorem 4 to the bound on B from Corollary 1 when the support on latent WTP is induced from the support on shock-contaminated WTP. In this case, Theorem 4 unambiguously yields tighter bounds because it exploits the assumption of mean-zero shocks.

Part 2 considers the reverse case. In this case, the bound on latent WTP uses a narrower support restriction, but it pays the cost of translating from $B+\varepsilon$ back to B. In general, neither effect dominates the other, and their relationship depends on the shape of the pooled curve near the boundaries.

A three-price example suffices to demonstrate this comparison. Consider the set of observed prices $\{0, 1, 2\}$, and suppose $\operatorname{supp}(B) \subseteq [0, 2]$. Fixing $c = 1$, the induced support for shock-contaminated WTP is $[-1, 3]$.

The bound on latent WTP is

$$\left[0 + \int_1^3 \underline{m}(t)\ \mathrm{d}t, 0 + \int_{-1}^1 \overline{m}(t)\ \mathrm{d}t\right].$$

Since the step function that bounds $m(p)$ from below on the observed price grid $\{0, 1, 2\}$ satisfies $\underline{m}(t) = m(2)$ for $t \in [1, 2)$ and $\underline{m}(t) = 0$ for $t \in [2, 3)$, we have $\int_1^3 \underline{m}(t)\ \mathrm{d}t = m(2)$. Similarly, since the step function that bounds $m(p)$ from above satisfies $\overline{m}(t) = 1$ for $t \in [-1, 0)$ and $\overline{m}(t) = m(0)$ for $t \in [0, 1)$, we have $\int_{-1}^1 \overline{m}(t)\ \mathrm{d}t = 1 + m(0)$.

The direct bound on shock-contaminated WTP under the induced support $[-1, 3]$ is

$$\left[-1 + \int_{-1}^3 \underline{m}(t)\ \mathrm{d}t, -1 + \int_{-1}^3 \overline{m}(t)\ \mathrm{d}t\right].$$

Since the step function that bounds $m(p)$ from below satisfies $\underline{m}(t) = m(0)$ for $t \in [-1, 0)$, $\underline{m}(t) = m(1)$ for $t \in [0, 1)$, $\underline{m}(t) = m(2)$ for $t \in [1, 2)$, and $\underline{m}(t) = 0$ for $t \in [2, 3)$, we have $\int_{-1}^{3} \underline{m}(t)\ \mathrm{d}t = m(0) + m(1) + m(2)$. Likewise, since the step function that bounds $m(p)$ from above satisfies $\overline{m}(t) = 1$ for $t \in [-1, 0)$, $\overline{m}(t) = m(0)$ for $t \in [0, 1)$, $\overline{m}(t) = m(1)$ for $t \in [1, 2)$, and $\overline{m}(t) = m(2)$ for $t \in [2, 3)$, we have $\int_{-1}^{3} \overline{m}(t)\ \mathrm{d}t = 1 + m(0) + m(1) + m(2)$.

If $m(0) = 0.9$, $m(1) = 0.8$, and $m(2) = 0.4$, then the first bound is $[0.4, 1.9]$ and the second one is $[1.1, 2.1]$. If instead $m(0) = 0.4$, $m(1) = 0.3$, and $m(2) = 0.2$, then the first bound is $[0.2, 1.4]$ and the second one is $[-0.1, 0.9]$. This illustrates how neither interval uniformly dominates the other. Thus, when the support of B is the primitive assumption, intersecting the two can yield strictly tighter bounds.

3.3 Multiple attributes

The results derived above are stated for the case of a single attribute. While this case is common in practice, many applications involve multi-dimensional attributes. Suppose each individual has a vector of WTP parameters $W_i \in \mathbb{R}^d$, and choices satisfy

$$y_{it} = \mathbb{1}_{\{W_i^\top \Delta x_{it} - \Delta p_{it} + \varepsilon_{it} \geq 0\}},$$

where $\Delta x_{it} \in \mathbb{R}^d$ represents the vector of non-price attribute differences, and $\Delta p_{it} \in \mathbb{R}$ denotes the price difference. The logic of our bounds extends directly to such settings.

Consider the subsample of choice situations with $\Delta x_{it} = a$ for some $a \in \mathbb{R}^d$. On this subsample, the pooled choice curve is

$$\begin{aligned} m_a(p) &:= \Pr(y_{it} = 1 \mid \Delta x_{it} = a, \Delta p_{it} = p) \\ &= \mathbb{E}[F(a^\top W - p)] \\ &= \Pr(a^\top W + \varepsilon \geq p). \end{aligned}$$

Therefore, for a fixed $\Delta x_{it} = a$, the problem is exactly the same as in the one-dimensional setting, with the WTP B replaced by $a^\top W$.[9] Applying the bounds from the previous sections to $m_a(\cdot)$ then yields corresponding bounds for $\mathbb{E}[a^\top W]$ and $Q_{a^\top W}(\tau)$.

Letting $\{a_j\}_{j \in 1, \ldots, J}$ denote the attribute-difference vectors considered in the design, we obtain a collection of linear restrictions $L_j \leq a_j^\top \mu \leq U_j$ on the mean WTP vector $\mu := \mathbb{E}[W]$. If the design includes a direction that isolates a single attribute (i.e., some a_j is a standard

---

[9] To interpret the bounds as applying to the population distribution of W_i, we view the design cells $(\Delta x_{it}, \Delta p_{it})$ used to form $m_a(\cdot)$ as being assigned independently of (W_i, ε_{it}), with a shock distribution that does not vary across design cells.

basis vector), then we have a direct bound on the mean WTP for that attribute. Such componentwise bounds can potentially be tightened based on the bounds obtained from other attribute differences in the design. In general, the mean WTP vector lies in the intersection of the half-spaces defined by these linear restrictions:

$$\mu \in \bigcap_{j=1}^{J} \left\{ \mu \in \mathbb{R}^d : L_j \leq a_j^\top \mu \leq U_j \right\}.$$

For any standard basis vector included in the design, we obtain bounds on quantiles of the WTP for the corresponding attribute. Theorem B.1 in Appendix B.6 provides a formal statement along with a proof. These quantile bounds, however, are unable to leverage additional linear restrictions across dimensions since quantiles do not combine linearly across dimensions.

This has important implications for experimental design in multi-attribute settings. Bounds on mean WTP for a given attribute can be tightened by including additional designs that involve that attribute, even if those designs do not isolate it. Obtaining nontrivial bounds on the median WTP under our approach, by contrast, requires a design that isolates that attribute.

4 Empirical illustration of bounds and pooling bias

4.1 Bounds using pooled choice probabilities

This section uses data from Mas and Pallais (2019) to demonstrate the performance of the bounds derived in the previous section. Mas and Pallais (2019) use a discrete-choice experiment over job offers to estimate the marginal value of time, or the compensating wage differentials workers require for working a job that requires an additional 5 hours per week.[10] This allows for estimation of a willingness to accept (WTA) the higher-hours position, rather than a willingness to pay. Accordingly, the choice probability is an increasing function of effective wage in this application.

The data consist of multiple treatment conditions, corresponding to varying the hours associated with the longer-hours job between 10 and 50 hours per week. Their experimental design varies the wage associated with a longer-hours job to trace out how the probability of choosing that job changes with its effective wage. The design includes a finite number of price differences, typically at $2 increments. The data contain a measure of respondents'

---

[10] See Garro-Marín, Thakral and Tô (2025) for a model of work hours as a job amenity in a compensating differentials framework.

inattentiveness based on whether they correctly recall the number of hours they chose. We focus on two subsamples: the preferred sample from Mas and Pallais (2019) consisting of unemployed jobseekers, and the sample of jobseekers who did not apply to the advertised position.

The estimated inattention rates in the two samples are 10.03 percent and 6.73 percent, respectively.[11] When computing bounds, we set α from Theorem 2 and Theorem 3 to these values, and we set c to the \$2 increment between observed price differences. This choice allows a fraction α of responses to be displaced by more than one observed wage cell in either direction, but rules out such large errors for the remaining $1 - \alpha$ fraction of responses. Thus, the inattention rate α governs how often consequential mistakes are allowed, while c governs the threshold for determining such mistakes. We also impose the independence assumption and median-zero normalization, which tighten the bounds from Theorems 2 and 3.

For Theorems 3 and 4, we assume a support of $[0, 35]$, except in the overtime treatments (50 vs. 45 and 45 vs. 40), where we set the upper bound to 50.[12] To compute the bounds, we estimate the pooled choice probability at each observed effective wage and apply the monotone rearrangement procedure of Chernozhukov, Fernández-Val and Galichon (2010) by sorting the estimated choice probabilities to convert these choice probabilities to a monotonic function, which is necessary for computing quantile bounds. This imposes the economically natural restriction that the probability of choosing the longer-hours job is weakly increasing in its effective wage. We use the monotone fitted values instead of observed choice probabilities when computing bounds, but in Table A1 we report more conservative bounds on the mean based directly on the (possibly non-monotone) observed choice probabilities.

For each sample and each treatment, we estimate a pooled logit regression to compute the mean/median WTA.[13] We also estimate a two-type finite-mixture logit and compute the median WTA using the posterior probabilities of the two latent types, and we compute average WTA as a weighted average of the implied WTAs for the two types. A two-type mixture

---

[11] To estimate the inattention rates, following Mas and Pallais (2019), we add the share of respondents who recall an incorrect choice divided by 0.9 (to account for the fact that this measure only captures 90 percent of inattentive types since there is a 10 percent chance of answering the question correctly by random chance) to the share of respondents who do not remember their choice.

[12] Figure 1 of Mas and Pallais (2019), which displays the share of applicants who choose the longer-hours job as a function of the effective wage, shows a nearly 100 percent choice probability by \$35 in the non-overtime treatments and around \$45 in the overtime treatments. For each set of treatments, we assume the same upper bound when operationalizing both theorems so that the differences between the bounding approaches come from their respective identifying assumptions rather than from different parameter choices.

[13] Our pooled logit estimates coincide with those reported by Mas and Pallais (2019) in all cases except for the overtime treatments. They implement a procedure to anchor the estimated intercepts in those treatments to ensure that the fitted curve is increasing in effective wage. Our estimates do not reflect this correction because our goal is to use these data to illustrate how standard pooled parametric models compare to the bounds rather than to replicate the implementation choices of the original paper.

is a natural mixed-logit specification in this setting because each respondent contributes only one binary choice, and estimating a standard continuous-mixture specification leads to convergence issues in this setting.[14]

The non-parametric bounds in Table 1 reveal several important patterns. Most notably, the pooled and mixed logit WTA estimates may fall outside the bounds. The fact that this occurs despite the bounds being quite wide suggests that (i) the bounds are not too wide to be informative, and (ii) the pooled and mixed logit estimates rely heavily on assumptions that the data do not support. In addition, the bounds following the approach of Theorem 3 are wider than those following the approach of Theorem 4, although Corollary 2 highlights that dominance does not necessarily hold in general. Despite this, the wider bounds following Theorem 3 still suffice to rule out the mixed logit mean WTA estimate in several cases.

The pooled or mixed logit median WTA estimate falls outside the bounds from Theorem 2 in six instances among the non-overtime treatments. In the sample of unemployed jobseekers (panel A), the median of the two-type mixture logit exceeds the upper bound in the 30 vs. 25 treatment. In the sample of jobseekers who did not apply for the position (panel B), the pooled WTA exceeds the upper bound in the 10 vs. 5 treatment. The median of the two-type mixture logit exceeds the upper bound in the 10 vs. 5 treatment, and it falls below the lower bound in the 40 vs. 35, 25 vs. 20, and 15 vs. 10 treatments. The overtime treatments typically do not provide informative lower bounds on median WTA in the sample of unemployed jobseekers because the share choosing the longer-hours job at the lowest effective wage exceeds 50 percent; we return to this issue in Section 4.2.

For mean WTA, at least one point estimate falls outside the bounds from Theorem 4 in all of the non-overtime treatments. In the sample of unemployed jobseekers (panel A), the pooled logit mean falls below the lower bound in the 30 vs. 25, 25 vs. 20, 20 vs. 15, and 15 vs. 10 treatments. The two-type mixture logit mean exceeds the upper bound in the 40 vs. 35 and 15 vs. 10 treatments, and falls below the lower bound in the 20 vs. 15 treatment.[15] The sample in panel B exhibits a similar pattern, with the pooled logit mean falling below the lower bound in the 25 vs. 20, 20 vs. 15, 15 vs. 10, and 10 vs. 5 treatments, and the two-type mixture logit mean falling below the lower bound in the 40 vs. 35 and 15 vs. 10 treatments, and exceeding the upper bound in the 35 vs. 30 and 30 vs. 25 treatments. Even if we use the wider bounds based on Theorem 3, the two-type mixture logit mean still falls outside

---

[14] With only a single choice per respondent, a standard continuous random-coefficients mixed logit is not anchored by within-person substitution patterns and depends substantially on functional-form assumptions.

[15] Additionally, in the 45 vs. 40 treatment, the pooled logit mean exceeds the upper bound, but the best-fit relationship for these data has the opposite of the expected sign. Mas and Pallais (2019) devise a strategy that anchors the intercept to ensure that the fitted curve is increasing in effective wage, which we do not implement here.

the bounds in the 35 vs. 30, 30 vs. 25, and 15 vs. 10 treatments. Across both samples, the pooled logit mean violations tends to fall below the lower bound, while the two-type mixture logit mean falls outside the bounds in either direction.

The logit estimates from Mas and Pallais (2019) imply that the value of non-work time, or the opportunity cost of a full-time job (e.g., due to lost leisure and household production), is 58.3 percent of pre-tax earnings.[16] However, the bounds on the marginal value of time from Table 1 imply a lower bound of 57.3 percent and an upper bound of 75.3 percent for the value of non-work time. Adjusting their estimates that fall outside the bounds to the nearest endpoint increases the implied value of non-work time by 3.9 percent.

There are several reasons why the logit estimates may fall outside the bounds. One interpretation is that the pooled parametric models are misspecified. Another possibility is that the bounds are too tight because of our implementation assumptions.

Specifically, the use of a monotone fit to the pooled choice probabilities may lead to the bounds being misplaced or otherwise too narrow. To address this concern, we compute a more conservative version based directly on the raw choice probabilities. Let $\hat{m}_j$ denote the observed choice share at price p_j, for $j = 1, \dots, K$. Let $\mathcal{G}$ denote the set of weakly increasing functions $g \colon \mathbb{R} \to \mathbb{R}$, and let

$$\begin{aligned} \mathcal{G}_- &:= \{g \in \mathcal{G} \mid g(p_j) \le \hat{m}_j,\ j = 1, \dots, K\} \\ \mathcal{G}_+ &:= \{g \in \mathcal{G} \mid g(p_j) \ge \hat{m}_j,\ j = 1, \dots, K\} \end{aligned}$$

be the subsets that lie weakly below and weakly above the raw shares at every observed price, respectively. Define

$$\begin{aligned} \underline{m}^c(p) &:= \sup_{g \in \mathcal{G}_-} g(p) \\ \overline{m}^c(p) &:= \inf_{g \in \mathcal{G}_+} g(p). \end{aligned}$$

as the highest weakly increasing function that never exceeds the raw shares, and the lowest weakly increasing function that is never below them, respectively. Equivalently, at the observed prices, $\underline{m}^c(p_j) = \min_{\ell \ge j} \hat{m}_\ell$ and $\overline{m}^c(p_j) = \max_{\ell \le j} \hat{m}_\ell$. We then use the corresponding step functions on the intervals $I_0, \dots, I_K$ as the lower and upper extremes of all weakly increasing fits consistent with the raw shares, rather than a single monotonic fit, when computing the conservative bounds. The results similarly show that pooled estimates can still fall outside the bounds (Table A1). In particular, among the non-overtime treatments, the pooled logit estimate remains below the conservative lower bound for the 30 vs. 25 hour treatment in

---

[16] This is calculated following Equation (2) in Mas and Pallais (2019).

the unemployed sample. The two-type mixture mean also remains outside the conservative latent-B bounds in the 35 vs. 30, 30 vs. 25, and 15 vs. 10 treatments.

Overly restrictive assumptions on the shock distribution may also lead to bounds that are too tight. If the logit fit implies shocks larger than those allowed by our choice of c, α, and the support bounds, then one either concludes that the logit provides a poor description of choice behavior under a reasonable calibration, or that the assumption on the errors unfairly rules out plausible choice patterns. While the economic justification for our choice of parameters suggests the former interpretation, it does not definitively rule out the latter. Thus, we explore the sensitivity of our conclusions to weakening the assumptions on the parameters used to compute the bounds. When the pooled logit estimate falls short of the Theorem 4 bounds, expanding the lower bound requires allowing for large negative values in the assumed support of shock-contaminated WTA. Even if we assume that people may accept an effective wage rate as low as −\$7 on the marginal 5 extra hours of the 25-hour job relative to the 20-hour job, the pooled logit mean WTA still falls below the lower bound. This is implausible since a negative effective wage means that someone would prefer to work the longer-hours job for lower total compensation than the shorter-hours job. The same logic also applies to the wider bounds from Theorem 3. When the mixed logit estimates fall short of these bounds, encompassing these estimates requires allowing for negative effective wages. In particular, in the 15 vs. 10 treatment, even if we quadruple both the inattention rate and the threshold for large errors, the logit estimate only falls within the bound if we allow for an effective wage of −\$4. When the mixed logit estimate exceeds the bound, as in the 35 vs. 30 treatment, allowing for effective wages as high as \$40 and doubling the threshold for large errors requires nearly doubling the rate of large errors to encompass the estimate within the bound. Aside from this suggestive evidence that the bounds are not unreasonably narrow, we provide direct evidence in favor of the model-misspecification interpretation in the next section.

4.2 Pooling bias due to misspecification

First, we show that pooling together attentive and non-attentive types in the Mas and Pallais (2019) data leads to a pooled estimate of the premium required to accept a job with longer hours that is lower than the estimated premium for each group separately. The issue stems from the fact that, when considering tradeoffs that involve working more than 40 hours, the support of wages in their experimental design does not include the region where respondents are approximately indifferent. Then, to broaden the scope of this conclusion, we draw from a recent meta-analysis of contingent-valuation studies in environmental economics by Parsons and Myers (2017) to document the prevalence of similar issues in practice.

We focus on the treatments in Mas and Pallais (2019) where respondents are asked to choose between one job that requires 45 hours of work per week and another that requires 50 hours. The experimental design varies the wage associated with the longer-hours job to trace out how the probability of choosing the 50-hour job changes with its effective wage. As the authors note, the experimental design covers a limited range of effective wages because the Fair Labor Standards Act "requires paying hourly workers an overtime premium of at least 1.5 times their hourly pay for hours above 40." Specifically, even at the lowest effective wage that could be offered, the share choosing the longer job exceeds one-half. As a result, the data do not directly reveal the effective wage at which the longer and shorter jobs are chosen with equal probability, and determining that indifference point involves an extrapolation.

We reanalyze these data by separating the sample into two types based on their level of attentiveness. Define attentive types as those who recall their choice correctly when asked to recall which job option they chose, which comprise 88 percent of the sample. Analyzing the attentive types and non-attentive types separately, the model implies WTAs of \$25.14 for attentive types and \$27.64 for non-attentive types, as Figure 1 shows. As expected, the figure additionally shows a steeper change in choice probabilities as a function of effective wages for attentive types. Due to the difference in scale parameters across the two types, pooling leads to a WTA estimate of \$25.00 in the full sample, which is below the WTAs for each type.

The same phenomenon occurs when separating the sample by application status. Those who do not apply for the job exhibit less choice consistency than those who do apply. The estimated WTAs for these groups (\$27.27 and \$25.32, respectively) again exceed the pooled WTA.

The mixed-logit specification changes the WTA estimates considerably, which itself highlights the sensitivity of extrapolated valuations to functional-form assumptions, but leaves our qualitative conclusion about pooling-related distortions unchanged. The estimated WTA in the pooled sample increases from \$25.00 to \$29.785. For the attentiveness split, the estimated average WTA is \$29.61 for attentive respondents and \$30.66 for non-attentive respondents. For the application-status split, the estimated average WTA is \$31.85 for those who do not apply and \$29.789 for those who apply, both of which exceed the pooled WTA. This demonstrates that allowing for discrete latent heterogeneity as in a mixed logit specification does not eliminate pooling distortions.

The pooling distortion that arises in this example occurs due to the limited range of effective wages in the experiment. While this occurs due to a legal constraint in the Mas and Pallais (2019) setting, this example exhibits a broader pattern that appears in many other settings as well. If the share willing to pay the highest observed price exceeds 50 percent, then the elicitation design does not reach the point at which respondents are equally likely to say yes or no. In that case, inferring the implied valuation necessarily involves extrapolating beyond the observed support of the data as in the example of choosing between working 45 hours and 50 hours at the effective wage rates offered by Mas and Pallais (2019). In addition to creating scope for design-driven distortions, such extrapolations appear quite sensitive to heterogeneity in choice inconsistency.

To investigate these issues, we draw upon a meta-analysis of contingent-valuation studies published in eight leading environmental economics journals by Parsons and Myers (2017). Figure 2 plots the share of respondents for each study in their sample who indicate a willingness to pay at the highest price included in the design. In 13 of the 86 studies, this exceeds 50 percent. That means that, in a substantial part of the stated-preference literature, the key WTP object is identified largely through extrapolation of the assumed model beyond the support of the design.[17] More generally, many designs induce limited variation in choice probabilities and remain above the region where responses would be well-approximated near a choice probability of 50 percent.

5 Pooling as projection under misspecification

5.1 Individual choice and the pooled curve

Consider a binary choice experiment in which individuals make a series of choices between pairs of alternatives that differ in their price and the presence of an attribute. Assume that utility from an alternative j with price p_j and attribute $a_j \in [0,1]$ takes the form $U_i(p,a) = -p + \beta_i \mathbb{1}_{\{a=1\}} + \varepsilon_i^j$, where the i.i.d. error terms ε_i^j follow a Type-I extreme value distribution. This specification normalizes the coefficient on price to -1 so that β_i represents individual i's willingness to pay (WTP) for the attribute and ε_i^j represents idiosyncratic utility shocks in the same unit as price.

The difference in utility between two alternatives $(p_0, 0)$ and $(p_1, 1)$ then takes the form

$$\Delta U_i = -\Delta p + \beta_i + \Delta\varepsilon_i,$$

where $\Delta p = p_1 - p_0$, and the errors $\Delta\varepsilon_i$ follow a logistic distribution with scale parameter σ_i. A higher value of the scale parameter corresponds to greater choice error or inconsistency.

---

[17] In these applications, unlike in the Mas and Pallais (2019) data, the choice probability is a decreasing function of price, so if more than half of respondents are willing to pay the highest price, then the design does not reach the point at which respondents are equally likely to say yes or no. In such cases, the upper bound on median WTP in Theorem 2 is infinite.

The logit cumulative distribution function (CDF) $F(z) = \frac{1}{1+\exp(-z)}$ provides the probability that individual i chooses the alternative with the attribute:

$$\Pr_i((p_1, 1) \mid \{(p_0, 0), (p_1, 1)\}) = \frac{1}{1 + \exp[(\Delta p - \beta_i)/\sigma_i]}. \tag{3}$$

In our single-logit specification with the price coefficient normalized to -1, the attribute coefficient β is in price units and coincides with the price difference at which the fitted choice probability equals 0.5.

Let μ denote the population distribution of price differences Δp induced by the experimental design (equivalently, the probability limit of the empirical distribution of pooled Δp's). Let the true pooled choice probability be $m(\Delta p) := \Pr(y = 1 \mid \Delta p)$. That is, $m(\cdot)$ represents the actual aggregate relationship between price differences and choice probabilities in the population. Consider fitting a single logit model with willingness to pay β and scale σ, i.e., $\Pr_{\beta,\sigma}(y = 1 \mid \Delta p) = F\left(\frac{\beta - \Delta p}{\sigma}\right)$, and estimating (β, σ) by maximum likelihood using the pooled data $\{y_n, \Delta p_n\}_{n=1}^{N}$.

Under misspecification, when heterogeneity makes the pooled choice probability non-logit, the pooled maximum likelihood estimator (MLE) converges to pseudo-true parameters $(\beta^*(\mu), \sigma^*(\mu))$: the logit curve that best approximates the true pooled choice probability under the experimental distribution μ of price differences. Equivalently, it selects the member of the logit family that minimizes the Kullback-Leibler divergence from the true Bernoulli model generated by $m(\cdot)$. Formally, given a design μ, define

$$Q_\mu(\beta, \sigma) := \int \left[m(\Delta p) \log F\left(\frac{\beta - \Delta p}{\sigma}\right) + (1 - m(\Delta p)) \log\left(1 - F\left(\frac{\beta - \Delta p}{\sigma}\right)\right)\right] d\mu(\Delta p). \tag{4}$$

This criterion weights each price difference Δp according to μ and evaluates how well the candidate logit curve matches the true pooled probability $m(\Delta p)$ at that point. A *pseudo-true pooled logit* under design μ is any maximizer

$$(\beta^*(\mu), \sigma^*(\mu)) \in \underset{\beta \in \mathbb{R}, \sigma > 0}{\arg\max}\, Q_\mu(\beta, \sigma).$$

Under standard regularity conditions for extremum estimators, the pooled MLE $(\widehat{\beta}, \widehat{\sigma})$ converges in probability to $(\beta^*(\mu), \sigma^*(\mu))$ (White, 1982; Newey and McFadden, 1994).[18] We refer to $\beta^*(\mu)$ as the *pooled-logit-implied WTP* under design μ.

---

[18] For logit in discrete-choice settings, see also Train (2009, Ch. 8).

5.2 Illustrative examples

Consider a population that consists of two types of individuals, a share π of type L and a share $1-\pi$ of type H, with WTPs β_L and β_H satisfying $\beta_L < \beta_H$ and error scale parameters σ_L and σ_H. The true population choice probability when facing $(p_0, 0)$ and $(p_1, 1)$ is then given by the mixture

$$m(\Delta p) = \pi F\left(\frac{\beta_L - \Delta p}{\sigma_L}\right) + (1-\pi) F\left(\frac{\beta_H - \Delta p}{\sigma_H}\right). \tag{5}$$

Suppose a researcher pools data from both types and estimates the WTP $\widehat{\beta}$ and scale parameter $\widehat{\sigma}$ from a single logit model, where the price coefficient remains normalized to -1. We provide a series of examples that illustrate the mismatch between the pooled estimate $\widehat{\beta}$ and the individual WTPs β_L and β_H.[19] The pooled logit specification may fit well in the tails of the choice probability function, or in the marginal region where respondents are close to indifferent and small price changes can flip their choices, depending on the price range in the experiment. However, it often fails to fit well across the entire range of price differences. This stems from the simple fact that the mixture of two logits does not follow a logit distribution. The examples below highlight the resulting bias, focusing on three sources of distortion: the primitives (WTPs and scale parameters), the population composition (shares of types), and the experimental design (the set of prices shown to respondents). In all examples except the last, we hold the design fixed, with prices drawn uniformly from $[0, 1]$.

5.2.1 Heterogeneity-driven distortion

Heterogeneous choice error alone can generate bias, even when all individuals have the same WTP. Consider two types with the same WTP $\beta_L = \beta_H = 0.75$, different scale parameters $\sigma_L = 0.5$ and $\sigma_H = 0.05$, and equal population shares $\pi = 0.5$. Every individual has the same true WTP of 0.75, as reflected in the fact that both types choose the alternative containing the attribute with probability 0.5 at price difference $\Delta p = 0.75$ (Figure 3).

Nevertheless, the pooled logit does not recover this value. A single logit curve cannot simultaneously capture both the thick tails arising from the high-σ type and the sharp change in choice probability arising from the low-σ type. The price variation in the design forces the pooled logit to accommodate the high-σ type's thicker tails, pushing the WTP estimate upward. The resulting pooled estimate of $\widehat{\beta} = 0.94$ overstates the true mean WTP

---

[19] We report pooled estimates from samples of 500,000 choices to numerically approximate the pseudo-true parameters.

by 25 percent.

WTP heterogeneity alone can generate bias, even when all individuals have the same choice error. For a particularly stark example, consider two types with different WTPs, $\beta_L = -0.75$ and $\beta_H = 0.75$, a common scale parameter $\sigma_L = \sigma_H = 0.05$, and a smaller share of low-WTP types $\pi = 0.25$. Both types exhibit the same degree of choice consistency, so the only source of heterogeneity is in preferences. The type-specific logit curves therefore have the same shape but are shifted horizontally relative to one another (Figure 4).

Yet the pooled logit implies a WTP far from $[\beta_L, \beta_H]$. The mixture probability function contains an approximately flat region in which the population choice probability stays near 0.75 over a wide range of Δp, reflecting the fact that the high-WTP type chooses the attribute with near certainty while the low-WTP type chooses not to have the attribute with near certainty. A single logit curve cannot reproduce this plateau with any choice of (β, σ). The pooled logit approximates the mixture with a curve that implies choice probabilities above 0.5 for the entire range of prices between the WTPs of the low and high types. As a result, the pooled estimate $\hat{\beta} = 1.18$ lies above even the high type's WTP $\beta_H = 0.75$.

Reducing WTP heterogeneity can increase bias. Consider two environments that differ only in the WTP of the low type, with scale heterogeneity given by $\sigma_L = 0.5$ and $\sigma_H = 0.05$, equal population shares $\pi = 0.5$, and the WTP of the high type fixed at $\beta_H = 0.75$. Let the WTP of the low type vary from $\beta_L = 0$ in the first environment to $\beta_L = 0.75$ in the second (Figure 5).

In the first environment, with greater WTP heterogeneity, the pooled logit and mixture distribution imply similar WTPs. At a price difference of $\Delta p = 0.375$ (the average WTP in the population), the high-WTP type chooses the attribute with near certainty while the low-WTP type chooses it slightly less than one-third of the time. Thus, almost two-thirds of the population chooses the attribute at a price difference equal to the average WTP in the population. The true mixture distribution thus crosses a choice probability of 0.5 at a price difference above 0.375, namely at $\Delta p = 0.68$. The pooled logit estimate of 0.66 aligns closely with this value.

In the second environment, with no WTP heterogeneity, both types have WTP $\beta_L = \beta_H = 0.75$. This results in the same setting as the first example above, which shows how scale heterogeneity alone leads the pooled estimate $\hat{\beta} = 0.94$ to overstate the common WTP. Thus, reducing WTP heterogeneity can increase the distortion in pooled WTP estimates.

5.2.2 Composition-driven distortion

Reducing the share of noisier types can increase bias. Consider two types with the same WTP $\beta_L = \beta_H = 0.9$ and different scales $\sigma_L = 0.5$ and $\sigma_H = 0.05$. With homogeneous WTPs, any mixture of the two types crosses a choice probability of 0.5 at $\Delta p = 0.9$. When $\pi = 0.5$, we obtain a pooled estimate of $\hat{\beta} = 1.29$, exceeding the common WTP as previous examples also show (Figure 6).

Now consider the effect of decreasing the population share of the noisier type to $\pi = 0.25$. The tails of the mixture become thinner, reflecting the greater weight on the less noisy type. Having a thinner tail leads to a high population probability of choosing the attribute at prices just below the common WTP. This in turn pushes the pooled logit estimate even higher, to $\hat{\beta} = 1.46$. Thus, even though the true average WTP remains unchanged and the share of noisier types falls, pooling bias increases.

Increasing the share of low-WTP types can push the pooled estimate further above the high type's WTP. Consider the same scale parameters as above ($\sigma_L = 0.5$ and $\sigma_H = 0.05$), but the WTPs of both types are now $\beta_L = 0.85$ and $\beta_H = 0.95$. Suppose the population share of the low-WTP type changes from $\pi = 0.05$ to $\pi = 0.25$ (Figure 7).

As the noisy low-WTP type becomes more common, the true mixture probability function becomes considerably less steep, as does the pooled logit approximation. This pushes the price difference at which the pooled logit crosses a choice probability of 0.5 up from $\Delta p = 1.54$ to $\Delta p = 1.64$.

5.2.3 Design-driven distortion

Holding preferences fixed, changing the price range can move the pooled estimate from below the low type's WTP to above the high type's WTP. Finally, pooling distortions can arise purely from the experimental design. Holding the population fixed at $\beta_L = 0.1$, $\beta_H = 0.9$, $\sigma_L = 6$, $\sigma_H = 0.2$, and $\pi = 0.5$, consider two designs that differ only in the range of prices that generate Δp in the experiment. The type-specific choice probabilities and the mixture choice probability thus remain the same (Figure 8).

Consider an experimental design in which prices fall within a relatively narrow range. Fix the price of the alternative without the attribute at $p_0 = 0$, and let the price of the alternative with the attribute vary uniformly over $p_1 \in [0, 1]$. Due to the presence of the noisier low-WTP type, the pooled logit naturally has thicker tails than the choice probability function of the high-WTP type. At the same time, the pooled logit prioritizes fitting the portion of the mixture choice probability function in a range of price differences where variation in Δp

leads to relatively sharp changes in choice probabilities (due to the low choice error of the high-WTP type). This pushes the pooled estimate above the high type's WTP to $\hat{\beta} \approx 0.95$.

Now consider a design in which the price difference varies over a much wider range, with $p_0 = 0$ and $p_1 \in [0, 8]$. This puts much more weight on observations with large price differences, in the tail regions of the choice probability function, where the noisy low-WTP type generates non-negligible choice probabilities even when the high-WTP type rarely chooses the attribute. This flattens the pooled logit's implied choice probability function and pushes the pooled WTP estimate downward to $\hat{\beta} = -0.03$, which lies below the low type's WTP.

Expanding the experimental price range thus flips the pooled logit from fitting the steep change in choice probabilities with a small range of price differences to fitting the tails with a large range of price differences. As a result, holding preferences and population composition fixed, the pooled logit approximates different parts of the same mixture curve under alternative designs, producing WTP estimates that can move from above the highest to below the lowest individual valuation.

5.3 General results on pooling distortions

The examples above reveal two distinct forces that move the pooled-logit-implied WTP.

First, mixtures of individual logit curves generate features in $m(\cdot)$ that a single logit cannot reproduce. Heterogeneity in willingness to pay and scale naturally create plateaus and thick tails that mechanically distort the location and slope parameters of any single-logit approximation. Composition effects also operate through this channel: changing population shares can alter tail thickness or the height of plateaus in $m(\cdot)$ and thereby change how severely the single-logit approximation distorts the crossing point.

Second, the implied WTP depends on how price differences are distributed in the experiment. Under misspecification, maximum likelihood selects the logit curve that best fits $m(\cdot)$ under the experimental distribution of price differences μ. When different designs place relatively more weight on the tails versus the marginal region, the pooled logit shifts toward fitting those regions, and the implied WTP moves accordingly. As a result, changing μ changes the objective function and can move $\beta^*(\mu)$ substantially even when preferences and population composition are held fixed.

We proceed in two steps. First, we characterize the pseudo-true pooled logit under a two-point design and use this characterization to derive a theorem showing how pooled-logit-implied WTP can lie outside the support of individual valuations as well as a corollary formalizing the plateau example. Second, we show that holding preferences fixed, changing only μ can move the implied WTP by an arbitrary amount.

5.3.1 Pooled WTP outside the range of individual WTPs

The first result gives an exact characterization of the pseudo-true pooled logit under a two-point design.

Lemma 1. *Let* $\ell(u) := \log \frac{u}{1-u}$, *and let* μ *put positive probability on exactly two distinct price differences satisfying* $a < b$. *Let* $p_a := m(a)$ *and* $p_b := m(b)$, *and suppose that* $0 < p_b < p_a < 1$. *Then the pooled-logit-implied parameters satisfy*

$$\sigma^*(\mu) = \frac{b-a}{\ell(p_a) - \ell(p_b)}, \tag{6}$$

$$\beta^*(\mu) = b + \frac{b-a}{\ell(p_a) - \ell(p_b)}\ell(p_b) \tag{7}$$

$$= a + \frac{b-a}{\ell(p_a) - \ell(p_b)}\ell(p_a). \tag{8}$$

Under a design that consists of exactly two price differences, the pseudo-true pooled logit is uniquely determined by the pooled choice probabilities at those two points. The proof in Appendix C.1 derives this characterization from the two-point likelihood problem. Under a two-point design, the pseudo-true pooled logit matches the pooled choice probabilities exactly at the two support points:

$$F\left(\frac{\beta^*(\mu) - a}{\sigma^*(\mu)}\right) = p_a,$$

$$F\left(\frac{\beta^*(\mu) - b}{\sigma^*(\mu)}\right) = p_b.$$

The next theorem, proven in Appendix C.2, shows that outside-support distortions are governed by how slowly the pooled choice probability moves in log-odds space between two design points inside the range of individual valuations. This yields conditions under which the pooled-logit-implied WTP lies outside the support of individual valuations.

Theorem 5. *Let* β_L *and* β_H *denote lower and upper bounds on individual willingness to pay in the population. Let* $\ell(u) := \log \frac{u}{1-u}$, *and let* μ *put positive probability on exactly two distinct price differences satisfying* $a < b$ *with* $a, b \in [\beta_L, \beta_H]$. *Let* $p_a := m(a)$ *and* $p_b := m(b)$, *and suppose that* $0 < p_b < p_a < 1$. *The pooled-logit-implied WTP* $\beta^*(\mu)$ *satisfies the following:*

1. *If* $p_b > \frac{1}{2}$, *then* $\beta^*(\mu) > \beta_H$ *if and only if* $\frac{\ell(p_a)}{\ell(p_b)} < \frac{\beta_H - a}{\beta_H - b}$.
2. *If* $p_a < \frac{1}{2}$, *then* $\beta^*(\mu) < \beta_L$ *if and only if* $\frac{\ell(p_a)}{\ell(p_b)} > \frac{a - \beta_L}{b - \beta_L}$.

The exact two-point formulas in Lemma 1 clarify the result in the theorem. To lie above β_H, the pooled choice probability curve must decline sufficiently slowly in log-odds space between a and b relative to the distances $\beta_H - a$ and $\beta_H - b$. To lie below β_L, it must decline sufficiently slowly between a and b relative to the distances $a - \beta_L$ and $b - \beta_L$. If the pooled probability remains sufficiently high at the larger design point b, then the pooled-logit-implied WTP exceeds β_H. If it remains sufficiently low at the smaller design point a, then the pooled-logit-implied WTP falls below β_L.

The following corollary, proven in Appendix C.3, formalizes the example with heterogeneous WTP and common choice error, which produces an extended region where $m(\Delta p)$ stays nearly constant for a range of price differences. When types are sufficiently separated relative to the common scale, the pooled choice probability remains high and nearly flat between two interior design points, and the pooled WTP lies above the valuation of the high type.

Corollary 3. *Consider the two-type mixture*

$$m(\Delta p) = \pi F\left(\frac{\beta_L - \Delta p}{\sigma}\right) + (1-\pi) F\left(\frac{\beta_H - \Delta p}{\sigma}\right),$$

with $\beta_L < \beta_H$, common scale parameter σ, and $\pi < \frac{1}{2}$. Consider a design μ with support consisting of exactly the two price differences $a := \beta_L + t\sigma$ and $b := \beta_H - t\sigma$, where $t > 0$ satisfies $(1-\pi)F(t) > \frac{1}{2}$ with $F(z) := \frac{1}{1+\exp(-z)}$ denoting the logit CDF. Then there exists K such that if

$$\frac{\beta_H - \beta_L}{\sigma} > K,$$

then $\beta^*(\mu) > \beta_H$.

To see why this holds, note that

$$m(a) = \pi F(-t) + (1-\pi) F\left(\frac{\beta_H - \beta_L}{\sigma} - t\right),$$

and

$$m(b) = \pi F\left(t - \frac{\beta_H - \beta_L}{\sigma}\right) + (1-\pi) F(t).$$

As $\frac{\beta_H - \beta_L}{\sigma}$ grows, these converge to $(1-\pi) + \pi F(-t)$ and $(1-\pi)F(t)$, respectively, both of which fall between $\frac{1}{2}$ and 1. This implies that $\frac{\ell(m(a))}{\ell(m(b))}$ converges to a finite positive constant. At the same time, $\frac{\beta_H - a}{\beta_H - b} = \frac{\beta_H - \beta_L}{t\sigma} - 1$ can be made arbitrarily large. Thus, for $\frac{\beta_H - \beta_L}{\sigma}$ large enough, we have $\frac{\ell(p_a)}{\ell(p_b)} < \frac{\beta_H - a}{\beta_H - b}$, so Theorem 5 implies $\beta^*(\mu) > \beta_H$.

5.3.2 Arbitrarily large design sensitivity

The previous results show how the shape of the pooled choice probability curve can force the pooled-logit-implied WTP outside the range of individual valuations. The next theorem turns to the role of the experimental design itself. We consider a two-type population with homogeneous WTP, yielding pooled choice probability

$$m(\Delta p) = \pi F\Big(\frac{\beta_0 - \Delta p}{\sigma_L}\Big) + (1-\pi) F\Big(\frac{\beta_0 - \Delta p}{\sigma_H}\Big).$$

Holding the population fixed, changing only where the design places weight can move the pooled-logit-implied WTP by an arbitrary amount.

Theorem 6. *Fix any $\pi \in (0,1)$ and any $\beta_0 \in \mathbb{R}$. For every $M > 0$, there exist scale parameters $\sigma_H > \sigma_L$ and $R > 0$ such that, if the support of μ_R^- consists of the two price differences $\beta_0 - R - 1$ and $\beta_0 - R$, and the support of μ_R^+ consists of the two price differences $\beta_0 + R$ and $\beta_0 + R + 1$, then*

$$\beta^*(\mu_R^-) > \beta_0 + M$$

and

$$\beta^*(\mu_R^+) < \beta_0 - M.$$

This theorem delivers two conclusions at once. First, even when all individuals share the same WTP β_0, the pooled-logit-implied WTP can be arbitrarily far from the truth. A left-tail design pushes the pooled estimate arbitrarily far above β_0, while a right-tail design pushes it arbitrarily far below β_0. Second, because these two distortions arise for the same population under different designs, changing only the experimental distribution of price differences can move the pooled-logit-implied WTP by an arbitrary amount.

To see why, place both design points far in the left tail or far in the right tail. In those regions, the pooled choice probability is asymptotically governed by the higher-scale type. As shown in the proof of Theorem 6, evaluating the pooled probabilities at the two support points under each design and applying the exact two-point characterization from Lemma 1 yields

$$\beta^*(\mu_R^-) \to \beta_0 - \sigma_H \log(1-\pi)$$

and

$$\beta^*(\mu_R^+) \to \beta_0 + \sigma_H \log(1-\pi).$$

Choosing σ_H large makes the left-tail limit arbitrarily far above β_0 and the right-tail limit arbitrarily far below it.

6 Discussion

6.1 When additional structure restores identification

Two standard kinds of restrictions restore point identification. First, specifying the function F in Equation (1) provides the additional structure needed to separate B from ε. The pooled curve $m(\cdot)$ identifies the distribution of shock-contaminated WTP $B + \varepsilon$. Given F, under standard deconvolution conditions, the marginal distribution of B can be recovered, so moments and quantiles of B are point identified. However, as our analysis in Section 4 cautions, commonly made assumptions may lead to point estimates that fall outside plausible bounds.

Second, observing multiple choices for the same individual yields additional information beyond the pooled choice probability curve. Learning the individual choice curves $q_i(\Delta p)$ makes it possible to estimate features of the WTP distribution even without specifying the distribution of shocks. In particular, the normalization $F(0) = \frac{1}{2}$ implies $\beta_i = q_i^{-1}(\frac{1}{2})$, so the marginal distribution of B is identified.

This discussion clarifies what mixed logit, latent class finite mixture, hierarchical Bayes, and related methods are doing. They overcome the problem by imposing structure on latent heterogeneity and on the shock process. Within a given model class, objects such as $\mathbb{E}[B]$ or $\mathrm{Med}[B]$ can be point identified. But that identification is relative to the maintained structure. Especially when price support is limited, distinct assumptions on heterogeneity and shocks can fit observed choice patterns similarly well while implying different WTP moments. Attempting to circumvent assumptions on latent heterogeneity by using these approaches to estimate WTP for each individual separately leads to other challenges: with few responses per person, individual likelihoods are often flat, and when a respondent almost always chooses one option at low prices and never at high prices, the maximum likelihood estimate can diverge or be extremely sensitive to small perturbations in the data. This observation explains why such approaches are rarely used in practice, given the limited number of choice opportunities that are typically observed per respondent. Adaptive experimental design offers a way to learn about WTP at the individual level in such situations (Drake et al., 2025; Drake, Thakral and Tô, 2025).

6.2 Reporting practices for applied work

Our analysis suggests several guidelines for improving the transparency and credibility of empirical results when analyzing pooled binary-choice data.

When using parametric specifications, plotting pooled choice probabilities as a function of price differences alongside parametric fits can reveal how well the assumed functional form captures the observed choice patterns. This would also show where the observed experimental designs lie relative to the estimated indifference point to assess the potential for bias resulting from misspecification and extrapolation. Comparing the pooled relationship with subgroup-specific fits, especially along dimensions that plausibly proxy for noisiness or attentiveness, also helps to illustrate whether the assumed parametric form appropriately aggregates heterogeneity in the population. Examining how estimates change when narrowing the support of the designs considered in the experiment would also be informative about how sensitive the reported pooled estimates are to extrapolation. Finally, reporting parametric estimates of mean or median WTP (whether from logit, mixed logit, hierarchical Bayes, or other methods) alongside the bounds in Theorems 2 to 4 provides a more complete picture of the range of estimates consistent with the data. Point estimates that do not fall within even a conservative set of bounds would raise questions about the assumed parametric structure and the credibility of the conclusions.

6.3 Concluding remarks

One of the messages of this paper is about misspecification. The idea that a mixture of logits is itself not a logit is well-understood in the choice modeling literature, even if many applications in economics estimate a fixed-coefficient logit rather than a mixed logit.[20] Our results push this idea further by (i) illustrating how bias from pooling can arise due to heterogeneity in preferences, heterogeneity in choice error, or population composition; and (ii) showing that pooling bias can lead to mean or median WTP estimates that fall outside the range of the individual-level WTPs and can be arbitrarily far away from the true population mean or median. In addition, the empirical support of prices can change estimates arbitrarily even when individual preferences are held fixed.

More fundamentally, even if the pooled curve were known exactly, pooling generally only identifies the distribution of shock-contaminated WTP rather than the WTP distribution itself. The bounding approach proves useful here. The bounds on mean and median WTP in our empirical application are tight enough to rule out the point estimates from the preferred parametric specification. They additionally highlight that designs in which median WTP remains unbounded occur in a variety of applied settings. If the goal is to point identify WTP, one needs additional structure, or a design that directly targets individual indifference points such as an adaptive choice experiment (Drake et al., 2025).

---

[20] Recent examples include Bustelo et al. (2023), Fuchsman, McGee and Zamarro (2023), Castro and Mang (2024), and Johnston (2025).

References

Ashraf, Nava, James Berry, and Jesse M Shapiro. 2010. “Can higher prices stimulate product use? Evidence from a field experiment in Zambia.” *American Economic Review*, 100(5): 2383–2413. 6

Barseghyan, Levon, Maura Coughlin, Francesca Molinari, and Joshua C Teitelbaum. 2021. “Heterogeneous choice sets and preferences.” *Econometrica*, 89(5): 2015–2048. 4

Bhattacharya, Debopam. 2015. “Nonparametric welfare analysis for discrete choice.” *Econometrica*, 83(2): 617–649. 3

Bhattacharya, Debopam. 2018. “Empirical welfare analysis for discrete choice: Some general results.” *Quantitative Economics*, 9(2): 571–615. 3

Bhattacharya, Debopam. 2021. “The empirical content of binary choice models.” *Econometrica*, 89(1): 457–474. 3

Bhattacharya, Debopam. 2024. “Nonparametric approaches to empirical welfare analysis.” *Journal of Economic Literature*, 62(2): 554–593. 3

Bustelo, Monserrat, Ana Maria Diaz, Jeanne Lafortune, Claudia Piras, Luz Magdalena Salas, and José Tessada. 2023. “What is the price of freedom? Estimating women’s willingness to pay for job schedule flexibility.” *Economic Development and Cultural Change*, 71(4): 1179–1211. 30

Carson, Richard T. 2000. “Contingent valuation: a user’s guide.” 1

Carson, Richard T, and W Michael Hanemann. 2005. “Contingent valuation.” *Handbook of environmental economics*, 2: 821–936. 1

Castro, Silvia, and Clarissa Mang. 2024. “Breaking the silence–Group discussions and the adoption of menstrual health technologies.” *Journal of Development Economics*, 169: 103264. 30

Chernozhukov, Victor, Iván Fernández-Val, and Alfred Galichon. 2010. “Quantile and probability curves without crossing.” *Econometrica*, 78(3): 1093–1125. 15

Chesher, Andrew, Adam M Rosen, and Konrad Smolinski. 2013. “An instrumental variable model of multiple discrete choice.” *Quantitative Economics*, 4(2): 157–196. 4

Compiani, Giovanni. 2022. “Market counterfactuals and the specification of multiproduct demand: A nonparametric approach.” *Quantitative Economics*, 13(2): 545–591. 4

Cosaert, Sam, and Thomas Demuynck. 2018. “Nonparametric welfare and demand analysis with unobserved individual heterogeneity.” *Review of Economics and Statistics*, 100(2): 349–361. 4

Daly, Andrew, Stephane Hess, and Kenneth Train. 2012. "Assuring finite moments for willingness to pay in random coefficient models." *Transportation*, 39(1): 19–31. 3

Desvousges, William H, F Reed Johnson, Richard W Dunford, Kevin J Boyle, Sara P Hudson, and K Nicole Wilson. 2010. *Measuring nonuse damages using contingent valuation: An experimental evaluation of accuracy.* RTI press. 1

Drake, Marshall, Fernando Payró, Neil Thakral, and Linh T Tô. 2025. "Bayesian Adaptive Choice Experiments." *Mimeo.* 29, 30

Drake, Marshall, Neil Thakral, and Linh T Tô. 2025. "Wage Differentials and the Price of Workplace Flexibility." *Mimeo.* 29

Fiebig, Denzil G, Michael P Keane, Jordan Louviere, and Nada Wasi. 2010. "The generalized multinomial logit model: accounting for scale and coefficient heterogeneity." *Marketing Science*, 29(3): 393–421. 3

Fuchsman, Dillon, Josh B McGee, and Gema Zamarro. 2023. "Teachers' willingness to pay for retirement benefits: A national stated preferences experiment." *Economics of Education Review*, 92: 102349. 30

Garro-Marín, César, Neil Thakral, and Linh T Tô. 2025. "Work Hours and Amenity Trade-Offs." Vol. 115, 282–286. 14

Hanemann, W Michael. 1984. "Welfare evaluations in contingent valuation experiments with discrete responses." *American Journal of Agricultural Economics*, 66(3): 332–341. 3

Hauber, A Brett, Juan Marcos González, Catharina GM Groothuis-Oudshoorn, Thomas Prior, Deborah A Marshall, Charles Cunningham, Maarten J IJzerman, and John FP Bridges. 2016. "Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR conjoint analysis good research practices task force." *Value in Health*, 19(4): 300–315. 3

Hausman, Jerry A, and Whitney K Newey. 2016. "Individual heterogeneity and average welfare." *Econometrica*, 84(3): 1225–1248. 4

Johnston, Andrew C. 2025. "Preferences, selection, and the structure of teacher pay." *American Economic Journal: Applied Economics*, 17(3): 310–346. 30

Johnston, Robert J, Kevin J Boyle, Wiktor Adamowicz, Jeff Bennett, Roy Brouwer, Trudy Ann Cameron, W Michael Hanemann, Nick Hanley, Mandy Ryan, Riccardo Scarpa, et al. 2017. "Contemporary guidance for stated preference studies." *Journal of the Association of Environmental and Resource Economists*, 4(2): 319–405. 3

Kamat, Vishal, and Samuel Norris. 2025. "Estimating Welfare Effects in a Non-Parametric Choice Model: The Case of School Vouchers." *Review of Economic Studies*, rdaf064. 3

Kitamura, Yuichi, and Jörg Stoye. 2018. "Nonparametric analysis of random utility models." *Econometrica*, 86(6): 1883–1909. 4

Lancsar, Emily, and Jordan Louviere. 2008. "Conducting discrete choice experiments to inform healthcare decision making: a user's guide." *Pharmacoeconomics*, 26(8): 661–677. 3

Manski, Charles F. 2003. *Partial identification of probability distributions.* Springer. 4

Mas, Alexandre, and Amanda Pallais. 2019. "Labor supply and the value of non-work time: Experimental estimates from the field." *American Economic Review: Insights*, 1(1): 111–126. 2, 14, 15, 16, 17, 18, 19, 20, 42, 43

Molinari, Francesca. 2008. "Partial identification of probability distributions with misclassified data." *Journal of Econometrics*, 144(1): 81–117. 4

Molinari, Francesca. 2020. "Microeconometrics with partial identification." *Handbook of Econometrics*, 7: 355–486. 4

Newey, Whitney K, and Daniel McFadden. 1994. "Large sample estimation and hypothesis testing." *Handbook of Econometrics*, 4: 2111–2245. 21

Parsons, George, and Kelley Myers. 2017. "Fat tails and truncated bids in contingent valuation: An application to an endangered shorebird species." In *Contingent valuation of environmental goods.* 17–42. Edward Elgar Publishing. 2, 18, 20, 35

Small, Kenneth A, and Harvey S Rosen. 1981. "Applied welfare economics with discrete choice models." *Econometrica: Journal of the Econometric Society*, 105–130. 3

Sonnier, Garrett, Andrew Ainslie, and Thomas Otter. 2007. "Heterogeneity distributions of willingness-to-pay in choice models." *Quantitative Marketing and Economics*, 5(3): 313–331. 3

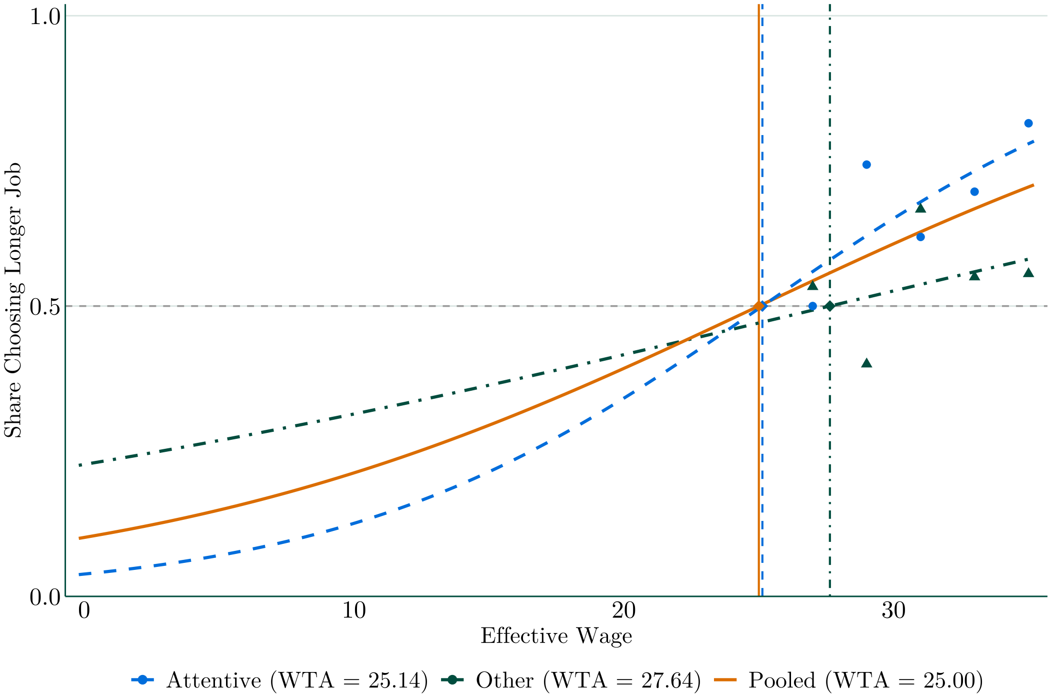
Tebaldi, Pietro, Alexander Torgovitsky, and Hanbin Yang. 2023. "Nonparametric estimates of demand in the california health insurance exchange." *Econometrica*, 91(1): 107–146. 4

Train, Kenneth, and Melvyn Weeks. 2005. "Discrete choice models in preference space and willingness-to-pay space." In *Applications of simulation methods in environmental and resource economics.* 1–16. Springer. 3

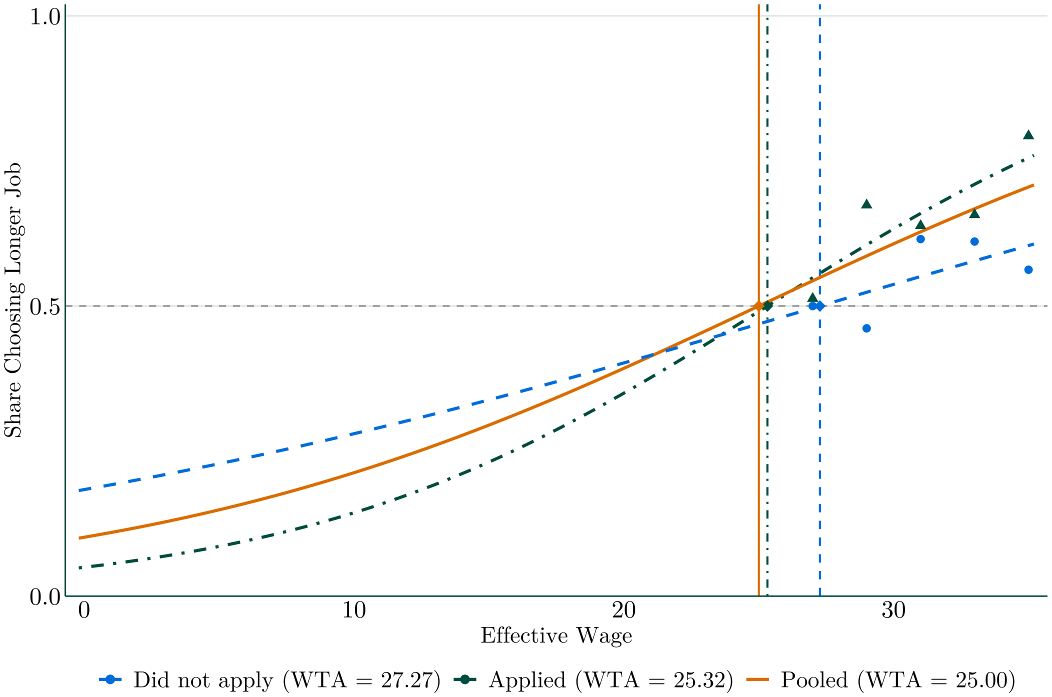
Train, Kenneth E. 2009. *Discrete choice methods with simulation.* Cambridge University Press. 21

White, Halbert. 1982. "Maximum likelihood estimation of misspecified models." *Econometrica: Journal of the econometric society*, 1–25. 3, 21

Figure 1: Pooling bias in compensating differential for longer hours



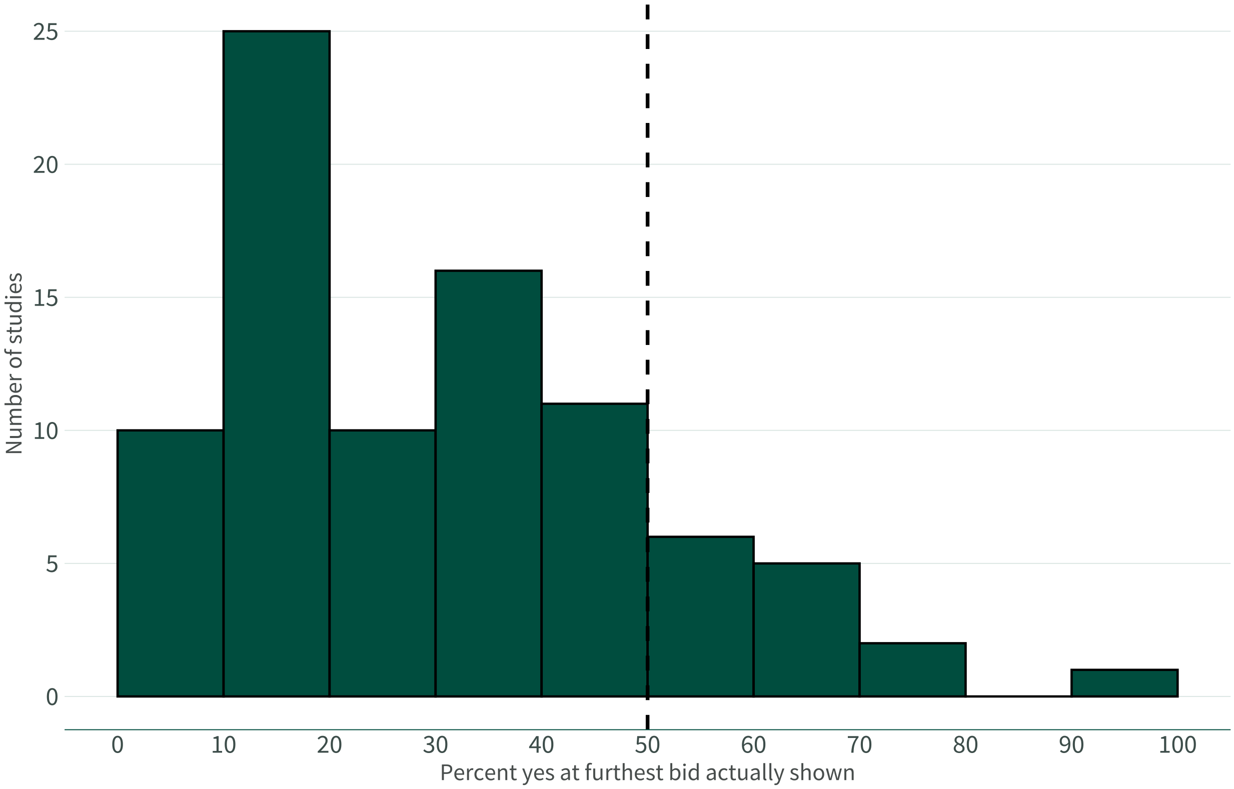
(a) Split by attentiveness



(b) Split by job application

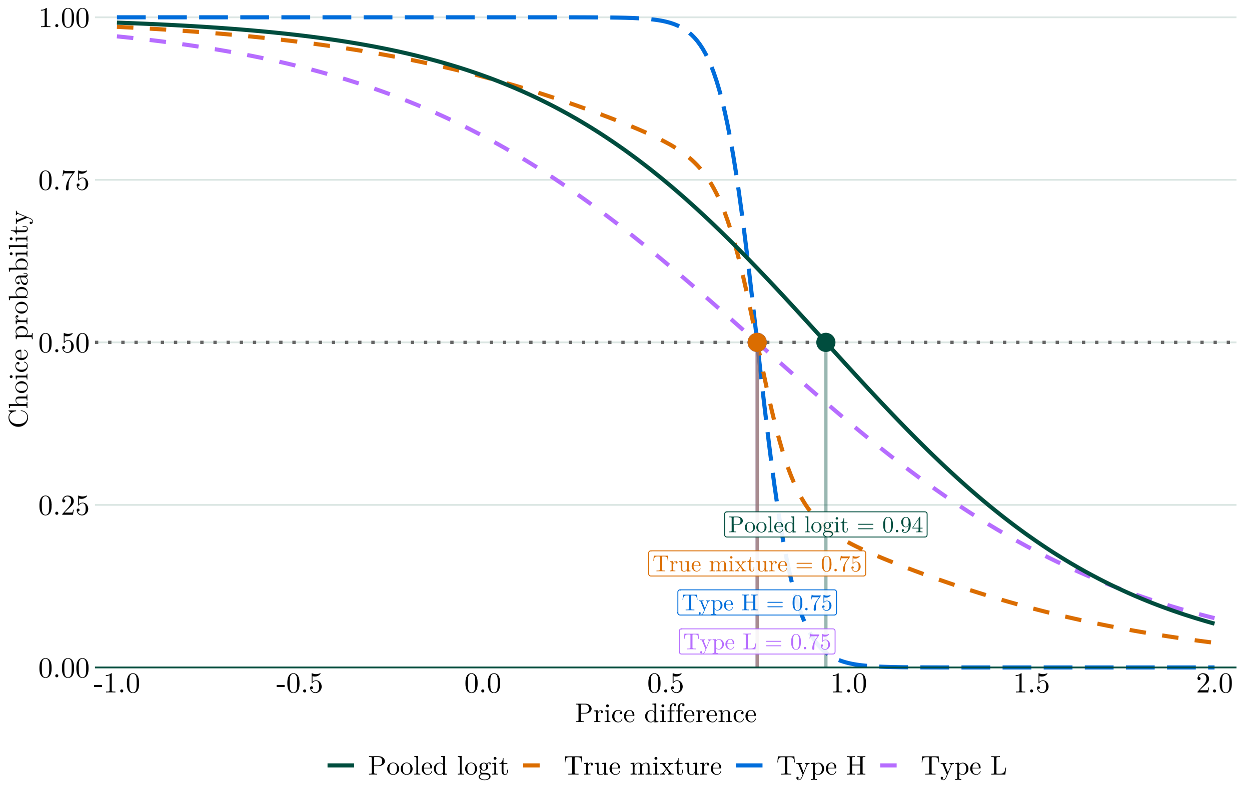
Note: The first figure splits the sample by attentiveness. The second figure splits the sample by whether the respondent applied for the job. In both figures, the pooled fitted crossing lies to the left of both subgroup-specific crossings.

Figure 2: Percent willing to pay at highest valuation in experiment



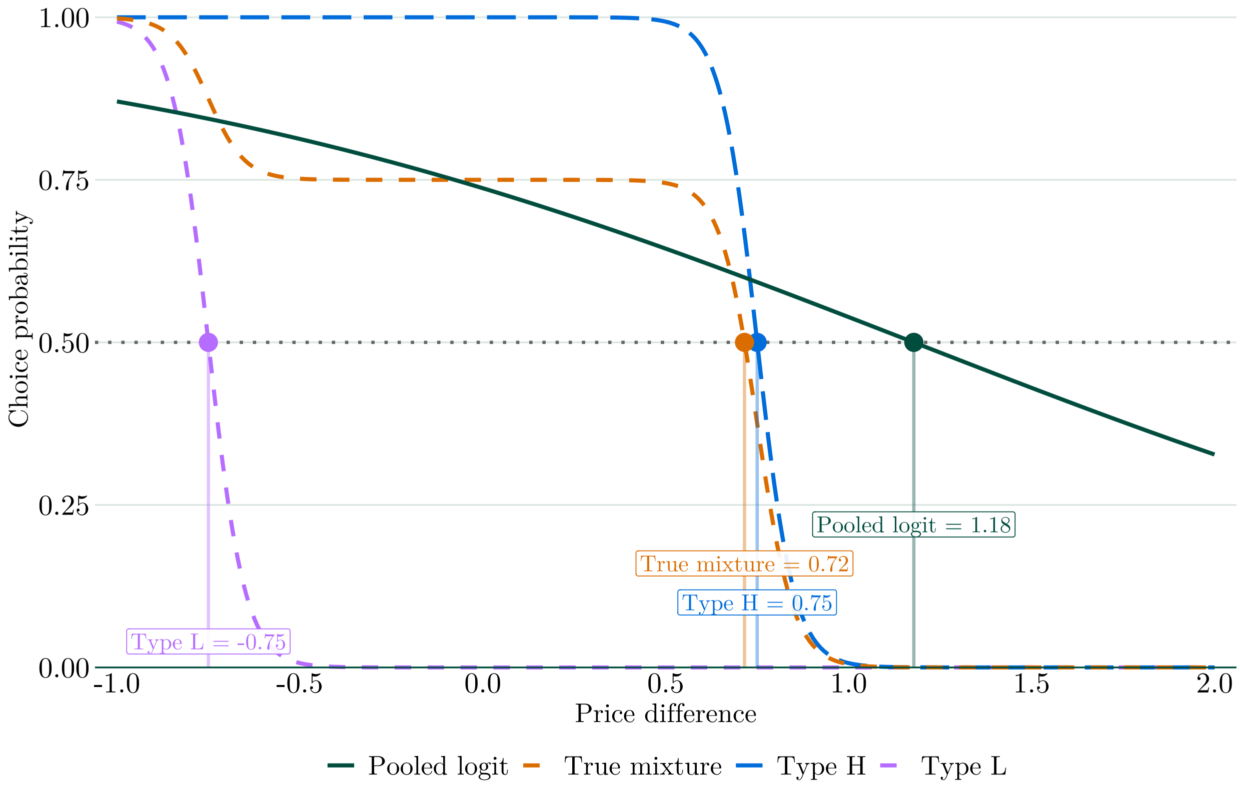
Note: This figure shows the percent of respondents indicating a willingness to pay the highest amount across studies in the meta-analysis by Parsons and Myers (2017). The vertical dashed line denotes a threshold of 50 percent.

Figure 3: Same WTP, heterogeneous choice error



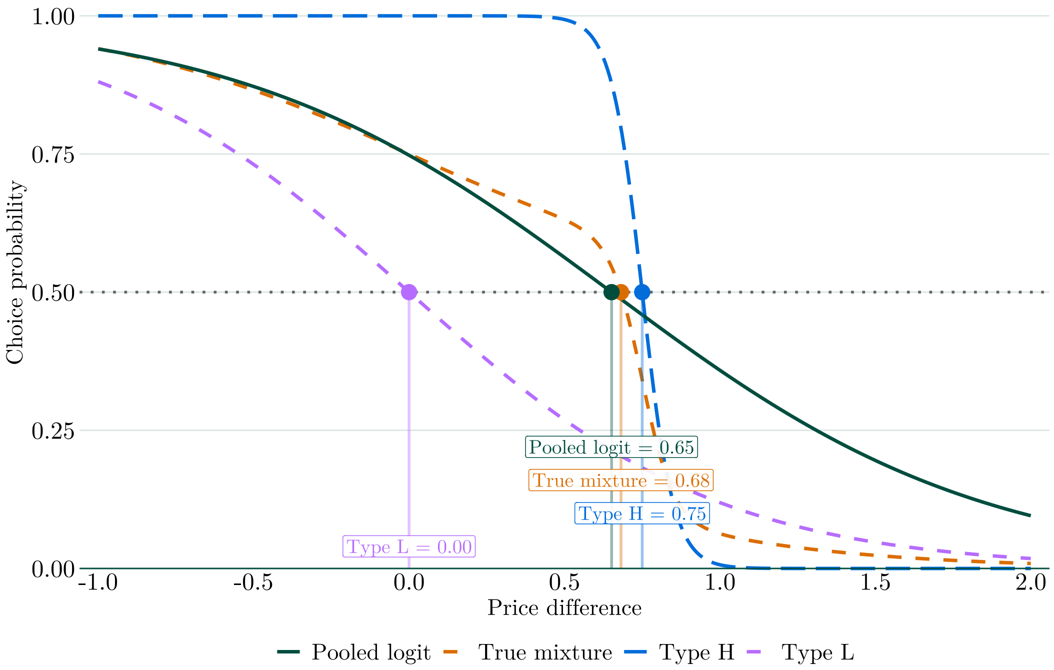
Note: The figure displays choice curves for two types following Equation (3) with preference parameters $\beta_L = \beta_H = 0.75$, $\sigma_L = 0.5$, $\sigma_H = 0.05$, and a share $\pi = 0.5$ of type L.

Figure 4: Heterogeneous WTP, same choice error

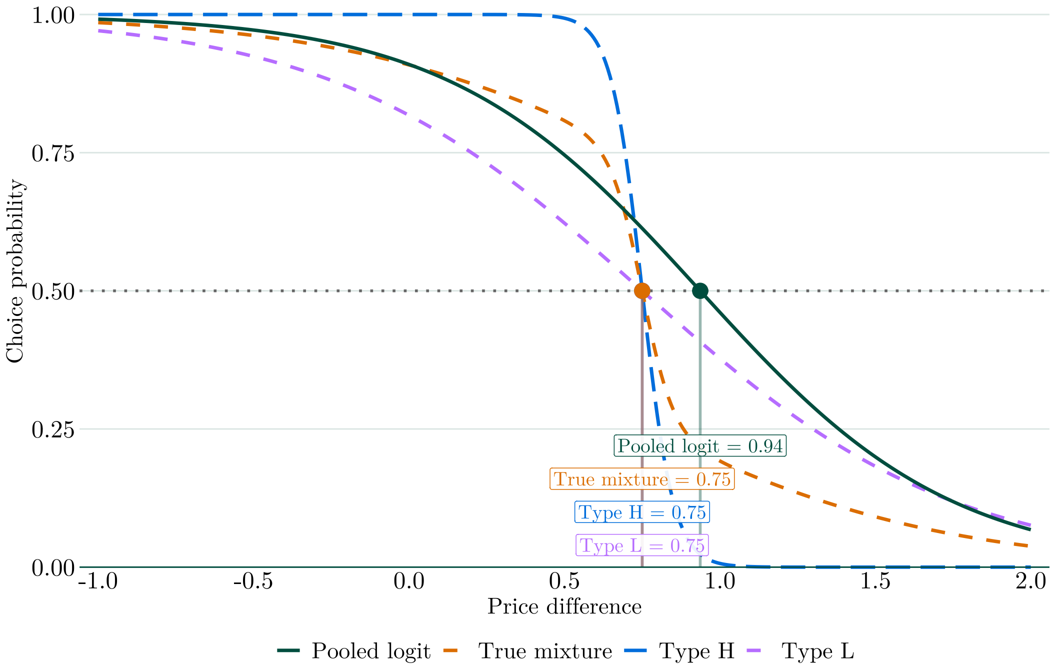


Note: The figure displays choice curves for two types following Equation (3) with preference parameters $\beta_L = -0.75$ and $\beta_H = 0.75$, $\sigma_L = \sigma_H = 0.05$, and a share $\pi = 0.25$ of type L.

Figure 5: Reducing WTP heterogeneity



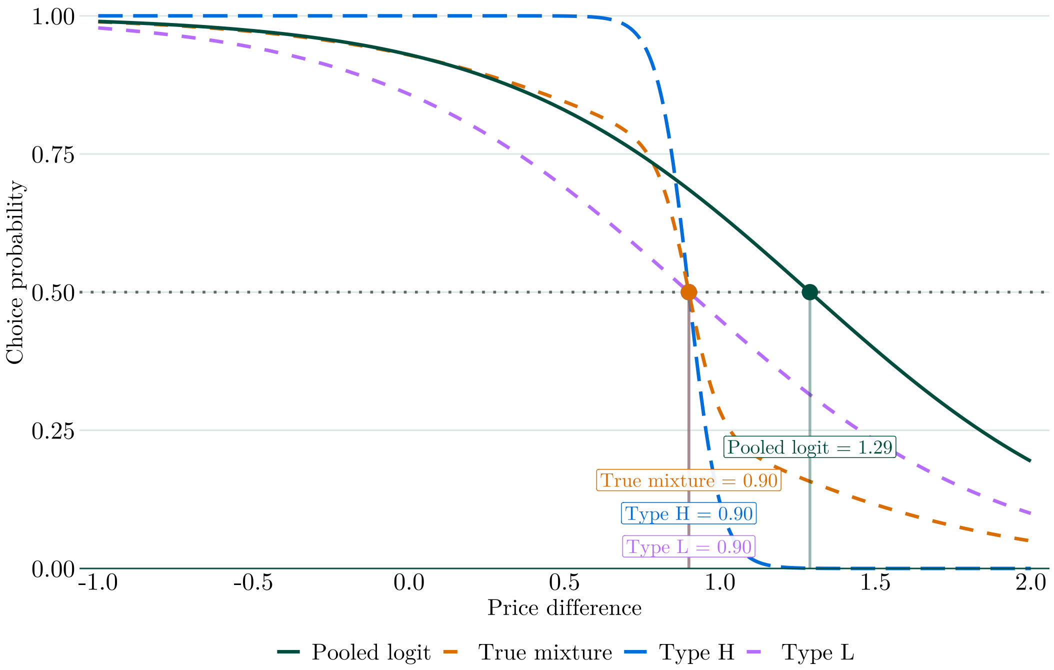
(a) Heterogeneous WTP



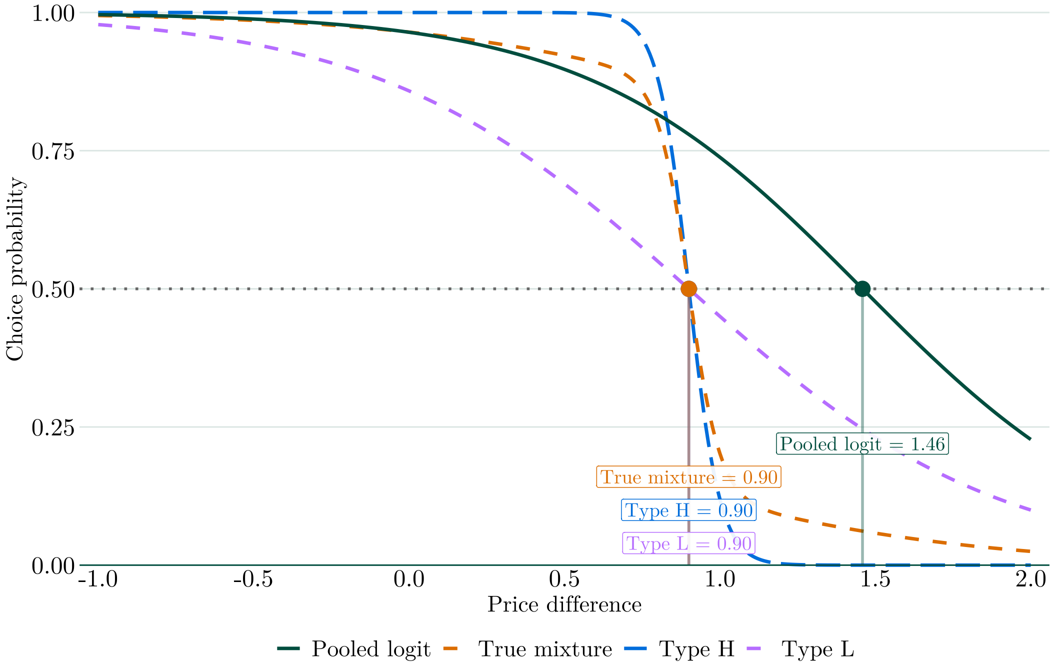
(b) Homogeneous WTP

Note: The first figure displays choice curves for two types following Equation (3) with preference parameters $\beta_L = 0$, $\beta_H = 0.75$, $\sigma_L = 0.5$, $\sigma_H = 0.05$, and a share $\pi = 0.5$ of type L. The second figure uses the same parameters except changes β_L from 0 to 0.75.

Figure 6: Reducing share of noisier type



(a) Larger share of noisier type

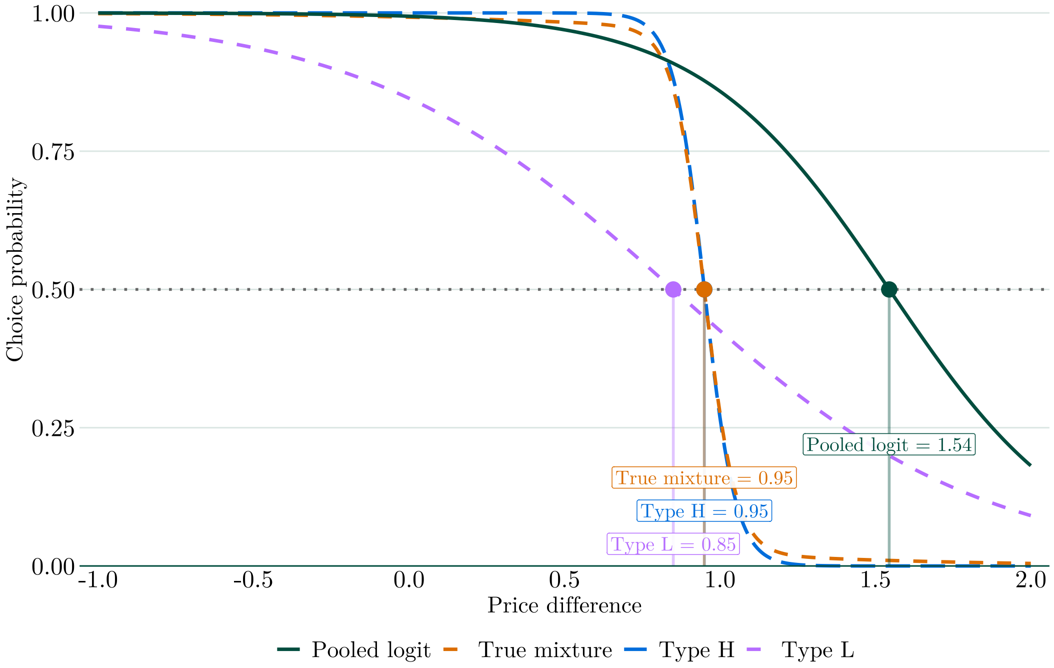


(b) Smaller share of noisier type

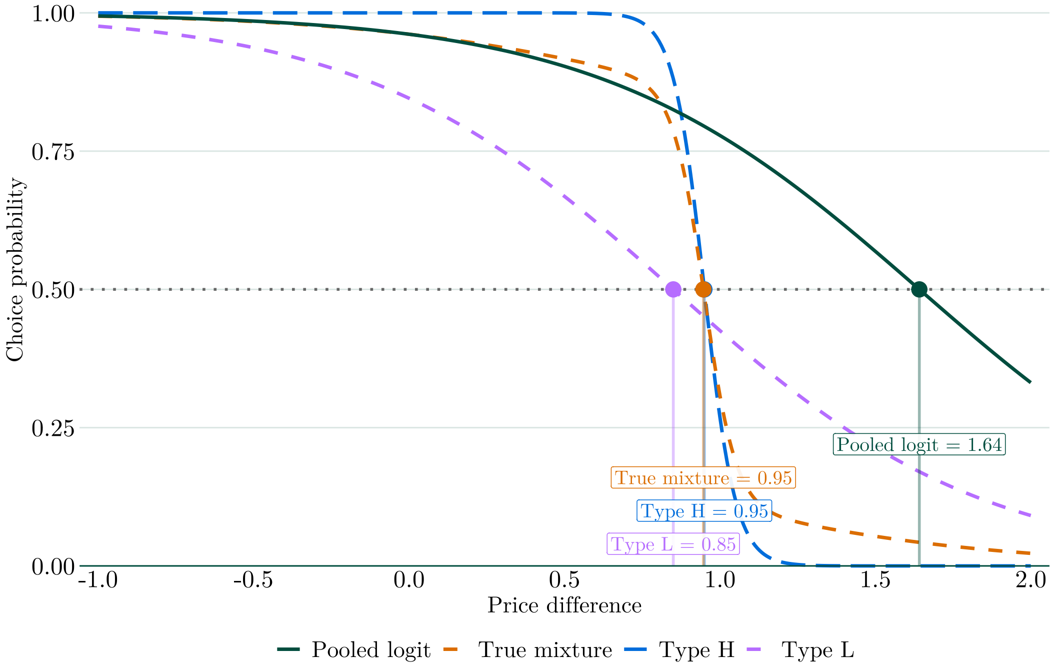
Note: Both figures display choice curves for two types following Equation (3) with preference parameters $\beta_L = \beta_H = 0.9$, $\sigma_L = 0.5$, and $\sigma_H = 0.05$. The share of type L changes from $\pi = 0.5$ in the first figure to 0.25 in the second figure.

Figure 7: Increasing share of low-WTP types

(a) Smaller share of low-WTP types



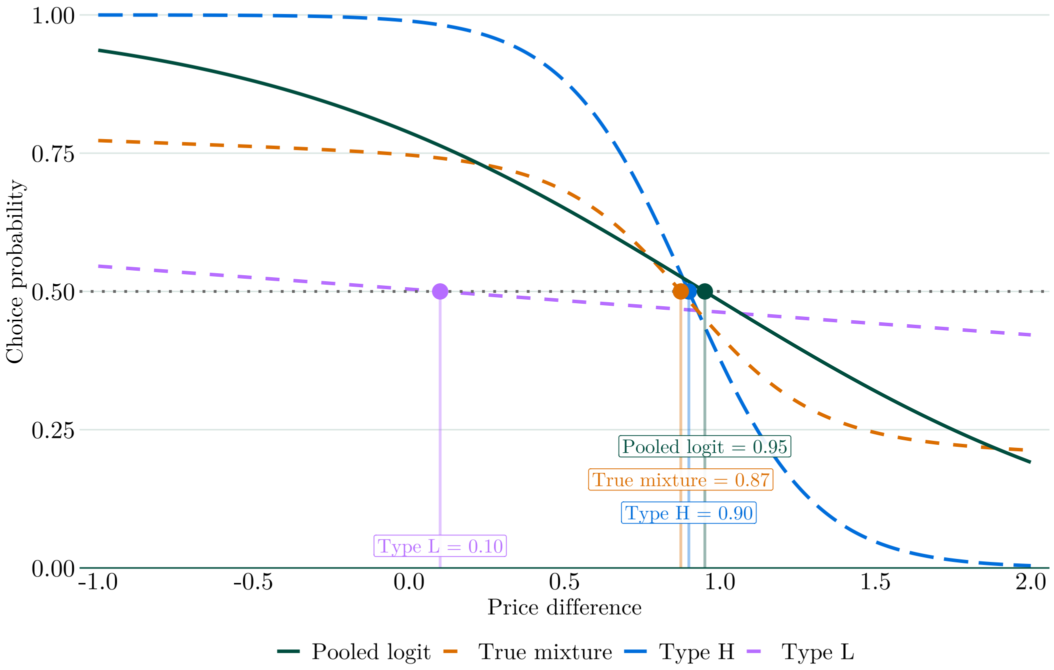
(b) Larger share of low-WTP types



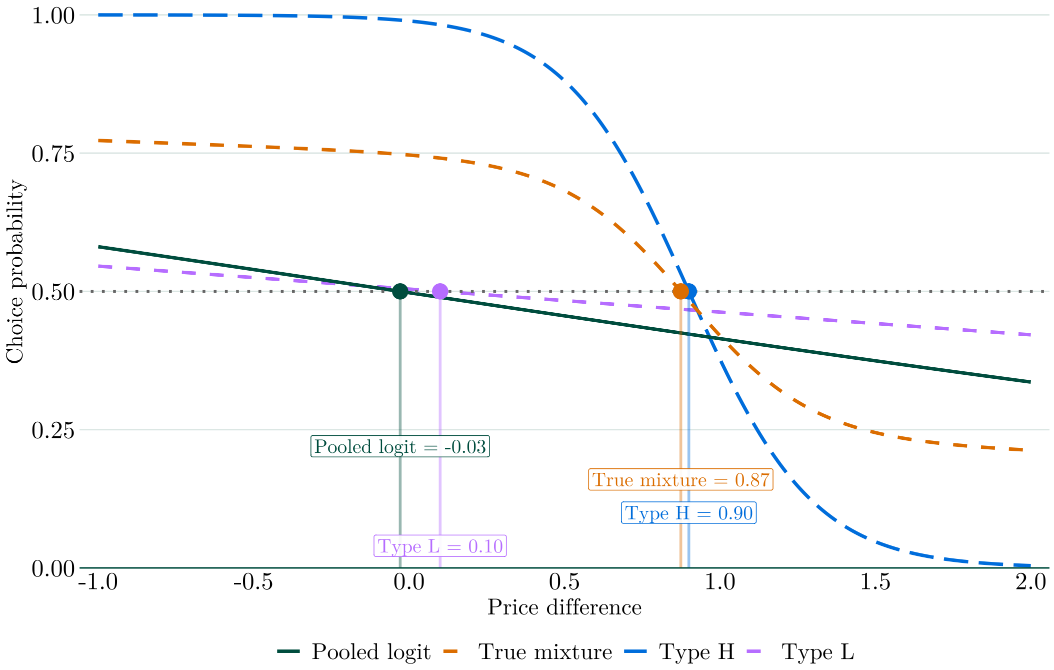
Note: Both figures display choice curves for two types following Equation (3) with preference parameters $\sigma_L = 0.5$, $\sigma_H = 0.05$, $\beta_L = 0.85$, and $\beta_H = 0.95$. The share of type L changes from $\pi = 0.05$ in the first figure to 0.25 in the second figure.

Figure 8: Changing the range of price differences

(a) Narrower range of price differences



(b) Wider range of price differences



Note: The figure displays choice curves for two types following Equation (3) with preference parameters $\beta_L = 0.1$, $\beta_H = 0.9$, $\sigma_L = 6$, $\sigma_H = 0.2$, and a share $\pi = 0.5$ of type L. The distribution of price differences in the experimental design changes from $\mathcal{U}[0, 1]$ in the first figure to $\mathcal{U}[0, 8]$ in the second figure.

Table 1: Bounds on WTA vs. pooled logit using Mas and Pallais (2019) data

	Bounds			Point estimates		
	Median	Mean of latent B	Mean of $B+\varepsilon$	Pooled logit	Median of two-type mixture	Mean of two-type mixture
Panel A: Sample of unemployed jobseekers						
50 vs. 45	$(-\infty, 31]$	[11.7029, 33.8985]	[14.7678, 31.8939]	24.6965	26.9949	29.5971
45 vs. 40	$(-\infty, 29]$	[10.9926, 34.9473]	[14.0668, 33.0192]	43.1008	26.5090	27.3805
40 vs. 35	[6.4, 18]	[7.9507, 16.8984]	[10.9356, 13.6829]	12.0000	14.0075	15.5584
35 vs. 30	[13.9, 19.8]	[9.9351, 19.5332]	[13.1770, 16.7510]	14.6843	19.0969	15.1035
30 vs. 25	[-2, 14]	[5.7787, 14.2898]	[8.6479, 11.3460]	7.9666	16.8080	9.2902
25 vs. 20	[4, 18]	[6.4866, 15.6400]	[9.6940, 12.5558]	8.9491	15.6817	11.1588
20 vs. 15	[2, 10]	[5.2037, 12.9759]	[8.0829, 10.1194]	7.0035	7.7741	7.3321
15 vs. 10	[1.9, 8]	[3.0451, 10.1433]	[5.7319, 7.5265]	5.2673	6.2395	7.9753
10 vs. 5	[0, 6]	[2.1329, 8.5310]	[4.4517, 6.4699]	4.7976	1.8711	5.8931
Panel B: Sample of jobseekers who did not apply						
50 vs. 45	[25, 33]	[15.7199, 38.3600]	[17.9605, 36.4989]	27.2673	32.3530	31.8544
45 vs. 40	$(-\infty, 29]$	[10.3041, 34.4622]	[12.6017, 32.9863]	19.4836		
40 vs. 35	[14.4, 22.1]	[12.2838, 21.1450]	[14.5806, 18.5494]	17.9224	11.3805	13.5476
35 vs. 30	[13.9, 20]	[11.5439, 20.4725]	[13.9023, 17.8575]	15.4323	19.3627	26.8034
30 vs. 25	[10, 20]	[9.0426, 16.4209]	[11.3818, 13.7536]	12.1391	17.7416	20.4904
25 vs. 20	[2, 14]	[6.5670, 13.8759]	[9.1524, 11.3619]	8.2289	1.5882	10.5006
20 vs. 15	[6, 18]	[7.2679, 14.3891]	[9.6885, 11.8406]	9.1580		
15 vs. 10	[6.1, 14]	[7.0696, 13.8341]	[9.1390, 11.4816]	8.5539	-12.1162	-1.1741
10 vs. 5	[0, 6]	[5.2176, 13.1948]	[7.6530, 10.6227]	7.2537	14.0958	10.1854

Note: The top panel reports estimates for the preferred sample of unemployed jobseekers (inattention rate 10.03 percent), while the bottom panel reports estimates for the sample of workers who did not apply to the advertised position (inattention rate 6.73 percent). Each row corresponds to a different treatment, defined by the number of hours of work in the longer position relative to the shorter position. The median bound and the mean bound for latent B impose the independence and median-zero assumptions from Theorem 2 and Theorem 3. The mean bound directly on $B+\varepsilon$ under the assumption of $\mathbb{E}[\varepsilon]=0$ follows Theorem 4. We set α equal to the inattention rate for the sample and $c=2$. We assume a support of $[0,35]$ for computing bounds on mean WTA, except for in the overtime treatments (50 vs. 45 and 45 vs. 40) where we set the upper bound to 50. The pooled logit column reports WTA estimates based on a logit regression with an indicator for choosing the longer position as the dependent variable and the effective wage as the explanatory variable. The last two columns report mean and median WTA estimates based on a latent two-type mixture model. Empty cells reflect convergence issues with the two-type mixture model.

Table A1: Conservative mean WTA bounds vs. pooled logit using Mas and Pallais (2019) data

	Bounds		Point estimates	
	Mean of latent B	Mean of $B + \varepsilon$	Pooled logit	Mean of two-type mixture
Panel A: Sample of unemployed jobseekers				
50 vs. 45	[11.4238, 33.9659]	[14.5167, 32.1104]	24.6965	29.5971
45 vs. 40	[6.5624, 37.8754]	[9.8143, 35.9750]	43.1008	27.3805
40 vs. 35	[7.5787, 18.2558]	[10.0403, 15.1263]	12.0000	15.5584
35 vs. 30	[9.6824, 19.6703]	[12.9498, 16.8874]	14.6843	15.1035
30 vs. 25	[5.4016, 14.6440]	[8.2223, 11.6742]	7.9666	9.2902
25 vs. 20	[4.5649, 18.0258]	[7.9650, 15.0603]	8.9491	11.1588
20 vs. 15	[3.7645, 17.2152]	[5.8761, 14.2749]	7.0035	7.3321
15 vs. 10	[2.4272, 12.1973]	[4.9527, 9.4590]	5.2673	7.9753
10 vs. 5	[1.6890, 9.8278]	[3.6487, 7.6833]	4.7976	5.8931
Panel B: Sample of jobseekers who did not apply				
50 vs. 45	[14.5664, 39.2839]	[16.8077, 37.4663]	27.2673	31.8544
45 vs. 40	[8.9295, 34.5721]	[11.2171, 33.0889]	19.4836	
40 vs. 35	[11.9660, 23.2186]	[14.0976, 20.4833]	17.9224	13.5476
35 vs. 30	[11.5133, 22.3290]	[13.8423, 20.1075]	15.4323	26.8034
30 vs. 25	[8.2821, 18.5838]	[10.0745, 15.9209]	12.1391	20.4904
25 vs. 20	[4.6399, 17.2253]	[6.4571, 14.6000]	8.2289	10.5006
20 vs. 15	[4.6390, 19.0108]	[6.0000, 16.5000]	9.1580	
15 vs. 10	[5.7075, 18.1162]	[7.6006, 15.6024]	8.5539	-1.1741
10 vs. 5	[3.9456, 16.2569]	[6.4333, 14.1061]	7.2537	10.1854

Note: This table expands the mean WTA bounds by using the highest weakly increasing function that never exceeds the raw choice probabilities and the lowest weakly increasing function that is never below them in Theorem 3 and Theorem 4 (with the independence and median-zero assumptions). See Table 1 for additional details on the samples, treatments, and specifications.