Applied Welfare Analysis for Discrete Choice with Interval-data on Income

Ying-Ying Lee*
University of California Irvine

Debopam Bhattacharya
University of Cambridge

January, 2019

Abstract

This paper concerns empirical measurement of Hicksian consumer welfare under interval-reported income. Bhattacharya (2015, 2018a) has shown that for discrete choice, welfare distributions resulting from a hypothetical price-change can be expressed as closed-form transformations of choice probabilities. However, when income is interval-reported, as is the case in many surveys, the choice probabilities, and hence welfare distributions are not point-identified. We derive bounds on average welfare in such scenarios under the assumption of a normal good. A finding of independent interest is a set of Slutsky-like shape restrictions which are linear in average demand, unlike those for continuous choice. A parametric specification of choice probabilities facilitates imposition of these Slutsky conditions, and leads to computationally simple inference for the partially identified features of welfare. In particular, the estimand is shown to be directionally differentiable, so that recently developed bootstrap methods can be applied for inference. Under mis-specification, our results provide a "best parametric approximation" to demand and welfare. These methods can be used for inference in more general settings where a class of set-identified functions satisfy linear inequality restrictions, and one wishes to conduct inference on functionals thereof. We illustrate our theoretical results using a simulation exercise based on a real dataset where actual income is observed. We artificially introduce interval-censoring of income, calculate bounds for the average welfare effects of a

*We are grateful to Jack Porter for valuable discussion that greatly improves this paper. We thank Bruce Hansen, Xiaoxia Shi, and Joachim Freyberger for very helpful comments. We also thank conference and seminar participants in IAAE 2015/2016, the Econometric Society World Congress 2015, 2016 North American Summer Meeting/Asian Meeting of the Econometric Society, cemmap, 2017 CEME Interaction Conference, University of Oxford, LSE, Universit catholique de Louvain, Academia Sinica, National Taiwan University, North Carolina State University, UC Irvine, University of Western Ontario, USC, UCSD, and Duke. E-mail: yingying.lee@uci.edu. JEL codes: C14, C25, C25 price-subsidy using our methods, and find that they perform favorably in comparison with estimates obtained using actual income.

Keywords: Binary choice, equivalent variation, interval-data, Slutsky restriction, set identified function, inference on functionals, directional differentiability.

1 Introduction

This paper concerns Hicksian welfare analysis of price changes in discrete choice settings, using micro-level demand data. For example, if the government raises the tax on rail-travel in a city where commuters choose between alternative modes of transport, then a theory-consistent way to measure its welfare effect is to calculate the compensating variation, i.e., the hypothetical income transfer required to maintain commuters' utilities Typically, individuals differ in their preferences, so that a price change produces a *distribution* of compensating variations and deadweight losses associated with the tax increase. Bhattacharya (2015, 2018a), has recently shown that in discrete choice settings with completely general heterogeneity, the distribution of Hicksian welfare resulting from price-changes can be expressed in terms of choice-probabilities which are functions of prices and consumer income. Estimating these probabilities requires knowledge of prices each individual in the sample faces and her income. However, in many datasets commonly used in discrete choice analysis, including the US Health and Retirement Study, the British Transportation Survey, the Nielsen database, the Current Population Survey, etc., individual income is reported in intervals, so that the choice probabilities cannot be calculated directly. This paper aims to develop econometric methods for welfare analysis in such scenarios.

Our theoretical approach is as follows. We assume that the alternative undergoing the price change (due to a tax, say) is normal on average, i.e., the probability of choosing it weakly rises with income, for fixed price. We choose a finite-dimensional parametric model for the choice probability, in a sense clarified below. We then derive bounds for welfare distributions that are closed-form functions of the parameters of the approximating model, subject to a set of new, Slutsky-like inequality restrictions. We develop the formal inference procedure for these bounds by adapting some recent results on bootstrap-inference for directionally differentiable functionals. Finally, we provide a "best parametric approximation" type interpretation of our estimates, to allow for the possibility that our parametric model is mis-specified.

Although developed in the context of welfare analysis, our inference procedures have wider applicability. In particular, consider *a function* $q(x)$ partially identified by a convex compact set of the form $[L(x), U(x)]$, where $q(x)$ satisfies a set of linear inequality restrictions. Our methods can be used to conduct inference for *functionals of* $q(\cdot)$. Such set identified functions are common in economics; for example, they arise when dependent variables in regressions are interval-censored or suffer from sample selectivity biases; see Manski (2003), Tamer (2010), and Chandrasekhar, Chernozhukov, Molinari, and Schrimpf (2012) for examples.

Description of the problem Consider an empirical setting of binary choice by a population of heterogeneous consumers. For a random sample drawn from this population, we observe each individual's choice among two alternatives 0 and 1, her demographic and other choice-relevant characteristics, and prices of the alternatives. We are interested in measuring the distribution of compensating variation, evaluated at income y, of a hypothetical price rise of alternative 1 from p_0 to p_1, e.g., one induced by a sales tax. Bhattacharya (2015) has shown that for this problem the expected compensating variation takes a closed-form expression, given by

$$\mathbb{E}[CV] = \int_0^{p_1 - p_0} q(p_0 + a, y + a)\ da, \tag{1}$$

where $q\left(p, y\right)$ is the structural choice probability, representing the fraction of the population that would choose alternative 1 if price and income were set to (p, y). If individual income is also observed in the micro-data, then evaluating the above expressions reduces to estimating the structural choice probabilities. But in many consumer datasets, individual income is recorded in fixed intervals to produce higher response rates. As the data do not report the continuous income variable, the conditional choice probability $q(p, y)$ in (1) cannot be estimated consistently without further assumption.

A common empirical approach is to make assumptions on the distribution of unobserved income within the observed intervals, and impute the missing income values thereof. A common short-cut is to simply use the interval's mid-point as the imputed income. If true income varies within the interval, as it must surely do in the real world, the midpoint imputation approach implies a strange behavioral assumption, viz. that there is no income effect on choice probabilities *within* the income interval that happens to have been fixed arbitrarily by the survey design, but there can be income effects *across* income intervals, as no restrictions are imposed on the income effect across intervals in the mid-point imputation approach. Of course, one can assume that there is no income effect anywhere; but this assumption is hard to justify in many contexts. Indeed, if one assumes away income effects, then the problem studied in this paper disappears, as true average welfare equals the usual Marshallian consumer surplus. In more general, non-demand settings, Hsiao (1983) shows that the common approaches of using the midpoint or dummy variable regression pose problems for statistical inference and for interpreting linear regression models; Manski and Tamer (2002) provide further discussion on this point. An alternative is to use the observations on interval-valued income to bound the conditional choice probability, which is the approach we take here.

Toward that end, assume that the alternative 1 is normal on average, meaning that its choice probability is increasing in income y for fixed p.[1] Then for given price p, the choice probability $q(p, y)$ can be bounded below [above] by the conditional choice probabilities averaged over an income interval containing values smaller [larger] than y. Such bounds can be estimated nonparametrically, producing a bounded convex set as the identification region for the partially identified choice probability.

Depending on the size of the income intervals, the bounds on the conditional choice probabilities may not be tight; so we turn to economic theory for another source of identifying power. We derive a Slutsky type restriction on the choice probabilities. These take the form of *linear* inequality restrictions on the choice probabilities – a result that is of independent theoretical interest (c.f. Bhattacharya (2018b)). We use these restrictions to tighten the bounds on average CV, and develop a method of inference for these bounds. To achieve these objectives, we use the following approach.

We first assume that the choice probability takes a probit form, develop tools of inference, and finally we provide an interpretation of this exercise as an approximation that minimizes mean squared error in an appropriate sense. Given the probit specification, the average CV in (1) is approximated by a nonlinear function of the probit coefficients. As the choice probability is partially identified, so are the corresponding probit coefficients. Our goal is to find bounds on this average CV subject to the Slutsky restriction, which we achieve in two steps. First, we characterize the interior points of the identification set of the probit coefficients. Second, we show how to impose the Slutsky restrictions, and translate the identification set characterization of the coefficients into bounds for the average CV. Specifically, we construct a convex and compact identified set for the probit/logit coefficients. This is done by using an approach developed in the recent literature on support function based identification (c.f., Horowitz and Manski (2006); Beresteanu and Molinari (2008); Bontemps, Magnac, and Maurin (2012); Chandrasekhar, Chernozhukov, Molinari, and Schrimpf (2012); Kline and Santos (2013); Kaido and Santos (2014); Escanciano and Zhu (2014); Kaido (2016)). The set is characterized by its boundary points which have closed-form expressions, and its interior points are convex combinations of these boundaries. These features facilitate the later imposition of shape restrictions.

The next step involves solving a stochastic programming problem that maximizes/minimizes a non-convex objective function subject to linear inequality restrictions. The resulting estimand is a non-differentiable function of the preliminary parameters, so inference on it is non-standard (c.f., Hirano and Porter (2012); Woutersen and Ham (2013); Fang and Santos (2019); Hong and Li (2018); Hansen (2017)). Interestingly, however, we can show that our estimand is directionally differentiable, and derive the asymptotic distribution of its sample

[1] This assumption cannot be directly tested, since true income are not reported. But one can verify that the price-conditioned choice probability on the income intervals are increasing as one moves to the higher intervals.

analog estimator. In particular, we show that the bootstrap method recently developed by Dümbgen (1993), Fang and Santos (2019), and Hong and Li (2018) provides a consistent approximation to the limiting distribution, and use this to construct confidence intervals for the average CV and for its identified set. Last, but not least, our sample-analog estimator preserves the simplicity of ordinary least squares and is easy to calculate in practice.

We study the effectiveness of our methods using a simulation exercise based on a real dataset. Specifically, we use survey data from India on teenagers choosing whether to attend school or not, and the ultimate object of interest is the average EV for a hypothetical tuition subsidy. To judge the efficacy of our bounds approach, we first compute the average EV using the *true* incomes reported in the dataset, using the method of Bhattacharya (2015). Then we artificially create the problem of interval-valued incomes by categorizing true income into intervals, and compare the identified set for average welfare, constructed using methods of this paper, with the estimates obtained by using the true income data and midpoint imputations. We find that the point estimates using midpoint imputations can be potentially misleading. This setting also allows us to examine the behavior of our estimators and confidence sets under different extents of interval-censoring, e.g., many small intervals versus few large intervals.

The rest of the paper is organized as follows. Section 2 introduces the setup. Section 3 begins with a parametric specification and overviews the estimation procedure. Section 4 presents the inference theory and bootstrap method. Section 5 provides some discussion of welfare analysis under endogeneity. In Section 6, we discuss the mis-specification and best parametric approximation. Section 7 demonstrates the efficacy of our methods through a simulation exercise using data on school-attendance in India. All technical proofs are collected in an Appendix.

2 Setup

We begin by laying out a general problem that includes as a special case our problem of interest, viz., welfare analysis with interval-valued income. Then, in Section 2.1, we describe the theory for the binary choice model and structural objects of interest by building on Bhattacharya (2015) and present a new, nonparametric Slutsky restriction for binary choice. Following this, in Section 2.2, we discuss the issue of partial identification of welfare measures in the context of interval-valued income.

The general statistical problem can be stated as follows. Consider a situation where the object of interest is $\bar{f}(q)$, a real-valued functional of a function $q(\cdot)$ that is partially identified, i.e., it is contained in an estimable, compact and convex set, and satisfies an inequality restriction $\bar{R}(q) \leq 0$. Specifically, the identified set for q has the following form:

$$\begin{aligned}
\mathcal{Q} &\equiv \mathcal{Q}_{data} \cap \mathcal{Q}_{model}, \text{where} \\
\mathcal{Q}_{data} &\equiv \{q \in \mathcal{C}(\mathcal{X}) : L(x) \leq q(x) \leq U(x) \text{ for all } x \in \mathcal{X}\} \text{ and} \\
\mathcal{Q}_{model} &\equiv \left\{q \in \mathcal{C}(\mathcal{X}) : \bar{R}(q(x)) \leq 0 \text{ for all } x \in \mathcal{X} \right\},
\end{aligned}$$

where $\mathcal{C}(\mathcal{X})$ denotes a space of continuous functions on $\mathcal{X} \subset \mathcal{R}^{d_x}$. The bounding functions L and U in $\mathcal{Q}_{data}$ can be estimated consistently from the *data*. The shape restriction functional $\bar{R} : \mathcal{C}(\mathcal{X}) \to \mathcal{R}^{d_R}$ is known, and is implied by the economic *model*. Our goal is to perform inference on $\bar{f}(q)$ and the set containing it, i.e.,[2]

$$\left[\inf_{q \in \mathcal{Q}} \bar{f}(q), \sup_{q \in \mathcal{Q}} \bar{f}(q)\right]. \tag{2}$$

In our welfare-analysis setting, q is the structural choice probability, $\bar{f}$ is the average CV, and $\bar{R}$ comes from Slutsky-type restrictions.

2.1 Binary choice model

Consider an individual with income Y, who faces the choice between two options labelled 1 and 0. Individual utilities from choosing 1 and 0 are respectively $U_1(W, \eta)$ and $U_0(W, \eta)$, where W is the quantity of numeraire, which the individual consumes in addition to the binary good, and η represents unobserved taste; other observed individual-specific characteristics are implicitly controlled for. We allow the unobserved heterogeneity η to be of unknown dimension and enter the utility functions in any arbitrary way. The budget constraint is $PQ + W = Y$, where $Q \in \{0, 1\}$ represents the binary choice with price P. So the individual chooses 1 if and only if $U_1(Y - P, \eta) > U_0(Y, \eta)$, i.e., $Q \equiv Q(P, Y, \eta) = \mathbf{1}\{U_1(Y - P, \eta) > U_0(Y, \eta)\}$.

Now suppose the price of option 1 increases from p_0 to p_1, with the marginal distribution of η remaining unchanged. We wish to calculate the marginal distributions of the welfare change evaluated at fixed income y_0 corresponding to this price change. In particular, the *compensating variation* measures the income CV to be given to an η type individual at income y_0, facing price p_1 so that her maximized utility with this additional income equals her maximized utility when the price was p_0 and income was y_0. Then CV as a function of

[2]In general, the set $\left[\inf_{q \in \mathcal{Q}} \bar{f}(q), \sup_{q \in \mathcal{Q}} \bar{f}(q)\right]$ is a superset or a convex hull of the identified set of $\bar{f}(q)$ defined by $\{\bar{f}(q) : q \in \mathcal{Q}\}$. We could call the former set in (2) an "outer" identified set. In the rest of the paper, we omit the term "outer" and use "identified set" for (2) for simplicity without loss of clarity. We thank one anonymous referee to point this out.

(p_0, p_1, y_0, η) solves the equation:

$$\max\{U_0(y_0 + CV, \eta), U_1(y_0 + CV - p_1, \eta)\} = \max\{U_0(y_0, \eta), U_1(y_0 - p_0, \eta)\}.$$

The *equivalent variation* measures the income EV to be subtracted from this individual so that her maximized utility at price p_0 equals that when price was p_1. The EV as a function of (p_0, p_1, y_0, η) solves the equation:

$$\max\{U_0(y_0 - EV, \eta), U_1(y_0 - EV - p_0, \eta)\} = \max\{U_0(y_0, \eta), U_1(y_0 - p_1, \eta)\}.$$

Our analysis focuses on the CV; the results for the EV are analogous.

Bhattacharya (2015) shows that the marginal distributions of individual CV and EV can be expressed as closed-form transformations of choice probabilities. Specifically, denoting the *structural* choice probability at a hypothetical price and income (p, y) as

$$q(p, y) \equiv \int \mathbf{1}\{U_1(y - p, \eta) > U_0(y, \eta)\} dF_\eta(\eta),$$

the main result in Bhattacharya (2015) is the following:

Result 1 (Theorem 1 in Bhattacharya (2015)) *Assume $U_0(W, \eta)$ and $U_1(W, \eta)$ are strictly increasing in W for each η. Consider a price rise from p_0 to p_1. Then across individuals with income y_0, the marginal distribution of the CV is given by*

$$\Pr(CV \leq a) = \begin{cases} 0 & \text{if } a < 0, \\ 1 - q(p_0 + a, y_0 + a) & \text{if } 0 \leq a < p_1 - p_0, \\ 1 & \text{if } a \geq p_1 - p_0. \end{cases}$$

Given this, we focus on the average CV: $\mathbb{E}[CV] = \int_{p_0}^{p_1} q\left(p, y_0 + p - p_0\right) dp$. More general structural objects of interest include functionals of $q(\cdot, \cdot)$, such as the marginal effect of changing income on the choice probability or quantiles of the CV.

Our first result, stated as Proposition 1, provides a Slutsky-type restriction for the choice probability.

Proposition 1 (Slutsky restriction) *Under the conditions of Result 1, monotonicity of the marginal distributions of the EV and the CV in Result 1 is equivalent to the restriction*

$$q(p, y) \geq q(p + b, y + c), \text{ for any } b \geq c \geq 0. \tag{3}$$

When $q(p,y)$ is differentiable, the restriction (3) holds if and only if

$$\frac{\partial}{\partial p}q(p,y)+\frac{\partial}{\partial y}q(p,y)\leq 0 \text{ and } \frac{\partial}{\partial p}q(p,y)\leq 0, \text{ for any } (p,y). \tag{4}$$

The Slutsky condition (3) follows from the assumption that $U_0(\cdot,\eta)$ and $U_1(\cdot,\eta)$ are strictly monotone. This result is new to the literature, to the best of our knowledge. It may be contrasted with the case of a continuous good with demand $q_c(p,y,\eta)$, where the conventional Slutsky equation is $\partial q_c(p,y,\eta)/\partial p + q_c(p,y,\eta)\partial q_c(p,y,\eta)/\partial y \leq 0$. Defining $Q(p,y,\tau)$ as the τth quantile of demand, Dette, Hoderlein, and Neumeyer (2016) have shown that the Slutsky condition also holds for quantile demand, i.e., for all p,y, and for all $\tau\in[0,1]$,

$$\frac{\partial}{\partial p}Q(p,y,\tau)+Q(p,y,\tau)\frac{\partial}{\partial y}Q(p,y,\tau)\leq 0.$$

Observe that this last inequality is *nonlinear* in Q, in contrast to the linear inequalities in (4) for the binary good.

2.2 Identification

We now state two assumptions, viz., exogeneity and monotonicity, under which we will derive our first set of bounds.

Assumption 1 *(i) Price and income are jointly independent of the unobserved preference heterogeneity η (ii) The structural choice probability $q(p,y)$ is increasing in y for each p and is differentiable in (p,y).*

Independence of preferences and budge sets (conditional on covariates) has been maintained in this literature (c.f., Hausman and Newey (2016)). Below, we provide a brief discussion on relaxing it. The second assumption states that alternative 1 is a normal good on average. That is, if income goes up with price remaining fixed, the probability of buying good 1 goes up. Note that we need this assumption to hold only on average; alternative 1 being a normal good for all consumers is sufficient but not necessary for this assumption to hold.

In order to implement the Bhattacharya (2015) formulae for calculating welfare distributions, one needs to observe the price faced by each consumer, as well as individual income. However, when income is censored into intervals, the structural choice probability $q(p,y)$, and consequently, the CV/EV distributions cannot be point-identified. The first step, therefore, is to construct the identified set for $q(p,y)$. We then show how revealed preference restrictions would make the set tighter.

Toward that end, suppose that incomes are recorded in fixed intervals. That is, the true income belongs to a sequence of intervals or brackets given deterministically by administrators and not chosen by the respondents. This setup of fixed intervals is precisely how incomes are recorded in many survey data. We assume that a set of nonrandom grid points $\{y^1, ..., y^K\}$ partitions the support of income into disjoint intervals: $\mathcal{Y} = [y^0, y^{K+1}] = \bigcup_{k=0}^{K} \mathcal{Y}_k$, where the interval $\mathcal{Y}_k \equiv [y^k, y^{k+1})$ for $k = 0, 1, ..., K-1$ and $\mathcal{Y}_K \equiv [y^K, y^{K+1}]$.

We observe the realizations of (Q, P), the income interval where the unobserved income Y belongs. When price and income are observed and Assumption 1(i) holds, the structural choice probability is point-identified by the conditional expectation of the binary outcome given the covariates $q(p, y) = \mathbb{E}[Q|P = p, Y = y]$. But given the sampling scheme involving interval censored income, we can point-identify the conditional choice probability given price p and interval $[y^1, y^2)$, which we denote by

$$\pi\big(p, [y^1, y^2)\big) \equiv \Pr\left(\text{choose } 1|P = p, Y \in [y^1, y^2)\right).$$

From the data, we can nonparametrically estimate $\pi\big(p, [y^1, y^2)\big)$ using individuals whose income lies in $[y^1, y^2)$.

We will assume that the choice probability also satisfies restrictions from economic theory, viz., the Slutsky restriction in Proposition 1 and monotonicity in Assumption 1(ii). Putting all of this together, we define the following identified set with shape restrictions:

$$\mathcal{Q} \equiv \mathcal{Q}_{data} \cap \mathcal{Q}_{model}, \text{ where}$$
$$\mathcal{Q}_{model} \equiv \left\{ q \in \mathcal{C}(\mathcal{P} \times \mathcal{Y}) : \frac{\partial}{\partial p} q(p, y) + \frac{\partial}{\partial y} q(p, y) \leq 0 \text{ and } \frac{\partial}{\partial y} q(p, y) \geq 0, \text{ for all } (p, y) \in \mathcal{P} \times \mathcal{Y} \right\}.$$

As the choice probabilities are partially identified, so are the welfare effects based on CV/EV, e.g., the average CV at income y_0 for a price change from p_0 to p_1 in (1).

Nonparametric Formulation In principle, one can attempt to nonparametrically estimate the set $\mathcal{Q}$ and explore inference for functionals with domain $\mathcal{Q}$. This problem is equivalent to the following constrained optimization problem. Let $f(\cdot, \cdot)$ denote the unobserved joint density function of (P, Y), and $\Pr(p, \mathcal{Y}_k)$ denote the observed joint distribution of $(P, \mathcal{Y}_k)$. For a price increase from p_0 to p_1, the upper bound of the expected CV can be obtained by solving the constrained, infinite-dimensional optimization problem:

$$\max_{f(\cdot,\cdot), q(\cdot,\cdot)} \int_0^{p_1 - p_0} q\left(p_0 + a, y_0 + a\right) da,$$

s.t.

$$\int_{\mathcal{Y}_k} q\left(p,y\right) \times f\left(p,y\right) dy = \underbrace{\pi(p,\mathcal{Y}_k)\Pr(p,\mathcal{Y}_k)}_{\text{Observed}},\ k=0,1,...,K,$$

$$\int_{\mathcal{Y}_k} f(p,y)dy = \underbrace{\Pr(p,\mathcal{Y}_k)}_{\text{Observed}},\ k=0,1,...,K, \tag{5}$$

$$\frac{\partial}{\partial y} q\left(p,y\right) > 0,\ \text{and}\ \frac{\partial}{\partial y} q\left(p,y\right) + \frac{\partial}{\partial p} q\left(p,y\right) < 0.$$

Since an admissible choice of f is where the entire probability mass within each observed income interval is concentrated at the right end-points, we can satisfy the first constraint in (5) by setting $q\left(p,y\right) = q\left(p,y^{k+1}\right)$ for all $y \in \mathcal{Y}_k$. Assumption 1(ii), the integrand $q\left(p,y_0+p-p_0\right)$ in the objective function can take a value at least as high as $q\left(p,y^{k+1}\right)$ where y^{k+1} is the right end point of the observed income interval containing y_0+p-p_0.[3] However, $q\left(p,y^{k+1}\right)$ is also unobserved, and we will need to find an upper bound on it. Note that the probability mass on the next interval to the right, i.e. $\left[y^{k+1},y^{k+2}\right]$ can be concentrated at y^{k+1}. Therefore, the sharp upper bound for $q\left(p,y_0+p-p_0\right)$ is given by $q\left(p,[y^{k+1},y^{k+2})\right)$, i.e. $\pi(p,\mathcal{Y}_{k+1})$. A similar idea works for the lower bound. For example, for $y \in [y^2,y^3)$, $\pi(p,[y^1,y^2)) \le q(p,y) \le \pi(p,[y^3,y^4))$, as illustrated in Figure 1. Formally, the identified set of q in the presence of interval-censored income and under Assumption 1 is

$$\begin{aligned}
&\mathcal{Q}_{data} \equiv \left\{q : L(p,y) \le q(p,y) \le U(p,y),\ \text{for all}\ (p,y) \in \mathcal{P}\times\mathcal{Y}\right\}, \text{where} \\
&L(p,y) = \sum_{k=1}^{K} \pi(p,\mathcal{Y}_{k-1})\mathbf{1}\left\{y \in \mathcal{Y}_k\right\}\ \text{and} \\
&U(p,y) = \sum_{k=0}^{K-1} \pi(p,\mathcal{Y}_{k+1})\mathbf{1}\left\{y \in (y^k,y^{k+1}]\right\} + \mathbf{1}\{y > y^K\}.
\end{aligned} \tag{6}$$

As the bounding functions $(L(p,y),U(p,y))$ are between zero and one, $\mathcal{Q}_{data}$ is bounded and convex. Under Assumption 1, the identified set $\mathcal{Q}_{data}$ describes all the information available from the data. The above argument shows $\mathcal{Q}_{data}$ is sharp (see Proposition 1 in Manski and Tamer (2002)).

The bounding functions (L,U) of q, defined in (6) can be used to compute the bounds for the average CV. However, since the bounding functions are averages of the conditional choice probabilities over the income interval $\pi(\cdot,\mathcal{Y}_k)$, it is nontrivial to impose the Slutsky restriction in nonparametric estimation of $\pi(p,\mathcal{Y}_k)$. But without the Slutsky restrictions, the resulting identified set for the average CV based on $\pi(\cdot,Y_k)$ might not be tight. To impose

[3]By a similar logic, the integrand will take a value no smaller than $q\left(p,y^L\left(p\right)\right)$ where $y^L\left(p\right)$ is the left end point of the observed income interval containing y_0+p-p_0, etc.

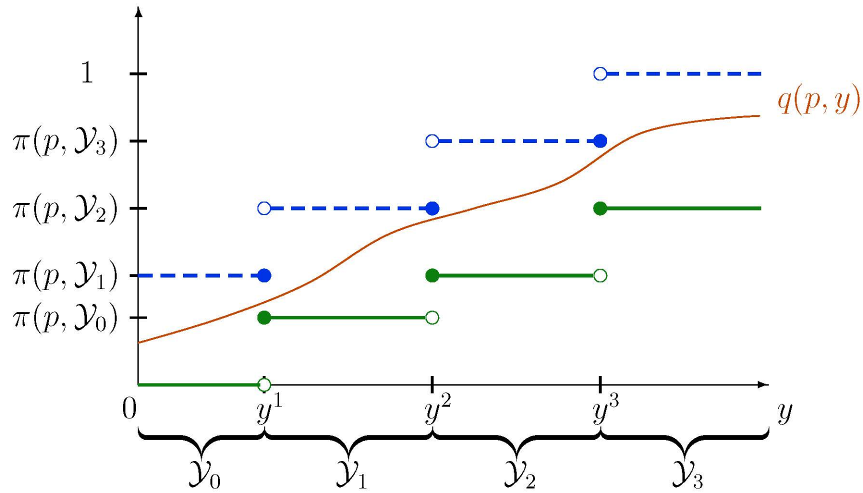

Figure 1: The curve is the true structural choice probability $q(p, y)$ at a price p. The horizontal axis is the income variable y, whose support is partitioned to four intervals by a set of grid points $\{y^1, y^2, y^3\}$. The upper bounding function $U(p, y)$ (blue dashed line) and the lower bounding function $L(p, y)$ (green solid line) are composed of the conditional choice probabilities $\pi(p, \mathcal{Y}_k) \equiv \mathbb{E}[Q|P = p, Y \in \mathcal{Y}_k]$ for $k = 0, 1, 2, 3$.

Slutsky restrictions while allowing for interval-data, we begin with a parametric specification of $q(\cdot)$ in Section 3.

3 Parametric modelling

We consider a parametric model $q(x) = \Phi(x^\top \beta)$, where Φ is a specified link function. For example, Φ can be the normal C.D.F. for a probit model. Assumption of parametric forms for the outcome equation is ubiquitous in the interval data literature, c.f. Manski and Tamer (2002), Wan and Xu (2015) (we discuss mis-specification issues in Section 6, below). Now, given the parametric model, since q depends only on β, we can re-write our functional of interest as an operator f on the parameter β by $f(\beta) \equiv \bar{f}(\Phi(x^\top \beta))$. The linear shape restriction $\bar{R}(q) \leq 0$ implies there exists a function R such that $R(\beta) \leq 0$ if and only if $\bar{R}(\Phi(X^\top \beta)) \leq 0$ almost surely. Then our problem in (2) becomes one of conducting inference on a nonlinear function of β that is partially identified by conditional moment inequalities and shape restrictions $R(\beta) \leq 0$.

In particular, for our binary choice model, the covariates $X = (1, P, Y)^\top$. A parametric specification of the choice probability $q(p, y)$ is given by $\Phi(\beta_0 + \beta_P p + \beta_Y y)$. Then the constraints in $\mathcal{Q}_{data}$ in (6) can be expressed by $2K$ conditional moment inequalities:

$$\mathbb{E}[(-Q+\Phi(\beta_0+\beta_P P+\beta_Y y^k))\mathbf{1}\{Y\in\mathcal{Y}_k\}|P]\le 0 \text{ and}$$
$$\mathbb{E}[(Q-\Phi(\beta_0+\beta_P P+\beta_Y y^{k+1}))\mathbf{1}\{Y\in\mathcal{Y}_k\}|P]\le 0, \text{ for } k=1,...,K. \quad (7)$$

We impose the shape restrictions implied by economic theory: $\beta_Y \geq 0$ and $\beta_P+\beta_Y \leq 0$, i.e., $R(\beta) = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix} (\beta_0, \beta_P, \beta_Y)^\top \leq 0$. This setup covers many general objects of interest. For example, for the marginal effect of income on the choice probability, $f(\beta) = \beta_Y \phi((1,p,y)\beta)$; for the average CV for a price change from p_0 to p_1 at income y_0, we have that $f(\beta)=\int_0^{p_1-p_0}\Phi(\beta_0+\beta_P(p_0+a)+\beta_Y(y_0+a))da$. Therefore the set in (2) becomes

$$\left[\inf_{q\in\mathcal{Q}}\bar{f}(q), \sup_{q\in\mathcal{Q}}\bar{f}(q)\right] = \left[\min_{\beta\in\mathcal{B}} f(\beta), \max_{\beta\in\mathcal{B}} f(\beta)\right], \text{where } \mathcal{B}\equiv\mathcal{B}_{data}\cap\mathcal{B}_{model},$$
$$\mathcal{B}_{data}\equiv\{\beta:(7)\text{ holds.}\} \text{ and } \mathcal{B}_{model}\equiv\{\beta:\beta_Y\geq 0,\beta_Y+\beta_P\leq 0\}. \quad (8)$$

We propose a tractable approach that transforms the problem in (2) to perform inference on $f(\beta)$ and its identified set, denoted by $\left[f^l, f^u\right]$. In the following, we first characterize the identified set for β. Then we search for the maximum and minimum of nonlinear functions of β over the identified set subject to shape restrictions.

Estimation Overview Given that $q(x)=\Phi\left(x^\top\beta\right)$, the parameter β can be written as the minimizer of a quadratic loss function,

$$\beta = \arg\min_{b\in\mathcal{R}^{d_x}} \mathbb{E}_\mu\left[\left(\Phi^{-1}(q(X))-X^\top b\right)^2\right] = \left(\mathbb{E}_\mu\left[XX^\top\right]\right)^{-1}\mathbb{E}_\mu\left[X\Phi^{-1}(q(X))\right], \quad (9)$$

where $E_\mu[g(X)]$ denotes the expectation of a known function $g(X)$, when X is distributed according to a continuous measure μ on the support of the covariates X. Since the true distribution of X is unknown due to interval censoring of income, there is no unique way to choose the measure μ (analogous to choosing the weighting matrix for constructing the criterion function in moment (in)equality models), except to ensure that μ is consistent with the observed feature of the data. We discuss more on the choice of μ below.

Given μ, as q is partially identified by $\mathcal{Q}=\mathcal{Q}_{data}\cap\mathcal{Q}_{model}$, we can define a set of observationally equivalent parameter vectors β as follows:

$$\begin{aligned}
\mathcal{B}_{\mu} &\equiv \mathcal{B}_{data,\mu} \cap \mathcal{B}_{model}, \text{where} \\
\mathcal{B}_{data,\mu} &\equiv \left\{ \beta = \left(\mathbb{E}_{\mu} \left[X X^{\top} \right] \right)^{-1} \mathbb{E}_{\mu} \left[X \Phi^{-1}(q(X)) \right] \text{ for some } q \in \mathcal{Q}_{data} \right\} \text{and} \\
\mathcal{B}_{model} &\equiv \{\beta : R(\beta) \leq 0\}.
\end{aligned} \tag{10}$$

This results in an identified set for $f(\beta)$:

$$\left[\min_{\beta \in \mathcal{B}_{\mu}} f(\beta), \max_{\beta \in \mathcal{B}_{\mu}} f(\beta) \right] \tag{11}$$

that contains the set defined in (8).

We now show how to calculate this identified set in a computationally tractable way. We use the well-known fact that the boundary of a convex, compact set is determined by the hyperplanes that are tangent to it (e.g., Rockafellar (1970), Chapter 13).

Accordingly, let the unit sphere in $\mathcal{R}^{d_x}$ be denoted $\mathcal{S} \equiv \{s \in \mathcal{R}^{d_x} : \|s\| = 1\}$. The *support function* of the set $\mathcal{B}_{data,\mu}$ is the set of tangent hyperplanes given by $\{\max_{b \in \mathcal{B}_{data,\mu}} s^{\top} b : s \in \mathcal{S}\}$. The boundary point of $\mathcal{B}_{data,\mu}$ that maximizes $s^{\top} b$ has a closed-form expression given by

$$\begin{aligned}
&\bar{\beta}(s) \equiv \arg \max_{b \in \mathcal{B}_{data,\mu}} s^{\top} b = \Sigma^{-1} \mathbb{E}_{\mu}[X \ \gamma_s(X)], \text{where} \\
&\Sigma \equiv \mathbb{E}_{\mu}[X X^{\top}] \text{ and } \gamma_s(X) \equiv \Phi^{-1}(U(X)) \mathbf{1}\{s^{\top} \Sigma^{-1} X \geq 0\} + \Phi^{-1}(L(X)) \mathbf{1}\{s^{\top} \Sigma^{-1} X < 0\}.
\end{aligned} \tag{12}$$

As $q(X)$ is partially identified by an interval $[L(X), U(X)]$, the constructed variable $\gamma_s(X)$ switches between the bounds $\Phi^{-1}(L(X))$ and $\Phi^{-1}(U(X))$ depending on the sign of $s^{\top} \Sigma^{-1} X$. Then the maximizer $\bar{\beta}(s)$ characterizes the boundary of $\mathcal{B}_{data,\mu}$ by tracing out all direction $s \in \mathcal{S}$.

Having the closed-form expression for the boundary points of $\mathcal{B}_{data,\mu}$, we now characterize the interior points. As the set $\mathcal{B}_{data,\mu}$ is strictly convex and compact by construction due to a continuous measure μ, there is a unique point on the boundary of $\mathcal{B}_{data,\mu}$ that intersects its supporting hyperplane in a given direction (e.g., Bontemps, Magnac, and Maurin (2012)). So each interior point is a convex combination of the boundary points. It follows that the set $\mathcal{B}_{data,\mu}$ in (10) can be expressed as

$$\mathcal{B}_{data,\mu} = \left\{ \beta(s,t) : \beta(s,t) \equiv t\bar{\beta}(s) + (1-t)\bar{\beta}(s_0), \text{ for } s \in \mathcal{S}, t \in [0,1] \right\}$$

for any $s_0 \in \mathcal{S}$. This is a key step for imposing shape restrictions on β over the parameter space $[0,1]$ and for finding the max/min of nonlinear functions of β. The optimization problem in (11) becomes

$$\left[\min_{\substack{(s,t)\in\mathcal{S}\times[0,1]\\ R(\beta(s,t))\le 0}} f(\beta(s,t)),\ \max_{\substack{(s,t)\in\mathcal{S}\times[0,1]\\ R(\beta(s,t))\le 0}} f(\beta(s,t))\right] \equiv \left[f^l, f^u\right].$$

The expression based on $\beta(s,t)$ changes the parameter space of the optimization problem to a nonrandom set $\mathcal{S}\times[0,1]$. Figure 2 illustrates the idea by projecting the identified set to the joint identified set for (β_Y, β_P).

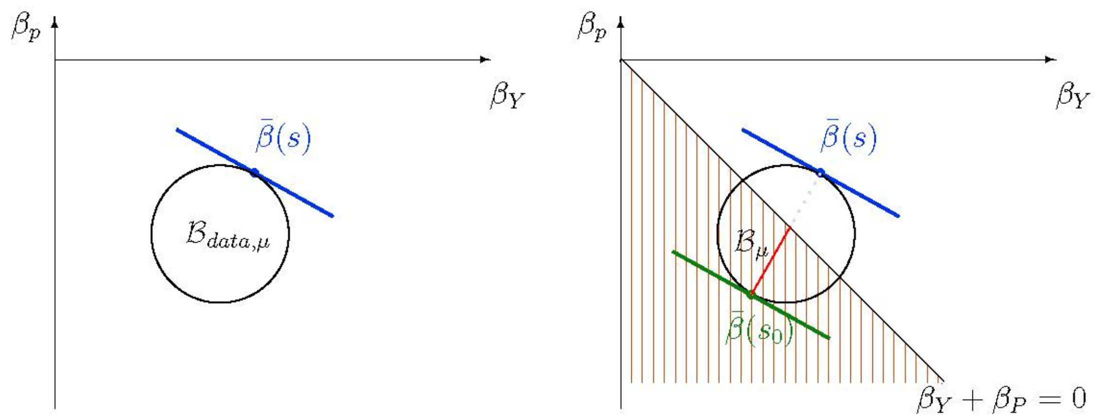

Figure 2: The joint identified set for (β_Y, β_P). In the left panel, the boundary point $\bar{\beta}(s) = \arg\max_{b\in\mathcal{B}_{data,\mu}} s^\top b$ for $s = \left(0, 0.8, 0.6\right)^\top$. In the right panel, the shaded area satisfies the shape restriction $R(\beta) = (-\beta_Y, \beta_P + \beta_Y)^\top \le 0$. The red solid line of the line segment of $\bar{\beta}(s_0)$ and $\bar{\beta}(s)$ lies in $\mathcal{B}_\mu$.

In sum, the estimands are the lower endpoint given by

$$f^l \equiv \min_{\beta\in\mathcal{B}_\mu} f(\beta) = \min_{s\in\mathcal{S},t\in[0,1]} f(\beta(s,t)) \text{ subject to } R(\beta(s,t)) \le 0, \tag{13}$$

and the upper endpoint f^u that is defined analogously by changing min to max in (13). Our estimator is a straightforward sample analog as outlined in the procedure below. To perform inference, we introduce a bootstrap method in Section 4.2.

Estimation Detail

Step 1. **Bounding functions L and U.**

For each income interval $\mathcal{Y}_k$, estimate the conditional choice probability given price p by a parametric probit/logit or a nonparametric kernel/sieve estimator, denoted by $\hat{\pi}(p, \mathcal{Y}_k)$.

Obtain $\Phi^{-1}(\hat{\pi}(p, \mathcal{Y}_k))$ by a transformation of the inverse link function, which constitutes the bounding functions $\hat{l}(p, y)$ and $\hat{u}(p, y)$ for $\Phi^{-1}(q(p, y))$.

Step 2. **Identified set for β.** For each direction on the unit sphere $s \in \mathcal{S} = \{s \in \mathcal{R}^{d_x} : \|s\| = 1\}$, the boundary point of the identified set for β is estimated by an OLS-type estimator

$$\hat{\bar{\beta}}(s) = \Sigma^{-1}\mathbb{E}_\mu[X\ \hat{\gamma}_s(X)], \text{where } \Sigma = \mathbb{E}_\mu\left[XX^\top\right] \text{ and}$$
$$\hat{\gamma}_s(X) = \hat{u}(X)\mathbf{1}\{s^\top\Sigma^{-1}X \geq 0\} + \hat{l}(X)\mathbf{1}\{s^\top\Sigma^{-1}X < 0\}.$$

Estimate $\mathbb{E}_\mu[g(X)]$ by the sample analogue $m^{-1}\sum_{j=1}^m g(X_{\mu j})$ with a random sample $\{X_{\mu j} : j = 1, ..., m\}$ from a continuous measure μ.

Step 3. **Identified set for $f(\beta)$.** Fixing one direction $s_0 \in \mathcal{S}$, define a convex combination $\hat{\beta}(s, t) \equiv t\hat{\bar{\beta}}(s) + (1-t)\hat{\bar{\beta}}(s_0)$ for $t \in [0, 1]$. The lower endpoint is estimated by

$$\hat{f}^l = \min_{s\in\mathcal{S}} \min_{t\in[0,1]} f(\hat{\beta}(s,t)) \text{ subject to } R(\hat{\beta}(s,t)) \leq 0.$$

The upper endpoint $\hat{f}^u$ is estimated the same by changing min to max.

In our welfare analysis for the average CV for a price change from p_0 to p_1 at income y_0, $X = (1, P, Y)^\top$ and

$$\hat{f}^l = \min_{s\in\mathcal{S}} \min_{t\in[0,1]} \int_0^{p_1-p_0} \Phi\left((1, p_0 + a, y_0 + a)\,\hat{\beta}(s,t)\right) da$$
$$\text{subject to } \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix} \hat{\beta}(s,t) \leq 0.$$

Choice of μ Researchers can specify a continuous measure μ, provided that the resulting set $\mathcal{B}_{data,\mu}$ is strictly convex and compact. For the general problem in (2), a natural estimator of μ would be the empirical distribution of the covariates X (e.g., Kline and Santos (2013)). Such choice is not feasible in our application, as income is interval-censored. In our setup, we can use the empirical distribution of the price and interval-valued income, assign uniform density within each income interval,[4] and discard the extreme income intervals $\mathcal{Y}_0$ and $\mathcal{Y}_K$. This is because the lower bound of $q(p, y)$ for $y \in \mathcal{Y}_0$ is the worst-case bound, 0. The transformations by the inverse link function $\Phi^{-1}(0)$ might not be finite, e.g., a normal C.D.F. Similarly, the upper bound $q(p, y)$ for $y \in \mathcal{Y}_K$ is the worst-case bound, 1. These bounds are

[4]More specifically, this measure μ specifies the joint probability of (p, y) to be $\sum_{k=0}^K f_{P|\mathcal{Y}_k}(p, y)\mathbf{1}\{y \in \mathcal{Y}_k\}/|\mathcal{Y}_k|$, where $|\mathcal{Y}_k|$ is the length of the deterministic interval $\mathcal{Y}_k$. The conditional density function of price given income interval $f_{P|\mathcal{Y}_k}(p, y)$ is consistently estimated by its empirical counterpart using the empirical distribution function. And for each observation we draw income value from a uniform distribution over that income interval.

not informative and implied by the link function bounded between 0 and 1. So the measure μ also summarizes the information contained in $\mathcal{Q}_{data}$. Instead of simulating from a uniform distribution within each income interval, we could also use the marginal distribution of income that might be available from other datasets. In practice, we can check the sensitivity of the estimates to different μ. In our empirical application in Section 7, we compare the estimates using the empirical measure described above and using an uniform measure that assigns equal weight over the support of income and price.

We view the measure μ as somewhat similar in spirit to the choice of instrument functions that transform conditional moment inequality/equality restrictions to unconditional ones (Andrews and Shi (2013), Chen (2007)), and the choice of weighting matrix in forming the optimization criterion function thereof. In theory, one would aim to exhaust all possible μ such that $\cap_\mu B_{data,\mu} = B_{data} \equiv \{\beta : (7) \text{ holds.}\}$ to avoid losing information in the conditional moment inequalities. But it is not obvious as to how one may implement this in practice.

3.1 Relation to existing literature

An important feature of our method is that we characterize the interior of the identified set of β, because the optimizer might locate at the interior. Toward that end, we build on prior work by Beresteanu and Molinari (2008), Bontemps, Magnac, and Maurin (2012), Chandrasekhar, Chernozhukov, Molinari, and Schrimpf (2012), and Kline and Santos (2013). Their focus is on the identified set for the underlying function (i.e., $q(\cdot)$ known up to β), which they construct by characterizing the boundary of the identified set of β using the support functions. In contrast, our interest is in functionals of partially identified functionals of $q(\cdot)$, subject to $q(\cdot)$ satisfying a set of shape restrictions. Thus the above methods cannot be directly applied to our problem. Our approach, instead, is to express the interior points in terms of support functions, and then to obtain a closed-form expression for the interior points of the identified set for β. These, in turn, yield simple expressions for bounds on the scalar parameter of interest, viz. average welfare.

Some results on inference for functions of partially identified parameters appear in Bugni, Canay, and Shi (2017) and Kaido, Molinari, and Stoye (2017) for *unconditional* moment inequality models; these cannot be applied directly to our *conditional* moment inequality setting in (7). It may be possible to transform our conditional moment inequalities to a finite number of unconditional moment inequalities (c.f. Andrews and Shi (2013)), and then to apply one of these methods for valid inference. But such extensions appear more complicated for our set-up, relative to a direct and simpler alternative, which is feasible here due to the special structure of our problem that delivers a compact and convex identified set.

In a related paper, Chernozhukov, Newey, and Santos (2015) study inference for functionals of parameters defined by very general conditional moment restrictions and shape constraints. Their sieve generalized method of moments J-test can be applied to both a parametrically specified $q(\cdot)$ in (8) and nonparametrically specified $q(\cdot)$ in (2), and encompasses the case where the function q is point-identified. However, the test-inversion approach can be computationally demanding, as it requires searching for all points not rejected by a hypothesis-test. On the other hand, this approach can cover more general types of identified sets, beyond what we need for our problem. In particular, in our setting, the function of interest is set-identified by a *compact convex* set (c.f. the set $\mathcal{Q}_{data}$ in (6)), which motivates the support-function approach, and enables us to provide a direct inference procedure that is computationally simpler.[5] Finally, when the parametric model is mis-specified, our approach can be interpreted via a "best-parametric approximation" route (see Section 6); the interpretation of "pseudo-true" parameters using the above-mentioned alternative methods is less obvious.

For the case where the function q is *point-identified* with shape restrictions, Horowitz and Lee (2018) investigate estimation and inference under shape restrictions. That is not the case in our set-up, where interval censoring of income prevents the choice probability from being point-identified even without any shape restrictions.

4 Inference

This section presents the theory of inference based on the estimation method described in Section 3. Estimating the endpoints $\left[f^l, f^u\right]$ in (13) is a constrained stochastic optimization problem. Our main result is the limiting distribution of $\left(\hat{f}^l, \hat{f}^u\right)$ in Theorem 1 in Section 4.1. The limiting distribution is not pivotal and depends on the binding constraints. Nonetheless, Hadamard directional differentiability and a corresponding delta method allow us to apply the bootstrap proposed by Dümbgen (1993), Fang and Santos (2019), and Hong and Li (2018). Theorem 2 in Section 4.2 shows the validity of this bootstrap method in our context.

[5] To obtain a confidence region, our bootstrap-based inference method with B bootstrap replications involves $2(B+1)$ nonlinear optimization problems, each of which can be easily solved by routines available in standard packages. In the test-inversion approach, designed for general identified sets, the bootstrap procedure to simulate the critical values is often repeated over all the hypothetical values of $f(\beta)$ in the null hypothesis.

4.1 Asymptotic theory

We express our estimator via a constrained stochastic optimization problem. Recall that our estimand is

$$f^l = \min_{s\in\mathcal{S},t\in[0,1]} f(\beta(s,t)) \text{ subject to } R(\beta(s,t)) \leq 0, \text{ where } \beta(s,t) = t\bar{\beta}(s) + (1-t)\bar{\beta}(s_0) \tag{14}$$

for a $s_0 \in \mathcal{S}$ and $\bar{\beta}(s)$ is defined in (12). For ease of exposition, the presentation focuses on the lower endpoint $\hat{f}^l$. The result for the upper endpoint bound $\hat{f}^u$ is symmetric by replacing the objective function with $-f$ in (14). Notice that when the objective function f is convex and nonlinear, $-f$ is generally nonconvex. To deal with these cases, we allow for a *nonconvex* objective function for the minimization problem in (14). Our main theoretical contribution is Theorem 1 that shows the asymptotic properties of $\hat{f}^l$ by Hadamard directional differentiability.

The estimand f^l can be expressed as a mapping from a function space to the real line $\mathbb{D}_\phi \to \mathcal{R}$:

$$\phi(\theta) \equiv \min_{\substack{(s,t)\in\mathcal{S}\times\mathcal{T} \\ t\in[0,1],\ \theta^{(2)}(s,t)\leq 0}} \theta^{(1)}(s,t), \tag{15}$$

where $\theta = \left(\theta^{(1)}, \theta^{(2)\top}\right)^\top \in \mathbb{D}_\phi \equiv \mathcal{C}(\mathcal{S}\times[0,1]) \times \cdots \times \mathcal{C}(\mathcal{S}\times[0,1]) \subset \mathcal{R}^{d_R+1}$ and $\mathcal{T} \subset \mathcal{R}$ is a convex compact set. Denote the true $\theta_0(s,t) \equiv \left(f\left(\beta(s,t)\right), R(\beta(s,t))^\top\right)^\top$ and the estimator $\hat{\theta}(s,t) = \left(f\big(\hat{\beta}(s,t)\big), R\big(\hat{\beta}(s,t)\big)^\top\right)^\top$. Then the estimands are $f^l = \phi(\theta_0)$ and $f^u = \phi\left(\left(-\theta_0^{(1)}, \theta_0^{(2)\top}\right)^\top\right)$. Our sample analog estimator outlined in Section 3 is denoted by $\phi(\hat{\theta})$.

To apply the delta method to analyze $\phi(\hat{\theta})$, we require $\sqrt{n}(\hat{\beta}(s,t) - \beta(s,t))$ to weakly converge to a tight Gaussian process indexed by (s,t), as in Fang and Santos (2019). In order to devote maximum space to what is new in our work, we assume the availability of such a preliminary estimator for the support function process, e.g., Chandrasekhar, Chernozhukov, Molinari, and Schrimpf (2012).[6] Assumption 2 states the corresponding high-level assumption.

Let $l^\infty(\mathcal{X})$ be a space of bounded functions on $\mathcal{X}$.

Assumption 2 (Support function process) *(i) $\mathcal{B}_{data,\mu}$ is non-empty, strictly convex, and*

[6] The literature mostly focuses on the support function $s^\top\bar{\beta}(s)$ rather than the boundary point $\bar{\beta}(s)$, e.g, Chandrasekhar, Chernozhukov, Molinari, and Schrimpf (2012) and Kline and Santos (2013). The result in Chandrasekhar, Chernozhukov, Molinari, and Schrimpf (2012) should be modified to the weak convergence of $\sqrt{n}(\hat{\bar{\beta}}(s) - \bar{\beta}(s))$.

compact; (ii) There exist consistent estimators of $(l(x), u(x))$ *such that* $\sqrt{n}(\hat{\bar{\beta}} - \bar{\beta}) \Longrightarrow \mathbb{G}_b$, *where* $\mathbb{G}_b$ *is a tight Gaussian process on* $l^{\infty}(\mathcal{S}) \times \cdots \times l^{\infty}(\mathcal{S}) \subset \mathcal{R}^{d_x}$*; (iii) The derivatives of* f *and* R *exist and are non-zero and continuously differentiable such that* $\sqrt{n}(\hat{\theta} - \theta_0) \Longrightarrow \mathbb{G}_0$, *where* $\mathbb{G}_0$ *is a tight Gaussian process on* $l^{\infty}(\mathcal{S} \times [0,1]) \times \cdots \times l^{\infty}(\mathcal{S} \times [0,1]) \subset \mathcal{R}^{d_R+1}$ *with a non-degenerate covariance function.*

Assumption 2(i) is implied by properly choosing the measure μ as discussed in the previous section. So $\Sigma = \mathbb{E}_{\mu}\left[XX^{\top}\right]$ exists and $\bar{\beta}(s)$ is well-defined for all $s \in \mathcal{S}$.[7] For the preliminary nonparametric estimators $(\hat{l}(x), \hat{u}(x))$, Chandrasekhar, Chernozhukov, Molinari, and Schrimpf (2012) suggest using series logit estimation in Hirano, Imbens, and Ridder (2003).[8] By the standard delta method, Assumption 2(ii) and (iii) imply $\theta_0(s,t) = \left(f(\beta(s,t), R(\beta(s,t))^{\top}\right)^{\top}$ has a regular estimator $\hat{\theta}(s,t) = \left(f(\hat{\beta}(s,t), R(\hat{\beta}(s,t))^{\top}\right)^{\top}$ that weakly converges to a tight non-degenerate Gaussian process indexed by (s,t).

Using the delta method for Hadamard directionally differentiable functionals and building on the preliminary result in Assumption 2, we derive the limiting distribution of $\hat{f}^l$ in Theorem 1 below. The limiting distribution is not pivotal, depending on the binding constraints whose Lagrangian multipliers are not zero. Our main theoretical result will follow from the next constraint qualification condition for Hadamard directional differentiability.

Assumption 3 *(i) There exists* $s_0 \in \mathcal{S}$ *such that* $R(\bar{\beta}(s_0)) \leq 0$*; (ii)* $R(\beta)$ *is linear in* β*.*

The optimization literature, e.g., Theorem 4.25 in Bonnans and Shapiro (2013), has provided general results for optimization problems with nonlinear nonconvex objective functions and constraints.[9] Assumption 3 gives the low-level constraint qualification conditions under our setup; see more technical detail in the Appendix. Assumption 3(i) implies that the constrained space defined by the shape restriction from $\mathcal{B}_{model}$ is not a strict subset of $\mathcal{B}_{data,\mu}$, meaning that the data provides informative bounds. Assumption 3(i) is potentially testable given the asymptotic theory of the support function estimator in Chandrasekhar, Chernozhukov, Molinari, and Schrimpf (2012). We choose the fixed direction s_0 in our estimands (f^l, f^u) to satisfy Assumption 3(i). So for each $s \in \mathcal{S}$, there exists $t \in [0,1]$ such that $R(\beta(s,t)) \leq 0$ and the optimal solution exists.

[7] In the presence of discrete covariates, we could apply the approach in Chandrasekhar, Chernozhukov, Molinari, and Schrimpf (2012) and introduce a conservative distortion to our inference method. It is known that when X contains discrete covariates, there are exposed faces on $\mathcal{B}_{data,\mu}$ and $\bar{\beta}(s)$ is not everywhere differentiable in $s \in \mathcal{S}$ (Bontemps, Magnac, and Maurin, 2012). Consequently, the estimator of $\bar{\beta}(s)$ does not weakly converge to a Gaussian process as in Assumption 2(ii). Chandrasekhar, Chernozhukov, Molinari, and Schrimpf (2012) propose a jittered estimation by adding a small noise to discrete covariates. Heuristically, they construct a super-set of $\mathcal{B}_{data,\mu}$ and obtain a uniform limiting distribution of the corresponding support function estimator.

[8] Alternatively, we might use a kernel estimator as in Lee (2018). It is known that $\sqrt{n}(\hat{\bar{\beta}}(s) - \bar{\beta}(s))$ can be shown to converge to a Gaussian process by controlling the bias of the nonparametric preliminary estimators.

[9] We thank one anonymous referee for the reference.

For each $s \in \mathcal{S}$, let the set of the optimal solutions be denoted by $\bar{\mathcal{T}}(s) \equiv \arg\min_{t\in\mathcal{T}} f(\beta(s,t))$ subject to $R(\beta(s,t)) \leq 0$ and $t \in [0,1]$. Define the Lagrangian $L(s,t,\lambda) \equiv f(\beta(s,t)) + \lambda_r^\top R(\beta(s,t)) - \lambda_0 t + \lambda_1(t-1)$ for the Lagrangian multiplier $\lambda = (\lambda_r^\top, \lambda_0, \lambda_1)^\top$. For each $s \in \mathcal{S}$ and $\bar{t} \in \bar{\mathcal{T}}(s)$, let the corresponding set of the Lagrangian multipliers be $\Lambda(s,\bar{t}) \equiv \{\lambda \in \mathcal{R}^{d_R+2} : L(s,\bar{t},\lambda) = \min_{t\in\mathcal{T}} L(s,t,\lambda),\ \lambda \geq 0,\ \lambda_r^\top R(\beta(s,\bar{t})) = 0,\ \lambda_0 \bar{t} = 0, \text{ and } \lambda_1(\bar{t}-1) = 0\}$. Let the set of the optimal solutions be denoted by $\bar{\mathcal{S}} \equiv \arg\min_{s\in\mathcal{S}} \min_{t\in\bar{\mathcal{T}}(s)} f(\beta(s,t))$.

The following definition is adapted from Fang and Santos (2019).

Definition 1 (Hadamard directional differentiability) *Let $\mathbb{D}$ and $\mathbb{E}$ be Banach spaces, and $\phi : \mathbb{D}_\phi \subseteq \mathbb{D} \to \mathbb{E}$. The map ϕ is said to be Hadamard directionally differentiable at $\theta \in \mathbb{D}_\phi$ tangentially to a set $\mathbb{D}_0 \subset \mathbb{D}$, if there is a continuous map $\phi'_\theta : \mathbb{D}_0 \to \mathbb{E}$ such that:*

$$\lim_{n\to\infty} \left\| \frac{\phi(\theta + \varepsilon_n h_n) - \phi(\theta)}{\varepsilon_n} - \phi'_\theta(h) \right\|_E = 0,$$

for all sequence $\{h_n\} \subset \mathbb{D}$ and $\{\varepsilon_n\} \subset \mathcal{R}_+$ such that $\varepsilon_n \downarrow 0$, $h_n \to h \in \mathbb{D}_0$ as $n \to \infty$ and $\theta + \varepsilon_n h_n \in \mathbb{D}_\phi$ for all n.

When ϕ'_θ is linear for all sequences $\{\varepsilon_n\} \subset \mathcal{R}$ such that $\varepsilon_n \to 0$, the map ϕ is Hadamard differentiable at $\theta \in \mathbb{D}_\phi$ tangentially to a set $\mathbb{D}_0 \subset \mathbb{D}$.

Theorem 1 (Asymptotic distribution) *Suppose Assumptions 2 and 3 hold. Then*

$$\begin{aligned}\sqrt{n}\left(\hat{f}^l - f^l\right) = \sqrt{n} \min_{s\in\bar{\mathcal{S}}} \min_{t\in\bar{\mathcal{T}}(s)} \max_{(\lambda_r^\top,\lambda_0,\lambda_1)^\top\in\Lambda(s,t)} &\Big(f(\hat{\beta}(s,t)) - f(\beta(s,t)) \\ &+ \lambda_r^\top \left(R\left(\hat{\beta}(s,t)\right) - R\left(\beta(s,t)\right)\right)\Big) + o_p(1) \\ \xrightarrow{L} \phi'_{\theta_0}\left(\mathbb{G}_0\right),\end{aligned}$$

where the Hadamard directional derivative in the direction $h = \left(h^{(1)}, h^{(2)\top}\right)^\top \in \mathbb{D}_\phi$ at θ_0 is $\phi'_{\theta_0}(h) = \min_{s\in\bar{\mathcal{S}}} \min_{t\in\bar{\mathcal{T}}(s)} \max_{\lambda\in\Lambda(s,t)} \left(h^{(1)}(s,t) + \lambda_r^\top h^{(2)}(s,t)\right)$.

When the optimal solution is unique ($\bar{\mathcal{S}}$ and $\bar{\mathcal{T}}(s)$ for any $s \in \bar{\mathcal{S}}$ are singletons), the Hadamard directional derivative of the mapping $\phi(\theta)$ in the direction h is linear in h, i.e., the mapping ϕ is Hadamard differentiable. So $\hat{f}^l = \phi(\hat{\theta})$ is a regular estimator and is asymptotically normal. When the optimal solution is not unique, the Hadamard directional derivative $\phi'_{\theta_0}(h)$ is not linear in h, i.e., the mapping is not differentiable. Consequently the estimator is not regular and the limiting distribution is not pivotal. The non-unique optimal solutions could occur when there is an inequality restriction imposed directly on the objective function, i.e., $f \in R$. Figure 3 illustrates one example when $\bar{\mathcal{S}}$ is not a singleton.

Inference by directly estimating or simulating the asymptotic distribution is difficult. This

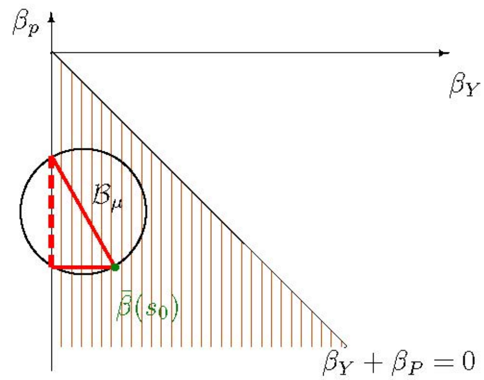

Figure 3: Consider the income coefficient $f(\beta) = (0,0,1)\beta = \beta_Y$ and the lower endpoint of its identified set $\arg\min_{\beta\in\mathcal{B}_\mu} \beta_Y = 0$. The dashed line indicates the set of the optimal solutions $\bar{\mathcal{S}} = \arg\min_{s\in\mathcal{S}}(0,0,1)\beta(s,\bar{t}(s))$.

is because we need to consistently estimate the Lagrangian multipliers, which depend on whether the constraints bind and might not be unique. But the Hadamard directional differentiability serves for the bootstrap procedure recently developed by Fang and Santos (2019) and Hong and Li (2018). Using the bootstrap in the next section, we will not need to estimate the optimal solution and Lagrangian multipliers.

4.2 Bootstrap

When the optimal solution is unique and hence ϕ is Hadamard differentiable, Fang and Santos (2019) show that the estimator is regular and the standard bootstrap is valid. When the optimal solutions are not unique and ϕ is only Hadamard directionally differentiable, the conventional bootstrap fails. They propose an alternative bootstrap using a consistently estimated Hadamard *directional* derivative. Specifically, suppose there is a valid bootstrap procedure for $\hat{\theta}$ such that the bootstrap sample $\sqrt{n}(\hat{\theta}^* - \hat{\theta})$ consistently estimates the limiting distribution of $\sqrt{n}(\hat{\theta} - \theta)$ conditional on the data. Fang and Santos (2019) show that the standard bootstrap using $\sqrt{n}\left(\phi(\hat{\theta}^*) - \phi(\hat{\theta})\right)$ is not valid when ϕ is not Hadamard differentiable. They propose an alternative bootstrap method: given a consistent estimator $\hat{\phi}'_{\theta_0}(\cdot)$ for the Hadamard directional derivative, $\hat{\phi}'_{\theta_0}\left(\sqrt{n}(\hat{\theta}^* - \hat{\theta})\right)$ consistently estimates the limiting distribution of $\sqrt{n}\left(\phi(\hat{\theta}) - \phi(\theta)\right)$.

However in our case, it is hard to estimate the Hadamard directional derivative consistently since it involves estimating the Lagrangian multipliers and which constraint binds. Hong and Li (2018) propose the direct analog estimator of numerical differentiation: $\hat{\phi}'_{\theta_0}(h) = (\phi(\hat{\theta} + h\varepsilon_n) - \phi(\hat{\theta}))/\varepsilon_n$, for a sequence $\varepsilon_n \downarrow 0$ and $\varepsilon_n\sqrt{n} \to \infty$. Then the Fang-Santos alternative bootstrap uses

$$\hat{\phi}'_{\theta_0}\left(\sqrt{n}(\hat{\theta}^* - \hat{\theta})\right) = \frac{\phi\left(\hat{\theta} + \varepsilon_n \sqrt{n}(\hat{\theta}^* - \hat{\theta})\right) - \phi(\hat{\theta})}{\varepsilon_n},$$

which is the rescaled bootstrap in Dümbgen (1993). We choose $\varepsilon_n = c(\log n)^{-1/2}$ for some constant c and assess the choice of c via simulation. Hong and Li (2018) show that their proposed estimator is consistent for $\phi'_{\theta_0}(h)$ when ϕ is Lipschitz continuous. Theorem 2 formalizes this statement in our context.

Theorem 2 *Assume $\theta^{(2)}$ is convex in t. Then ϕ as defined in (15) is Lipschitz continuous and its Hadamard directional derivative is consistently estimated by $\hat{\phi}'_{\theta_0}(h) = (\phi(\hat{\theta} + h\varepsilon_n) - \phi(\hat{\theta}))/\varepsilon_n$, for $h \in \mathcal{C}(\mathcal{S} \times [0,1])$, $\varepsilon_n \downarrow 0$ and $\varepsilon_n \sqrt{n} \to \infty$.*

Now we outline the numerical delta method bootstrap procedure following the estimation procedure in Section 3.

Step B1. Suppose a valid bootstrap procedure for $\hat{\beta}(s,t)$. Generate a bootstrap sample $\{X_i^*\}$ and follow Step 1 to 2 to calculate $\hat{\beta}^*(s,t)$. Repeat B times to obtain the bootstrap estimators $\{\hat{\beta}_b^*(s,t)\}_{b=1}^B$ such that $\sqrt{n}\big(\hat{\beta}^*(s,t) - \hat{\beta}(s,t)\big)$ estimate the limiting distribution of $\sqrt{n}\big(\hat{\beta}(s,t) - \beta(s,t)\big)$ consistently, conditional on the data.[10]

Step B2. Choose ε_n satisfying $\varepsilon_n \downarrow 0$ and $\varepsilon_n \sqrt{n} \to \infty$. Compute the bootstrap samples of

$$Z^{l*} = \frac{\phi\left(\hat{\theta} + \varepsilon_n \sqrt{n}(\hat{\theta}^* - \hat{\theta})\right) - \phi(\hat{\theta})}{\varepsilon_n},$$

where $\phi(\hat{\theta}) = \hat{f}^l$ and

$$\begin{aligned}\phi\Big(\hat{\theta} + \varepsilon_n \sqrt{n}(\hat{\theta}^* - \hat{\theta})\Big) \equiv \; & \min_{s \in \mathcal{S}, t \in [0,1]} f\left(\hat{\beta}(s,t)\right) + \varepsilon_n \sqrt{n}\left(f\left(\hat{\beta}^*(s,t)\right) - f\left(\hat{\beta}(s,t)\right)\right) \\ & \text{subject to} \\ & R\left(\hat{\beta}(s,t)\right) + \varepsilon_n \sqrt{n}\left(R\left(\hat{\beta}^*(s,t)\right) - R\left(\hat{\beta}(s,t)\right)\right) \le 0.\end{aligned}$$

The bootstrap samples for the upper endpoint Z^{u*} are computed analogously by changing min to max.

Step B3. A $(1-\alpha)$-level confidence interval

[10] Fang and Santos (2019) formally provide conditions for the bootstrap estimator in their Assumption 3. The nonparametric, Bayesian, block, score, and weighted bootstrap are included. Chandrasekhar, Chernozhukov, Molinari, and Schrimpf (2012) propose a Bayesian bootstrap.

- for $f(\beta)$ is $\left[\hat{f}^l - \hat{c}^l_\alpha/\sqrt{n}, \hat{f}^u + \hat{c}^u_\alpha/\sqrt{n}\right]$, where the critical values $\hat{c}^l_\alpha$ and $\hat{c}^u_\alpha$ are defined by $\Pr\left(Z^{l*} \leq \hat{c}^l_\alpha\right) = 1 - \alpha$ and $\Pr\left(Z^{u*} \geq -\hat{c}^u_\alpha\right) = 1 - \alpha$.
- for the identified set $\left[f^l, f^u\right]$ is $\left[\hat{f}^l - \hat{C}^l_\alpha/\sqrt{n},\ \hat{f}^u + \hat{C}^u_\alpha/\sqrt{n}\right]$, where the critical values $\hat{C}^l_\alpha$ and $\hat{C}^u_\alpha$ are defined by $\Pr\left(Z^{l*} \leq \hat{C}^l_\alpha, Z^{u*} \geq -\hat{C}^u_\alpha\right) = 1 - \alpha$.

An asymptotically valid pointwise $1-\alpha$ confidence interval for $f(\beta)$ is the intersection of one-sided confidence interval for $\hat{f}^u$ and $\hat{f}^l$. The confidence interval for the identified set $\left[f^l, f^u\right]$ might be of interest when we allow mis-specification and do not assume there is a unique true $f(\beta) = \bar{f}(g)$ in $\left[f^l, f^u\right]$ (Tamer, 2010). Notice that the limiting distribution of f^l and f^u are not independent of one another and the bootstrap procedure consistently estimates the joint asymptotic distribution. The construction of the confidence intervals is based on Imbens and Manski (2004), Stoye (2009), Chandrasekhar, Chernozhukov, Molinari, and Schrimpf (2012), and Freyberger and Horowitz (2015).

5 Endogeneity

In this section, we first provide some discussion of welfare analysis under endogeneity. We first consider the case where income is both endogenous *and* interval censored, and we have access to an instrumental variable Z that is correlated with income but independent of unobserved preference heterogeneity. The identified set for the structural choice probability can now be constructed as follows.

Let the outcome variable $Q = g(P, Y, \eta)$, where g is an unknown structural function of the observed variables (P, Y) and the unobserved heterogeneity η. For the binary choice model in Section 2.1, $Q = \mathbf{1}\{U_1(Y - P, \eta) - U_0(Y, \eta) \geq 0\}$. We now state a proposition that shows how to obtain a set that contains the true structural choice probability (SCP) $q(p, y) = \int g(p, y, \eta) dF_\eta(\eta)$ at a hypothetical price and income. To state this proposition, let Z denote the excluded exogenous variables and P be the included exogenous variables. We assume that the good under consideration is normal, and allow Y to be endogenous in the sense that $F_{\eta|Y} \neq F_\eta$. Assumption 4 formally states the conditions.

Assumption 4 (Instrumental variable) *(i) The function $g(p, y, \eta)$ is increasing in y for each p and η; (ii) The instrumental vector Z is excluded from the function g; (iii) η is independent of P and Z; (iv) There exist (y, z) such that $F_{Y|Z}(y|z) \neq F_Y(y)$.*

The conditions in the above assumption can be used to derive bounds for the SCP. For ease of exposition, we suppress P in the following discussion. Recall that Y is observed in intervals of the form $\mathcal{Y}_k \equiv [y^k, y^{k+1})$ for $k = 0, 1, ..., K-1$, and $\mathcal{Y}_K$ is the interval $[y^K, y^{K+1}]$.[11] By the monotonicity condition (i), the conditional choice probability given income interval, i.e., $\mathbb{E}[Q|Y \in \mathcal{Y}_k]$ can be consistently estimated from the data and satisfies

$$\begin{aligned} \mathbb{E}[Q|Y \in \mathcal{Y}_k] &\equiv \int_{\mathcal{Y}_k} \mathbb{E}[g(Y,\eta)|Y=y]dF_Y(y)/\Pr(Y \in \mathcal{Y}_k) \\ &\leq \int_{y^k}^{y^{k+1}} \mathbb{E}[g(y^{k+1},\eta)|Y=y]dF_Y(y)/\Pr(Y \in \mathcal{Y}_k) \\ &\leq \mathbb{E}[g(y,\eta)|Y \in \mathcal{Y}_k] \text{ for } y \geq y^{k+1}, \end{aligned} \tag{16}$$

which is the *conditional* SCP at $y \geq y^{k+1}$ for the subpopulation whose income lies in the interval $\mathcal{Y}_k$. When Y is endogenous, the conditional SCP would be different from the *unconditional* SCP $\mathbb{E}[g(y,\eta)]$. So the conditional choice probability given income interval $\mathbb{E}[Q|Y \in \mathcal{Y}_k]$ bounds only the conditional SCP given $\mathcal{Y}_k$ rather than the unconditional SCP, which is our object of interest. Figure 4 illustrates this idea. This is in contrast to the exogenous income case where we only need the monotonicity Assumption 1(ii) on the structural choice probability $q(p,y) = \mathbb{E}[g(p,y,\eta)] = \mathbb{E}[g(P,Y,\eta)|P=p,Y=y]$. Note that the last equality does not hold for the endogenous income. When the income variable is endogenous, we assume the good to be normal for everyone, i.e., the binary choice $g(p,y,\eta) = \mathbf{1}\{U_1(y-p,\eta) > U_0(y,\eta)\}$ is increasing in y for any η and p in Assumption 4(i).

Interestingly, however, a bound for the unconditional SCP can be obtained by averaging over all income intervals:

$$q(y^j) = \sum_{k=0}^{K} \mathbb{E}[g(y^j,\eta)|Y \in \mathcal{Y}_k]\Pr(Y \in \mathcal{Y}_k) \underset{\text{from (16)}}{\geq} \sum_{k=0}^{j-1} \mathbb{E}[Q|Y \in \mathcal{Y}_k]\Pr(Y \in \mathcal{Y}_k).$$

Furthermore, by the exogeneity condition Assumption 4(iii), the instrumental variable Z satisfies $q(y^j) = \mathbb{E}[g(y^j,\eta)] = \mathbb{E}[g(y^j,\eta)|Z]$. So the inequalities in (16) hold when Z is included in the conditioning covariates. Thus, by taking intersection over values of Z, we obtain a tighter identified set for q in the presence of both interval-censored and endogenous income. This is formally stated in the following proposition.

Proposition 2 *Suppose Assumption 4 holds. Then for $j = 1, .2, ..., K$ and $y \in \mathcal{Y}_j \equiv$*

[11] A control function approach is occasionally used to account for endogeneity in nonlinear models, e.g., Blundell and Powell (2003), Imbens and Newey (2009), Petrin and Train (2010). However, when Y is only observed in intervals, the control function is not point-identified, making this approach infeasible.

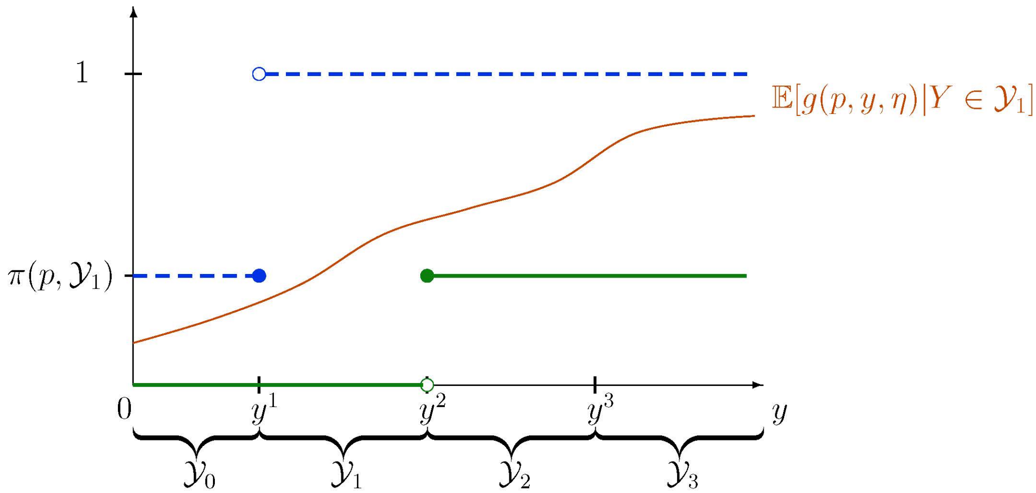

Figure 4: The curve is the true conditional structural choice probability (SCP) for the subpopulation whose income lies in the interval $\mathcal{Y}_1$, $\mathbb{E}[g(p,y,\eta)|Y \in \mathcal{Y}_1]$ at a price p. The horizontal axis is the income variable y, whose support is partitioned to four intervals by a set of grid points $\{y^1, y^2, y^3\}$. The blue dashed line is the upper bounding function and the green solid line is the lower bounding function. When income Y is endogenous, the conditional choice probability given income interval $\mathcal{Y}_1$, $\pi(p,\mathcal{Y}_1) \equiv \mathbb{E}[Q|P = p, Y \in \mathcal{Y}_1]$, provides bounds for only the conditional SCP given $\mathcal{Y}_1$ rather than the unconditional SCP.

$[y^j, y^{j+1})$, *the structural choice probability* $q(p,y)$ *satisfies*

$$\begin{aligned}
&\sup_z \mathbb{E}\big[Q\big|P=p, Z=z, Y \in \big[y^0, y^j\big)\,\big]\Pr\big(Y \in \big[y^0, y^j\big)\,\big|P=p, Z=z\big)\\
&\le\ q(p,y^j) \le\ q(p,y) \le\ q(p,y^{j+1})\\
&\le \inf_z \mathbb{E}\big[Q\big|P=p, Z=z, Y \in [y^{j+1}, y^{K+1})\big]\Pr\big(Y \in [y^{j+1}, y^{K+1})\big|P=p, Z=z\big)\\
&\quad + \Pr\big(Y \le y^{j+1}\big|P=p, Z=z\big).
\end{aligned}$$

One can easily extend this approach to include other covariates along with the included exogenous P, and to obtain bounds on welfare estimates from the identified set of choice probabilities.[12] If Z does not determine Y, i.e., the instrument relevance Assumption 4(iv) is violated, then $\mathbb{E}[Q|P = p, Z = z, Y \in \big[y^0, y^j\big)]\Pr(Y \in \big[y^0, y^j\big)\,|P = p, Z = z) = \mathbb{E}\left[Q|P = p, Y \in \big[y^0, y^j\big)\right]\Pr\left(Y \in \big[y^0, y^j\big)\,|P = p\right)$. Proposition 2 still provides the identified set, but the tightening obtained via intersections over Z is no longer viable.

[12]Substantively similar ideas are discussed in Manski (1994) and, at a somewhat abstract level, in the generalized instrumental variable model in Chesher and Rosen (2017).

Price Endogeneity When we have individual level data, endogeneity of price is typically of lesser concern, because an individual's choice or her omitted characteristics are unlikely to affect the market price she faces. If price is suspected of being endogenous (e.g., due to omitted product quality), and one has a potential IV[13], then a control-function methods used in the first step of our estimation. In particular, suppose there exists a control variable V such that P is independent of η conditional on V. For example, Blundell and Powell (2003) assume price $P = \mathbb{E}[P|W]+e$, where the IV W is independent of (η, e) and the disturbance e is continuously distributed with C.D.F. strictly increasing on the support of e. Then the residual $V = P - \mathbb{E}[P|W]$ is a valid control variable. By the conditional independence assumption and the normal good Assumption 4(i), the conditional choice probability given income interval, price, and the control variable provides a bound for the conditional SCP,

$$\begin{aligned}\mathbb{E}[Q|Y \in \mathcal{Y}_k, P = p, V = v] \leq \mathbb{E}[g(y^{k+1}, p, \eta)|Y \in \mathcal{Y}_k, P = p, V = v] &= \mathbb{E}[g(y^{k+1}, p, \eta)|Y \in \mathcal{Y}_k, V = v] \\ &\leq \mathbb{E}[g(y, p, \eta)|Y \in \mathcal{Y}_k, V = v] \quad \text{for } y \geq y^{k+1}.\end{aligned}$$

Next, an average taken over the marginal distribution of V in the income interval $\mathcal{Y}_k$ under the standard common support assumption yields

$$\begin{aligned}\pi_V(p, \mathcal{Y}_k) &\equiv \int \mathbb{E}[Q|Y \in \mathcal{Y}_k, P = p, V = v] dF_{V|Y \in \mathcal{Y}_k}(v) \\ &\leq \mathbb{E}[g(y, p, \eta)|Y \in \mathcal{Y}_k] \quad \text{for } y \geq y^{k+1}.\end{aligned}$$

We therefore obtain a lower bound for the conditional structural choice probability given $\mathcal{Y}_k$, which is the unconditional SCP $q(y, p)$ for the exogenous income case.

6 Mis-specification and best parametric approximation

Going back to the original problem of estimating the endpoints of the identified set in (2), notice that the parameter space $\mathcal{Q}$ of the optimization problem is, in general, nonparametric and infinite-dimensional. Our parametric model can be viewed as an approximation, designed to make this constrained stochastic optimization problem tractable. We now provide a brief interpretation of our parametric estimates and related inference with reference in terms of a *best parametric approximation*. In particular, for our binary choice model, the unobserved heterogeneity may not have a normal or logistic C.D.F., but the corresponding inference

[13] Typical examples used in the literature include cost-shifters on the supply side, size of the market and average price of the same alternative in other markets (Hausman, 1997; Petrin and Train, 2010).

theory presented in Section 4 also works under the weaker assumption that the parameter belonging to $\mathcal{B}_\mu$ in (10) simply provides a best parametric fit. In other words, the parametric model may be mis-specified in the sense that there is no $\beta \in \mathcal{R}^{d_x}$ such that $q(X) = \Phi(X^\top \beta)$ almost surely, so there might not be a β satisfying the conditional moment inequalities in (7). Since the random utility model with unrestricted heterogeneity in Section 2.1 does not specify a parametric structural choice probability, our estimand may be interpreted as arising from a parametric approximation. That is, the parameter β is a best parametric approximation to the choice probability q in the sense that it minimizes a quadratic loss function $\mathbb{E}_\mu\left[\left(\Phi^{-1}(q(X)) - X^\top\beta\right)^2\right]$, where μ is a continuous measure on the support of X, with respect to which "best approximation" is to be defined and interpreted. Parameters of this type have been studied, for example, in Horowitz and Manski (2006), where Φ is a logistic C.D.F. and $\Phi(X^\top\beta)$ is the best logit predictor under square loss of the conditional log-odds of a binary random variable. Other examples discussed in Chandrasekhar, Chernozhukov, Molinari, and Schrimpf (2012) include distribution and duration regressions with an interval-valued dependent variable. Therefore each parameter in $\mathcal{B}_{data,\mu}$ provides an approximation to a function q that is partially identified by a bounded convex set $\mathcal{Q}_{data}$ in (6). Consequently, the problem of calculating bounds for $\bar{f}(q)$ is approximated by bounds for $f(\beta)$ in (11).

One may compare this to mis-specification in *point-identified* models. Indeed, in the point-identification literature, a pseudo-true parameter is often defined as some approximation to the truth, e.g., Chamberlain (1994). For example, the OLS coefficient estimates $\arg\min_b \mathbb{E}[(g(X) - X^\top b)^2]$, where the conditional mean function $g(x) = \mathbb{E}[Y|X = x]$ is point-identified. We can define an alternative pseudo-true parameter as the best linear approximation to $g(x)$ by $\beta_\mu \equiv \arg\min_b \mathbb{E}_\mu[(g(X) - X^\top b)^2]$, where the loss function is $\int_{\mathcal{X}} (g(x) - x^\top b)^2 f_\mu(x)dx = \mathbb{E}[(g(X) - X^\top b)^2 f_\mu(X)/f(X)]$ by a different measure μ of X with density function $f_\mu(\cdot)$ rather than the true density function $f(\cdot)$. If the conditional mean function is correctly specified by a linear model $g(x) = x^\top\beta^*$, then $\beta^* = \beta_\mu$ for any measure μ. Analogously in the *partially identified* setting, when the parametric model is correctly specified, the identified set $\mathcal{B}_{data} \equiv \{\beta : (7) \text{ holds}\}$ is a subset of $\mathcal{B}_{data,\mu}$ defined in (10), for any measure μ, i.e., we estimate $\min/\max_{\beta\in\mathcal{B}_\mu} f(\beta)$ in (11) that is a superset of the identified set $\min/\max_{\beta\in\mathcal{B}} f(\beta)$ in (8). Indeed, as pointed out by a referee, a practically appealing feature of our best parametric approximation approach is that $\mathcal{B}_{data,\mu}$ is in general non-empty, non-singleton, and interpretable, even if the model for the structural choice probability is mis-specified.[14]

Our inference procedure applies to specifications that can include higher order terms. Theoretically speaking, including higher order terms would result in an infinite number of

[14]As such, the caveat that under mis-specification, the estimated identified set assuming a parametric model could be too small (c.f. Ponomareva and Tamer (2011) and Kline and Santos (2013)) also applies to our analysis.

inequality restrictions. For example, for a quadratic approximation $\Phi^{-1}(q(p,y)) \approx \beta_0+\beta_{P1}p+\beta_{P2}p^2+\beta_Y y$, the Slutsky-type restriction $\partial q/\partial p+\partial q/\partial y \leq 0$ implies $\beta_{P1}+2\beta_{P2}p+\beta_Y \leq 0$ for all p. Operationally, one can evaluate the restrictions at a finite number of grid points over the support of the regressors, and apply our our inference method, which would lead to conservative but valid inference.

When the parametric model is mis-specified, the parameter β_μ that provides a best parametric approximation to a $q \in \mathcal{Q}$ might not satisfy the shape restrictions $\beta_Y \geq 0$ and $\beta_Y+\beta_P \leq 0$, i.e., $\beta_\mu \notin \mathcal{B}_{model}$. On the other hand, imposing and not imposing the shape restrictions on β_μ provide different approximations to $\mathcal{Q}$. It seems debatable as to whether one should impose shape restrictions for the best parametric approximation, i.e., whether one should use the sets $\mathcal{B}_\mu = \mathcal{B}_{data,\mu} \cap \mathcal{B}_{model}$ and $\mathcal{B}_{data,\mu}$. So we also provide inference theory for the case when we do not impose the shape restrictions in Corollary 1. Now the endpoints of the identified set are the solutions to $\min/\max_{(s,t)\in\mathcal{S}\times[0,1]} f(\beta(s,t))$. We present a formal statement of this result, which is an application of Lemma S.4.7 in the Supplemental Appendix of Fang and Santos (2019).

Corollary 1 (Asymptotic distribution without shape restriction) *Suppose Assumption 2 holds. Let the set of the optimal solutions be denoted by* $\mathcal{A} \equiv \arg\min_{s\in\mathcal{S},t\in[0.1]} f(\beta(s,t))$. *Then* $\sqrt{n}\left(\hat{f}^l - f^l\right) = \sqrt{n}\min_{(s,t)\in\mathcal{A}} f(\hat{\beta}(s,t)) - f(\beta(s,t)) + o_p(1) \xrightarrow{L} \phi'_{\theta_0}(\mathbb{G}_0)$, *where the Hadamard directional derivative* $\phi'_{\theta_0}(h) = \min_{(s,t)\in\mathcal{A}} h(s,t)$ *for* $h \in \mathcal{C}(\mathcal{S}\times[0,1])$.

7 Application to welfare analysis of tuition subsidy

In this section, we examine the empirical efficacy of our methods through a simulation exercise using data from a large-scale household survey, conducted by the Indian National Sample Survey Organization in 2004-5. The context is to estimate average welfare effects resulting from a hypothetical price subsidy toward school attendance of teenagers. In the dataset, individual wealth, proxied by monthly per capita household expenditure (as is standard in household surveys from developing countries) is actually measured, and thus we can point-estimate the average welfare effects of price changes. In order to see how our estimates work for the interval-censored income, we artificially generate the problem by dividing the real income variables into intervals. We can then compare our set-estimates using interval-valued income with point estimates using the actual income variable and midpoint imputations. This strategy also enables us to check how our estimates perform under greater or lesser degrees of censoring. The empirical results from this exercise suggest that in real settings (i) shape restrictions on choice probabilities, implied by the economic theory, can substantially tighten the identified set of welfare effects, and (ii) point estimates using midpoint imputations can be imprecise and potentially misleading in a substantive sense.

Data The data are drawn from a random sample of nearly 25000 households with exactly one teenage child. Each household makes the binary decision of whether to send the child to school or not, which we denote by Q; the income (i.e., the monthly per capita expenditure) is denoted by Y. The price P faced by a household in relation to school attendance is observed for those attending school. For those not attending, the potential price is taken to be the median tuition-related expense (divided by 12) per school-going child across all households in the *same income stratum* in the village or urban block where the household resides.[15] This enables us to get around the problem that in any village/block, the relatively wealthy would typically choose a high-fee private school while the poor would choose a cheaper state-school. Furthermore, when deciding whether school-attendance is affordable, a family is likely to ask around their neighbors to get an estimate of the potential costs. Therefore using the median tuition seems to us to be a reasonable proxy for the "potential" fees as envisaged by households who are not sending their children to school.

The following analysis uses a subsample of households that have a single female child of secondary-school age (15-18 years), belong to the Hindu religious group, do not belong to the historically disadvantaged Hindu castes, have household size between 4 and 6, and have adult literacy rate larger than 55%. We restrict attention to 647 observations with income smaller than Rs 6250 (74 percentile) and price between Rs 67 (8.5 percentile) and Rs 712 (78.8 percentile).[16] In what follows, all money amounts are expressed as Indian rupees per month (Rs), with 1 Indian rupee = 0.02 US dollars in 2004. Table 1 presents some descriptive statistics.

We consider a hypothetical price subsidy of 500 causing the price of schooling to decrease from $p_1 = 600$ to $p_0 = 100$ (73.8 to 15.6 percentile). We focus on the equivalent variation (EV) that measures how much income needs to be given to households so that their maximized utility at the high price Rs 600 is the same as at the low price Rs 100. Since we are dealing with a price reduction, our previous formulae for compensating variation of a price rise become the formulae for the Equivalent Variation; in particular, for a price decrease from p_1 to p_0, the average EV is $\bar{f}(q) = \int_0^{p_1 - p_0} q(p_0 + a, y_0 + a)da$.

Since we observe the actual income data in this sample, we can calculate average EV exactly, and this will serve as our benchmark result. We will contrast this with results

[15] The NSS stratifies each village/block by income, and samples independently from each stratum. This design makes sure that for each household observed in the sample, there is also included in the sample a set of households from the same village/block with similar income levels.

[16] The price range $[67, 712]$ constitutes the common support of price for each artifically created income interval (see below). Specifically, the conditional choice probabilities $\pi(p, \mathcal{Y}_{k-1})$ and $\pi(p, \mathcal{Y}_{k+1})$ are valid bounds for the structural choice probability $q(p, y) = \mathbb{E}[Q|P = p, Y = y]$ for $y \in \mathcal{Y}_k$ and p lying in the common support of P given $\mathcal{Y}_j$ for $j = k-1, k, k+1$.

obtained by artificially generating interval-censored income, using six intervals constructed by five equidistant intervals between Rs 2500 (10 percentile) and Rs 6250 (74 percentile): $\{[0, 2500), [2500, 3250), [3250, 4000), ..., [5500, 6250)\}$, and then applying the methods developed in this paper to produce bounds on the average EV. These results will be contrasted with those obtained using coarser partitioning of income, and those obtained via midpoint imputations.

Calculations and Results To calculate the average EV for a range of incomes y_0, we follow the methods described in Section 2.3 and 5.1. The details are as follows. 0.6cm0cm

Step 1. The conditional choice probability given price p and for each income interval $\mathcal{Y}_k$ is estimated by a Nadaraya-Watson kernel estimator $\hat{\pi}(p, \mathcal{Y}_k)$. We use a second-order Gaussian kernel and the Silverman's Rule-of-Thumb bandwidth $n^{-0.2} \times std.dev(P) \times C$ calculated using observations in the income interval $\mathcal{Y}_k$; the constant C is varied over a range from 0.1 to 10 for robustness check. We then obtain the bounding functions for the transformation of the inverse link function of the structural choice probability $\Phi^{-1}(q(p, y))$:

$$\hat{l}(x) \equiv \Phi^{-1}(\hat{\pi}(p, \mathcal{Y}_{k-1})) \leq \Phi^{-1}(q(p, y)) \leq \Phi^{-1}(\hat{\pi}(p, \mathcal{Y}_{k+1})) \equiv \hat{u}(x)$$

for $y \in \mathcal{Y}_k$, $k = 1, ..., K-1$, and $x \equiv (1, p, y)^\top$. For our design of income intervals, $\mathcal{Y}_0 = [0, 2500)$ and $\mathcal{Y}_K = [6250, \infty)$ with $K = 6$.

Step 2. We define $x^\top \beta = \beta_0 + \beta_P p + \beta_Y y$ to be the best linear approximation to $\Phi^{-1}(q(p, y))$, so the boundary point of the identified set of β is estimated by an OLS-type estimator, where the dependent variables are the bounding functions $\hat{l}(x), \hat{u}(x)$ from Step 1. That is, for a direction s on the unit sphere $\mathcal{S} = \{s \in \mathcal{R}^3 : \|s\| = 1\}$, the boundary point of the identified set of β is estimated by

$$\hat{\bar{\beta}}(s) = \Sigma^{-1}\mathbb{E}_\mu[X\ \hat{\gamma}_s(X)], \text{where } \Sigma = \mathbb{E}_\mu\left[XX^\top\right] \text{ and}$$
$$\hat{\gamma}_s(X) = \hat{u}(X)\mathbf{1}\{s^\top\Sigma^{-1}X \geq 0\} + \hat{l}(X)\mathbf{1}\{s^\top\Sigma^{-1}X < 0\},$$

The measure μ (denoted μ_1 for later reference) uses the empirical distribution of the price and interval-valued income and assigns uniform density within each income interval. So for a known function $g(X)$, $\mathbb{E}_\mu[g(X)] = n^{-1}\sum_{i=1}^n g\left((1, P_i, \tilde{Y}_i)\right)$, where $\tilde{Y}_i$ is drawn from $Uniform[y^k, y^{k+1}]$ if the interval-censored income of the observation i lies in $\mathcal{Y}_k = [y^k, y^{k+1})$.

Step 3. Fixing one direction $s_0 \in \mathcal{S}$, define a convex combination $\hat{\beta}(s,t) \equiv t\hat{\bar{\beta}}(s) + (1 - t)\hat{\bar{\beta}}(s_0)$ for $t \in [0, 1]$ and $s \in \mathcal{S}$ to characterize the identified set of β. Let $\hat{\theta} =$

$\left(f(\hat{\beta}(s,t)), R(\hat{\beta}(s,t)))\right)^{\top}$, where $f(\beta)$ is the average EV given by $\int_0^{500} \Phi(\beta_0 + \beta_P(100 + a) + \beta_Y(y_0 + a))da$, and $R(\beta)$ incorporates the normal-good and Slutsky-type shape restrictions, viz. $R(\beta) \equiv \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix}\beta = (-\beta_Y, \beta_Y + \beta_P)^{\top} \leq 0$. The lower endpoint of the identified set for $f(\beta)$ is estimated by the mapping

$$\hat{f}^l = \phi(\hat{\theta}) \equiv \min_{s\in\mathcal{S}, t\in[0,1]} f(\hat{\beta}(s,t)) \text{ subject to } R(\hat{\beta}(s,t)) \leq 0.$$

The minimum is searched over a grid of values s on $\mathcal{S}$ by a polar coordinate and a grid of values t on $[0,1]$. The upper endpoint $\hat{f}^u$ is estimated in the same way by changing min to max.

Step 4. For constructing the confidence region, we obtain the critical values by the numerical delta method bootstrap procedure outlined in Section 5.1. We generate five hundred bootstrap samples $\{X_i^*\}$ and follow the above Steps to obtain a bootstrap sample for $\hat{\theta}$, denoted by $\hat{\theta}^*$. As increment for the numeric derivative, we use $\varepsilon_n = 0.1/\sqrt{\log n} = 0.038$. We then compute the bootstrap samples of

$$Z^{l*} = \frac{\phi\left(\hat{\theta} + \varepsilon_n\sqrt{n}(\hat{\theta}^* - \hat{\theta})\right) - \phi(\hat{\theta})}{\varepsilon_n}.$$

The bootstrap samples for the upper endpoint Z^{u*} are computed analogously. A $(1-\alpha)$-level confidence interval for $f(\beta)$ is $\left[\hat{f}^l - \hat{c}_\alpha^l/\sqrt{n}, \hat{f}^u + \hat{c}_\alpha^u/\sqrt{n}\right]$, where the critical values $\hat{c}_\alpha^l$ and $\hat{c}_\alpha^u$ are defined by $\Pr\left(Z^{l*} \leq \hat{c}_\alpha^l\right) = 1-\alpha$ and $\Pr\left(Z^{u*} \geq -\hat{c}_\alpha^u\right) = 1-\alpha$.

Results In Figure 5, the solid lines present the estimated identified set for the average EV over a range of income y_0. The dashed lines are the 90% pointwise confidence region for the average EV. For ease of interpreting the results, Table 2 reports the estimated identified set and 90% confidence region for the average EV at the 25th percentile and the median incomes using the constant $C = 5$ for the bandwidth. In particular, at the median income (Rs 3732) of our subsample, a tuition subsidy of Rs 500 is equivalent to increasing income on average by an amount ranging between Rs 413.18 to Rs 453.00, i.e., 11.07% to 12.14% of the median income, at a 90% confidence level.[17] In general, the average EV rises with income, reflecting the fact that school-attendance is more prevalent at higher incomes, so that the subsidy is more fully utilized.

[17] We find that the least-squares cross-validation suggests a very large bandwidth for some income interval, which implies the conditional probability might not vary with price. Moreover in Step 1, the nonparametric estimation of the conditional choice probability $\pi(p, \mathcal{Y}_k)$ requires undersmoothing (Chandrasekhar, Chernozhukov, Molinari, and Schrimpf, 2012). So we choose the bandwidth by $n^{-0.2} \times std.dev(P) \times C$ and vary the constant C for robustness check. For C over a range from 3 to 10, the lower bound estimate varies between 419.37 and 423.95 and the upper bound estimate varies between 439.58 and 445.95.

Next, we consider a coarser partition of two intervals $\{[0, 3750), [3750, 6250)\}$, where 3750 is the 38th percentile of actual income. In Figure 6, we compare our set estimators obtained by the probit maximum likelihood (ML) estimators using the *actual incomes* and the *midpoint imputations*. Our set estimate for the two-interval case shows that at the median income (Rs 3732) of our subsample, a tuition subsidy of Rs 500 is equivalent to increasing income on average by an amount ranging between Rs 384.41 to Rs 484.62, i.e., 10.3% to 12.99% of the median income, at a 90% confidence level. For the two-interval case, we see that the Midpoint estimates are different from the Actual-income estimates. In particular, the confidence region by the Actual-income estimates does not cover the estimated identified set and confidence region by the Midpoint estimates. So depending on the pattern of interval data measurement, inference using the midpoint imputations could be misleading. To check the robustness of our substantive results to the inclusion of higher-order terms, Figure 7 presents probit ML estimates using the *actual incomes* under different specifications of choice probabilities. It is evident that the average EV is substantively unaffected by the inclusion of higher order terms of price and income in the regression.

To see how the shape restrictions $\beta_Y \geq 0$ and $\beta_P + \beta_Y \leq 0$ tighten the identified set, we compare our estimates with the those obtained without the restriction. In the top panel of Figure 8 for the two-interval case, the estimated identified set with restrictions are smaller than those without restrictions, so the restrictions bind at low incomes. The upper endpoint estimates with and without restrictions give the same values at high incomes, i.e., the restrictions do not bind in estimating the upper endpoints. For the two-interval case, our set estimate under these shape restrictions shows that at the median income (Rs 3732) of the subsample, a tuition subsidy of Rs 500 is equivalent to increasing income on average by an amount ranging between Rs 317.98 to Rs 483.24, at a 90% confidence level. In contrast, our set estimate without restrictions shows that at the median income, a tuition subsidy of Rs 500 is equivalent to increasing income on average by an amount ranging between Rs 328.19 to Rs 499.59, at a 90% confidence level, that is, the shape restrictions lead to a shrinking of the identified set by about $100 \times 6.14/500 = 1.23\%$ of the subsidy amount. When the true parameter satisfies the constraints with strict inequalities, the constrained estimator would entail additional noise resulting from the implicit testing involved in its construction. Therefore the restrictions do not necessary tighten the confidence regions.

To see how the measure μ affects the corresponding identified set, we perform our calculations using a *different* measure μ_2, viz. one that assigns equal weight over a range of income and price. Specifically, in Step 2 of the estimation procedure, we use a simulated sample of size 10000 from a uniform distribution of income in $[2500, 9000]$ and price in $[83.33, 625.67]$ after trimming 10% of the observations at the tails. Figure 9 shows the estimates corresponding to these two different measures μ_1 and μ_2 which provide two different approximations.

The estimated identified sets and 90% confidence regions using these two measures are both seen to contain the actual-income estimates. It is also evident from the Figure that our best-probit-approximation methodology is not sensitive to these two measures in this empirical application. If one believes the parametric model is correctly specified, one could take the intersection of the estimated identified set and confidence regions using various measures.

Finally we consider potential endogeneity of income, i.e., income might depend on unobserved heterogeneity. We use *land_owned* as an instrumental variable for income, and obtain the identified set of the structural choice probability characterized in Proposition 2. The idea is that a household's taste for education would likely be correlated with income but not on its breakdown into sources; that is, the only way land holding could be correlated with taste for schooling is through its impact on total income. One potential concern is that households owning more land might benefit more from sending their children to work in agriculture as opposed to schooling. But regressing working on age and land-holding, controlling for income, yields a t-stat of 0.08, suggesting that this concern is not empirically relevant in our data. On the other hand, the F-statistic from regressing income on land-holding (i.e., the first stage) is 1129.4, suggesting a strong instrument.

The results from using the IV technique (described in Proposition 2) are reported in Figure 10. The red dashed lines present the estimated identified set for the average EV using the instrumental variable. Comparing with the previous estimation under the exogenous income assumption, the solid lines in Figure 10 represent the estimated identified set, which had previously appeared in Figure 5. We see that imposing a stronger exogenous income assumption results in a tighter identified set. In particular, the estimated identified set under the exogenous income assumption suggests that at the median income, a tuition subsidy of Rs 600 is equivalent to increasing income on average by an amount ranging between Rs 423.73 to Rs 445.68, i.e., 11.4% to 11.9% of the median income. On the other hand, upon dispensing with the exogenous income assumption, our set estimate using the IV produces a different and wider identified set for average EV, suggesting that a tuition subsidy of Rs 600 is equivalent to increasing income on average by an amount ranging between Rs 266.92 to Rs 401.77, i.e., 7.15% to 10.77% of the median income.

8 Conclusion

In this paper, we have investigated the problem of empirical welfare analysis in a binary choice setting when income values are interval-censored. In this case, money-metric welfare effects of price changes, such as the average equivalent and compensating variation, cannot be point-identified. We show how to obtain bounds on these quantities under the assumption of a normal good, and subject to choice probabilities satisfying a set of theory-consistent

Slutsky-like restrictions. These restrictions are *linear* in choice probabilities, as opposed to the standard nonlinear Slutsky conditions for continuous choice.

Starting with a parametric approach, we develop the theory of inference for the endpoints of the resulting identified set under shape restrictions which lead to non-differentiability of the estimand, and make the inference theory nonstandard. Nonetheless, we can show that our estimator is directionally differentiable, so that confidence intervals can be constructed by applying recently developed bootstrap methods for such cases. The inference theory is developed without assuming that the probit/logit model is a correct specification for the binary choice model, so that the corresponding estimates have a "best parametric approximation" type interpretation. Finally, in a simulation exercise using real survey data from India, our methods yield promising results, in that they provide ranges of estimates for average welfare that are consistent with the true point-identified estimates.

The methods proposed here have wider applicability beyond interval data in binary choice models. In particular, our analysis applies to a general function $q(x)$ that (i) is partially identified by a convex compact set of the form $[L(x), U(x)]$, and (ii) is subject to shape restrictions. More generally in applications, one may use our methods to conduct inference on the identified set for other functionals subject to the shape restrictions, e.g., the average price derivative $\bar{f}(q) = \int (\partial q(p,y)/\partial p)\, dF\,(p,y)$.

Table 1: Descriptive statistics

	min	max	mean	std.dev
Binary outcome (Q)	0	1	0.85	0.36
Price (P)	69.17	708.33	278.19	156.47
Actual-income (Y)	$1,247$	$6,238$	$3,835.71$	$1,127.48$
Land_owned (Z)	0	$12,468$	694.27	1483.90

Notes: Summary statistics for 647 households that have a single female child aged 15-18, belong to the Hindu religious group, do not belong to the historically disadvantaged Hindu castes, have household size between 4 and 6, have adult literacy rate larger than 55%, have income smaller than Rs 6250, and have price between Rs 67 and Rs 712.

Table 2: Average EV in Rupees (6 intervals)

Income (y_0)	Rs 2,997 (25th percentile)	Rs 3,732 (median)
Estimated Identified Set	[390.73, 435.38]	[423.73, 445.68]
(90% Confidence Region)	(379.16, 447.26)	(413.18, 453.00))
Probit Point Estimate (Actual-income)	414.32	433.93
(90% Confidence Region)	(397.85, 430.30)	(420.66, 446.55)

Notes: Estimates and 90% confidence regions for the average EV resulting from a tuition subsidy of Rs 500 causing a fall in price from Rs 600 to Rs 100. The average EV is computed for individuals at the 25th percentile and the median income, respectively.

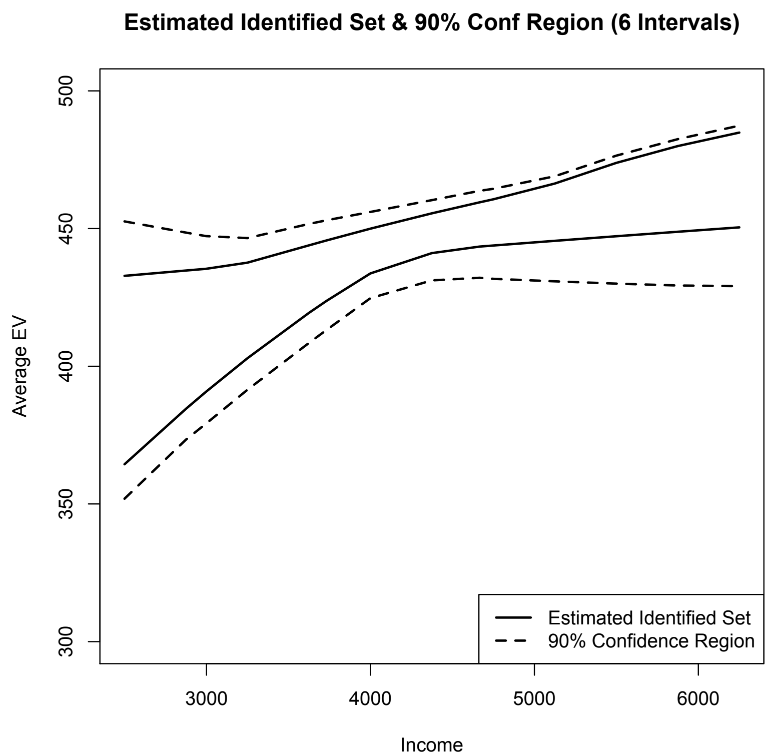

Figure 5: The average EV resulting from a tuition subsidy causing a fall in price from Rs 600 to Rs 100, for a range of income, when the income is observed in 8 intervals. The solid lines are the estimated identified sets and the dashed lines are the pointwise 90% confidence regions for the average EV. The average EV measures how much income needs to be given to the household so that the maximized utility at the higher income and the original high price of Rs 600 is the same as that at the original income and new lower price of Rs 100. Note that the average EV of a price decrease equals the average CV of a reverse price increase.

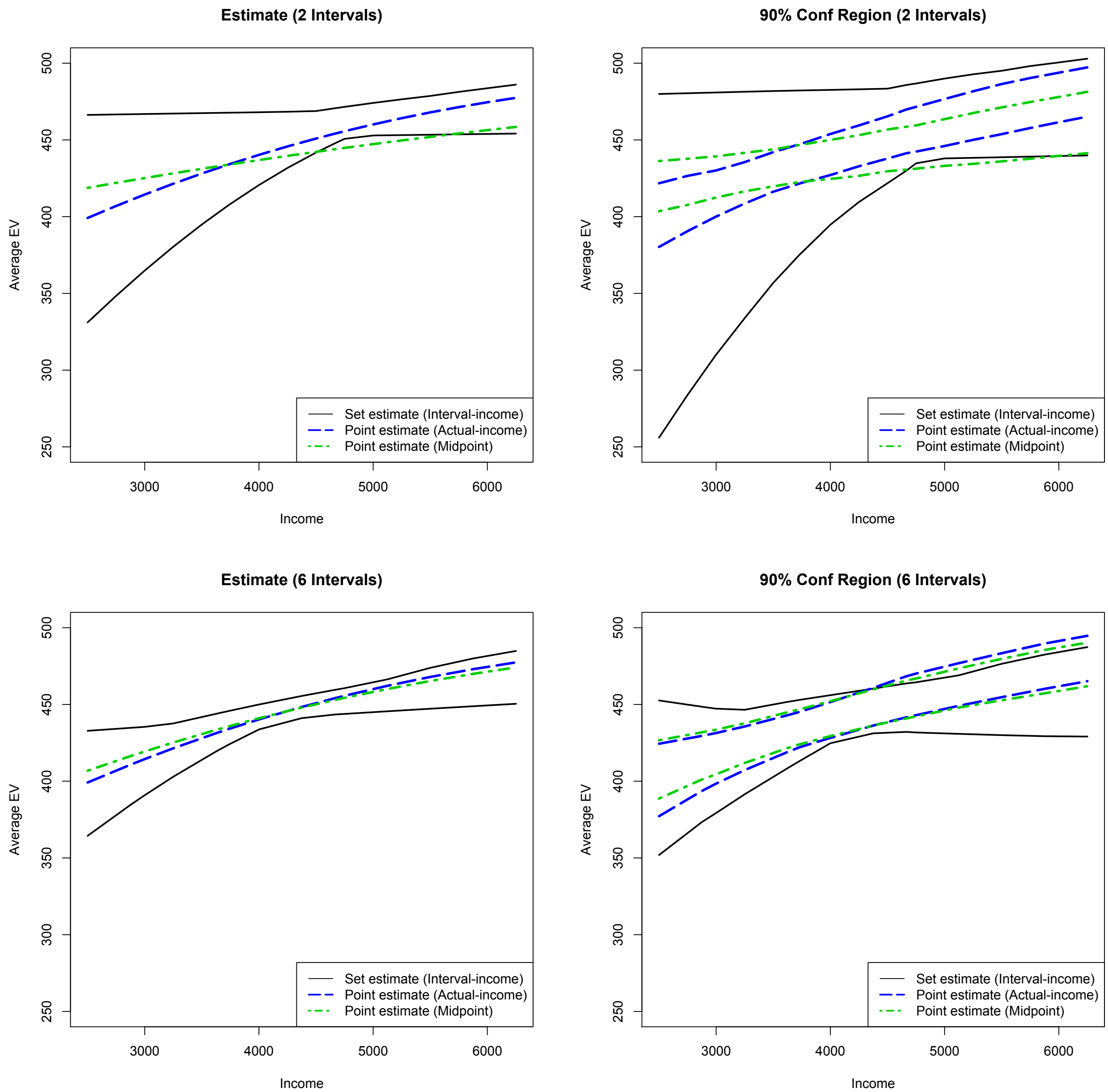

Figure 6: The average EV resulting from a Rs 500 price-subsidy, calculated over a range of income. The left panels present the estimates of our set estimator using *Interval-income* under normal-good and Slutsky-type shape restrictions, the probit ML point estimators using *Actual-income*, and using *Midpoint*. The right panels present the pointwise 90% confidence regions for the average EV. The top panels are for the coaser partition of two intervals and the bottom panels are for the finer partition of six intervals.

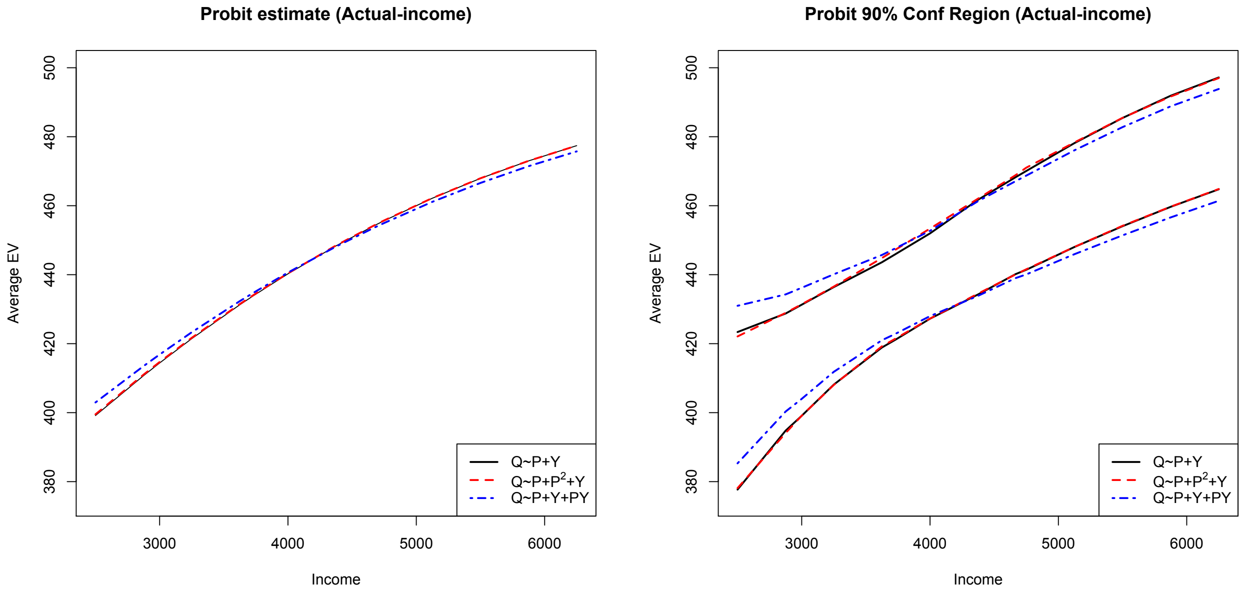

Figure 7: The average EV resulting from a Rs 500 price-subsidy, calculated over a range of income. The left panel presents the estimates of probit ML estimator using *Actual-income* for different specifications: linear, quadratic, and with an interaction term. The right panel presents the pointwise 90% confidence regions for the average EV.

9 Appendix

The Appendix is organized as follows. Section 9.1 is the proof of Proposition 1 in Section 2. Section 9.2 presents the proofs of the asymptotic theorems in Section 4. We prove the Hadamard directional differentiability of a mapping that minimizes a nonconvex objective function subject to convex inequality constraints in Lemma 1. Lemma 1 serves as a preliminary result and can be of independent interest.

9.1 Proof of Proposition 1

The proof contains two parts. The first part derives the revealed preference inequality (3). The second part shows the first result is equivalent to the monotonicity conditions of the C.D.F.s of the CV/EV.

1. Consider a type-η individual under the price p_0 and income y_0. If $U_0(y_0, \eta) > U_1(y_0 - p_0, \eta)$, she chooses option 0, i.e., $Q(p_0, y_0, \eta) = 0$. Consider a price increase to p_1 and an income compensation smaller than the price change, i.e., under the high price p_1, she has a higher income $y_1 \leq y_0 + p_1 - p_0$ and $y_1 \geq y_0$.

   The assumption that the utility functions $U_0(W, \eta)$ and $U_1(W, \eta)$ are strictly increasing in W implies $U_0(y_1, \eta) > U_0(y_0, \eta)$ and $U_1(y_0 - p_0, \eta) > U_1(y_1 - p_1, \eta)$. So the event $\{Q(p_0, y_0, \eta) = 0\} = \{U_0(y_0, \eta) > U_1(y_0 - p_0, \eta)\} \subseteq \{U_0(y_1, \eta) > U_1(y_1 - p_1, \eta)\} = \{Q(p_1, y_1, \eta) = 0\}$. That is to say if she does not choose option 1 at the low price, she will not choose option 1 at the high price under an income compensation smaller than the price increase.

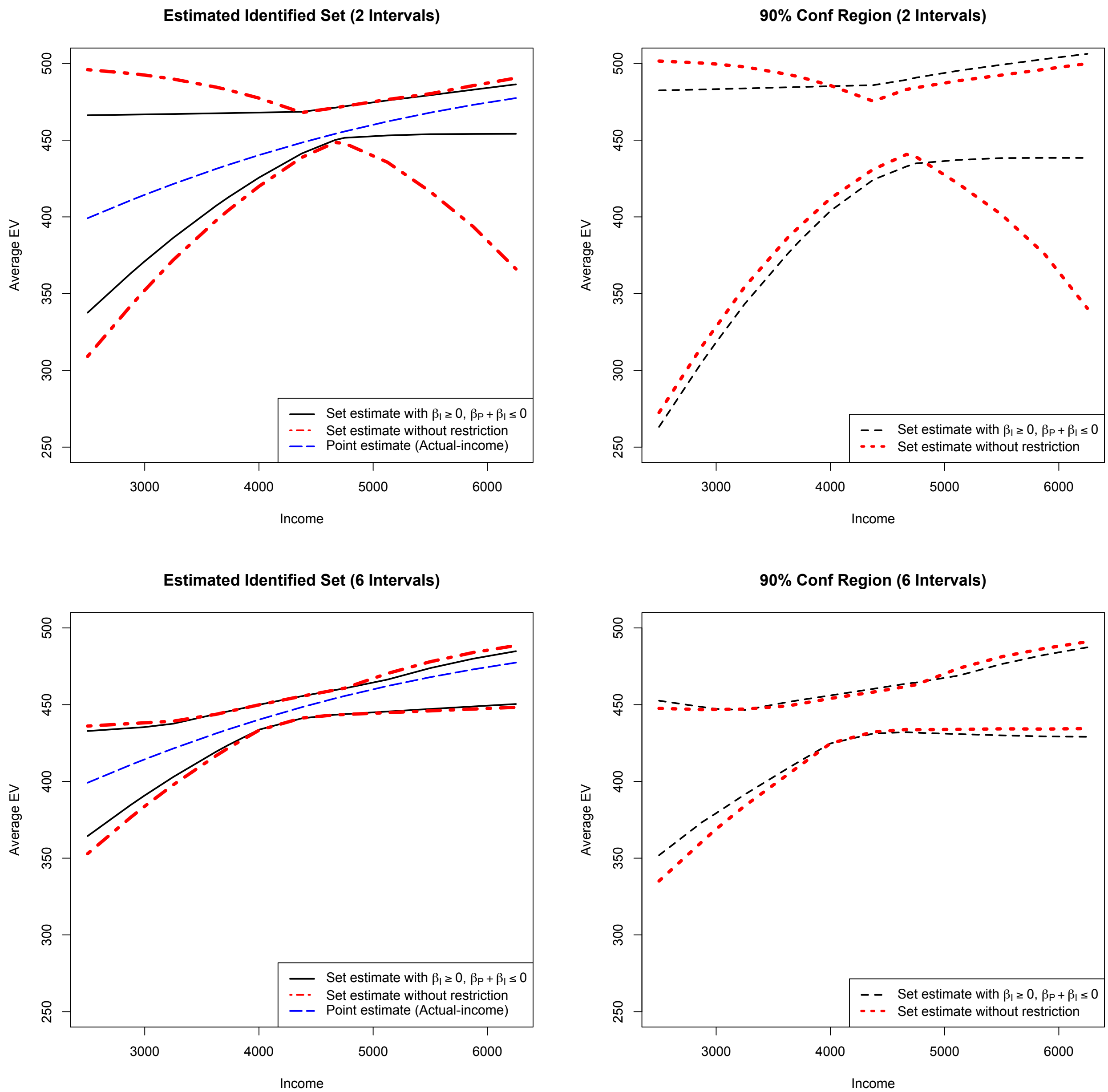

Figure 8: The average EV resulting from a Rs 500 price-subsidy, calculated under normal-good and Slutsky-type shape restrictions, and then without these restrictions. The left panel shows the estimates of the identified sets. The right panel shows the pointwise 90% confidence regions for the average EV. The top panels are for the coaser partition of two intervals. The bottom panels are for the finer partition of six intervals.

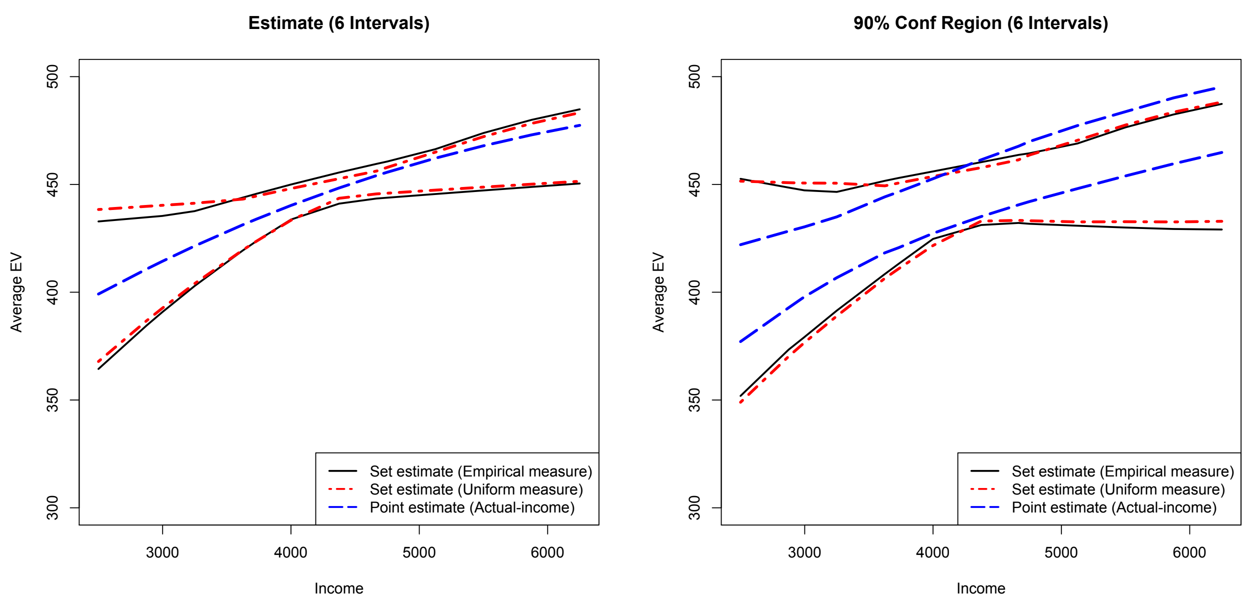

Figure 9: The average EV resulting from a Rs 500 price-subsidy, calculated over a range of income by our set esimator using two measures μ: Empirical measure uses the empirical distribution of the price and interval-valued income and assigns uniform density within each income interval; Uniform measure uses a simulated sample of size 10000 from a uniform distribution of price and income. The left panel shows the estimates of the identified sets. The right panel shows the pointwise 90% confidence regions for the average EV.

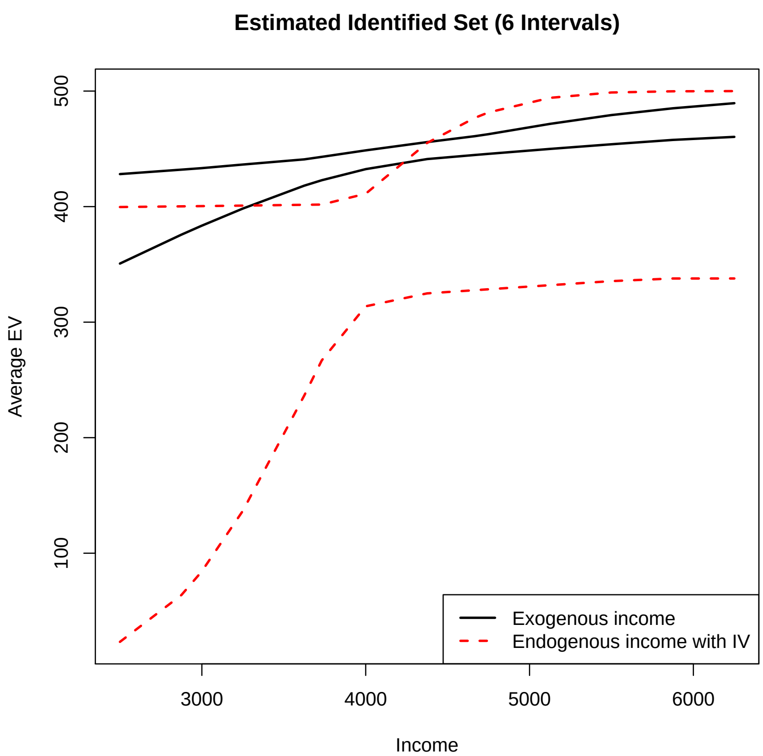

Figure 10: The average EV resulting from a Rs 500 price-subsidy for a range of income, by our set esimator under exogenous income assumption and then allowing for endogenous income. The instrumental variable is *land_owned*.

On the other hand, if $Q(p_0, y_0, \eta) = 1$, the same argument implies $Q(p_1, y_1, \eta)$ can be 0 or 1.

In sum, for any $y_1 \in [0, y_0 + p_1 - p_0]$, we obtain $Q(p_1, y_1, \eta) \leq Q(p_0, y_0, \eta)$. Calculating the expected demand by integrating with respect to the marginal distribution of η, we obtain $q(p_1, y_1) \leq q(p_0, y_0)$. Letting $y_1 = y_0 + c$ and $p_1 = p_0 + b$, we obtain the inequality (3).

An alternative proof is to decompose the total change in demand

$$Q(p_1, y_0, \eta) - Q(p_0, y_0, \eta) = \big(Q(p_1, y_0, \eta) - Q(p_1, y_1, \eta)\big) + \big(Q(p_1, y_1, \eta) - Q(p_0, y_0, \eta)\big)$$

for $y_1 \in [y_0, y_0 + p_1 - p_0]$. The first difference $Q(p_1, y_0, \eta) - Q(p_1, y_1, \eta)$ is the income effect. From the above argument of revealed preference, we know the second difference $Q(p_1, y_1, \eta) - Q(p_0, y_0, \eta)$ is negative. So we obtain the same inequality

$$Q(p_1, y_0, \eta) - Q(p_0, y_0, \eta) \leq Q(p_1, y_0, \eta) - Q(p_1, y_1, \eta).$$

Notice that for $y_1 = y_0 + p_1 - p_0$, we can interpret $Q(p_1, y_1, \eta)$ as the Slutsky compensated demand function if this individual had enough income to afford her original choice.

2. The C.D.F.s of CV and EV are monotone if and only if (*iff*) $q(p, y) \in [0, 1]$,

$$q(p + a, y + a) \leq q(p, y) \quad \text{and} \tag{17}$$

$$q(p + a, y) \leq q(p, y), \tag{18}$$

for any $p, y, a \geq 0$. (3) implies (17) by setting $b = c$. Setting $c = 0$ implies (3) *iff* (18). For the other direction, (17) and (18) implies $q(p, y) \geq q(p + c, y + c) \geq q(p + b, y + c)$, which is (3).

When q is differentiable, the result is a trivial application of calculus. Proving (18) *iff* the second object in (4) is immediate. To see (17) implies the first object in (4), $\lim_{a \to 0} a^{-1}\big(q(p + a, y + a) - q(p + a, y)\big) \leq \lim_{a \to 0} a^{-1}\big(q(p, y) - q(p + a, y)\big)$. To show the reverse direction, $q(p + a, y + a) - q(p, y) = \int_0^a \partial q(p + s, y + s)/\partial p + \partial q(p + s, y + s)/\partial y ds \leq 0$. $\square$

9.2 Proofs in Section 4

We first give an overview of the proof of Theorem 1 and discuss some technical aspects. Consider the mapping $\phi(\theta_0) = (\varphi \circ \psi)(\theta_0)$, where the inner mapping $\psi : \mathbb{D}_\phi \to \mathcal{C}(\mathcal{S})$ is defined by $\psi(\theta(s, \cdot)) \equiv \min_{t \in [0,1]} \theta^{(1)}(s, t)$ subject to $\theta^{(2)}(s, t) \leq 0$. And the outer mapping $\varphi : \mathcal{C}(\mathcal{S}) \to \mathcal{R}$ is defined by $\varphi(\vartheta) \equiv \min_{s \in \mathcal{S}} \vartheta(s)$, for any $\vartheta \in \mathcal{C}(\mathcal{S})$. As φ is an unconstrained optimization problem, showing φ is Hadamard directionally differentiable is known in the literature, e.g., Lemma S.4.7 in the Supplemental Appendix of Fang and Santos (2019). We show the inner mapping ψ is Hadamard directionally differentiable in Lemma 1 below.

The complication occurs in the inner mapping ψ when multiple constraints bind. In such a case, the Lagrangian multipliers could have multiple solutions satisfying the first order necessary conditions described in $\Lambda(s, t)$. To derive the Hadamard directional derivative, we need to consider sequences of constraints that nearly bind. Thus the directional derivative of the mapping can be well defined for all sequences; see Lemma 2 below. Under our setup, we provide low-level conditions for the high-level conditions in Theorem 4.25 in Bonnans and Shapiro (2013) for general results with nonlinear nonconvex objective functions and constraints.[18]

[18]In particular, Lemma 2 verifies the conditions (iii) and (iv) of Theorem 4.25 in Bonnans and Shapiro (2013). Assumption 3(i) implies the condition (ii) of Theorem 4.25 in Bonnans and Shapiro (2013). By Assumption 3(ii) and $t \in [0, 1]$, the directional regularity condition (ii) of Theorem 4.25 in Bonnans and Shapiro (2013) holds.

Let $\theta'(t)$ be a vector of the first order derivatives of the elements of the vector $\theta(t)$. Let $l^\infty(\mathcal{X})$ be a space of bounded functions on $\mathcal{X}$ endowed with the sup-norm $\|f\|_\infty = \sup_{x\in\mathcal{X}}|f(x)|$.

Lemma 1 (Hadamard directional differentiability) *Consider $\psi(\theta) = \min_{t\in\mathcal{T}}\theta^{(1)}(t)$ subject to $\theta^{(2)}(t) \le 0$, where (i) $\theta = (\theta^{(1)}, \theta^{(2)\top})^\top \in \mathbb{D}_\psi \equiv \mathcal{C}(\mathcal{T}) \times \cdots \times \mathcal{C}(\mathcal{T}) \subset \mathcal{R}^{d_r+1}$ for a convex compact set $\mathcal{T} \subset \mathcal{R}$. Assume (ii) $\theta_0^{(2)} : \mathcal{T} \to \mathcal{R}^{d_r}$ is linear; (iii) There is an interior point $t \in \mathcal{T}$ satisfying $\theta_0^{(2)}(t) < 0$; (iv) There exists a set of optimal solutions $\bar{\mathcal{T}}$ in the interior of $\mathcal{T}$ such that $\psi(\theta_0) = \theta_0^{(1)}(\bar{t})$ for $\bar{t} \in \bar{\mathcal{T}}$; (v) $\theta_0^{(1)}$ is continuously differentiable on $\mathcal{T}$. For $\bar{t} \in \bar{\mathcal{T}}$, the Lagrangian multiplier $\lambda \in \Lambda(\bar{t}) \equiv \{\lambda \in \mathcal{R}^{d_r} : \theta_0^{(1)\prime}(\bar{t}) + \lambda^\top\theta_0^{(2)\prime}(\bar{t}) = 0, \lambda \ge 0, \lambda^\top\theta_0^{(2)}(\bar{t}) = 0\}$.*

Then $\psi(\theta_0)$ is Hadamard directionally differentiable at θ_0 tangentially to $\mathbb{D}_\psi$ and for $h \in \mathbb{D}_\psi$, $\psi'_{\theta_0}(h) = \min_{t\in\bar{\mathcal{T}}}\max_{\lambda\in\Lambda(t)} h^{(1)}(t) + \lambda^\top h^{(2)}(t)$.

Lemma 2 *Let the conditions in Lemma 1 hold. Consider a sequence of functions $\theta_n = (f_n, R_n^\top)^\top \equiv \theta_0 + \varepsilon_n h_n \in \mathbb{D}_\psi$, where $\theta_0 = \left(f, R^\top\right)^\top \in \mathbb{D}_\psi$, for $\varepsilon_n \downarrow 0$, $\{h_n\} \in \mathbb{D}_\psi$, and $h_n \to h \in \mathbb{D}_\psi$ uniformly on $\mathcal{T}$. Assume $J_+ \equiv \left\{j : R^{(j)}(\bar{t}) = 0\right\}$ to be non-empty. Define a sequence $\{v_n\}$ by $\Pi_{j\in J_+} R_n^{(j)}(v_n) = 0$ and $R_n(v_n) \le 0$. Then (i) for n large enough, $\max_{\lambda\in\Lambda(\bar{t})}\lambda^\top R_n(v_n) = 0$; (ii) $|v_n - \bar{t}| = O(\varepsilon_n)$; (iii) for any $\lambda \in \Lambda(\bar{t})$, $L(v_n, \lambda, \theta_0) - L(\bar{t}, \lambda, \theta_0) = o(\varepsilon_n)$, where $L(t, \lambda, \theta) \equiv \theta^{(1)}(t) + \lambda^\top\theta^{(2)}(t)$.*

Proof of Lemma 1 Let $\theta_0 = \left(\theta_0^{(1)}, \theta_0^{(2)\top}\right)^\top \equiv \left(f, R^\top\right)^\top \in \mathbb{D}_\psi$, for notational simplicity. Consider any $\bar{t} \in \bar{\mathcal{T}}$. Then $\psi(\theta_0) = \theta_0^{(1)}(\bar{t}) = f(\bar{t}) \le f(t), \forall t : R(t) \le 0$. Condition (iii) implies $\Lambda(\bar{t})$ is non-empty and compact.

Consider a sequence of approximating functions $\theta_n \equiv (f_n, R_n^\top)^\top = \theta_0 + \varepsilon_n h_n \in \mathbb{D}_\psi$ for $\varepsilon_n \downarrow 0$, $\{h_n\} \in \mathbb{D}_\psi$, and $h_n \to h \in \mathbb{D}_\psi$ uniformly on $\mathcal{T}$. Let c denote a generic positive constant. So for any $c \in \mathcal{R}_+^{d_r+1}$ and n large enough, $\sup_{t\in\mathcal{T}}|\theta_n(t) - \theta_0(t)| \le c\varepsilon_n$. Let a sequence of optimal solution $t_n \equiv \arg\min_{t\in\mathcal{T}} f_n(t)$ subject to $R_n(t) \le 0$. So $\psi(\theta_n) = f_n(t_n)$ and the set of Lagrangian multiplier is

$$\Lambda_n \equiv \left\{\lambda \in \mathcal{R}^{d_r} : L(t_n, \lambda, \theta_n) = \min_{t\in\mathcal{T}} L(t, \lambda, \theta_n), \lambda \ge 0, \lambda^\top\theta_n^{(2)}(t_n) = 0\right\},$$

where $L(t, \lambda, \theta) \equiv \theta^{(1)}(t) + \lambda^\top\theta^{(2)}(t)$.[19]

The proof is made up of three steps by showing (I) $|t_n - \bar{t}| = o(1)$ and the following inequalities (II) and (III):

$$\min_{\bar{t}\in\bar{\mathcal{T}}}\max_{\lambda\in\Lambda(\bar{t})} L(\bar{t}, \lambda, h) \overset{\text{(II)}}{\le} \liminf_{n\to\infty}\frac{\psi(\theta_n) - \psi(\theta_0)}{\varepsilon_n} \le \limsup_{n\to\infty}\frac{\psi(\theta_n) - \psi(\theta_0)}{\varepsilon_n} \overset{\text{(III)}}{\le} \min_{\bar{t}\in\bar{\mathcal{T}}}\max_{\lambda\in\Lambda(\bar{t})} L(\bar{t}, \lambda, h).$$

We start with a preliminary result. Claim 1 and the compactness of $\mathcal{T}$ ensure that t_n exists and Λ_n is non-empty.

Claim 1: The constrained set $\{t \in \mathcal{T} : R_n(t) \le 0\}$ is non-empty for n large enough.

Proof. By condition (iii), there exists a point $v \in \mathcal{T}$ such that $R(v) < -\eta$ for some $\eta > 0$. By condition (ii), $R(\bar{t} + \zeta(v - \bar{t})) = (1 - \zeta)R(\bar{t}) + \zeta R(v) \le -\eta\zeta$ for all $\zeta \in [0, 1]$. There exists a sequence $\zeta_n \to 0^+$ and $v_n = \bar{t} + \zeta_n(v - \bar{t}) \to \bar{t}$ such that for n large enough, $R_n(v_n) \le R(v_n) + c\varepsilon_n \le -\eta\zeta_n + c\varepsilon_n \le 0$ by choosing $c = \eta/2$ and $\zeta_n = \varepsilon_n$. □

[19] In Section 4, we suppress θ_0 in the Lagrangian function $L(s, t, \lambda)$ with an abuse of notation without loss of clarity.

When some constraint binds, we need to control the sequence of constraints $\{R_n(t_n)\}$ that nearly binds and $\{R_n(\bar{t})\}$ that has positive elements. Define $J_+ \equiv \{j : R^{(j)}(\bar{t}) = 0\}$ that collects the binding constraints. Consider the case when some constraint binds $R^{(b)}(\bar{t}) = 0$ for $b \in J_+ \neq \emptyset$ and $\Lambda(\bar{t})$ might not be a singleton. For any positive integer M, there exists some $n > M$, such that **Case A**: $R_n^{(b)}(\bar{t})$ is positive or **Case B**: $R_n(t_n)$ is negative, i.e., all constraints are slack. To account for these two cases, we define an auxiliary sequence $\{v_n\}$ by $\Pi_{j \in J_+} R_n^{(j)}(v_n) = 0$ and $R_n(v_n) \leq 0$. By Claim 1, v_n exists for n large enough. The properties of $\{v_n\}$ in Lemma 2 are key to the proof.[20]

For the rest of the proof, all arguments hold for n large enough, unless noted otherwise. We do not repeat the statement "for n large enough" for ease of exposition and without loss of clarity. Now we show (I), (II), and (III).

(I) We first consider the case when some constraints bind, i.e., $J_+ \neq \emptyset$.

Claim 2: $f(t_n) - f(\bar{t}) \geq -o(1)$.
Proof. To see if t_n is a feasible solution to solve $\psi(\theta_0)$, observe that $R_n(t_n) - c\varepsilon_n \leq R(t_n) \leq R_n(t_n) + c\varepsilon_n \leq c\varepsilon_n$.

- **Case 1:** $R(t_n) \leq 0$**.** It must be that $f(\bar{t}) \leq f(t_n)$.
- **Case 2:** $R(t_n) \in (0, c\varepsilon_n]$**.** For any $b \in J_+$, without loss of generality, let $R^{(b)\prime} > 0$. $R^{(b)}(t_n) \in (R^{(b)}(\bar{t}), c\varepsilon_n]$ implies $t_n - \bar{t} \in (0, c\varepsilon_n / R^{(b)\prime}]$. By the mean value theorem and condition (v), $f(t_n) = f(\bar{t}) + f'(\tilde{t}_n)(t_n - \bar{t})$, where $\tilde{t} \in [\bar{t}, t_n]$. When $f'(\tilde{t}_n) \geq 0$, $f(t_n) \geq f(\bar{t})$. When $f'(\tilde{t}_n) \leq 0$, $f(t_n) \geq f(\bar{t}) + f'(\tilde{t}_n) c\varepsilon_n / R^{(b)\prime} = f(\bar{t}) - o(1)$.

Combining the two cases, we obtain $f(t_n) - f(\bar{t}) \geq -o(1)$. □

Claim 3: $f_n(t_n) \leq f_n(\bar{t}) + o(1)$.
Proof. By $R_n(v_n) \leq 0$ and Lemma 2(ii), $f_n(t_n) \leq f_n(v_n) = f_n(\bar{t}) + o(1)$. We use v_n to account for **Case A** where $f_n(t_n)$ might be larger than $f_n(\bar{t})$. □

By Claim 2 and Claim 3, $-o(1) \leq f(t_n) - f(\bar{t}) \leq f_n(t_n) - f_n(\bar{t}) + f(t_n) - f_n(t_n) + f_n(\bar{t}) - f(\bar{t}) \leq o(1)$. So $|t_n - \bar{t}| = o(1)$.

Now we consider the case when all the constraints are slack $R(\bar{t}) < 0$, i.e., $J_+ = \emptyset$. Since $f \in \mathcal{C}(\mathcal{T})$ and R is linear, there exists a constant $\eta > 0$ such that $f(\bar{t}) < f(t)$ for all $t \neq \bar{t}$ satisfying $R(t) \leq \eta$. Since $R_n(t_n) \leq 0$, there exists a positive constant c such that $R(t_n) \leq c\varepsilon_n \leq \eta$. So $f(t_n) \geq f(\bar{t})$, as Claim 2 above.

For any positive constant c, $\varepsilon_n \leq \min\{-R(\bar{t})/c\}$ and $R_n(\bar{t}) \leq R(\bar{t}) + c\varepsilon_n \leq 0$. So $f_n(t_n) \leq f_n(\bar{t})$, as Claim 3 above.

By the same argument for the case $J_+ \neq \emptyset$, $|t_n - \bar{t}| = o(1)$.

(II) **Claim 4:** For any $\lambda \in \Lambda(\bar{t})$, $L(t_n, \lambda, \theta_0) - L(\bar{t}, \lambda, \theta_0) = o(\varepsilon_n)$.
Proof. For any $\lambda_n \in \Lambda_n$,

$$\begin{aligned} 0 &\geq L(t_n, \lambda_n, \theta_n) - L(\bar{t}, \lambda_n, \theta_n) = f_n(t_n) - f_n(\bar{t}) + \lambda_n^\top (R_n(t_n) - R_n(\bar{t})) \\ &\geq f(t_n) - f(\bar{t}) + \lambda_n^\top (R(t_n) - R(\bar{t})) - c\varepsilon_n = o(1) \end{aligned} \tag{19}$$

[20] Because of the linear constraints in Assumption 3(i), we can explicitly construct $\{v_n\}$ that has the properties in Lemma 2. With nonconvex constraints, Shapiro (1991) assumes a unique Lagrangian multiplier and uses a constraint qualification.

by $\|\theta_n - \theta_0\|_\infty = O(\varepsilon_n)$ and (I). For any $\lambda \in \Lambda(\bar{t})$, by (I),

$$0 \geq L(\bar{t}, \lambda, \theta_0) - L(t_n, \lambda, \theta_0) = f(\bar{t}) - f(t_n) + \lambda^\top (R(\bar{t}) - R(t_n)) = o(1). \tag{20}$$

The summation of (19) and (20) yields $(\lambda_n - \lambda)^\top (R(t_n) - R(\bar{t})) = O(\varepsilon_n)$. Together with condition (ii), $|t_n - \bar{t}| = O(\varepsilon_n)$.[21] For any $\lambda \in \Lambda(\bar{t})$, $L(t_n, \lambda, \theta_0) - L(\bar{t}, \lambda, \theta_0) = f(t_n) - f(\bar{t}) + \lambda^\top (R(t_n) - R(\bar{t})) = \left(f'(\bar{t}) + \lambda^\top R'(\bar{t})\right)(t_n - \bar{t}) + o(|t_n - \bar{t}|) = o(\varepsilon_n)$. □

For any $\lambda \in \Lambda(\bar{t})$, $\psi(\theta_n) - \psi(\theta_0) \geq f_n(t_n) + \lambda^\top R_n(t_n) - f(\bar{t}) - \lambda^\top R(\bar{t}) = L(t_n, \lambda, \theta_n) - L(t_n, \lambda, \theta_0) + L(t_n, \lambda, \theta_0) - L(\bar{t}, \lambda, \theta_0) = \varepsilon_n L(t_n, \lambda, h_n) + o(\varepsilon_n)$ by Claim 4. Therefore $\liminf_{n\to\infty}(\psi(\theta_n) - \psi(\theta_0))/\varepsilon_n \geq \liminf_{n\to\infty} L(t_n, \lambda, h_n) + o(\varepsilon_n) \geq \min_{\bar{t}\in\bar{\mathcal{T}}} L(\bar{t}, \lambda, h)$. The inequalities hold for any arbitrary $\lambda \in \Lambda(\bar{t})$, so we prove (II).

(III) When all the constraints are slack, $\Lambda(\bar{t}) = \{\mathbf{0} \in \mathcal{R}^{d_r}\}$ is a singleton. Since $R(\bar{t}) < 0$, $R_n(\bar{t}) \leq R(\bar{t}) + c\varepsilon_n \leq 0$ by choosing c such that $\varepsilon_n \leq \min\{-R(\bar{t})/c\}$ for n large enough. So $\psi(\theta_n) - \psi(\theta_0) \leq f_n(\bar{t}) - f(\bar{t}) = L(\bar{t}, \lambda, \theta_n) - L(\bar{t}, \lambda, \theta_0) = \varepsilon_n L(\bar{t}, \lambda, h_n)$. So (III) holds.
Now for the case when some constraint binds $R^{(b)}(\bar{t}) = 0$ for $b \in J_+$. When $\mathbf{0} \notin \Lambda(\bar{t})$, $\max_{\lambda\in\Lambda(\bar{t})} \lambda^\top R_n(t_n)$ can be negative due to **Case B** above. So we use $\{v_n\}$ and Lemma 2(i) to derive $\psi(\theta_n) = f_n(t_n) \leq f_n(v_n) \leq f_n(v_n) + \max_{\lambda\in\Lambda(\bar{t})} \lambda^\top R_n(v_n) = \max_{\lambda\in\Lambda(\bar{t})} L(v_n, \lambda, \theta_n)$. Then

$$\begin{aligned}\psi(\theta_n) - \psi(\theta_0) &\leq \max_{\lambda\in\Lambda(\bar{t})} L(v_n, \lambda, \theta_0) + \varepsilon_n \max_{\lambda\in\Lambda(\bar{t})} L(v_n, \lambda, h_n) - \max_{\lambda\in\Lambda(\bar{t})} L(\bar{t}, \lambda, \theta_0) \\ &\leq \varepsilon_n \max_{\lambda\in\Lambda(\bar{t})} L(v_n, \lambda, h_n) + \max_{\lambda\in\Lambda(\bar{t})} \{L(v_n, \lambda, \theta_0) - L(\bar{t}, \lambda, \theta_0)\}.\end{aligned}$$

By Lemma 2(iii), continuity of L, and $\bar{t}$ being an arbitrary point of $\bar{\mathcal{T}}$, we prove (III).

□

Proof of Lemma 2 By Claim 1 in the proof of Lemma 1, v_n exists for n large enough.

(i) For any $j \in J_+$, define $\lambda^j \in \mathcal{R}^{d_r}$ by letting its jth element be $-f'(\bar{t})/R^{(j)\prime}$ and its kth element be 0 for $k \neq j$. To show $\lambda^j \in \Lambda(\bar{t})$, it suffices to show $-f'(\bar{t})/R^{(j)\prime} \geq 0$. Without loss of generality, suppose $R^{(j)\prime} > 0$. $R^{(j)}(\bar{t}) = 0$ implies $R^{(j)}(\bar{t} - \eta) \leq 0$ for any small $\eta > 0$. $f(\bar{t}) \leq f(\bar{t} - \eta)$ implies $f'(\bar{t}) \leq 0$. So $\lambda^j \in \Lambda(\bar{t})$ for any $j \in J_+$.
If $R_n^{(j)}(v_n) = 0$ and $R_n^{(k)}(v_n) \leq 0$ for $k \neq j$, then $\max_{\lambda\in\Lambda(\bar{t})} \lambda^\top R_n(v_n) = \lambda^{j\top} R_n(v_n) = 0$.

(ii) By the mean value theorem, $0 = \Pi_{j\in J_+} R_n^{(j)}(v_n) - \Pi_{j\in J_+} R^{(j)}(v_n) + \Pi_{j\in J_+} R^{(j)}(v_n) - \Pi_{j\in J_+} R^{(j)}(\bar{t}) = O(\varepsilon_n) + (v_n - \bar{t}) d\left(\Pi_{j\in J_+} R^{(j)}(\tilde{t}_n^j)\right)/dt$, where $\tilde{t}_n^j$ is between $\bar{t}$ and v_n. $(v_n - \bar{t}) \sum_{j\in J_+} R^{(j)\prime} \cdot \Pi_{k\neq j} \cdot R^{(k)}(\tilde{t}_n^k) = O(\varepsilon_n)$ implies $|v_n - \bar{t}|^{dim(J_+)} = O(\varepsilon_n)$. So $|v_n - \bar{t}| = O(\varepsilon_n)$.

(iii) For any $\lambda \in \Lambda(\bar{t})$, $L(v_n, \lambda, \theta_0) - L(\bar{t}, \lambda, \theta_0) = f(v_n) - f(\bar{t}) + \lambda^\top (R(v_n) - R(\bar{t})) = \Big(f'(\bar{t}) + \lambda^\top R'(\bar{t})\Big)(v_n - \bar{t}) + o(|v_n - \bar{t}|) = o(\varepsilon_n)$.

□

[21] First note that we do not assume f_n to be continuously differentiable at t_n. So the necessary condition of λ_n only implies the first inequality in (19). Second, because the Lagrangian multiplier might not be unique, $|\lambda_n - \lambda|$ might not be $o(1)$.

Proof of Theorem 1 We transform the original problem to $\phi(\theta_0) \equiv (\varphi \circ \psi)(\theta_0)$ and show φ and ψ are two Hadamard directionally differentiable functions. Then the proof is completed by the chain rule for the Hadamard directionally differentiable maps and the delta method.

For the inner optimization, define $\psi : \mathbb{D}_\psi \to \mathcal{C}(\mathcal{S})$ by $\psi(\theta(s,\cdot)) \equiv \min_{t\in[0,1]} \theta^{(1)}(s,t)$ subject to $\theta^{(2)}(s,t) \leq 0$, for any $\theta(s,t) = \left(\theta^{(1)}(s,t), \theta^{(2)}(s,t)^\top\right)^\top \in \mathbb{D}_\psi$. Let $\theta_0(s,t) \equiv \left(f\left(\beta(s,t)\right), R(\beta(s,t))^\top\right)^\top$. To apply Lemma 1, for any $s \in \mathcal{S}$, let $\theta^{(2)}(s,\cdot) : \mathcal{T} \to \mathcal{C}(\mathcal{T}) \times \cdots \times \mathcal{C}(\mathcal{T}) \subset \mathcal{R}^{d_R+2}$ include the deterministic constraints $t-1 \leq 0$ and $-t \leq 0$. Lemma 1 implies for any $s \in \mathcal{S}$, $\psi'_{\theta_0}(h(s,\cdot)) = \min_{\bar{t}\in\bar{\mathcal{T}}(s)} \max_{\lambda\in\Lambda(s)} h^{(1)}(s,\bar{t}) + \lambda^\top h^{(2)}(s,\bar{t})$.

For the outer optimization, define $\varphi : \mathcal{C}(\mathcal{S}) \to \mathcal{R}$ by $\varphi(\vartheta) \equiv \min_{s\in\mathcal{S}} \vartheta(s)$, for any $\vartheta \in \mathcal{C}(\mathcal{S})$. Let $\vartheta_0(s) \equiv \min_{t\in[0,1]} f(\beta(s,t))$ subject to $R(\beta(s,t)) \leq 0$. We show $\vartheta_0(s)$ at the end of the proof. Then we apply the result in Lemma S.4.7 in the Supplemental Appendix of Fang and Santos (2019) to obtain the Hadamard directional derivative of φ at ϑ_0 tangentially to $\mathcal{C}(\mathcal{S})$ to be $\varphi'_{\vartheta_0}(h) = \min_{s\in\bar{\mathcal{S}}} h(s)$ for $h \in \mathcal{C}(\mathcal{S})$.

By Proposition 3.6 in Shapiro (1990), the chain rule gives the Hadamard directional derivative of $\phi(\theta_0)$: $\phi'_{\theta_0}(h) = (\varphi \circ \psi)'_{\theta_0}(h) = (\varphi'_{\vartheta_0} \circ \psi'_{\theta_0})(h) = \min_{s\in\bar{\mathcal{S}}} \min_{\bar{t}\in\bar{\mathcal{T}}(s)} \max_{\lambda\in\Lambda(s,\bar{t})} h^{(1)}(s,\bar{t}) + \lambda h^{(2)}(s,\bar{t})$.

By the delta method in Theorem 2.1 in Fang and Santos (2019),

$$\sqrt{n}\left(\phi(\hat{\theta}) - \phi(\theta_0)\right) = \phi'_{\theta_0}\left(\sqrt{n}(\hat{\theta} - \theta_0)\right) + o_p(1) \xrightarrow{L} \phi'_{\theta_0}\left(\mathbb{G}_0\right).$$

Proof of $\vartheta_0(s) \in \mathcal{C}(\mathcal{S})$. For $s_1, s_2 \in \mathcal{S}$, consider any $\bar{t}_1 \equiv \bar{t}(s_1) \in \bar{\mathcal{T}}(s_1)$ and $\bar{t}_2 \equiv \bar{t}(s_2) \in \bar{\mathcal{T}}(s_2)$. We will show that for any $\delta > 0$, there exists $\epsilon > 0$ such that $|s_1 - s_2| \leq \epsilon$ implies $f(\beta(s_2,\bar{t}_2)) - \delta < f(\beta(s_1,\bar{t}_1)) < f(\beta(s_2,\bar{t}_2)) + \delta$ in the following three cases that exhaust all possible situations.

Note that for each s the linear constraint $R(\beta(s,t)) \leq 0$ has an equivalent expression $t \in [l(s), u(s)] \subseteq [0,1]$ for $l, u \in \mathcal{C}(\mathcal{S})$.

- **Case 1:** Consider the case when $u(s_1) = \bar{t}_1$ and $u(s_2) = \bar{t}_2$. By $u \in \mathcal{C}(\mathcal{S})$, $|\bar{t}_1 - \bar{t}_2| = |u(s_1) - u(s_2)| = o(s_1 - s_2)$. By $f(\beta(s,t)) \in \mathcal{C}(\mathcal{S}\times[0,1])$, $f(\beta(s_2,\bar{t}_2)) - \delta \leq f(\beta(s_1,\bar{t}_1)) \leq f(\beta(s_2,\bar{t}_2)) + \delta$.
- **Case 2:** Consider the case when $u(s_2) = \bar{t}_2 > u(s_1) > \bar{t}_1$ or $u(s_1) > u(s_2) = \bar{t}_2 > \bar{t}_1$. By $f(\beta(s,t)) \in \mathcal{C}(\mathcal{S} \times [0,1])$, $u \in \mathcal{C}(\mathcal{S})$, and the unique optimal solution, for any $\delta > 0$ there exists $\epsilon > 0$ such that $\max\{|s_1 - s_2|, |u(s_1) - u(s_2)|\} \leq \epsilon$ implies

$$\begin{aligned} f(\beta(s_2,\bar{t}_2)) - \delta < f(\beta(s_2,\bar{t}_1)) - \delta &\leq f(\beta(s_1,\bar{t}_1)) \\ &< f(\beta(s_1,u(s_1))) \leq f(\beta(s_1,u(s_2))) + \delta/2 \leq f(\beta(s_2,\bar{t}_2)) + \delta. \end{aligned}$$

- **Case 3:** Consider the case when the constraints are slack $R(\beta(s,\bar{t}(s))) < 0$. By $f(\beta(s,t)) \in \mathcal{C}(\mathcal{S} \times [0,1])$, linear R, and the unique optimal solution $\bar{t}(s)$ for each s, there exists a constant $\eta > 0$ such that $f(\beta(s,\bar{t}(s))) < f(\beta(s,t))$ for all $t \neq \bar{t}(s)$ satisfying $R(\beta(s,t)) \leq \eta$. There exists $\epsilon > 0$ such that $|s_1 - s_2| \leq \epsilon$ implies$R(\beta(s_1,\bar{t}_2)) \leq R(\beta(s_2,\bar{t}_2)) + \eta \leq \eta$. So $f(\beta(s_1,\bar{t}_1)) < f(\beta(s_1,\bar{t}_2))$. The same argument yields $f(\beta(s_2,\bar{t}_2)) < f(\beta(s_2,\bar{t}_1))$.

  By continuity of $f(\beta(s,t))$ in s, $|s_1 - s_2| \leq \epsilon$ implies $f(\beta(s_1,\bar{t}_2))) \leq f(\beta(s_2,\bar{t}_2)) + \delta$ and $f(\beta(s_2,\bar{t}_1)) - \delta \leq f(\beta(s_1,\bar{t}_1))$.

  Combining the above inequalities yields $f(\beta(s_2,\bar{t}_2)) - \delta < f(\beta(s_1,\bar{t}_1)) < f(\beta(s_2,\bar{t}_2)) + \delta$.

The case when the solution is at the lower bound is included in Case 1 and Case 2. □

Proof of Theorem 2 As in the proof of Lemma 1, we first suppress the notation s for the inner optimization. Define $\mathcal{T} \equiv [0,1] \cap \{t : \theta^{(2)}(t) \leq 0\}$ and $\mathcal{T}^\dagger \equiv [0,1] \cap \{t : \theta^{(2)\dagger}(t) \leq 0\}$. Since $\theta^{(2)}(t) = R(\beta(s,t))$ is convex in t, we can write $\mathcal{T} = [\underline{t}, \bar{t}]$ and $\mathcal{T}^\dagger = [\underline{t}^\dagger, \bar{t}^\dagger]$. Without loss of generality, we consider the case when $\bar{t} < \bar{t}^\dagger$, $\theta^{(2)}(\bar{t}) = 0$, $\theta^{(2)\dagger}(\bar{t}^\dagger) = 0$, and $d_R = 1$.

By the Mean Value Theorem, $\theta^{(2)}(\bar{t}^\dagger) - \theta^{(2)}(\bar{t}) = \theta^{(2)'}(t)(\bar{t}^\dagger - \bar{t}) + \theta^{(2)''}(\tilde{t})(\bar{t}^\dagger - \bar{t})^2/2$, where $\tilde{t} \in [\bar{t}, \bar{t}^\dagger]$. For any $\delta \geq 0$, $\theta^{(2)''}(\tilde{t})(\bar{t}^\dagger - \bar{t})^2/2 + \theta^{(2)'}(t)(\bar{t}^\dagger - \bar{t}) + \big(\theta^{(2)\dagger}(\bar{t}^\dagger) - \theta^{(2)}(\bar{t}^\dagger)\big) \leq \delta$. Notice that $\theta^{(2)\dagger}(\bar{t}^\dagger) - \theta^{(2)}(\bar{t}^\dagger)$ is negative. Convexity of $\theta^{(2)}$ implies $\theta^{(2)''}(t) \geq 0$ for all $t \in [0,1]$. So $|\bar{t}^\dagger - \bar{t}| \leq C|\theta^{(2)\dagger}(\bar{t}^\dagger) - \theta^{(2)}(\bar{t}^\dagger) - \delta| = C\big(-\theta^{(2)\dagger}(\bar{t}^\dagger) + \theta^{(2)}(\bar{t}^\dagger) + \delta\big)$. Since δ is chosen arbitrarily, we obtain $|\bar{t}^\dagger - \bar{t}| \leq C\|\theta^{(2)\dagger} - \theta^{(2)}\|_\infty$. This result is trivial when $\theta^{(2)}$ is linear.

Then the Hausdorff distance is

$$d_H(\mathcal{T}, \mathcal{T}^\dagger) \leq \max\left\{\left|\bar{t} - \bar{t}^\dagger\right|, \left|\underline{t} - \underline{t}^\dagger\right|\right\} \leq C\|\theta^{(2)} - \theta^{(2)\dagger}\|_\infty.$$

Now we can prove Lipschitz continuity of ψ:

$$\begin{aligned}
\left|\psi(\theta) - \psi(\theta^\dagger)\right| &= \left|\min_{t\in\mathcal{T}} \theta^{(1)}(t) - \min_{t\in\mathcal{T}^\dagger} \theta^{(1)\dagger}(t)\right| \\
&\leq \left|\min_{t\in\mathcal{T}} \theta^{(1)}(t) - \min_{t\in\mathcal{T}^\dagger} \theta^{(1)}(t)\right| + \left|\min_{t\in\mathcal{T}^\dagger} \theta^{(1)}(t) - \min_{t\in\mathcal{T}^\dagger} \theta^{(1)\dagger}(t)\right| \\
&\leq C\|\theta^{(1)}\|_\infty \; d_H(\mathcal{T}, \mathcal{T}^\dagger) + C\|\theta^{(1)} - \theta^{(1)\dagger}\|_\infty \leq C\|\theta - \theta^\dagger\|_\infty.
\end{aligned}$$

For the outer optimization, the Lipschitz continuity of $\varphi(\vartheta) \equiv \min_{s\in\mathcal{S}} \vartheta(s)$ follows the above proof. Therefore, $\phi(\theta) = (\vartheta \circ \psi)(\theta)$ is Lipschitz continuous.

By Definition 1, consider any $h_n \to h$ and any $\varepsilon_n \downarrow 0$ for $\hat{\phi}'_{\theta_0}(h)$.

$$\begin{aligned}
\left|\hat{\phi}'_{\theta_0}(h) - \phi'_{\theta_0}(h)\right| &= \left|\frac{1}{\varepsilon_n}\left(\phi(\hat{\theta} + h\varepsilon_n) - \phi(\hat{\theta})\right) - \lim_{n\to\infty} \frac{1}{\varepsilon_n}\left(\phi(\theta_0 + h_n\varepsilon_n) - \phi(\theta_0)\right)\right| \\
&\leq \left|\frac{\sqrt{n}}{\sqrt{n}\varepsilon_n}\left(\phi(\hat{\theta} + h\varepsilon_n) - \phi(\theta_0 + h_n\varepsilon_n)\right)\right| + \left|\frac{\sqrt{n}}{\sqrt{n}\varepsilon_n}\left(\phi(\hat{\theta}) - \phi(\theta_0)\right)\right| \\
&\quad + \left|\frac{1}{\varepsilon_n}\phi(\theta_0 + h_n\varepsilon_n) - \lim_{n\to\infty} \frac{1}{\varepsilon_n}\phi(\theta_0 + h_n\varepsilon_n)\right| + \left|\frac{\phi(\theta_0)}{\varepsilon_n} - \lim_{n\to\infty} \frac{\phi(\theta_0)}{\varepsilon_n}\right| \\
&\leq C\frac{1}{\sqrt{n}\varepsilon_n}\left\|\sqrt{n}\left(\hat{\theta} - \theta_0\right) + \sqrt{n}\left(h - h_n\right)\varepsilon_n\right\|_\infty + o_p(1) = o_p(1)
\end{aligned}$$

by choosing $\varepsilon_n \downarrow 0$ and $\sqrt{n}\varepsilon_n \to \infty$. □

References

Andrews, Donald W. K. and Xiaoxia Shi (2013). Inference based on conditional moment inequalities. *Econometrica*, 81:609–666.

Bonnans, J Frederic and Alexander Shapiro (2013). *Perturbation analysis of optimization problems*. Springer Science & Business Media.

Beresteanu, Arie and Francesca Molinari (2008). Asymptotic properties for a class of partially identified models. *Econometrica*, 76(4):763–814.

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

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

Bhattacharya, Debopam (2018b). The empirical content of binary choice models. SSRN working paper No. 2960282.

Blundell, R. and J. L. Powell (2003). Endogeneity in Nonparametric and Semiparametric Regression Models. *Advances in Economics and Econometrics, Theory and Applications, Eighth World Congress*, II, ed. M. Dewatripont, L.P. Hansen, and S.J.Turnovsky, page 312–57. Cambridge University Press, Cambridge, U.K.

Bontemps, Christian, Thierry Magnac, and Eric Maurin (2012). Set identified linear models. *Econometrica*, 80(3):1129–1155.

Bugni, Federico A., Ivan A. Canay, and Xiaoxia Shi (2017). Inference for functions of partially identified parameters in moment inequality models. *Quantitive Econometrics*, 8(1):1-38.

Chamberlain, Gary (1994). *Quantile regression, censoring and the structure of wages*, page 171–209. Cambridge University Press, Cambridge, U.K.

Chandrasekhar, Arun, Victor Chernozhukov, Francesca Molinari, and Paul Schrimpf (2012). Inference for best linear approximations to set identified functions. cemmap working paper cwp 43/12.

Chen, Xiaohong, (2007). In: Heckman, J.J., Leamer, E.E. (Eds.), Large Sample Sieve Estimation of Semi-Nonparametric Models. In: *Handbook of Econometrics*, vol. 6B. North-Holland, Amsterdam, (Chapter 76).

Chernozhukov, Victor, Whitney K. Newey, and Angres Santos (2015). Constrained conditional moment restriction models. cemmap working paper cwp59/15.

Chesher, Andrew and Adam M. Rosen (2017). Characterizations of identified sets delivered by structural econometric models. *Econometrica*, 85(3):959–989.

Dette, Holger, Stefan Hoderlein, and Natalie Neumeyer (2016). Testing multivariate economic restrictions using quantiles: The example of Slutsky negative semidefiniteness. *Journal of Econometrics*, 191(1):129–144.

Dümbgen, Lutz (1993). On nondifferentiable functions and the bootstrap. *Probability Theory and Related Fields*, 95(1):125–140.

Escanciano, Juan Carlos and Lin Zhu (2014). Set inferences and sensitivity analysis in semiparametric conditionally identified models. cemmap working paper cwp55/13.

Fang, Zheng and Andres Santos (2019). Inference on directionally differentiable functions. *The Review of Economic Studies*, 86(1):377–412.

Freyberger, Joachim and Joel L. Horowitz (2015). Identification and shape restrictions in nonparametric instrumental variables estimation. *Journal of Econometrics*, 189(1):41–53.

Hansen, Bruce E. (2017). Regression kink with an unknown threshold. *Journal of Business and Economic Statistics*, (35)2:228–240.

Hausman, Jerry (1997). Valuation of New Goods under Perfect and Imperfect Competition. *The Economics of New Goods*, ed. Timothy F. Bresnahan and Robert J. Gordon, page 207–248. National Bureau of Economic Research, University Chicago Press.

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

Hirano, Keisuke and Jack R. Porter (2012). Impossibility results for nondifferentiable functionals. *Econometrica*, 80(4):1769–1790.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4):1161–1189.

Hong, Han and Jessie Li (2018). The numerical directional delta method. *Journal of Econometrics*, 206(2):379–394.

Horowitz, Joel and Sokbae Lee (2018). Nonparametric estimation and inference under shape restrictions. *Journal of Econometrics*, 201(1):108–126.

Horowitz, Joel L. and Charles F. Manski (2006). Identification and estimation of statistical functionals using incomplete data. *Journal of Econometrics*, 132(2):445–459.

Hsiao, Cheng (1983). *Regression Analysis with a Categorized Explanatory Variable*. New York:Academic Press.

Imbens, Guido W. and Charles F. Manski (2004). Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857.

Imbens, Guido W. and Whitney K. Newey (2009). Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity. *Econometrica*, 77(5):1481–1512.

Kaido, Hiroaki (2016). A dual approach to inference for partially identified econometric models. *Journal of Econometrics*, 192(1):269–290.

Kaido, Hiroaki and Andres Santos (2014). Asymptotically efficient estimation of models defined by convex moment inequalities. *Econometrica*, 82(1):387–413.

Kaido, Hiroaki, Francesca Molinari, and Jörg Stoye (2017). Confidence Intervals for Projections of Partially Identified Parameters. cemmap working paper cwp49/17.

Kline, Patrick and Andres Santos (2013). Sensitivity to missing data assumptions: Theory and an evaluation of the U.S. wage structure. *Quantitative Economics*, 4(2):231–267.

Lee, Ying-Ying (2018). Efficient propensity score regression estimators of multivalued treatment effects for the treated. *Journal of Econometrics*, 204(2):207–222.

Manski, Charles F. (1994). The selection problem, in *Advances in econometrics*, Sixth World Congress, ed. by C. Sims, pp. 143–170. Cambridge, UK: Cambridge University Press.

Manski, Charles F. (2003). *Partial Identification of Probability Distributions*. Springer Series in Statistics. Springer.

Manski, Charles F. and Elie Tamer (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica*, 70(2):519–546.

Petrin, A. and Train, K., 2010. A control function approach to endogeneity in consumer choice models. *Journal of marketing research*, 47(1), 3–13.

Ponomareva, Maria and Elie Tamer (2011). Misspecification in moment inequality models: back to moment equalities? *The Econometrics Journal*, 14(2):186–203.

Rockafellar, R. Tyrrell (1970). *Convex Analysis*. Princeton landmarks in mathematics and physics. Princeton University Press.

Shapiro, Alexander (1990). On concepts of directional differentiability. *Journal of Optimization Theory and Applications*, 66(3):477–487.

Shapiro, Alexander (1991). Asymptotic analysis of stochastic programs. *Annals of Operations Research*, 30(1):169–186.

Stoye, Jörg (2009). More on confidence intervals for partially identified parameters. *Econometrica*, 77(4):1299–1315.

Tamer, Elie (2010). Partial identification in econometrics. *The Annual Review of Economics*, 2:167–195.

Wan, Yuanyuan and Haiqing Xu (2015). Inference in semiparametric binary response models with interval data. *Journal of Econometrics*, 184(2):347–360.

Woutersen, Tiemen and John Ham (2013). Calculating confidence intervals for continuous and discontinuous functions of parameters. cemmap working papers cwp 23/13.