Demand Analysis with interval-valued sales

Please click here to access the most recent version of this paper

COLIN NDONFACK

Ph.D. in Economics, Université de Montréal

Abstract

Most of existing empirical demand studies assume that market shares are fully observed. However, market sales can be partially observed when they are recorded in interval values, as it is often the case in the recent and growing industry for apps where downloads/installs data can be disclosed by app providers in intervals (e.g. 100+, 1000+, 10k+, 100k+, 1M+). Interval-valued market sales can also be derived when aggregate market shares are observed. This paper is an attempt to use recent developments in partial identification methods to run empirical demand analysis when observed sales are interval-valued. I provide simulation evidence of potential bias when using the middle of interval as the true value for sales to obtain point estimation. An illustration is done in a conditional logit framework using data on recent casual video games available on the Steam Platform. In that illustration, I could rule out some values from the space grid of parameters - including values around the boundary of the space grid - when computing the confidence set by inverting a test for moment inequalities. Moreover, the estimated confidence region allows negative values for the price coefficient, while these are ruled out by the confidence intervals computed when using the midpoint of interval as true value for sales.

Keywords Discrete choice demand; interval data; moment inequalities

1. INTRODUCTION

Existing methods in discrete-choice demand analysis (Berry, 1994; Berry and Haile, 2014; Berry et al., 1995, 2004; Nevo, 2001 have mainly assumed that market shares are point-observed, ruling out the possibility of having censored sales. Censored sales occur when sales are reported in intervals. Demand estimates in such a context are related to censored data regression.

A censored data issue occurs when the "true" value of a variable of interest is not known for some observations and is replaced with some partial information. When those unknown true values are all below (resp. above) a certain known threshold and are all replaced by a constant value, they are said to be left (resp. right) censored. When the censored variable is a continuous one, many empirical works have relied on Tobit regression. This method has been particularly popular in wage analysis, as the potential wages for unemployed people are left-censored and replaced by 0. Tobit regression relies on the main assumption that the distribution of the unobvervables is known (and is gaussian). However, this assumption is not common in studies on discrete choice demand where analysts have an interest in estimating the distribution of the unobservables, i.e. the unobserved product attributes (Berry et al., 1995).

In addition to left and right-censored data, data can also be censored when the unknown true value is replaced by a closed interval where it almost certainly belongs. Interval data regressions cover such cases. In some income analyses, household incomes are often interval-reported in surveys. In that case, one simple approach often used is to take a "representative" point within the interval, generally the mid-point. However, the underlying assumption is that the midpoint is fairly representative of the whole interval. This is problematic when the censored variable has an asymmetric distribution around the midpoint of the interval. This is typically true for income and sales whose distributions are known to be skewed. A second simple approach - an extension of the Tobit model to the interval data regression - would be to address this interval data issue by assuming a known distribution of the unobservables and then running regressions based on the likelihood function to point identify the parameters of interest (Bettin and Lucchetti, 2012; Stewart, 1983). However, in demand analysis, as mentioned above, it is not convenient to make such an assumption as the distribution of these unobservables is insightful to researchers. Instead, it is common to make the less restrictive assumption of conditional mean independence of demand shocks and some instruments (Berry and Haile, 2014; Berry et al., 1995).

Analysts can carry out regression procedures that do not require to assume that they know the distribution of demand shocks and that rely on partial identification. Partial observation of sales inherently leads to partial identification of demand parameters when no restrictions that can sustain point identification are added.

Interval-valued sales can occur in various cases. In the recent and fast-growing app sector, app providers often disclose their sales data - represented by the volume of downloads or installs - in intervals. These intervals are in the following format for apps that are on the Play Store platform: 100+, 1000+, 10k+, 100k+, 1M+, 10M+, etc. As another example, downloads of digital video games on Steam are accessible through SteamPy and Steam API in the form of interval values. Interval-valued sales can be derived when aggregate sales are recorded by the empiricist while product-specific characteristics are observed. In this case, the quantity of sales for each product in a given market is not observed, but is known to ultimately be upper-bounded by the market aggregate sales.

Since interval-valued demand may lead to moment inequalities without point identification, this paper is an attempt to apply partial identification methods to discrete choice demand estimation when sales are observed in interval values, including sales for the outside good. I provide evidence of a potential bias when using the midpoint of the interval as the true value for sales and performing point identification. From an illustration with data on recent casual video game apps from the action-adventure genre on Steam, I was able to compute a confidence set that allows a negative price coefficient to be considered plausible, whereas this is ruled out by the confidence intervals obtained by considering the midpoint as the actual sales value.

The rest of the paper is organized as follows. Section 2 briefly reviews the literature on moment inequalities and partial identification and their application in empirical industrial organization. Section 3 presents the model of discrete choice demand under the conditional logit setting with partial observation of sales, and describes the inference method carried out. Section 4 then implements this method in simulations and compares it with the point estimate obtained by using the midpoint of the interval as the actual sales value. Finally, the section 5 illustrates this method using data from recent casual video games in the action-adventure genre.

2. LITERATURE

From old topics - such as the measurement error leading to bounds in the estimation of coefficients in linear models (Duncan and Davis, 1953; Fréchet, 1951; Frisch, 1934; Marschak and Andrews, 1944) - to recent issues of identification and inference (D. W. Andrews and Shi, 2013; Beresteanu and Molinari, 2008; Bontemps et al., 2012; Chernozhukov et al., 2007, 2013, 2019; Cox and Shi, 2023; Kaido et al., 2019; Manski and Tamer, 2002; Pakes et al., 2015), there are several existing studies in the literature on partial identification. The most contemporary - starting with Chernozhukov et al., 2007 - develop test inversion procedures to obtain the confidence region of parameters based on moment inequalities (D. W. Andrews and Shi, 2013; Chernozhukov et al., 2013, 2019; Cox and Shi, 2023; Kaido et al., 2019; Pakes et al., 2015). The challenge that these studies typically attempt to overcome is to develop an inference strategy with less computational cost, more power and less conservative confidence regions. For a comprehensive literature review, we refer to Canay et al., 2023; Molinari, 2020; Tamer, 2010.

The computation cost-effectiveness sought by some studies (I. Andrews et al., 2023; Cox and Shi, 2023; Kaido et al., 2019; Pakes et al., 2015) consists in reducing the computational load in the case there are several parameters. This is convenient in demand analysis when there are several relevant product attributes to consider. Other studies (Chernozhukov et al., 2019) have the merit to be robust to an increase in the number of moment inequalities, even above the sample size. Since the number of moments increases with the number of instruments, we often have this many moments setting in demand analysis when we use BLP instruments - i.e. rival products' characteristics (Berry et al., 1995). This happens when there are many product characteristics that the researcher considers as relevant to consumers.

Empirical industrial organization economists have used moment inequality restrictions and partial identification for reasons other than partially observed sales. In English auctions, the intrinsic valuations that players assign to the object of the auction are unknown but needed as researchers are interested in the distribution of those private valuations. According to economic theory-based restrictions, the private valuation of a player is lower-bounded by their revealed bid and upper-bounded by the winner's bid. These restrictions are enough to partially identify the distribution of private valuations (Haile and Tamer, 2003; Tamer, 2010).

Another example where moment inequality restrictions and partial identification are used is related to entry games under complete information and pure strategy (Ciliberto and Tamer, 2009; Molinari, 2020). In such games, each firm decides whether to enter the market or not; and the subgame perfect Nash equilibrium (SPNE) can be unique or multiple depending on the firms' payoff shifters. For regions of observable and unobservable payoff shifters where many outcomes are predicted in equilibrium, the economic theory fails to select a single outcome. Therefore, there is a partial identification of outcomes in such regions. As a consequence, if the researcher is interested in the probability that a specific outcome is realized given the observable payoff shifters, if that outcome is often predicted as part of a multiple equilibrium, the probability of interest cannot be point-identified without additional restrictions. However, Ciliberto and Tamer, 2009 suggests that the probability of interest is lower-bounded by the model-implied probability that the outcome is a unique equilibrium; and upper-bounded by the model-implied probability that the outcome is an equilibrium[1]. This leads inherently to moment inequality restrictions and partial identification.

Beyond these economic theory-based restrictions, the revealed preference restrictions widely lead to partial identification with many moment inequalities when they're not associated with an assumption of a known distribution for the unobserved heterogeneity.

The present article is not the first study of demand when sales are observed in formats that do not meet the requirements of most existing discrete choice demand models. Gandhi et al., 2023 dealt with many zero market shares and developed an inference strategy that relies on moment inequalities while still using point identification. Indeed, their moment inequalities derived from the deviations between observed sales and equilibrium demand become asymptotic equalities. However, this asymptotic point identification cannot be applied when all observations have interval-valued market shares, as is the case in this study. Therefore, I contribute to censored demand regression models by applying recent developments

[1] See Molinari, 2020 for more details

in the literature of partial identification from moment inequalities (Chernozhukov et al., 2019) to estimate demand when sales are observed in intervals, including sales for the outside option.

3. MODEL

3.1. Setting

In a market $t \in \{1, ..., T\}$ with J_t products, the payoff that the consumer i gains from getting product j is

$$U_{ijt} = X'_{jt}\theta + \xi_{jt} + \varepsilon_{ijt} \equiv \delta_{jt} + \varepsilon_{ijt} \tag{1}$$

X_{jt} is the observed product characteristics (including endogenous characteristics such as prices) that affect the mean utilities. ξ_{jt} represents the unobserved product heterogeneity - known to the consumer. Their distribution is assumed to be unknowned to the analyst. δ_{jt} refers to the mean utility from product j in market t. ε_{ijt} is the unobserved consumer preference. The analyst observes the data $W_{jt} = (S_{jtL}, S_{jtU}, X_{jt}, Z_{jt})_{j,t}$, where S_{jtL}, S_{jtU} refer to the lower and upper bounds for S_{jt}, the market sales quantity for good j ($j = 0$ for the outside good) and Z_{jt} is a vector of instruments. Denotes M_t the value for the market size. The main assumptions are stated as follows.

Assumption 1. *Data Independence (DI)*

Demand shocks $\{\xi_{jt}\}_{j,t}$ *are independent across* j, t

Assumption 2. *Conditional Logit (CL)*

Consumers' shocks $\{\varepsilon_{jt}\}_{j,t}$ *are i.i.d type-1 extreme value errors*

Assumption 3. *Partial Observation (PO) of sales*

$$Prob\,(S_{jtL} \leq S_{jt} \leq S_{jtU}) = 1 \quad \forall j \in \{0, 1, ..., J_t\}$$

Where bounds S_{jtL}, S_{jtU}*;* $\forall j \in \{0, 1, ..., J_t\}$*; are known by the analyst.*

Assumption 4. *Conditional Mean Independence (CMI) of demand shocks*

$$E\,(\xi_{jt}|Z_{jt}) = 0$$

Where Z_{jt} *is a set of instruments*

While assumptions 2, 3, 4 are useful for the identification, the assumption 1 is useful for the inference. Another implicit assumption relevant for the inference is that there are enough variations in the data, i.e. in all partially and point-observed variables related to the demand equation.

Assumption 2 implies a conditional logit specification for the demand function $\sigma_j(\delta_t)$, where $\delta_t = (\delta_{1t}, ..., \delta_{J_t t})$ is the vector of mean utilities across products in market t. The inverse $\sigma_j^{-1}(\mathbf{s_t})$ is the mean utility δ_{jt} obtained by inverting the vector of market shares $\mathbf{s_t} \equiv \frac{S_t}{M_t}$ (see Berry, 1994; Berry et al., 1995 for the inversion). The inverted demand is $\sigma_j^{-1}(\mathbf{s_t}) = Log(S_{jt}) - Log(S_{ot})$. Since δ_{jt} is also a function of product attributes ($\delta_{jt} = X'_{jt}\theta + \xi_{jt}$), these two expressions for the mean utility - namely $\sigma_j^{-1}(\mathbf{s_t})$ and $X'_{jt}\theta + \xi_{jt}$ - lead to the following demand equation:

$$Log(S_{jt}) - Log(S_{0t}) = X'_{jt}\theta + \xi_{jt} \tag{2}$$

Assumption 3 states that bounds for sales are observed, including bounds for sales quantity of the outside good. Bounds for sales of the outside good and market size bounds are related through: $M_{tb} = S_{otb} + \sum_{l\neq 0} S_{ltb}$; $b \in \{L, U\}$. If the bounds for the market size were directly observed and the bounds for the outside good were not, we could derive bounds for the outside good as $S_{otL} = max(M_{tL} - \sum_{l\neq 0} S_{ltU}, 1)$, $S_{otU} = M_{tU} - \sum_{l\neq 0} S_{ltL}$. In both cases, bounds for market size and bounds for sales of the outside good are observed or derived by the analyst.

3.2. Identification and Estimation method

3.2.1 Market share inversion and bounds for mean utilities

In the following, I describe a pseudo-market share inversion. This is a market share inversion - from Berry, 1994; Berry et al., 1995 - extended to a setting of interval-valued sales with a focus on the conditional logit model.

From assumption 3 we observe bounds of sales S_{jtL}, S_{jtU} $\forall j$, including bounds of sales quantity for the outside good. We know that $\delta_{jt} = Log(S_{jt}) - Log(S_{ot})$ is strictly increasing in S_{jt} and strictly decreasing in S_{ot}. Hence, the Lower bound

$$\begin{aligned}
\delta_{jtL} \equiv & \inf_{\{S_{ot}, S_{1t}, \dots, S_{J_t t}\}} Log(S_{jt}) - Log(S_{ot}) \\
s|t \quad & S_{otL} \leq S_{ot} \leq S_{otU} \\
& S_{tL} \leq S_{lt} \leq S_{tU} \quad \forall l \in \{1, ..., J_t\}
\end{aligned}$$

is reached at

$$\delta_{jtL} = Log(S_{jtL}) - Log(S_{otU}) \tag{3}$$

if $S_{jtL}, S_{otU} > 0$ and $\delta_{jtL} = -\infty$ otherwise.

The upper bound

$$\begin{aligned}\delta_{jtU} \equiv & \sup_{\{S_{ot}, S_{1t}, ..., S_{J_t t}\}} Log(S_{jt}) - Log(S_{ot}) \\ s|t \quad & S_{otL} \leq S_{ot} \leq S_{otU} \\ & S_{tL} \leq S_{lt} \leq S_{tU} \quad \forall l \in \{1, ..., J_t\}\end{aligned}$$

is reached at

$$\delta_{jtU} = Log(S_{jtU}) - Log(S_{otL}) \tag{4}$$

if $S_{jtU}, S_{otL} > 0$ and $\delta_{jtU} = \infty$ otherwise.

We need the lower bounds S_{jtL}, S_{otL} of the sales quantities to be strictly positive. Otherwise the mean utilities can not be bounded above and bellow simultanously. Moreover, the sales quantity of the outside good is relevant in shaping the size of the confidence region. Indeed, the range of the interval for the mean utility is given by $Log(S_{jtU}) - Log(S_{jtL}) + Log(S_{otU}) - Log(S_{otL})$; which is the sum of the range for log-sales quantity of own-product and the range for log-sales quantity of outside good.

3.2.2 Identified set and confidence region

Denote Θ the parameter space. Under assumptions 2, 3 and 4, the identified set as defined by conditional moment inequalities is given by:

$$\Theta_I = \{\theta \in \Theta : E(\delta_{jtL}|Z_{jt}) \leq E(X_{jt}|Z_{jt})'\theta \leq E(\delta_{jtU}|Z_j)\} \tag{5}$$

Considering that I apply an inference method based on unconditional moment inequalities, as the one proposed by Chernozhukov et al., 2019, the inference is actually about an outer set close to Θ_I and defined by unconditional moment restrictions as follows.

$$\Theta_{\mathcal{G}} = \left\{\theta \in \Theta : E[\eta_b g(Z_{jt})(X'_{jt}\theta - \delta_{jtb})] \leq 0 \quad \forall g \in \mathcal{G}, \forall b \in \{L, U\}\right\} \tag{6}$$

Where $\eta_b = -1$ if $b = L$, $\eta_b = 1$ if $b = U$. $\mathcal{G}$ is a convenient countable set of non-negative instrumental functions. Although Θ_I and $\Theta_{\mathcal{G}}$ are identical only under strong restrictions imposed on $\mathcal{G}$, I consider $\Theta_I \approx \Theta_{\mathcal{G}}$. In the present study, $\mathcal{G}$ is built as hypercube indicator functions. This way of obtaining instrumental functions was proposed and shown by D. W. Andrews and Shi, 2013 as suitable to keep information contained in conditional moment restrictions when transforming them into unconditional ones.

Graphs in Figure 1 show some identified sets simulated from three examples of data generating process (DGP). As Θ_I is defined by conditional moment restrictions; its set-value represented in Figure 1 - by black-colored shapes - is computed by solving the linear system of inequalities implied by those restrictions. More precisely, it is built by sampling the subsequent linear inverse problem through Markov chain Monte Carlo (Van den Meersche et al., 2009).

The goal is to estimate the confidence region $C(\alpha)$ of size α for $\theta \in \Theta_I$ from data. Methods developed in recent studies in partial identification generally proceed by inverting a test for moment inequalities:

$$\begin{cases} H_0(\theta) : & \max_{v \in \mathcal{V}} E(m_v(\theta, W_{jt})) \leq 0 \\ H_1(\theta) : & \max_{v \in \mathcal{V}} E(m_v(\theta, W_{jt})) > 0 \end{cases} \tag{7}$$

where:

- $v = (b, g)$; $\mathcal{V} = \{L, U\} \times \mathcal{G}$.
- $W_{jt} = (X_{jt}, S_{tL}, S_{tU})$ is an observed data row.
- $m_v(\theta, W_{jt}) = \eta_b g(Z_{jt})(X'_{jt}\theta - \delta_{jtb})$ is the (unconditional) moment function.

The common inference strategy consists of running the test for each of the parameter values in a grid of parameter space, and then gathering values that fail to reject H_0 as part of the confidence region. The test statistics are in general from the following categories:

- the max statistic: maximum across standardized average moments
- quasi-likelihood ratio statistic
- modified method of moment statistic

In this study, I opt for the max statistic from Chernozhukov et al., 2019 (hence CCK test statistic)[2]:

$$T_n(\theta) = \max_{1 \leq v \leq k} \frac{\sqrt{n}\bar{m}_v(\theta)}{\hat{\sigma}_{n,v}(\theta)}$$

Where:

- $\bar{m}_v(\theta) = \frac{1}{n}\sum_{j,t} m_v(\theta, W_{jt})$ is the average of the lth moment.
- $\hat{\sigma}_{n,v} = \sqrt{\frac{1}{n}\sum_{j,t}(m_v(\theta, W_{jt}) - \bar{m}_v(\theta))^2}$ is its standard deviation.
- n represents the sample size ($n = \sum_{t=1}^{T} J_t$).

This max-type statistic has some robustness properties as its critical value increases slowly with the number of moment inequalities. As a result, it remains suitable in models with large moment inequalities - relative to sample size n. Moreover, it has some power as it increases

[2] In a context of data independence, i.e. when the error term ξ_{jt} is independent across observations j, t

weakly with the value of each moment. See Canay et al., 2023; Chernozhukov et al., 2019 for more on the test statistics and on the suitable properties of the max statistic.

Chernozhukov et al., 2019 propose different test methods to compute critical values: a self-normalized (SN) critical value, which is based on a pivotal statistic (a standard normal); and two bootstrap-based critical values: the multiplier bootstrap (MB) and the empirical bootstrap (EB). They propose three variants for each of these three methods: a one-step variant without moment selection; a two-step variant that discards uninformative moment inequalities that are statistically not binding; and a three-step variant where the authors propose a new selection of "weakly informative inequalities that are potentially binding but do not provide first-order information"(Chernozhukov et al., 2019). To conciliate the computation effectiveness of the SN method with the power and low conservativeness of the bootstrap methods, they also propose hybrid methods.

In this paper, in the simulation exercises carried out in section 4, I am using the two-step hybrid method with multiplier bootstrap critical values in the second step. This choice is motivated by its conciliatory advantage between the computational effectiveness of the SN method in the first step and the power of the bootstrap method in the second step. In the illustration presented in section 5, which uses data on a small number of products from the Steam platform, I consider only a few product attributes, which solves the computational burden and allows me to use the two-step method with multiplier bootstrap critical values in both the first and second steps.

The algorithm from Chernozhukov et al., 2019 in the two-step MB method is described as follows:

Step 0: Generate B vectors $\left\{\{U_{jt}^1\}_{j,t}, ..., \{U_{jt}^b\}_{j,t}, ..., \{U_{jt}^B\}_{j,t}\right\}$ of independent standard normal random variables independent of the data $\{W_{jt}\}_{j,t}$.

Step 1: Moment selection using MB critical value

1. Construct the multiplier bootstrap test statistic for the first step critical value:

$$T_n^b(\theta) = \max_{1\leq v\leq k} \frac{\sqrt{n}\tilde{m}_v^b(\theta)}{\hat{\sigma}_{n,v}(\theta)} \quad \forall b \in \{1, ..., B\}$$

   where

$$\tilde{m}_v^b(\theta) = \frac{1}{n}\sum_{j,t} U_{jt}^b \left(m_v(\theta, W_{jt}) - \bar{m}_v(\theta)\right)$$

2. Compute the related first step critical value $c_n^1(\theta, \beta_n)$ of size β_n, as the $1-\beta_n$-quantile of $\{T_n^b(\theta)\}_{b=1}^{b=B}$; where β_n is a chosen such that $0 < \beta_n < \alpha/2$[3]

[3] Chernozhukov et al., 2019 suggest $\beta_n = \alpha/50$ based on their simulation exercices. Canay et al., 2023, pp 13, also mention this rule of thumb.

3. Get an estimated set of informative moments

$$\hat{\mathcal{V}} = \{v \in \mathcal{V} : \frac{\sqrt{n}\bar{m}_v(\theta)}{\hat{\sigma}_{n,v}(\theta)} > -2c_n^1(\theta, \alpha)\}$$

Step 2: Compute the critical value for the test 7.

1. Construct the multiplier bootstrap test statistic for the second step critical value:

$$Y_n^b(\theta) = \begin{cases} \max_{v \in \hat{\mathcal{V}}} \frac{\sqrt{n}\tilde{m}_v^b(\theta)}{\hat{\sigma}_{n,v}(\theta)} & \text{if } |\hat{\mathcal{V}}| > 0 \\ 0 & \text{if } |\hat{\mathcal{V}}| = 0 \end{cases}$$

2. Compute the critical value $c_n(\theta, \alpha)$ of size α for the test 7, as the $(1 - \alpha + 2\beta_n)$-quantile of $\{Y_n^b(\theta)\}_{b=1}^{b=B}$.

3. Conclude: do not reject $H_0(\theta)$ if $T_n(\theta) \leq c_n(\theta, \alpha)$.

The algorithm for the two-step hybrid method with MB in the second step is similar to the previous algorithm except that there is no step 0 and the first step critical value c_n^1 is from the SN method and has a closed-form expression:

$$c_n^1(\theta, \beta_n) \equiv c_n^1((\beta_n) = \frac{\Phi^{-1}(1 - \alpha/|\mathcal{V}|)}{\sqrt{1 - \Phi^{-1}(1 - \alpha/|\mathcal{V}|)^2/n}}$$

where Φ is the cumulative distribution function of a standard normal random variable, and its inverse Φ^{-1} is the corresponding quantile function.

The level $(1 - \alpha)$-confidence region for $\theta \in \Theta_I$ is defined by the set $CR_n(1 - \alpha)$ of all θ from the parameter space for which $H_0(\theta)$ is not rejected by the test 7. More precisely,

$$CR_n(1 - \alpha) = \{\theta \in \Theta : T_n(\theta) \leq c_n(\theta, \alpha)\}$$

where $T_n(\theta)$ is the test statistic and $c_n(\theta, \alpha)$ is the corresponding critical value such that

$$\theta \in \Theta_I \implies plimProb(\theta \in CR_n(1 - \alpha)) \geq 1 - \alpha$$

Hence, we can assert with a probability weakly greater than $1 - \alpha$ asymptotically that a given parameter from the identified set is an element of $CR(1 - \alpha)$. Chernozhukov et al., 2019[4] provide regularity conditions - related to appropriate variations and bounds in moments of studentized transformations of $m_v(\theta, W_{jt})$ - under which $CR_n(1 - \alpha)$ is asymptotically honest

[4]Theorem A.1. in their online supplement. Chernozhukov et al., 2014 also study asymptotically honest confidence regions.

with a polynomial rate. $CR_n(1-\alpha)$ is asymptotically honest - with a polynomial rate - to a suitable sequence $\mathcal{P}_n$ of classes of distributions P of random vectors W_{jt} if $CR_n(1-\alpha)$ has a coverage being correct uniformly in $\mathcal{P}_n$. This means,

$$\exists c, C > 0 \quad \text{s.t.} \quad \forall n \geq 1, \quad \inf_{P \in \mathcal{P}_n} \inf_{\theta \in \Theta_I(P)} P(\theta \in CR_n(1-\alpha)) \geq 1 - \alpha - Cn^{-c} \tag{8}$$

Hence, asymptotically, the minimum coverage rate guaranteed is weakly greater than $1-\alpha$. In the following, $CR_n(1-\alpha)$ is considered as asymptotically honest.

4. SIMULATIONS

Three simple examples of DGP are simulated to implement the estimation method described in the previous section and compare it with point-estimation obtained when using the midpoint of the interval as the true value of sales. The assumed DGP is described as follows:

- $J_t = 5$ products
- $X_{jt} = (1, P_{jt})'$
- true parameter $\theta_0 = (\beta_0, \beta_p) = (-7, -1.5)$
- Two instruments $Z_{jt} = (Z_j^1, Z_{jt}^2)$ generated from Bernouilli $Ber(0.5)$
  - Z_j^1 is constant across markets
- With $\lambda_0 = 1$ and $\lambda_p = (-1, 1)'$; the price is generated in 3 different cases according to the level of skewness of its distribution (see Figure 5):
  - Example 1: $P_{jt} = \left|\lambda_0 + Z_{jt}'\lambda_p + \mu_{jt}\right| + 0.1$
  - Example 1: $P_{jt} = \left(\lambda_0 + Z_{jt}'\lambda_p + \mu_{jt}\right)^2 + 0.1$
  - Example 1: $P_{jt} = \left|\lambda_0 + Z_{jt}'\lambda_p + \mu_{jt}\right|^3 + 0.1$
- $\xi_{jt} = \mu_{jt} + \nu_{jt}$
- $\mu_{jt} = max\left\{-1, min\{\eta_{jt}, 1\}\right\}$; where η_{jt} is a mixture of 5 normal distributions with the following vectors of mixture probabilities $(0.15, 0.3, 0.2, 0.3, 0.15)$, means $(-0.50, -0.25, 0.00, 0.25, 0.50)$ and standard deviations $(1, 1, 1, 1, 1)$.
- ν_{jt} is from $Uniform(-1, 1)$
- The cutoffs used to get intervals for sales are: 1e-5, 2e4, 5e4, 1e5, 1.5e5, 2e5, 5e5, 1e6, 1.5e6, 2e6, 5e6, 1e7, 1.5e7, 2e7, 5e7, 1e8, 1.5e8, 2e8, 5e8, 1e9

- A draw of 100 markets
  - In each draw t, $M_{tL} = M_{tU} = M_t = 10^9$. Hence bounds for sales of outside good are derived: $S_{otL} = max(M_{tL} - \sum_{l \neq 0} S_{ltU}, 0)$ and $S_{otU} = M_{tU} - \sum_{l \neq 0} S_{ltL}$.

The above three examples of DGP only differ in the price distribution. The underlying correlation between simulated prices P_{jt}, instruments Z_{jt} and demand shocks ξ_{jt} from those examples is reported in Table 4, appendix A. For each example of DGP, the identified set and a confidence region from a single simulated data sample are computed and presented in Figure 1. The confidence intervals that are derived for each coordinate of the parameter θ are presented in Table 1 and compared with confidence intervals computed from a 2SLS point estimation obtained when taking the midpoint of the interval as the true value for sales. One can see that the confidence intervals computed using interval midpoints may not contain the true parameter values. Using a single data sample, the midpoint of the interval leads to confidence intervals that fail to include the true parameter value for examples 2 and 3.

Table 1: Estimates and 95% Confidence Intervals from examples

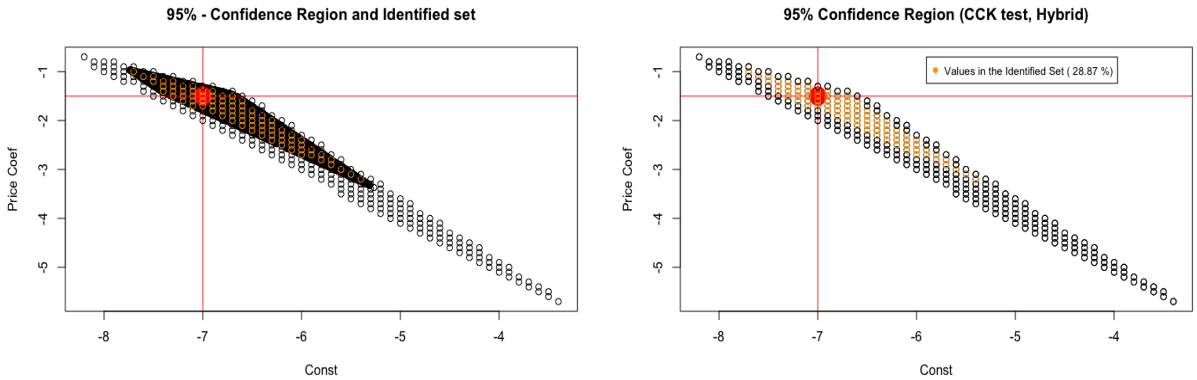
	2SLS with midpoint of interval$^{(a)}$			CCK (Hybrid)		True Values
	Coef	Lower	Upper	Lower	Upper	(θ_0)
Example 1						
Const	-7.025	-7.271	-6.779	-8.2	-3.4	-7
Price Coef	-1.454	-1.634	-1.273	-5.7	-0.7	-1.5
Example 2						
Const	-7.721	-7.886	-7.555	-8.2	-3.6	-7
Price Coef	-0.790	-0.856	-0.724	-5.6	-0.5	-1.5
Example 3						
Const	-8.118	-8.241	-7.995	-8.4	-6.4	-7
Price Coef	-0.273	-0.294	-0.252	-1.9	-0.2	-1.5

Notes These are marginal (or projected) confidence intervals computed from a single simulation for each example of the DGP. The corresponding CCK confidence regions are shown in figure 1. The three examples of DGP only differ by the equation that generates prices. In Example 1: $P_{jt} = \left|\lambda_0 + Z'_{jt}\lambda_p + \mu_{jt}\right| + 0.1$. In Example 2: $P_{jt} = \left(\lambda_0 + Z'_{jt}\lambda_p + \mu_{jt}\right)^2 + 0.1$. In Example 3: $P_{jt} = \left|\lambda_0 + Z'_{jt}\lambda_p + \mu_{jt}\right|^3 + 0.1$
(a) $\delta_{jt} = (Log(S_{jtL} + S_{jtU})) - (Log(S_{otL} + S_{otU}))$

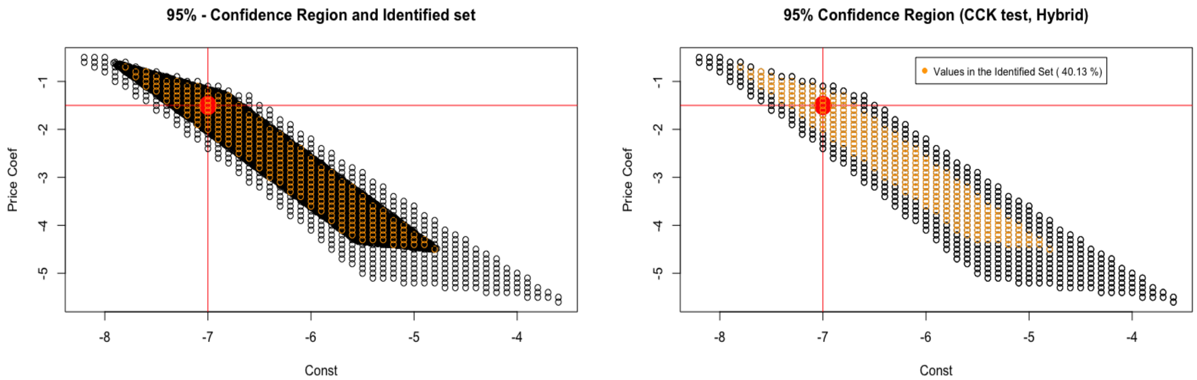
These observations generalize as a midpoint bias introduced when running point estimation based on the midpoint of the interval-valued sales, and the extent of the bias may depend on the DGP. When running 500 Monte Carlo simulations in examples 2 and 3, one gets systematic bias in both the price coefficient and the intercept if one uses the midpoint of the interval as the true value for sales (see Table 2). Indeed, when using the midpoint of the interval in example 2 or example 3, there is 0% probability that the confidence intervals for the price coefficient and the intercept cover the their true values. Meanwhile, the confidence region

Figure 1: Identified set and confidence region

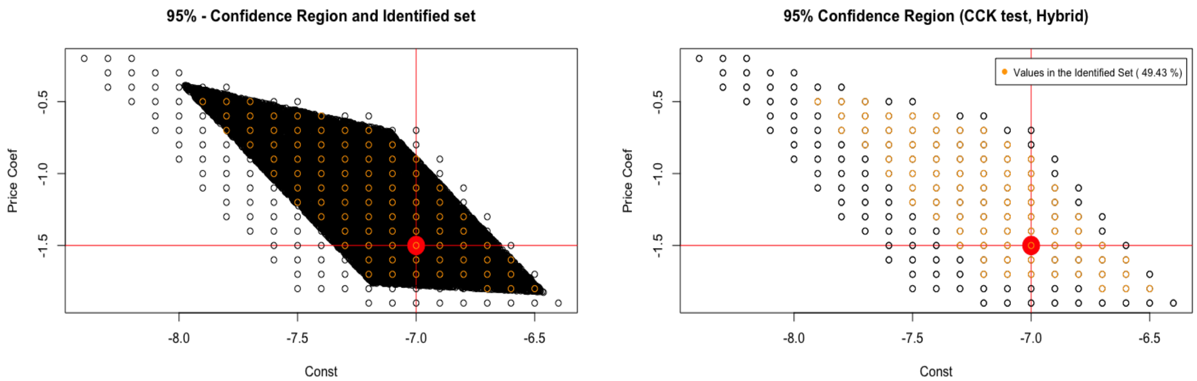
(a) Example 1: $P_{jt} = \left(\lambda_0 + Z'_{jt}\lambda_p + \mu_{jt}\right)^2 + 0.1$



(b) Example 2: $P_{jt} = \left(\lambda_0 + Z'_{jt}\lambda_p + \mu_{jt}\right)^2 + 0.1$



(c) Example 3: $P_{jt} = \left|\lambda_0 + Z'_{jt}\lambda_p + \mu_{jt}\right|^3 + 0.1$



Notes For each example, the identified set is represented by the full and black-colored shape; and the confidence region computed from a single simulation is represented by the sparse dots-shape. The three examples of DGP only differ in the equation that generates prices.

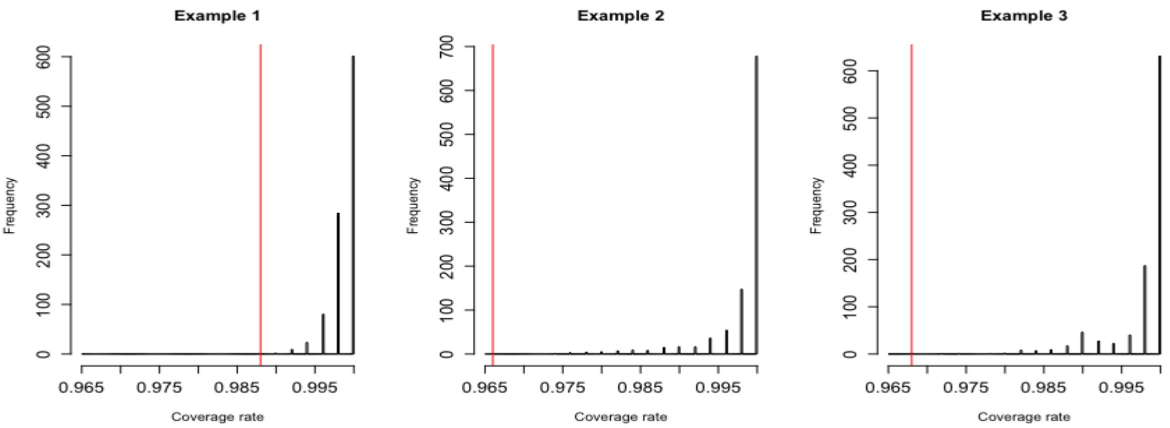
Table 2: 95 %-Confidence Intervals: Averages across 500 simulations

	2SLS with midpoint of interval[a]			CCK (Hybrid)		True Values
	Coef	Lower	Upper	Lower	Upper	(θ_0)
Example 1						
Const	-6.9	-7.147	-6.653	-8.091	-2.722	-7
Price Coef	-1.539	-1.722	-1.356	-6.197	-0.763	-1.5
Coverage rate[b] in example 1						
Const		90.2%		98.8%		
Price Coef		95.6%				
Example 2						
Const	-7.715	-7.874	-7.556	-8.227	-3.894	-7
Price Coef	-0.774	-0.836	-0.711	-5.25	-0.471	-1.5
Coverage rate[b] in example 2						
Const		0%		96.6%		
Price Coef		0%				
Example 3						
Const	-8.129	-8.262	-7.995	-8.316	-6.057	-7
Price Coef	-0.308	-0.331	-0.285	-2.215	-0.23	-1.5
Coverage rate[b] in example 3						
Const		0%		96.8%		
Price Coef		0%				

(a) $\delta_{jt} = (Log(S_{jtL} + S_{jtU})) - (Log(S_{otL} + S_{otU}))$

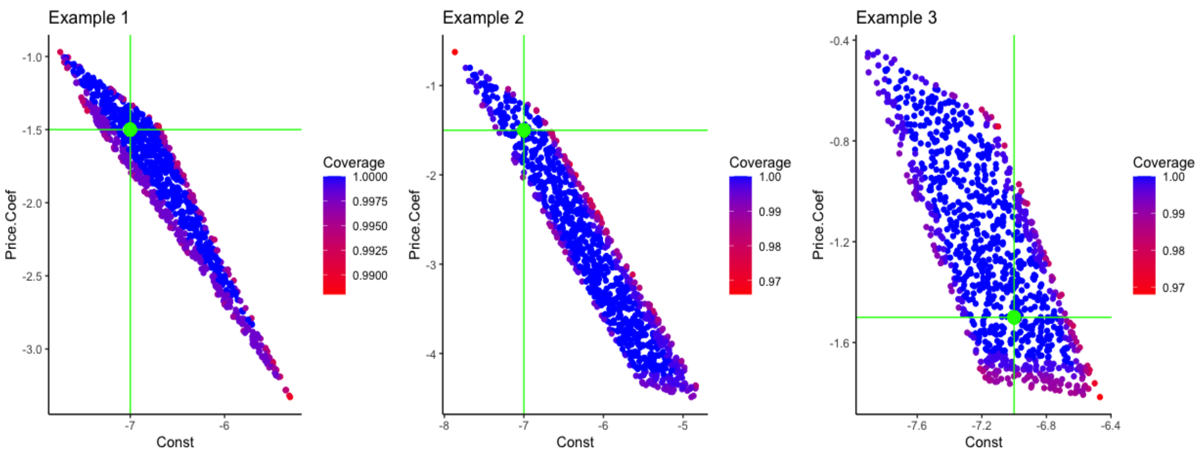
(b) For the 2SLS estimates using the midpoint of interval as the true value for sales, this is the % of simulations with the confidence intervals covering the true parameter value: the coverage rate is provided for each parameter coordinate. For the CCK estimates, this is the minimum % of simulations with the confidence region covering $\theta \in \Theta_I$: the coverage rate is computed pointwise for each $\theta \in \Theta_I$ and then minimized across parameters from Θ_I. The distribution of these pointwise coverage rates is represented in Figure 2. A coverage rate of the identified set, namely the probability that Θ_I is entirely covered by the CCK confidence region, is also computed and its value is 99.9% for example 1, 99.9% for example 2, and 100% for example 3.

Figure 2: Histogram of pointwise coverage rates across parameters in the identified set



Notes Red lines delimit the min values reached by the coverage rate (in the x-axis). The coverage rate is computed for 1000 parameter values randomly drawn from the identified set. A coverage rate for a given parameter $\theta \in \Theta_I$ is the ratio of simulations that give a CCK confidence region covering θ. In each DGP example, most of parameters from Θ_I have 100% coverage rate.

Figure 3: Identified set and pointwise coverage rates



Notes The green point represents the true parameter $\theta = (-7, -1.5)$; which has 100% coverage rate in all three examples of DGP. In all three examples, only the boundary of the identified set Θ_I determines the minimum coverage rate.

computed using the CCK test covers each parameter from the identified set with a probability weakly greater than 0.966 in both examples (see Figure 2). This minimum coverage rate is higher, at 0.988, for example 1. As shown in Figure 3, only parameters in the neighborhoods of boundaries determine this minimum coverage rate. Moreover, since the true value of the price coefficient -1.5 is below the expected confidence interval - $[-0.836, -0.711]$ for example 2 and $[-0.331, -0.285]$ for example 3 - the price coefficient is therefore systematically overestimated in both examples of DGP when using the midpoint as true value for sales. Likewise, the intercept is underestimated. In the simplified setting of conditional logit hypothesized here, a DGP with a longer tail of the price distribution leads to a more positive bias in the price coefficient. However, despite being less conservative, the confidence intervals projected from the CCK confidence region has a good coverage of values in the identified set, including the true parameter.

5. ILLUSTRATION WITH DATA ON STEAM APPS

5.1. Data

An industry of interest for implementing the suggested method in a demand analysis is that of digital apps, as app providers often disclose download numbers in intervals. This is the case for apps that are available on the Google's Playstore platform. I focus on another case, that of personal computer (PC) video games that are available on the Steam platform. Data was collected monthly from the Steam API and SteamPy over 10 months on all apps available on Steam: from October 2023 to July 2024. Table 5 summarizes the main observed outcomes such as prices, interval-valued sales and user bases. Intervals for sales are derived from intervals for cumulative sales (i.e. base of users).

Markets are given by the observation periods for sales and prices, spanning November 2023 to July 2024; and sales units are consumer purchase occasions. To illustrate the suggested method in demand analysis, I focus on markets for recent video game apps from popular genres, i.e. apps from popular genres that were released between January 2023 and July 2024 inclusive. This choice is motivated by the fact that we expect lower variations in sales from old apps as they have reached the last phase in their life cycle. Only very few ones from these old video game apps are still able to attract new users. Only apps from well-known genres are considered in consumer-facing options. These popular genres are the followings: action, adventure, role-playing-game, simulation, sports,racing and education. As selection issues might arise from the heterogeneity in the product launch period and in the product deprecation period, the consumer choice problem is restricted to products that were released during the first quarter of 2023 and that were still available in July 2024. After the previous selection of products to consider, the number of remaining apps ranges from 3417 in October 2023 to 7070 in July 2024. Of these few thousand selected apps, 90 belong to the action-adventure genre and reached at least once the threshold of 20K users between 03-2023 and 07-2024. These 90 apps are the substitutable products that I consider in the consumer choice set. The other apps that are not from the action-adventure genre or never reached 20K users are pooled in the outside option. This focus on the 90 popular apps is driven by the quest for enough variability in (interval-valued) sales and user bases. Figure 7 is indicative of the much concentrated distribution of (interval-valued) user base when we consider all apps available on Steam: As of October 2023, more than half of all apps had the same interval of 0-20000. Moreover, by focusing on a quite specific group of products (recent, casual video games from action-adventure genre), the demand analysis doesn't require us to take many product attributes into account. I am assuming that the 90 applications selected are differentiated by two observed attributes - price and online/multiplayer mode status - and by their unobserved quality.

A summary of product attributes considered in this analysis is provided in table 6. The price (P_{jt}) is an endogenous regressor in the demand equation. Exogenous variations come from the multiplayer/online mode status[5] (MO_{jt}) and two excluded instruments for the price: the number - in log - of total apps the firm has on Steam and a dummy variable for whether or not the firm has a single app available on Steam. The online/multiplayer mode status for rival products couldn't be used as instrument because of low variations. These instruments relate to BLP instruments since they represent aggregate attributes across other products of the same firm.

The mean utility is given by:

$$\delta_{jt} = \theta_c + \theta_p P_{jt} + \theta_{om} MO_{jt} + \xi_{jt} \quad \text{and} \quad \delta_{jt} = Log(S_{jt}) - Log(S_{ot})$$

Sales quantities S_{jt} and S_{ot} are partially observed in intervals. The derived intervals for mean utilities δ_{jt} have bounds that vary as shown in Figure 8. There are enough variations in the upper bounds to make inference on demand possible.

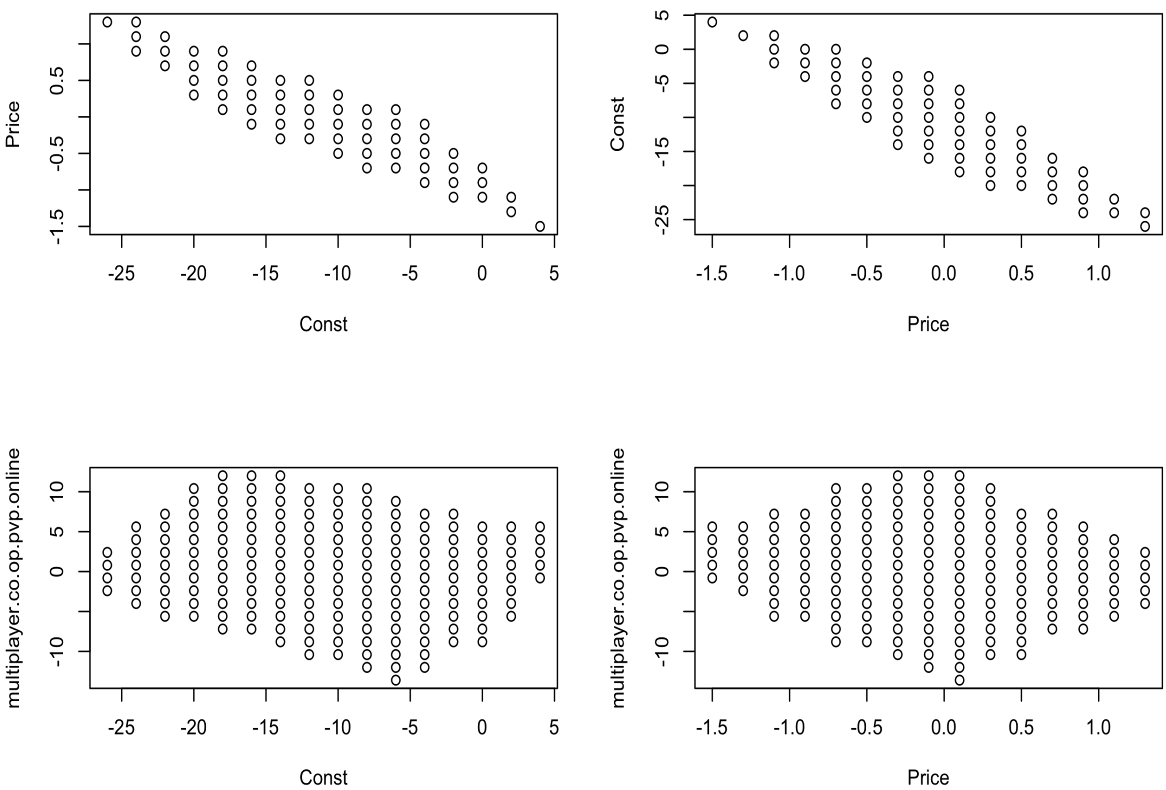
Monthly time periods are considered as markets. Market units are purchase occasions and the sales quantity of a product is the number of purchase occasions in which consumers chose that product. Therefore, given the details previously given on markets, the market size is the total purchase occasions on Steam that led to actual purchases of popular genre apps, published between January 2023 and July 2024. Purchase occasions are assumed to be independent in their respective demand shocks, in line with assumption 1. As already mentioned, I focus on recent and casual apps from the action-adventure genre that were able to gather at least 20K units downloaded from the Steam platform. Assuming that all

[5] More precisely, this is a tag variable indicating to users that they have access to at least one of the following features if they get the product: multiplayer, person-versus-person (pvp), co-op and online gameplay

purchase occasions result in an actual purchase on the platform, the outside option includes other recent and casual apps from big genres that were unable to gather 20K downloads by July 2024.

5.2. Confidence set estimation

Figure 4: Projected 95% Confidence Region



Notes "Price" refers to the price coefficient and "multiplayer.co.op.pvp.online" refers to the coefficient of the multiplayer/online dummy variable.

I perform the proposed approach to build confidence regions for parameters in the demand for recent casual video games from the action-adventure genre; namely the sensitivity to price, the sensitivity to multiplayer/online options and the average taste - for unobserved attributes - of the considered products w.r.t the outside good (i.e. the constant term in θ). To visualize its shape, the estimated three-dimensional confidence set is projected in a joined space of two of its coordinates: first, the (intercept, price coefficient) space; and then, the (intercept, multiplayer/online coefficient) space (Figure 4). Table 3 shows the projected confidence intervals in the space of each coordinate of the confidence set. I also computed point estimates and confidence intervals from a 2SLS method when using the midpoint of the interval as the true value for sales.

Table 3: 95% projected confidence intervals

Coef	Grid		CCK 95%-CI		2SLS 95%-CI$^{(a)}$		
	lower	upper	lower	upper	Coef	lower	upper
Const	-40	10	-26.0	4.0	-8.4430	-8.5742	-8.3118
Price	-2.5	2.5	-1.5	1.3	0.0126	0.0009	0.0242
multiplayer/online	-20	20	-13.6	12.0	0.0476	-0.1563	0.2514

(a) Using the midpoint of interval as the true value for sales, 2SLS-point estimate with confidence interval are provided for each coordinate of the parameter vector. CI refers to "confidence interval".

In a first observation, a 2SLS estimation using the midpoint as the true value for sales provides a confidence interval for θ_p that rules out negative values, which is not convenient for a price sensitivity parameter in the demand function. These negative price coefficients are plausible under the confidence set computed by inverting the CCK test though. Since this confidence set includes zero, the information carried by the data are not enough to sustain that consumers have systematically negative sensitivity to price nor positive sensitivity to the multiplayer/online mode. However, one can at least infer from the data that if the average taste for unobserved attributes - i.e. the intercept - were high (> -5), consumers would have negative sensitivity to price. Furthermore, consumers seem to be not significantly sensitive to the multiplayer/online mode as values of θ_{mo} always include zero for any value of θ_p in the confidence set. Point observation of sales data or additional restrictions in the demand model would be required to infer on the exact and unique sensitivity parameters. Based on the simulations presented in the previous section, considering the midpoint of the interval as the actual value of sales is a restriction that exposes us to potential bias, although it does lead to fairly precise estimates.

6. CONCLUSION

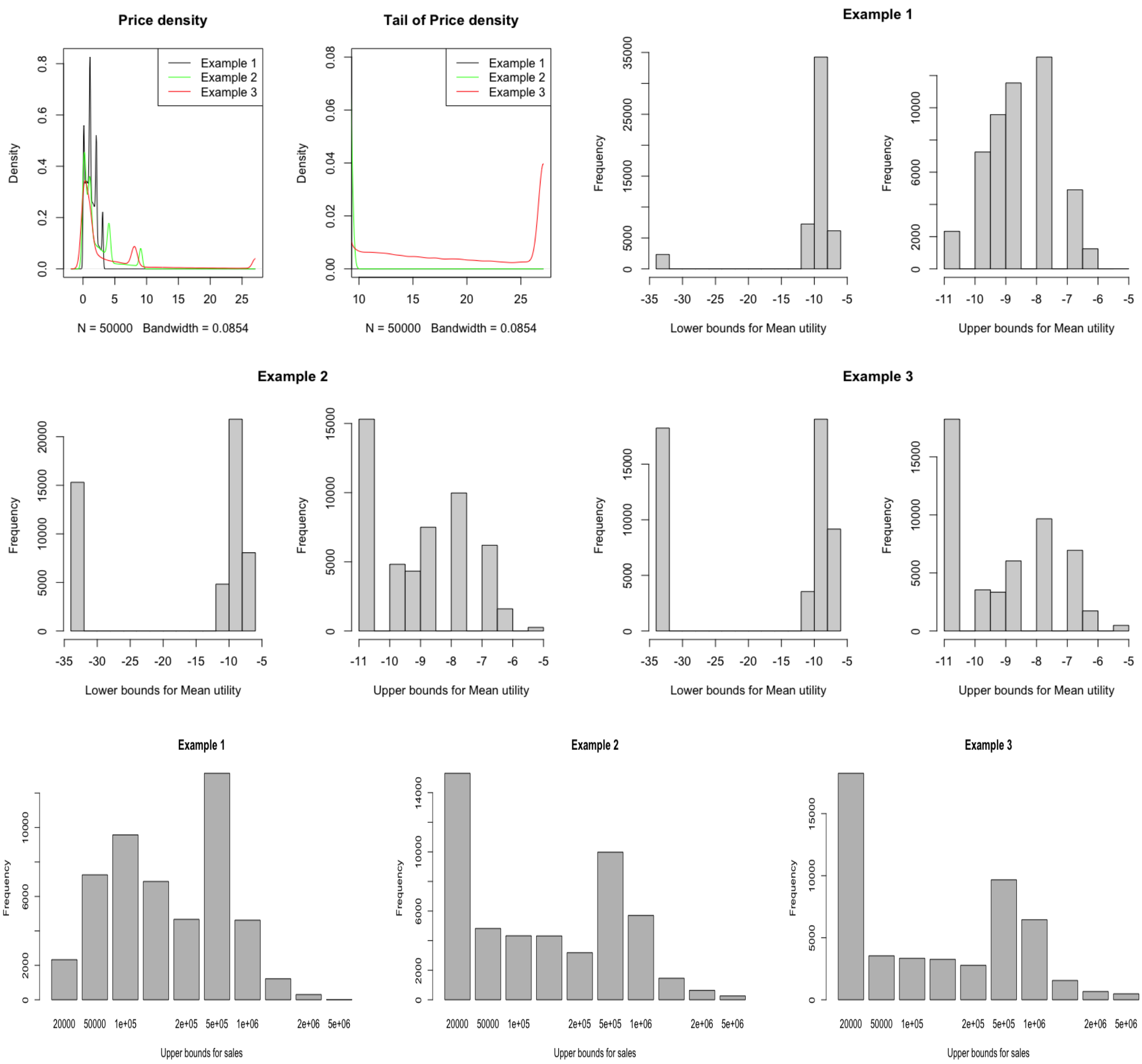
The aim of this study is to estimate demand when market sales are partially observed in interval values. I propose a method which uses a recent moment inequality-based inference strategy as developed by Chernozhukov et al., 2019 to compute the confidence region of the identified set of parameters. I compare the projected confidence intervals from this confidence region with confidence intervals that would be obtained assuming point identification of demand using the midpoint of the interval as the true value of sales. To do so, I run simulations using three simple DGPs that allow the price to be the sole relevant attribute for consumers and those three DGPs differ in the length of the tail for the price distribution. I provide simulation evidence that point estimates of demand using the midpoint of the interval as the true value for sales leads to systematic bias for some DGPs. In particular, among the DGPs used, that bias is greatest for the DGP with the longest tail in the price distribution. For this reason, we can rely more on the confidence region obtained by inverting a test for moment inequalities implied by partial identification. I also illustrate the estimation strategy using demand data for recent and casual PC video games that belong to the action-adventure genre and that reached at the threshold of 20K users by July 2024. In this illustration, I was able to exclude certain values from the parameter space grid - including values around the boundary of the space grid - when computing the confidence set using the CCK test for moment inequalities. This is a sign that the identified set exists as a bounded subset of the parameter space. Moreover, the estimated confidence region allows negative values for the price coefficient, while these are ruled out by the confidence intervals computed when using the midpoint of the interval as the true value for sales.

A. SUPPLEMENT ON SIMULATIONS

Table 4: Correlations from simulation examples

	Example 1	Example 2	Example 3
$Corr(P_{jt}, \xi_{jt})$	0.5173	0.4819	0.4367
$Corr(P_{jt}, Z^1_{jt})$	-0.4059	-0.42	-0.3994
$Corr(P_{jt}, Z^2_{jt})$	0.4082	0.42	0.3967
$Corr(Z^1_{jt}, \xi_{jt})$	0.0016	0.0025	-0.00018
$Corr(Z^2_{jt}, \xi_{jt})$	-0.0023	-0.0014	-0.0029

Figure 5: Distributions of simulated prices; interval-valued sales and mean utilities



Notes In Example 1: $P_{jt} = \left|\lambda_0 + Z'_{jt}\lambda_p + \mu_{jt}\right| + 0.1$. In Example 2: $P_{jt} = \left(\lambda_0 + Z'_{jt}\lambda_p + \mu_{jt}\right)^2 + 0.1$. In Example 3: $P_{jt} = \left|\lambda_0 + Z'_{jt}\lambda_p + \mu_{jt}\right|^3 + 0.1$

B. SUPPLEMENT ON DATA ILLUSTRATION

Figure 6: Price distribution

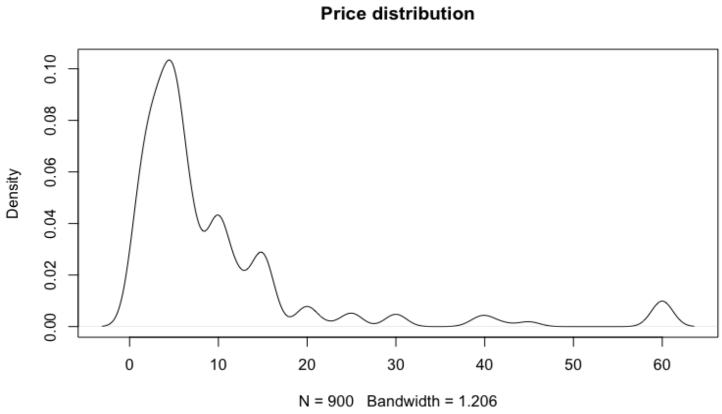


Figure 7: Distribution of interval-valued Base of user for all apps available in Oct 2023

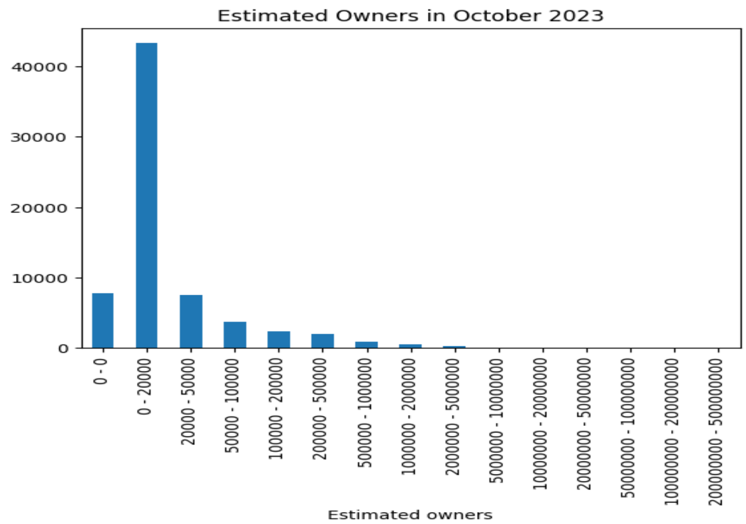


Figure 8: Distribution of Lower and Upper Bounds for δ_{jt}

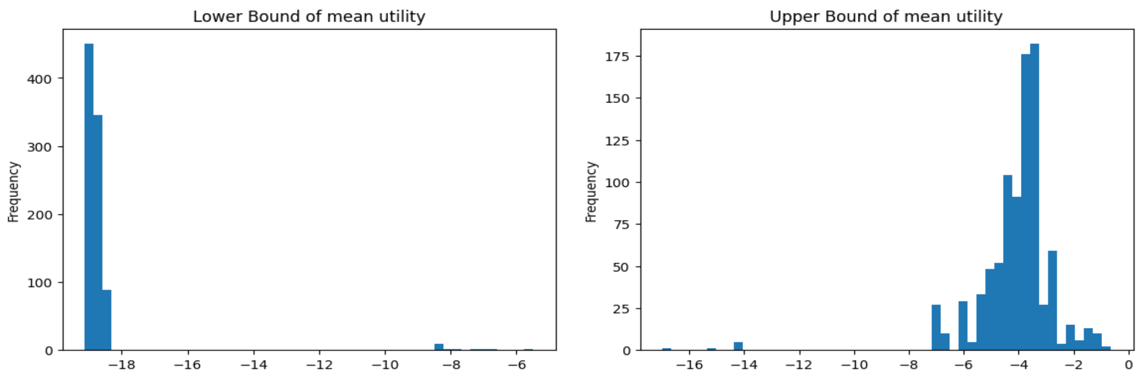


Table 5: Summary of demand data[a]

Collect date	11-2023	12-2023	01-2024	02-2024	03-2024	04-2024	05-2024	06-2024	07-2024
Nb Apps	3937	4330	4594	4713	5317	5774	6216	6621	7070
Nb in outside option	3847	4240	4504	4623	5227	5684	6126	6531	6980
Nb in inside options	90	90	90	90	90	90	90	90	90
Inside options: casual and PTP apps[b] from action-adventure genre, with at least 20K users reached in 07-2024)									
Avg Price	9.018	10.460	10.344	9.672	6.938	9.813	10.231	9.772	9.913
Free to play ratio	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Avg Middle for Base (1e5)	0.577	0.579	0.587	0.644	0.676	0.676	0.747	0.796	0.817
Avg Lower Bound for Base (1e5)	0.339	0.341	0.348	0.391	0.413	0.413	0.460	0.490	0.507
Avg Upper Bound for Base (1e5)	0.816	0.817	0.827	0.898	0.938	0.938	1.033	1.101	1.128
Avg Lower Bound for Sales (1e3)	0.333	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Avg Upper Bound for Sales (1e3)	45.778	47.778	48.556	55.000	54.667	52.444	62.000	64.111	63.778
Ratio of zero Lower Bound for Sales	0.989	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Outside option (apps from other major genres that do not satisfy criteria for inside option apps)									
Avg Price	8.383	7.943	7.981	6.244	7.949	8.078	8.371	8.183	
FTP ratio	0.014	0.014	0.016	0.017	0.017	0.016	0.016	0.017	0.017
Avg Middle Base (1e5)	0.216	0.219	0.225	0.229	0.229	0.224	0.228	0.227	0.227
Avg Lower Bound for Base (1e5)	0.078	0.079	0.082	0.085	0.086	0.082	0.086	0.085	0.085
Avg Upper Bound for Base (1e5)	0.354	0.359	0.367	0.372	0.371	0.365	0.371	0.369	0.369
Avg Lower Bound for Sales (1e3)	1.175	0.330	0.180	0.344	0.631	0.167	0.424	0.214	0.202
Avg Upper Bound for Sales (1e3)	28.869	28.811	29.203	29.161	29.585	28.561	29.483	28.898	28.910
Ratio of zero Lower Bound for Sales	0.988	0.994	0.996	0.996	0.990	0.999	0.993	0.994	0.996

(a): Focus on recent apps (released since Q1-2023) with positive number of users and from major genres
(b): Pay to Play apps, i.e. apps that are accessible through some fees. Oppositely, free to play (FTP) apps have free access.

Table 6: Summary statistics

Variables	count	mean	std	min	25%	50%	75%	max
Lower Bound for Sales (1e3)	810	0.038	1.054	1e-3	1e-3	1e-3	1e-3	30
Upper Bound for Sales (1e3)	810	54.901	72.700	1e-3	30	30	50	80
Lower Bound for Sales of Outside good (1e5)	810	19.97778	11.6286	8.1	14	14.1	26	45.2
Upper Bound for Sales of Outside good (1e5)	810	1541.85556	298.30033	1110.6	1315.3	1546.4	1806.1	2017.9
Price (USD)	810	9.573	11.609	0.490	2.990	4.990	9.990	59.990
Instruments for price								
multiplayer/co-op/pvp/online	810	0.133	0.340	0	0	0	0	1
Log-Nb Releases	810	1.614	1.661	0	0	0.896	3.091	5.215
Single Release	810	0.344	0.475	0	0	0	1	1

Table 7: Grid of parameter space

	Grid												
Const	-40	-38	-36	-34	-32	-30	-28	-26	-24.0	-22.	-20	-18	-16
	-14	-12	-10	-8	-6	-4	-2	0	2	4	6	8	10
Price Coef	-2.5	-2.3	-2.1	-1.9	-1.7	-1.5	-1.3	-1.1	-0.9	-0.7	-0.5	-0.3	-0.1
	0.1	0.3	0.5	0.7	0.9	1.1	1.3	1.5	1.7	1.9	2.1	2.3	2.5
multiplayer/co-op/	-20	-18.4	-16.8	-15.2	-13.6	-12	-10.4	-8.8	-7.2	-5.6	-4	-2.4	-0.8
online/pvp Coef	0.8	2.4	4	5.6	7.2	8.8	10.4	12	13.6	15.2	16.8	18.4	20

Figure 9: Projected 90% confidence region

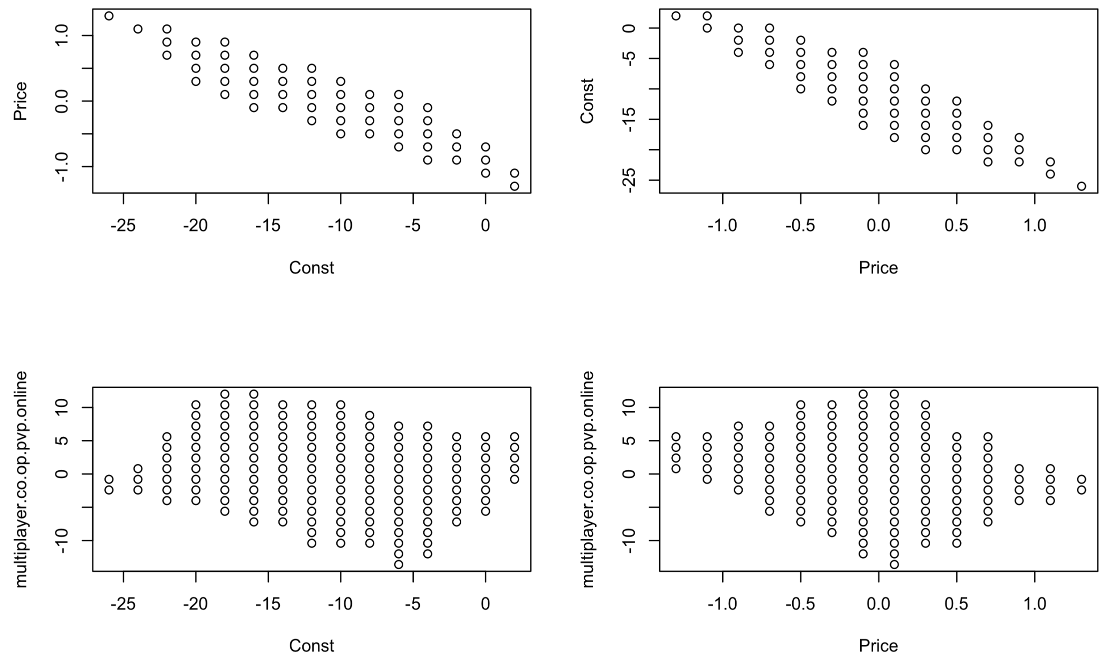
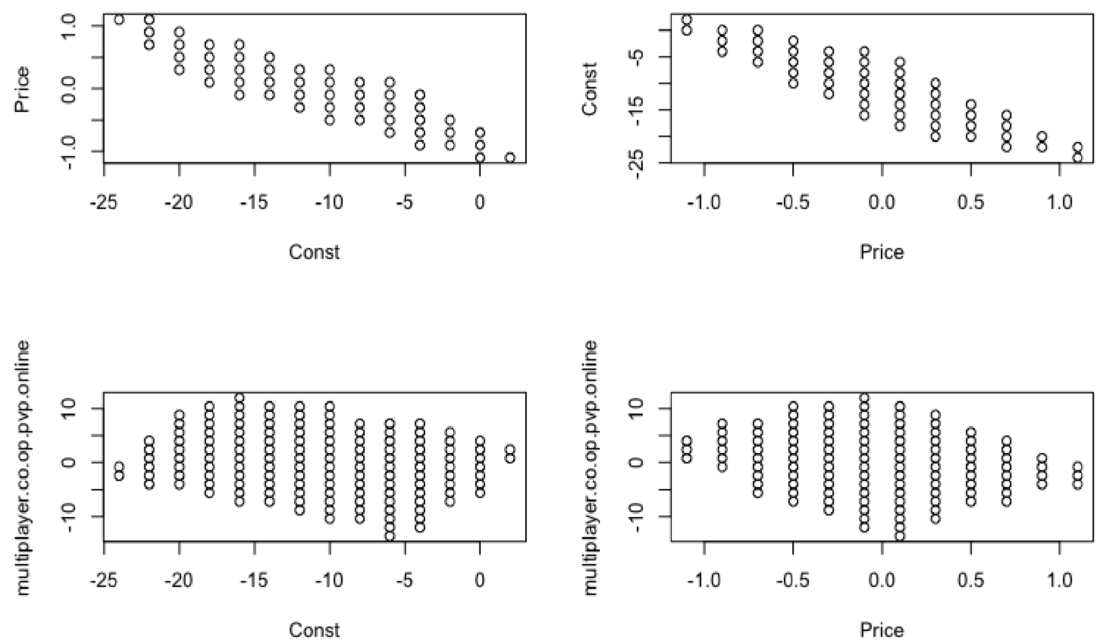


Figure 10: Projected 80% confidence region



References

Andrews, D. W., & Shi, X. (2013). Inference based on conditional moment inequalities. *Econometrica*, *81*(2), 609–666.

Andrews, I., Roth, J., & Pakes, A. (2023). Inference for linear conditional moment inequalities. *Review of Economic Studies*, *90*(6), 2763–2791.

Beresteanu, A., & Molinari, F. (2008). Asymptotic properties for a class of partially identified models. *Econometrica*, *76*(4), 763–814.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.

Berry, S. T., & Haile, P. A. (2014). Identification in differentiated products markets using market level data. *Econometrica*, *82*(5), 1749–1797.

Berry, S. T., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Berry, S. T., Levinsohn, J., & Pakes, A. (2004). Differentiated products demand systems from a combination of micro and macro data: the new car market. *Journal of political Economy*, *112*(1), 68–105.

Bettin, G., & Lucchetti, R. (2012). Interval regression models with endogenous explanatory variables. *Empirical Economics*, *43*, 475–498.

Bontemps, C., Magnac, T., & Maurin, E. (2012). Set identified linear models. *Econometrica*, *80*(3), 1129–1155.

Canay, I. A., Illanes, G., & Velez, A. (2023). A user's guide for inference in models defined by moment inequalities. *Journal of Econometrics*, 105558.

Chernozhukov, V., Chetverikov, D., & Kato, K. (2014). Anti-concentration and honest, adaptive confidence bands.

Chernozhukov, V., Chetverikov, D., & Kato, K. (2019). Inference on causal and structural parameters using many moment inequalities. *The Review of Economic Studies*, *86*(5), 1867–1900.

Chernozhukov, V., Hong, H., & Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models 1. *Econometrica*, *75*(5), 1243–1284.

Chernozhukov, V., Lee, S., & Rosen, A. M. (2013). Intersection bounds: estimation and inference. *Econometrica*, *81*(2), 667–737.

Ciliberto, F., & Tamer, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica*, *77*(6), 1791–1828.

Cox, G., & Shi, X. (2023). Simple adaptive size-exact testing for full-vector and subvector inference in moment inequality models. *The Review of Economic Studies*, *90*(1), 201–228.

Duncan, O. D., & Davis, B. (1953). An alternative to ecological correlation. *American sociological review*, *18*(6).

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon, 3ˆ e serie, Sciences, Sect. A*, *14*, 53–77.

Frisch, R. (1934). Statistical confluence analysis by means of complete regression systems. *(No Title)*.

Gandhi, A., Lu, Z., & Shi, X. (2023). Estimating demand for differentiated products with zeroes in market share data. *Quantitative Economics*, *14*(2), 381–418.

Haile, P. A., & Tamer, E. (2003). Inference with an incomplete model of english auctions. *Journal of Political Economy*, *111*(1), 1–51.

Kaido, H., Molinari, F., & Stoye, J. (2019). Confidence intervals for projections of partially identified parameters. *Econometrica*, *87*(4), 1397–1432.

Manski, C. F., & Tamer, E. (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica*, *70*(2), 519–546.

Marschak, J., & Andrews, W. H. (1944). Random simultaneous equations and the theory of production. *Econometrica, Journal of the Econometric Society*, 143–205.

Molinari, F. (2020). Microeconometrics with partial identification. *Handbook of econometrics*, *7*, 355–486.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, *69*(2), 307–342.

Pakes, A., Porter, J., Ho, K., & Ishii, J. (2015). Moment inequalities and their application. *Econometrica*, *83*(1), 315–334.

Stewart, M. B. (1983). On least squares estimation when the dependent variable is grouped. *The review of economic studies*, *50*(4), 737–753.

Tamer, E. (2010). Partial identification in econometrics. *Annu. Rev. Econ.*, *2*(1), 167–195.

Van den Meersche, K., Soetaert, K., & Van Oevelen, D. (2009). Xsample (): an r function for sampling linear inverse problems. *Journal of Statistical Software*, *30*, Snippet–1.