Supplemental Appendix
Not Intended for Publication

This Supplemental Appendix is organized as follows. Section M.1 contains computational details on the implementation of our test. Section M.2 contains a Monte Carlo experiment examining the performance of our proposed procedure.

M.1 Computational Details

In this appendix, we provide details on how we compute our test statistic, T_n, defined in (11), the restricted estimator $\hat{\beta}_n^{\mathrm{r}}$, defined in (15), and obtain a critical value. One computational theme that we found important in our simulations is that the pseudoinverse $A^\dagger$ can be poorly conditioned. As we show below, however, it is possible to implement our procedure without computing $A^\dagger$ explicitly.

First, we need to select an estimator $\hat{x}_n^\star$. In the mixed logit simulation in Section M.2, the parameter $\beta(P)$ can be decomposed into $\beta(P) = (\beta_{\mathrm{u}}(P)', \beta_{\mathrm{k}}')'$, where $\beta_{\mathrm{u}}(P) \in \mathbf{R}^{p_{\mathrm{u}}}$ and $\beta_{\mathrm{k}} \in \mathbf{R}^{p_{\mathrm{k}}}$ is a known constant for all $P \in \mathbf{P}_0$. Similarly, we decompose any $b \in \mathbf{R}^p$ into $b = (b_{\mathrm{u}}', b_{\mathrm{k}}')'$ with $b_{\mathrm{u}} \in \mathbf{R}^{p_{\mathrm{u}}}$ and $b_{\mathrm{k}} \in \mathbf{R}^{p_{\mathrm{k}}}$, and partition the matrix A into the corresponding submatrices A_{u} (dimension $p_{\mathrm{u}} \times d$) and A_{k} (dimension $p_{\mathrm{k}} \times d$). In our simulations, we then set $\hat{x}_n^\star$ to be a solution to

$$\min_{x \in \mathbf{R}^d} (\hat{\beta}_{\mathrm{u},n} - A_{\mathrm{u}} x)' \Xi^{-1} (\hat{\beta}_{\mathrm{u},n} - A_{\mathrm{u}} x) \text{ s.t. } A_{\mathrm{k}} x = \beta_{\mathrm{k}}, \tag{M.1}$$

where $\hat{\beta}_n = (\hat{\beta}_{\mathrm{u},n}', \beta_{\mathrm{k}}')'$ and Ξ is an estimate of the asymptotic variance matrix of $\hat{\beta}_{\mathrm{u},n}$. While the solution to (M.1) may not be unique, any two minimizers x_1 and x_2 of (M.1) must satisfy $Ax_1 = Ax_2$. Since in our reformulations below $\hat{x}_n^\star$ only enters through $A\hat{x}_n^\star$, the specific choice of minimizer in (M.1) is immaterial.

Throughout, we let Ω^{e} be the sample standard deviation matrix of $\hat{\beta}_n$. Note that, since $\hat{\beta}_n = (\hat{\beta}_{\mathrm{u},n}', \beta_{\mathrm{k}}')'$ and β_{k} is non-stochastic, Ω^{e} has the form

$$\Omega^{\mathrm{e}} = \begin{bmatrix} \Xi^{1/2} & 0 \\ 0 & 0 \end{bmatrix}. \tag{M.2}$$

We further let Ω^{i} be the sample standard deviation of $A\hat{x}_n^\star$, although this choice of studentization plays no special computational role in what follows.

Consider the first component of T_n (see (11)), which we reproduce here as

$$T_n^{\mathrm{e}} \equiv \sup_{s \in \mathcal{V}^{\mathrm{e}}} \sqrt{n} \langle s, \hat{\beta}_n - A\hat{x}_n^\star \rangle \text{ where } \mathcal{V}^{\mathrm{e}} \equiv \{s \in \mathbf{R}^p : \|\Omega^{\mathrm{e}} s\|_1 \leq 1\}. \tag{M.3}$$

As in the main text, the superscript "e" alludes to the relation to the "equality" condition in Theorem 3.1. As noted in the main text, $\hat{\beta}_n = A\hat{x}_n^\star$ and hence $T_n^{\mathrm{e}} = 0$ whenever A is full rank and $d \geq p$. In other cases, we use the fact that $\hat{x}_n^\star$, as the solution to (M.1), must satisfy $A_{\mathrm{k}}\hat{x}_n^\star = \beta_{\mathrm{k}}$, and that our choice of Ω^{e} in (M.2) has $\Xi^{1/2}$ as its upper left block. From these observations, we deduce that

$$\begin{aligned} T_n^{\mathrm{e}} &= \sup_{s_{\mathrm{u}} \in \mathbf{R}^{p_{\mathrm{u}}}} \sqrt{n} \langle s_{\mathrm{u}}, \hat{\beta}_{\mathrm{u},n} - A_{\mathrm{u}}\hat{x}_n^\star \rangle \text{ s.t. } \|\Xi^{1/2} s_{\mathrm{u}}\|_1 \leq 1 \\ &= \|\sqrt{n}\Xi^{-1/2}(\hat{\beta}_{\mathrm{u},n} - A_{\mathrm{u}}\hat{x}_n^\star)\|_\infty. \end{aligned} \tag{M.4}$$

Thus, T_n^{e} can be computed by taking the maximum of a vector of length p_{u}.

The second component of T_n, defined in (11), is reproduced here as

$$T_n^{\mathrm{i}} \equiv \sup_{s \in \mathcal{V}^{\mathrm{i}}} \sqrt{n} \langle A^\dagger s, \hat{x}_n^\star \rangle \text{ where } \mathcal{V}^{\mathrm{i}} \equiv \{s \in \mathbf{R}^p : A^\dagger s \leq 0 \text{ and } \|\Omega^{\mathrm{i}}(AA')^\dagger s\|_1 \leq 1\}, \tag{M.5}$$

and the superscript "i" alludes to the relation to the "inequality" condition in Theorem 3.1. To compute T_n^{i} without explicitly using $A^\dagger$, we first note

$$A^\dagger = A'(AA')^\dagger, \tag{M.6}$$

see, e.g., Proposition 6.11.1(9) in Luenberger (1969). Then, we observe that

$$\mathrm{range}\{(AA')^\dagger\} = \mathrm{null}\{AA'\}^\perp = \mathrm{range}\{AA'\} = \mathrm{range}\{A\}, \tag{M.7}$$

where the first equality is a property of pseudoinverses, see Luenberger (1969, pg. 164). The second equality is a standard result in linear algebra, see Theorem 6.6.1 in Luenberger (1969). This result is also used in the third equality, which uses the following logic: if $t = As$ for some $s \in \mathbf{R}^p$, then also $t = As_1$, where $s_1 \in \mathrm{null}\{A\}^\perp = \mathrm{range}\{A'\}$ is determined from the orthogonal decomposition $s = s_0 + s_1$ with $s_0 \in \mathrm{null}\{A\}$, and hence $t \in \mathrm{range}\{AA'\}$ implying $\mathrm{range}\{A\} \subseteq \mathrm{range}\{AA'\}$. Since $\mathrm{range}\{AA'\} \subseteq \mathrm{range}\{A\}$ the third equality follows. Thus,

$$T_n^{\mathrm{i}} = \sup_{s \in \mathbf{R}^p} \sqrt{n} \langle A'(AA')^\dagger s, \hat{x}_n^\star \rangle \text{ s.t. } A'(AA')^\dagger s \leq 0 \text{ and } \|\Omega^{\mathrm{i}}(AA')^\dagger s\|_1 \leq 1,$$

$$= \sup_{x \in \mathbf{R}^d} \sqrt{n} \langle A'Ax, \hat{x}_n^\star \rangle \text{ s.t. } A'Ax \le 0 \text{ and } \|\Omega^{\mathrm{i}} Ax\|_1 \le 1,$$

$$= \sup_{x \in \mathbf{R}^d, s \in \mathbf{R}^p} \sqrt{n} \langle s, A\hat{x}_n^\star \rangle \text{ s.t. } Ax = s,\ A's \le 0 \text{ and } \|\Omega^{\mathrm{i}} s\|_1 \le 1, \qquad \text{(M.8)}$$

where the first equality follows from (M.6), the second from (M.7), and in the third we substituted $s = Ax$. The final program in (M.8) can be written explicitly as a linear program by introducing non-negative slack variables, so that

$$T_n^{\mathrm{i}} = \sup_{x \in \mathbf{R}^d, s \in \mathbf{R}^p, \phi^+ \in \mathbf{R}_+^p, \phi^- \in \mathbf{R}_+^p} \sqrt{n} \langle s, A\hat{x}_n^\star \rangle$$
$$\text{s.t. } Ax = s,\ A's \le 0,\ \langle \mathbf{1}_p, \phi^+ \rangle + \langle \mathbf{1}_p, \phi^- \rangle \le 1,\ \phi^+ - \phi^- = \Omega^{\mathrm{i}} s, \quad \text{(M.9)}$$

where $\mathbf{1}_p \in \mathbf{R}^p$ is the vector with all coordinates equal to one. Note that if $d \ge p$ and A has full rank, then the constraint $Ax = s$ is redundant since Ax ranges across all of $\mathbf{R}^p$ as x varies across $\mathbf{R}^d$. In these cases, the constraint $Ax = s$ together with the variable x can be entirely removed from the linear program in (M.9). Taking the maximum of (M.4) and (M.9) yields our test statistic T_n.

Turning to our bootstrap procedure, we first show how to solve (15) to find $\hat{\beta}_n^{\mathrm{r}}$. The optimization problem to solve is here reproduced as:

$$\min_{\tilde{x} \in \mathbf{R}_+^d, b = (b_{\mathrm{u}}', b_{\mathrm{k}}')'} \left\{ \sup_{s \in \mathcal{V}^{\mathrm{i}}} |\langle A^\dagger s, \hat{x}_n^\star - A^\dagger b \rangle| \right\} \text{ s.t. } b_{\mathrm{k}} = \beta_{\mathrm{k}},\ A\tilde{x} = b. \qquad \text{(M.10)}$$

With probability tending to one, the inner problem is finite when evaluated at $b = \beta(P)$, and hence we may restrict attention to b for which the inner problem is finite. Moreover, the inner problem has the same structure as (M.5), but with $\hat{x}_n^\star$ replaced by $\hat{x}_n^\star - A^\dagger b$. Hence, applying the same logic employed in (M.8) allows us to rewrite the inner problem in (M.10) as being equal to

$$\sup_{x \in \mathbf{R}^d} |\langle A'Ax, \hat{x}_n^\star - A^\dagger b \rangle| \text{ s.t. } A'Ax \le 0 \text{ and } \|\Omega^{\mathrm{i}} Ax\|_1 \le 1. \qquad \text{(M.11)}$$

It is in turn possible to establish that the optimization problem in (M.11) equals

$$\sup_{x \in \mathbf{R}^d} \langle A'Ax, \hat{x}_n^\star - A^\dagger b \rangle \text{ s.t. } x \in \mathrm{co}\{v \in \mathbf{R}^d : A'Av \le 0,\ \|\Omega^{\mathrm{i}} Av\|_1 \le 1\}, \qquad \text{(M.12)}$$

where $\mathrm{co}\{\cdot\}$ denotes the convex hull of a set. By introducing slack variables as in

(M.9), we may rewrite (M.12) explicitly as a linear program

$$\sup_{v_1,v_2\in\mathbf{R}^d,\phi_1^+,\phi_1^-,\phi_2^+,\phi_2^-\in\mathbf{R}^p_+,a_1,a_2\in\mathbf{R}_+} \langle A'A(v_1+v_2),\hat{x}_n^\star - A^\dagger b\rangle$$
$$\text{s.t.}\quad A'Av_1\le 0,\ -A'Av_2\le 0,\ \langle\mathbf{1}_p,\phi_1^+\rangle+\langle\mathbf{1}_p,\phi_1^-\rangle\le a_1,\ \langle\mathbf{1}_p,\phi_2^+\rangle+\langle\mathbf{1}_p,\phi_2^-\rangle\le a_2,$$
$$\phi_1^+-\phi_1^- = \Omega^{\mathrm{i}}Av_1,\ \phi_2^+-\phi_2^-=\Omega^{\mathrm{i}}Av_2,\ a_1+a_2=1. \tag{M.13}$$

In turn, the dual of the linear program in (M.13) can be shown to be equal to

$$\inf_{\phi_0\in\mathbf{R},\phi_1^s,\phi_2^s\in\mathbf{R}^d_+,\phi_1^n,\phi_2^n\in\mathbf{R}_+,\phi_1^e,\phi_2^e\in\mathbf{R}^p} \phi_0 \text{ s.t. } A'A\phi_1^s - A'\Omega^{\mathrm{i}}\phi_1^e = A'A(\hat{x}_n^\star - A^\dagger b),$$
$$-A'A\phi_2^s - A'\Omega^{\mathrm{i}}\phi_2^e = A'A(\hat{x}_n^\star - A^\dagger b),\ \phi_1^n\mathbf{1}_p+\phi_1^e\ge 0,\ \phi_1^n\mathbf{1}_p-\phi_1^e\ge 0,$$
$$\phi_2^n\mathbf{1}_p+\phi_2^e\ge 0,\ \phi_2^n\mathbf{1}_p-\phi_2^e\ge 0,\ -\phi_1^n+\phi_0\ge 0,\ -\phi_2^n+\phi_0\ge 0. \tag{M.14}$$

Let $V\equiv\text{range}\{AA'\}$ and $\Pi_V(b)$ denote the $\|\cdot\|_2$-projection of b onto V. Then note that $A^\dagger = A'(AA')^\dagger$ (see Proposition 6.11.1(8) in Luenberger (1969)) implies $A'AA^\dagger b = A'AA'(AA')^\dagger b = A'\Pi_V b$. However, by result (M.7), $V\equiv\text{range}\{AA'\} = \text{range}\{A\} = \text{null}\{A'\}^\perp$, where the final equality follows by Theorem 6.6.1 in Luenberger (1969). Hence, $A'\Pi_V b = A'b$ and (M.14) equals

$$\inf_{\phi_0\in\mathbf{R},\phi_1^s,\phi_2^s\in\mathbf{R}^d_+,\phi_1^n,\phi_2^n\in\mathbf{R}_+,\phi_1^e,\phi_2^e\in\mathbf{R}^p} \phi_0 \text{ s.t. } A'A\phi_1^s - A'\Omega^{\mathrm{i}}\phi_1^e = A'(A\hat{x}_n^\star - b),$$
$$-A'A\phi_2^s - A'\Omega^{\mathrm{i}}\phi_2^e = A'(A\hat{x}_n^\star - b),\ \phi_1^n\mathbf{1}_p+\phi_1^e\ge 0,\ \phi_1^n\mathbf{1}_p-\phi_1^e\ge 0,$$
$$\phi_2^n\mathbf{1}_p+\phi_2^e\ge 0,\ \phi_2^n\mathbf{1}_p-\phi_2^e\ge 0,\ -\phi_1^n+\phi_0\ge 0,\ -\phi_2^n+\phi_0\ge 0. \tag{M.15}$$

Substituting (M.15) back into the inner problem in (M.10) then yields a single linear program that determines $\hat{\beta}_n^{\mathrm{r}}$. Given $\hat{\beta}_n^{\mathrm{r}}$ it is then straightforward to compute our bootstrap statistic. For instance, in the simulations in Section M.2, we let

$$\hat{\mathbb{G}}_n^{\mathrm{e}} = \sqrt{n}\{(\hat{\beta}_{b,n} - A\hat{x}_{b,n}^\star) - (\hat{\beta}_n - A\hat{x}_n^\star)\} \qquad \hat{\mathbb{G}}_n^{\mathrm{i}} = \sqrt{n}A(\hat{x}_{b,n}^\star - \hat{x}_n^\star)$$

where $\hat{\beta}_{b,n}$ and $\hat{x}_{b,n}^\star$ are nonparametric bootstrap analogues to $\hat{\beta}_n$ and $\hat{x}_n^\star$. Arguing as in result (M.4) it is then straightforward to show that

$$\sup_{s\in\mathcal{V}^{\mathrm{e}}}\langle s,\hat{\mathbb{G}}_n^{\mathrm{e}}\rangle = \|\sqrt{n}\Xi^{-1/2}\hat{\mathbb{G}}_n^{\mathrm{e}}\|_\infty. \tag{M.16}$$

In analogy to (M.4), (M.16) equals zero whenever A is full rank and $d\ge p$. Next, we may employ the same arguments as in (M.8) and (M.9) and note $AA^\dagger\hat{\mathbb{G}}_n^{\mathrm{i}} = \hat{\mathbb{G}}_n^{\mathrm{i}}$

and $AA^{\dagger}\hat{\beta}_n^{\mathrm{r}} = \hat{\beta}_n^{\mathrm{r}}$ because $\hat{\mathbb{G}}_n^{\mathrm{i}}$ and $\hat{\beta}_n^{\mathrm{r}}$ are on the range of A to obtain

$$\begin{aligned}
&\sup_{s\in\mathcal{V}^{\mathrm{i}}}\langle A^{\dagger}s, A^{\dagger}(\hat{\mathbb{G}}_n^{\mathrm{i}} + \sqrt{n}\lambda_n\hat{\beta}_n^{\mathrm{r}})\rangle \\
&\qquad = \sup_{x\in\mathbf{R}^d, s\in\mathbf{R}^p, \phi^+\in\mathbf{R}_+^p, \phi^-\in\mathbf{R}_+^p} \langle s, \hat{\mathbb{G}}_n^{\mathrm{i}} + \sqrt{n}\lambda_n\hat{\beta}_n^{\mathrm{r}}\rangle \\
&\qquad\quad \text{s.t. } Ax = s,\ A's \le 0,\ \langle \mathbf{1}_p, \phi^+\rangle + \langle \mathbf{1}_p, \phi^-\rangle \le 1,\ \phi^+ - \phi^- = \Omega^{\mathrm{i}}s. \qquad \text{(M.17)}
\end{aligned}$$

As in (M.9), we note that if A is full rank and $d \ge p$, then the constraint $Ax = s$ and the variable x may be dropped from (M.17). The critical value is then obtained by computing the $1-\alpha$ quantile of the maximum of (M.16) and (M.17) across bootstrap iterations. Finally, we note that the problem (M.23) used to determine λ_n^{b} is equivalent to (M.9) with $A\hat{x}_n^{\star}$ replaced by $\hat{\mathbb{G}}_n^{\mathrm{i}}$.

M.2 Simulations with a Mixed Logit Model

M.2.1 The Model

Example 2.1 is an example of a class of mixture models considered by Fox et al. (2011). A simpler example with the same structure is a static, binary choice logit with random coefficients. In this model, a consumer chooses $Y \in \{0,1\}$ by

$$Y = 1\{C_0 + C_1W - U \ge 0\},$$

where W is an observed variable which we will think of as the price of buying a good ($Y=1$), and $V \equiv (C_0, C_1)$ and U are latent variables. The unobservable U is assumed to follow a standard logistic distribution, independently of (V, W).

A consumer of type $v = (c_0, c_1)$ facing price w buys the good with probability

$$P(Y = 1|W = w, V = v) = \frac{1}{1+\exp(-c_0 - c_1w)} \equiv \ell(w, v). \qquad \text{(M.18)}$$

Bajari et al. (2007) and Fox et al. (2011) assume V is independent of W and approximate the distribution of V using a discrete distribution with known support points $(v_1, \ldots, v_d)$ and unknown respective probabilities $x \equiv (x_1, \ldots, x_d)$. Under these assumptions, (M.18) can be aggregated into a conditional moment equality:

$$P(Y=1|W=w)=\sum_{j=1}^{d} x_j \ell(w,v_j). \tag{M.19}$$

A natural quantity of interest in this model is the price elasticity of purchase probability. For a consumer of type $v=(c_0,c_1)$ facing price $\bar{w}$, this is

$$\epsilon(v,\bar{w}) \equiv \frac{\partial}{\partial w}\ell(v,w)\Big|_{w=\bar{w}} \times \frac{\bar{w}}{\ell(\bar{w},v)} = c_1\bar{w}(1-\ell(\bar{w},v)).$$

The cumulative distribution function (c.d.f.) of this elasticity is

$$F_\epsilon(t|\bar{w}) \equiv P(\epsilon(V,\bar{w})\le t) = \sum_{j=1}^{d} 1\{\epsilon(v_j,\bar{w})\le t\}x_j \equiv a(t,\bar{w})'x, \tag{M.20}$$

where $a(t,\bar{w}) \equiv (a_1(t,\bar{w}),\ldots,a_d(t,\bar{w}))'$ with $a_j(t,\bar{w}) \equiv 1\{\epsilon(v_j,\bar{w})\le t\}$. We take the c.d.f. $F_\epsilon(\cdot|\bar{w})$ as our parameter of interest in the discussion ahead.

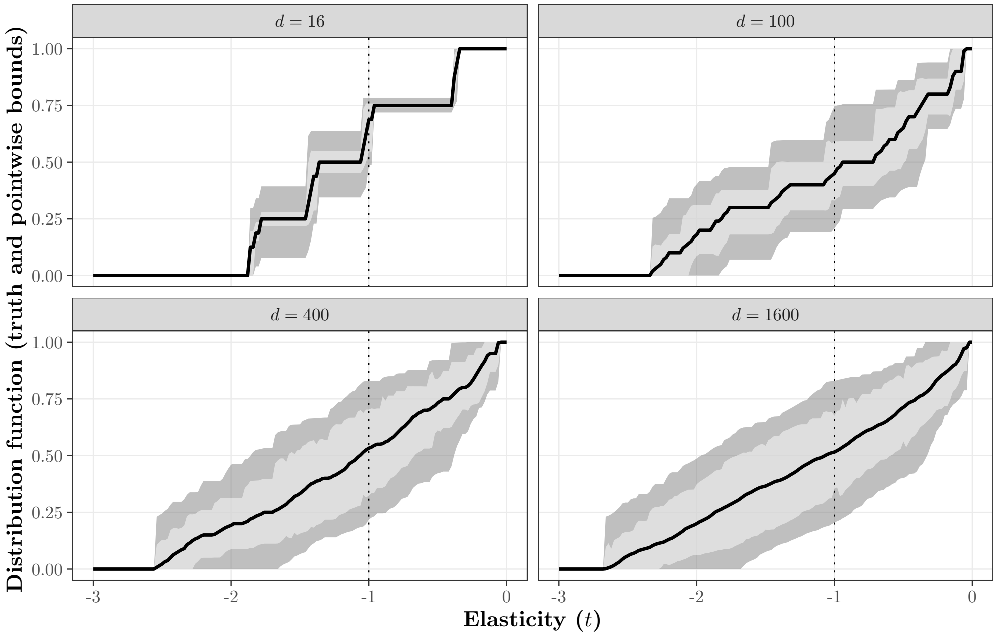
M.2.2 Data Generating Processes

In our simulations we generate data from a class of mixed logit models parameterized as follows. The distribution of W is uniform over $p-2$ evenly spaced points between 0 and 2, inclusive. The known support of C_0 is generated by taking a Sobol sequence of length $\sqrt{d}$ and rescaling it to lie in $[.5, 1.0]$. Similarly, the support of C_1 is a Sobol sequence of length $\sqrt{d}$ rescaled to $[-3, 0]$. The distribution of $V \equiv (C_0, C_1)$ is taken to be uniform over the product of the two marginal supports, so that it has d support points.

Fox et al. (2012) provide identification results that apply to the binary mixed logit model. However, their conditions require W to be continuously distributed. When W is discretely distributed, one might expect the distributions of V and thus of $\epsilon(V,\bar{w})$ are only partially identified. We explore this conjecture computationally. We denote the identified set for the distribution of V as

$$\mathbb{X}^\star(P) \equiv \{x\in\mathbf{R}^d_+ : \sum_{j=1}^{d} x_j = 1,\ \sum_{j=1}^{d} x_j\ell(w,v_j) = P(Y=1|W=w) \text{ for all } w\in\mathcal{W}\},$$

Figure 1: Bounds on the distribution of price elasticity $F_{\epsilon}(t|1)$



These plots are based on the data generating processes described in Section M.2.2. The solid black line is the actual value of $F_{\varepsilon}(t|1)$. The lighter color is $\mathbb{A}^{\star}(t, 1|P)$ when the support of W has sixteen points. The darker color is the same set when the support of W has only four points. The dotted vertical is the value $t = -1$ used in the Monte Carlo simulations in Section M.2.4.

where $\mathcal{W}$ is the support of W. In addition, for any $t \in \mathbf{R}$, we denote the identified set for $F_{\epsilon}(t|\bar{w})$ by $\mathbb{A}^{\star}(t, \bar{w}|P)$, which simply equals the projection of $\mathbb{X}^{\star}(P)$ under the linear map introduced in (M.20):

$$\mathbb{A}^{\star}(t, \bar{w}|P) \equiv \{a(t, \bar{w})'x : x \in \mathbb{X}^{\star}(P)\}.$$

Since $\mathbb{X}^{\star}(P)$ is a system of linear equalities and inequalities, and $x \mapsto a(t, \bar{w})'x$ is scalar-valued and linear, $\mathbb{A}^{\star}(t, \bar{w}|P)$ is a closed interval (see, e.g. Mogstad et al., 2018, for a similar argument). The left endpoint of this interval is given by

$$\min_{x \in \mathbf{R}_{+}^{d}} a(t, \bar{w})'x \text{ s.t. } \sum_{j=1}^{d} x_j = 1, \quad \sum_{j=1}^{d} x_j \ell(w, v_j) = P(Y = 1|W = w) \text{ for all } w \in \mathcal{W}, \tag{M.21}$$

and the right endpoint is equal to its maximization counterpart.

Figure 1 depicts $\mathbb{A}^{\star}(t, \bar{w}|P)$ as a function of t for $\bar{w} = 1$. The outer and inner bands depict the identified set when the support of W has four and sixteen points, respectively, while the solid line indicates the distribution under the actual data generating process. The identified sets are non-trivial and widen with the number of support points d for the unobservable V. For $d = 16$, the bounds when W has sixteen support points are narrow, but numerically distinct from a point. This is because the system of moment equations that defines $\mathbb{X}^\star(P)$, while known to be nonsingular in principle, is sufficiently close to singular to matter numerically.

M.2.3 Test Implementation

As in Example 2.1, we may use our results to test whether a hypothesized $\gamma \in \mathbf{R}$ belongs to the identified set for $F_\epsilon(t|\bar{w})$. Using (M.19) and recalling W was set to have $p-2$ support points, we may then map such hypothesis into (1) by setting

$$\beta(P) = \begin{pmatrix} P(Y=1|W=w_1) \\ \vdots \\ P(Y=1|W=w_{p-2}) \\ 1 \\ \gamma \end{pmatrix} \quad A = \begin{pmatrix} \ell(w_1, v_1) & \cdots & \ell(w_1, v_d) \\ \vdots & \vdots & \vdots \\ \ell(w_{p-2}, v_1) & \cdots & \ell(w_{p-2}, v_d) \\ 1 & \cdots & 1 \\ a_1(t, \bar{w}) & \cdots & a_d(t, \bar{w}) \end{pmatrix}.$$

We take $\hat{\beta}_n \equiv (\hat{\beta}_{\mathrm{u},n}, 1, \gamma)' \in \mathbf{R}^p$, where $\hat{\beta}_{\mathrm{u},n}$ is the sample analogue to the first $p-2$ components of $\beta(P)$. We set $\hat{x}_n^\star = A^\dagger \hat{\beta}_n$ for designs with $d \geq p$, and let

$$\hat{x}_n^\star \equiv \arg\min_{x \in \mathbf{R}^d} (\hat{\beta}_{\mathrm{u},n} - A_{\mathrm{u}} x)' \Xi^{-1} (\hat{\beta}_{\mathrm{u},n} - A_{\mathrm{u}} x) \quad \text{s.t.} \quad \sum_{j=1}^d x = 1 \quad \text{and} \quad a(t, \bar{w})' x = \gamma,$$

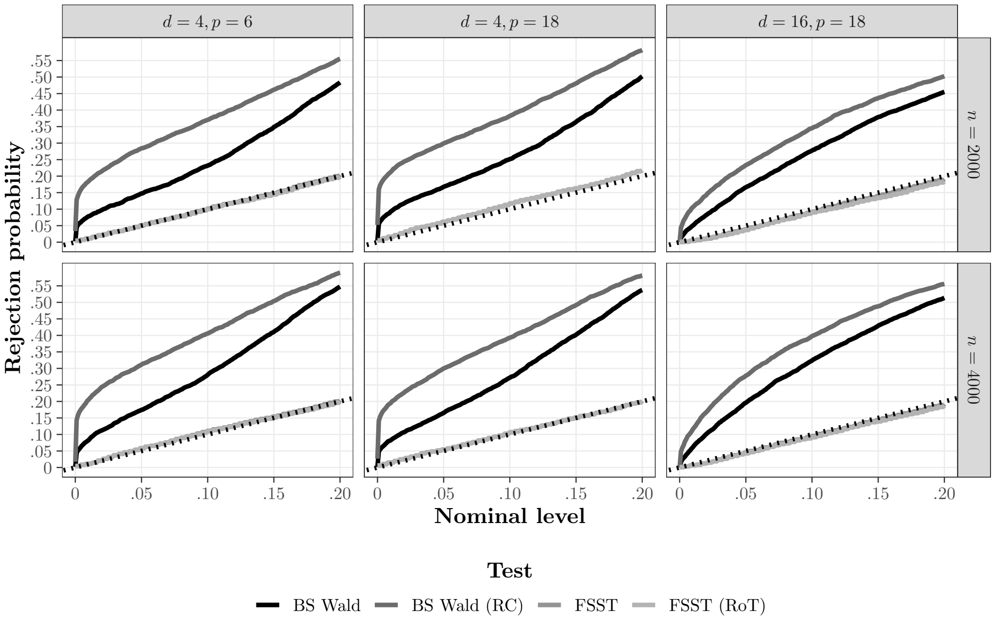
when $d < p$, where A_{u} corresponds to the first $p-2$ rows of A and Ξ is the sample analogue estimator of asymptotic variance matrix of $\hat{\beta}_{\mathrm{u},n}$. We let Ω^{e} be the sample standard deviation matrix of $\hat{\beta}_n$, and Ω^{i} be the sample standard deviation of $\sqrt{n} A \hat{x}_n^\star$ computed from 250 draws of the nonparametric bootstrap.

We explore two rules for selecting λ_n. To motivate them, we note that an important theoretical restriction on λ_n is that, uniformly in $P \in \mathbf{P}_0$, it satisfy

$$\lambda_n \sqrt{n} \sup_{s \in \mathcal{V}^{\mathrm{i}}} \langle A^\dagger s, A^\dagger A(\hat{x}_n^\star - x^\star(P)) \rangle = o_P(1); \tag{M.22}$$

see Lemma A.1. Employing our coupling $\sqrt{n} A(\hat{x}_n^\star - x^\star(P)) \approx \mathbb{G}_n^{\mathrm{i}}(P)$ and Ω^{i} being an estiamte of the standard deviation matrix of $\mathbb{G}_n^{\mathrm{i}}(P)$ suggests selecting λ_n to satisfy $\lambda_n \sqrt{\log(e \vee p)} = o(1)$ – here $a \vee b \equiv \max\{a, b\}$. For a concrete choice of λ_n, we rely on the law of iterated logarithm and let $\lambda_n^{\mathrm{r}} = 1/\sqrt{\log(e \vee p) \log(e \vee \log(e \vee n))}$.

Figure 2: Null rejection probabilities for (nearly) point-identified designs



The dotted line is the 45 degree line. "FSST" refers to the test developed in this paper with λ_n^{b}, whereas "FSST (RoT)" uses the rule of thumb choice λ_n^{r}. "BS Wald" corresponds to a Wald test using bootstrap estimates of the standard errors. "BS Wald (RC)" is the same procedure but with standard errors based on bootstrapping with a re-centered GMM criterion. The null hypothesis is that $F_\epsilon(-1|1)$ is equal to its true value. In the case of $d = 16, p = 18$, which is set identified but with a very narrow identified set, we test the null hypothesis that $F_\epsilon(-1|1)$ is equal to the lower bound of the identified set.

As an alternative to λ_n^{r}, we employ the bootstrap to approximate the law of (M.22). In particular, for some $\delta_n \downarrow 0$ we let $\lambda_n^{\mathrm{b}} \equiv \min\{1, 1/\hat{\tau}_n(1-\delta_n)\}$ where $\hat{\tau}_n(1-\delta_n)$ denotes the $1-\delta_n$ quantile of

$$\sup_{s \in \mathcal{V}^{\mathrm{i}}} \langle A^\dagger s, A^\dagger \hat{\mathbb{G}}_n^{\mathrm{i}} \rangle \tag{M.23}$$

conditional on the data. For concreteness we let $\delta_n = 1/\sqrt{\log(e \vee \log(e \vee n))}$.

In Appendix M.1, we describe the computation of our test in more detail. In particular, we show how to reformulate all optimization problems into linear programming problems that do not require explicitly computing $A^\dagger$. An R package for implementing our test is available at https://github.com/conroylau/lpinfer.

Table 1: Null rejection probabilities for a nominal 0.05 test

		(a) Results for λ_n^{b}							(b) Results for λ_n^{r}						
		d							d						
n	p	100	400	1600	4900	100^2	225^2	317^2	100	400	1600	4900	100^2	225^2	317^2
1000	6	.036	.034	.034	.037	.038	.036	.036	.020	.019	.021	.021	.022	.019	.021
	18	.040	.035	.036	.041	.039	.038	.036	.037	.029	.029	.033	.033	.031	.030
2000	6	.042	.042	.049	.046	.047	.052	.061	.030	.025	.033	.032	.033	.027	.039
	18	.031	.028	.032	.032	.030	.030	.028	.023	.021	.028	.027	.025	.027	.020
	38	.053	.046	.051	.052	.052	.067	.053	.048	.039	.043	.045	.047	.062	.046
4000	6	.045	.048	.049	.054	.058	.051	.065	.034	.034	.038	.042	.046	.035	.058
	18	.028	.031	.029	.028	.030	.038	.035	.023	.026	.024	.022	.025	.032	.028
	38	.031	.034	.039	.036	.040	.035	.037	.026	.029	.033	.032	.035	.032	.033
	51	.042	.051	.051	.040	.047	.047	.030	.038	.044	.045	.034	.042	.041	.027
8000	6	.049	.055	.056	.048	.054	.055	.073	.040	.046	.048	.040	.046	.050	.061
	18	.034	.035	.036	.030	.032	.040	.041	.028	.028	.032	.025	.027	.032	.034
	38	.033	.035	.035	.037	.037	.025	.047	.027	.029	.030	.032	.032	.021	.043
	51	.034	.043	.035	.040	.037	.035	.038	.029	.036	.028	.034	.033	.030	.031
	83	.043	.042	.050	.048	.042	.054	.046	.038	.035	.046	.041	.034	.048	.042

The null hypothesis is that $F_\epsilon(-1|1)$ is equal to the lower bound of the population identified set.

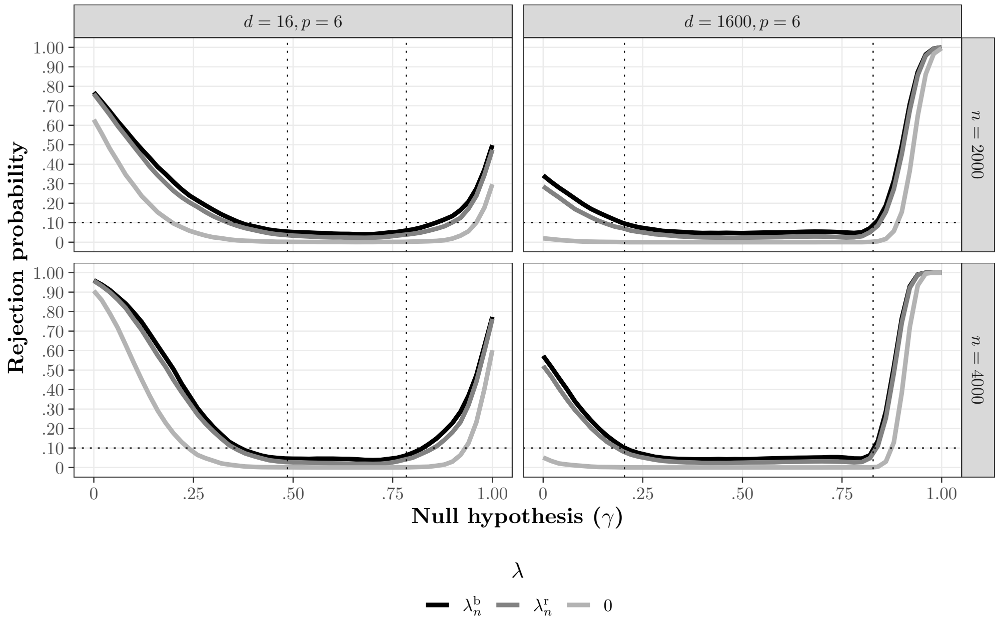
M.2.4 Monte Carlo Simulations

We start by examining the null rejection probabilities of our testing procedure by setting γ to be the lower bound of the population identified set computed via (M.21) with $t = -1$ and $\bar{w} = 1$. In unreported simulations we found setting γ to be the upper bound of the identified set yielded similar results. We consider sample sizes of $n = 1000$, 2000, 4000, and 8000 for each of the data generating processes discussed in Section M.2.2. Results with $d \leq 10000$ are based on 5000 Monte Carlo replications and 250 nonparametric bootstrap draws. When $d > 10000$, we use 1000 Monte Carlo replications.

We first consider the designs in which $p - 2 \geq d$ so that $F_\epsilon(-1|1)$ is (nearly) point identified. In this case, one might alternatively consider estimating probability weights x_0 satisfying the moment restrictions in (M.19) by constrained GMM, and then conducting inference on $F_\epsilon(-1|1)$ using a bootstrapped Wald test. For example, this is the approach that appears to have been taken by Nevo et al. (2016) in the related setting discussed in Example 2.1. However, the non-negativity constraints on x_0 imply that the bootstrap will generally not be consistent in this case (Fang and Santos, 2018).

We demonstrate this point in Figure 2 with plots of the actual and nominal level for both our (FSST) and for the bootstrapped Wald test based on constrained

Figure 3: Power curves for FSST nominal 0.10 test



The vertical dotted lines indicate the lower and upper bounds of the population identified set. The horizontal dotted line indicates the nominal level (0.10).

GMM. The latter exhibits large size distortions. For example the GMM test with nominal level 5% rejects in over 15% of draws $d = 16, p = 18$ and $n = 2{,}000$, and a nominal level 10% test rejects in over 25% of draws when $d = 4, p = 18$, and $n = 4{,}000$. Re-centering the GMM criterion before conducting this test (e.g. Hall and Horowitz, 1996) leads to even greater over-rejection. In contrast, FSST has nearly equal nominal and actual levels across the examined designs.

In Table 1, we report empirical rejection rates for our procedure using partially identified designs that range in size from relatively small ($d = 100, p = 6$) to enormous ($p = 83, d = 317^2 \approx 10^5$). We note that in this application, p/n should be small because otherwise we will draw samples (or bootstrap samples) that do not contain all the support points of W. Reflecting this constraint, in Table 1 we let p grow with n but keep the largest values of p/n at approximately .01. No such restriction is imposed on d and we consider designs in which d far exceeds n (e.g., with d/n as large as 100). Across all different data generating processes and sample sizes, even in the largest models, we find the null rejection probabilities remain approximately no greater than the nominal level.

Comparing panels (a) and (b) of Table 1, we see that the occasional (and mild) over-rejections can be controlled by using λ_n^{r} instead of λ_n^{b}. Figure 3 illustrates the impact that the choice of λ_n has on power for two of the smaller designs. Both λ_n^{b} and λ_n^{r} provide considerable power gains over the conservative choice of $\lambda_n = 0$.

References

Bajari, P., Fox, J. T. and Ryan, S. P. (2007). Linear Regression Estimation of Discrete Choice Models with Nonparametric Distributions of Random Coefficients. *American Economic Review*, **97** 459–463.

Fang, Z. and Santos, A. (2018). Inference on directionally differentiable functions. *The Review of Economic Studies*, **86** 377–412.

Fox, J. T., il Kim, K., Ryan, S. P. and Bajari, P. (2012). The random coefficients logit model is identified. *Journal of Econometrics*, **166** 204–212.

Fox, J. T., Kim, K. I., Ryan, S. P. and Bajari, P. (2011). A simple estimator for the distribution of random coefficients. *Quantitative Economics*, **2** 381–418.

Hall, P. and Horowitz, J. L. (1996). Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators. *Econometrica*, **64** 891–916.

Luenberger, D. G. (1969). *Optimization by Vector Space Methods*. Wiley, NY.

Mogstad, M., Santos, A. and Torgovitsky, A. (2018). Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters. *Econometrica*, **86** 1589–1619.

Nevo, A., Turner, J. L. and Williams, J. W. (2016). Usage-based pricing and demand for residential broadband. *Econometrica*, **84** 411–443.