Measuring the Sensitivity of Parameter Estimates to Estimation Moments

Isaiah Andrews	Matthew Gentzkow	Jesse M. Shapiro
MIT and NBER	*Stanford and NBER*	*Brown and NBER*

March 2017

Online Appendix

Contents

List of Tables

List of Figures

1 Sample Sensitivity with Perturbed Weight Matrix

Our discussion of sample sensitivity in section 5 of the main text assumes that the perturbation parameter μ affects the moments $\hat{g}(\theta,\mu)$ but not the weight matrix. This section provides a result for the more general case where μ also enters the weight matrix (for example through a first-stage estimator), which nests the result stated in equation (5) in the main text.

Define a family of perturbed moments

$$\hat{g}(\theta,\mu) = \hat{g}(\theta) + \mu \cdot \eta,$$

where the scalar μ and vector η control the magnitude and direction of the perturbation, respectively. Note that the perturbations considered in the main text are of this form, with $\eta = -g(a)$. Correspondingly, define a perturbed weight matrix $\hat{W}(\mu)$ with $\hat{W}(0) = \hat{W}$. Define the resulting estimator $\hat{\theta}(\mu)$ to solve

$$\min_{\theta\in\Theta} \hat{g}(\theta,\mu)' \hat{W}(\mu)\, \hat{g}(\theta,\mu). \qquad \text{(OA1)}$$

We assume that $\hat{g}(\theta)$ is twice continuously differentiable in θ, and that $\hat{W}(\mu)$ is differentiable on a ball $\mathscr{B}_\mu$ around zero. If we define

$$\hat{A} = \left[\begin{array}{ccc} \left(\frac{\partial}{\partial\theta_1}\hat{G}(\hat{\theta})'\right)\hat{W}\hat{g}(\hat{\theta}) & \ldots & \left(\frac{\partial}{\partial\theta_P}\hat{G}(\hat{\theta})'\right)\hat{W}\hat{g}(\hat{\theta}) \end{array}\right],$$

and

$$\hat{B} = \hat{G}(\hat{\theta})' \frac{\partial}{\partial\mu}\hat{W}(0)\, \hat{g}(\hat{\theta}),$$

then we obtain the following result.

Online Appendix Proposition 1. *Provided $\hat{\theta}$ lies in the interior of Θ and is the unique solution to (1) in the main text,*

$$\frac{\partial}{\partial\mu}\hat{\theta}(0) = -\left(\hat{G}(\hat{\theta})'\hat{W}\hat{G}(\hat{\theta}) + \hat{A}\right)^{-1}\left(\hat{G}(\hat{\theta})'\hat{W}\frac{\partial}{\partial\mu}\hat{g}(\hat{\theta},0) + \hat{B}\right),$$

whenever $\hat{G}\left(\hat{\theta}\right)'\hat{W}\hat{G}\left(\hat{\theta}\right)+\hat{A}$ *is non-singular.*

Proof. We know that $\hat{g}(\theta,\mu)\to\hat{g}(\theta)$ uniformly in θ as $\mu\to 0$. Thus $\hat{g}(\theta,\mu)'\hat{W}(\mu)\hat{g}(\theta,\mu)$ converges uniformly to $\hat{g}(\theta)'\hat{W}\hat{g}(\theta)$ as $\mu\to 0$. Since $\hat{\theta}$ is the unique solution to (1), this implies that, for any $\varepsilon>0$, there exists $\mu(\varepsilon)>0$ such that $\left\|\hat{\theta}(\mu)-\hat{\theta}(0)\right\|<\varepsilon$ whenever $|\mu|<\mu(\varepsilon)$, where $\hat{\theta}(\mu)$ is the unique solution to (OA1).

For any μ such that $\hat{\theta}(\mu)$ belongs to the interior of Θ, $\hat{\theta}(\mu)$ satisfies the first-order condition (in θ)

$$f\left(\hat{\theta}(\mu),\mu\right)=\hat{G}\left(\hat{\theta}(\mu)\right)'\hat{W}(\mu)\hat{g}\left(\hat{\theta}(\mu),\mu\right)=0,$$

where we have used the fact that

$$\frac{\partial}{\partial\theta}\hat{g}(\theta,\mu)=\hat{G}(\theta,\mu)=\hat{G}(\theta).$$

Finally, note that

$$\frac{\partial}{\partial\theta}f\left(\hat{\theta}(\mu),\mu\right)=\hat{G}\left(\hat{\theta}(\mu)\right)'\hat{W}(\mu)\hat{G}\left(\hat{\theta}(\mu)\right)+\hat{A}(\mu),$$

for

$$\hat{A}(\mu)=\left[\begin{array}{ccc}\left(\frac{\partial}{\partial\theta_1}\hat{G}\left(\hat{\theta}(\mu)\right)'\right)\hat{W}(\mu)\hat{g}\left(\hat{\theta}(\mu),\mu\right) & \dots & \left(\frac{\partial}{\partial\theta_P}\hat{G}\left(\hat{\theta}(\mu)\right)'\right)\hat{W}(\mu)\hat{g}\left(\hat{\theta}(\mu),\mu\right)\end{array}\right],$$

where $\frac{\partial}{\partial\theta}f\left(\hat{\theta}(\mu),\mu\right)$ has full rank for μ sufficiently small. By the implicit function theorem, for μ in an open neighborhood of zero we can define a unique differentiable function $\tilde{\theta}(\mu)$ such that $f\left(\tilde{\theta}(\mu),\mu\right)\equiv 0$. By the argument at the beginning of this proof $\hat{\theta}(\mu)=\tilde{\theta}(\mu)$ for μ sufficiently small. Thus, again by the implicit function theorem,

$$\frac{\partial}{\partial\mu}\tilde{\theta}(0)=-\left(\hat{G}\left(\hat{\theta}\right)'\hat{W}\hat{G}\left(\hat{\theta}\right)+\hat{A}\right)^{-1}\left(\hat{G}\left(\hat{\theta}\right)'\hat{W}\frac{\partial}{\partial\mu}\hat{g}\left(\hat{\theta},0\right)+\hat{B}\right),$$

for $\hat{A}=\hat{A}(0)$, which establishes the claim. □

2 Results Under Non-Vanishing Misspecification

In section 3 of the main text, we showed that the sensitivity matrix Λ allows us to characterize the first-order asymptotic bias of the estimator $\hat{\theta}$ under local perturbations to our data generating process. In this section, we show that analogous results hold for the probability limit of $\hat{\theta}$ under small, non-vanishing perturbations to the data generating process.

As in section 3 of the main text, we define a family of perturbed distributions $F(\cdot|\theta,\psi,\mu)$, where μ controls the degree of perturbation and $F(\cdot|\theta,\psi,0) = F(\cdot|\theta,\psi)$. Let $F_n(\mu) \equiv \{\times_n F(\cdot|\theta_0,\psi_0,\mu)\}_n$. When $\mu \neq 0$, the model is misspecified in the sense that under $F_n(\mu)$, $\hat{g}(\theta_0) \overset{p}{\nrightarrow} 0$. Online appendix proposition 2 below shows that Λ relates changes in the population values of the moments to changes in the probability limit of the estimator.

Online Appendix Assumption 1. *For a ball $\mathscr{B}_\mu$ around zero, we have that under $F_n(\mu)$ for any $\mu \in \mathscr{B}_\mu$, (i) $\hat{g}(\theta)$ and $\hat{G}(\theta)$ converge uniformly in θ to functions $g(\theta,\mu)$ and $G(\theta,\mu)$ that are continuously differentiable in (θ,μ) on $\Theta \times \mathscr{B}_\mu$, and (ii) $\hat{W} \overset{p}{\to} W(\mu)$ for $W(\mu)$ continuously differentiable on $\mathscr{B}_\mu$.*

Online Appendix Proposition 2. *Under online appendix assumption 1, there exists a ball $\mathscr{B}^*_\mu \subset \mathscr{B}_\mu$ around zero such that for any $\mu \in \mathscr{B}^*_\mu$, $\hat{\theta}$ converges in probability under $F_n(\mu)$ to a continuously differentiable function $\theta(\mu)$, and*

$$\frac{\partial}{\partial \mu}\theta(0) = \Lambda \frac{\partial}{\partial \mu} g(\theta_0, 0).$$

Proof. By the differentiability of $g(\theta,\mu)$ in μ, we know that $g(\theta,\mu) \to g(\theta)$ pointwise in θ as $\mu \to 0$. Moreover, since $G(\theta,\mu)$ is continuous in $(\theta,\mu) \in \Theta \times \mathscr{B}_\mu$, for $\tilde{\mathscr{B}}_\mu \subset \mathscr{B}_\mu$ a closed ball around zero, we know that $\sup_{(\theta,\mu)\in\Theta\times\tilde{\mathscr{B}}_\mu} \lambda_{\max}\left(G(\theta,\mu)' G(\theta,\mu)\right)$ is bounded, where $\lambda_{\max}(A)$ denotes the maximal eigenvalue of a matrix A. This implies that $g(\theta,\mu)$ is uniformly Lipschitz in θ for $\mu \in \tilde{\mathscr{B}}_\mu$, and thus that $g(\theta,\mu) \to g(\theta)$ uniformly in θ as $\mu \to 0$. Thus, $g(\theta,\mu)' W(\mu) g(\theta,\mu)$ converges uniformly to $g(\theta)' W g(\theta)$ as $\mu \to 0$. Since θ_0 is the unique solution to $\min_\theta g(\theta)' W g(\theta)$, this implies that, for any $\varepsilon > 0$, there exists $\mu(\varepsilon) > 0$ such that $\|\theta(\mu) - \theta_0\| < \varepsilon$ whenever $|\mu| < \mu(\varepsilon)$, where $\theta(\mu)$ is the unique solution to

$$\min_{\theta\in\Theta} g(\theta,\mu)' W(\mu) g(\theta,\mu).$$

Moreover, standard consistency arguments (e.g., theorem 2.1 in Newey and McFadden 1994) imply that $\hat{\theta} \overset{p}{\to} \theta(\mu)$ under $F_n(\mu)$.

Next, note that for any μ such that $\theta(\mu)$ belongs to the interior of Θ, $\theta(\mu)$ satisfies the first-order condition (in θ)

$$f(\theta(\mu),\mu) = G(\theta(\mu),\mu)' W(\mu) g(\theta(\mu),\mu) = 0.$$

Note that

$$\frac{\partial}{\partial \theta} f(\theta(\mu),\mu) = G(\theta(\mu),\mu)' W(\mu) G(\theta(\mu),\mu) + A(\mu),$$

for

$$A(\mu)=\left[\begin{array}{ccc}\left(\frac{\partial}{\partial\theta_1}G(\theta(\mu),\mu)'\right)W(\mu)g(\theta(\mu),\mu) & \ldots & \left(\frac{\partial}{\partial\theta_P}G(\theta(\mu),\mu)'\right)W(\mu)g(\theta(\mu),\mu)\end{array}\right].$$

Since we have assumed that $G'WG$ is non-singular and $\frac{\partial}{\partial\theta}f(\theta,\mu)$ is continuous in θ and μ, $\frac{\partial}{\partial\theta}f(\theta(\mu),\mu)$ has full rank for μ sufficiently close to zero. By the implicit function theorem, for μ in an open neighborhood of zero we can define a unique differentiable function $\tilde{\theta}(\mu)$ such that $f\left(\tilde{\theta}(\mu),\mu\right)\equiv 0$. By the argument at the beginning of this proof $\theta(\mu)=\tilde{\theta}(\mu)$ for μ sufficiently small. Thus, again by the implicit function theorem,

$$\frac{\partial}{\partial\mu}\tilde{\theta}(\mu)=\begin{array}{c}-\left(G(\theta(\mu),\mu)'W(\mu)G(\theta(\mu),\mu)+A(\mu)\right)^{-1}\\ \times\left(G(\theta(\mu),\mu)'W(\mu)\frac{\partial}{\partial\mu}g(\theta(\mu),\mu)+B(\mu)+C(\mu)\right)\end{array},$$

for

$$B(\mu)=G(\theta(\mu),\mu)'\left(\frac{\partial}{\partial\mu}W(\mu)\right)g(\theta(\mu),\mu)$$

$$C(\mu)=\left(\frac{\partial}{\partial\mu}G(\theta(\mu),\mu)'\right)W(\mu)g(\theta(\mu),\mu).$$

Since $\theta(\mu)=\tilde{\theta}(\mu)$ for μ sufficiently small, $\tilde{\theta}(0)=\theta(0)=\theta_0$. Thus, since $g(\theta(0),0)=g(\theta_0)=0$, $A(0)$, $B(0)$, and $C(0)$ are all equal to zero, from which the conclusion follows immediately for $\mathscr{B}_\mu^*$ sufficiently small. □

If we define $F(\cdot|\theta,\psi,\mu)$ such that

$$\frac{\partial}{\partial\mu}g(\theta_0,0)=g(a)-g(a_0),$$

for $g(a)$ the probability limit of $\hat{g}(\theta_0)$ under assumptions a, then for $\theta(a)$ the probability limit of $\hat{\theta}$ under assumption a, online appendix proposition 2 implies the first-order approximation

$$\begin{aligned}\theta(a)-\theta_0 &\approx \Lambda[g(a)-g(a_0)]\\ &= \Lambda g(a)\end{aligned}$$

discussed in section 3 of the main text.

Sections 4 and 6 of the main text develop sufficient conditions to apply our results and estimate sensitivity for the case of local perturbations. We next develop analogous results for models with a fixed degree of misspecification as studied in online appendix proposition 2. We first revisit the special cases considered in section 4 and then consider estimation of Λ.

2.1 Special Cases

We begin by developing the analogue of lemma 1 in the appendix to the main text.

Online Appendix Lemma 1. *Suppose that under* $F_n(\mu)$

$$\hat{g}(\theta) = \hat{a}(\theta) + \hat{b},$$

where the distribution of $\hat{a}(\theta)$ *is the same under* $F_n(0)$ *and* $F_n(\mu)$ *for every n,* $\hat{a}(\theta)$ *converges to a twice continuously differentiable function* $a(\theta)$*, and* $\hat{b}$ *converges in probability to* $b(\mu)$*, which is continuously differentiable in* μ*, and* $b(0) = 0$*. Suppose also that* $\hat{W}$ *either does not depend on the data or is equal to* $w\left(\hat{\theta}^{FS}\right)$ *for* $w(\cdot)$ *a continuously differentiable function and* $\hat{\theta}^{FS}$ *a first-stage estimator that solves (1) for a fixed positive semi-definite weight matrix* W^{FS} *not dependent on the data. Then online appendix assumption 1 holds.*

Proof. By assumption, under F_n we have that $\hat{g}(\theta) = \hat{a}(\theta)$ and $\hat{G}(\theta) = \frac{\partial}{\partial\theta}\hat{a}(\theta)$ converge uniformly to $g(\theta)$ and $G(\theta)$. Since $\hat{b}$ does not depend on θ, $\hat{g}(\theta)$ converges uniformly to $g(\theta,\mu) = g(\theta) + b(\mu)$ under $F_n(\mu)$, while $\hat{G}(\theta)$ converges uniformly to $G(\theta,\mu) = G(\theta)$. As we have assumed that $b(\mu)$ is continuously differentiable, we see that $g(\theta,\mu)$ and $G(\theta,\mu)$ are continuously differentiable in (θ,μ), as we wanted to show. By applying online appendix proposition 2 with $\hat{W} = W^{FS}$ (which satisfies the remaining condition of online appendix assumption 1 by construction), we can establish that, under $F_n(\mu)$, $\hat{\theta}^{FS} \xrightarrow{p} \theta^{FS}(\mu)$, which is differentiable in μ in a neighborhood of zero. Thus, $\hat{W} \xrightarrow{p} W(\mu) = w\left(\theta^{FS}(\mu)\right)$, which is continuously differentiable in μ by the chain rule. □

Applying this result, we can extend proposition 2 of the main text to describe the behavior of minimum distance estimators under a fixed level of misspecification.

Online Appendix Proposition 3. *Suppose that* $\hat{\theta}$ *is a CMD estimator and, under* $F_n(\mu)$*,* $\hat{s} = \tilde{s} + \mu\hat{\eta}$*, where* $\hat{\eta}$ *converges in probability to a vector of constants* η *and the distribution of* $\tilde{s}$ *does not depend on* μ*. Suppose that* $\hat{W}$ *takes the form given in online appendix lemma 1. Then, for* $\theta(\mu)$ *the probability limit of* $\hat{\theta}$ *under* $F_n(\mu)$*, we have that* $\frac{\partial}{\partial\mu}\theta(0) = \Lambda\eta$*.*

Proof. By online appendix lemma 1, online appendix assumption 1 holds for CMD estimators. The result then follows from online appendix proposition 2. □

Next, we turn to nonlinear IV models and develop the analogue of lemma 2 in the appendix to the main text. For clarity, here we make explicit the dependence of $\hat{\zeta}_i$ on the data D_i and write $\hat{\zeta}_i(\theta) = \hat{\zeta}(Y_i, X_i; \theta)$.

Online Appendix Assumption 2. *The observed data $D_i = [Y_i, X_i]$ consist of i.i.d. draws of endogenous variables Y_i and exogenous variables X_i, where $Y_i = h(X_i, \zeta_i; \theta)$ is a continuous one-to-one transformation of the vector of structural errors ζ_i given X_i and θ, with inverse $\hat{\zeta}(Y_i, X_i; \theta) = \hat{\zeta}_i(\theta)$. There is also an unobserved (potentially omitted) variable V_i. Under F_n, for a ball $\mathscr{B}_\mu$ around zero: (i) $\mathbb{E}\left(\hat{\zeta}(h(X_i, \zeta_i + \mu V_i; \theta_0), X_i; \theta)\right)$ and $\mathbb{E}\left(\frac{\partial}{\partial\theta}\hat{\zeta}(h(X_i, \zeta_i + \mu V_i; \theta_0), X_i; \theta)\right)$ are continuously differentiable in $(\theta, \mu) \in \Theta \times \mathscr{B}_\mu$; (ii) there exists a random variable $d(D_i)$ such that both*

$$\sup_{(\theta,\mu)\in\Theta\times\mathscr{B}_\mu} \left\| Z_i \hat{\zeta}(h(X_i, \zeta_i + \mu V_i; \theta_0), X_i; \theta) \right\| \leq d(D_i)$$

and

$$\sup_{(\theta,\mu)\in\Theta\times\mathscr{B}_\mu} \left\| Z_i \frac{\partial}{\partial\theta} \hat{\zeta}(h(X_i, \zeta_i + \mu V_i; \theta_0), X_i; \theta) \right\| \leq d(D_i)$$

with probability one, and $\mathbb{E}(d(D_i))$ is finite; and (iii) $\hat{W}$ either does not depend on the data or is equal to $\hat{W}\left(\hat{\theta}^{FS}\right)$, where, under $F_n(\mu)$, $\hat{W}(\theta)$ converges uniformly in θ to $W(\theta, \mu)$ which is differentiable in (θ, μ), and $\hat{\theta}^{FS}$ is a first-stage estimator that solves (1) for $\hat{W}^{FS}$ which depends on the data only through X_i.

Online Appendix Lemma 2. *Suppose that online appendix assumption 2 holds and that, under $F_n(\mu)$ with $\mu \in \mathscr{B}_\mu$, we have $\hat{\zeta}_i(\theta_0) = \tilde{\zeta}_i + \mu V_i$, where the distribution of $\left(\tilde{\zeta}_i, X_i, V_i\right)$ does not depend on μ. Then online appendix assumption 1 holds.*

Proof. By assumption, the distribution of (Y_i, X_i) under $F_n(\mu)$ is the same as the distribution of $(h(X_i, \zeta_i + \mu V_i; \theta_0), X_i)$ under $F_n(0)$. Thus, by the uniform law of large numbers (see lemma 2.4 of Newey and McFadden 1994), part (ii) of online appendix assumption 2 implies that, under $F_n(\mu)$, $\hat{g}(\theta) \xrightarrow{p} g(\theta, \mu)$ and $\hat{G}(\theta) \xrightarrow{p} G(\theta, \mu)$, both uniformly in θ. Part (i) of online appendix assumption 2 then implies that both of these limits are continuously differentiable in (θ, μ). Since $\hat{W}^{FS}$ does not depend on μ, we see that $\hat{W}^{FS} \xrightarrow{p} W^{FS}$ under $F_n(\mu)$. Thus, for this weight matrix, we have verified all the conditions of online appendix assumption 1, so online appendix proposition 2 implies that $\hat{\theta}^{FS} \xrightarrow{p} \theta^{FS}(\mu)$, which is differentiable in μ in some neighborhood of zero. Thus, we see that

$$\hat{W} = \hat{W}\left(\theta^{FS}(\mu)\right) \xrightarrow{p} W(\theta(\mu), \mu),$$

where the limit is differentiable in μ. Thus, we have verified the conditions of online appendix assumption 1 in a neighborhood of zero. □

Using this result, we can now develop the analogue of proposition 3 in the main text.

Online Appendix Proposition 4. *Suppose that $\hat{\theta}$ is an IV estimator satisfying online appendix assumption 2 and that, under $F_n(\mu)$, we have $\hat{\zeta}_i(\theta_0) = \tilde{\zeta}_i + \mu V_i$, where scalar V_i is an omitted* *variable with* $\frac{1}{n}Z'V \xrightarrow{p} \Omega_{ZV} \neq 0$ *and the distribution of* $\tilde{\zeta}_i$ *does not depend on* μ. *Then, for* $\theta(\mu)$ *the probability limit of* $\hat{\theta}$ *under* $F_n(\mu)$, *we have* $\frac{\partial}{\partial \mu}\theta(0) = \Lambda\Omega_{ZV}$.

Proof. By online appendix lemma 2, online appendix assumption 1 holds. Online appendix proposition 2 thus implies that

$$\frac{\partial}{\partial \mu}\theta(0) = \Lambda \frac{\partial}{\partial \mu}\mathbb{E}\left(Z_i \hat{\zeta}\left(h\left(X_i, \zeta_i + \mu V_i; \theta_0\right), X_i; \theta\right)\right)\Big|_{\mu=0}.$$

However, online appendix assumption 2 part (ii) implies that $Z_i \hat{\zeta}\left(h\left(X_i, \zeta_i + \mu V_i; \theta_0\right), X_i; \theta\right)$ is uniformly integrable for μ in a neighborhood of zero. Thus, we can exchange integration and differentiation to obtain that

$$\frac{\partial}{\partial \mu}\left|\mathbb{E}\left(Z_i \hat{\zeta}\left(h\left(X_i, \zeta_i + \mu V_i; \theta_0\right), X_i; \theta\right)\right)\right|_{\mu=0} = \mathbb{E}\left(Z_i V_i \frac{\partial}{\partial \zeta_i}\hat{\zeta}\left(h\left(X_i, \zeta_i; \theta_0\right), X_i; \theta\right)\right)$$

$$= \mathbb{E}\left(Z_i V_i\right) = \Omega_{ZV},$$

since $Y_i = h\left(X_i, \zeta_i; \theta\right)$ is a one-to-one function with inverse $\hat{\zeta}\left(Y_i, X_i; \theta\right)$. □

2.2 Estimation of Sensitivity Under Non-Vanishing Misspecification

This section considers estimation of sensitivity and develops results analogous to propositions 4 and 5 in the main text. We first consider plug-in sensitivity $\hat{\Lambda}$.

Online Appendix Lemma 3. *Under online appendix assumption 1,* $\hat{\Lambda} \xrightarrow{p} \Lambda(\mu)$ *under* $F_n(\mu)$ *for* $\mu \in \tilde{\mathscr{B}}_\mu \subset \mathscr{B}^*_\mu$, *where* $\Lambda(\cdot)$ *is continuous and* $\Lambda(0) = \Lambda$.

Proof. We have assumed that $\hat{g}(\theta) \xrightarrow{p} g(\theta, \mu)$ and $\hat{G}(\theta) \xrightarrow{p} G(\theta, \mu)$, both uniformly in θ. By online appendix proposition 2, we know that $\hat{\theta} \xrightarrow{p} \theta(\mu)$ for $\mu \in \mathscr{B}^*_\mu$, where $\theta(\mu)$ is continuous in μ. We have assumed, moreover, that $g(\theta, \mu)$, $G(\theta, \mu)$, and $W(\mu)$ are continuous in (θ, μ), so

$$G(\theta(\mu), \mu)' W(\mu) G(\theta(\mu), \mu)$$

is continuous in μ as well. Thus, since $G'WG$ has full rank, $\left(G(\theta(\mu), \mu)' W(\mu) G(\theta(\mu), \mu)\right)^{-1}$ is continuous for $\mu \in \mathscr{B}^*_\mu$. Define $\tilde{\mathscr{B}}_\mu \subset \mathscr{B}^*_\mu$ to be a ball of sufficiently small radius and note that, for

$$\Lambda(\mu) = -\left(G(\theta, \mu)' W(\mu) G(\theta, \mu)\right)^{-1} G(\theta, \mu)' W(\mu),$$

$\Lambda(\cdot)$ is continuous on $\tilde{\mathscr{B}}_\mu$, $\Lambda(0) = \Lambda$, and by the continuous mapping theorem $\hat{\Lambda} \xrightarrow{p} \Lambda(\mu)$ under $F_n(\mu)$. □

Analogous to proposition 5, we thus see that plug-in sensitivity is consistent under the assumptions of online appendix proposition 2. By contrast, we require an additional assumption to establish consistency of sample sensitivity $\hat{\Lambda}_S$ for Λ.

Online Appendix Assumption 3. *For $1 \leq p \leq P$, under $F_n(\mu)$ for any $\mu \in \mathscr{B}_\mu$, $\frac{\partial}{\partial \theta_p}\hat{G}(\theta)$ converges to $\frac{\partial}{\partial \theta_p}G(\theta,\mu)$ uniformly in θ on a ball $\mathscr{B}_\theta$ around θ_0.*

Online Appendix Lemma 4. *Suppose that online appendix assumption 1 and online appendix assumption 3 are satisfied. Then there exists a ball $\check{\mathscr{B}}_\mu$ around zero such that, under $F_n(\mu)$ for $\mu \in \check{\mathscr{B}}_\mu$, $\hat{\Lambda}_S \xrightarrow{p} \Lambda_S(\mu)$ with $\Lambda_S(0) = \Lambda$.*

Proof. By online appendix proposition 2, we know that $\hat{\theta} \xrightarrow{p} \theta(\mu)$ under $F_n(\mu)$, and that $\theta(\mu) \in \mathscr{B}_\theta$ for μ sufficiently small. Thus, since we have assumed that $\hat{g}(\theta)$, $\hat{G}(\theta)$, and $\frac{\partial}{\partial \theta_p}\hat{G}(\theta)$ converge uniformly to $g(\theta,\mu)$, $G(\theta,\mu)$, and $\frac{\partial}{\partial \theta_p}G(\theta,\mu)$ on $\mathscr{B}_\theta$, we see that, for μ sufficiently small,

$$\left(\hat{g}\left(\hat{\theta}\right), \hat{G}\left(\hat{\theta}\right), \frac{\partial}{\partial \theta_p}\hat{G}\left(\hat{\theta}\right), \hat{W}\right) \xrightarrow{p} \left(g\left(\theta\left(\mu\right),\mu\right), G\left(\theta\left(\mu\right),\mu\right), \frac{\partial}{\partial \theta_p}G\left(\theta\left(\mu\right),\mu\right), W\left(\mu\right)\right).$$

Thus, by the continuous mapping theorem, we see that, for μ sufficiently small,

$$\hat{G}\left(\hat{\theta}\right)'\hat{W}\hat{G}\left(\hat{\theta}\right) + \hat{A} \xrightarrow{p} G\left(\theta\left(\mu\right),\mu\right)'W\left(\mu\right)G\left(\theta\left(\mu\right),\mu\right) + A\left(\mu\right),$$

where since $G'WG$ is non-singular and $\theta(\mu) \to \theta_0$, $A(\mu) \to 0$ as $\mu \to 0$, the right hand side is non-singular for μ sufficiently small. Applying the continuous mapping theorem thus yields the result for $\check{\mathscr{B}}_\mu$ a ball of sufficiently small radius. □

Overall, we see that the results under a fixed degree of misspecification agree with the results under local perturbations: $\hat{\Lambda}$ is consistent for sensitivity under the assumptions of online appendix proposition 2, while an additional assumption is required to establish consistency of $\hat{\Lambda}_S$ for sensitivity.

3 Results for Non-Smooth Moments

As noted in section 3 of the main text, many of our results on sensitivity extend to models where $\hat{g}(\theta)$ is not differentiable in θ, allowing us to accommodate a range of additional estimators including quantile regression and many simulation-based approaches. To formalize this, following section 7.1 of Newey and McFadden (1994) we assume that the estimator $\hat{\theta}$ satisfies

$$\hat{g}\left(\hat{\theta}\right)'\hat{W}\hat{g}\left(\hat{\theta}\right) \leq \inf_{\theta \in \Theta} \hat{g}\left(\theta\right)'\hat{W}\hat{g}\left(\theta\right) + o_p\left(\frac{1}{n}\right).$$

Thus, $\hat{\theta}$ need not exactly minimize the objective function (which may be impossible in some models with non-smooth moments), but should come close. We further assume that under $F_n(0)$, (i) $\sqrt{n}\hat{g}(\theta_0) \xrightarrow{d} N(0,\Omega)$; (ii) $\hat{W}$ converges in probability to a positive semi-definite matrix W; (iii) there is a differentiable function $g(\theta)$ with derivative $G(\theta)$ such that $g(\theta) = 0$ only at $\theta = \theta_0$ and $\hat{g}(\theta)$ converges to $g(\theta)$ uniformly in θ; (iv) $G'WG$ is nonsingular with $G = G(\theta_0)$; and (v) for any sequence $\kappa_n \to 0$,

$$\sup_{\|\theta-\theta_0\|\leq\kappa_n} \frac{\sqrt{n}\,\|\hat{g}(\theta) - \hat{g}(\theta_0) - g(\theta)\|}{1+\sqrt{n}\,\|\theta-\theta_0\|} \xrightarrow{p} 0.$$

See section 7 of Newey and McFadden (1994) for a discussion of sufficient conditions for (v).

Under these assumptions, theorems 2.1 and 7.2 of Newey and McFadden (1994) imply that $\hat{\theta}$ is consistent and asymptotically normal with variance $\Sigma \equiv (G'WG)^{-1} G'W\Omega WG (G'WG)^{-1}$. Since the moments are non-smooth the results on sensitivity derived in the main text no longer directly apply, but it is still interesting to relate perturbations of the moments to perturbations of the parameter estimates.

The approach based on sample sensitivity is no longer feasible for non-smooth moments, since the estimates will not in general change smoothly with the moments when the moments are non-smooth. Our results under non-vanishing misspecification, on the other hand, go through nearly unchanged in this case.

Online Appendix Assumption 4. *For a ball $\mathscr{B}_\mu$ around zero, we have that under $F_n(\mu)$ for any $\mu \in \mathscr{B}_\mu$, $\hat{g}(\theta)$ converges uniformly in θ to a function $g(\theta,\mu)$ with $\frac{\partial}{\partial\theta}g(\theta,\mu) = G(\theta,\mu)$ such that both $g(\theta,\mu)$ and $G(\theta,\mu)$ are continuously differentiable in (θ,μ) on $\Theta \times \mathscr{B}_\mu$ and $\hat{W} \xrightarrow{p} W(\mu)$ for $W(\mu)$ continuously differentiable on $\mathscr{B}_\mu$.*

Online Appendix Proposition 5. *Under online appendix assumption 4, there exists a ball $\mathscr{B}^*_\mu \subset \mathscr{B}_\mu$ around zero such that for any $\mu \in \mathscr{B}^*_\mu$, $\hat{\theta}$ converges in probability under $F_n(\mu)$ to a continuously differentiable function $\theta(\mu)$, and*

$$\frac{\partial}{\partial\mu}\theta(0) = \Lambda \frac{\partial}{\partial\mu} g(\theta_0, 0).$$

Proof. The proof is exactly the same as that of online appendix proposition 2. □

Thus, even when the moments are non-smooth, sensitivity characterizes the derivative of the probability limit of $\hat{\theta}$ with respect to perturbations of the moments. The results in online appendix section 2.2 can likewise be extended to the case with non-smooth moments, but we instead follow the main text and focus on results under local perturbations.

In particular, for models with non-smooth moments we say that a sequence $\{\mu_n\}_{n=1}^{\infty}$ is a *local perturbation* if under $F_n(\mu_n)$: (i) $\hat{\theta} \xrightarrow{p} \theta_0$; (ii) $\sqrt{n}\hat{g}(\theta_0)$ converges in distribution to a random variable $\tilde{g}$; (iii) $\hat{g}(\theta)$ converges uniformly in probability to $g(\theta)$; (iv) $\hat{W} \xrightarrow{p} W$; and (v) for any sequence $\kappa_n \to 0$,

$$\sup_{\|\theta-\theta_0\|\leq\kappa_n} \frac{\sqrt{n}\,\|\hat{g}(\theta)-\hat{g}(\theta_0)-g(\theta)\|}{1+\sqrt{n}\,\|\theta-\theta_0\|} \xrightarrow{p} 0.$$

As before, any sequence of alternatives μ_n such that $F_n(\mu_n)$ is contiguous to $F_n(0)$ and under which $\sqrt{n}\hat{g}(\theta_0)$ has a well-defined limiting distribution will be a local perturbation. As in proposition 1 of the main text, we can derive the asymptotic distribution of $\hat{\theta}$ under local perturbations.

Online Appendix Proposition 6. *For any local perturbation $\{\mu_n\}_{n=1}^{\infty}$, $\sqrt{n}\left(\hat{\theta}-\theta_0\right)$ converges in distribution under $F_n(\mu_n)$ to a random variable $\tilde{\theta}$ with*

$$\tilde{\theta} = \Lambda\tilde{g}$$

almost surely. This implies in particular that the first-order asymptotic bias $\mathbb{E}\left(\tilde{\theta}\right)$ is equal to $\Lambda\mathbb{E}(\tilde{g})$.

Proof. By the same argument as in the proofs of theorems 7.1 and 7.2 of Newey and McFadden (1994) applied under $F_n(\mu_n)$,

$$\sqrt{n}\left(\hat{\theta}-\theta_0+\left(G'WG\right)^{-1}G'W\hat{g}(\theta_0)\right) \xrightarrow{p} 0,$$

and thus

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right) = \Lambda\sqrt{n}\hat{g}(\theta_0)+o_p(1).$$

Consequently, by the continuous mapping theorem,

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right) \xrightarrow{d} \tilde{\theta} = \Lambda\tilde{g},$$

which proves the claim. □

As in the case with smooth moments, we next give sufficient conditions for a sequence of data generating processes to be a local perturbation.

3.1 Special Cases

We first consider additive perturbations of the moments as in lemma 1 in the appendix to the main text. The statement of the resulting lemma is the same as that of lemma 1 in the appendix to the main text, but the proof differs slightly so we re-state the result for completeness.

Online Appendix Lemma 5. *Consider a sequence* $\{\mu_n\}_{n=1}^{\infty}$*. Suppose that under* $F_n(\mu_n)$

$$\hat{g}(\theta) = \hat{a}(\theta) + \hat{b},$$

where the distribution of $\hat{a}(\theta)$ *is the same under* $F_n(0)$ *and* $F_n(\mu_n)$ *for every n and* $\sqrt{n}\hat{b}$ *converges in probability. Also,* $\hat{W} \xrightarrow{p} W$ *under* $F_n(\mu_n)$*. Then* $\{\mu_n\}_{n=1}^{\infty}$ *is a local perturbation.*

Proof. Uniform convergence of $\hat{g}(\theta)$ follows from uniform convergence of $\hat{a}(\theta)$ and the fact that $\hat{b} \xrightarrow{p} 0$. Convergence in distribution of $\sqrt{n}\hat{g}(\theta_0)$ follows from the fact that $\sqrt{n}\hat{a}(\theta_0)$ converges in distribution and $\sqrt{n}\hat{b}$ converges in probability. That $\hat{\theta} \xrightarrow{p} \theta_0$ then follows from the observation that $\hat{g}(\theta)' \hat{W} \hat{g}(\theta)$ converges uniformly to $g(\theta)' W g(\theta)$. Finally, that

$$\sup_{\|\theta - \theta_0\| \leq \kappa_n} \frac{\sqrt{n}\|\hat{g}(\theta) - \hat{g}(\theta_0) - g(\theta)\|}{1 + \sqrt{n}\|\theta - \theta_0\|} \xrightarrow{p} 0$$

follows from the fact that $\hat{b}$ differences out of this expression, and we have assumed that this holds under $F_n(0)$. □

Applying online appendix lemma 5 and online appendix proposition 6 again yields a simple characterization of the first-order asymptotic bias of misspecified CMD estimators with non-smooth moments, which is again the same as the corresponding result for the case with differentiable moments (proposition 2 in the main text).

Online Appendix Proposition 7. *Suppose that* $\hat{\theta}$ *is a CMD estimator and, under* $F_n(\mu)$*,* $\hat{s} = \tilde{s} + \mu\hat{\eta}$*, where* $\hat{\eta}$ *converges in probability to a vector of constants* η *and the distribution of* $\tilde{s}$ *does not depend on* μ*. Take* $\mu_n = \frac{1}{\sqrt{n}}$*, and suppose that* $\hat{W} \xrightarrow{p} W$ *under* $F_n(\mu_n)$*. Then taking* $\mu_n = \frac{1}{\sqrt{n}}$*, we have* $\mathbb{E}\left(\tilde{\theta}\right) = \Lambda\eta$*.*

Proof. That $\{\mu_n\}_{n=1}^{\infty}$ is a local perturbation follows from online appendix lemma 5 with $\hat{a}(\theta) = \tilde{s} - s(\theta)$ and $\hat{b} = \mu_n\hat{\eta}$. The expression for $\mathbb{E}\left(\tilde{\theta}\right)$ then follows by online appendix proposition 6. □

To extend the results of lemma 2 to the case with non-smooth moments, we need to incorporate the definition of local perturbations for the non-smooth case, but we don't need to modify assumption 1 in the main text at all.

Online Appendix Lemma 6. *Consider a sequence* $\{\mu_n\}_{n=1}^{\infty}$ *with* $\mu_n = \frac{\mu^*}{\sqrt{n}}$ *for a constant* μ^**. Suppose that assumption 1 in the main text holds and that, under* $F_n(\mu)$*, we have* $\hat{\zeta}_i(\theta_0) = \tilde{\zeta}_i + \mu V_i$*, where the distribution of* $\left(\tilde{\zeta}_i, X_i, V_i\right)$ *does not depend on* μ*. Then* $\{\mu_n\}_{n=1}^{\infty}$ *is a local perturbation.*

Proof. By the proof of lemma 2 in the appendix to the main text, the sequence of data generating processes $F_n(\mu_n)$ is contiguous to $F_n(0)$, and

$$\sqrt{n}\hat{g}(\theta_0) \xrightarrow{d} N(\mu^*\Xi, \Omega)$$

under $F_n(\mu_n)$. Here, as in the proof of lemma 2 in the appendix to the main text, $\mu^*\Xi$ is the asymptotic covariance between the moments $\hat{g}(\theta_0)$ and the log-likelihood ratio $\log \frac{dF_n(\mu_n)}{dF_n(0)}$. By contiguity, convergence in probability under $F_n(0)$ implies convergence in probability to the same limit under $F_n(\mu_n)$, which suffices to verify the other conditions for $\{\mu_n\}_{n=1}^{\infty}$ to be a local perturbation. □

Analogous to proposition 3, applying online appendix lemma 6 and online appendix proposition 6 allows us to characterize the effects of misspecification in the class of nonlinear IV estimators where $\hat{\zeta}(\theta)$ may be non-smooth in θ.

Online Appendix Proposition 8. *Suppose that $\hat{\theta}$ is an IV estimator satisfying assumption 1 in the main text and that, under $F_n(\mu)$, we have $\hat{\zeta}_i(\theta_0) = \tilde{\zeta}_i + \mu V_i$, where V_i is an omitted variable with $\frac{1}{n}\sum_i Z_i \otimes V_i \xrightarrow{p} \Omega_{ZV} \neq 0$ and the distribution of $\tilde{\zeta}_i$ does not depend on μ. Then taking $\mu_n = \frac{1}{\sqrt{n}}$, we have $\mathbb{E}(\tilde{\theta}) = \Lambda\Omega_{ZV}$.*

Proof. That $\{\mu_n\}_{n=1}^{\infty}$ is a local perturbation follows from online appendix lemma 6. The expression for $\mathbb{E}(\tilde{\theta})$ then follows from online appendix proposition 6. □

Thus, we see that the sufficient conditions for local perturbations developed in section 3 of the main text extend to non-smooth models. We next show that our results on the estimation of sensitivity can likewise be extended.

3.2 Estimation of Sensitivity with Non-Smooth Moments

To estimate sensitivity, we require an estimator of $G = G(\theta_0)$. Unlike in the case with smooth moments, we cannot simply differentiate $\hat{g}(\hat{\theta})$. Instead, we follow section 7.3 of Newey and McFadden (1994) and consider an estimator based on numerical derivatives. In particular, consider the matrix $\hat{G}(\theta)$ of numerical derivatives with j^{th} column

$$\left(\hat{g}\left(\theta + e_j\varepsilon_n\right) - \hat{g}\left(\theta - e_j\varepsilon_n\right)\right)/(2\varepsilon_n),$$

where e_j is the j^{th} standard basis vector and ε_n is a nonzero scalar. As in the smooth case, we can use this to define plug-in sensitivity.

Definition. Define **plug-in sensitivity** as

$$\hat{\Lambda} = -\left(\hat{G}\left(\hat{\theta}\right)' \hat{W} \hat{G}\left(\hat{\theta}\right)\right)^{-1} \hat{G}'\left(\hat{\theta}\right) \hat{W}.$$

Online Appendix Lemma 7. *Assuming* $\varepsilon_n \to 0$ *and* $\varepsilon_n \sqrt{n} \to \infty$ *as* $n \to \infty$, $\hat{\Lambda} \xrightarrow{p} \Lambda$ *under* $F_n(\mu_n)$ *for any local perturbation* $\{\mu_n\}_{n=1}^{\infty}$.

Proof. The proof of theorem 7.4 of Newey and McFadden (1994) implies that $\hat{G}\left(\hat{\theta}\right) \xrightarrow{p} G$. Since $G'WG$ has full rank by assumption the result follows by the continuous mapping theorem. □

While this result is obtained for a particular numerical derivative estimator $\hat{G}(\theta)$, analogous results can be established for alternative estimators (and a derivative estimator is usually required to compute standard errors in models with non-smooth moments).

Online Appendix Table 1: Standard deviations of excluded instruments in BLP (1995)

	Standard deviation
Demand-side instruments	
Other cars by same firm:	
Number of cars	11.9210
Sum of horsepower/weight	4.7548
Number of cars with AC standard	4.4708
Sum of miles/dollar	27.0197
Other cars by rival firms:	
Number of cars	23.5870
Sum of horsepower/weight	11.7205
Number of cars with AC standard	21.3580
Sum of miles/dollar	90.8318
Supply-side instruments	
Other cars by same firm:	
Number of cars	11.9210
Sum of log horsepower/weight	11.6288
Number of cars with AC standard	4.4708
Sum of log miles/gallon	8.4861
Sum of log size	4.1672
Sum of time trend	181.5812
Other cars by rival firms:	
Number of cars	23.5870
Sum of log horsepower/weight	19.5929
Number of cars with AC standard	21.3580
Sum of log miles/gallon	24.9750
Sum of log size	4.6629
This car:	
Miles/dollar	0.6981

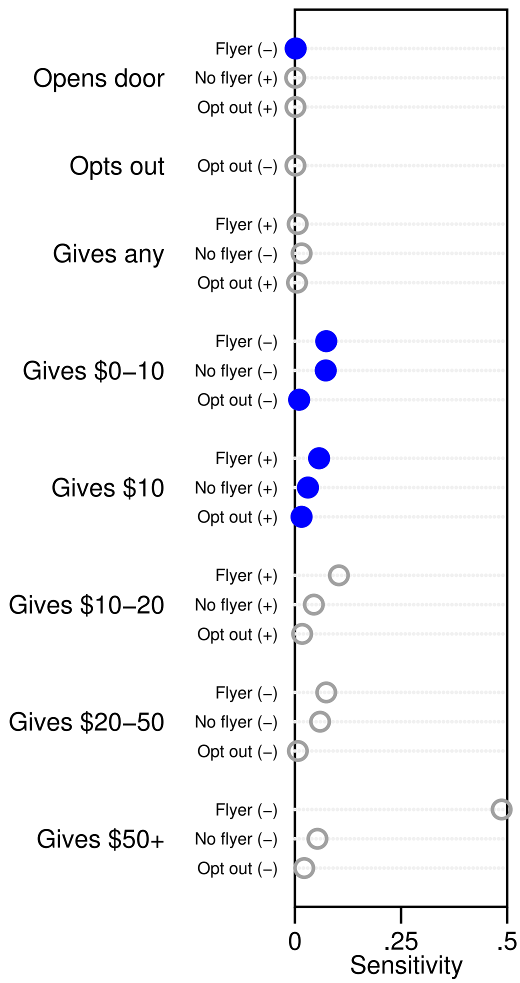
Note: Table shows the standard deviation, across the entire sample, of the excluded instruments Z_s, Z_d.

Online Appendix Table 2: Global sensitivity of average markup in BLP (1995)

	First-order bias in average markup	Global sensitivity of average markup	Difference
Violation of supply-side exclusion restrictions:			
Removing own car increases average marginal cost	-0.1731	-0.1284	-0.0446
by 1% of average price	(0.0433)	(0.0238)	(0.0482)
Removing rival's car increases average marginal cost	0.2095	-0.1895	0.3989
by 1% of average price	(0.0689)	(0.0348)	(0.0753)
Violation of demand-side exclusion restrictions:			
Removing own car decreases average willingness to pay	-0.1277	0.0289	-0.1566
by 1% of average price	(0.0915)	(0.0874)	(0.1271)
Removing rival's car decreases average willingness to pay	0.2515	0.1703	0.0812
by 1% of average price	(0.1285)	(0.1024)	(0.1634)
Baseline estimate	0.3272	0.3272	
	(0.0392)	(0.0392)	

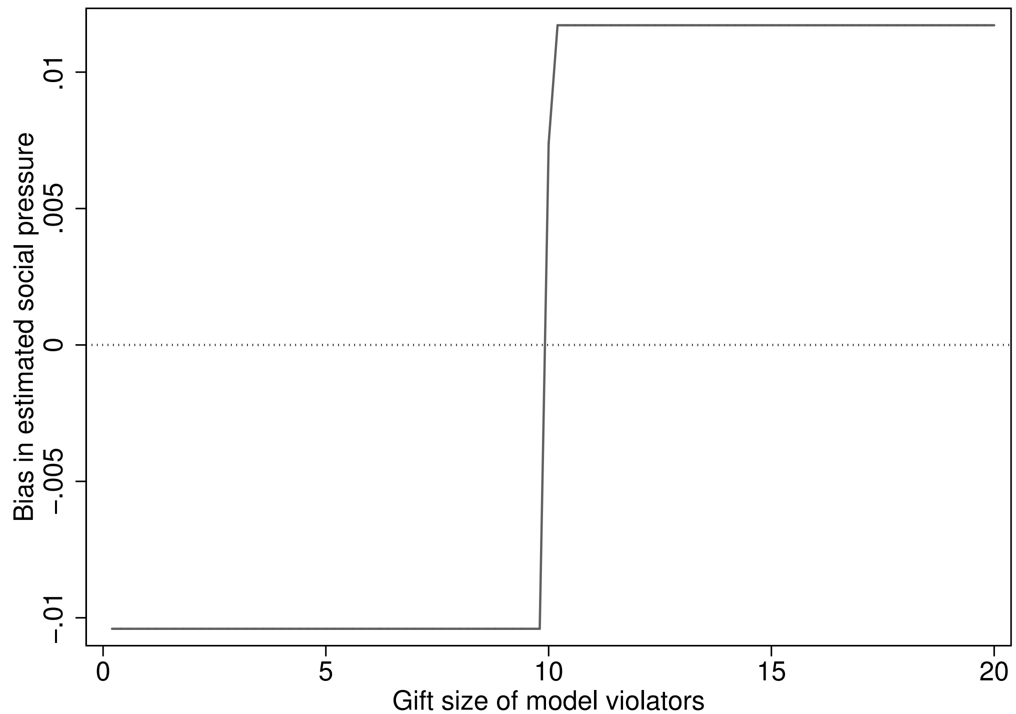
Note: The average markup is the average ratio of price minus marginal cost to price across all vehicles. The first column of the table reproduces the estimates of the first-order asymptotic bias from table 2 in the main paper. The second column of the table estimates the sensitivity of the estimated average markup to modifications of the moment conditions analogous to those in the first column. The third column of the table reports the difference between the values in the first and second columns. In all cases, standard errors in parentheses are based on a non-parametric block bootstrap over sample years with 70 replicates. The modifications to the moment conditions that we contemplate are as follows. In the first two rows, we modify the component of the supply-side moment condition corresponding to each car i to be $Z_{si}' \otimes \left(\hat{\omega}_i(\theta) + 0.01\left(\overline{P}/\overline{mc}\right) Num_i\right)$, where Num_i is the number of cars produced by the [same firm / other firms] as car i in the respective year, $\overline{mc}$ is the sales-weighted mean marginal cost over all cars i in 1980, and $\overline{P}$ is the sales-weighted mean price over all cars i in 1980. In the second two rows, we modify the component of the demand-side moment condition corresponding to each car i to be $Z_{di}' \otimes \left(\hat{\xi}_i(\theta) - 0.01\left(\overline{P}/K_{\xi}\right) Num_i\right)$, where K_{ξ} is the derivative of willingness to pay with respect to ξ for a 1980 household with mean income. Global sensitivity is calculated as the difference between the baseline estimate and the estimate under the corresponding modified moment condition. We treat the average price $\overline{P}$, the marginal cost $\overline{mc}$, and the derivative K_{ξ} as constant across bootstrap replicates and moment conditions.

Online Appendix Figure 1: Sensitivity of La Rabida social pressure cost in DellaVigna et al. (2012) to local violations of identifying assumptions



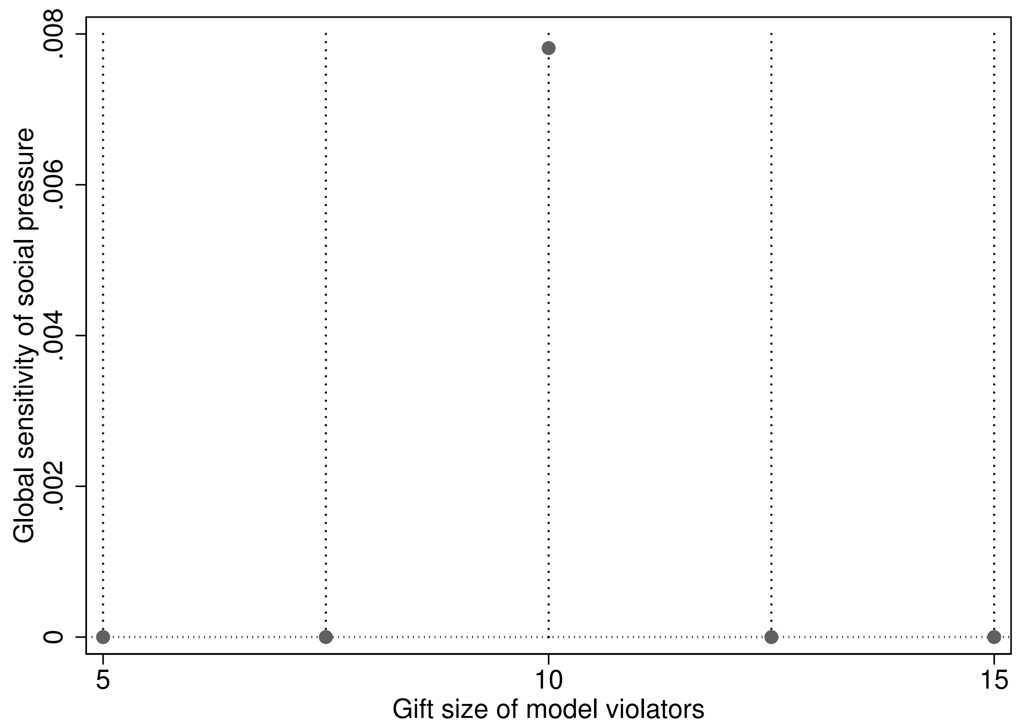
Notes: The plot shows one-hundredth of the absolute value of plug-in sensitivity of the social pressure cost of soliciting a donation for the La Rabida Children's Hospital (La Rabida) with respect to the vector of estimation moments, with the sign of sensitivity in parentheses. While sensitivity is computed with respect to the complete set of estimation moments, the plot only shows those corresponding to the La Rabida treatment. Each moment is the observed probability of a response for the given treatment group. The leftmost axis labels in larger font describe the response; the axis labels in smaller font describe the treatment group. Filled circles correspond to moments that DellaVigna et al. (2012) highlight as important for the parameter.

Online Appendix Figure 2: Sensitivity of La Rabida social pressure cost in DellaVigna et al. (2012) to exogenous gift levels



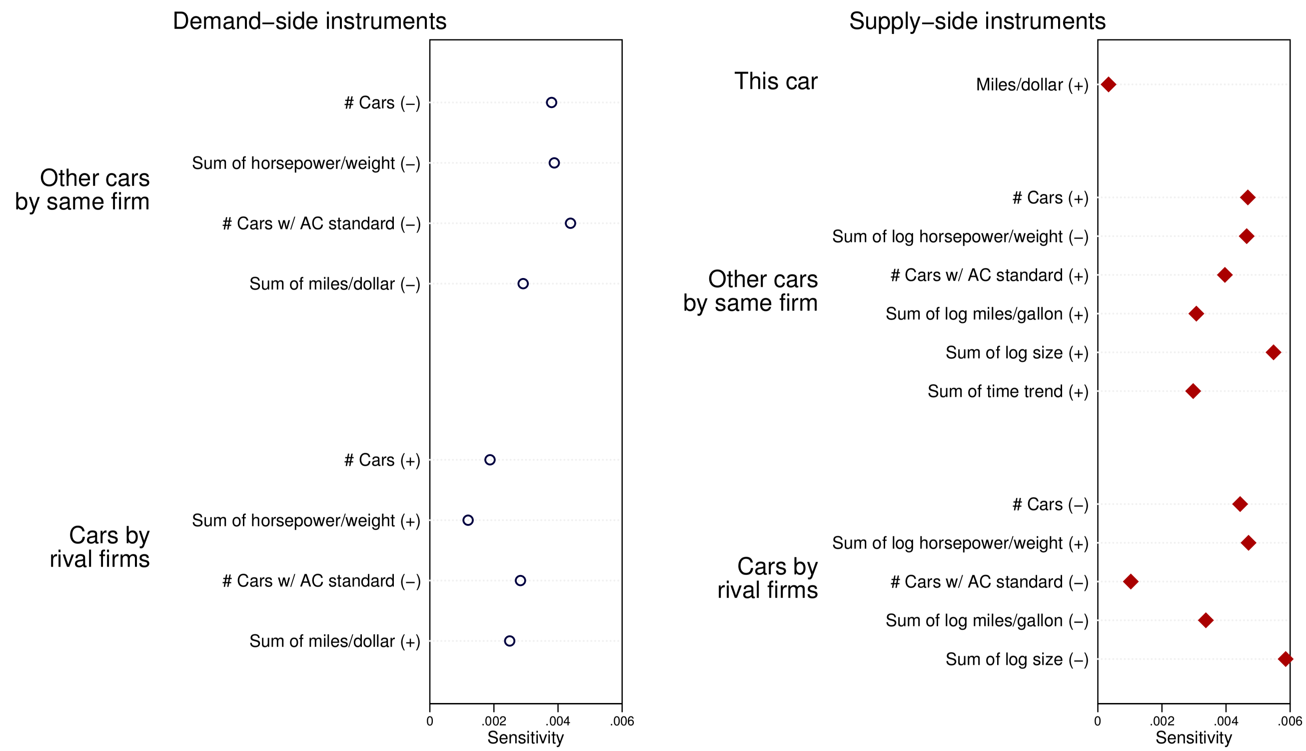
Notes: The plot shows the estimated first-order asymptotic bias in DellaVigna et al.'s (2012) published estimate of the per-dollar social cost of not giving to the La Rabida Children's Hospital under various levels of misspecification, as implied by proposition 2. Our calculations use the plug-in estimator of sensitivity. We consider perturbations under which a share $\left(1-\frac{0.01}{\sqrt{n}}\right)$ of households follow the paper's model and a share $\frac{0.01}{\sqrt{n}}$ give with the same probabilities as their model-obeying counterparts but always give an amount $\tilde{d}$ conditional on giving. First-order asymptotic bias is computed for values of $\tilde{d}$ in \$0.20 increments from \$0 to \$20 and interpolated between these increments. Values of $\tilde{d}$ are shown on the x axis.

Online Appendix Figure 3: Global sensitivity of ECU social pressure cost in DellaVigna et al. (2012)



Notes: The plot shows the estimated global sensitivity of DellaVigna et al.'s (2012) published estimate of the per-dollar social cost of not giving to the East Carolina Hazard Center (ECU) under various levels of misspecification. We consider perturbations under which a share 0.99 of households follow the paper's model and a share 0.01 give with the same probabilities as their model-obeying counterparts but always give an amount $\tilde{d}$ conditional on giving. Let $\tilde{s}(\theta, \tilde{d})$ denote the predicted value of each statistic in $\hat{s}$ under the alternative model. We solve (1) with moments $\hat{s} - \tilde{s}(\theta, \tilde{d})$ and the weight matrix provided to us by the authors. To solve (1), we use Matlab's *patternsearch* routine with the constraints, within 10^{-12}, outlined in DellaVigna et al. (2012, 36) and the additional constraints that the per-dollar social pressure cost parameters for both charities are less than $1 - 10^{-12}$. Though not explicitly mentioned in DellaVigna et al. (2012), we also constrain the curvature of the altruism function to be greater than 10^{-12}, to avoid taking the log of zero or a negative value in optimization. We use the published estimate $\hat{\theta}$ as the starting value of the optimization routine. Global sensitivity is calculated as the difference between the published estimate and the estimate under the alternative model. We compute global sensitivity for $\tilde{d} \in \{5, 7.5, 10, 12.5, 15\}$. Values of $\tilde{d}$ are shown on the x axis.

Online Appendix Figure 4: Sample sensitivity of average markup in BLP (1995) to local violations of the exclusion restrictions



Notes: The plot shows the absolute value of the estimate of $C\hat{\Lambda}_S\tilde{\Omega}_{ZZ}K$, where C is the gradient of the average markup with respect to model parameters, $\hat{\Lambda}_S$ is sample sensitivity of parameters to estimation moments, $\tilde{\Omega} = \begin{bmatrix} \mathbb{E}(Z_{di}Z'_{di}) & 0 \\ 0 & \mathbb{E}(Z_{si}Z'_{si}) \end{bmatrix}$, and K is a diagonal matrix of normalizing constants. We use plug-in estimates of C, $\tilde{\Omega}_{ZZ}$, and K. The sign of $C\hat{\Lambda}_S\Omega_{ZZ}K$ is shown in parentheses. For the demand-side instruments on the left, the diagonal elements of K are chosen so that the plotted values can be interpreted as sample sensitivity of the markup to beliefs about the effect of a one-standard-deviation increase in each instrument on the willingness-to-pay of a household with mean income in 1980, expressed as a percent of the sales-weighted mean price over all cars i in 1980. For the supply-side instruments on the right, the diagonal elements of K are chosen so that the plotted values can be interpreted as sample sensitivity of the markup to beliefs about the effect of a one-standard-deviation increase in each instrument on the marginal cost of a car with the sales-weighted average marginal cost in 1980, expressed as a percent of the sales-weighted mean price over all cars i in 1980. While sample sensitivity is computed with respect to the complete set of estimation moments, the plot only shows those corresponding to the excluded instruments (those that do not enter the utility or marginal cost equations directly).