NBER WORKING PAPER SERIES

MEASURING THE SENSITIVITY OF PARAMETER ESTIMATES TO SAMPLE STATISTICS

Matthew Gentzkow
Jesse M. Shapiro

Working Paper 20673
http://www.nber.org/papers/w20673

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2014

Conversations with Kevin M. Murphy inspired and greatly improved this work. We are grateful also to Josh Angrist, Steve Berry, Alan Bester, Raj Chetty, Tim Conley, Ron Goettler, Brett Gordon, Phil Haile, Christian Hansen, Matt Taddy, E. Glen Weyl, and seminar audiences at the University of Chicago, Berkeley, Harvard, the University of Michigan, MIT, Northwestern, Princeton, Stanford and Yale for advice and suggestions, and to our dedicated research assistants for important contributions to this project. We thank the following authors for their assistance in working with their code and data: Stefano DellaVigna, John List, and Ulrike Malmendier; Ron Goettler and Brett Gordon; Pierre-Olivier Gourinchas and Jonathan Parker; Chris Knittel and Konstantinos Metaxoglou; Michael Mazzeo; Boris Nikolov and Toni Whited; and Amil Petrin. This research was funded in part by the Initiative on Global Markets, the George J. Stigler Center for the Study of the Economy and the State, the Ewing Marion Kauffman Foundation, the Centel Foundation / Robert P. Reuss Faculty Research Fund, the Neubauer Family Foundation, and the Kathryn C. Gould Research Fund, all at the University of Chicago Booth School of Business, the Alfred P. Sloan Foundation, and the National Science Foundation. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Measuring the Sensitivity of Parameter Estimates to Sample Statistics
Matthew Gentzkow and Jesse M. Shapiro
NBER Working Paper No. 20673
November 2014
JEL No. C1,C52

ABSTRACT

Empirical papers in economics often describe heuristically how their estimators map specific data features to parameters or other magnitudes of interest. We propose a formal, quantitative measure of this relationship that can be computed at negligible cost even for complex models. We show that our measure of sensitivity to particular sample statistics can be informative about the importance of particular identifying assumptions, providing one rationale for the attention that sensitivity receives in applied research. We apply our measure to empirical papers in industrial organization, macroeconomics, public economics, and finance.

Matthew Gentzkow
University of Chicago
Booth School of Business
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
gentzkow@chicagobooth.edu

Jesse M. Shapiro
University of Chicago
Booth School of Business
5807 S. Woodlawn Avenue
Chicago, IL 60637
and NBER
jesse.shapiro@chicagobooth.edu

An Online Appendix is available at:
http://faculty.chicagobooth.edu/matthew.gentzkow/research/trans_online.pdf

1 Introduction

An estimator is a mapping from data to parameters of interest. In many cases, it is possible to show analytically how the estimates change as the data vary along particular dimensions. In other cases, the estimator is sufficiently complicated that interrogating the mapping through brute-force computation, or through direct inspection of the economic and econometric assumptions, is prohibitively costly. In this paper, we introduce a quantitative measure of an estimator's sensitivity to specific features of the data that is easy to compute even for complex models. We then apply the measure to gain new insight into empirical models in industrial organization, macroeconomics, public economics, and finance.

Throughout the paper, we consider the following abstract setting. A researcher begins with an estimator $\hat{\theta}$ of some economic parameters with true value θ_0. The researcher also computes a vector of statistics $\hat{\gamma}$ that summarize the data features of interest. These may be the moments used in estimating $\hat{\theta}$ in a GMM procedure, descriptive statistics such as means or variances, or estimates of the parameters of an auxiliary model. The statistics $\hat{\theta}$ and $\hat{\gamma}$ are jointly asymptotically normal, and $\hat{\theta}$ is consistent for θ_0.

We define the *sensitivity* of $\hat{\theta}$ to $\hat{\gamma}$ to be the expected coefficient from a regression of $\hat{\theta}$ on $\hat{\gamma}$ in data drawn from the asymptotic distribution. In other words, suppose we: (i) draw samples of size n from the true DGP of the data; (ii) compute $\hat{\theta}_s$ and $\hat{\gamma}_s$ for each sample s; (iii) stack the $\hat{\theta}_s$ and $\hat{\gamma}_s$ and regress each column of the former on the latter plus a constant term. Then sensitivity is the limit of the expected regression coefficients as the sample size n grows large. In the special case in which $\hat{\theta}$ is fully determined by $\hat{\gamma}$, sensitivity corresponds to the derivative of $\hat{\theta}$ with respect to $\hat{\gamma}$ at the probability limit of $\hat{\gamma}$. Sensitivity thus measures how $\hat{\theta}$ is related to the elements of $\hat{\gamma}$ across alternative realizations of the data.

Computing a consistent estimate of sensitivity does not require actually performing this resampling experiment. When $\hat{\gamma}$ is the vector of moments used in estimating $\hat{\theta}$, sensitivity can be estimated at essentially no computational cost using only the objects used to estimate asymptotic standard errors. In a large class of remaining cases, sensitivity can be estimated using empirical influence statistics that are available at low computational cost and without any simulation or re-estimation of the model. Our measure can also be trivially extended to the case where the economic quantity of interest is not a parameter itself but a function of underlying parameters: an elasticity, a welfare effect, or a summary of a counterfactual simulation, for example.

Our main motivation for considering such a measure is to complement the discussions of identification that have become an increasingly important part of applied economics. Empirical papers often devote whole sections to discussing, formally or informally, the features of the data that make it possible to pin down the values of particular parameters.[1] One may understand many of these as attempts to establish constructive proofs of parametric or nonparametric identification in the sense of Matzkin (2007; 2013) and Berry and Haile (forthcoming).

Our sensitivity measure adds value to these discussions in two ways. First, the data features highlighted in a constructive proof of identification need not be the ones that actually drive the estimator used in practice. For example, most papers that offer discussions of nonparametric identification still take a parametric model to the data. We know of no generally accepted way to compare the behavior of the actual estimator to that of the hypothetical one constructed in the proof. We propose sensitivity as a way to fill this gap: sensitivity quantifies which features of the data the estimator depends on, and thus permits a comparison to the intuitive mapping laid out in the discussion of identification. Second, identification in a formal sense is a binary characteristic—either a feature of the data is sufficient to infer the value of a parameter or it is not. Yet applied researchers often stop short of this sharp claim, instead making more quantitative statements, such as that some aspect of the data is the "primary" or most "important" source of identification for a particular parameter.[2] To our knowledge no one has offered a formal way to understand these statements. We propose to re-cast them as claims about sensitivity: across alternative realizations of the data, variation in the estimated parameter will be explained only or mostly by variation in a particular data feature.

Our sensitivity measure also allows us to develop a formal argument for *why* we might want to know how a particular estimator maps data to parameters. We show that for a broad class of estimators, sensitivity to empirical moments is formally equivalent to sensitivity to small violations of identifying assumptions. A decision maker who entertains the possibility that the model's assumptions may not hold exactly can use sensitivity to map beliefs about assumptions to beliefs about the asymptotic behavior of the parameter estimates. We think that this reasoning explains at least part of the effort in the literature devoted to discussing specific sources of identification.

Throughout the paper, we draw on simple examples to illustrate the mechanics of our measure in familiar settings. Consider, for example, a two-stage least squares regression with dependent variable y, one endogenous regressor x, and two instruments z_1 and z_2, where the researcher wishes to assess the relative importance of the two instruments. To map this example to our framework, we define $\hat{\theta}$ to be the 2SLS coefficient and $\hat{\gamma}$ to be the respective covariances of z_1 and z_2 with

[1] See, for example, the discussions of identification in Einav et al. (2013), Berry et al. (2013), and Bundorf et al. (2012).

[2] For example, Goettler and Gordon (2011) write, "The demand-side parameters ... are primarily identified by the pricing moments, the Intel share equation moments, and the mean ownership quality relative to the frontier quality" (p. 1161). Lee (2013) writes that "[a parameter] is primarily identified" by a particular source of variation (p. 2978). Lim (2013) writes that particular dynamics in the data "play an important role in the separate identification of a judge's preference and reelection incentives" (p. 1378). Crawford and Yurukoglu (2012) write that "One may casually think of [a set of moments] as 'empirically identifying' [a set of parameters]" (p. 662).

y and x. We show that the sensitivity of $\hat{\theta}$ to the covariances involving z_1 and z_2 respectively is proportional to the respective coefficients from the first-stage regression. Moreover, sensitivity to these covariances is equivalent to sensitivity to small violations of the identifying assumptions $E(z_1'\varepsilon) = 0$ and $E(z_2'\varepsilon) = 0$, where ε is the disturbance in the second stage equation. What we recommend for more complex models thus boils down in this special case to the common practice of inspecting the first-stage estimates to learn which exclusion restrictions are most important in driving the second-stage estimate.

In the final sections of the paper, we present estimates of sensitivity for a number of empirical papers. We begin with an extended application to Berry et al.'s (1995) empirical model of the automobile market. We quantify the importance of demand-side and supply-side estimation moments in driving the estimated markup. We also show that estimates of a much simpler model—a logit with no unobserved heterogeneity—do a poor job of capturing the information in the data that pins down Berry et al.'s (1995) estimated parameters.

We turn next to an application to two models of intertemporal choice. Applying sensitivity to Gourinchas and Parker's (2002) model of lifecycle consumption and saving, we show how information on consumption at different ages permits simultaneous inference about time and risk preference. In an application to De Nardi et al.'s (2010) model of post-retirement saving, we show how a government-policy parameter not present in Gourinchas and Parker's (2002) model is pinned down by data on the asset holdings of rich and poor households.

After these detailed applications we present shorter applications to Goettler and Gordon's (2011) study of competition between AMD and Intel, DellaVigna et al.'s (2012) model of charitable giving, and Nikolov and Whited's (2014) model of corporate investment. For each paper we show the sensitivity of key parameters to the empirical moments used in estimation and compare our results to the authors' discussions of identification. In most cases, our analysis accords with the authors' stated intuitions, but we also find cases in which parameter estimates depend on information in the data that the authors did not highlight as important.

Our final applications are to Mazzeo's (2002) model of motel entry, Gentzkow et al.'s (2014) model of political competition in newspaper markets, and Gentzkow's (2007) model of competition between print and online newspapers. Because these papers use maximum likelihood estimators, we focus on estimating the sensitivity of their parameter estimates or key counterfactuals to descriptive statistics rather than to the estimation moments. We find that there is often a tight link between the estimates of structural parameters and the corresponding descriptive statistics. For example, in the case of Mazzeo's (2002) model, we show that estimates of an analogous linear regression model capture more than 80 percent of the information in the data that is used to estimate key parameters. This finding suggests a way in which sensitivity can be used to build up linear intuitions for the inner workings of nonlinear models.

An important limitation of our formal approach is that, because we focus on properties of the asymptotic distribution, the notion of sensitivity that we consider is intrinsically local. The approximations that we work with have the same mechanics and hence the same limitations as those commonly used to compute asymptotic standard errors. Generalizing our approach to more global exploration of model properties is conceptually straightforward but may be computationally expensive. In our concluding section, we provide some guidance on how a researcher might minimize computational costs in practice.

A second limitation is that the units of sensitivity are contingent on the units of $\hat{\gamma}$. We suggest a normalization that serves as a useful default for many practical applications but acknowledge that the appropriate scaling of sensitivity may be application-specific.

Our paper contributes to a methodological conversation about structural vs. "reduced-form" or "program evaluation" methods. At the center of this conversation is a perceived tradeoff between the realism of a model's economic assumptions and the transparency of its mapping from data to parameters.[3] Our sensitivity measure makes this tradeoff shallower by permitting a precise characterization of the dependence of a structural estimate on intuitive features of the data and on small violations of modeling assumptions.[4] Because our sensitivity measure correctly identifies cases in which only a subset of empirical moments is needed to answer a question of interest, sensitivity analysis may also be seen as a complement to the "sufficient statistics" approach of Chetty (2009), Einav et al. (2010), and Jaffe and Weyl (2013).

Our measure facilitates traditional sensitivity analysis (Leamer 1983) by showing how data map into parameters and by showing how small violations of model assumptions affect inference. In this sense our paper also complements recent research on inference in the presence of possibly misspecified exclusion restrictions (Conley et al. 2012; Nevo and Rosen 2012). Our work is also closely related to the large literature on sensitivity analysis for scientific models (Sobol 1993; Saltelli et al. 2008).[5]

The remainder of the paper is organized as follows. Section 2 defines our sensitivity measure. Section 3 discusses its properties and interpretation, and section 4 shows how to estimate it. Section 5 compares our approach to alternatives. Sections 6 and 7 apply the measure to several empirical papers. Section 8 concludes with a discussion of how to generalize the measure so that it is not

[3]Heckman (2010) writes that "The often complex computational methods that are required to implement [structural estimation] make it less transparent" (p. 358). Angrist and Pischke (2010) write that "in [Nevo's (2000)] framework, it's hard to see precisely which features of the data drive the ultimate results" (p. 21).

[4]In this sense our measure exploits the fact that structural models often "make the relationship between the economic model and the estimating equations transparent" (Pakes 2003, p. 193).

[5]Linear regression of model outputs on model inputs is a standard tool for model interrogation in the physical sciences. Our primary contribution is to show that the asymptotic properties of common estimators used in economics make it possible to perform such an analysis without repeatedly re-estimating or simulating the model, thus sparing substantial computational expense.

local to a particular sample.

2 Measure of Sensitivity and Sufficiency

2.1 Definitions

An econometrician possesses a sample of size n. She computes (i) a $(P \times 1)$ estimator $\hat{\theta}$ of a parameter θ with true value θ_0; and (ii) a $(J \times 1)$ vector of auxiliary statistics $\hat{\gamma}$ with population value γ_0. Both $\hat{\theta}$ and $\hat{\gamma}$ are functions of the data.[6]

We assume that there exist random variables $(\tilde{\theta}, \tilde{\gamma})$ such that

$$(1) \qquad \sqrt{n}\begin{pmatrix} \hat{\theta} - \theta_0 \\ \hat{\gamma} - \gamma_0 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \tilde{\theta} - \theta_0 \\ \tilde{\gamma} - \gamma_0 \end{pmatrix} \sim N(0, \Sigma),$$

for some finite Σ. We assume throughout that the submatrix $\Sigma_{\gamma\gamma}$ of Σ corresponding to the variance of $\tilde{\gamma}$ is nonsingular.

From equation (1) it follows that the conditional expectation of $\tilde{\theta}$ given $\tilde{\gamma}$ is linear. Letting $\Sigma_{\theta\gamma}$ denote the submatrix of Σ corresponding to the covariance of $\tilde{\theta}$ and $\tilde{\gamma}$, we have:

$$(2) \qquad \mathrm{E}\left(\tilde{\theta} - \theta_0 | \tilde{\gamma}\right) = \Sigma_{\theta\gamma}\Sigma_{\gamma\gamma}^{-1}\left(\tilde{\gamma} - \gamma_0\right).$$

Definition. *The* **sensitivity** *of $\hat{\theta}$ to $\hat{\gamma}$ is*

$$\Lambda = \Sigma_{\theta\gamma}\Sigma_{\gamma\gamma}^{-1}.$$

Sensitivity Λ is the expected coefficient from a regression of realizations of $\tilde{\theta}$ on realizations of $\tilde{\gamma}$. An element Λ_{pj} of Λ is the effect of changing the realization of a particular $\tilde{\gamma}_j$ on the expected value of a particular $\tilde{\theta}_p$, holding constant the other elements of $\tilde{\gamma}$.[7]

Note that, although we have defined sensitivity as a property of the statistics $\hat{\theta}$ and $\hat{\gamma}$, sensitivity depends only on these statistics' joint asymptotic distribution. Thus, two different estimators of θ that converge to the same $\tilde{\theta}$ must have the same sensitivity to $\hat{\gamma}$.

It is also helpful to be able to measure how much of the asymptotic variation in $\hat{\theta}$ is explained by $\hat{\gamma}$:

[6] Depending on the context, we use $\hat{\theta}$ and $\hat{\gamma}$ to refer to both the random variables with a sample of size n and the sequences of such random variables indexed by n.

[7] It follows from equation (2) that Λ_{pj}^2 is the partial derivative of the variance of $\mathrm{E}\left(\tilde{\theta}_p|\tilde{\gamma}\right)$ with respect to the variance of $\tilde{\gamma}_j$. In this sense, Λ captures not only the impact of $\tilde{\gamma}$ on $\tilde{\theta}$, but also the impact of *uncertainty* about γ_0 on uncertainty about θ_0.

Definition. *The* **sufficiency** *of $\hat{\gamma}$ for an element $\hat{\theta}_p$ of $\hat{\theta}$ is*

$$\Delta_p = \frac{\mathrm{Var}\left(\mathrm{E}\left(\tilde{\theta}_p|\tilde{\gamma}\right)\right)}{\mathrm{Var}\left(\tilde{\theta}_p\right)} = \frac{\left(\Lambda\Sigma_{\gamma\gamma}\Lambda'\right)_{pp}}{\left(\Sigma_{\theta\theta}\right)_{pp}}.$$

We let Δ denote the column vector of Δ_p.

We will say that $\hat{\gamma}$ is **sufficient for** $\hat{\theta}_p$ *if $\Delta_p = 1$ and that $\hat{\gamma}$ is* **sufficient for** $\hat{\theta}$ *if $\Delta = 1$.*

The value $\Delta_p \in [0,1]$ is the probability limit of the R^2 of a regression of realizations of $\tilde{\theta}_p$ on realizations of $\tilde{\gamma}$, as the number of realizations grows large.

Most of our applications will have $\Delta = 1$ or $\Delta \approx 1$, and as we discuss below we believe Λ is most informative when this is the case. When $\Delta = 1$, $\hat{\gamma}$ fully determines $\hat{\theta}$ asymptotically. When $\Delta \approx 1$, $\hat{\gamma}$ is "almost sufficient" for $\hat{\theta}$ in the sense that, asymptotically, knowing $\hat{\gamma}$ allows the econometrician to predict $\hat{\theta}$ with little error. Sufficiency Δ_p provides a quantitative way to evaluate a researcher's claim that some low-dimensional representation of the data $\hat{\gamma}$ captures the key information that drives her estimator $\hat{\theta}_p$.[8]

The applications we will discuss are all examples of minimum distance estimators (MDE), a class that includes generalized method of moments (GMM), maximum likelihood (MLE), and classical minimum distance (CMD), as well as simulated analogues such as simulated minimum distance (SMD) and simulated method of moments (SMM). Formally:

Definition. $\hat{\theta}$ *is a* **minimum distance estimator** *(MDE) if we can write*

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \hat{g}(\theta)' \hat{W}_g \hat{g}(\theta), \tag{3}$$

where $\hat{g}(\theta)$ is a function of parameters and data, $\hat{g}(\theta_0)$ is asymptotically normal with mean 0 and variance Ω_{gg}, and the weight matrix $\hat{W}_g$ is positive semi-definite, as is its probability limit W_g. Except where stated we will assume further regularity conditions so that $\hat{\theta}$ is consistent for θ_0 and asymptotically normal with asymptotic variance $\left(G'W_gG\right)^{-1} G'W_g\Omega_{gg}W_gG\left(G'W_gG\right)^{-1}$, where G is the Jacobian of an appropriate limit of $\hat{g}$ evaluated at θ_0.[9]

We will say that Λ is **sensitivity to moments** *when $\hat{\theta}$ is an MDE and $\hat{\gamma} = \hat{g}(\theta_0)$ is the vector of estimation moments evaluated at the true paramter value.*

An immediate consequence of these definitions is:

[8]We abuse the probability-theoretic term "sufficient" to emphasize that, when $\Delta = 1$, asymptotically $\hat{\theta}$ contains no information about the data that is not also in $\hat{\gamma}$.

[9]We allow some flexibility in the definition of G here so that our definition of MDE includes both standard cases where $\hat{g}(\theta)$ is a smooth function of parameters (as in GMM or CMD), and cases where $\hat{g}(\theta)$ is not smooth (as in SMM). For the precise regularity conditions and definition of G, see Newey and McFadden (1994) Theorem 3.2 for the smooth case and Theorem 7.2 for the non-smooth case.

Remark 1. If $\hat{\theta}$ is an MDE and Λ is sensitivity to moments, then $\Delta = 1$ and $\Lambda = -\left(G'W_gG\right)^{-1} G'W_g$.

Note that although we have focused attention for simplicity on the case in which $\hat{\theta}$ is a parameter, it is easy to extend our approach to cases in which the quantity of interest is a function of the underlying parameters, such as a welfare calculation or other unobservable or counterfactual implication. To emphasize this possibility, we establish the following definition:

Definition. *We will say that $c\left(\hat{\theta}\right)$ is a* **counterfactual** *if the function $c()$ does not depend on the data, and is continuously differentiable with non-zero gradient C at θ_0.*

It is immediate from the delta method that:

Remark 2. The sensitivity of counterfactual $c\left(\hat{\theta}\right)$ to statistics $\hat{\gamma}$ is given by $C\Lambda$, where Λ is the sensitivity of $\hat{\theta}$ to $\hat{\gamma}$.

It is also easy to extend our approach to consider sensitivity to a function of the statistics $\hat{\gamma}$:

Definition. *We will say that $a(\hat{\gamma})$ is a* **transformed statistic** *if the function $a()$ does not depend on the data, and is continuously differentiable with non-zero gradient A at γ_0.*

It is again immediate from the delta method that:

Remark 3. The sensitivity of $\hat{\theta}$ to transformed statistic $a(\hat{\gamma})$ is given by $\Sigma_{\theta\gamma}A'\left(A\Sigma_{\gamma\gamma}A'\right)^{-1}$.

In our applications, we use this result, for example, to consider sensitivity of an MDE to an average of conceptually related estimation moments.

2.2 Preliminaries

We now establish some useful properties of sensitivity and sufficiency. These help to build intuition and to streamline derivations in the examples we present below.

Our main formal result is that when $\hat{\gamma}$ is sufficient for $\hat{\theta}_p$, sensitivity can be interpreted as the derivative of an appropriately defined function of $\hat{\gamma}$:

Proposition 1. *The statistics $\hat{\gamma}$ are sufficient for $\hat{\theta}_p$ if and only if there exists a continuously differentiable function $f()$ with nonzero gradient at γ_0 such that $\sqrt{n}\left[\hat{\theta}_p - f(\hat{\gamma})\right]$ converges in probability to* 0. *If such an $f()$ exists, its partial derivative at γ_0 is $\Lambda_{p\cdot}$.*

Proof. Suppose that $\hat{\gamma}$ is sufficient for $\hat{\theta}_p$. Define $f()$ such that $f(\hat{\gamma}) = \theta_{0p} + \Lambda_{p\cdot}(\hat{\gamma} - \gamma_0)$. That $\sqrt{n}\left[\hat{\theta}_p - f(\hat{\gamma})\right] \xrightarrow{p} 0$ then follows from standard limit results and properties of normal random variables. The other properties of $f()$ are immediate.

Now suppose that there exists an $f()$ satisfying the given conditions. Observe that

$$\sqrt{n}\begin{pmatrix} f(\hat{\gamma}) - f(\gamma_0) \\ \hat{\gamma} - \gamma_0 \end{pmatrix} = \sqrt{n}\begin{pmatrix} \hat{\theta}_p - \theta_{0p} \\ \hat{\gamma} - \gamma_0 \end{pmatrix} - \sqrt{n}\begin{pmatrix} \hat{\theta}_p - f(\hat{\gamma}) \\ 0 \end{pmatrix}.$$

The first term on the right-hand side converges in distribution to $\left(\tilde{\theta}_p - \theta_{0p}, \tilde{\gamma} - \gamma_0\right)'$. The second term on the right-hand side converges in probability to zero by construction. Therefore the term on the left-hand side converges in distribution to $\left(\tilde{\theta}_p - \theta_{0p}, \tilde{\gamma} - \gamma_0\right)'$. That $\Delta_p = 1$ and that $f()$ has partial derivative at γ_0 given by $\Lambda_{p\cdot}$ then follow from standard limit results and the definitions of Δ_p and Λ. $\square$

A case of special interest is when $\hat{\theta}_p = f(\hat{\gamma})$ for some suitable function $f()$:

Definition. *The statistics $\hat{\gamma}$ are* ***strongly sufficient*** *for $\hat{\theta}_p$ if there exists continuously differentiable function $f()$, not dependent on θ_0, such that $\hat{\theta}_p = f(\hat{\gamma})$ and $f()$ has nonzero gradient at γ_0.*

Corollary 1. *If $\hat{\gamma}$ is strongly sufficient for $\hat{\theta}_p$ then $\hat{\gamma}$ is sufficient for $\hat{\theta}_p$.*

Strong sufficiency obtains, for example, when $\hat{\theta}$ is a CMD estimator with fixed weight matrix W and $\hat{\gamma}$ is the vector of empirical moments used in estimation.

Proposition 1 is also useful in an important class of cases in which sufficiency obtains although strong sufficiency does not:

Corollary 2. *Suppose that we can write $\hat{\theta}_p - \theta_{0p} = \hat{M} \cdot [h(\hat{\gamma}) - h(\gamma_0)]$ for some function $h()$ that is continuously differentiable with non-zero partial derivatives at γ_0, and for some function $\hat{M}$ of the data that has probability limit M_0. Then $\Delta_p = 1$ and $\Lambda_{p\cdot} = M_0 \nabla h(\gamma_0)$.*

Proof. Observe that $\sqrt{n}\left(\left(\hat{\theta}_p - \theta_{0p}\right) - M_0[h(\hat{\gamma}) - h(\gamma_0)]\right) = \left(\hat{M} - M_0\right)\left(\sqrt{n}[h(\hat{\gamma}) - h(\gamma_0)]\right) \xrightarrow{p} 0$, where the last step follows from Slutsky's theorem. The result then follows from proposition 1. $\square$

As we show below, corollary 2 provides a convenient method to derive Λ in familiar examples.

2.3 Examples

In this section, we study several "pen and paper" examples, many of which are familiar. Our goal is not to shed new light on these examples. Rather, we want to see whether sensitivity delivers reasonable intuitions in simple cases that we already understand. If so, this may give us greater confidence in applying sensitivity to more complex models.

Example. (OLS) $\hat{\theta} = \begin{bmatrix} \hat{\alpha} & \hat{\beta} \end{bmatrix}'$ is the constant term and coefficient from an OLS regression of y on a scalar x. We assume standard conditions for the consistency and asymptotic normality of $\hat{\theta}$. Define $\hat{\gamma} = \begin{bmatrix} \hat{\mu}_y & \hat{\sigma}_{xy} & \hat{\sigma}_x^2 & \hat{\mu}_x \end{bmatrix}'$, where $\hat{\mu}_y$ is the sample mean of y, $\hat{\sigma}_{xy}$ is the sample covariance of x and y, $\hat{\sigma}_x^2$ is the sample variance of x, and $\hat{\mu}_x$ is the sample mean of x. We can write $\hat{\alpha} = \hat{\mu}_y - \hat{\beta}\hat{\mu}_x$ and $\hat{\beta} = \hat{\sigma}_{xy}/\hat{\sigma}_x^2$, so by proposition 1 $\Delta = 1$ and we can solve for Λ by evaluating the partial derivatives of the estimates at the population values of $\hat{\gamma}$. Focusing on the first two columns of Λ, which give sensitivity to $\hat{\mu}_y$ and $\hat{\sigma}_{xy}$ respectively, we have:

$$\Lambda = \begin{bmatrix} 1 & -\frac{\mu_x}{\sigma_x^2} & \dots \\ 0 & \frac{1}{\sigma_x^2} & \dots \end{bmatrix},$$

where μ_x and σ_x^2 are the population mean and variance of x. Consistent with intuition, we find that when the mean of x is zero, the constant $\hat{\alpha}$ is sensitive only to $\hat{\mu}_y$ and $\hat{\beta}$ is sensitive only to $\hat{\sigma}_{xy}$. When the mean of x is not zero, $\hat{\alpha}$ is also sensitive to $\hat{\sigma}_{xy}$ because this affects $\mathrm{E}\left(\hat{\beta}x\right)$. Note that it is straightforward to generalize the example to multivariate regression.[10]

Example. (2SLS) $\hat{\theta}$ is the coefficient $\hat{\beta}$ from a two-stage least squares regression with dependent variable y, one endogenous regressor x, and two instruments z_1 and z_2. For simplicity, we assume all variables are mean zero. We define $\hat{\gamma} = \frac{1}{n}\begin{bmatrix} y'z & x'z \end{bmatrix}'$, with $z = \begin{bmatrix} z_1 & z_2 \end{bmatrix}$. We assume standard conditions for consistency and asymptotic normality of $\hat{\theta}$. Letting $\hat{x} = z\hat{\phi}$ be the predicted values from the first stage (where $\hat{\phi} = (z'z)^{-1} z'x$ with probability limit ϕ), we have:

$$\begin{aligned} \hat{\beta} - \beta &= \left(\hat{x}'\hat{x}\right)^{-1} \hat{x}' \left(y - \hat{x}\beta\right) \\ &= \left(\hat{x}'\hat{x}\right)^{-1} \hat{\phi}' \left(z'y - z'x\beta\right). \end{aligned}$$

where β is the true value of $\hat{\beta}$. Letting $\Omega_{\hat{x}\hat{x}} = \operatorname{plim} \frac{1}{n}\hat{x}'\hat{x}$, by corollary 2 we have that $\Delta = 1$ and

$$\Lambda = \Omega_{\hat{x}\hat{x}}^{-1} \begin{bmatrix} \phi' & -\phi'\beta \end{bmatrix},$$

where $\phi = \begin{bmatrix} \phi_1 & \phi_2 \end{bmatrix}'$. Consistent with intuition, the sensitivities of $\hat{\beta}$ to $\frac{1}{n}z_1'y$ and $\frac{1}{n}z_2'y$ respectively are proportional to the first-stage coefficients ϕ_1 and ϕ_2. The sensitivities to $\frac{1}{n}z_1'x$ and $\frac{1}{n}z_2'x$ have the opposite sign and are scaled by β. Note that the example can be generalized to a model with multiple endogenous regressors.[11]

[10] Let independent variables be $X = [x_1, ..., x_J]$ and let $\hat{\gamma} = \frac{1}{n}\begin{bmatrix} y'X & x_1'x_1 & x_1'x_2 & \cdots & x_j'x_k & \cdots & x_J'x_J \end{bmatrix}'$ with $j \in \{1, ..., J\}$ and $k \in \{j, ..., J\}$ (so there are no redundant elements). Then $\Delta = 1$ and

$$\Lambda = \begin{bmatrix} \Omega_{XX}^{-1} & \cdots \end{bmatrix},$$

where Ω_{XX} is the probability limit of $\frac{1}{n}X'X$. The columns of Λ equal to Ω_{XX}^{-1} give the sensitivity of $\hat{\beta}$ to $\frac{1}{n}y'X$.

[11] Let endogenous regressors be $X = [x_1, ..., x_J]$ and instruments be $Z = [z_1, ..., z_K]$ with $K \geq J$. If we let $\hat{\gamma} = \frac{1}{n}\begin{bmatrix} y'Z & x_1'z_1 & x_1'z_2 & \cdots & x_j'z_k & \cdots & x_J'z_K \end{bmatrix}'$ where j, k are defined so there are no redundant elements (in case some elements of X are included in Z), then

$$\Lambda = \begin{bmatrix} \Omega_{\hat{X}\hat{X}}^{-1}\phi' & \cdots \end{bmatrix},$$

Example. (Sample Mean) $\hat{\theta}$ is the average of a scalar variable y in a random sample. The population mean of y is θ_0. The population consists of J groups indexed by j, each of which represents share ρ_j of the population, and we wish to assess the sensitivity of $\hat{\theta}$ to the sample means $\hat{\gamma}_j$ in each group. Note that $\hat{\theta} = \hat{\rho}'\hat{\gamma}$, where $\hat{\rho} = [\hat{\rho}_1, ..., \hat{\rho}_J]'$ is the vector of sample shares. In this case, we cannot write that $\hat{\theta} = f(\hat{\gamma})$ because the mapping from $\hat{\gamma}$ to $\hat{\theta}$ depends on the sample shares $\hat{\rho}$. However, we can write $\hat{\theta} - \theta_0 = \hat{\rho}'\left(\hat{\gamma}_j - \theta_0\right)$, and so invoke corollary 2 to find that $\Delta = 1$ and $\Lambda = \rho$. Consistent with intuition, the sensitivity of $\hat{\theta}$ to the sample mean of group j is proportional to its population share ρ_j.

3 Interpretation of Sensitivity and Sufficiency

3.1 Identification

Figure 1 shows the dramatic increase in the number of articles published in top economic journals containing a claim that some estimator is "identified by" some feature of the data. In 2013, the *American Economic Review* published 14 empirical papers that include structural models; of these, 10 contain a section or subsection with "identification" in the title, while two others provide similar discussion without breaking it out into a separate subsection.[12] Consistent with figure 1, these discussions typically relate specific variation or data features to the identification of specific parameters.[13]

We can interpret these discussions as efforts to build a constructive argument for identification in the sense defined by Matzkin (2007; 2013). Such an argument reveals a functional that maps the true distribution of the data into the economic quantities of interest. As Matzkin (2013) observes, such a construction provides a candidate estimator of the quantities of interest, obtained by simply applying the functional to the observed data. Identification is "nonparametric" if the functional is valid without restrictions on functional forms other than those derived from economic theory.

To make this precise, we introduce a stripped-down version of Matzkin's (2013) framework. A *model* is a family of primitive functions and distributions S (e.g., utility or production functions, distributions of unobservable shocks, etc.), and a mapping that for any $\zeta \in S$ determines the joint population distribution of all observable variables. Two primitives ζ and ζ' are *observationally*

where $\Omega_{\hat{X}\hat{X}}$ is the probability limit of $\frac{1}{n}\hat{X}'\hat{X}$ and ϕ is the matrix of first-stage coefficients.

[12]The online appendix lists these articles and shows how we classify them.

[13]For example, Barseghyan et al. (2013) write "given three or more deductible options, it is exogenous variation in premiums for a fixed [claim probability] that allows us to pin down [the coefficient of absolute risk aversion] and [the probability distortion function]" (p. 2511). Fan (2013) writes that "Identification of [the diminishing utility parameter] comes from the variation in the number of newspapers in a county" (p. 1610). Kawai and Watanabe (2013) write "we use the systematic difference between the predicted vote share and the actual vote share to partially identify the fraction of strategic voters" (p. 643).

equivalent if they imply the same joint distribution. An *economic quantity* c (e.g., a vector of elasticities, welfare effects, etc.) is a functional of ζ,[14] and is *identified* if for any observationally equivalent ζ and ζ', $c(\zeta) = c(\zeta')$.

This definition says that c is identified when the complete distribution of the observed data is enough to pin down c's value. We can naturally extend this definition to say what it means for c to be identified by a particular statistic or vector of features γ of the population distribution of the observable data, noting that any such vector can itself be written as a functional $\gamma(\zeta)$ of the underlying primitives ζ.

Definition. *Economic quantity c is **identified by** data features γ if for any primitives ζ and ζ' such that $\gamma(\zeta) = \gamma(\zeta')$, we have $c(\zeta) = c(\zeta')$.*

If economic quantity c is identified by data features γ, there must be some function Φ such that $c = \Phi(\gamma)$. A natural estimator for c is then $\hat{c} = \Phi(\hat{\gamma})$.

Applied researchers' treatments of identification depart from this roadmap in two main ways. First, the estimator they take to the data is usually different from the Φ implicitly defined by their discussion of identification. For example, it is common for authors to provide an argument for nonparametric identification, then go on to use a parametric model due to data limitations or other practical concerns.[15] The estimators used in practice may therefore depend on features of the data other than γ, and may be valid only under stronger assumptions than those needed to justify Φ. Second, researchers sometimes discuss identification as a quantitative rather than a qualitative property, arguing that a given feature γ "mostly" or "primarily" identifies a given economic quantity c.[16] Such claims do not have a formal meaning in Matzkin's framework.

Sensitivity and sufficiency complement analysis of identification and speak directly to these two common departures. As quantitative measures, they can support the claim that a particular estimated quantity $\hat{c}_p$ (or parameter $\hat{\theta}_p$) depends *primarily* on a particular data feature $\hat{\gamma}_j$ ($\Delta_p \approx 1$ and $\Lambda_{pk} \approx 0$ for $k \neq j$). Yet they also retain a tight connection with identification in the limiting case of strong sufficiency:

[14] Matzkin (2007; 2013) refers to such quantities as "features" of a model, and uses a definition that includes the case where the object of interest is a distribution function or other high-dimensional object. We use different language to avoid confusion with the term "features of the data" which we use as a synonym for sample statistics above.

[15] Einav et al., "Selection on Moral Hazard," (2013) write that "our actual data depart from the ideal data ... we thus make additional parametric assumptions to aid us in identification" (p. 201). Kawai and Watanabe (2013) write that "while our identification argument does not rely on the particular functional form [of preferences], our estimation does impose these functional forms" (p. 639).

[16] Lee (2013) writes that "[the price sensitivity of hardware] is primarily identified as hardware sales responds to variation in software utility (both within and across platforms), which in turn is caused by variation in software availability and sales over time" (pp. 2978-2979). Lim (2013) writes that "the dynamics of sentencing relative to changes in political climate and the variation in judges' sentencing across different stages of their career play an important role in the separate identification of a judge's preference and reelection incentives" (p. 1378).

Proposition 2. *Suppose that there is a statistic $\hat{\gamma}$ that is consistent for data feature γ and asymptotically normal. Then the following are equivalent: (i) economic quantity c is identified by γ through a continuously differentiable function Φ; (ii) there exists consistent and asymptotically normal estimator $\hat{c}$ for which $\hat{\gamma}$ is strongly sufficient.*

Proof. That (i) implies (ii) follows by setting $\hat{c} = \Phi(\hat{\gamma})$. To see that (ii) implies (i), pick ζ, ζ' such that $\gamma(\zeta) = \gamma(\zeta') = \gamma_0$. There is some consistent estimator $\hat{c}$ and continuously differentiable function $f()$ such that $\hat{c} = f(\hat{\gamma})$. By standard limit results $\hat{c} \xrightarrow{p} f(\gamma_0)$. Because $\hat{c}$ is consistent, $f(\gamma_0) = c(\zeta) = c(\zeta')$, and we can set $\Phi = f$. $\square$

A crucial distinction is that sensitivity and sufficiency are properties of an estimator, whereas identification is a property of the underlying model. This means that when a given model admits many different estimators that in turn depend on different data features γ, identification arguments alone cannot show which features are being exploited in practice. Sensitivity and sufficiency, on the other hand, do show which features are being exploited, and they capture the effect of any parametric assumptions that are used in estimation but not in the constructive proof of identification.

In fact, *all* models admit multiple estimators which differ in their sensitivities to different data features, including estimators that depend on features unrelated to model identification:

Remark 4. Suppose that the hypothesis of proposition 2 holds, feature γ_j identifies quantity c_p, and feature γ_k does not. Then there exists a consistent and asymptotically normal estimator $\hat{c}_p$ that is sensitive to both $\hat{\gamma}_j$ and $\hat{\gamma}_k$.

Proof. By proposition 2 there exists $\hat{c}_p^*$ for which $\hat{\gamma}_j$ is strongly sufficient. Then $\hat{c}_p = \hat{c}_p^* + \alpha(\hat{\gamma}_k - \gamma_k)$ is sensitive to both $\hat{\gamma}_j$ and $\hat{\gamma}_k$ for some appropriate $\alpha > 0$. $\square$

Remark 4 shows the limits to proposition 2: while there always exists *some* estimator for which an identifying feature is sufficient, there also exist some estimators for which the identifying feature is not sufficient. It is therefore valuable to show what statistics an estimator is sensitive to, even if it can be shown that identification requires only a subset of the corresponding features.

3.2 Misspecification

If the model is correctly specified and the estimator has desirable econometric properties (consistency, efficiency, etc.), we might not care which features identify which parameters. We believe the reason that researchers devote so much attention to these issues is that empirical models are typically not correctly specified. Important substantive assumptions such as exclusion restrictions are almost always open to reasonable doubt. And nearly every empirical model includes some "whimsical assumptions" that are made purely for convenience and are not defensible on economic grounds (Leamer 1983). Knowing which data features drive the estimates provides valuable insight into the relative importance of these assumptions, and thus the robustness of the results.

In this subsection, we show formally that the sensitivity of an estimator to particular statistics is informative about the sensitivity of the estimator to violations of particular assumptions. In showing this, we justify the emphasis in applied work on discussing the way an estimator maps data to parameters, and we offer an additional reason to be interested in estimating sensitivity.

We motivate our analysis with the following example:

Example. (Standard Deviation of Exponential Random Variable) We wish to use a random sample to estimate the population standard deviation θ_0 of an exponentially distributed random variable x. Let μ_x and σ_x be the population mean and standard deviation, respectively, with sample analogues $\hat{\mu}_x$ and $\hat{\sigma}_x$. Clearly σ_x identifies θ_0 but so too does μ_x because $\sigma_x = \mu_x$ for x exponential. Let $\hat{\theta}$ be an MDE with diagonal weight matrix and estimation moments $\hat{g}(\theta_0) = \begin{bmatrix} (\hat{\mu}_x - \theta_0) & (\hat{\sigma}_x - \theta_0) \end{bmatrix}'$, which implies that $\hat{\theta} = \hat{w}\hat{\mu}_x + (1-\hat{w})\hat{\sigma}_x$ for some $\hat{w} \in [0,1]$ with probability limit w. If Λ is sensitivity to moments, $\Lambda = \begin{bmatrix} w & 1-w \end{bmatrix}$. If $\hat{\theta}$ is MLE, then $w = 1$.

That the estimate of the population standard deviation depends on the sample mean does not bother us if we are confident that x is exponential. But if we entertain the possibility that x is not exponential, and therefore that the population mean and standard deviation may be quite different, then the more sensitive $\hat{\theta}$ is to $\hat{\mu}_x$, the more concerned we will be about using $\hat{\theta}$ as an estimator.

More precisely, suppose that x is not exponential, and $(\hat{\sigma}_x - \theta_0) \xrightarrow{p} 0$ but $(\hat{\mu}_x - \theta_0) \xrightarrow{p} \epsilon \neq 0$. Then $\left(\hat{\theta} - \theta_0\right) \xrightarrow{p} w\epsilon$. Under the maximum likelihood estimator, $w = 1$ and misspecification translates one-for-one into asymptotic bias; under a method-of-moments estimator with $w = 0$, misspecification has no asymptotic impact. Therefore w, the sensitivity of $\hat{\theta}$ to $\hat{\mu}_x$, measures the impact of misspecification of the moment condition on asymptotic bias.

Put differently, we may think of Λ as a linear map that translates misspecification of assumptions (moment conditions) into asymptotic bias of the estimator. This property of Λ holds in a large class of MDEs. To see this we must first generalize our definition of an MDE to allow for the possibility of misspecification:

Definition. *The estimator $\hat{\theta}_\epsilon$ is a* **misspecified minimum distance estimator** *(MMDE) if we can write*

$$\hat{\theta}_\epsilon = \underset{\theta \in \Theta}{\arg\min}\, \hat{g}(\theta)' \hat{W}_g \hat{g}(\theta),$$

with $\hat{g}(\theta_0) \xrightarrow{p} \epsilon$. We assume that the function $\hat{g}(\theta)$ is continuously differentiable and converges uniformly in probability to a continuously differentiable function $g(\theta)$, the gradient $\hat{G}(\theta)$ of $\hat{g}(\theta)$ converges uniformly in probability to the gradient $G(\theta)$ of $g(\theta)$, the weight matrix $\hat{W}_g$ is positive *semi-definite, as is its probability limit W_g, and $g(\theta)' W_g g(\theta)$ is uniquely minimized at some θ_ϵ in the interior of the (compact) domain of θ.*

With this definition in hand the intuition from the exponential example generalizes easily:

Proposition 3. *Suppose $\hat{\theta}_\epsilon$ is a MMDE with $\hat{g}(\theta)$ affine in θ. Then $\left(\hat{\theta}_\epsilon - \theta_0\right) \xrightarrow{p} \Lambda\epsilon$, where $\Lambda \equiv -\left(G(\theta_0)' W_g G(\theta_0)\right)^{-1} G(\theta_0)' W_g$ is the sensitivity to moments at θ_0.*

Proof. Because $\hat{g}(\theta)$ is affine we can write $\hat{g}(\theta) = \hat{G}\theta + \hat{s}$ for gradient $\hat{G}$ and some statistic $\hat{s}$ not dependent on θ. An immediate implication is that $\hat{g}\left(\hat{\theta}_\epsilon\right) = \hat{g}(\theta_0) + \hat{G}\left(\hat{\theta}_\epsilon - \theta_0\right)$. The first-order condition for $\hat{\theta}_\epsilon$ is $\hat{G}' \hat{W}_g \hat{g}\left(\hat{\theta}_\epsilon\right) = 0$. Combining the preceding expressions, we have $\left(\hat{\theta}_\epsilon - \theta_0\right) = \hat{\Lambda}\hat{g}(\theta_0)$ where $\hat{\Lambda} = -\left(\hat{G}'\hat{W}_g\hat{G}\right)^{-1} \hat{G}'\hat{W}_g$. Note that $\hat{g}(\theta_0) \xrightarrow{p} \epsilon$, $\hat{G} \xrightarrow{p} G(\theta_0)$, and $\hat{W}_g \xrightarrow{p} W_g$, so $\hat{\Lambda} \xrightarrow{p} \Lambda$ and $\left(\hat{\theta}_\epsilon - \theta_0\right) \xrightarrow{p} \Lambda\epsilon$ by the continuous mapping theorem. $\square$

Proposition 3 applies in our 2SLS example with one endogenous regressor and two instruments. Either instrument alone is sufficient for identification of θ_0 but estimators differ in the weight they give to each. We can think of the 2SLS estimator as an MDE with estimation moments $\hat{g}(\beta) = \frac{1}{n} z' (y - x\beta)$. Sensitivity Λ allows us to translate beliefs about violations of the identifying assumptions $\mathrm{E}(z'(y - x\beta)) = 0$ into beliefs about the asymptotic bias of $\hat{\beta}$.[17]

Building intuition about robustness to misspecification is especially difficult in nonlinear models where, unlike in our exponential and 2SLS examples, estimators cannot typically be written as closed-form functions of intuitive statistics. Importantly, the interpretation of Λ suggested by proposition 3 generalizes to nonlinear models under small misspecification:

Proposition 4. *Let $\hat{\theta}$ be an MDE with weights $\hat{W}_g$, distance function $\hat{g}(\theta)$, and sensitivity to moments Λ. Suppose there is a sequence of MMDEs $\hat{\theta}_\epsilon$ indexed by $\epsilon \to 0$ with weights $\hat{W}_g$ and distance functions $\hat{g}(\theta) + \epsilon$. Then for any ϵ, $\hat{\theta}_\epsilon - \theta_0 \xrightarrow{p} \Lambda_\epsilon \epsilon$ for some Λ_ϵ, with $\Lambda_\epsilon \to \Lambda$ as $\epsilon \to 0$.*

Proof. Pick some ϵ. By assumption $\hat{\theta}_\epsilon \xrightarrow{p} \theta_\epsilon$. Taking the probability limit of the first-order condition for $\hat{\theta}_\epsilon$ yields $G(\theta_\epsilon)' W_g (g(\theta_\epsilon) + \epsilon) = 0$. Note that $g(\theta_\epsilon) = g(\theta_0) + G\left(\bar{\theta}_\epsilon\right)(\theta_\epsilon - \theta_0)$ where $\bar{\theta}_\epsilon$ is a mean value and $g(\theta_0) = 0$. Combining the preceding expressions, we have $(\theta_\epsilon - \theta_0) = \Lambda_\epsilon \epsilon$ where

$$\Lambda_\epsilon = -\left(G(\theta_\epsilon)' W_g G\left(\bar{\theta}_\epsilon\right)\right)^{-1} G(\theta_\epsilon)' W_g.$$

Therefore $\left(\hat{\theta}_\epsilon - \theta_0\right) \xrightarrow{p} \Lambda_\epsilon \epsilon$. Because $g()$ and $G()$ are continuous in θ, θ_ϵ is continuous in ϵ, which implies $\theta_\epsilon \to \theta_0$ as $\epsilon \to 0$. Together with the continuity of $G()$ this yields $\Lambda_\epsilon \to \Lambda$. $\square$

[17] In particular, for 2SLS, sensitivity to moments Λ is $\mathrm{plim}\left(\left(\frac{x'z}{n}\left(\frac{z'z}{n}\right)^{-1}\frac{z'x}{n}\right)^{-1}\left(\frac{x'z}{n}\right)\left(\frac{z'z}{n}\right)^{-1}\right)$ (Conley et al. 2012).

In other words, Λ is a linear map from misspecification to bias, local to a correctly specified model. This means Λ can serve as a fast tool for obtaining intuitions about the sensitivity of an estimator to various types of misspecification. Of course, in any given instance a Bayesian with a fully-specified prior over the space of models could in principle compute the parameter value implied by each model and then form the correct posterior. In practice, however, this is usually prohibitively costly, whereas computing a valid estimate of Λ is trivial.

3.3 Descriptive Statistics

Much of the preceding discussion focuses on cases, such as sensitivity to moments, in which $\hat{\gamma}$ is sufficient for $\hat{\theta}$. It is also common for researchers to discuss the relationship between structural parameters and model-free "descriptive statistics."[18] In principle, indirect inference makes it possible to base estimation solely on such descriptive statistics (Gourieroux et al. 1993; Smith 1993). In practice, either for econometric or computational reasons, researchers often choose not to base estimation directly on descriptive statistics. In such cases the link between parameter estimates and descriptive statistics is typically not formalized.

Sensitivity and sufficiency quantify the relationship between descriptive statistics and parameter estimates. Suppose that $\hat{\theta}_p$ estimates a structural parameter of interest and that $\hat{\gamma}_j$ is a descriptive statistic—say, a regression coefficient—that seems intuitively useful for estimating parameter p. A researcher can test this claim by computing the sensitivity of $\hat{\theta}_p$ to $\hat{\gamma}_j$: if sensitivity is very small or has an unexpected sign, then the researcher's intuition is likely misleading. Likewise, analysis of sufficiency allows a test of whether a vector of statistics $\hat{\gamma}$ captures most of the information in the data that is used to estimate structural parameters. If $\hat{\gamma}$ is almost sufficient for $\hat{\theta}$, then in a large sample a researcher in possession of $\hat{\gamma}$ (and Λ) can predict $\hat{\theta}$ very well. We will show examples of real applications in which simple descriptive statistics can be used to "emulate" structural estimates in exactly this way.

It can also happen, of course, that the vector $\hat{\gamma}$ of interesting descriptive statistics has low sufficiency for $\hat{\theta}$. In that case, $\hat{\theta}$ depends on the data in ways that are not well captured by $\hat{\gamma}$. This raises the possibility of omitted variable bias: because Λ is just a measure of covariance, it will "pick up" the influence on $\hat{\theta}$ of statistics correlated with those in $\hat{\gamma}$. We therefore think Λ is most reliable and meaningful when Δ is close to 1, though it remains well-defined even in cases with low Δ.

[18] Einav et al., "Selection on Moral Hazard," (2013), for example, relate the identification of the moral hazard parameters in their model to a preceding difference-in-difference analysis housed in a section called "Descriptive Evidence of Moral Hazard" (p. 192). Lim (2013) relates the identification of key parameters in her model to evidence contained in a data plot that is "not dependent on any particular modeling decision or estimated parameter values of the model" (p. 1378).

4 Estimation of Sensitivity and Sufficiency

In this section we show that it is easy to estimate sensitivity Λ and sufficiency Δ even for computationally difficult models. We focus on the case in which $\hat{\theta}$ is an MDE, a class that as noted above encompasses a large range of estimators including GMM and MLE. We do not explicitly discuss the case in which the magnitude of interest is a counterfactual or the statistics of interest are transformations of $\hat{\gamma}$, but we note that the results below extend immediately to those cases following the logic of remarks 2 and 3.

4.1 Sensitivity to Moments

If Λ is sensitivity to moments, then by remark 1, we know that $\Delta = 1$ and $\Lambda = -\left(G' W_g G\right)^{-1} G' W_g$. By assumption the researcher possesses $\hat{W}_g$, a consistent estimate of W_g. A consistent estimate $\hat{G}$ of G is typically in hand to estimate the asymptotic variance of $\hat{\theta}$.[19] Therefore in typical applications estimating Λ imposes no additional computational burden beyond the estimation of the asymptotic variance.

Remark 5. If $\hat{\theta}$ is an MDE and $\hat{G}$ is a consistent estimate of G then $\hat{\Lambda} = -\left(\hat{G}' \hat{W}_g \hat{G}\right)^{-1} \hat{G}' \hat{W}_g$ is a consistent estimate of sensitivity to moments. If the researcher has computed a plug-in estimator of $\mathrm{Var}\left(\tilde{\theta} - \theta_0\right)$, then computing $\hat{\Lambda}$ requires only matrix algebra and no additional simulation or estimation.

4.2 Sensitivity to Descriptive Statistics

If Λ is not sensitivity to moments, then the most convenient way to estimate Λ depends on how $\hat{\gamma}$ is defined. We assume throughout that $\hat{\gamma}$ is also an MDE, an assumption that is not very restrictive as most common summary statistics—first or second moments, many functions of estimation moments, regression coefficients, etc.—can be represented as MDE.

Let $\hat{m}(\gamma)$, M, and W_m denote the analogues of $\hat{g}(\theta)$, G, and W_g respectively that are used to estimate $\hat{\gamma}$. We assume conditions so that $\hat{g}(\theta)$ and $\hat{m}(\gamma)$ can be "stacked" to form an MDE $\left(\hat{\theta}, \hat{\gamma}\right)$, in particular that $\hat{g}(\theta_0)$ and $\hat{m}(\gamma_0)$ are jointly asymptotically normal with variance Ω. We let Ω_{gg}, Ω_{mm}, and Ω_{gm} denote the sub-matrices of Ω corresponding to the variance of $\hat{g}(\theta_0)$, the variance of $\hat{m}(\gamma_0)$, and the covariance of $\hat{g}(\theta_0)$ and $\hat{m}(\gamma_0)$ respectively.

Under these assumptions, it is straightforward to show that

$$\Sigma_{\theta\gamma} = \left(G' W_g G\right)^{-1} G' W_g \Omega_{gm} W_m M \left(M' W_m M\right)^{-1}.$$

[19] In CMD or SMD where $\hat{g}(\theta) = \hat{\pi} - h(\theta)$, $H = -G$ where H is the jacobian of $h()$ at the true value θ_0.

Standard estimators $\hat{\Sigma}_{\theta\theta}$ and $\hat{\Sigma}_{\gamma\gamma}$ are available for $\Sigma_{\theta\theta}$ and $\Sigma_{\gamma\gamma}$. If we can construct an estimator $\hat{\Sigma}_{\theta\gamma}$ for $\Sigma_{\theta\gamma}$, we can form consistent estimators $\hat{\Lambda} = \hat{\Sigma}_{\theta\gamma}\hat{\Sigma}_{\gamma\gamma}^{-1}$ and $\hat{\Delta}_p = \left(\hat{\Lambda}\hat{\Sigma}_{\gamma\gamma}\hat{\Lambda}'\right)_{pp} / \left(\hat{\Sigma}_{\theta\theta}\right)_{pp}$ for Λ and the elements of Δ.

Of the components of $\Sigma_{\theta\gamma}$, W_g and W_m are consistently estimated by $\hat{W}_g$ and $\hat{W}_m$ which are in hand from estimation, and G and M are consistently estimated by the sample analogues $\hat{G} = G\left(\hat{\theta}\right)$ and $\hat{M} = M\left(\hat{\gamma}\right)$. All that remains is to estimate Ω_{gm}. In cases such as CMD or SMD, it is common to use a bootstrap to estimate Ω_{gg}; in such cases the same bootstrap can typically be used to estimate Ω_{gm}.

Remark 6. If $\hat{\theta}$ and $\hat{\gamma}$ are MDEs and the researcher has computed plug-in estimators of $\text{Var}\left(\tilde{\theta} - \theta_0\right)$ and $\text{Var}\left(\tilde{\gamma} - \gamma_0\right)$ then computing a consistent estimate $\hat{\Lambda}$ requires only computing a consistent estimate $\hat{\Omega}_{gm}$ of the asymptotic covariance of the moment conditions.

An important special case is when $\hat{\theta}$ and $\hat{\gamma}$ are both estimated via GMM (Hansen 1982). (Recall that this case includes MLE.) Then $\hat{g}(\theta) = \frac{1}{n}\sum_{i=1}^{n} g(z_i, \theta)$ and $\hat{m}(\gamma) = \frac{1}{n}\sum_{i=1}^{n} m(z_i, \gamma)$ for i.i.d. data z_i and functions $g(z, \theta)$ and $m(z, \gamma)$ satisfying $\text{E}(g(z, \theta_0)) = \text{E}(m(z, \gamma_0)) = 0$. In this case a consistent estimator for Ω_{gm} is $\hat{\Omega}_{gm} = \frac{1}{n}\sum_{i=1}^{n} g\left(z_i, \hat{\theta}\right) m\left(z_i, \hat{\gamma}\right)'$.

An alternative representation of the estimator for $\hat{\Lambda}$ is useful for building intuition in this case.

Definition. Let $\tilde{g}_i = -\left(\hat{G}'\hat{W}_g\hat{G}\right)^{-1}\hat{G}'\hat{W}_g g\left(z_i, \hat{\theta}\right)$ *and define* $\tilde{m}_i$ *analogously. These* $(P \times 1)$ *and* $(J \times 1)$ *vectors are the* influence *of observation* i *on* $\hat{\theta}$ *and* $\hat{\gamma}$ *respectively (Hampel et al. 1986; Ronchetti and Trojani 2001).*

Intuitively, through the first-order condition $\tilde{g}_i$ tells us how much (and in what direction) observation i affects $\hat{\theta}$. The same property holds for $\tilde{m}_i$. Then by regressing $\tilde{g}_i$ on $\tilde{m}_i$ we recover how the influence of an observation on $\hat{\gamma}$ relates to its influence on $\hat{\theta}$, and hence how $\hat{\gamma}$ and $\hat{\theta}$ are related under the data-generating process:

Proposition 5. *The transposed coefficient matrix* $\hat{\Lambda} = (\tilde{g}'\tilde{m})(\tilde{m}'\tilde{m})^{-1}$ *from a regression of* $\tilde{g}_i'$ *on* $\tilde{m}_i'$ *is a consistent estimator of the sensitivity* Λ *of* $\hat{\theta}$ *to* $\hat{\gamma}$*. The* R^2 *from the regression is a consistent estimator of* Δ_p*.*

Proof. Let $\tilde{g}$ and $\tilde{m}$ denote the matrices whose rows are $\tilde{g}_i'$ and $\tilde{m}_i'$, respectively. The first statement follows from the continuous mapping theorem and the definition of sensitivity after noting that $\hat{\Lambda} = (\tilde{g}'\tilde{m})(\tilde{m}'\tilde{m})^{-1}$, $\frac{1}{n}\tilde{g}'\tilde{m} \xrightarrow{p} \Sigma_{\theta\gamma}$ and $\frac{1}{n}\tilde{m}'\tilde{m} \xrightarrow{p} \Sigma_{\gamma\gamma}$. The second statement follows from the continuous mapping theorem and definition of sufficiency after noting that:

$$R^2 = \frac{\left(\hat{\Lambda} \cdot \left(\frac{1}{n}\tilde{m}'\tilde{m}\right) \cdot \hat{\Lambda}'\right)_{pp}}{\left(\frac{1}{n}\tilde{g}'\tilde{g}\right)_{pp}}$$

and that $\frac{1}{n}\tilde{g}'\tilde{g} \xrightarrow{p} \Sigma_{\theta\theta}$. □

Example. (Sensitivity of MLE to Sample Mean) Suppose the data are $z_i \in \mathbb{R}^D$, with elements z_{di}, the parameter of interest θ is a scalar, and $\hat{\theta}$ is an MLE with likelihood function $f(z_i|\theta)$:

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \ln f(z_i|\theta).$$

Suppose we wish to assess sensitivity to the means of the elements of z_i, so we define $\hat{\gamma} \equiv \bar{z} \equiv \frac{1}{n}\sum_{i=1}^{n} z_i$.

We can interpret $\hat{\theta}$ as a GMM estimator with moment functions $g(z_i|\theta) = \partial \ln f(z_i|\theta)/\partial\theta$, weight matrix $W_g = I$, and Jacobian $G(\theta) = E\left(\partial^2 \ln f(z_i|\theta)/\partial\theta^2\right)$. We can interpret $\hat{\gamma}$ as a GMM estimator with moment functions $m(z_i|\gamma) = z_i - \gamma$, weight matrix $W_m = I$, and Jacobian $M(\gamma) = -I$. We can consistently estimate Λ with the coefficients from a regression of the (scaled) score of observation i:

$$\tilde{g}_i = -\frac{1}{\hat{G}} \cdot \left.\frac{\partial \ln f(z_i|\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}}$$

on the deviation from the mean of observation i:

$$\tilde{m}_i = (z_i - \bar{z}).$$

Intuitively, $\hat{\theta}$ is more sensitive to the mean of a particular variable z_{di} when observations with high values of z_{di} have high values of the score (holding the other elements of z_i constant).

This approach is easily extended to look at the sensitivity of $\hat{\theta}$ to higher-order moments of the data. Suppose we define $\hat{\gamma}$ to be all first and (centered) second empirical moments of z_i. Then the elements of $\tilde{m}_i$ corresponding to the mean of the various z_{di} are equal to $(z_{di} - \bar{z}_d)$, the elements corresponding to the variances of the various z_{di} are equal to $(z_{di} - \bar{z}_d)^2$, and the elements corresponding to the covariances of various z_{di} and $z_{d'i}$ are equal to $(z_{di} - \bar{z}_d)(z_{d'i} - \bar{z}_{d'})$. We can consistently estimate Λ with the coefficients from a regression of $\tilde{g}_i$ on this redefined $\tilde{m}_i$. By continuing to add empirical moments, we can flexibly explore the way $\hat{\theta}$ depends on the moments of the data.[20]

4.3 Units of Measurement

We have noted that Λ has an interpretation as the probability limit of coefficients from a regression of $\hat{\theta}$ on $\hat{\gamma}$. As with any regression coefficients, the elements of Λ depend on the units of measurement of the regressors $\hat{\gamma}$. Determining which element of $\hat{\gamma}$ is most "important" for a given $\hat{\theta}_p$ therefore requires judgment. The problem of assessing the relative importance of regressors is

[20]It may also be possible to use variable selection methods such as the Lasso to search for a low-dimensional set of moments with high predictive power.

age-old and no solution is satisfactory in all situations (Kim and Ferree 1981; Bring 1994; Gelman 2008). But it is helpful to have a default. For this we propose the analogue of the standardized regression coefficient:

Definition. *The* **standardized sensitivity** *of* $\hat{\theta}_p$ *to* $\hat{\gamma}_j$ *is*

$$\tilde{\Lambda}_{pj} = \Lambda_{pj}\sqrt{\frac{\mathrm{Var}\left(\tilde{\gamma}_j\right)}{\mathrm{Var}\left(\tilde{\theta}_p\right)}}.$$

Standardized sensitivity measures how much a one-standard-deviation change in the realization of $\tilde{\gamma}_j$ affects the expected value of $\tilde{\theta}_p$, fixing other elements of $\tilde{\gamma}$, in units of the standard deviation of $\tilde{\theta}_p$. If the elements of $\hat{\gamma}$ are asymptotically independent (i.e., if $\Sigma_{\gamma\gamma}$ is diagonal) then the matrix $\tilde{\Lambda}$ of standardized sensitivities is the correlation matrix of $\tilde{\theta}$ with $\tilde{\gamma}$.

An attractive property of standardized sensitivity is that it is invariant to changes in units. Formally, for vectors a, c and strictly positive diagonal matrices B, D the standardized sensitivity of $a + B\hat{\theta}$ to $c + D\hat{\gamma}$ is equal to the standardized sensitivity of $\hat{\theta}$ to $\hat{\gamma}$. This means that, for example, if we switched from measuring an element of $\hat{\gamma}$ in dollars to measuring it in euros, our conclusions about the relative importance of different moments would be unchanged.[21]

Comparisons in units of standard deviations will not always be appropriate or necessary. If two statistics are in comparable economic units, it may be meaningful to compare their unstandardized sensitivities directly. For example, in the sample means case presented in section 2.3, the empirical moments are averages of the same variable in different populations, making unstandardized comparisons very natural. Nevertheless, abstracting from any particular context it seems attractive to have a unitless measure as a default, and we will report estimates of standardized sensitivity for all of our applications.

5 Alternatives

Here we compare our sensitivity measure to two alternative methods of developing intuition for the mapping from data to parameter estimates.

5.1 Inverse Sensitivity

Our sensitivity measure asks how the expected values of the parameters change as we vary the data features of interest. An alternative way to investigate what drives an estimator would be to ask how

[21] There are other transformations of Λ, such as the matrix of partial correlations of $\tilde{\theta}_p$ with $\tilde{\gamma}_j$ (conditional on $\tilde{\gamma}_{\sim j}$), that would also exhibit this invariance property.

the expected values of the data features change when we vary the parameters. Intuitively, we might say that a particular $\hat{\theta}_p$ will depend heavily on a particular $\hat{\gamma}_j$ if varying θ_p in the model causes large changes in the expected value of $\hat{\gamma}_j$. This approach can easily be implemented by simulating data from the model at alternative parameter values. Goettler and Gordon (2011), Kaplan (2012), Morten (2013), and Berger and Vavra (forthcoming) are examples of papers that refer to such simulations in their discussions of identification.[22]

This approach can be thought of as the "inverse" of our proposed sensitivity measure. To see why, suppose that $\hat{\theta}$ is a GMM estimator and Λ is sensitivity to moments. The alternative approach would infer that the j-th moment $\hat{\gamma}_j = \frac{1}{n}\sum_{i=1}^{n} g_j(z_i, \theta_0)$ is an important driver of $\hat{\theta}_p$ if the absolute value of $\frac{\partial}{\partial \theta_p} \mathrm{E}\left[g_j(z_i, \theta)\right]\Big|_{\theta=\theta_0}$ is large. Notice that the matrix of these partial derivatives is simply the Jacobian G. Since $\Lambda = -\left(G' W_g G\right)^{-1} G' W_g$, we have $-\Lambda G = I$, and so when Λ is square $G = -\Lambda^{-1}$.

How does the intuition delivered by G about the relative importance of moments compare to the intuition delivered by Λ? There is a special case where they yield the same intuition: when the model has a single parameter ($P = 1$), and the weight matrix $W_g = I$. In this case, $(G' W_g G)^{-1}$ is simply a constant, so $|\Lambda| \propto |G|$. If $\hat{\gamma}_j$ changes more than $\hat{\gamma}_k$ when we vary the single parameter θ, $\hat{\theta}$ will be more sensitive to $\hat{\gamma}_j$ than to $\hat{\gamma}_k$.

Outside of this special case, the intuitions from Λ and G can be very different. While examining G can be a useful way to build economic intuition about a model, we argue that it can actually be very misleading if interpreted as a guide to the sensitivity properties of an estimator or to sources of identification in the sense defined above.

The reason that G is not a good guide to the sensitivity properties of an estimator is that it is not a property of an estimator; rather, it is a (local) property of a model. An easy way to see this is to note that G does not depend on the weight matrix W_g. For an overidentified model, this means that G can't tell us which features of the data drive a particular $\hat{\theta}$. Consider our earlier example in which θ_0 is the population standard deviation of an exponential random variable. In this case, G tells us that θ is equally related to the mean and the standard deviation, because under the model both change by the same amount when we vary θ. By contrast, Λ reveals the weights employed by the estimator we are using, and so shows us which empirical moments drive the result.

The reason that G is not a good guide to identification is that the relationship established in

[22] Goettler and Gordon (2011) describe specific parameters as "primarily identified by" particular moments if those moments respond sharply to changes in those parameters (p. 1161). Kaplan (2012) writes: "I address the question of identification in three ways ... Third, below I provide an informal argument that each of the parameters has influence on a subset of the chosen moments and give some intuition for why this is the case" (p. 478). Morten (2013) writes: "As a check on how well the identification arguments for the simple model apply ... I simulate the dynamic model for a range of parameter values. I vary each parameter ... and then plot the responses of each of the main moments as the parameter changes" (p. 33).

proposition 2 does not hold for G: it can easily happen that G assigns zero sensitivity to features that are needed for identification, and non-zero sensitivity to features that are not needed for identification. Recall our OLS example in which $\hat{\gamma} = \left[\begin{array}{cccc} \hat{\mu}_y & \hat{\sigma}_{xy} & \hat{\sigma}_x^2 & \hat{\mu}_x \end{array} \right]'$. The coefficient β is identified by σ_{xy} and σ_x^2 alone. Consistent with this, the row of Λ corresponding to $\hat{\beta}$ has non-zero entries for $\hat{\sigma}_{xy}$ and $\hat{\sigma}_x^2$ and zeros elsewhere. The corresponding column of G, however, has non-zero entries only for μ_y and σ_{xy} (assuming $\mu_x \neq 0$).[23] Changing β affects the mean of y and its covariance with x, but leaves the mean and variance of x unchanged; however, β is not identified by the mean of y and its covariance with x alone, and the mean of y is not necessary for identification of β.

5.2 Dropping Moments

In the case of an overidentified MDE, an alternative way to check sensitivity to an empirical moment is to drop the moment and re-estimate the model. To fix ideas, assume that equation (3) has a solution when the j^{th} element of $\hat{g}(\theta)$ is excluded, and denote the resulting estimator by $\hat{\theta}^{\sim j}$. Comparing the parameters estimated with and without moment j amounts to calculating $\left(\hat{\theta} - \hat{\theta}^{\sim j}\right)$.

Suppose that the j^{th} moment (and only the j^{th} moment) is possibly misspecified. Then the following corollary of proposition 4 shows that the measure $\left(\hat{\theta} - \hat{\theta}^{\sim j}\right)$ combines information about sensitivity Λ with information about the degree of misspecification:

Corollary 3. *Suppose that $\hat{\theta}$ is an MMDE with $\hat{g}_j(\theta_0) \xrightarrow{p} \epsilon_j$ and that $\hat{\theta}^{\sim j}$ is an MDE. Then $\hat{\theta} - \hat{\theta}^{\sim j} \xrightarrow{p} \Lambda_\epsilon \epsilon_j$ for some Λ_ϵ with $\lim_{\epsilon_j \to 0} \Lambda_\epsilon = \Lambda_{\cdot j}$ where $\Lambda_{\cdot j}$ is the sensitivity to moment j of the MDE corresponding to $\hat{\theta}$.*

Proof. By the continuous mapping theorem $\operatorname{plim}\left(\hat{\theta} - \hat{\theta}^{\sim j}\right) = \operatorname{plim}\left(\hat{\theta}\right) - \operatorname{plim}\left(\hat{\theta}^{\sim j}\right)$. By consistency $\operatorname{plim}\left(\hat{\theta}^{\sim j}\right) = \theta_0$. By proposition 4 $\operatorname{plim}\left(\hat{\theta}\right) = \theta_0 + \Lambda_\epsilon \epsilon_j$ with $\lim_{\epsilon_j \to 0} \Lambda_\epsilon = \Lambda_{\cdot j}$. □

When $\epsilon_j = 0$, $\hat{\theta} - \hat{\theta}^{\sim j} \xrightarrow{p} 0$ and therefore the limiting behavior of $\left(\hat{\theta} - \hat{\theta}^{\sim j}\right)$ is unrelated to sensitivity. Sensitivity will continue to play a role in any finite sample, however, in the sense that $\left(\hat{\theta} - \hat{\theta}^{\sim j}\right)$ will be related both to Λ's finite-sample analogue and to the realization of $\hat{g}_j(\theta_0)$.

[23] To restate this example as an MDE, let $\theta = \left[\begin{array}{cccc} \alpha & \beta & \sigma_x^2 & \mu_x \end{array} \right]'$ and $\hat{g}(\theta) = \hat{\gamma} - h(\theta)$ where $h(\theta) = \left[\begin{array}{cccc} \alpha + \beta\mu_x & \beta\sigma_x^2 & \sigma_x^2 & \mu_x \end{array} \right]'$. Then

$$G = \left[\begin{array}{cccc} -1 & -\mu_x & 0 & -\beta \\ 0 & -\sigma_x^2 & -\beta & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{array} \right].$$

6 Main Applications

6.1 Automobile Demand

Our first application is to Berry et al.'s (1995) model of automobile demand. We follow Berry et al. (1995) closely, using their data and SMM procedure, with moments from both the demand and supply sides of the model.[24]

The estimation moments are derived from two sets of identifying assumptions. On the demand side, the model assumes that the expected unobserved quality ξ_j of car j is zero conditional on instruments z_j^d. Berry et al. (1995) construct z_j^d from a set of demand-side variables: a constant, a dummy for whether car j has air conditioning, and car j's horsepower-per-weight, miles-per-dollar of gasoline, and size. For each variable, z_j^d includes: (i) the value of the variable for car j; (ii) the sum of the variable across other cars produced by the firm that produces car j; (iii) the sum of the variable across cars produced by rival firms. The demand-side moments are the product of ξ_j (computed as a residual inverted from market shares) with each element of z_j^d.

On the supply side, the model assumes that the expected unobserved cost component ω_j of car j is zero conditional on instruments z_j^s. Berry et al. (1995) construct z_j^s in the same way as z_j^d, but using instead a set of supply-side variables: a constant, a dummy for whether car j has air conditioning, a time trend, and the logarithms of car j's horsepower-per-weight, miles-per-gallon of gasoline, and size. In addition, z_j^s includes an excluded demand variable, miles-per-dollar of gasoline for car j (but not the sums of this variable across other cars). The supply-side moments are the product of ω_j (computed as a residual inverted from estimated marginal costs) with each element of z_j^s.

We first apply our method to analyze the relative importance of the various instruments in driving estimated markups. We define the counterfactual $c\left(\hat{\theta}\right)$ of interest to be the average estimated markup across all cars.[25] We define $\hat{\gamma}$ to be the complete set of estimation moments $\left(z_j^{d\prime}\xi_j \quad z_j^{s\prime}\omega_j \right)$, but plot sensitivity only for moments involving the "excluded" instruments—i.e., those that do not enter the utility or cost j directly.[26]

These results are presented in figure 2. We find that markups are overall more sensitive to the

[24]Since the code for Berry et. al (1995) does not appear to be publicly available, we extract automobile data and guide our implementation using the GAUSS code for Berry et al. (1999), downloaded from the Internet Archive's April 2005 web capture of James Levinsohn's (now defunct) website at the University of Michigan. Table 1 from Berry et al. (1995) and table 2 from Berry et al. (1999) imply that the two use the same dataset. We used code from Petrin (2002), Dubé et al. (2012), and Knittel and Metaxoglou (2014) as additional references.

[25]The markup is tightly related to the own-price elasticity, and, therefore, to the price coefficient. We show in the online appendix that the pattern of sensitivities for either the average own-price elasticity or the price coefficient are similar to those we present here.

[26]The online appendix reports the complete standardized sensitivity matrix $\tilde{\Lambda}$, with values for all parameters and moments.

supply moments than the demand moments. We also find that the supply-side instruments that play the largest role in driving markups is simply the number of other products produced by the same firm (i.e., the sum of the constant term across other products produced by the same firm), the gas mileage of these cars, and the number of products produced by rival firms. To build intuition for this, recall that the model is estimated using data from 1971-1990, a period that saw the large-scale entry of Japanese cars and a shift toward higher mileage. A possible interpretation, therefore, is that the model is using the changes in prices predicted as a result of this increased competitiveness and product differentiation and as the key exogenous source of variation.

Next, we apply our method to ask to what extent the relationship of moments to estimated elasticities in the full BLP model is well approximated by the relationship of moments to estimated elasticities in the aggregate logit version of the model. The latter model, which has no random coefficients, can be estimated by two-stage least squares. Berry et al. (1995) present estimates from this logit model as a point of departure. We define the counterfactuals $c\left(\hat{\theta}\right)$ to be mean elasticities of demand with respect to price and product attributes implied by the estimated parameters of the full BLP model, and transformed statistics $a\left(\hat{\gamma}\right)$ to be the same mean elasticities implied instead by the estimated parameters of the logit model. We compute sensitivity of each elasticity in the full model to the elasticities implied by the logit model.

These results are presented in figure 3. We find that sufficiency Δ_p is low for all elasticities, ranging from 0.02 for the air conditioning dummy to 0.13 for miles-per-dollar. Moreover, there is no systematic pattern in which the estimated demand elasticity to a particular attribute in the full model is primarily related to the estimated demand elasticity for that same attribute in the logit model. This suggests, consistent with the discussion in Berry et al. (1995), that carrying forward simple intuitions from the logit model is not a useful way to understand the full model.

6.2 Life-cycle Consumption and Savings

6.2.1 Gourinchas and Parker (2002)

Our next application is to Gourinchas and Parker's (2002) model of life-cycle consumption. Gourinchas and Parker (2002) model the behavior of a consumer with a time-separable constant-relative-risk-aversion felicity function and a stochastic income process. The parameters of the income process are estimated in a first step and are taken as given in a second step, in which preference parameters are estimated from moments corresponding to mean consumption at different ages (adjusted for family size and business cycle shocks).

Figure 4 presents results for the second-step model's two key preference parameters: the discount factor and the coefficient of relative risk aversion.[27] The plot reveals three periods of life

[27]The baseline specification in Gourinchas and Parker (2002) has two additional parameters, which govern a

with different implications for the parameter estimates. In the first period, roughly ages 26-36, and in the third period, roughly ages 62-65, higher consumption implies a higher discount factor and a lower coefficient of relative risk aversion. In the second period, roughly ages 37-61, higher consumption implies a lower discount factor and a higher coefficient of relative risk aversion.

A stylized intuition is as follows. The consumer saves for retirement and for precautionary reasons. The strength of retirement saving motives is governed by the discount factor, and the strength of precautionary motives by the coefficient of relative risk aversion. Both a higher discount factor and a higher coefficient of relative risk aversion predict more delay of consumption, i.e., lower consumption early in life and greater consumption later in life. The two parameters are separately identified because of their different quantitative implications.

In the first period of life, saving is primarily precautionary, so risk aversion matters comparatively more than discounting, and higher consumption is interpreted as evidence of low risk aversion. In the second period, saving is primarily for retirement, so discounting matters comparatively more, and higher consumption is interpreted as evidence of impatience. In the third period, retirement looms and income uncertainty has essentially vanished, so high consumption is evidence that the household has already accumulated substantial retirement wealth, i.e., that the household is patient.

The fact that the two plots are essentially inverse to one another arises because both a higher discount factor and a higher coefficient of relative risk aversion imply the same qualitative change in the consumption profile. Therefore a change in consumption at a given age that implies a high discount factor must be offset by a lower coefficient of relative risk aversion in order to hold consumption at other ages constant.

6.2.2 De Nardi et al. (2010)

De Nardi et al. (2010) model consumption and saving by retired, nonworking households with uninsurable mortality and medical expense risk. Households have a time-separable constant-relative-risk aversion felicity function and a consumption floor guaranteed by the government. The parameters of the mortality and medical expense processes are estimated in a first step and are taken as given in a second step, in which the discount factor, coefficient of relative risk aversion, and consumption floor are estimated using SMM from moments corresponding to median assets for different cohorts, ages, and permanent income levels. We use the results in remark 3 to compute the sensitivity of second-step parameters to the means of related groups of estimation moments.[28]

reduced-form retirement consumption function. We fix these at their estimated values for the purposes of our analysis. The online appendix reports the numerical values of standardized sensitivity for the discount factor and the coefficient of relative risk aversion.

[28] The online appendix reports the standardized sensitivity of second-step parameters to the full set of (untransformed) estimation moments.

The first two plots in figure 5 present the sensitivity of the consumption floor and the coefficient of relative risk aversion to the mean of the asset holdings by income quintile. The consumption floor is sensitive primarily to the savings of households in the lowest income quintile: the less these households save, the greater is the inferred consumption floor. The coefficient of relative risk aversion rises with the savings of the rich and falls with the savings of the poor. This pattern matches closely the intuition in De Nardi et al. (2010):

> The coefficient of relative risk aversion is identified by differences in saving rates across the income distribution, in combination with the consumption floor. Low-income households are relatively more protected by the consumption floor and will thus have lower [variance of consumption growth] and hence weaker precautionary motives. The parameter helps the model explain why individuals with high permanent income typically display less asset decumulation (p. 59).

The third plot in figure 5 presents the sensitivity of the discount factor to the mean of the asset holding moments by age. As expected, the estimator interprets large asset holdings at younger ages as evidence of patience.

7 Other Applications

7.1 Sensitivity to Moments

In this subsection we apply our measure of the sensitivity to moments to several empirical papers that use MDEs. In each case we obtain plug-in estimators $\hat{G}$, $\hat{W}_g$, and $\hat{\Omega}_{gg}$ either directly from the authors or from replication files posted by the authors. For papers that estimate multiple specifications we use the baseline or main specification reported in the paper.

We present our findings as plots of standardized sensitivity for all moments for each of a set of key parameters. In the online appendix we report the complete standardized sensitivity matrix $\tilde{\Lambda}$ for each paper. Each plot indicates the key moments that the authors highlight as important for the identification of the given parameter. When we quote from a given paper, we replace mathematical symbols with italicized phrases in order to avoid notational conflict or confusion.

Goettler and Gordon (2011)

Goettler and Gordon (2011) model innovation in the market for a durable good. In the model, each of a set of firms maintains a position on a quality ladder. The chance of moving up the ladder is greater the more the firm invests in R&D and the further the firm is from the technological frontier. Marginal costs are increasing in product quality. Consumers value quality and treat the firms' products as vertically and horizontally differentiated. Both firms and consumers are forward-looking.

The model is estimated on data from the market for computer microprocessors. The main research question is whether the market leader, Intel, innovates more or less than it would in a counterfactual world without its main competitor, AMD. Seven parameters are estimated from 15 empirical moments using SMD. The authors provide a discussion of identification that follows the "inverse sensitivity" approach discussed in section 5: they suggest a parameter is identified by a moment if the value of that moment responds strongly to changes in the parameter.

Figure 6 presents results for three key parameters. We follow Goettler and Gordon (2011) in dividing moments into groups of "demand-side" and "supply-side" groups.

The first two parameters that we consider are demand parameters: the price coefficient, which reflects the disutility of higher prices, and the quality coefficient, which reflects the utility from higher quality. Regarding these parameters Goettler and Gordon (2011) write:

> The demand-side parameters (*price coefficient*, *quality coefficient*, *Intel fixed effect*, and *AMD fixed effect*) are primarily identified by the pricing moments, the Intel share equation moments, and the mean ownership quality relative to the frontier quality. The pricing moments respond sharply to changes in any of these four parameters. The market share equation is primarily sensitive to *quality coefficient* and *Intel fixed effect - AMD fixed effect*. The mean *upgrading moment* decreases if consumers upgrade more quickly and is akin to an outside share equation that identifies the levels of the *Intel fixed effect* and the *AMD fixed effect* (p. 1161).

In figure 6, we find that the price coefficient is primarily sensitive to the average prices of Intel and AMD. This is intuitive because Goettler and Gordon (2011) have a direct measure of marginal cost. Given the assumption of dynamically optimal pricing, the higher is the observed price, the less price-sensitive consumers are estimated to be. The quality coefficient is primarily sensitive to the potential upgrade gains, a measure of the difference between the average CPU quality of the computer stock and the frontier quality available. Again, this is intuitive: the more sensitive consumers are to quality, the more often consumers will upgrade their PCs and the smaller will be the gap between average and frontier quality.

The third parameter that we consider is the innovation spillover, a measure of the extent to which innovation is easier the further the firm lies inside the technological frontier. Goettler and Gordon (2011) write:

> The supply-side parameters (*Intel innovation efficiency*, *AMD innovation efficiency*, and *innovation spillover*), which govern the investment process, are primarily identified by observed innovation rates, quality differences, and investment levels. The

> investment efficiencies are chosen such that the observed investment levels (per unit revenue) yield innovation at the observed rates. The spillover parameter *innovation spillover* is chosen to match the mean difference in quality across firms: a high spillover keeps the qualities similar (p. 1161).

We find that the innovation spillover is very responsive to the mean quality difference as expected. However, it responds slightly more to the average Intel price, and in general is very responsive to demand moments.

DellaVigna et al. (2012)

DellaVigna et al. (2012) model a household's charitable giving. In the model, a household may give to charity either out of altruism or because of social pressure. DellaVigna et al. (2012) conduct a field experiment in which they solicit charitable donations door-to-door. In some treatments they alert the household in advance that they will be coming to solicit. Households' response to this warning provides evidence on the motivations for giving and allows DellaVigna et al. (2012) to assess the welfare effects of charitable solicitations.

The model is estimated using 70 moments corresponding to the empirical frequencies of opening the door and giving different amounts of money in different treatment conditions. The model has 15 parameters estimated via CMD, using quadrature to approximate the expected value of the empirical moments as a function of the parameters.

Figure 7 presents results for two parameters. For each parameter, we show the standardized sensitivity to all moments, indicating key moments highlighted by the authors in red.

The first parameter, the baseline probability of being home, has a very simple relationship to the empirical moments. DellaVigna et al. (2012) explain that:

> The baseline probabilities of answering the door ... are identified by the observed probabilities of opening the door in treatments without flyer (p. 37).

Our plot bears out this discussion, showing that the empirical probabilities of being home in no-flier conditions are the most important drivers of this parameter.

The second parameter, the social cost of giving less than $10 to the East Carolina Hazard Center (ECU), has a richer economic structure. DellaVigna et al. (2012) write:

> Finally, the social pressure ... is identified from two main sources of variation: home presence in the flyer treatment ... and the distribution of small giving (the higher the social pressure, the more likely is small giving and in particular bunching at *the threshold of $10*) (p. 38).

The authors define the social cost of giving \$$X$ as $S \times \max\{10 - X, 0\}$, where S is a parameter. We report sensitivity values for the cost of giving \$0, which is $10S$. The sensitivity values closely match the authors' discussion: Giving at the \$10 threshold increases the inferred level of social pressure, as does failing to open the door when warned in advance by a flier. (The only exception is that giving less than \$10 is found to decrease rather than increase the estimated level of social pressure, perhaps because this level of giving does not allow the household to avoid feeling socially pressured.)

Nikolov and Whited (2014)

Nikolov and Whited (2014) model an infinitely lived firm whose manager makes decisions in discrete time about both the level of real investment and the extent of external financing. Capital is subject to depreciation and a convex adjustment cost. External financing imposes a real cost on the firm. The manager has an equity stake, a profit stake, and an ability to "tunnel" resources that are held as cash. The profit stake and the ability to tunnel lead to a divergence between the manager's interests and those of the shareholders.

The model has 8 estimated parameters, corresponding to features of the production and investment technology, the external financial environment, and the manager's incentives. These parameters are estimated via SMM based on empirical moments that contain information on investment, financing, and compensation in a sample of firms.

Figure 8 presents standardized sensitivity for three select parameters. We follow Nikolov and Whited (2014) in dividing the moments loosely into "real" moments related to the investment decision, "financial" moments related to cash vs. external finance, and "incentives" moments related to managerial compensation and incentives.

The first parameter we study is the rate of depreciation of capital. Nikolov and Whited (2014) report that this parameter is identified by the mean rate of investment:

> The first two non-financial or "real" moments are the first and second central moments of the rate of investment ... The first moment identifies the capital depreciation rate (p. 1899).

The economic logic here is that in a deterministic steady-state, the rate of investment is equal to the rate of depreciation of capital. The sensitivity values for the depreciation parameter bear out this intuition: the mean rate of investment is by far the most important moment in determining the estimated depreciation rate.

The second parameter that we study is the profit-sharing parameter, which corresponds to the fraction of after-tax operating earnings that accrue to the manager. Nikolov and Whited (2014) report that this parameter is identified principally by the average bonus paid to the firm's CEO:

> Finally, we discuss the identification of the profit-sharing parameter. ... First, without our data on ownership and compensation, we would have to infer the value of this parameter solely from firm decisions. In this case, a high value of *the profit-sharing parameter* implies low average profitability because the manager acts as if the firm is more profitable than it actually is and makes distorted investment decisions. However, many other parameters affect average profitability, so this moment alone cannot help identify *the profit-sharing parameter*. Fortunately, this parameter corresponds directly to one moment from our compensation data: the average bonus (p. 1900).

The authors also note that Tobin's q is useful in identifying this parameter. The sensitivity measure agrees with the authors' discussion. By far the most important driver of the estimated profit-sharing parameter is the average bonus. The average profit level is also relevant, and has the sign predicted by the model.

The third and final parameter that we study is the tunneling parameter, which corresponds to the fraction of the current stock and flow of cash that the manager consumes privately. Nikolov and Whited (2014) write:

> Not surprisingly, the moment that is most important for identifying resource diversion is the mean of Tobin's q: the more resource diversion, the lower q (p. 1900).

The sensitivity plot shows that greater Tobin's q does correspond to a lower inferred tunneling. Other moments also play an important role, however. Both lower investment and greater average profits imply greater tunneling. A possible explanation is that lower investment and greater profits imply a greater flow of resources, so for a fixed distribution to shareholders, managerial resource diversion must adjust to enforce the accounting identity that determines distributions to shareholders.

7.2 Sensitivity to Descriptive Statistics

Our final set of applications are cases in which the economic model is estimated via MLE. Formally, an MLE is an MDE in which the moments are first-order conditions. In the applications below these first-order conditions do not have a clear economic interpretation. We therefore define $\hat{\gamma}$ to be a set of descriptive statistics, typically those presented by the authors to provide a summary of key features of the data. We compute standardized sensitivity of key parameters or counterfactuals using the empirical influence components as described in section 4. (Recall that these calculations do not require re-estimation of the model.) In addition to $\tilde{\Lambda}$, we report an estimate of sufficiency Δ. Unlike in the case of sensitivity to moments, Δ need not be equal to one; its magnitude summarizes how well the descriptive statistics $\hat{\gamma}$ capture the information in the data that drives the estimated parameters $\hat{\theta}$. We present our findings in plots; the online appendix contains the corresponding numerical estimates.

Mazzeo (2002)

Mazzeo (2002) models entry into motel markets along US interstates. In the variant of Mazzeo's model that we consider, a set of anonymous potential entrants to a local market make sequential decisions either not to enter the market, to enter as low quality, or to enter as high quality. Following the entry decision, firms realize payoffs that depend on observable market characteristics, the number of firms of each type, and a normally distributed profit shock that is specific to each firm type and local market and is unknown to the econometrician. Mazzeo (2002) estimates the model by maximum likelihood using data on the number and quality of motels along rural interstates in the United States.

Figure 9 reports the sensitivity of Mazzeo's (2002) estimates of the effect of market characteristics on firm profits. Here we let $\hat{\gamma}$ be the coefficients from regressions of the number of low- and high-quality firms on observable market characteristics. Intuitively, we would expect the structural parameter governing the effect of a given characteristic on profitability to be tightly related to that characteristic's effect on the number of firms. We find that this is indeed the case, and that the regression coefficients are almost sufficient for the structural parameters. In all cases, knowing the regression coefficients would allow us to predict more than 80 percent of the variation in the structural parameter under the asymptotic distribution.

Figure 10 reports the sensitivity of a counterfactual $c\left(\hat{\theta}\right)$. We define the n-firm differentiation effect as the probability that the n^{th} entrant's choice of type is affected by the choices of the prior entrants, in the sense that the choice would be different (fixing unobservables) if the prior entrants' choices were different. This counterfactual summarizes the importance of incentives to differentiate, which is a key economic magnitude in the model.

We consider sensitivity to the share of markets with different configurations, which is a simple descriptive summary of the competitive patterns in the data. As expected, we find that the 2-firm differentiation effect is driven largely by the composition of 2-firm markets. The more markets have one firm of each type, the greater is the implied differentiation effect. Likewise, the 3-firm differentiation effect is mostly driven by the composition of 3-firm markets, though the composition of 2-firm markets also plays a role.[29]

[29] In some cases the sensitivity of the differentiation effect to the share of markets with only one type of firm has a positive sign. This is counterintuitive because such markets will be more common when the incentive to differentiate is weak. The main reason for these counterintuitive signs is that we are computing sensitivity to the unconditional shares of different market configurations, whereas the model's inference about competitive effects is ultimately driven by the observed market configurations *conditional* on observable market characteristics. In the online appendix, we support this interpretation by showing that most of the counterintuitive signs reverse (and the sufficiencies rise) when

Gentzkow et al. (2014)

Gentzkow et al. (2014) model households' demand for partisan newspapers. In their model, both households and newspapers are either Republican or Democrat. Each household has a preference for reading newspapers of its own type. Households may read multiple newspapers, and newspapers are more substitutable if they are of the same party than if their parties differ. Gentzkow et al. (2014) estimate this model using aggregate data on circulation in a set of small towns in 1924.

We focus our discussion on two parameters: the relative preference for own-type newspapers, and the extent of substitutability of same-type newspapers. Gentzkow et al. (2014) argue that the parameter governing the relative preference for same-type newspapers is identified by the correlation between the relative circulation of Republican newspapers and the share of households who vote Republican. They argue that the parameter governing the extent of substitutability of same-type newspapers is identified by the extent to which adding more Republican newspapers to the choice set disproportionately reduces demand for other Republican papers. Following Gentzkow et al.'s (2014) discussion of identification we define $\hat{\gamma}$ to be the coefficients from a regression of the relative circulation of Republican newspapers on the Republican share of the vote and the number of Republican and Democratic newspapers.

The first plot in figure 11 shows that the structural parameter governing the relative preference for same-type newspapers is highly sensitive to the coefficient from a regression of the relative circulation of Republican newspapers on the Republican share of the vote. This is in line with the discussion in Gentzkow et al. (2014). The second plot shows that the structural parameter governing the substitutability of same-type newspapers is sensitive to all regression coefficients to a similar extent. Intuitively, as the coefficient on the number of Republican papers grows, this parameter shrinks, and the opposite happens for the coefficient on the number of Democratic papers.

Sufficiency calculations show that variation in these four regression parameters is sufficient to explain the majority of the asymptotic variation in the structural parameters. This is striking in light of the fact that the underlying structural model has many additional parameters, and that the maximum likelihood estimator is in principle exploiting much more information than can be captured in a simple regression.

Gentzkow (2007)

Gentzkow (2007) uses survey data from a cross-section of individuals to estimate demand for print and online newspapers in Washington DC. A central goal of Gentzkow's (2007) paper is to estimate the extent to which online editions of papers crowd out readership of the associated print editions,

we constrain the model so that observable market characteristics do not affect profits.

which in turn depends on a key parameter governing the extent of print-online substitutability. We focus here on the substitutability of the print and online editions of the Washington Post.

A naive approach would estimate this parameter from the correlation between print and online consumption: if those who read the Post in print are less likely to read online, the two are substitutes; if they are more likely to read online, the two are complements. This approach will be invalid, however, if there are unobserved consumer characteristics correlated with the taste for both products.

Gentzkow (2007) exploits two features of the data to distinguish correlated tastes from true substitutability: (i) a set of variables—such as a measure of Internet access at work—that plausibly shift the utility of online papers but do not affect the utility of print papers; and (ii) a coarse form of panel data—separate measures of consumption in the last day and last seven days—that identifies stable individual preferences in a manner analogous to fixed or random effects in a linear model.

To capture these two features of the data, we define $\hat{\gamma}$ to consist of two components: (i) the coefficient from a 2SLS regression of last-five-weekday print readership on last-five-weekday online readership, instrumenting for the latter with the set of excluded variables such as Internet access at work; and (ii) the coefficient from an OLS regression of last-one-day print readership on last-one-day online readership controlling flexibly for readership of both editions in the last five weekdays. Each of these auxiliary models includes the standard set of demographic controls from Gentzkow (2007).

We define the counterfactual $c\left(\hat{\theta}\right)$ to be the change in readership of the Post print edition that would occur if the Post online edition were removed from the choice set (Gentzkow 2007, table 10).

The results are presented in figure 12. Sufficiency is 0.64, suggesting that these two features of the data capture much but not all of the variation that drives the counterfactual. Sensitivity is negative for both elements of $\hat{\gamma}$ as expected, reflecting the fact that a more positive relationship between print and online consumption implies less substitutability and thus a smaller gain of print readership. Finally, the results show that sensitivity to the panel variation is much larger than sensitivity to the IV variation, implying that the former is the more important driver of the estimated counterfactual.

8 Conclusions

We develop a measure Λ of the extent to which a given parameter is sensitive to a given feature of the data. The measure is trivial to compute in common applications and is interpretable as a measure of sensitivity to model misspecification.

An important limitation of our approach is that Λ is a local measure. It captures the way $\hat{\theta}$

varies with small perturbations of $\hat{\gamma}$ around its limiting value. Conceptually, relaxing this constraint is straightforward. Consider the following exercise: (i) simulate or otherwise obtain data with dispersed values of $\hat{\gamma}$; (ii) estimate $\hat{\theta}$ on each dataset; and (iii) regress $\hat{\theta}$ on $\hat{\gamma}$ across these datasets. Such a procedure delivers a "global Λ" as compared to the "local Λ" we work with in this paper.

We focus on the local Λ precisely because repeated simulation and estimation is often costly. We can, however, suggest approaches to minimizing this computational burden. First, for estimators whose cost of execution scales well with the size of the dataset, a researcher might use small-scale simulations to obtain the global Λ and to compare it to the local Λ. If the two are similar, this adds confidence to the use of the local Λ for sensitivity analysis.

Second, for cases where simulation from the data-generating process is cheaper than estimation, a researcher might simulate data from several possible values of θ and compute $\hat{\gamma}$ on the simulated data. Then, by regressing θ on $\hat{\gamma}$, one obtains a version of the global Λ that does not require repeated model estimation.

References

Angrist, Joshua D. and Jörn-Steffen Pischke. 2010. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives* 24(2): 3-30.

Barseghyan, Levon, Francesca Molinari, Ted O'Donoghue, and Joshua C. Teitelbaum. 2013. The nature of risk preferences: Evidence from insurance choices. *American Economic Review* 103(6): 2499-2529.

Berger, David and Joseph Vavra. Forthcoming. Consumption dynamics during recessions. *Econometrica.*

Berry, Steven and Philip A. Haile. Forthcoming. Identification in differentiated products markets using market level data. *Econometrica.*

Berry, Steven, Alon Eizenberg, and Joel Waldfogel. 2013. Optimal product variety in radio markets. The Hebrew University of Jerusalem mimeo.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* 63(4): 841-890.

—. 1999. Voluntary export restraints on automobiles: Evaluating a trade policy. *American Economic Review* 89(3): 400-430.

Bring, Johan. 1994. How to standardize regression coefficients. *The American Statistician* 48(3): 209-213.

Bundorf, M. Kate, Jonathan Levin, and Neale Mahoney. 2012. Pricing and welfare in health plan choice. *American Economic Review* 102(7): 3214-3248.

Chetty, Raj. 2009. Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annual Review of Economics* 1: 451-488.

Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi. 2012. Plausibly exogenous. *Review of Economics and Statistics* 94(1): 260-272.

Crawford, Gregory S. and Ali Yurukoglu. 2012. The welfare effects of bundling in multichannel television markets. *American Economic Review* 102(2): 643-685.

De Nardi, Mariacristina, Eric French, and John B. Jones. 2010. Why do the elderly save? The role of medical expenses. *Journal of Political Economy* 118(1): 39-75.

DellaVigna, Stefano, John A. List, and Ulrike Malmendier. 2012. Testing for altruism and social pressure in charitable giving. *Quarterly Journal of Economics* 127(1): 1-56.

Dubé, Jean-Pierre, Jeremy T. Fox, and Che-Lin Su. 2012. Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica* 80(5): 2231-2267.

Einav, Liran, Amy Finkelstein, and Mark R. Cullen. 2010. Estimating welfare in insurance markets using variation in prices. *Quarterly Journal of Economics* 125(3): 877-921.

Einav, Liran, Amy Finkelstein, Stephen P. Ryan, Paul Schrimpf, and Mark R. Cullen. 2013. Selection on moral hazard in health insurance. *American Economic Review* 103(1): 178-219.

Einav, Liran, Amy Finkelstein, and Paul Schrimpf. 2013. The response of drug expenditure to non-linear contract design: Evidence from Medicare Part D. NBER Working Paper No. 19393.

Fan, Ying. 2013. Ownership consolidation and product characteristics: A study of the US daily newspaper market. *American Economic Review* 103(5): 1598-1628.

Gelman, Andrew. 2008. Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine* 27(15): 2865-2873.

Gentzkow, Matthew. 2007. Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review* 97(3): 713-744.

Gentzkow, Matthew, Jesse M. Shapiro, and Michael Sinkinson. 2014. Competition and ideological diversity: Historical evidence from US newspapers. *American Economic Review* 104(10): 3073-3114.

Goettler, Ronald L. and Brett R. Gordon. 2011. Does AMD spur Intel to innovate more? *Journal of Political Economy* 119(6): 1141-1200.

Gourieroux, Christian S., Alain Monfort, and Eric Renault. 1993. Indirect inference. *Journal of Applied Econometrics* 8: S85-S118.

Gourinchas, Pierre-Olivier and Jonathan A. Parker. 2002. Consumption over the life cycle. *Econometrica* 70(1): 47-89.

Hampel, Frank R., Elvezio M. Ronchetti, Peter J. Rousseeuw, and Werner A. Stahel. 1986. *Robust statistics: The approach based on influence functions.* New York: Wiley-Interscience.

Hansen, Lars P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50(4): 1029-1054.

Heckman, James J. 2010. Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature* 48(2): 356-398.

Jaffe, Sonia and E. Glen Weyl. 2013. The first-order approach to merger analysis. *American Economic Journal: Microeconomics* 5(4): 188-218.

Kaplan, Greg. 2012. Moving back home: Insurance against labor market risk. *Journal of Political Economy*. 120(3): 446-512.

Kawai, Kei and Yasutora Watanabe. 2013. Inferring strategic voting. *American Economic Review* 103(2): 624-662.

Kim, Jae-On and G. Donald Ferree Jr. 1981. Standardization in causal analysis. *Sociological Methods Research* 10(2): 187-210.

Knittel, Christopher R. and Konstantinos Metaxoglou. 2014. Estimation of random-coefficient demand models: Two empiricists' perspective. *Review of Economics and Statistics* 96(1): 34-59.

Leamer, Edward E. 1983. Let's take the con out of econometrics. *American Economic Review* 73(1): 31-43.

Lee, Robin S. 2013. Vertical integration and exclusivity in platform and two-sided markets. *American Economic Review* 103(7): 2960-3000.

Lim, Claire S. H. 2013. Preferences and incentives of appointed and elected public officials: Evidence from state trial court judges. *American Economic Review* 103(4): 1360-1397.

Matzkin, Rosa L. 2007. Nonparametric identification. In Ch. 73 of J. Heckman and E. Leamer (eds.), *Handbook of Econometrics* 6B: 5307-5368. Elsevier.

—. 2013. Nonparametric identification in structural economic models. *Annual Review of Economics* 5: 457-486.

Mazzeo, Michael J. 2002. Product choice and oligopoly market structure. *RAND Journal of Economics* 33(2): 221-242.

Morten, Melanie. 2013. Temporary migration and endogenous risk sharing in village India. Stanford mimeo.

Nevo, Aviv. 2000. Mergers with differentiated products: The case of the ready-to-eat cereal industry. *RAND Journal of Economics* 31(3): 395-442.

Nevo, Aviv and Adam M. Rosen. 2012. Identification with imperfect instruments. *Review of Economics and Statistics* 94(3): 659-671.

Newey, Whitney K. and Daniel McFadden. 1994. Large sample estimation and hypothesis testing. In R. Engle and D. McFadden (eds.), *Handbook of Econometrics* 4: 2111-2245. Amsterdam: North-Holland.

Nikolov, Boris and Toni M. Whited. 2014. Agency conflicts and cash: Estimates from a dynamic model. *Journal of Finance* 69(5): 1883-1921.

Pakes, Ariel. 2003. Common sense and simplicity in empirical industrial organization. *Review of Industrial Organization* 23(3): 193-215.

Petrin, Amil. 2002. Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy* 110(4): 705-729.

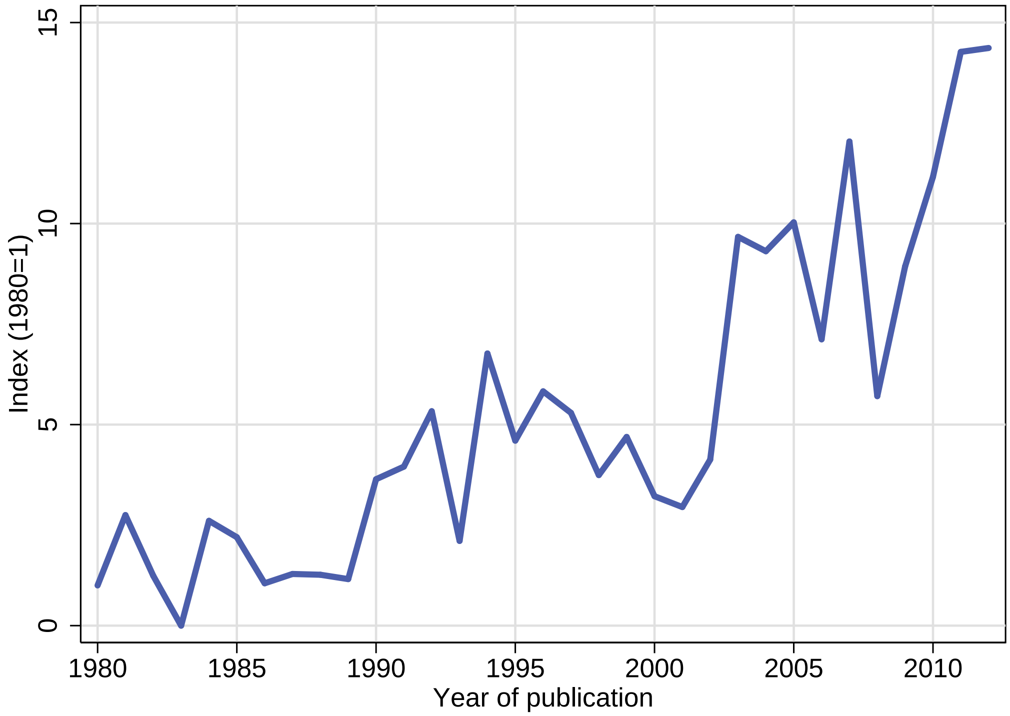
Ronchetti, Elvezio and Fabio Trojani. 2001. Robust inference with GMM estimators. *Journal of Econometrics* 101(1): 37-69.

Saltelli, Andrea, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. 2008. *Global sensitivity analysis: The primer*. West Sussex, UK: Wiley-Interscience.

Smith, Anthony A. 1993. Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics* 8: S63-S84.

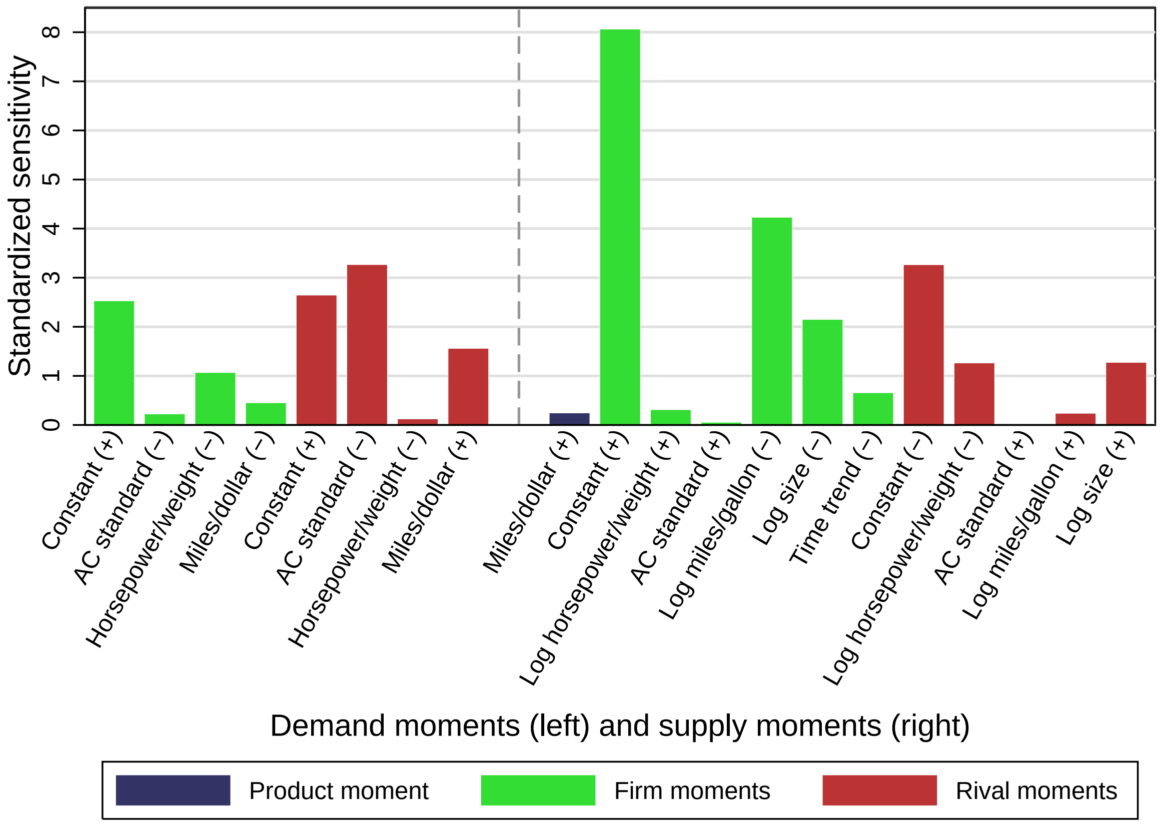
Sobol, Ilya. M. 1993. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiments* 1(4): 407-414.

Figure 1: Share of top journal articles containing the phrase "identified by"



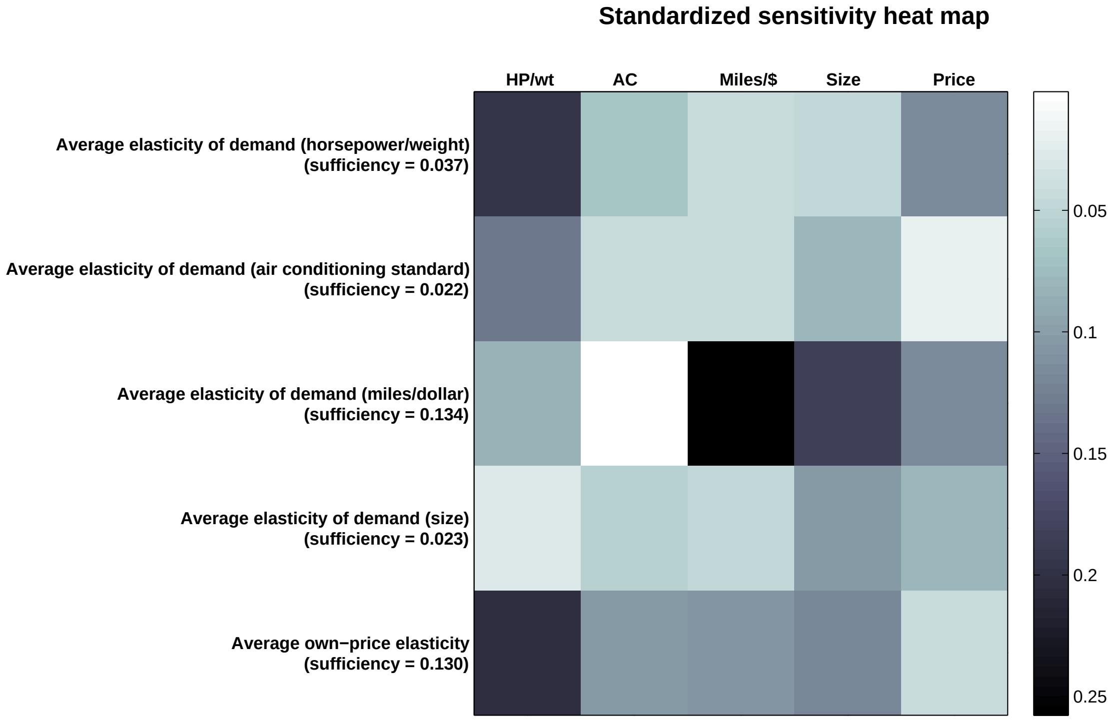
Notes: The plot shows an annual index of the share of articles published in the *American Economic Review,* the *Journal of Political Economy,* the *Quarterly Journal of Economics,* the *Review of Economic Studies,* and *Econometrica* containing the phrase "is identified by" or "are identified by" along with the word "data," among all articles containing the word "data." Cases where the word "identified" is not used in the econometric sense are manually excluded.

Figure 2: Standardized sensitivity for average markup in Berry et al. (1995)



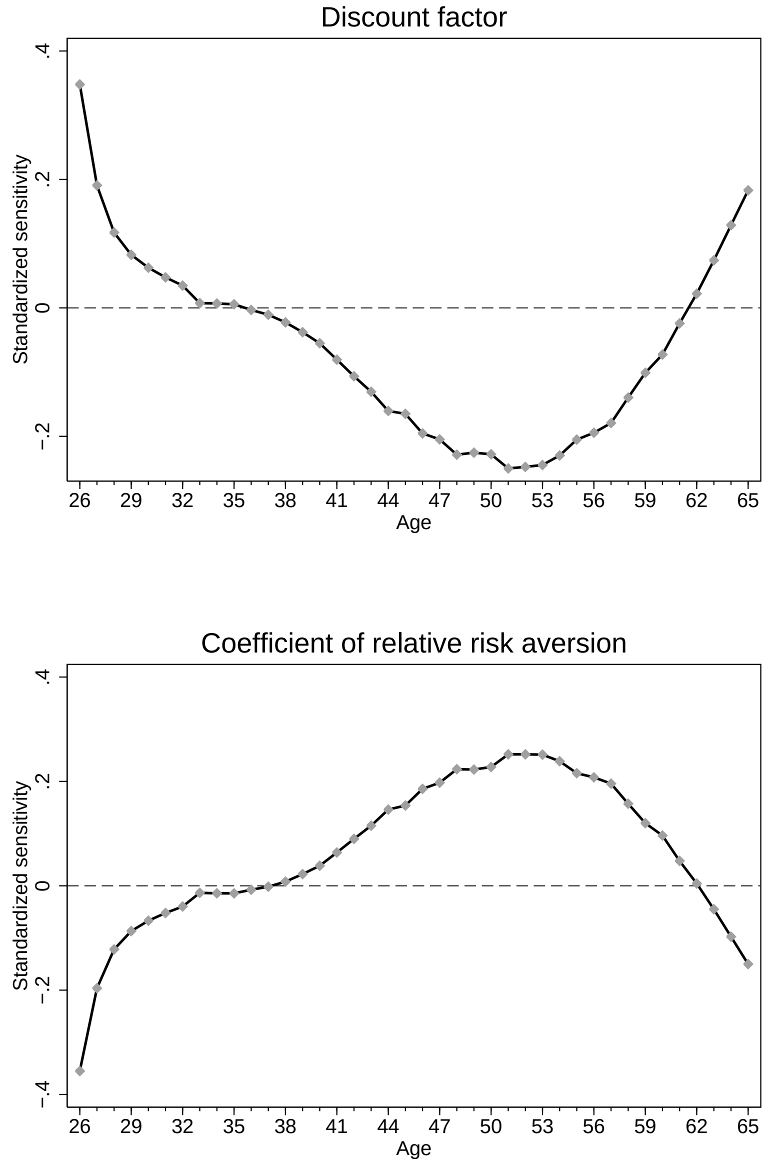
Notes: The plot shows the absolute value of the standardized sensitivity of the implied average markup with respect to the estimation moments, with the sign of sensitivity in parentheses. The left section of the plot shows sensitivity to demand-side moments, while the right section shows sensitivity to supply-side moments. All estimation moments (except for those corresponding to the "excluded" constant instruments) use demeaned instruments. While sensitivity is computed with respect to the complete set of estimation moments, the plot only shows those corresponding to the "excluded" instruments. As a result of collinearity in the instrument matrix, we drop three instruments (two demand-side and one supply-side) prior to estimation: "Firm products: Size," "Rival products: Size," and "Rival products: Time trend."

Figure 3: Standardized sensitivity for elasticities of demand in Berry et al. (1995)



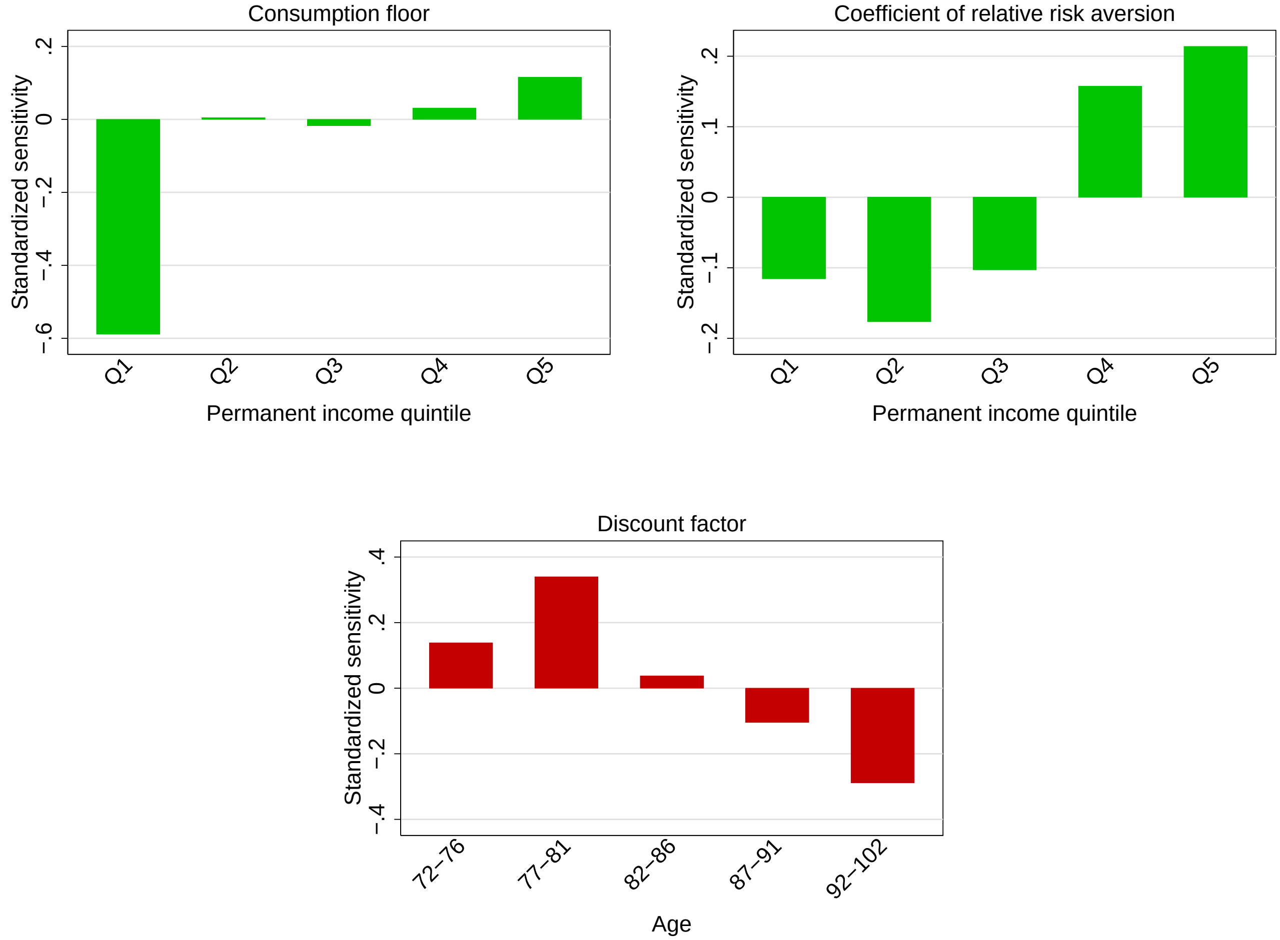
Notes: The plot shows a heat map of the absolute value of standardized sensitivity of the average own-price or own-characteristic elasticity of demand from the BLP model (in rows) with respect to the vector of analogous elasticities from a logit model with the same excluded instruments as the BLP model (in columns). The number in parentheses in each row is the sufficiency of the vector of logit model elasticities for the BLP model elasticity.

Figure 4: Standardized sensitivity for select parameters in Gourinchas and Parker (2002)



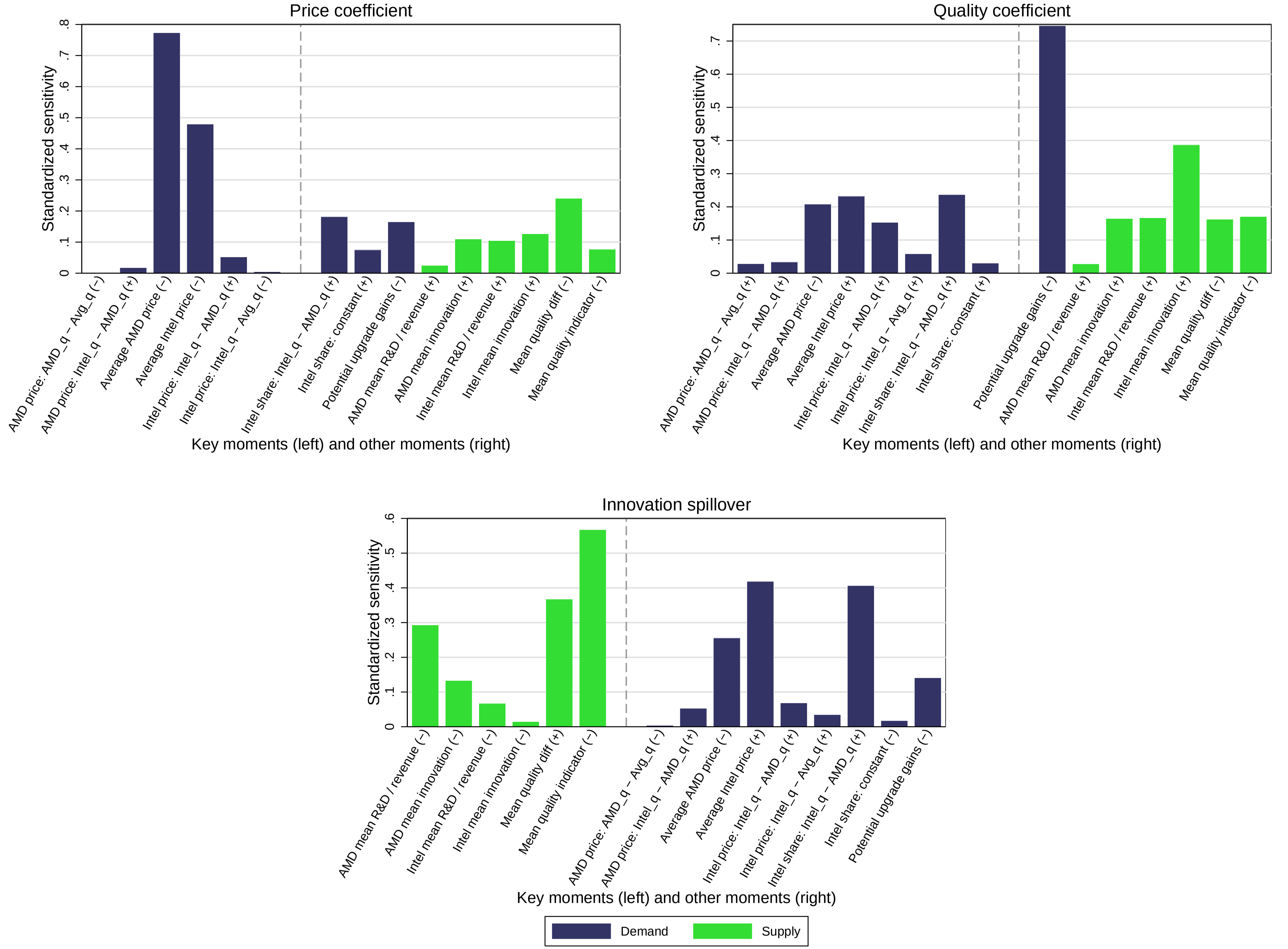
Notes: Each plot shows the standardized sensitivity of the parameter named in the plot title with respect to the full vector of estimation moments, which are the mean adjusted consumption levels at each age.

Figure 5: Standardized sensitivity for select parameters in De Nardi et al. (2010)



Notes: Each plot shows the standardized sensitivity of the parameter named in the plot title with respect to median asset holdings averaged by permanent income quintile (for the consumption floor and risk aversion) or averaged by age group (for the discount factor).

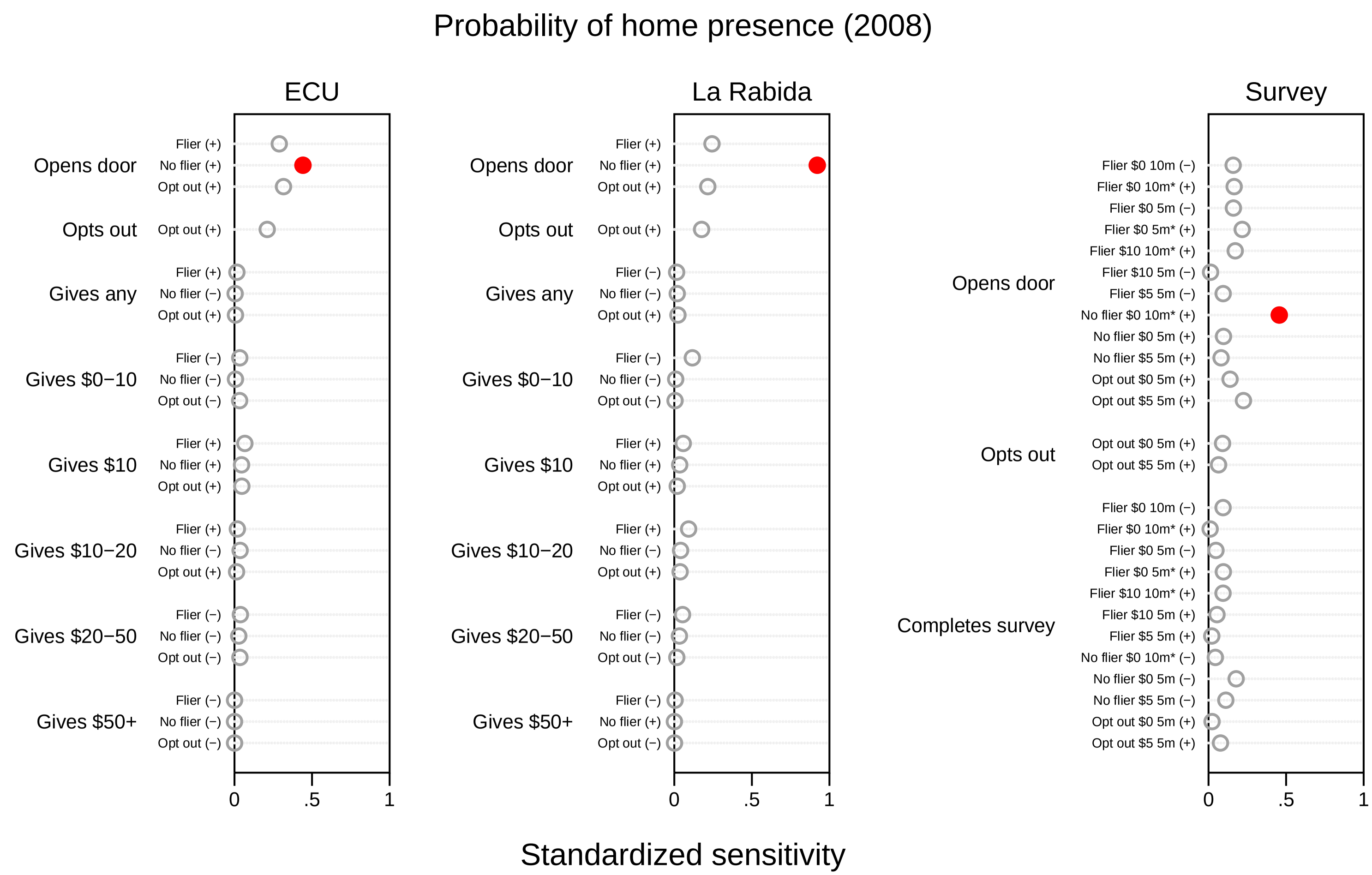
Figure 6: Standardized sensitivity for select parameters in Goettler and Gordon (2011)



Notes: Each plot shows the absolute value of standardized sensitivity of the parameter named in the plot title with respect to the vector of estimation moments, with the sign of sensitivity in parentheses. A “key moment” is a moment that Goettler and Gordon (2011) highlight as especially sensitive to the given parameter. The plots use the following shorthand: “q” stands for quality of the product offere; “Avg_q” stands for the mean quality of products currently owned; “Intel share: X” stands for the coefficient on X from a regression of Intel share on the difference in quality between Intel and AMD; and “Y price: X” stands for the coefficient on X from a regression of firm Y’s price on both the difference in quality between Intel and AMD, and the difference between Y’s quality and the mean quality of owned products.

Figure 7: Standardized sensitivity for select parameters in DellaVigna et al. (2012)

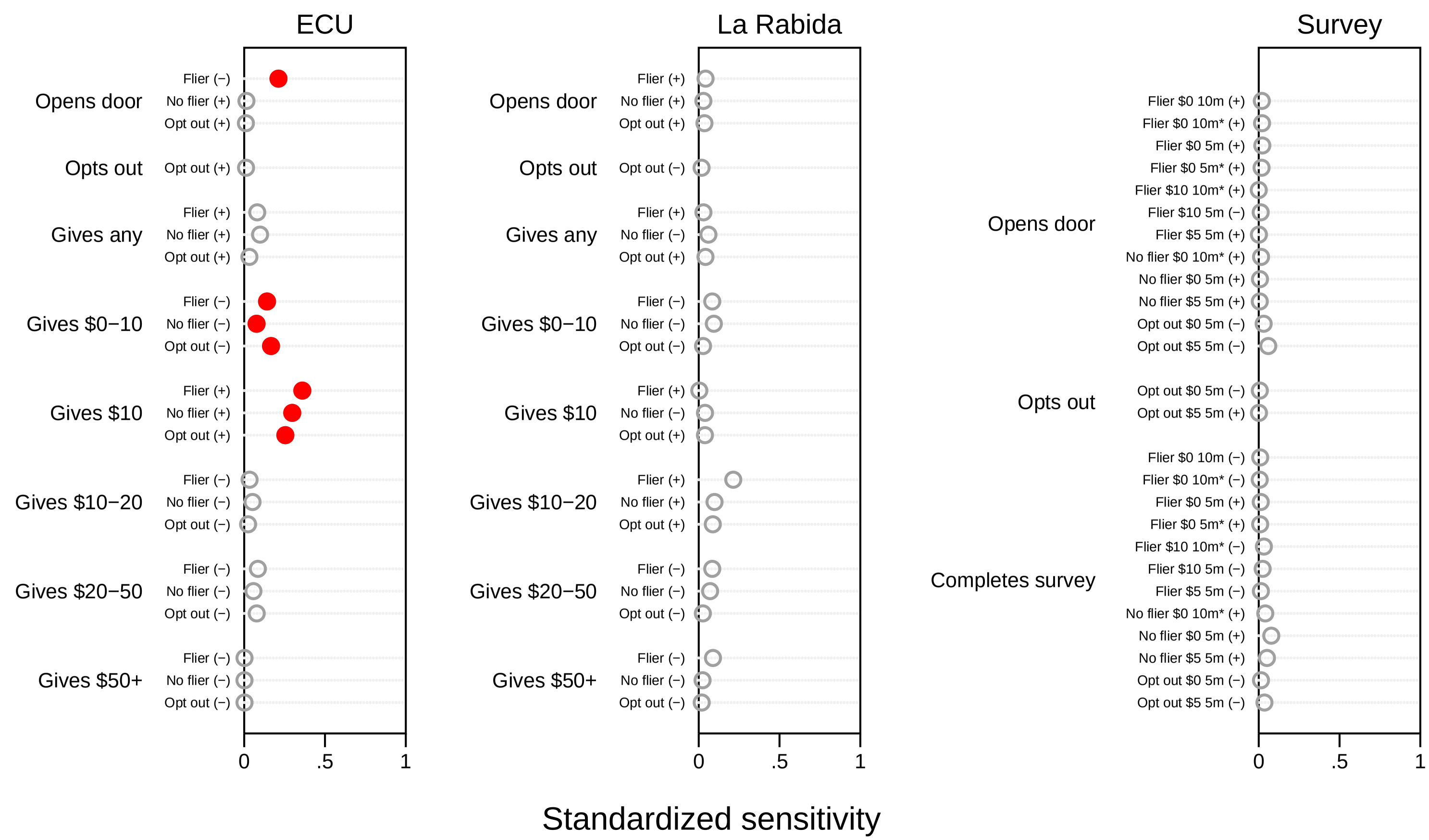
Panel A



Notes: Each plot shows the absolute value of standardized sensitivity of the parameter named in the plot title with respect to the full vector of estimation moments, with the sign of sensitivity in parentheses. A "key moment" is a moment that DellaVigna et al. (2012) highlight as important for identifying the given parameter. 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.

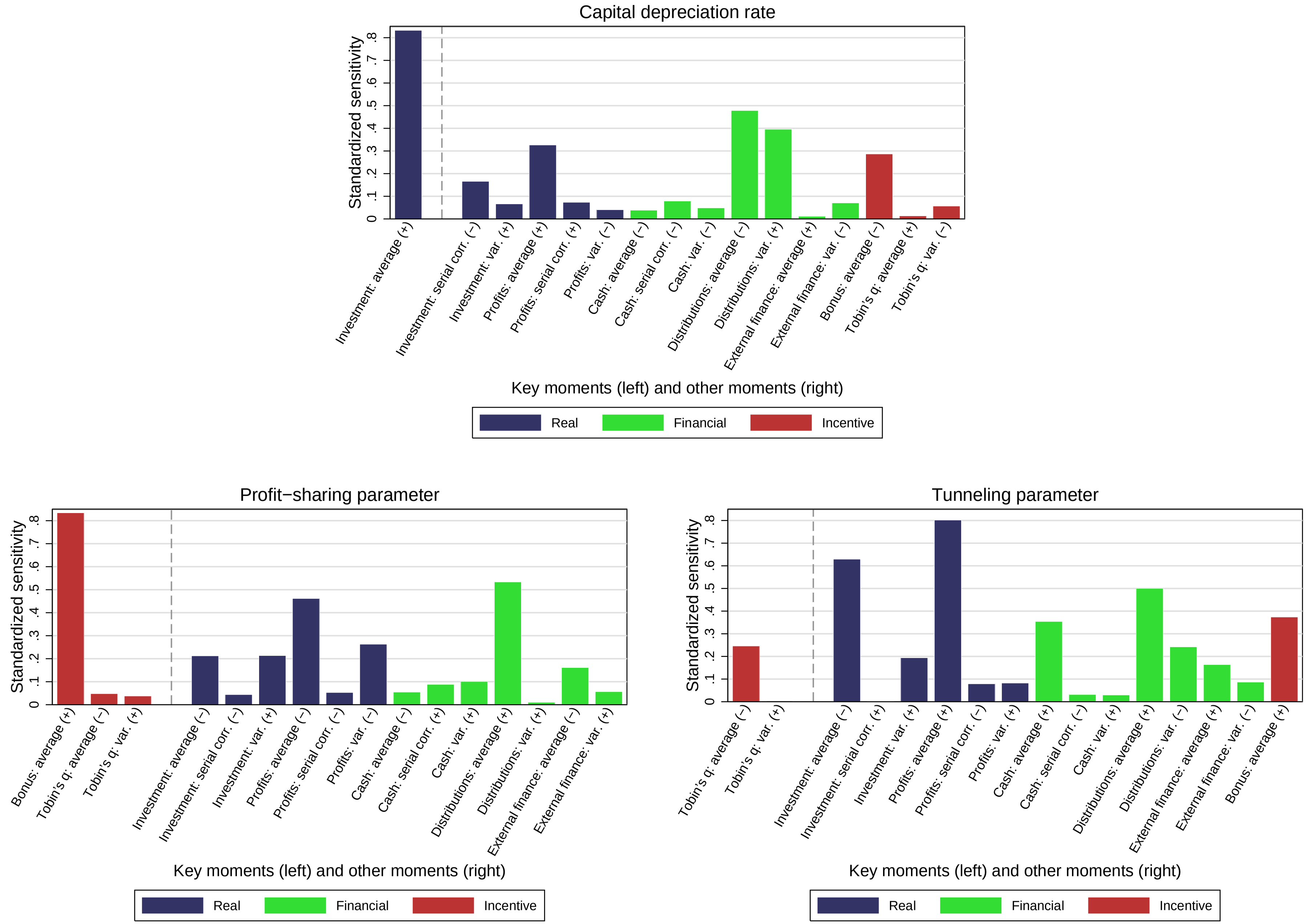
Panel B

Social pressure cost of giving 0 in person (ECU)



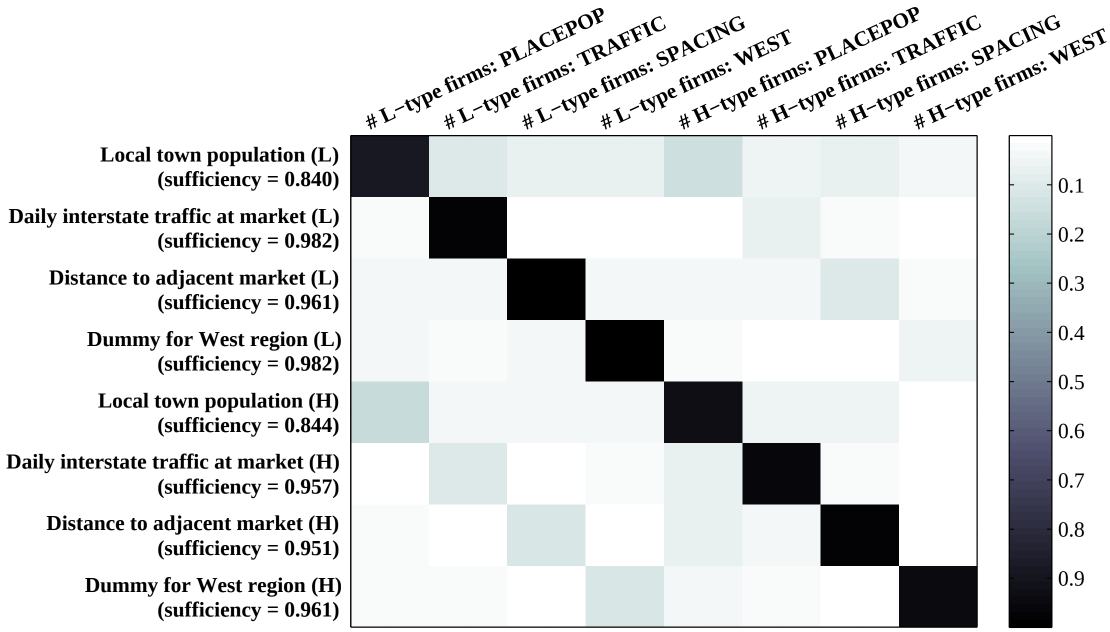
Notes: Each plot shows the absolute value of standardized sensitivity of the parameter named in the plot title with respect to the full vector of estimation moments, with the sign of sensitivity in parentheses. A “key moment” is a moment that DellaVigna et al. (2012) highlight as important for identifying the given parameter. 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.

Figure 8: Standardized sensitivity for select parameters in Nikolov and Whited (2014)



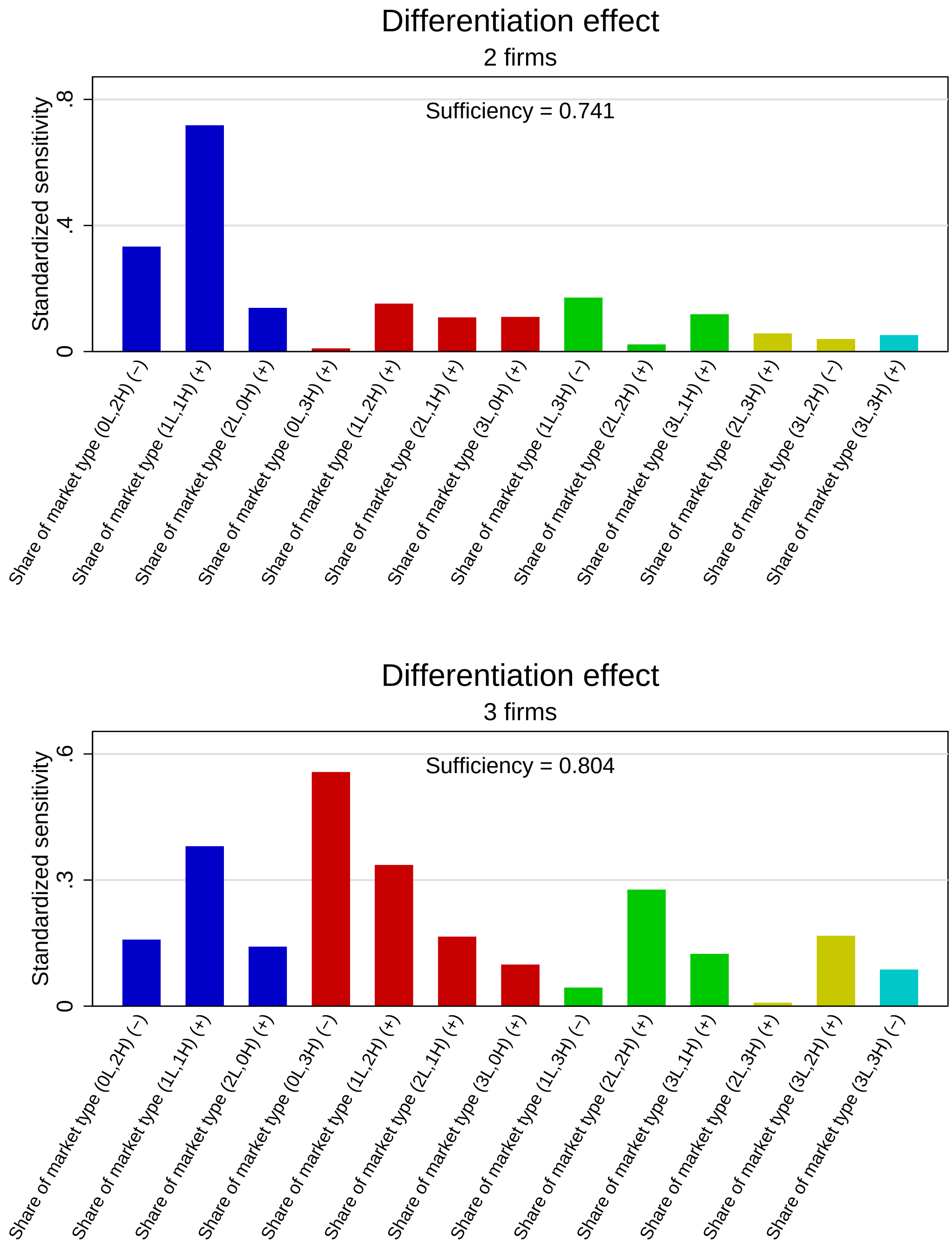
Notes: Each plot shows the absolute value of standardized sensitivity of the parameter named in the plot title with respect to the full vector of estimation moments, with the sign of sensitivity in parentheses. A "key moment" is a moment that Nikolov and Whited (2014) highlight as important for identifying the given parameter.

Figure 9: Standardized sensitivity for market characteristics parameters in Mazzeo (2002)



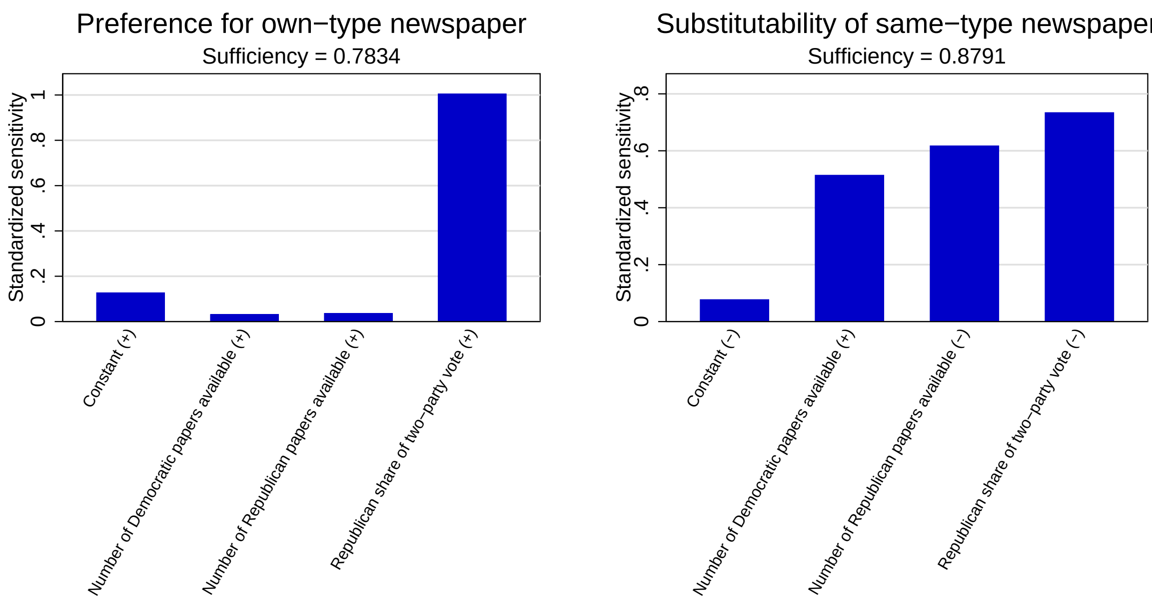
Notes: The plot shows a heat map of the absolute value of the standardized sensitivity of a model parameters (in rows) with respect to a vector of descriptive statistics (in columns). Each row also shows the sufficiency of the vector of statistics for the given parameter. Parameter names ending in "(L)" refer to effects on low-type payoffs, and parameter names ending in "(H)" refer to effects on high-type payoffs. The descriptive statistics are the coefficients from regressions of the number of low- and high-type firms on observable market characteristics. The model is the two-type Stackelberg model defined and estimated in Mazzeo (2002).

Figure 10: Standardized sensitivity for differentiation effect in Mazzeo (2002)



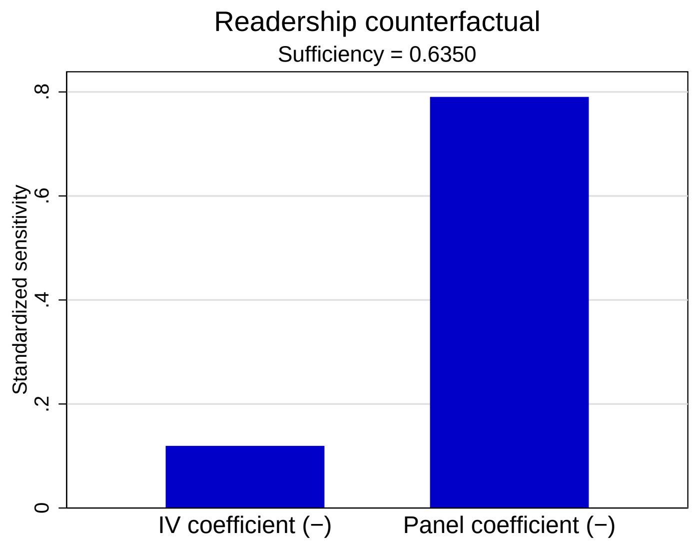
Notes: Each plot shows the absolute value of standardized sensitivity of the counterfactual named in the plot title with respect to the vector of descriptive statistics listed on the x-axis, with the sign of sensitivity in parentheses and the sufficiency of the vector of descriptive statistics for the given counterfactual listed above the plot. The $n-$firm differentiation effect is the probability that the n^{th} entrant would choose type L in a market in which the $(n-1)$ prior entrants are type H and would choose type H if the $(n-1)$ prior entrants are type L. The descriptive statistics are the empirical shares of 2-6 firm markets with each configuration (number of low and high type firms). The model is the two-type Stackelberg model defined and estimated in Mazzeo (2002).

Figure 11: Standardized sensitivity for select parameters in Gentzkow et al. (2014)



Notes: Each plot shows the absolute value of the standardized sensitivity of the parameter named in the plot title with respect to the vector of descriptive statistics listed on the x-axis, with the sign of sensitivity in parentheses and the sufficiency of the vector of descriptive statistics for the given parameter listed above the plot. The descriptive statistics are the estimated parameters from a regression of the log of the ratio of Republican to Democratic circulation in a given town on a constant, the number of Republican newspapers available in the town, the number of Democratic newspapers available in the town, and the Republican share of the two-party vote.

Figure 12: Standardized sensitivity for counterfactual estimate in Gentzkow (2007)



Notes: The plot shows the absolute value of standardized sensitivity for the readership counterfactual with respect to the two descriptive statistics listed on the x-axis, with the sign of sensitivity in parentheses and the sufficiency of the vector of descriptive statistics for the given parameter listed above the plot. The readership counterfactual is the change in readership of the print edition of the *Washington Post* when the post.com is removed from the choice set (Gentzkow 2007, table 10). The IV coefficient is the estimated coefficient from a two-stage least squares regression of last-five-weekday *Washington Post* print readership on last-five-weekday post.com readership, with a set of excluded instruments including Internet access at work (reported in Gentzkow 2007, table 4, IV specification (1)). The panel coefficient is the coefficient from an OLS regression of last-one-day print readership on last-one-day online readership controlling flexibly for readership of both editions in the last five weekdays. Each of these auxiliary regressions includes the standard set of demographic controls from Gentzkow (2007).