Measuring the Sensitivity of Parameter Estimates to Sample Statistics

Matthew Gentzkow*
Jesse M. Shapiro
Chicago Booth and NBER

First Version: December 2012
This Version: October 2013

Abstract

We propose a measure of the dependence of a parameter estimate on specific features of the data. The measure can be computed at negligible cost even for computationally difficult models. It can be interpreted as a measure of sensitivity to model misspecification. The measure delivers satisfactory intuitions for pen-and-paper examples. We apply the measure to recent empirical papers in industrial organization.

1 Introduction

An estimator is a mapping from data to parameters of interest. Knowing the form of this mapping—how the parameters change as the data vary along particular dimensions—is valuable, as it provides intuition for the workings of the estimator and the role of underlying assumptions. It is often difficult to interrogate an estimator's behavior directly, however, as many interesting econometric models are computationally expensive to estimate.

In this paper, we formulate a measure of the sensitivity of an estimator to specific features of the data that can be approximated at little or no cost even for computationally difficult models. We show that our measure matches common intuitions in well-understood cases and can be interpreted as a measure of sensitivity to model misspecification. We apply our measure to two recent empirical papers in industrial organization, where it delivers useful and precise intuitions about the behavior of the estimators.

Throughout the paper we consider the following abstract setting. A researcher computes an estimate $\hat{\theta}$ of some economic parameters with true value θ_0. The researcher also computes a vector of statistics $\hat{\gamma}$ that summarize the data. 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 $\hat{\gamma}$ as the expected coefficient from a regression of $\hat{\theta}$ on $\hat{\gamma}$ in data drawn from the asymptotic distribution. Sensitivity measures how $\hat{\theta}$ is related to the elements of $\hat{\gamma}$ across alternative realizations of the data.

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 asymptotic mean of $\hat{\gamma}$. If $\hat{\theta}$ depends only on a subset of elements of $\hat{\gamma}$, then only elements of that subset will have nonzero sensitivity. Moreover, if $\hat{\theta}$ is connected to $\hat{\gamma}$ through a set of moment conditions, sensitivity to $\hat{\gamma}$ is equivalent to sensitivity to small perturbations to the corresponding moment conditions. In this sense, our measure simultaneously captures both sensitivity to the data and sensitivity to model misspecification.

Our measure delivers appropriate intuitions for linear models. For example, in a regression of y on two orthogonal covariates x_1 and x_2, it implies that the coefficient on x_1 is sensitive to the covariance between y and x_2 only to the extent that x_1 and x_2 are correlated. In a two-stage least squares regression with one endogenous variable x and two orthogonal instruments z_1 and z_2, it implies that the sensitivity of the coefficient on x to a single-instrument IV estimate is proportional to the instrument's strength in the first-stage regression.

Our measure can be readily applied to computationally expensive nonlinear models. When $\hat{\gamma}$ are the moments used to estimate $\hat{\theta}$, a consistent estimator of sensitivity can be obtained through transformations of objects typically used to compute the asymptotic variance of $\hat{\theta}$. When $\hat{\gamma}$ is some other statistic not directly involved in estimation of $\hat{\theta}$, sensitivity can be estimated via an OLS regression of empirical influence components, which are in turn available at little or no cost for a wide class of estimators.

We apply our measure to Goettler and Gordon's (2011) study of the effect of competition from AMD on Intel's incentive to invest in R&D. Goettler and Gordon's (2011) model is economically rich and computationally difficult, yet it is trivial to measure their estimator's sensitivity to empirical moments. Estimated sensitivity provides quantitative grounding for Goettler and Gordon's (2011) discussion of the empirical identification of their model.

We also apply our measure to Gentzkow et al.'s (2013) model of newspaper demand with multiple readership.

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 variances. Generalizing our approach to more global exploration of model properties is conceptually straightforward but may be computationally expensive. 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 provide a normal-

ization that serves as a useful default for most practical applications but acknowledge that the appropriate scaling of sensitivity may be model-specific.

Our paper contributes mainly to two literatures in economics. The first is an applied literature using structural models for empirical inference. It is common in such papers to provide a heuristic discussion of a model's "empirical identification"—how the model's parameters are related to intuitive features of the data—sometimes including a descriptive or reference model.[1] Our sensitivity measure allows researchers to make precise, quantitative statements about the empirical determinants of structural parameters and their connection to useful descriptive statistics.

The second is a methodological conversation about the appropriate role of structural economic models in inference. At the center of this conversation is a perceived tradeoff between the credibility of a model's economic assumptions and the transparency of its mapping from data to parameters.[2] Our sensitivity measure makes this tradeoff shallower by permitting a precise characterization of the dependence of an estimate on intuitive features of the data and on modeling assumptions.[3]

Our sensitivity measure complements the "sufficient statistics" approach, in which answers to questions of interest are posed as functions of empirical objects rather than structural parameters (Chetty 2009; Einav et al. 2010; Jaffe and Weyl forthcoming). Our sensitivity measure correctly identifies cases in which only a subset of empirical moments affect a given parameter. Our measure is typically less expensive to obtain than a sufficient statistics representation, though the resulting quantitative measure is less general than the theoretical results typically offered in the sufficient statistics approach.

Our measure facilitates traditional sensitivity analysis (Leamer 1983) by showing how data map into parameters and by showing which of a set of identifying assumptions matters most for a given parameter. In this sense our paper complements recent research on inference in the presence of possibly misspecified exclusion restrictions (Conley et al. 2012; Nevo and Rosen 2012).

Our work is closely related to the large literature on sensitivity analysis for scientific models (Sobol 1993, Saltelli et al. 2008). 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.

The remainder of the paper is organized as follows. Section 2 defines our measure. Section 3 discusses its properties and interpretation, and section 4 shows how to estimate sensitivity. Section 5 applies the measure to pen-and-paper examples. Section 6 compares our approach to an alternative based on the sensitivity of moments to parameters rather than the sensitivity of parameters to moments. Section 7 applies the mea-

---

[1]For example, Cohen and Einav (2007) devote a section of their paper to "provid[ing] intuition for which features of the data allow [them] to identify particular parameters of the model." Crawford and Yurukoglu (2012) note that "one may casually think of [a set of empirical moments] as 'empirically identifying' [a set of related parameters]." Berry, Levinsohn and Pakes (2004) present a mix of formal results and intuition to characterize the role of different empirical moments in their analysis. Nevo (2001) presents analysis of a homogeneous logit model as a "useful tool in getting a feel for the data."

[2]Heckman (2010) writes that "The often complex computational methods that are required to implement [structural estimation] make it less transparent." 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."

[3]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).

sure to two empirical papers in industrial organization. Section 8 concludes with a discussion of how to generalize the measure so that it is not local to a particular dataset.

2 Measure of Sensitivity

An econometrician possesses a sample of size n. She computes (i) an estimate $\hat{\theta}$ of a parameter θ and (ii) an auxiliary statistic $\hat{\gamma}$. It is helpful in some cases to think of $\hat{\theta}$ as a scalar and $\hat{\gamma}$ as a vector, but we do not restrict attention to this case.

We assume that

$$
(1) \qquad \sqrt{n}\left(\begin{array}{c} \hat{\theta} - \theta_0 \\ \hat{\gamma} - \gamma_0 \end{array}\right) \xrightarrow{d} N(0, \Sigma),
$$

where θ_0 is the true value of parameter θ, γ_0 is some constant, and Σ is finite. In what follows statistical operators (expectations, variances) are with respect to the asymptotic distribution except where stated.

It follows that the conditional expectation of $\hat{\theta}$ given $\hat{\gamma}$ is linear. Let $\Sigma_{\gamma\gamma}$ and $\Sigma_{\theta\gamma}$ denote the submatrices of Σ corresponding to the variance of $\hat{\gamma}$ and the covariance of $\hat{\theta}$ and $\hat{\gamma}$ respectively. Then,

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

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

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

Our measure of sensitivity Λ is the matrix relating the expected value of $\hat{\theta}$ to $\hat{\gamma}$. This is the expected coefficient from a regression of $\hat{\theta}$ on $\hat{\gamma}$ across repeated draws from their joint asymptotic distribution. An element of Λ is the effect of changing a single element of $\hat{\gamma}$ on the expected value of a particular parameter, holding constant the other elements of $\hat{\gamma}$.

Because the units of $\hat{\gamma}$ may differ across its elements, it is often useful to normalize Λ. Let Λ_{ij} be a typical element of Λ corresponding to the sensitivity of $\hat{\theta}_i$ to $\hat{\gamma}_j$. We will call $\tilde{\Lambda}_{ij} = \Lambda_{ij}\sqrt{\text{Var}(\hat{\gamma}_j)/\text{Var}(\hat{\theta}_i)}$ the *scaled sensitivity*. Scaled sensitivity measures how much a one-standard-deviation change in $\hat{\gamma}_j$ affects the expected value of $\hat{\theta}_i$, fixing other elements of $\hat{\gamma}_j$, in units of the asymptotic standard deviation of $\hat{\theta}_i$. In general the appropriate normalization will depend on the economic application, but this one provides a sensible default.

Special Cases

To illustrate the properties of our measure, it is helpful to consider several specific classes of estimators. We follow Newey and McFadden's (1994) notation throughout.

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

$$
(3) \qquad \hat{\theta} \;=\; \underset{\theta \in \Theta}{\arg\min}\, \hat{g}_n(\theta)' \hat{W}_g \hat{g}_n(\theta)
$$

4

where $\hat{g}_n(\theta)$ is a function of parameters and data and the weight matrix $\hat{W}_g$ is positive semi-definite, as is its probability limit W_g.

Except where stated we will assume that (i) $\hat{\theta} \xrightarrow{P} \theta_0$, (ii) $\hat{g}_n(\theta)$ is continuously differentiable in θ with a gradient $\hat{G}(\theta)$ that converges uniformly in probability to some continuous $G(\theta)$ with $G = G(\theta_0)$, (iii) the limit θ_0 is interior to the compact set Θ, (iv) the matrix $(G'WG)$ is nonsingular, and (v) the statistic $\hat{g}_n(\theta_0)$ is asymptotically normal with mean 0 and variance Ω_{gg}. Under these conditions an MDE is asymptotically normal with mean θ_0 (Newey and McFadden, Theorem 3.2).

It is often useful to consider the case where the statistics $\hat{\gamma}$ are the empirical moments.

Definition. Λ is the *sensitivity to moments* if $\hat{\theta}$ is an MDE and $\hat{\gamma} = \hat{g}_n(\theta_0)$.

We will often make use of the exactly identified linear instrumental variables model to build intuition.

Example. $\hat{\theta}$ is a *linear instrumental variables estimator* (IV) if it is an MDE with $\hat{W}_g = I$ and $\hat{g}_n(\theta) = \frac{1}{n}(z'(y - x\theta))$ where z is an $n \times k$ array of instruments, x is an $n \times k$ array of regressors, and y is an $n \times 1$ vector of outcomes.

3 Interpretations of Sensitivity

3.1 Model Misspecification

Although we have written Λ as a measure of sensitivity to data, it is also a measure of sensitivity to assumptions. To see this, suppose that $\hat{\theta}$ is an MDE, but that the model is misspecified in that $\hat{g}_n(\theta_0)$ has asymptotic mean ε in a neighborhood of 0. Define θ_ε so that $\hat{g}_n(\theta_\varepsilon) \xrightarrow{P} 0$. Then if $\hat{\theta} \xrightarrow{P} \theta_\varepsilon \in \text{int}(\Theta)$, $\hat{\theta}$ is asymptotically normal with bias $b(\varepsilon)$. Let B be the gradient of $b(\varepsilon)$ with respect to ε at $\varepsilon = 0$. The value B measures the sensitivity of the bias of $\hat{\theta}$ to a small perturbation of the moment conditions. It is straightforward to show that $B = \Lambda$ where Λ is the sensitivity to moments.

If $\hat{\theta}$ is an IV, then $\frac{1}{n}(z'(y - x\theta)) \xrightarrow{P} \varepsilon$ implies that $\hat{\theta} \xrightarrow{P} \theta_0 + \Lambda\varepsilon$. That is, the bias $b(\varepsilon) = \Lambda\varepsilon$, where Λ is the sensitivity and ε is the asymptotic covariance between the instruments and the error term. A researcher can therefore use Λ to evaluate the credibility of the estimates of $\hat{\theta}$ as a function of her priors over the credibility of each exclusion restriction.

In the case of an overidentified model an alternative route to evidence on sensitivity to misspecification is to re-estimate the model dropping a suspicious moment condition. This approach has a close relationship to Λ. Suppose only the jth moment is misspecified, i.e. $\varepsilon_k = 0 \forall k \neq j$. Let $\hat{\theta}_{\sim j}$ be the estimate that solves equation (3) excluding the jth component of $\hat{g}_n(\theta)$. Assume that the model is overidentified in the sense that $\hat{\theta}_{\sim j} \to \theta_0$. Then the derivative of the asymptotic mean of $(\hat{\theta} - \hat{\theta}_{\sim j})$ with respect to ε_j is the jth column of Λ.

3.2 Almost Sufficient Statistics

Our measure is useful in identifying cases in which a subset of data elements are sufficient to recover $\hat{\theta}$.[4] Observe that if $\hat{\theta} = f(\hat{\gamma})$ for some continuously differentiable function $f()$ with derivatives $F()$, then $\Lambda = F(\gamma_0)$. An immediate implication is that if $f()$ can be written as a function of a subset of the elements of $\hat{\gamma}$, then Λ assigns a coefficient of zero to all elements of $\hat{\gamma}$ not in the subset.

If $\hat{\theta}$ is an MDE and Λ is the sensitivity to moments, this result says that Λ will communicate which moments contribute to a particular parameter. Suppose for example that $\hat{\theta}$ is an IV. Suppose that $\left(\frac{1}{n}z'x\right)$ is diagonal, i.e. each instrument is correlated with one and only one regressor. Then the ith instrument can be thought of as sufficient for $\hat{\theta}_i$, in the sense that $\hat{\theta}_i$ can be written as a function of the ith moment condition alone.

A similar logic applies when $\hat{\theta}$ cannot be written as a function of $\hat{\gamma}$. Define the *sufficiency* of $\hat{\gamma}$ for an element of $\hat{\theta}$ as

$$(4) \qquad \Delta = \frac{\mathrm{Var}\left(\mathrm{E}\left(\hat{\theta} - \theta_0 | \hat{\gamma}\right)\right)}{\mathrm{Var}\left(\hat{\theta} - \theta_0\right)}.$$

The value $\Delta \in [0,1]$ is the probability limit of the R^2 of a regression of $\hat{\theta}$ on $\hat{\gamma}$ across repeated draws form their joint asymptotic distribution, as the number of draws grows large. If $\hat{\theta}$ and $\hat{\gamma}$ are asymptotically independent, then $\Delta = 0$. If $\hat{\theta} = f(\hat{\gamma})$, then $\Delta = 1$.

When Δ is close to 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. In such a case, if Λ reveals that $\hat{\theta}$ is meaningfully sensitive only to a subset of the elements of $\hat{\gamma}$, that subset can be thought of as almost sufficient.

As a simple example, consider the case of estimating the mean θ of an i.i.d. scalar random variable. Let $\hat{\gamma}$ denote a vector of means of independent samples j of size $\phi_j n$ with $\sum_j \phi_j = 1$. Then the efficient MDE $\hat{\theta} = \Phi\hat{\gamma}$ where Φ is a vector whose j^{th} element is ϕ_j. It follows that sensitivity $\Lambda = \Phi$ and $\Delta = 1$. The sufficiency of any given element $\hat{\gamma}_j$ is ϕ_j, i.e. a given element $\hat{\gamma}_j$ is almost sufficient for $\hat{\theta}$ if sample j contains the majority of the data.

3.3 Value of Information

From equation (2), it is immediate that Λ_{ij}^2 is the partial derivative of the variance in $\mathrm{E}\left(\hat{\theta}_i | \hat{\gamma}\right)$ with respect to the variance of $\hat{\gamma}_j$. In this sense, Λ captures not only the impact of $\hat{\gamma}$ on $\hat{\theta}$, but also the impact of *uncertainty* about γ_0 on uncertainty about θ_0. In the special case where Λ is sensitivity to moments, it shows how the standard error on a given parameter estimate varies with the precision of each moment.

To give this a more concrete interpretation, consider a decision-maker who observes $\hat{\gamma}$ but not $\hat{\theta}$. The decision-maker submits a guess for θ_{0i} and pays a quadratic penalty $\mathrm{E}(\theta_i - \theta_{0i})^2$. If the decision-maker chooses $\theta_i = \mathrm{E}\left(\hat{\theta}_i | \hat{\gamma}\right)$ as her guess, the derivative of her loss with respect to the variance of $\hat{\gamma}_j$ is Λ_{ij}^2.

Suppose that $\hat{\theta}$ is an MDE and Λ is the sensitivity to moments. Suppose that the moments $\hat{g}_n(\theta_0)$ are asymptotically independent so that Ω_{gg} is diagonal, and that the decision-maker can influence the data

---

[4]We follow Chetty (2009) in borrowing the term *sufficient* from statistics with apologies for its abuse.

collection process to improve the precision of a given moment holding all else constant. Then the value of a reduction in the variance of the jth moment is given by Λ_{ij}^2.

4   Computing Sensitivity

In this section we show that consistent estimators for sensitivity Λ are typically readily available and inexpensive to compute even for computationally difficult models. We focus on the case in which $\hat{\theta}$ is obtained as an MDE, a class that encompasses generalized method of moments (GMM), maximum likelihood (MLE), and classical minimum distance (CMD).

Begin with the case in which Λ is the sensitivity to moments. It is straightforward to show that $\Lambda = \left(G'W_g G \right)^{-1} G'W_g$.

By assumption the researcher possesses a consistent estimate of W_g, namely $\hat{W}_g$. The sample analogue $\hat{G} = \hat{G}(\hat{\theta})$ is a consistent estimate of G and is typically already in hand to compute the asymptotic variance of $\hat{\theta}$.[5] Therefore in typical applications computing an estimate of Λ imposes no additional computational burden once the asymptotic variance of the estimator has been computed.

In the more general case in which Λ is not sensitivity to moments, the most convenient way to estimate Λ depends on how $\hat{\gamma}$ is obtained. As a general framework we assume that $\hat{\gamma}$ is an MDE with analogues $\hat{m}_n(\theta)$, $\hat{M}$ and $\hat{W}_m$ of $\hat{g}_n(\theta)$, $\hat{G}$ and $\hat{W}_g$, respectively.

An important special case is where both $\hat{\theta}$ and $\hat{\gamma}$ are estimated via GMM (Hansen 1982). Then $\hat{g}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} g(z_i, \theta)$ and $\hat{m}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} m(z_i, \gamma)$ for i.i.d. data z_i and functions $g(z, \theta)$ and $m(z, \theta)$ satisfying $E(g(z, \theta_0)) = E(m(z, \gamma_0)) = 0$. Let $\tilde{g}_i = \left(\hat{G}'\hat{W}_g\hat{G}\right)^{-1} \hat{G}'\hat{W}_g g(z_i, \hat{\theta})$ and define $\tilde{m}_i$ analogously. Then a consistent estimator for Λ is obtained via a regression of $\tilde{g}_i$ on $\tilde{m}_i$.[6] The intuition for this result is that $\tilde{g}_i$ and $\tilde{m}_i$ are plug-in estimators for the influence of observation i, a measure of the observation's effect on the value of the estimator (Hampel et al. 1986; Ronchetti and Trojani 2001). A regression of $\tilde{g}_i$ on $\tilde{m}_i$ thus recovers 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.

Because many estimators (e.g., MLE, OLS, 2SLS) can be formulated as GMM, the case above is general enough to accommodate most applications of practical interest. However it is worth noting that a convenient procedure will often be available even when a GMM representation is not. Recall that $\Lambda = \Sigma_{\theta\gamma}\Sigma_{\gamma\gamma}^{-1}$. A plug-in estimator of $\hat{\Sigma}_{\gamma\gamma}$ of $\Sigma_{\gamma\gamma}$ will typically be available. What remains is to estimate $\Sigma_{\theta\gamma}$, which is a function of G, M, W_g, W_m and the asymptotic covariance Ω_{gm} of $\hat{g}_n(\theta_0)$ and $\hat{m}_n(\gamma_0)$. Estimators of all but Ω_{gm} are in hand if one has calculated $\hat{\Sigma}_{\gamma\gamma}$. In the case of CMD, it is common to use a bootstrap to estimate Ω_{gg}; in such cases the same bootstrap can often be used to estimate Ω_{gm}.

---

[5] Asymptotically, $\mathrm{Var}(\hat{\theta}) = \left(G'W_g G\right)^{-1} G'W_g \Omega_{gg} W_g G \left(G'W_g G\right)^{-1}$. This variance is typically estimated using a plug-in estimator with components $\hat{W}_g$ and $\hat{G}$.

[6] More precisely, a regression of $\tilde{g}_i$ on $\tilde{m}_i$ yields coefficients

$$\hat{\Lambda} = \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{m}_i'\tilde{m}_i \right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{m}_i'\tilde{g}_i \right)$$

where $\hat{\Lambda} \xrightarrow{p} \Lambda$.

5 Example: Linear Regression

Here we use several special cases of linear instrumental variables to build intuition. For simplicity, we assume throughout that all variables have mean zero.

Begin with the case of an OLS regression with two right-hand side variables ($k = 2$ and $z = x$). The two moments are then the sample covariance of y and x_1 and the sample covariance of y and x_2. Intuitively, the second moment should affect $\hat{\theta}_1$ only to the extent that x_2 is correlated with x_1. If x_2 and x_1 are independent, the estimate $\hat{\theta}_1$ should not be sensitive to the covariance between y and x_2. By contrast if small changes in x_2 predict large changes in x_1, then for given covariance of y and x_1, changes in the covariance of y and x_2 should make a large difference to the estimate $\hat{\theta}_1$.

Our measure of sensitivity matches these intuitions. If Λ is the sensitivity to moments, its first row is

$$(5) \qquad \left[\frac{1}{\mathrm{Var}\,(x_1) - \mathrm{Cov}\,(x_1, \beta x_2)}, -\frac{\beta}{\mathrm{Var}\,(x_1) - \mathrm{Cov}\,(x_1, \beta x_2)} \right],$$

where β is the coefficient from a univariate regression of x_1 on x_2. The denominators of the two terms are the same so, as expected, the relative sensitivity to the two covariances scales directly with β.

The intuition is similar in the case of two-stage least squares with two endogenous regressors and two instruments ($k = 2$, $z \neq x$, $\dim(x) = 2$). The first row of Λ is the same as in equation (5), with β replaced by the coefficient from a univariate IV regression of x_1 on x_2 instrumenting for x_2 with z_2. The sample covariance of y and z_2 will be relatively important for determining $\hat{\theta}_1$ to the extent that the IV estimates imply a strong *causal* effect of x_2 on x_1. If x_1 is strongly correlated with x_2, but orthogonal to z_2, $\hat{g}_2$ will play no role in determining $\hat{\theta}_1$.

Consider next the case of an overidentified model with one endogeneous regressor and multiple instruments ($k \geq 2$, $z \neq x$, $\dim(x) = 1$). Assume that the k instruments z are asymptotically independent with unit variance. Let $\hat{\theta}$ be the two-stage least squares (2SLS) estimate. Let $\hat{\gamma}$ be the vector whose i^{th} element is the ith IV estimate, i.e. the estimate of θ based on the ith instrument only. Let $\hat{\beta} = \left(\frac{1}{n} z'z \right)^{-1} \left(\frac{1}{n} z'x \right)$ be the first stage coefficients and let $\beta = \mathrm{plim}\left(\hat{\beta} \right)$. Then Λ is a row vector whose i^{th} element is $\beta_i^2 / \sum \beta_j^2$. That is, the sensitivity of the 2SLS estimate to a given IV estimate is proportional to the strength of the first stage for the corresponding instrument.

If some instruments are endogenous, then $\mathrm{E}\left(\hat{\theta} - \theta_0 \right) = \Lambda \mathrm{E}\left(\hat{\gamma} - \theta_0 \right)$, so Λ measures the extent to which the bias in a given element of $\hat{\gamma}$ translates into bias in the 2SLS estimator. If we instead define $\hat{\gamma} = \frac{1}{n} z'\varepsilon$, then Λ's i^{th} element is $\beta_i / \sum_j \beta_j^2$, and $\mathrm{E}\left(\hat{\theta} - \theta_0 \right) = \Lambda \mathrm{E}(\hat{\gamma})$. In this last case, Λ coincides with the matrix A that Conley, Hansen and Rossi (2012) define in order to perform inference for the 2SLS estimator in the presence of uncertainty about the exogeneity of the instrument.

Finally, consider an example in which the sample statistics of interest $\hat{\gamma}$ are not the moments used in estimation. Suppose $k = 1$ and $z = x$. Suppose that x is correlated with a variable v and a researcher is concerned that v may affect y, thus calling into question the exogeneity of x. The researcher wishes to know the sensitivity of the OLS estimator $\hat{\theta}$ to such endogeneity.

Let x_1 denote the orthogonal projection of x on v and let $x_2 = x - x_1$ denote its residual. Let $\hat{\gamma} = (\hat{\gamma}_1, \hat{\gamma}_2)$ be the vector of OLS coefficients from regressions of y on x_1 and x_2, respectively. Then $\Lambda =$

$\left(\frac{\text{plim}\left(\frac{1}{n}x_1'x_1\right)}{\text{plim}\left(\frac{1}{n}x'x\right)}, \frac{\text{plim}\left(\frac{1}{n}x_2'x_2\right)}{\text{plim}\left(\frac{1}{n}x'x\right)} \right)'$. That is, the estimate $\hat{\theta}$ is sensitive to the "bad" variation in x (that is correlated with v) to the extent that this constitutes a large share of the variance in x. If the researcher includes v as a control in the regression, then $\Lambda = (1,0)'$, i.e. the covariance between x_2 and y no longer influences $\hat{\theta}$.

6 Sensitivity and its Inverse

Define the matrix Γ so that $\mathrm{E}\left(\hat{\gamma} - \gamma_0 | \hat{\theta}\right) = \Gamma\left(\hat{\theta} - \theta_0\right)$. If $\hat{\gamma}$ is sufficient for $\hat{\theta}$, then $\Lambda\Gamma = I$. In this sense Γ is the inverse of Λ: rather than measuring the sensitivity of parameters to data, it measures the sensitivity of data to parameters.

The objects Γ and Λ are equally easy to estimate, and in some cases they communicate very similar intuitions. For example, if statistic $\hat{\gamma}_i$ is sufficient for parameter $\hat{\theta}_i$ for all i, then Λ and Γ are both diagonal. That is, parameter $\hat{\theta}_i$ is sensitive only to statistic $\hat{\gamma}_i$, and vice versa.

We focus on Λ because Γ may in general provide misleading intuitions about the empirical determinants of the parameters. One reason for this is the role of weights in estimation. If $\hat{\theta}$ is an MDE, then $\Gamma = G$ and (recall) $\Lambda = \left(G'W_g G\right)^{-1} G'W_g$. That is, Γ does not depend on weights W_g, so inspection of Γ may not correctly capture which moments "drive" the estimates.

A deeper issue is that Γ does not generally inherit the sufficiency property of Λ that we highlight in section 3. To see this consider the following abstract setting . Suppose that $\hat{\gamma}$ are moments and that $\hat{\gamma}$ and $\hat{\theta}$ are of length k. Suppose that $\hat{\theta}_1$ is a nontrivial function of all moments, that $\hat{\theta}_2$ is a function of all moments except the first, and so on, with $\hat{\theta}_k$ a function of only the last moment. Then Λ and Γ are upper triangular, with the rows of Λ and the columns of Γ corresponding to parameters. Taking Γ as a measure of sensitivity of parameters to moments, we conclude that only the first moment is relevant for $\hat{\theta}_1$ and that all moments are relevant for $\hat{\theta}_k$: exactly the opposite of the estimator's actual mechanics.

As a simple example, suppose that $\hat{\theta}$ is an IV with $k = 2$ and $z = x = \begin{bmatrix} 1 & x \end{bmatrix}$ where we abuse notation to define x as a vector and 1 as a vector of 1s. Let $\hat{\gamma} = \begin{bmatrix} \hat{\mathrm{E}}(y) & \hat{\mathrm{Cov}}(y,x) \end{bmatrix}'$ consist of the sample mean of y and the sample covariance of y and x. Then

$$\Lambda = \frac{1}{\mathrm{Var}(x)} \begin{bmatrix} \mathrm{Var}(x) & -\mathrm{E}(x) \\ 0 & 1 \end{bmatrix}.$$

That is, fixing the covariance of y and x, the coefficient on x is not sensitive to the mean of y. The covariance is sufficient for the coefficient. The constant term, by contrast, is sensitive to both moments, and neither is sufficient.

Consider the inverse:

$$\Gamma = \begin{bmatrix} 1 & \mathrm{E}(x) \\ 0 & \mathrm{Var}(x) \end{bmatrix}.$$

Observe that changing either parameter changes $\mathrm{E}(y)$ as long as $\mathrm{E}(x) \neq 0$. That is, while the mean of y does not affect the coefficient on x, the coefficient on x does affect the mean. This might suggest a misleading inference that the covariance is not sufficient for the coefficient.

An analogous argument implies that Γ does not inherit Λ's interpretation as a measure of the sensitivity of the estimator to model misspecification. That a given parameter affects a given empirical moment's asymptotic mean does not imply that misspecification of the relevant moment condition will result in an inconsistent estimate of the parameter.

7   Applications

7.1   Goettler and Gordon (2011)

We apply our measure to Goettler and Gordon's (2011) study of the effect of competition from AMD on Intel's level of innovation. We compute scaled influence for all parameters using the weight matrix $\hat{W}$ and the approximations $\hat{G}$ and $\hat{\Omega}_{gg}$ that Goettler and Gordon (2011) use to compute the asymptotic variance of their parameters. Although Goettler and Gordon's (2011) model is difficult to compute, scaled influence is computationally costless once the asymptotic variance has been computed.

Figures 1 and 2 present scaled influence for Goettler and Gordon's (2011) demand-side and supply-side parameters, respectively. We visually separate moments important for identifying demand-side and supply-side parameters based on Goettler and Gordon's (2011) discussion of model identification. Appendix table 1 reports influence in matrix form.

Our findings are broadly consistent with the intuition that Goettler and Gordon (2011) provide about the identification of their model. Below we discuss the identification of particular parameters in more detail in light of our calculations.

Consider first the demand-side parameters in figure 1, about which Goettler and Gordon (2011) write:[7]

> 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*. We interpret the *Intel fixed effect* as a hassle cost of upgrading one's computer and *Intel fixed effect - AMD fixed effect* as a brand effect.

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.

---

[7]In this and the subsequent quote, we substitute phrases in italics for the mathematical symbols used in the original paper.

Consider next the supply-side parameters in figure 2, about which 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.

We find that the innovation efficiencies are greatly affected by the observed innovation rates. AMD's investment efficiency is also identified by the observed quality difference between Intel and AMD CPUs: the more of a gap AMD allows, the greater is AMD's inferred disadvantage in innovating.

7.2 Gentzkow et al. (2013)

We next apply our measure to Gentzkow et al.'s (2013) model of newspaper demand with multiple readership. 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. (2013) 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. (2013) 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.

In contrast to our analysis of Goettler and Gordon (2011), here we do not study sensitivity to the moments used to estimate Gentzkow et al.'s (2013) model. As the model is estimated via maximum likelihood, the moments are first-order conditions with limited interpretability. Instead, we focus on the influence of a set of descriptive statistics, namely 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. Although the formal model is nontrivial to estimate, computing influence is instantaneous.

Figure 3 presents our results graphically. Appendix table 2 presents our findings in matrix form.

The first plot in figure 3 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. (2013). 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.

Both parameters have a high sufficiency Δ, indicating 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 the maximum likelihood estimator is in principle exploiting much more information than can be captured in a simple regression.

8 Extensions and 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, (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.

Another limitation to stress is that evaluating the relative sensitivity to different data features requires comparing the units of different elements of $\hat{\gamma}$. The scaled measure we propose, $\tilde{\Lambda}_{ij}$, may be natural in many contexts, and has the advantage of being invariant to the units in which $\hat{\gamma}$ and $\hat{\theta}$ are measured. In certain applications, however, alternative approaches to scaling may be more appropriate. For example, an element of $\hat{\gamma}$ may be known with near statistical certainty, so that a one-standard-deviation change in its value is of no economic significance. In such a case, a researcher may wish to scale Λ by some appropriate notion of economic significance rather than by statistical precision.

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.

Berry, Steven, James Levinsohn, and Ariel Pakes. 2004. Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy* 112(1): 68-105.

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

Cohen, Alma and Liran Einav. 2007. Estimating risk preferences from deductible choice. *American Economic Review* 97(3):745-788.

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.

Einav, Liran, Amy Finkelstein, and Mark R. Cullen. 2010. Estimating welfare in insurance markets using variation in prices. *Quarterly Journal of Economics* 75(3): 877-921.

Gentzkow, Matthew, Jesse Shapiro and Michael Sinkinson. 2013. Competition and ideological diversity: Historical evidence from US newspapers. NBER Working Paper No. 18234.

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

Hampel, Frank R., Elvezio M. Ronchetti, Peter J. Rosseeuw, 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. Forthcoming. The first-order approach to merger analysis. *American Economic Journal: Microeconomics.*

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

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

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: pp. 2111-2245. Amsterdam: North-Holland.
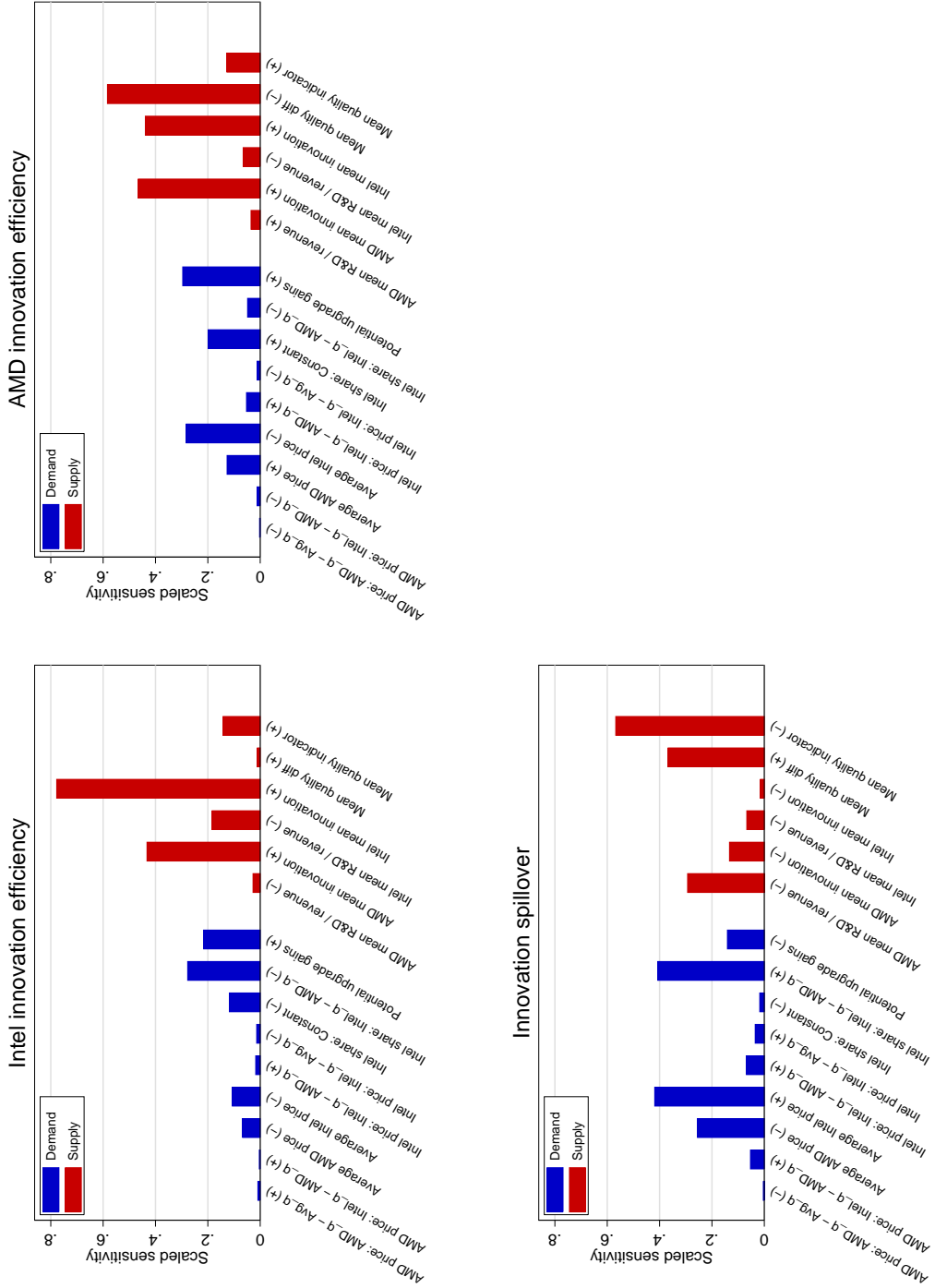
Pakes, Ariel. 2003. Common sense and simplicity in empirical industrial organization. *Review of Industrial Organization* 23(3): 193-215.

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, 2008.

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

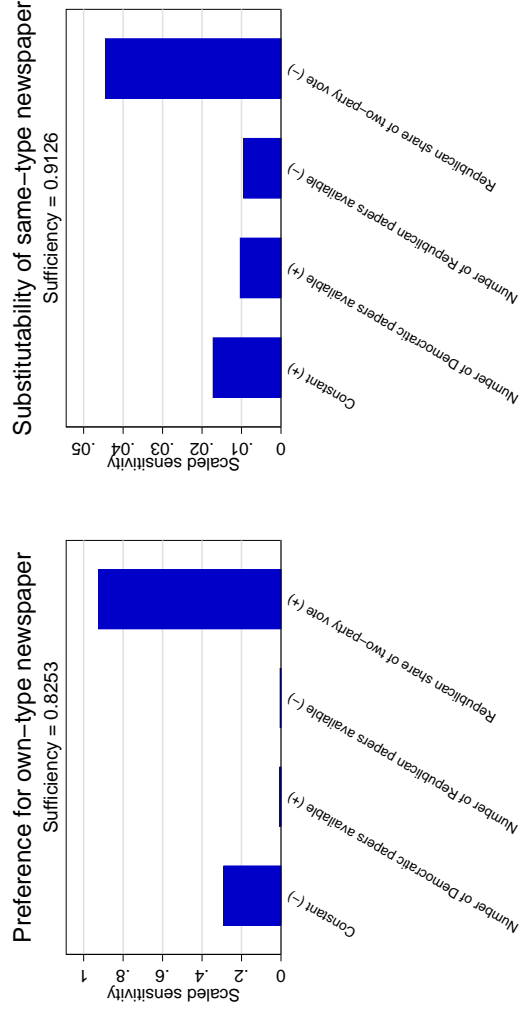Figure 1: Scaled sensitivity for demand parameters in Goettler and Gordon (2011)
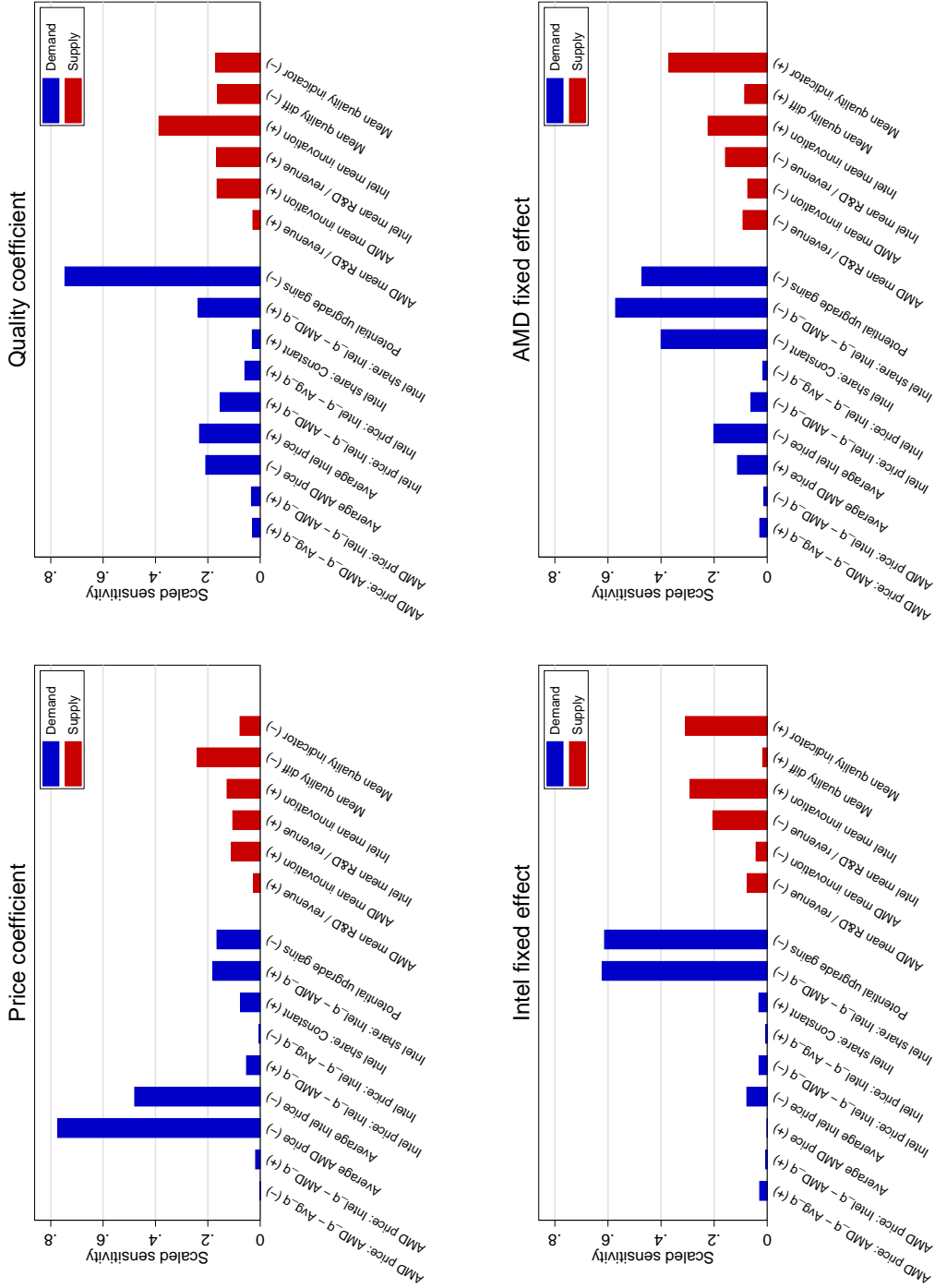
Notes: Each plot shows the absolute value of scaled sensitivity $|\tilde{\Lambda}_{ij}|$ for all moments j for a given parameter i, with the sign of $\tilde{\Lambda}_{ij}$ given in parentheses.

Figure 2: Scaled sensitivity for supply parameters in Goettler and Gordon (2011)

Figure 3: Scaled sensitivity for select parameters in Gentzkow, Shapiro, and Sinkinson (2013)

Notes: Each plot shows the absolute value of scaled sensitivity $|\tilde{\Lambda}_{ij}|$ for statistics j for a given parameter i, with the sign of $\tilde{\Lambda}_{ij}$ given in parentheses. The sufficiency for each parameter is shown above the plot.

17

Appendix Table 1: Scaled sensitivity for all parameters in Goettler and Gordon (2011)

	Price coefficient	Quality coefficient	Intel fixed effect	AMD fixed effect	Intel innovation efficiency	AMD innovation efficiency	Innovation spillover
AMD price: $AMD_q - Avg_q$	0.00	0.03	0.03	0.03	0.01	0.00	0.00
AMD price: $Intel_q - AMD_q$	0.02	0.03	0.01	0.01	0.00	0.01	0.05
Average AMD price	0.77	0.21	0.00	0.11	0.07	0.13	0.26
Average Intel price	0.48	0.23	0.08	0.20	0.11	0.28	0.42
Intel price: $Intel_q - AMD_q$	0.05	0.15	0.03	0.06	0.02	0.05	0.07
Intel price: $Intel_q - Avg_q$	0.00	0.06	0.01	0.02	0.01	0.01	0.03
Intel share: Constant	0.07	0.03	0.03	0.40	0.12	0.20	0.02
Intel share: $Intel_q - AMD_q$	0.18	0.24	0.62	0.57	0.28	0.05	0.41
Potential upgrade gains	0.17	0.75	0.61	0.47	0.22	0.30	0.14
AMD mean R&D / revenue	0.02	0.03	0.08	0.09	0.03	0.04	0.29
AMD mean innovation	0.11	0.16	0.04	0.07	0.43	0.47	0.13
Intel mean R&D / revenue	0.10	0.17	0.20	0.16	0.18	0.07	0.07
Intel mean innovation	0.13	0.39	0.29	0.22	0.78	0.44	0.01
Mean quality diff	0.24	0.16	0.02	0.09	0.01	0.58	0.37
Mean quality indicator	0.08	0.17	0.31	0.37	0.14	0.13	0.57

Note: The table shows a matrix with elements corresponding to the absolute value of scaled sensitivity $\left|\tilde{\Lambda}_{ij}\right|$ for parameters i (in columns) and moments j (in rows).

Appendix Table 2: Scaled sensitivity for Gentzkow, Shapiro and Sinkinson (2013)

Coefficient from a regression of relative circulation of Republican papers on:	Structural parameter	
	Preference for own-type newspaper	Substitutability of same-type newspapers
Constant	-0.2889	0.0172
Republican share of two-party vote	0.9247	-0.0445
Number of Republican papers available	-0.0041	-0.0094
Number of Democratic papers available	0.0077	0.0103
Sufficiency	0.8253	0.9126

Note: The table shows a matrix with elements corresponding to scaled sensitivity $\tilde{\Lambda}_{ij}$ for parameters i (in columns) and statistics j (in rows), along with sufficiency Δ.