IDENTIFICATION WITH IMPERFECT INSTRUMENTS

Aviv Nevo and Adam M. Rosen*

Abstract—Dealing with endogenous regressors is a central challenge of applied research. The standard solution is to use instrumental variables that are assumed to be uncorrelated with unobservables. We instead allow the instrumental variable to be correlated with the error term, but we assume the correlation between the instrumental variable and the error term has the same sign as the correlation between the endogenous regressor and the error term and that the instrumental variable is less correlated with the error term than is the endogenous regressor. Using these assumptions, we derive analytic bounds for the parameters. We demonstrate that the method can generate useful (set) estimates by using it to estimate demand for differentiated products.

I. Introduction

ENDOGENEITY is a central issue in nonexperimental studies. A common method for dealing with it is to use an instrumental variable (IV). In linear models, an IV has to be correlated with the endogenous covariate and uncorrelated with the econometric unobservable. The former condition is known as relevance, or strength of the IV, and the latter as exogeneity, or validity. Unfortunately, even when great care is taken, the validity of the IV often relies on what might seem like arbitrary assumptions, and empirical findings are often called into question as a result of debate over this assumption.

In this paper, we ask what, if anything, can be learned about the parameters of interest if the validity assumption fails. We provide (set) identification results for the parameters of a single regression equation with endogenous regressors and an IV that fails to satisfy the usual exogeneity condition. For a broad class of parametric models, we provide a set of inequalities that fully characterize the identified set, and for an often used special case—the linear model—we show that these inequalities can be characterized analytically using well-known estimators.

In place of the usual exogeneity assumption, we assume that some of the instruments are imperfect (they are imperfect instrumental variables, IIV) in the sense that they may be correlated with the error term. Without further assumptions or conditions on the data, these variables do not help in narrowing the possible values of the parameters. Indeed, they do not even identify the direction of bias of OLS estimates.

Received for publication April 30, 2009. Revision accepted for publication October 27, 2010.

*Nevo: Northwestern University and NBER; Rosen: UCL, IFS, and CEMMAP

We thank Joel Horowitz, Chuck Manski, Rob Porter, Elie Tamer, the editor, three anonymous referees, and seminar participants at Northwestern University, University College London, SITE, University of Paris I, University of Pennsylvania, University of Toronto, Yale, University of Naples Federico II, and the UK Competition Commission for comments and suggestions. This paper is a revised version of Chapter 3 of Adam Rosen's 2006 Northwestern Ph.D. dissertation, subsequently CEMMAP working paper no. CWP16/08. A.R. gratefully acknowledges financial support from the Center for the Study of Industrial Organization, the Eisner Memorial Fellowship at Northwestern University, and the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001. We are solely responsible for any and all errors.

If the value of the correlation between the IIV and the error term were known, then the IIV could be used to form a valid moment condition. It is not clear, however, how a researcher can learn and justify a particular value of this correlation. An alternative is to assume that the correlation is bounded between two values, for example, between -0.1 and 0.1, and use this information to bound the parameters of interest. While in principle this can help check the sensitivity of results to the validity assumption, it is not clear how to come up with reasonable values for bounding the correlation. For example, is the range of -0.1 to 0.1 reasonable? Is it too wide or not wide enough? It is not clear to us how to answer these questions in a systematic, non–ad hoc way.

Instead we consider an alternative. We first assume that the sign of the correlation between the IIV and the error term is (weakly) the same sign as the correlation between the endogenous variable and the error term. This assumption nests the standard IV assumption, since the sign of the correlation is weakly the same. Although this does not bound the value of the correlation between the IIV and the error term, we show that if the IIV and the endogenous variable are negatively correlated, the parameters can be bounded from both above and below.[1]

There is a sense in which the above requirements parallel the standard assumptions for IV. The standard assumptions require zero correlation between the IV and the error term and nonzero correlation between the IV and the endogenous variables, with no restrictions on the sign of this correlation. We relax the exogeneity assumption, instead requiring a weak inequality. As a result, however, we need to make stronger restrictions on the correlation of the observables if we are to obtain two-sided bounds. Hence we are substituting a weaker assumption on the unobservables for a stronger assumption on the observables. The advantage, of course, is that the assumption on the observables can be verified.

Next, we add an assumption that the correlation between the IIV and the error term is less than the correlation between the endogenous variable and the error term. That is, the IIV is thought to be correlated with the unobservable in the equation of interest, but less so than the endogenous regressor. This assumption seems intuitive: the IIV is not perfect, but it is better than the assumption justifying OLS (in the sense that the correlation with the error term is lower). This assumption allows us to bound the degree of correlation of the IIV and the error term between 0 and the correlation of the endogenous regressor and the error term. We show that this assumption allows us to tighten the bounds on the parameters of interest.

[1] Requiring that the correlation between the IIV and the error term have the same sign as the correlation between the endogenous variable and the error term can be viewed as a normalization. We could, of course, satisfy it by multiplying the IIV by -1 if it were not initially satisfied, but then the sign of the correlation between the IIV and the endogenous variable would flip, and the IIV would not provide two-sided bounds.

The Review of Economics and Statistics, August 2012, 94(3): 659–671
© 2012 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/REST_a_00171 by guest on 27 September 2021

We extend our results beyond the simple linear model to allow additional regressors and several IIVs. The case of multiple IIVs is of particular interest if none of these variables provides two-sided bounds on its own. We show how differencing the variables to create a new IIV can generate more informative two-sided bounds.

Our approach provides the range of values of the parameter of interest consistent with the data and the modeling assumptions. If the bounds are tight, the researcher may be able to draw conclusions using the weaker assumptions justifying the imperfect IV. If the bounds are too wide to allow definite conclusions, then one might want to use the stronger assumptions justifying standard IV methods. However, in this case, the justification of the instrument validity restriction is essential for the credibility of the results. Either way, our approach clarifies the relative role of data and assumptions. In section V, we demonstrate our method in an application to estimation of the demand for differentiated products and show that the approach can generate (what we believe to be) useful bounds.[2] While our motivation lies in IO applications such as these, we believe the method is more generally applicable in all fields of economics.

Like us, several papers provide bounds when the standard IV assumption fails (Frisch, 1934; Leamer, 1981; Klepper & Leamer, 1984; Hotz, Mullin, & Sanders, 1997; Altonji, Elder, & Taber, 2005; Bontemps, Magnac, & Maurin, 2006). Manski and Pepper (2000) characterize the identification region for model parameters when, instead of the usual exogeneity condition, the expectation of the outcome variable conditional on the instrument is assumed to be monotone for any given value of the endogenous covariate. Their analysis applies to nonparametric models, while we focus on a linear (parametric) model.[3] Because we focus on the linear model, we can employ weaker restrictions on correlations rather than assumptions on conditional means. The benefit of our assumptions is that we derive analytic bounds, which take on a very simple form and are easy to compute with standard regression software. (For a more complete comparison of our results to Manski & Pepper, 2000, see Nevo & Rosen, 2008.)

A different related literature provides sensitivity analysis in the context of estimation of treatment effects. (See, for example, Angrist, Imbens, & Rubin, 1996; Imbens, 2003; Small, 2007; Rosenbaum & Small, 2008. For a detailed overview and additional examples, see Rosenbaum, 2002.) For research on the implications of instrument invalidity in the linear model, outside the study of treatment effects, see, for example, Ashley (2009), Hahn and Hausman (2003), Berkowitz, Caner, and Fang (2008), Conley, Hansen, and Rossi (2006), and Kraay (2008).

In section II, we lay out the general setup and develop conditions that define the identified set. In section III, we first focus attention on the special case of a simple linear regression with one imperfect instrument and characterize analytically the identification region for the slope parameter β. We then extend the results to the multiple regression model as well as the case of several imperfect IVs and provide an analytic characterization for the bounds on all parameters. Section IV then discusses estimation and inference. Estimation is straightforward, and for inference, we illustrate how the method of Chernozhukov et al. (2009) can be used to compute confidence intervals for each of the model parameters individually. Section V provides the empirical illustration, and section VI concludes.

II. The Model

A. *The Setup*

The researcher is interested in identifying the parameters of a regression with at least one endogenous variable and an arbitrary number of exogenous variables. For each observation, the outcome variable Y is a function of an endogenous covariate X, a $1 \times k_w$ vector of additional covariates W, and U, an additively separable mean zero error unobserved by the researcher. We assume a parametric specification,

$$Y = m(X, W, \theta) + U,$$

where θ is a vector of parameters and $m(X, W, \theta)$ is twice continuously differentiable in θ.[4]

We assume that there exists an observable k_z-dimensional random vector Z of imperfect instruments—in the sense formally defined in our assumptions 3 and 4 below. These assumptions restrict the relationship between observables X and Z by imposing that X and components of Z have the same direction of correlation with the error and that the correlation between any component of Z and the error is less than that between X and the error. The key distinguishing feature of covariates W relative to X is that no such assumptions are made regarding Z and W. In addition, we assume the existence of a $1 \times k_w$ vector of valid instruments Z^w, which may include elements of W if some of the regressors are themselves exogenous. If the number of valid instruments exceeded the dimension of W, then model parameters would be locally point-identified under the usual rank condition, and they could be consistently estimated by standard instrumental variable methods such as GMM. For this reason, we restrict the number of valid instruments to be equal to the dimension of W.

We assume the econometrician observes a size n sample of realizations of (Y, X, W, Z, Z^w) satisfying the following assumptions. We use the subscript i to denote individual observations, and subscript j to denote individual elements of these vectors:

[2] An additional application to production function estimation is available in Nevo & Rosen (2008).

[3] See Manski and Pepper (1998) for an application to a linear model.

[4] For simplicity we present the results in this section for a scalar X. The basic ideas can be extended to a multivariate variable, as we show in the next section for the linear model.

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/REST_a_00171 by guest on 27 September 2021

Assumption 1 (*sampling process*). *The observations* $(y_i, x_i, w_i, z_i, z_i^w)'$, $i = 1, \ldots, n$ *are stationary and weakly dependent.*

Assumption 1 is a standard assumption, commonly used to guarantee consistency of parameter estimates in conventional cases with sufficiently many instruments for point identification. Here, it similarly serves to guarantee consistency of our estimated bounds. Note that this assumption allows heteroskedasticity and serial correlation. In assumption 2, we further assume that variables Z^w are exogenous, in the sense that each component of Z^w is uncorrelated with the unobservable U:

Assumption 2 (Z^w *exogenous*). $\mathbb{E}(Z^{w\prime}U) = \mathbf{0}$.

The variable X may be correlated with U, in which case identification and estimation of the model parameters θ typically rely on the use of a vector of instrumental variables. The assumptions justifying the IVs are often called into question in empirical work, and in some cases they may be untestable. We assume instead that the econometrician has observations of some imperfect instruments, Z, that are also correlated with the unobservable U, but less so than the endogenous regressor X, which we formalize below in assumptions 3 and 4. We assume that k_z, the number of imperfect instruments, is fixed.

Formally, the assumptions we impose on our imperfect instruments (IIV) are as follows. We use the notations σ_{ab} and ρ_{ab} throughout to denote the covariance and correlation, respectively, between any two random variables A, B, and σ_a to denote the standard deviation of a random variable A.

Assumption 3 (*same direction of correlation*).

$$\rho_{xu}\rho_{z_ju} \geq 0, j = 1, \ldots, k_z.$$

Assumption 3 asserts that the endogenous regressor X and the IIV have the same direction of correlation with the error term. Since we require a weak inequality, assumption 3 is a weakening of the usual assumption that $\rho_{z_ju} = 0$. Like the classic validity assumption, this assumption is also about unobserved variables and will not always be satisfied. We provide an application in section V where we believe this assumption is satisfied (see Nevo & Rosen, 2008, for an additional application).

Assumption 4 (*instruments less endogenous than x*).

$$|\rho_{xu}| \geq |\rho_{z_ju}|, j = 1, \ldots, k_z.$$

Assumption 4 adds the condition that the IIV is less correlated with the error term than is the endogenous regressor. To us, this seems like an intuitive assumption for numerous applications where one may believe that the instrument Z is not necessarily exogenous but is "better" or "less endogenous" than the endogenous regressor. For example, in the context of demand estimation in our application in section V, the market demand equation contains an unobservable market-level component. We assume that the unobservable component of demand in any given market is less correlated with the product's price in other markets than the own-market price. As we will see below, this assumption will help us tighten the bounds on the parameters.

We also make use of the standard rank conditions needed for the probability limits of the OLS and 2SLS estimators to be well-defined in the linear model.

Assumption 5 (*rank and order*). $Rank(\mathbb{E}[(Z, Z^w)'(Z, Z^w)]) = k_z + k_w$, $rank(\mathbb{E}[(X, Z^w)'(X, Z^w)]) = k_w + 1$, *and* $rank(\mathbb{E}[(Z, Z^w)'(X, W)]) = k_w + 1$, $k_z \geq 1$.

B. *Identification*

The above assumptions yield a system of moment equalities and inequalities that restrict the feasible values of model parameters θ. In general, these restrictions will not be sufficient for point identification but may still provide information regarding the value of θ.

A useful construct for the analysis is the ratio of correlations of the instrument Z_j and regressor X with the econometric error term. Denote this ratio as

$$\lambda_j^* \equiv \rho_{z_ju}/\rho_{xu}, \tag{1}$$

and when $\rho_{xu} = \rho_{zu} = 0$, we adopt the convention that $\lambda_j^* = 0$. By definition,

$$\mathbb{E}\left[\left(\sigma_x Z_j - \lambda_j^* \sigma_{z_j} X\right) U\right] = \sigma_x \sigma_{z_ju} - \lambda_j^* \sigma_{z_j} \sigma_{xu} = 0,$$

when assumptions 3 and 4 hold. Therefore, if λ_j^* were known, it could be used to construct a weighted average of Z_j and X that is uncorrelated with the error term.

Of course, λ_j^* is unknown, but our assumptions bound its value. That is, assumptions 3 and 4 together imply that $\lambda_j^* \in [0, 1]$. Our hope is to use the bounds on λ_j^* to bound the parameters of interest, θ. To do this, define the function $V_j(\cdot)$ as

$$V_j(\lambda) \equiv \sigma_x Z_j - \lambda \sigma_{z_j} X. \tag{2}$$

Note that when evaluated at λ_j^* the function generates a variable, $V_j(\lambda_j^*)$, that is uncorrelated with U and satisfies the moment condition $\mathbb{E}[(Y - m(X, W, \theta)) \times V_j(\lambda_j^*)] = 0$. In other words, the function evaluated at λ_j^* yields a valid instrumental variable. We do not know the value λ_j^*, but given our assumptions, we know that $\lambda_j^* \in [0, 1]$. Hence, the identified set is computed by evaluating the moment condition for the full range of $0 \leq \lambda_j \leq 1$. As we show in the next section, a much simpler characterization holds in the linear model.

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/REST_a_00171 by guest on 27 September 2021

Formally, the implied restrictions of assumptions 2 to 4 are:

$$\mathbb{E}\left[Z_j^{w\prime}(Y - m(X, W, \theta))\right] = \mathbf{0}, \tag{3a}$$

$$\mathbb{E}\left[(Y - m(X, W, \theta)) \times V_j\left(\lambda_j^*\right)\right] = 0, j = 1, \ldots, k_z, \tag{3b}$$

$$\lambda_j^* \in [0, 1], j = 1, \ldots, k_z, \tag{3c}$$

The model parameters must satisfy the $k_w + k_z$ moment conditions (3a) and (3b). If $\Lambda^* \equiv (\lambda_1^*, \cdots, \lambda_{k_z}^*)$ were known and the standard rank conditions were satisfied, these moment conditions would identify θ locally (globally, if the model is linear). However, since Λ^* is known only to belong to the unit cube in $\mathbb{R}^{k_z}$, θ will generally not be point-identified.

The identified set for θ is by definition the set of parameter values that satisfy restrictions (3a) to (3c). The identified set can be computed numerically by iterating over all the possible values of Λ^* and computing the implied value of the parameters. Estimation and inference can be performed using this characterization and by any of a variety of methods (see, for example, Chernozhukov et al., 2007, and the references therein).

III. The Linear Model

Conditions (3a) to (3c) fully characterize the identified set, but they do not provide an easy way to compute the set, nor do they tell us when the set is bounded. In this section, in order to obtain an analytic characterization of the identified set, we restrict attention to the linear model. We show that in the linear model, our modeling restrictions lead to a straightforward characterization of the identified set that can be exploited to perform (set) estimation with standard linear regression methods. We start with the simple linear model and then generalize the results to multiple IIVs as well as additional regressors.

A. *The Simple Linear Model with One Imperfect Instrument*

Consider the simple linear model,

$$Y = \alpha + X\beta + U,$$

where $\mathbb{E}[U] = 0$. Define β^{OLS} and β_z^{IV} to be the probability limits of the standard OLS and IV estimators for β, respectively:

$$\beta^{OLS} \equiv \frac{\sigma_{xy}}{\sigma_x^2} = \beta + \frac{\sigma_{xu}}{\sigma_x^2} \tag{4}$$

and

$$\beta_z^{IV} \equiv \frac{\sigma_{zy}}{\sigma_{xz}} = \beta + \frac{\sigma_{zu}}{\sigma_{xz}}. \tag{5}$$

$\widehat{\beta}^{OLS}$ and $\hat{\beta}_z^{IV}$ are taken to be their corresponding estimators. Applying equation (5) to the case where $V(1)$ is the instrument defines

$$\beta_{v(1)}^{IV} \equiv \frac{\sigma_x\sigma_{zy} - \sigma_z\sigma_{xy}}{\sigma_x(\sigma_{xz} - \sigma_z\sigma_x)}. \tag{6}$$

This is the probability limit of the traditional IV (2SLS) estimator for β when $V(1)$ is used as an instrument for X, where $V(\cdot)$ is as defined in equation (2), that is, $V(1) \equiv \sigma_x Z - \sigma_z X$.

The true value of the parameter, β, can be bounded when assumptions 1 to 3 are imposed.

Lemma 1. *Let assumptions 1 to 3 hold. If $\sigma_{xz} < 0$, then β is between β^{OLS} and β_z^{IV} and we have a two-sided bound ($\beta_z^{IV} \le \beta \le \beta^{OLS}$ if σ_{xu}, $\sigma_{zu} \ge 0$, and $\beta^{OLS} \le \beta \le \beta_z^{IV}$ if σ_{xu}, $\sigma_{zu} \le 0$). If instead, $\sigma_{xz} > 0$, then we have a one-sided bound given by $\beta \le \min\{\beta^{OLS}, \beta_z^{IV}\}$ if σ_{xu}, $\sigma_{zu} \ge 0$ and $\beta \ge \max\{\beta^{OLS}, \beta_z^{IV}\}$ if σ_{xu}, $\sigma_{zu} \le 0$.*

Lemma 1 gives a simple characterization of the identified set for β, which we denote as B^*. Given assumptions 1 to 3, β^{OLS} and β_z^{IV} provide either two-sided or one-sided bounds, depending on the correlation between X and Z, which can be consistently estimated from the data. Note that when $\sigma_{xz} < 0$, an assumption on the sign of σ_{xu} is not needed for two-sided bounds. When instead $\sigma_{xz} > 0$, there is only a one-sided bound, and a sign assumption for σ_{xu} and σ_{zu} is needed to determine whether this is an upper or a lower bound. When $\sigma_{xz} = 0$, β_z^{IV} is either positive or negative infinity, depending on the sign of σ_{zu}, so that the corresponding bound is uninformative. More generally, if σ_{xz} is "small," then the bound given from β_z^{IV} is large, which is a direct manifestation of the problem of weak instruments.

If $\sigma_{xz} < 0$, assumption 3 can provide finite, and, we hope, economically helpful, bounds on the parameter of interest. Moreover, inspection of these bounds reveals that when $\sigma_{xz} < 0$ or $\rho_{xz} > \lambda^*$, the bias of β^{OLS} has the same sign as $\beta^{OLS} - \beta_z^{IV}$. In general, however, an IIV will not even identify the direction of the bias or necessarily correct the OLS estimate in the "right" direction. A common intuition—that even if the IV is invalid, the IV estimate corrects the OLS estimate in the right direction—probably comes from the case when λ^* is small. Recall that $\lambda^* = 0$ is the valid IV case. So λ^* close to 0 means the IV is "almost" valid and then $\rho_{xz} > \lambda^*$ is likely to hold. But if neither $\sigma_{xz} < 0$ nor $\rho_{xz} > \lambda^*$ holds, then an IIV will not even identify the direction of the bias of β^{OLS}.

Up to this point we have not imposed assumption 4. We now ask what we get from imposing it. As the following result shows, by bounding λ^* between 0 and 1, as the assumption implies, we obtain sharper identification results:

Proposition 1. *Let assumptions 1 to 4 hold. If $\sigma_{xz} < 0$, then β is between $\beta_{v(1)}^{IV}$ and β_z^{IV}, and we have a two-sided bound ($\beta_z^{IV} \le \beta \le \beta_{v(1)}^{IV}$ if σ_{xu}, $\sigma_{zu} \ge 0$ and $\beta_{v(1)}^{IV} \le \beta \le \beta_z^{IV}$ if σ_{xu}, $\sigma_{zu} \le 0$). If instead $\sigma_{xz} > 0$, then we have a one-sided bound given by $\beta \le \min\{\beta_{v(1)}^{IV}, \beta_z^{IV}\}$ if $\sigma_{xu}, \sigma_{zu} \ge 0$ or $\beta \ge \max\{\beta_{v(1)}^{IV}, \beta_z^{IV}\}$ if σ_{xu}, $\sigma_{zu} \le 0$. These bounds are sharp.*

In cases where $\sigma_{xz} > 0$, proposition 1 shows that imposing assumption 4 improves on the bounds of lemma 1 but that the bounds remain one-sided. When instead $\sigma_{xz} < 0$, proposition 1 improves on the two-sided bounds given in lemma 1. So

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/REST_a_00171 by guest on 27 September 2021

while assumptions 3 and 4 jointly bound λ^* between 0 and 1, they do not in general provide two-sided bounds for the parameter of interest, β. To see why, note that

$$\beta^{IV}_{v(\lambda)} = \frac{\sigma_x\sigma_{zy} - \lambda\sigma_z\sigma_{xy}}{\sigma_x\sigma_{xz} - \lambda\sigma_z\sigma_x^2} = \frac{\sigma_y(\rho_{zy} - \lambda\rho_{xy})}{\sigma_x(\rho_{xz} - \lambda)}. \tag{7}$$

If $\lambda = \rho_{xz}$, then $\beta^{IV}_{v(\lambda)}$ is not defined, so even though λ^* is bounded, β might not be bounded. If $\sigma_{xz} > 0$, we cannot rule out that $\lambda = \rho_{xz}$. If $\sigma_{xz} < 0$, then for all $0 \leq \lambda \leq 1$, $\lambda \neq \rho_{xz}$, and therefore β is bounded from above and below.

Thus, while assumption 4 does not give us two-sided bounds where they did not already exist, it does tighten the bounds. Corollary 1 applies to the case where $\sigma_{xz} < 0$ and where proposition 1 achieves two-sided bounds for β. This corollary characterizes the degree to which use of $\beta^{IV}_{v(1)}$ improves on the bounds provided by lemma 1, which are just β^{OLS} and β^{IV}_z. The implication is that the greater the magnitude of the correlation between X and Z, the tighter the bound achieved using $\beta^{IV}_{v(1)}$ instead of β^{OLS}, with maximal improvement obtained when $\rho_{xz} = -1$, in which case the size of the bounds is halved.

Corollary 1. *If $\sigma_{xz} < 0$, then*

$$\beta^{IV}_{v(1)} - \beta^{IV}_z = \frac{1}{1 - \rho_{xz}}\left(\beta^{OLS} - \beta^{IV}_z\right).$$

Notice that if $\mathbb{E}[ZU] \neq 0$, it is possible for β^{OLS} to be closer to β than β^{IV}_z, even in the case when the instrument Z is less correlated with U than is the endogenous regressor X and $\sigma_{xz} < 0$ (the IV estimate identifies the sign of the OLS bias). Basically, it is possible that β^{IV}_z overcorrects, so that it may not be the case that the instrumental variables estimator offers an improvement over β^{OLS}. This result accords with the finding of Hahn and Hausman (2003).

B. Additional Regressors

This section generalizes beyond the simple linear model, allowing the presence of additional regressors, as well as valid instruments, in addition to the IIV. The regression of interest is

$$Y = X\beta + W\delta + U, \tag{8}$$

where X is univariate, W is a $1 \times k_w$ vector of additional regressors (possibly including a constant), and $\mathbb{E}[U] = 0$. Z is a univariate IIV that satisfies assumptions 3 and 4 with respect to the endogenous regressor X. The additional regressors W may include exogenous and endogenous components, but we assume that assumption 2 holds, so that there is a $1 \times k_w$ vector of valid instruments Z^w such that $\mathbb{E}[Z^{w\prime}W]$ is nonsingular and $\mathbb{E}[Z^{w\prime}U] = 0$. If any components of W are exogenous, then those will also be included as components of Z^w. Note that if the dimension of Z^w exceeded that of W, there would be more valid instruments than regressors. Then, as long as $\mathbb{E}[(X, W)'Z^w]$ had full rank, $\theta = (\beta, \delta')'$ would be point-identified and consistently estimable by the usual IV estimation procedures. Thus, we restrict attention to the case where Z^w has the same dimension as W.

In principle, the analysis is the same as before. As shown in section IIB, assumptions 3 and 4 are equivalent to the assertion that $V(\lambda) = \sigma_x Z - \lambda\sigma_z X$ is a valid instrument for X, for some unknown value of $\lambda \in [0, 1]$, which we denote λ^*. Thus, the identified set can be computed by computing $\theta(\lambda)$ over all $\lambda \in [0, 1]$ with $\theta(\lambda)$ given by the usual IV formula:

$$\theta(\lambda) = \mathbb{E}[(V(\lambda), Z^w)'(X, W)]^{-1}\mathbb{E}[(V(\lambda), Z^w)'Y]. \tag{9}$$

The remainder of this section first generalizes our results from the simple linear model to derive analytic bounds on β. We then construct the identification interval for each individual component of δ.

Bounds on β. We follow a standard way of deriving results for multivariate regressions by netting out the effects of W. The residuals of an IV regression of X and Y on W using Z^w as the instruments are given by

$$\tilde{X} \equiv X - W\mathbb{E}[Z^{w\prime}W]^{-1}\mathbb{E}[Z^{w\prime}X], \tag{10a}$$

$$\tilde{Y} \equiv Y - W\mathbb{E}[Z^{w\prime}W]^{-1}\mathbb{E}[Z^{w\prime}Y]. \tag{10b}$$

We show in the appendix that

$$\tilde{Y} = \beta\tilde{X} + U. \tag{11}$$

The result of running an IV regression of $\tilde{Y}$ on $\tilde{X}$ with $V(\lambda)$ as instrument consistently estimates the first component of θ in the IV regression above given by $\beta^{IV}_{V(\lambda)} \equiv \mathbb{E}[V(\lambda)\tilde{X}]^{-1}\mathbb{E}[V(\lambda)\tilde{Y}]$. Applying our analysis of the simple linear model to this regression then delivers the following bounds for β, where $\sigma_{z\tilde{y}}$ denotes the covariance of Z and $\tilde{Y}$, $\sigma_{z\tilde{x}}$ that between Z and $\tilde{X}$, and so on.

Proposition 2. *Let assumptions 1 to 5 hold. If $(\sigma_{\tilde{x}x}\sigma_z - \sigma_x\sigma_{\tilde{x}z})\sigma_{\tilde{x}z} < 0$, then β is between $\beta^{IV}_{v(1)}$ and β^{IV}_z and we have a two-sided bound ($\beta^{IV}_z \leq \beta \leq \beta^{IV}_{v(1)}$ if $\sigma_{xu}\sigma_{\tilde{x}z} < 0$ and $\beta^{IV}_{v(1)} \leq \beta \leq \beta^{IV}_z$ if $\sigma_{xu}\sigma_{\tilde{x}z} \geq 0$). If instead $(\sigma_{\tilde{x}x}\sigma_z - \sigma_x\sigma_{\tilde{x}z})\sigma_{\tilde{x}z} > 0$, then we have a one-sided bound given by $\beta \geq \max\{\beta^{IV}_z, \beta^{IV}_{v(1)}\}$ if $\sigma_{xu}\sigma_{\tilde{x}z} < 0$ or $\beta \leq \min\{\beta^{IV}_z, \beta^{IV}_{v(1)}\}$ if $\sigma_{xu}\sigma_{\tilde{x}z} \geq 0$. These bounds are sharp.*

The statement of proposition 2 follows closely that of proposition 1, but the conditions differ because we need to account for the partialing-out of exogenous variables in the first stage. Proposition 2 relies on the original statement of assumptions 3 and 4, which is in terms of sign restrictions on unconditional correlations. When $Z^w = W$, a sufficient condition for the required inequality is $\sigma_{\tilde{x}z} < 0$. An alternative approach would be to impose assumptions 3 and 4 on conditional correlations, with $\rho_{\tilde{x}u}$ and $\rho_{\tilde{z}u}$ replacing ρ_{xu} and

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/REST_a_00171 by guest on 27 September 2021

ρ_{zu}. This would lead to the condition $\sigma_{\tilde{x}z} < 0$ for two-sided bounds, but the bounds would differ from β_z^{IV} and $\beta_{v(1)}^{IV}$.

Bounds on other coefficients. In this section we shift focus from β to other regression coefficients and show that these coefficients are also interval identified. Whether their identification regions are one- or two-sided corresponds to whether the identification region for β is one- or two-sided.

To facilitate the analysis, define $Y^* \equiv Y - X\beta$, so that $Y^* = W\delta + U$. Given the assumption that the instruments Z^w are valid, it follows that $\delta = \mathbb{E}[Z^{w\prime}W]^{-1}\mathbb{E}[Z^{w\prime}Y^*]$. Then the jth component of δ is given by

$$\delta_j = \mathbb{E}\big[Z_j^{w\prime}\tilde{W}_j\big]^{-1}\mathbb{E}\big[Z_j^{w\prime}\tilde{Y}^*\big], \tag{12}$$

where

$$\tilde{W}_j \equiv W_j - W_{-j}\mathbb{E}\big[Z_{-j}^{w\prime}W_{-j}\big]^{-1}\mathbb{E}\big[Z_{-j}^{w\prime}X\big],$$
$$\tilde{Y}^* \equiv Y^* - W_{-j}\mathbb{E}\big[Z_{-j}^{w\prime}W_{-j}\big]^{-1}\mathbb{E}\big[Z_{-j}^{w\prime}Y^*\big],$$

and subscript $-j$ denotes a vector with its jth component removed. That is, $\tilde{W}_j$ and $\tilde{Y}^*$ correspond to the residuals of IV regressions of W_j and Y^* on W_{-j} employing Z_{-j}^w as instruments. The latter regression is infeasible as β, and therefore Y^* are unknown. However, β is interval-identified, and the identification region for δ_j can be obtained by tracing out the implied values of δ_j over the identified set for β.

Proposition 3. *Assume assumptions 1 to 5. Then the identification region for any $\delta_j, j = 1, \ldots, k_w$ is given by the interval ranging from δ_{j0} to δ_{j1} where*

$$\delta_{j0} \equiv \mathbb{E}\big[Z_j^{w\prime}\tilde{W}_j\big]^{-1}\mathbb{E}\big[Z_j^{w\prime}\tilde{Y}\big] - \mathbb{E}\big[Z_j^{w\prime}\tilde{W}_j\big]^{-1}\mathbb{E}\big[Z_j^{w\prime}\tilde{X}\big]\beta_L,$$
$$\delta_{j1} \equiv \mathbb{E}\big[Z_j^{w\prime}\tilde{W}_j\big]^{-1}\mathbb{E}\big[Z_j^{w\prime}\tilde{Y}\big] - \mathbb{E}\big[Z_j^{w\prime}\tilde{W}_j\big]^{-1}\mathbb{E}\big[Z_j^{w\prime}\tilde{X}\big]\beta_U,$$

and where β_L and β_U are the extreme points of the identification region for β.

C. Multiple Imperfect Instruments

Up to this point we have assumed that we have a single imperfect IV. We now ask what the researcher gains from multiple imperfect IVs for the endogenous regressor X. We show that this can help tighten the identification region for β and with additional assumptions can be used to obtain two-sided bounds where they were previously only one-sided.

When there are multiple imperfect instruments that satisfy assumptions 3 and 4, that is, $k_z > 1$ and $Z = (Z_1, \ldots, Z_{k_z})$, then proposition 2 can be used to derive bounds on β (one- or two-sided, depending on the sign of σ_{xz_r}) for each Z_r, $r = 1, \ldots, k_z$. For each r, denote the bounds implied by proposition 2 as $B_r^* = [\beta_{l,r}, \beta_{u,r}]$ (where one of the two is possibly $\pm\infty$). In addition, let $D_{j,r}^* = [\delta_{l,r}^j, \delta_{u,r}^j]$ denote the bounds on δ_j for each $j = 1, \ldots, k_w$ given by proposition 3. It follows that the identified set for β is the intersection of all of the intervals B_r^*, and the identified set for each δ_j is the intersection over $r = 1, \ldots, k_z$ of the intervals $D_{j,r}^*$.

Proposition 4. *Assume assumptions 1 to 5. Then the identification region for β is $B^* = [\max_r \beta_{l,r}, \min_r \beta_{u,r}]$, and the identification region for δ_j, each $j = 1, \ldots, k_w$ is $D_j^* = [\max_r \delta_{l,r}^j, \min_r \delta_{u,r}^j]$. These bounds are sharp.*

This proposition is a result of applying propositions 2 and 3 to each of the instruments $Z_j, j = 1, \ldots, J$. Furthermore, this exploits all the identifying power of the multiple IIVs, in the sense that there is no additional identifying power from imposing assumptions 3 and 4 jointly with respect to multiple instruments. This is because for every value of $\beta \in B^*$, there is an admissible data generation process that satisfies all of our modeling assumptions. The sharpness of the bounds for each element of δ is a direct result of their characterization in proposition 3 as functions of β. Note that in principle, B^* and the intervals D_j^* could be empty, which can be used to serve as the basis for a specification test, as we discuss in the following section.

A potential drawback is that when $(\sigma_{x\tilde{x}}\sigma_{z_j} - \sigma_{\tilde{x}z_j}\sigma_x)\sigma_{\tilde{x}z_j} \geq 0$ for each IIV Z_j, proposition 4 may provide only a one-sided bound. In some cases, the researcher may be willing to assert that one instrument is better than another in the sense that it is both more relevant and more valid. In particular, in the case where the bounds of proposition 4 are one-sided, such an assumption can be used to generate an IIV that provides two-sided bounds by constructing a weighted difference of the original IIVs. The intuition is that by subtracting the more relevant IIV from the less relevant one, one can generate a new IIV that is negatively correlated with X. This can be used to obtain two-sided bounds by the application of proposition 2 as long as the differencing does not flip the direction of the new IIVs' correlation with the error term, relative to the original ones.

Formally, let there be two instruments, Z_1 and Z_2, each satisfying assumption 3, and $(\sigma_{x\tilde{x}}\sigma_{z_j} - \sigma_{\tilde{x}z_j}\sigma_x)\sigma_{\tilde{x}z_j} \geq 0$, for $j = 1, 2$. Define the following weighted average of Z_1 and Z_2:

$$\omega(\gamma) = \gamma Z_2 - (1 - \gamma)Z_1,$$

where $\gamma \in (0, 1)$.

Proposition 5. *Assume assumptions 1, 2, and 5, and let assumption 3 hold for both Z_1 and Z_2. Suppose that $\sigma_{z_j u} \geq 0$ and $(\sigma_{x\tilde{x}}\sigma_{z_j} - \sigma_{\tilde{x}z_j}\sigma_x)\sigma_{\tilde{x}z_j} \geq 0$ for $j = 1, 2$. Assume that for some known $\gamma^* \in (0, 1)$, $\sigma_{\omega(\gamma^*)u} \geq 0$ and $(\sigma_{x\tilde{x}}\sigma_{\omega(\gamma^*)} - \sigma_{\omega(\gamma^*)\tilde{x}}\sigma_x)\sigma_{\omega(\gamma^*)\tilde{x}} < 0$. This yields the following bounds for β:*

$$\beta_{\omega(\gamma^*)}^{IV} \leq \beta \leq \min\big\{\beta_{z_1}^{IV}, \beta_{z_2}^{IV}, \beta^{OLS}\big\}.$$

If, in addition, assumption 4 holds for both Z_1 and Z_2, then

$$\beta_{\omega(\gamma^*)}^{IV} \leq \beta \leq \min\big\{\beta_{z_1}^{IV}, \beta_{z_2}^{IV}, \beta_{v_1(1)}^{IV}, \beta_{v_2(1)}^{IV}, \beta_{v^*(1)}^{IV}\big\},$$

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/REST_a_00171 by guest on 27 September 2021

where $\beta^{IV}_{v_1(1)}$, $\beta^{IV}_{v_2(1)}$, $\beta^{IV}_{\omega(\gamma^*)}$, *and* $\beta^{IV}_{v^*(1)}$ *are defined by using* $V_1(1)$, $V_2(1)$, $\omega(\gamma^*)$, *and* $V^*(1) \equiv \sigma_x\omega(\gamma^*) - \sigma_{\omega(\gamma^*)}X$ *as instruments in the definition given in equation (5).*

The first set of bounds given by the proposition simply says that if there exists a γ^* such that the conditions on $\omega(\gamma^*)$ are met, then we can apply lemma 1 to obtain a two-sided bound. If assumption 4 holds as well, then the bounds of proposition 2 also apply with Z_1, Z_2, and $\omega(\gamma^*)$ as IIVs, yielding sharper bounds. Note that the quantities $\beta^{IV}_{v_j(1)}, j = 1, 2$, correspond to the probability limits of IV estimators that employ $V_j(1) \equiv \sigma_x Z_j - \sigma_{z_j}X$ as an instrument, and $\beta^{IV}_{\omega(\gamma^*)}$ and $\beta^{IV}_{v^*(1)}$ to those employing $\omega(\gamma^*)$ and $V^*(1)$ as instruments, and are thus easily estimated via linear IV regression. The following lemma provides more basic conditions that guarantee the existence of such a γ^* and that are testable in the case where the additional regressors W are exogenous.

Lemma 2. *Assume assumptions 1, 2, and 5, and let assumption 3 hold for both* Z_1 *and* Z_2. *Suppose that* $\sigma_{z_j u} > 0$ *and* $(\sigma_{x\tilde{x}}\sigma_{z_j} - \sigma_{\tilde{x}z_j}\sigma_x)\sigma_{\tilde{x}z_j} > 0$ *for* $j = 1, 2$, *and that* $W = Z^w$. *Then the following statements are equivalent: (1) there exists* $\gamma^* \in (0, 1)$ *such that* $\sigma_{\omega(\gamma^*)u} \geq 0$ *and* $\sigma_{\omega(\gamma^*)\tilde{x}} < 0$; *(2) there exists* $\gamma^* \in (0, 1)$, *such that* $\frac{\sigma_{\tilde{x}z_1}}{\sigma_{\tilde{x}z_2}} > \frac{\gamma^*}{1-\gamma^*} > \frac{\sigma_{z_1 u}}{\sigma_{z_2 u}}$; *and (3)* $\sigma_{z_1\tilde{y}}\sigma_{\tilde{x}z_2} < \sigma_{z_2\tilde{y}}\sigma_{\tilde{x}z_1}$. *Furthermore, a sufficient condition for the inequality* $(\sigma_{x\tilde{x}}\sigma_{\omega(\gamma^*)} - \sigma_{\omega(\gamma^*)\tilde{x}}\sigma_x)\sigma_{\omega(\gamma^*)\tilde{x}} < 0$ *is that the partial correlation between* X *and* $\omega(\gamma^*)$ *controlling for* W *is negative.*

The lemma shows several things. Part 1 of the lemma restates the conditions required, on the weighted average $\omega(\gamma^*)$, to satisfy both assumption 3 and the conditions in proposition 1. Part 2 states a condition on the correlations of the two IIVs, Z_1 and Z_2. It says that the ratio of the IIVs' correlations with $\tilde{X}$ must exceed the ratio of their correlations with the error term. Finally, part 3 of the lemma states a condition on observable quantities that allows us to test whether such a γ^* exists.

Note that while proposition 5 provides two-sided bounds, it relies on knowing γ^*. Lemma 2 shows that by checking if $\sigma_{z_1\tilde{y}}\sigma_{\tilde{x}z_2} < \sigma_{z_2\tilde{y}}\sigma_{\tilde{x}z_1}$ holds, we can test if there exists some value of γ^* such that the required conditions hold. However, it does not reveal for which values of γ^* this holds; it reveals only that a set of such values exists. To exploit the results of the proposition, we need to assume a value for γ^*. For example, assuming $\gamma^* = 0.5$ implies $\sigma_{xz_1} > \sigma_{xz_2}$ and $\sigma_{z_1 u} < \sigma_{z_2 u}$, so the more relevant variable is also weakly better in terms of validity.

D. Multiple Endogenous X

In this section we consider the case of $k_x > 1$ endogenous regressors with exactly one IIV for each of them. For notational simplicity, assume there are no exogenous regressors except for the intercept term. The model is

$$Y = \beta_0 + X\beta + U. \tag{13}$$

Define

$$\lambda_j^* \equiv \frac{\rho_{z_j u}}{\rho_{x_j u}}, \quad j = 1, \ldots, k_x,$$

where, as before, we define $\lambda_j^* = 0$ if $\rho_{z_j u} = \rho_{x_j u} = 0$. Assumptions 3 and 4 together correspond to $\lambda_j^* \in [0, 1]$, $j = 1, 2$. Since $E[U] = 0$, we can write the model as

$$\overline{Y} = \overline{X}\beta + U,$$

where, for any random variable W, $\overline{W} \equiv W - E[W]$. By definition of λ_j^* it follows that $E[\overline{V}_j(\lambda_j^*)U] = 0, j \in \{1, \ldots k_x\}$, where

$$\overline{V}_j(\lambda_j) \equiv \sigma_{x_j}\overline{Z}_j - \lambda_j\sigma_{z_j}\overline{X}_j.$$

Following our usual notation, define $\overline{V}(\Lambda) \equiv (\overline{V}_1(\lambda_1), \ldots, \overline{V}_{k_x}(\lambda_{k_x}))$. The IV estimand using $\overline{V}(\Lambda)$ as the IV is given by

$$\beta^{IV}_{\overline{V}(\Lambda)} = E[\overline{V}(\Lambda)'\overline{X}]^{-1}E[\overline{V}(\Lambda)'\overline{Y}].$$

The identified set for β is unbounded if and only if the inverse $E[\overline{V}(\Lambda)'\overline{X}]^{-1}$ does not exist or, equivalently, if and only if the determinant of $E[\overline{V}(\Lambda)'\overline{X}]$ can take the value 0 over feasible values of $\Lambda \equiv (\lambda_1, \ldots, \lambda_J)$. For $k_x = 2$, a sufficient condition for the determinant not to equal 0 for all feasible values of Λ is that $(\rho_{x_2 z_1}\rho_{x_1 z_2} - \rho_{x_1 z_1}\rho_{x_2 z_2}) < 0$, $(\rho_{x_1 z_1} - \rho_{x_2 z_1}\rho_{x_1 x_2}) < 0$, and $(\rho_{x_2 z_2} - \rho_{x_1 z_2}\rho_{x_2 x_1}) < 0$.

IV. Estimation and Inference

So far we have focused solely on identification. However, these identification results are constructive, naturally leading to consistent estimators since the derived bounds are probability limits of OLS and IV estimators under assumptions 1, 2, and 5. Consistent estimators of our bounds can thus be computed using standard regression software.

Regarding statistical inference, a variety of methods from the literature are applicable, including Pakes et al. (2005), Chernozhukov et al. (2007), and Andrews and Guggenberger (2009). In section V, we employ a variant of the inferential procedure proposed by Chernozhukov et al. (2009). The method is particularly well suited for settings where the identified set is the intersection of many intervals.

Specifically, we use a sample analog estimator for the identified set of the form of $[L_n, U_n]$, where $L_n^* \equiv \max\{L_{n1}, \ldots, L_{nR}\}$, and $U_n^* \equiv \min\{U_{n1}, \ldots, U_{nS}\}$, and where each L_{nr}, U_{ns} is a consistent estimator of all the lower and upper bounds on the parameter of interest, respectively. To construct confidence intervals, we first construct confidence bands for each of the estimated bounds and then take the intersection of these. Intuitively, this adjusts each of the estimates by an amount that depends on the precision with which it is

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/REST_a_00171 by guest on 27 September 2021

estimated. Those estimates with high standard errors require a larger adjustment than those with lower standard errors. The precise procedure is as follows:

1. Compute $\mathcal{J}_l^* = \{j : L_{nj} \geq L_n^* - 2\bar{\sigma}_l\sqrt{\log n}\}$ and $\mathcal{J}_u^* = \{j : U_{nj} \leq U_n^* + 2\bar{\sigma}_u\sqrt{\log n}\}$ where

$$\bar{\sigma}_l = \max_{j=1,\ldots,R} \hat{s}_{lj}, \quad \bar{\sigma}_u = \max_{j=1,\ldots,S} \hat{s}_{uj},$$

   and where $\hat{s}_{lj}$ and $\hat{s}_{uj}$ are the standard errors of the estimates L_{nj}, U_{nj}, respectively.
2. Define $\Delta_n \equiv |U_n^* - L_n^*|_+$ and $p_n \equiv 1 - \Phi(\log n \times \Delta_n)\alpha$. Let $\hat{\Omega}_l$ and $\hat{\Omega}_u$ be consistent estimators for the variance covariance of $\{L_{nj} : j \in \mathcal{J}_l^*\}$ and $\{U_{nj} : j \in \mathcal{J}_u^*\}$, respectively. Such estimates may be obtained either by analytic calculation of asymptotic variance formulas or by bootstrapping. Let Z^l and Z^u denote $|\mathcal{J}_l^*|$ and $|\mathcal{J}_u^*|$ mean-zero multivariate normal random variables with variances $\hat{\Omega}_l$ and $\hat{\Omega}_u$, respectively.
3. Compute the p_n quantiles of $\max\{Z^l\}$ and $\max\{Z^u\}$, denoted $q^l(p_n)$ and $q^u(p_n)$, respectively. In the case where $|\mathcal{J}_l^*|$ or $|\mathcal{J}_u^*|$ equals 1, this can be done by inverting the normal CDF. Otherwise these may be computed via multivariate normal simulation.
4. A $1-\alpha$ confidence interval for the parameter of interest is then given by $CI_{1-\alpha} = [L_{1-\alpha}, U_{1-\alpha}]$, where

$$L_{1-\alpha} \equiv \max_{j \in \mathcal{J}_l^*} \{L_{nj} - \hat{s}_{lj} q^l(p_n)\},$$
$$U_{1-\alpha} \equiv \min_{j \in \mathcal{J}_u^*} \{U_{nj} + \hat{s}_{uj} q^u(p_n)\}.$$

The confidence set $CI_{1-\alpha}$ provides an asymptotic confidence interval for the bounded parameter of interest, which is uniformly valid over points in the identified set. The first step estimates the set of lower and upper bounds that are sufficiently close to binding to affect the asymptotic distribution of the maximum and minimum boundary estimators. This step may be skipped and the full set of boundary estimates used instead of the subsets $\mathcal{J}_l^*$ and $\mathcal{J}_u^*$, in which case the confidence set is valid but in general conservative. One may also adapt this procedure to provide either a $1-\alpha$ confidence interval for the entire identified set, by setting $p_n = 1 - \alpha/2$, or a pointwise (but not uniformly valid) $1-\alpha$ confidence interval for the parameter of interest by setting $p_n = 1-\alpha$. In our application, we report the confidence interval for $p_n = 0.975$ for inference on the identified set. Taking $p_n = 1 - \Phi(\log n \times \Delta_n)\alpha$ as above with $\alpha = 0.05$ provides an intermediate value between 0.95 and 0.975 that depends on the size of the estimated identified set so as to achieve uniformity, in similar spirit to Imbens and Manski (2004) and Stoye (2009). (For further details, refer to Chernozhukov et al., 2009). A by-product is that this also provides a specification test: if the lower and upper bounds of the confidence set computed with $p_n = 0.975$ cross, then the model is rejected at the 0.05 level. However, such a test does not provide information regarding the source of misspecification.[5]

V. An Application

In this section we provide an application of our method to the estimation of demand for differentiated products, which has played a central role in industrial organization as well other fields over the past few years. The working paper version of this paper presents an additional application to the estimation of a Cobb-Douglas production function.

Assume that the indirect utility consumer i gets from product j in market t is given by

$$u_{ijt} = p_{jt}\beta + w_{jt}'\Gamma + \xi_{jt} + \epsilon_{ijt},$$

where w_{jt}, p_{jt}, and ξ_{jt} are observable characteristics, price, and unobservable characteristics of product j in market t. ϵ_{ijt} is an unobservable stochastic term that captures the idiosyncratic portion of consumer i's taste for product j in market t. We assume that when making purchases, each consumer chooses exactly one good and also has the option to choose an "outside" good, that is, not to buy any of the products. We normalize the mean value of the outside good to be 0 so that $u_{i0t} = \epsilon_{i0t}$. Furthermore, ϵ_{ijt} is assumed to be a distributed iid extreme value, from which it follows that each product j has market share s_{jt} in market t, where

$$s_{jt} = \frac{\exp(p_{jt}\beta + w_{jt}'\Gamma + \xi_{jt})}{1 + \sum_{k=1}^{J} \exp(p_{kt}\beta + w_{kt}\Gamma + \xi_{kt})},$$

and

$$\log(s_{jt}) - \log(s_{0t}) = p_{jt}\beta + w_{jt}'\Gamma + \xi_{jt}. \tag{14}$$

If price p_{jt} and market characteristics w_{jt} were uncorrelated with the random unobservable ξ_{jt}, the parameters of this equation, β and Γ, could be estimated by ordinary least squares. However, it is commonly thought in these markets that any given product's price is correlated with unobservable shifters. In general, the error term may include unobserved product quality or promotional activities, and both are likely to be correlated with price. In this application, we control for unobserved product characteristics that are fixed over time by using product fixed effects. Thus, the error term includes mainly unobserved promotional activities.

We employ scanner data from the ready-to-eat cereal industry at the brand-quarter-MSA (metropolitan statistical area) level, obtained from the IRI Infoscan Data Base at the University of Connecticut. We have observations from twenty quarters, and for this application, we focus attention on the top 25 brands (in terms of market share) and the San Francisco

[5] Crossing of the boundary estimates would reflect that, given the maintained linear specification, the IIV assumptions that have been imposed are not mutually consistent. This is analogous to overidentification tests for GMM as in Hansen (1982). Indeed, assumptions 3 and 4 nest the possibility that the IIVs are valid instruments.

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/REST_a_00171 by guest on 27 September 2021

TABLE 1.—SUMMARY STATISTICS

Variable	Mean	Median	Standard Deviation	Minimum	Maximum	Brand Variation	City Variation	Quarter Variation
Price (¢ per serving)	20.5	20.0	4.9	8.5	40.9	88.5%	6.3%	1.8%
Advertising (Mil.$ per quarter)	3.6	3.0	2.0	0.0	9.8	65.9	NA	1.8
% share within Cereal Market	2.2	1.7	1.4	0.3	7.9	90.3	0.1	0.0

Source: IRI Infoscan Data Base, University of Connecticut, Food Marketing Center.

TABLE 2.—LOGIT DEMAND ESTIMATES

	(1)	(2)	(3)	(4)	(5)
				IIV	
	OLS	IV	2SLS	Only A3	A3 and A4
Price	−2.21	−4.08	−4.47	[−8.69, −4.08]	[−8.69, −5.99]
	(0.72)	(0.89)	(0.89)	(−11.44, −2.32)	(−11.44, −4.04)
Advertising/10	0.31	0.30	0.30	[0.28, 0.30]	[0.28, 0.29]
	(0.01)	(0.06)	(0.06)	(0.16, 0.41)	(0.16, 0.41)
First-stage F-statistic		1,780	930	380	1,116

The dependent variable in all columns is $\log(s_{jt}) - \log(s_{0t})$. The sample includes 990 observation: 25 brands in two cities, Boston and San Francisco, over twenty quarters (one of the brands was introduced after five quarters). The regressions also include brand fixed effects, a dummy variable for San Francisco, and twenty quarterly time dummy variables. Column 2 reports results using the average price in the region as an IV. Column 3 reports two-stage least-squares results using price in the region and the average price in the other city as IVs. Column 4 reports results using the difference between the average price in the other city and the average price in the region as an IIV. Column 5 imposes assumption 4 in addition to assumption 3 to sharpen the bounds. Standard errors are reported in parentheses. In columns 4 and 5 we report 95% confidence intervals for the identified set using the method described in section IV. Finally, the first-stage F-statistic tests the exclusion of the IVs from the (first-stage) regression of price on the instruments and all the exogenous variables.

and Boston markets. The key variables observed for each product, market, and quarter combination are quantity sold, total revenue, and brand-level advertising. We define price as the revenue divided by quantity. (For additional information on the data source and the details of the ready-to-eat cereal industry, see Nevo, 2000, 2001.) Table 1 provides a descriptive summary of the data.

The standard approach for dealing with the endogeneity of price in this setting is to use prices of the product in other markets as an instrumental variable (Hausman, Leonard, & Zona, 1994; Hausman, 1997; Nevo, 2001). The idea is that the IV is correlated with price through common marginal cost shocks. Assuming that the errors in demand are independent across markets, these instruments are valid. This latter assumption has been challenged (for example, the discussion of Hausman, 1997, by Bresnahan, 1997). The demand shocks could be correlated across cities for several reasons. For example, advertising could be at the regional or national levels. Alternatively, the brand preferences could change over time. For instance, if in the middle of the sample, fiber-rich cereals are found to be healthy, the preferences for these cereals could change. Any of these stories would render the IV, and the implied estimates, invalid. There is evidence that despite these theoretical concerns, the IV are valid (Nevo, 2001); nevertheless, some concerns linger over the validity of the estimates.

We use prices in other cities as IIVs. The typical concern in demand estimation is that price, the endogenous variable, is positively correlated with the error term. In order to satisfy assumption 3, we require that prices in other markets, the IIVs, be positively correlated with the error. The examples suggest that usually we worry about positive association between demand errors in different markets. Thus, it is natural to assume that the correlation of prices in other cities with the error term is positive. Thus, assumption 3 is likely satisfied. Unfortunately, in our data, it is also the case that prices in other markets are positively correlated with price. Therefore, these IIV will yield only one-sided bounds. However, we can exploit the fact that we have multiple cities to generate a valid IIV, using the results in section IIIC. For each of our markets, Boston and San Francisco, we use two IIVs. Denote by Z_1 the average price in the other markets in the region—New England for Boston and northern California for San Francisco—and by Z_2 the average price in the other city. In Nevo & Rosen (2008), we provide a model that justifies the required assumptions. The intuitive idea is as follows. The average price in the region, Z_1, is assumed to be more correlated with price because marginal cost shocks are common. It is also assumed to be less correlated with demand because the composition of demand is assumed to be more similar between Boston and San Francisco than with their surrounding (less urban) regions.

In the sample, the partial correlations, $\rho_{pz_1} = 0.81$ and $\rho_{pz_2} = 0.48$ satisfy the first of these assumptions. Furthermore, lemma 2 allows us to verify the second assumption. Indeed, we find that $\sigma_{\tilde{y}z_1}\rho_{\tilde{x}z_2} < 0 < \sigma_{\tilde{y}z_2}\rho_{\tilde{x}z_1}$, so by the lemma, there exists a range of γ such that $\omega(\gamma) = \gamma Z_2 - (1-\gamma)Z_1$ satisfies the conditions required to get a two-sided bound. The results in table 2 assume $\gamma = 0.5$. We also explored $\gamma = \sigma_{z_1}/(\sigma_{z_1} + \sigma_{z_2})$ and found almost identical results. For both values of γ we can verify that the correlation between price and $\omega(\gamma) = \gamma Z_2 - (1-\gamma)Z_1$ is negative.

In some of the results, we also use assumption 4 on the differenced IV. In the context of the demand application, assumption 4 is quite intuitive. It implies that the price in another market, say Z_1, is less correlated with the error term

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/REST_a_00171 by guest on 27 September 2021

than is the price in the same market. In Nevo & Rosen (2008) we provide a model that formalizes this intuition. If we assume that assumptions 3 and 4 hold for both Z_1 and Z_2, then assumption 4 will also hold for the weighted difference, and the second set of bounds from proposition 5 applies.

The results are presented in table 2. The dependent variable in all columns is $\log(s_{jt}) - \log(s_{0t})$. The first column presents results from regressing this variable on price, brand, and quarter dummy variables and a city of San Francisco dummy variable. The estimated price coefficient is negative, and the advertising coefficient is positive, as expected. However, the own price elasticities are less than 1 in absolute value. Once we use the average regional price as an IV, the price coefficient becomes more negative, as expected. The same is true when we use both the average regional price and the price in the other city jointly.

The next two columns use weighted differences between the average regional price and the price in the other city as IIV. The results in column 4 are from only imposing assumption 3 on the differenced IV, while in column 5, we present results when we additionally impose assumption 4.

The results yield a fairly consistent picture. If we do not impose assumption 4, then the price coefficient is between -8.7 and -4 (with a confidence interval of -11.4 and -2.3). If we impose assumption 4, the price coefficient is between -8.7 and -6 (with a confidence interval of -11.4 and -4).[6] In all cases, the OLS point estimate is outside the confidence interval. When we impose assumption 4, the IV estimate is very close to the boundary of the confidence interval. Note that in all cases, the excluded variables are highly significant in the first stage, as indicated by the reported F-statistics.

In the logit model price elasticities are proportional to the price coefficient. So going from column 1 to 2 doubles the price elasticities by roughly a factor of two. Further moving to the lower bound of the identified set increases the price elasticities by another factor of two. Generally the results suggest that the IV price elasticities are too low in absolute value.

There are two common uses for demand elasticities in the IO literature. Often the elasticities are used in a first-order condition, typically from a Bertrand pricing game, in order to compute price cost margins (PCM). PCM computed in this way are used to test among different supply models (Nevo, 2001). Our results suggest that the estimates of PCM using the standard IV assumption might be too high. Another use of demand estimates is for simulation of the effects of mergers (Hausman et al., 1994, and Nevo, 2000). The results in table 2 suggest that estimates using the standard IV assumption would tend to underestimate the effects of a merger, because they will tend to underestimate the substitution among products.

[6] The confidence intervals reported in the table are for the identified set. We also computed the confidence interval for the estimated parameters, which are essentially identical and therefore not reported. For example, for column 4, the confidence interval for the parameter was $[-11, -2.6]$.

VI. Conclusion

In this paper we study identification of the parameters of a single regression equation with endogenous regressors and instruments that fail to satisfy the usual exogeneity condition. Instead, we consider cases where the instruments are assumed to have the same direction of correlation with the error as the endogenous regressor, but where the instruments are less correlated with the error term than is the endogenous regressor.

Under our assumptions, we first derive an abstract characterization of the identified set for all parameters in a general class of models and then focus primarily on set identification of model parameters with a linear specification. Consistent estimates of these bounds can be computed by standard OLS and linear 2SLS regressions under the usual rank conditions. We found that the bounds obtained, and in particular whether they form an open or closed interval depends on the correlation between the endogenous regressor and the instrument, which is point-identified. Furthermore, depending on both the sign of σ_{xz} and its magnitude relative to σ_{xu} and σ_{zu}, β may be closer to either β^{OLS} or β_z^{IV} even if Z is less endogenous than X, in the sense that our assumptions 3 and 4 are satisfied.

Relative to conventional instrumental variable assumptions, the cost of our approach is that the weaker assumptions we impose generally yield partial identification rather than point identification of model parameters. The benefit, however, is that inferences made are robust to a lack of instrument exogeneity and thus may be more credible in some circumstances. Additionally, for cases in which the applied researcher wishes to impose instrument exogeneity, our approach provides one answer to the question of how much this assumption drives their results.

Our focus has been entirely on parametric models, with specialized results for linear models. While much empirical work relies on such models, an interesting extension would be to perform a similar analysis in a nonparametric model. However, with a nonparametric functional form, it is doubtful that our assumptions on the correlations of endogenous regressors and imperfect instruments with econometric errors would prove anywhere near as fruitful. More promising would be an extension of our assumptions to conditional expectation analogs. Indeed, the identifying power of MIV assumptions was examined in a nonparametric context by Manski and Pepper (2000). Their MIV assumption is a conditional expectation analog of our assumption that the instrument is positively correlated with the error. One could also posit a conditional mean version of our assumption that the instrument is less correlated with the latent variate than is the endogenous regressor. In light of the positive results of Manski and Pepper, it would seem that such an assumption

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/REST_a_00171 by guest on 27 September 2021

could have significant identifying power in a nonparametric model.

In the linear case, our focus was a model with a fixed parameter. The heterogeneous effects model has attracted much interest lately. For example, Angrist et al. (1996) show that the standard IV estimate can be interpreted as a local average treatment effect (LATE) under a standard exclusion restriction and a monotonicity condition on the effect of the IV on the endogenous variable. Angrist et al. (1996) show the bias in the estimate if the exclusion restriction is violated. It would be interesting to extend our analysis to this setup.

REFERENCES

Altonji, J. G., T. E. Elder, and C. R. Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy* 113 (2005), 151–184.

Andrews, D. W. K., and P. Guggenberger, "Validity of Subsampling and Plug-In Asymptotic Inference for Parameters Defined by Moment Inequalities," *Econometric Theory* 25 (2009), 669–709.

Angrist, J., G. Imbens, and D. B. Rubin, "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association* 91 (1996), 444–455.

Ashley, R., "Assessing the Credibility of Instrumental Variables Inference with Imperfect Instruments via Sensitivity Analysis," *Journal of Applied Econometrics* 24 (2009), 325–337.

Berkowitz, D., M. Caner, and Y. Fang, "Are 'Nearly Exogenous' Instruments Reliable?" *Economics Letters* 101 (2008), 20–23.

Bontemps, C., T. Magnac, and E. Maurin, "Set Identified Linear Models," IDEI, working paper (2006).

Bresnahan, T. F., "Comment on Valuation of New Goods under Perfect and Imperfect Competition" (pp. 237–247), in T. F. Bresnahan and R. J. Gordon (Eds.), *The Economics of New Goods* (Chicago: University of Chicago Press, 1997).

Chernozhukov, V., H. Hong, and E. Tamer, "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75 (2007), 1243–1284.

Chernozhukov, V., S. Lee, and A. Rosen, "Intersection Bounds, Estimation and Inference," CEMMAP working paper no. CWP19/09 (2009).

Conley, T., C. Hansen, and P. E. Rossi, "Plausibly Exogenous," Chicago Graduate School of Business, working paper (2006).

Frisch, R., *Statistical Confluence Analysis by Means of Complete Regression Systems* (Oslo, Norway: University Institute for Economics, 1934).

Hahn, J., and J. Hausman, "IV Estimation with Valid and Invalid Instruments," MIT working paper (2003).

Hansen, L. P., "Large Sample Properties of Generalized Method of Moment Estimators," *Econometrica* 50 (1982), 1029–1054.

Hausman, J. A. "Valuation of New Goods under Perfect and Imperfect Competition" (pp. 209–237), in T. F. Bresnahan and R. J. Gordon (Eds.), *The Economics of New Goods* (Chicago: University of Chicago Press, 1997).

Hausman, J. A., G. K. Leonard, and J. D. Zona, "Competitive Analysis with Differentiated Products," *Annales D'Économie et de Statistique* 34 (1994), 159–180.

Hotz, V. J., C. H. Mullin, and S. G. Sanders, "Bounding Causal Effects Using Data from a Contaminated Natural Experiment: Analysing the Effects of Teenage Child Bearing," *Review of Economic Studies* 64 (1997), 575–603.

Imbens, G., "Sensitivity to Exogeneity Assumptions in Program Evaluation," *American Economic Review* 93 (2003), 126–132.

Imbens, G., and C. F. Manski, "Confidence Intervals for Partially Identified Parameters," *Econometrica* 72 (2004), 1845–1857.

Klepper, S., and E. E. Leamer, "Consistent Sets of Estimates for Regressions with Errors in All Variables," *Econometrica* 52 (1984), 163–184.

Kraay, A., "Instrumental Variables Regressions with Honestly Uncertain Exclusion Restrictions," World Bank working paper (2008).

Leamer, E. E., "Is It a Demand Curve, or Is It a Supply Curve? Partial Identification through Inequality Constraints," this REVIEW 63 (1981), 319–327.

Manski, C. F., and J. Pepper, "Monotone Instrumental Variables: With an Application to the Returns to Schooling," NBER working paper (1998).

——— "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica* 68 (2000), 997–1010.

Nevo, A., "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry," *Rand Journal of Economics* 31 (2000), 395–421.

——— "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica* 69 (2001), 307–342.

Nevo, A., and A. Rosen, "Identification with Imperfect Instruments," CEMMAP working paper no. 16/08 (2008).

Pakes, A., J. Porter, K. Ho, and J. Ishii, "The Method of Moments with Inequality Constraints," Harvard University, working paper (2005).

Rosenbaum, P. R., *Observational Studies* (New York: Springer, 2002).

Rosenbaum, P., and D. S. Small, "War and Wages: The Strength of Instrumental Variables and Their Sensitivity to Unobserved Biases," *Journal of the American Statistical Association* 103 (2008), 925–933.

Small, D. S., "Sensitivity Analysis for Instrumental Variables Regression with Overidentifying Restrictions," *Journal of the American Statistical Association* 102 (2007), 1049–1058.

Stoye, J., "More on Confidence Regions for Partially Identified Parameters," *Econometrica* 77 (2009), 1299–1315.

APPENDIX A

Proofs

Lemma 1

Proof. The result follows directly from inspection of the expressions for β^{OLS} and β_z^{IV} given by equations (4) and (5), respectively.

Proposition 1

Proof. We note that if $\sigma_{xu}, \sigma_{zu} \geq 0$, then assumption 4 gives

$$\rho_{xu} \geq \rho_{zu} \geq 0 \Leftrightarrow \sigma_z\sigma_{xu} \geq \sigma_x\sigma_{zu}$$
$$\Leftrightarrow \sigma_z(\sigma_{xy} - \sigma_x^2\beta) \geq \sigma_x(\sigma_{zy} - \sigma_{xz}\beta) \Leftrightarrow \beta_{v(1)}^{IV} \geq \beta.$$

If instead $\sigma_{xu}, \sigma_{zu} \leq 0$, then $\sigma_z\sigma_{xu} \leq \sigma_x\sigma_{zu}$ and $\beta_{v(1)}^{IV} \leq \beta$. Notice that $\beta_{v(1)}^{IV} = \gamma\beta^{OLS} + (1-\gamma)\beta_z^{IV}$, where $\gamma \equiv \sigma_z\sigma_x/(\sigma_z\sigma_x - \sigma_{xz}) = 1/(1-\rho_{xz})$.

First consider the case where $\sigma_{xz} < 0$. Then $\gamma > 0$ and $1-\gamma > 0$, implying that $\beta_{v(1)}^{IV}$ lies between β^{OLS} and β_z^{IV}. If $\sigma_{xu} \geq 0$, lemma 1 implies that $\beta \in [\beta_z^{IV}, \beta^{OLS}]$. Since in this case, we also have that $\beta_{v(1)}^{IV} \geq \beta$, and $\beta_{v(1)}^{IV} \leq \beta^{OLS}$, it follows that $\beta_{v(1)}^{IV}$ provides a smaller upper bound for β: $\beta \in [\beta_z^{IV}, \beta_{v(1)}^{IV}]$. If instead $\sigma_{xu} < 0$, then similar logic leads to $\beta \in [\beta_{v(1)}^{IV}, \beta_z^{IV}]$.

Now suppose $\sigma_{xz} > 0$. Then if $\sigma_{xu} \geq 0$, $\beta_{v(1)}^{IV} \geq \beta$, and lemma 1 gives $\beta \leq \min\{\beta^{OLS}, \beta_z^{IV}\}$. An immediate implication is that $\beta \leq \min\{\beta^{OLS}, \beta_z^{IV}, \beta_{v(1)}^{IV}\}$, but it turns out that β^{OLS} is redundant in that $\min\{\beta^{OLS}, \beta_z^{IV}, \beta_{v(1)}^{IV}\} = \min\{\beta_z^{IV}, \beta_{v(1)}^{IV}\}$. This is because if $\beta^{OLS} < \beta_z^{IV}$, and then $\beta_{v(1)}^{IV} \leq \beta^{OLS}$, while clearly if instead $\beta^{OLS} \geq \beta_z^{IV}$, then $\min\{\beta^{OLS}, \beta_z^{IV}\} = \beta_z^{IV}$. The claim that $\beta^{OLS} < \beta_z^{IV} \Rightarrow \beta_{v(1)}^{IV} \leq \beta^{OLS}$ holds because $\sigma_{xz} > 0$ implies that $1-\gamma = -\sigma_{xz}/(\sigma_z\sigma_x - \sigma_{xz}) < 0$, so that

$$\beta_{v(1)}^{IV} = \gamma\beta^{OLS} + (1-\gamma)\beta_z^{IV} \leq \gamma\beta^{OLS} + (1-\gamma)\beta^{OLS} = \beta^{OLS}.$$

Symmetric reasoning applied to the case where $\sigma_{xz} > 0$ and $\sigma_{xu} < 0$ gives that $\beta \geq \max\{\beta_z^{IV}, \beta_{v(1)}^{IV}\}$. The sharpness of the bounds follows as an implication of proposition 4, which covers the case of multiple imperfect instruments and is proven below.

Corollary 1

Proof. Using the above notation,

$$\beta_{v(1)}^{IV} - \beta_z^{IV} = \gamma\beta^{OLS} + (1-\gamma)\beta_z^{IV} - \beta_z^{IV} = \gamma(\beta^{OLS} - \beta_z^{IV}).$$

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/REST_a_00171 by guest on 27 September 2021

Proposition 2

Proof. Assumptions 3 and 4 and $\rho_{xu} > 0$ give

$$\begin{aligned}&\rho_{xu} \geq \rho_{zu} \geq 0 \Leftrightarrow \sigma_{xu}\sigma_z \geq \sigma_{zu}\sigma_x \geq 0\\ &\Leftrightarrow \sigma_{x\tilde{y}}\sigma_z - \sigma_{x\tilde{x}}\sigma_z\beta \underset{(1)}{\geq} \sigma_{z\tilde{y}}\sigma_x - \sigma_{z\tilde{x}}\sigma_x\beta \underset{(2)}{\geq} 0,\end{aligned}$$

where the third line uses $\tilde{Y} = \beta\tilde{X} + U$, which is shown in lemma 3 below. Inequality 1 provides an upper bound on β if $\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z$ is negative, and a lower bound if this expression is positive. Inequality 2 provides an upper bound if $\sigma_{z\tilde{x}}$ is positive and a lower bound if $\sigma_{z\tilde{x}}$ is negative. Combining these inequalities and a symmetrical derivation for the case $\rho_{xu} < 0$ gives the conclusion of the proposition.

Lemma 3. *Let the conditions of proposition 2 hold. Then* $\tilde{Y} = \beta\tilde{X} + U$.

Proof. Subtracting $WE[Z^{w\prime}W]^{-1}E[Z^{w\prime}Y]$ from both sides of equation (8) gives

$$\begin{aligned}\tilde{Y} &= X\beta + W\delta - WE[Z^{w\prime}W]^{-1}E[Z^{w\prime}Y] + U\\ &= X\beta + W\delta - WE[Z^{w\prime}W]^{-1}E[Z^{w\prime}(X\beta + W\delta + U)] + U\\ &= X\beta + W\delta - WE[Z^{w\prime}W]^{-1}E[Z^{w\prime}X]\beta - WE[Z^{w\prime}W]^{-1}E[Z^{w\prime}W]\delta + U\\ &= \tilde{X}\beta + U.\end{aligned}$$

Proposition 3

Proof. Starting with the definition of $\tilde{Y}^*$, and making use of $Y^* = Y - X\beta$, we have that

$$\begin{aligned}\tilde{Y}^* &= Y^* - W_{-j}E\big[Z_{-j}^{w\prime}W_{-j}\big]^{-1}E\big[Z_{-j}^{w\prime}Y^*\big]\\ &= Y - X\beta - W_{-j}E\big[Z_{-j}^{w\prime}W_{-j}\big]^{-1}E\big[Z_{-j}^{w\prime}Y\big]\\ &\quad + W_{-j}E\big[Z_{-j}^{w\prime}W_{-j}\big]^{-1}E\big[Z_{-j}^{w\prime}X\big]\beta = \tilde{Y} - \tilde{X}\beta,\end{aligned}$$

where $\tilde{Y} = Y - W_{-j}E[Z_{-j}^{w\prime}W_{-j}]^{-1}E[Z_{-j}^{w\prime}Y]$. Now plugging this into equation (12) gives

$$\begin{aligned}\delta_j &= E\big[Z_j^{w\prime}\tilde{W}_j\big]^{-1}E\big[Z_j^{w\prime}\tilde{Y}^*\big]\\ &= E\big[Z_j^{w\prime}\tilde{W}_j\big]^{-1}E\big[Z_j^{w\prime}\tilde{Y}\big] - E\big[Z_j^{w\prime}\tilde{W}_j\big]^{-1}E\big[Z_j^{w\prime}\tilde{X}\big]\beta,\end{aligned}$$

so that δ_j is linear in β, which delivers the conclusion of the proposition, since all other components of the above expression are point-identified.

Proposition 4

Proof. That the true population parameter β belongs to B^* as defined in the statement of the proposition follows immediately from proposition 2. This is because proposition 2 can be applied to each Z_j individually, yielding that β must fall in the intersection of each of the bounds obtained with each Z_j individually.

It remains to show that these bounds are sharp—that any value of $b \in B^*$ is feasible. Choose an arbitrary $b \in B^*$. To prove sharpness, it must be shown that there exists a random variable $\tilde{U}$ such that the joint distribution of (Y, X, W, Z, Z^w, U) with $U = \tilde{U}$ satisfies all of our assumptions if $\beta = b$. Define $\tilde{U} \equiv \tilde{Y} - \tilde{X}b$, where $\tilde{Y}$ and $\tilde{X}$ are as defined in equations (10a) and (10b). By definition $E[Z^w\tilde{U}] = 0$. Next, adopt the shorthand $Z = Z_j$ for any j and define

$$\lambda^*(b) \equiv \frac{\rho_{z\tilde{u}}}{\rho_{x\tilde{u}}} = \frac{\sigma_x\sigma_{\tilde{y}z} - b\sigma_x\sigma_{\tilde{x}z}}{\sigma_z\sigma_{x\tilde{y}} - b\sigma_z\sigma_{\tilde{x}x}}. \tag{A1}$$

To complete the proof of sharpness, we now show that $\lambda^*(b) \in [0, 1]$, for arbitrary choice of j, which implies that assumptions 3 and 4 are both satisfied. Note that the derivative of $\lambda^*(b)$ is

$$\frac{d\lambda^*(b)}{db} = \sigma_x\sigma_z\frac{\sigma_{\tilde{x}x}\sigma_{\tilde{y}z} - \sigma_{\tilde{x}z}\sigma_{x\tilde{y}}}{(\sigma_z\sigma_{x\tilde{y}} - b\sigma_z\sigma_{\tilde{x}x})^2},$$

implying that $\lambda^*(b)$ is monotone in the direction of $\sigma_{\tilde{x}x}\sigma_{\tilde{y}z} - \sigma_{\tilde{x}z}\sigma_{x\tilde{y}}$.

First, suppose $(\sigma_{\tilde{x}x}\sigma_z - \sigma_x\sigma_{\tilde{x}z})\sigma_{\tilde{x}z} < 0$. Then by proposition 2, b lies between β_z^{IV} and $\beta_{v(1)}^{IV}$. These values correspond to values of $\lambda^*(b)$ of 0 and 1, respectively, and by monotonicity of $\lambda^*(b)$, it follows that $\lambda^* \in [0, 1]$. Now suppose instead that $(\sigma_{\tilde{x}x}\sigma_z - \sigma_x\sigma_{\tilde{x}z})\sigma_{\tilde{x}z} \geq 0$. We consider two subcases corresponding to the sign of $\sigma_{\tilde{x}z}$:

Case 1. $\sigma_{z\tilde{x}} < 0$. Then $b \in [\max\{\beta_z^{IV}, \beta_{v(1)}^{IV}\}, \infty)$ and $\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z \geq 0$. If $\beta_{v(1)}^{IV} > \beta_z^{IV}$, then it follows that $\sigma_{z\tilde{y}}\sigma_{x\tilde{x}} - \sigma_{z\tilde{x}}\sigma_{x\tilde{y}} < 0$, $b \in [\beta_{v(1)}^{IV}, \infty)$, and $\frac{d\lambda^*(b)}{db} < 0$. Thus, the upper bound on $\lambda^*(b)$ is 1, achieved when $b = \beta_{v(1)}^{IV}$. Using L'Hôpital's rule, if follows that the limit of $\lambda^*(b)$ as $b \to \infty$ is $\lambda^*(b) = \frac{\sigma_x\sigma_{\tilde{x}z}}{\sigma_z\sigma_{\tilde{x}x}} = \frac{\rho_{\tilde{x}z}}{\rho_{\tilde{x}x}}$, which is ≥ 0 from the inequalities $\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z \geq 0$ (since $(\sigma_{\tilde{x}x}\sigma_z - \sigma_x\sigma_{\tilde{x}z})\sigma_{\tilde{x}z} \geq 0$) and $\sigma_{z\tilde{x}} < 0$. If instead $\beta_{v(1)}^{IV} \leq \beta_z^{IV}$, then $\sigma_{z\tilde{y}}\sigma_{x\tilde{x}} - \sigma_{z\tilde{x}}\sigma_{x\tilde{y}} \geq 0$, $b \in \left[\frac{\sigma_{z\tilde{y}}}{\sigma_{z\tilde{x}}}, \infty\right)$, and $\frac{d\lambda^*(b)}{db} \geq 0$. Thus, the lower bound on $\lambda^*(b)$ is 0, achieved when $b = \beta_z^{IV}$. The upper bound on $\lambda^*(b)$ is the limit of expression (A1) as $b \to \infty$, which is $\lambda^*(b) = \frac{\rho_{\tilde{x}z}}{\rho_{\tilde{x}x}} \leq 1$, again since $\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z \geq 0$.

Case 2. $\sigma_{z\tilde{x}} > 0$. Then $b \in (-\infty, \min\{\beta_z^{IV}, \beta_{v(1)}^{IV}\}]$ and $\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z \leq 0$. Following the same logic as when $\sigma_{z\tilde{x}} < 0$, we have that when $\beta_z^{IV} < \beta_{v(1)}^{IV}$, $\frac{d\lambda^*(b)}{d\beta} < 0$, and $\lambda^*(b)$ as a function of b takes values on the interval $[0, \frac{\rho_{\tilde{x}z}}{\rho_{\tilde{x}x}})$. When instead $\beta_z^{IV} \geq \beta_{v(1)}^{IV}$, $\lambda^*(b) \in \left(\frac{\rho_{\tilde{x}z}}{\rho_{\tilde{x}x}}, 1\right]$. In both cases, $\frac{\rho_{\tilde{x}z}}{\rho_{\tilde{x}x}} \in [0, 1]$ follows from the inequalities $\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z \leq 0$ and $\sigma_{z\tilde{x}} > 0$.

Lemma 2

Proof. We assume that condition 1 holds and show it is equivalent to 2. Condition 1 holds if and only if there exists $\gamma^* \in [0, 1]$ such that $\gamma^*\sigma_{z_2u} - (1 - \gamma^*)\sigma_{z_1u} \geq 0$ and $\gamma^*\sigma_{\tilde{x}z_2} - (1 - \gamma^*)\sigma_{\tilde{x}z_1} < 0$, or, equivalently,

$$\gamma^* \geq \frac{\sigma_{z_1u}}{\sigma_{z_1u} + \sigma_{z_2u}} \geq 0 \tag{A2}$$

and

$$\gamma^* < \frac{\sigma_{\tilde{x}z_1}}{\sigma_{\tilde{x}z_1} + \sigma_{\tilde{x}z_2}} \leq 1, \tag{A3}$$

from which it follows that

$$\frac{\sigma_{\tilde{x}z_1}}{\sigma_{\tilde{x}z_1} + \sigma_{\tilde{x}z_2}} - \frac{\sigma_{z_1u}}{\sigma_{z_1u} + \sigma_{z_2u}} \geq 0.$$

Substituting $\sigma_{z_ju} = \sigma_{z_j\tilde{y}} - \sigma_{\tilde{x}z_j}\beta$ for $j = 1, 2$ and collecting terms gives $\sigma_{z_1\tilde{y}}\sigma_{\tilde{x}z_2} < \sigma_{z_2\tilde{y}}\sigma_{\tilde{x}z_1}$, condition 3. From the inequalities $\sigma_{z_ju} \geq 0$ and $\sigma_{\tilde{x}z_j} > 0$ (implied by $(\sigma_{x\tilde{x}}\sigma_{z_j} - \sigma_{\tilde{x}z_j}\sigma_x)\sigma_{\tilde{x}z_j} > 0$) it also follows that condition 3 $\Rightarrow$ condition 1. The equivalence of conditions 1 and 2 follows since equation (A2) holds if and only if $\frac{\sigma_{\tilde{x}z_1}}{\sigma_{\tilde{x}z_2}} > \frac{\gamma^*}{1-\gamma^*}$, and equation (A3) holds if and only if $\frac{\gamma^*}{1-\gamma^*} > \frac{\sigma_{z_1u}}{\sigma_{z_2u}}$. That the inequality delivering two-sided bounds is implied by the partial covariance between X and $\omega(\gamma^*)$ being negative (which is part of condition 1) follows from first noting that this partial covariance is equal to $\sigma_{\tilde{x}\omega(\gamma^*)}$, and then by using the fact that $\sigma_{x\tilde{x}} > 0$. This implies that $\sigma_{\omega(\gamma^*)\tilde{x}} < 0 \Rightarrow (\sigma_{\omega(\gamma^*)\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_{\omega(\gamma^*)}) < 0$, and consequently that $(\sigma_{\omega(\gamma^*)\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_{\omega(\gamma^*)})\sigma_{\omega(\gamma^*)\tilde{x}} > 0$.

Proposition 5

Proof. The proof follows by application of lemma 1 and proposition 4 with $\omega(\gamma^*)$ as an IIV for X. All that needs to be shown for the latter set of bounds is that if assumption 4 holds for both Z_1 and Z_2, then it also holds for $\omega(\gamma^*)$, $\rho_{xu} \geq \rho_{\omega u}$, or, equivalently, $\sigma_{xu}\sigma_\omega \geq \sigma_{\omega u}\sigma_x$. By definition we have

$$\sigma_x\sigma_{\omega u} = \gamma\sigma_x\sigma_{z_2u} - (1 - \gamma)\sigma_x\sigma_{z_1u}.$$

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/REST_a_00171 by guest on 27 September 2021

Note that by the Cauchy-Schwarz inequality,

$$\begin{aligned}\text{var}(\omega) &= \gamma^2\sigma_{z_2}^2 + (1-\gamma)^2\sigma_{z_1}^2 - 2\gamma(1-\gamma)\sigma_{z_1 z_2} \\ &\geq \gamma^2\sigma_{z_2}^2 + (1-\gamma)^2\sigma_{z_1}^2 - 2\gamma(1-\gamma)\sigma_{z_1}\sigma_{z_2} \\ &= (\gamma\sigma_{z_2} - (1-\gamma)\sigma_{z_1})^2,\end{aligned}$$

which implies that $\sigma_\omega \geq |\gamma\sigma_{z_2} - (1-\gamma)\sigma_{z_1}|$. If $\gamma\sigma_{z_2} - (1-\gamma)\sigma_{z_1} > 0$, we have that

$$\sigma_{xu}\sigma_\omega \geq \sigma_{xu}(\gamma\sigma_{z_2} - (1-\gamma)\sigma_{z_1}) \geq \gamma\sigma_{xu}\sigma_{z_2},$$

since $\sigma_{xu} \geq 0$, while if $\gamma\sigma_{z_2} - (1-\gamma)\sigma_{z_1} < 0$, we have that

$$\sigma_{xu}\sigma_\omega \geq \sigma_{xu}((1-\gamma)\sigma_{z_1} - \gamma\sigma_{z_2}) \geq (1-\gamma)\sigma_{xu}\sigma_{z_1} > \gamma\sigma_{xu}\sigma_{z_2}.$$

So in both cases,

$$\sigma_{xu}\sigma_\omega \geq \gamma\sigma_{xu}\sigma_{z_2} \geq \gamma\sigma_x\sigma_{z_2 u} \geq \gamma\sigma_x\sigma_{z_2 u} - (1-\gamma)\sigma_x\sigma_{z_1 u} = \sigma_x\sigma_{\omega u},$$

where the second inequality follows by assumption 4 for Z_2, the third since $\sigma_{z_1 u} \geq 0$, and the equality by definition of ω.

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/REST_a_00171 by guest on 27 September 2021