NBER WORKING PAPER SERIES

IDENTIFICATION WITH IMPERFECT INSTRUMENTS

Aviv Nevo
Adam M. Rosen

Identification with Imperfect Instruments
Aviv Nevo and Adam M. Rosen
NBER Working Paper No. 14434
October 2008
JEL No. C30,C31,C33

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 assume (i) the correlation between the instrument and the error term has the same sign as the correlation between the endogenous regressor and the error term, and (ii) that the instrument is less correlated with the error term than is the endogenous regressor. Using these assumptions, we derive analytic bounds for the parameters. We demonstrate the method in two applications.

Aviv Nevo
Department of Economics
Northwestern University
2001 Sheridan Road
Evanston, IL 60208-2600
and NBER
nevo@northwestern.edu

Adam M. Rosen
Department of Economics
Gower Street
London WC1E 6BT
adam.rosen@ucl.ac.uk

Identification with Imperfect Instruments [*]

Aviv Nevo[†]   Adam M. Rosen[‡]

Northwestern University and NBER   UCL, IFS, and CEMMAP

July 16, 2008

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 assume (i) the correlation between the instrument and the error term has the same sign as the correlation between the endogenous regressor and the error term, and (ii) that the instrument is less correlated with the error term than is the endogenous regressor. Using these assumptions, we derive analytic bounds for the parameters. We demonstrate the method in two applications.

1   Introduction

Dealing with endogeneity is one of the central issues in non-experimental studies. A common method for dealing with potential endogeneity issues in econometric models 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.

The exogeneity assumption cannot be tested in many cases and therefore the validity of the identifying restrictions is based on the subjective judgement of the researcher. Unfortunately, even when great care is

taken, the validity of the instruments chosen is often a matter of faith. Empirical findings are often called into question as a result of debate over the IV assumptions. Furthermore, the search for exogenous IVs sometimes leads the researcher towards IVs that are weak, i.e., are only weakly correlated with the endogenous variable.

In this paper we examine identification when we replace the standard exogeneity assumption with a weaker inequality. We set up the model in general, but in order to obtain analytic results we focus on a linear equation with one endogenous regressor, when the researcher has at least one imperfect instrumental variable (IIV). The IIV is allowed to be correlated with the error term. We show that without further assumptions this variable does not identify even the direction of the bias of the least squares estimator.

We assume that the sign of the correlation between the IIV and the error term is the same as the correlation between the endogenous variable and the error term. We show how this assumption can be used to partially identify the parameter of interest. In particular, we show that if the IIV and the endogenous variable are negatively correlated, the parameter of interest can be bounded both from above and below. We add the assumption that the correlation coefficient between the IIV and the error term is less than the correlation coefficient 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. Using this assumption we improve the bounds. Finally, we discuss the case where several IIVs are available. Of particular interest is the case where none of these variables provide useful bounds on their own. We show how differencing the variables to create an IIV can generate meaningful bounds.

A strength of our approach is its simplicity. By focusing on the linear model we are able to analytically characterize the identified set. Furthermore, the bounds we compute are well known estimators that are computed by standard econometric packages. Thus, the estimation and inference are simple to do, and require little additional programming. Indeed, many of the numbers we need to compute our bounds are available in many published papers.

We apply our estimator to two common examples in the Industrial Organization literature: production function estimation and demand estimation. While our motivation lies in applications such as these, the method is more generally applicable in practically all fields of economics. The applications demonstrate that, at least in the cases we examine, our approach can generate useful bounds on the parameters of interest.

1.1 Related Literature

Three related papers that also consider inference when instrument exogeneity fails are Conley, Hansen, and Rossi (2006), Hahn and Hausman (2003a), and Manski and Pepper (2000). Like our own work Conley,

Hansen, and Rossi (2006) consider a linear model with endogenous regressors and invalid instruments. However, they take an alternative approach to ours by parameterizing the degree of instrument endogeneity by its associated coefficient, denoted γ, in a linear regression of the outcome variable on regressors and instruments jointly. They first investigate the degree to which knowledge of only the support of γ can be used to infer the parameter of interest β. They then consider the added benefit of specifying a prior distribution for γ, and the implications of imposing a prior distribution over all model parameters jointly. This provides useful (set) identification results, inferential procedures, and methods of sensitivity analysis, though for a different set of modeling assumptions than those considered here. Rather than derive the identified set for model parameters in the presence of imperfect instruments, Hahn and Hausman (2003a) compare the mean-squared-error and bias of the OLS and 2SLS estimators when the instrument exogeneity assumption fails. 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. Unlike our analysis, their analysis applies to nonparametric models, but focuses on a bounded dependent variable. When a linear parametric model is imposed, as done in Manski and Pepper (1998), the Manski and Pepper assumptions neither nest nor are nested by ours. The Manski and Pepper monotone IV assumption implies monotonicity of the mean of the unobservable conditional on the instrument, whereas we employ weaker restrictions on correlations. However, in addition to this we also impose restrictions on the relative correlations of exogenous regressors with errors and instruments with errors, and there are no such restrictions imposed by Manski and Pepper. A benefit of the assumptions imposed in this paper, at least in the context of linear models, are analytic results on the bounds of the parameter of interest, which are easy to compute with standard regression software.

Related papers that bound parameters of a linear regression include those of Frisch (1934), Leamer (1981), Klepper and Leamer (1984), and Bontemps, Magnac, and Maurin (2006). Frisch (1934) develops bounds for the slope parameter in a simple linear regression model with measurement error. Klepper and Leamer (1984) generalize this result to multivariate regression with errors in all variables, deriving bounds on the true regression coefficient vector. Leamer (1981) considers an under-identified simultaneous equations model of supply and demand without instruments to overcome the simultaneity problem. He shows that knowledge of the signs of regression parameters, e.g. downward sloping demand, upward sloping supply, can be used to bound the slope of either demand or supply via directed and reverse regression. Bontemps, Magnac, and Maurin (2006) study set identification in linear models, providing an estimation procedure that relies

on an estimator for the identified set's support function. In addition, they generalize the Sargan test for overidentification to their setup, establishing a test of supernumerary restrictions whose intersection may in fact be a proper set, a point, or empty. While each of these papers provides bounds on the coefficients of a linear regression model, the models considered are quite different from the one we study in this paper.

Whereas this paper addresses the potential failing of the exogeneity condition for an instrument, the recent literature on weak instruments and weak identification confronts the possibility of the instruments having little relevance. This literature focusses on complications that arise when the relevance condition is not strongly satisfied, i.e. when the correlation between the endogenous regressor and the instruments is small. Standard methods for inference with instrumental variables are known to perform poorly when the instruments are weak (e.g. Rothenberg (1984), Nelson and Startz (1990), Bound, Jaeger, and Baker (1995), and Dufour (1997)) or when the necessary rank condition fails altogether, see Phillips (1989). The literature has thus sought to develop methods to test for weak IVs (e.g. Hahn and Hausman (2002) and Stock and Yogo (2005)), as well devise inferential methods that are robust to weak instruments, such as Anderson and Rubin (1949), Anderson and Rubin (1950), Staiger and Stock (1997), Wang and Zivot (1998), Zivot, Startz, and Nelson (1998), Stock and Wright (2000), Dufour and Jasiak (2001), Chioda and Jansson (2005), Dufour and Taamouti (2005), Guggenberger and Smith (2005), Kleibergen (2005), Andrews, Moreira, and Stock (2006), and Andrews and Marmer (2008). An extensive review of this literature is well beyond the scope of this paper, but surveys are given by Stock, Wright, and Yogo (2002), Dufour (2003), Hahn and Hausman (2003b), and Andrews and Stock (2007).

We focus on the possibility that the variables employed as instruments fail to satisfy the exogeneity condition rather than the relevance condition. Interestingly, we show that the correlation between the instrumental variable and the endogenous regressor plays a key role, much like it does in the case of weak IV. In our case, however, the key condition is that this correlation be negative. The larger its magnitude, the tighter are the bounds.

In section 2 we lay out the more general setup and develop conditions that define the identified set. In section 3 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 and discuss the case of several imperfect IVs. Section 4 then discusses estimation and inference. Section 5 provides two empirical illustrations and section 6 concludes.

2 The Model

2.1 The Setup

The econometrician is interested in identifying the parameters of a regression with 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 econometrician. We assume a parametric functional form.

Assumption AP (parametric functional form):

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

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

For some results, we focus on the case where the function $m(\cdot, \cdot, \theta)$ is linear in parameters.

Assumption AL (linear model): $\theta = (\beta, \delta')'$ and

$$Y = X\beta + W\delta + U.$$

We further assume that there exists an observable vector of random variables Z which has dimension k_z. These will be imperfect instruments in the sense defined in our assumptions A3 and A4 below. 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 point-identified under the usual rank condition for $Z^{w\prime}(X, W)$, and model parameters 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 number of elements of W.

We assume the econometrician observes a random sample of (Y, X, W, Z, Z^w) drawn from population $(\Omega, \mathcal{F}, \mathbb{P})$. We use the subscript i to denote observations drawn from this population, and subscript j to denote individual elements of these vectors.

Assumption A1 (random sampling): $(y_i, x_i, w_i, z_i, z_i^w, \epsilon_i)'$, $i = 1, ..., n$ are iid realizations from $\mathbb{P}$.

We assume that the distribution $\mathbb{P}$ is such that the valid instruments Z^w are exogenous, in the sense that each component of Z^w is uncorrelated with the unobservable U, but X is endogenous.

Assumption A2 (Z^w exogenous, X endogenous): $\mathbb{E}(Z^{w\prime}U) = \mathbf{0}$, $\mathbb{E}(XU) \neq 0$.

The variable X may be correlated with U. This can happen, for example, when the value of X is chosen by an economic agent who observes U prior to choosing X. In this case, in the linear model identification and estimation of the model parameters θ typically relies on the use a vector of instrumental variables that are correlated with X but not with U. Such IV assumptions are often called into question in empirical work, and in some cases 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 A3 and A4. 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 notation σ_{ab} and ρ_{ab} throughout to denote the covariance and correlation, respectively, between any two random variables a, b.

Assumption A3 (same direction of correlation):

$$\rho_{xu}\rho_{z_j u} \geq 0 \ , \ j = 1, ...k_z.$$

Assumption A4 (instruments "less endogenous" than x):

$$|\rho_{xu}| \geq \left|\rho_{z_j u}\right| \ , \ j = 1, ...k_z.$$

Assumption A3 asserts that the endogenous regressor X and the IIV have the same direction of correlation with the error term. Assumption A4 then adds the condition that the IIV be less correlated with the error term than the endogenous regressor. Were we to replace A3 and A4 with the stronger assumption that $\rho_{z_j u} = 0$, then Z_j would be a valid instrument in the classical sense.

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

Assumption A5 (rank and order): $\text{rank}\big(\mathbb{E}\left[(Z, Z^w)'(Z, Z^w)\right]\big) = k_z + k_w$, $\text{rank}\big(\mathbb{E}\left[(X, Z^w)'(X, Z^w)\right]\big) = k_w + 1$, and $\text{rank}\big(\mathbb{E}\left[(Z, Z^w)'(X, W)\right]\big) = k_w + 1$, $k_z \geq 1$.

2.2 Identification

We now show how the modeling assumptions can be used to identify the parameters. The 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 embody information regarding the value of θ.

A useful starting point is to assume we know the relative correlations of the instrument Z_j and regressor X with the econometric error term. Define this correlation as

$$\lambda_j^* \equiv \rho_{z_j u}/\rho_{xu}, \tag{2.1}$$

Assumptions A3 and A4 are equivalent to the restriction that $\lambda_j^* \in [0,1]$. If λ_j^* were known, then it could be used to construct a weighted average of Z_j and X that is uncorrelated with the error term. To see this, define the function $V_j(\cdot)$ as

$$V_j(\lambda) \equiv \sigma_x Z_j - \lambda \sigma_{z_j} X. \tag{2.2}$$

By definition, $V_j(\lambda_j^*)$ is uncorrelated with U, and satisfies the moment condition $\mathbb{E}\left[(Y - m(X,W,\theta)) \cdot V_j(\lambda_j^*)\right] = 0$. Although λ_j^* is unknown, it offers a convenient parameterization of the relative correlations of $\rho_{z_j \epsilon}$ and $\rho_{x\epsilon}$, which will prove useful for characterization of the identified set. The hope is that by varying λ_j^* between 0 and 1 we will be able to bound the parameters of interest.

The implied restrictions of assumption A1 - A4 are:

$$\mathbb{E}\left[Z_j^{w\prime}(Y - m(X,W,\theta))\right] = \mathbf{0}, \tag{2.3a}$$

$$\mathbb{E}\left[(Y - m(X,W,\theta)) \cdot V_j(\lambda_j^*)\right] = 0, \; j = 1, ..., k_z, \tag{2.3b}$$

$$\lambda_j^* \in [0,1], \; j = 1, ..., k_z, \tag{2.3c}$$

where λ_j^* and $V_j(\cdot)$ are as defined in (2.1) and (2.2), respectively. Thus, the model parameters must satisfy the $k_w + k_z$ moment conditions (2.3a), (2.3b). If $\Lambda^* = \left(\lambda_1^*, \cdots, \lambda_{k_z}^*\right)$ were known and the standard rank conditions were satisfied, these moment conditions would identify θ locally under the parametric model (AP) and globally in the linear model (AL). These conditions could then be used for consistent estimation via GMM. However, since Λ^* is only known to belong to the unit cube in $\mathbb{R}^{k_z}$, θ will generally not be point identified. The identified set for θ is the set of parameter values that satisfy these restrictions.

Proposition 1 *Let (AP), A1 - A4 hold. Then $\Theta^* = \{\theta : conditions\ (2.3a)-(2.3c)\ hold\}$*

Corollary 1 *Suppose that AL holds. Then Θ^* is convex.*

The characterization of Θ^* above can be used to perform estimation and inference on the identified set

7

using a variety of methods recently considered for models comprised of moment equalities and inequalities, see for example Chernozhukov, Hong, and Tamer (2007) and the references therein. For example, if the researcher is willing to sign σ_{xu}, one way this can be done is by rewriting (2.3b) and (2.3c) as $2k_z$ moment inequalities obtained by setting $\lambda_j^* = 0$ and $\lambda_j^* = 1$ into (2.3b) to obtain lower and upper bounds on $\mathbb{E}\left[(Y - m(X, W, \theta)) \cdot V_j(\lambda_j^*)\right]$. This allows for estimation and inference on the entire parameter vector θ, though in some cases one may be interested in inference on the marginal effect of X.

3 The Linear Model

The characterization of the identified set, provided in Proposition 1, is not very informative. For example, it does not tell us if the identified set is bounded in all directions. In other words, Assumptions 3 and 4 might not be very informative. 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.

3.1 The Simple Linear Model with One Imperfect Instrument

Consider the simple linear model

$$Y = \alpha + X\beta + U, \tag{3.4}$$

where $\mathbb{E}[U] = 0$.

For notational ease, we define β^{OLS} and β_z^{IV} to be the probability limits of the standard OLS and IV, with Z as the instrument, estimators for β, respectively. That is

$$\beta^{OLS} \equiv \frac{\sigma_{xy}}{\sigma_x^2} = \beta + \frac{\sigma_{xu}}{\sigma_x^2}, \tag{3.5}$$

and

$$\beta_z^{IV} \equiv \frac{\sigma_{zy}}{\sigma_{xz}} = \beta + \frac{\sigma_{zu}}{\sigma_{xz}}. \tag{3.6}$$

$\widehat{\beta}^{OLS}$ and $\hat{\beta}_z^{IV}$ are taken to be their corresponding estimators.

The true value of the parameter can be bounded when assumptions (A1)-(A3) are imposed.

Lemma 1 *Let (A1), (A2), and (A3) hold. If $\sigma_{xz} < 0$ then β lies between β^{OLS} and β_z^{IV}. β^{OLS} is the upper bound if $\sigma_{xu} > 0$, while β^{OLS} is the lower bound if $\sigma_{xu} < 0$. If instead $\sigma_{xz} \geq 0$, then if $\sigma_{xu} > 0$, $\beta \leq \min\left\{\beta^{OLS}, \beta_z^{IV}\right\}$, while if $\sigma_{xu} < 0$, $\beta \geq \max\left\{\beta^{OLS}, \beta_z^{IV}\right\}$.*

Lemma 1 gives a simple characterization of bounds for β. Given assumptions A1-A3, β^{OLS} and β_z^{IV} provide either two-sided or one-sided bounds for β, depending on the correlation between X and Z, which is identified and can be consistently estimated.

If $\sigma_{xz} < 0$ assumption A3 can provide finite, hopefully economically helpful, bounds on the parameter of interest. Before exploring the gains from imposing assumption A4, we ask under what conditions can the direction of bias of the OLS estimator to be ascertained based on the IV estimator.

Lemma 2 *Let A1-A3 hold. If $\sigma_{xz} < 0$ or $\rho_{xz} > \lambda^*$ then*

$$sgn\left(\beta^{OLS} - \beta_z^{IV}\right) = sgn\left(\beta^{OLS} - \beta\right).$$

When $\sigma_{xz} < 0$ it follows that the bias of β^{OLS} has the same sign as $\beta^{OLS} - \beta_z^{IV}$, which is identified. Since λ^* is unknown however, the second condition generally will not be verifiable. A common intuition – that, even if the IV is not valid, 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 zero means the IV is "almost" valid and then the conditions of the lemma are likely to hold. If the conditions of the Lemma do not hold then an IIV will not even identify the direction of the bias in OLS.

Up to this point we did not impose Assumption A4, and were able to get two-sided bounds only if $\sigma_{xz} < 0$. We now ask what we get from further imposing Assumption A4. As we said above the hope is that by bounding λ^* between 0 and 1 we obtain sharper identification results than lemmas 1 and 2. The following result shows exactly what we get from also imposing Assumption A4.

Proposition 2 *Let A1-A4 hold. If $\sigma_{xz} < 0$, then*

$$B^* = \begin{array}{ll} \left[\beta_z^{IV}, \beta_{v(1)}^{IV}\right] & \text{if } \sigma_{xu} > 0 \\ \left[\beta_{v(1)}^{IV}, \beta_z^{IV}\right] & \text{if } \sigma_{xu} < 0 \end{array}$$

where

$$\beta_{v(1)}^{IV} \equiv \frac{\sigma_z\sigma_{xy} - \sigma_x\sigma_{zy}}{\sigma_x\left(\sigma_z\sigma_x - \sigma_{xz}\right)}. \tag{3.7}$$

9

If, on the other hand, $\sigma_{xz} > 0$, then

$$B^* = \begin{cases} \left(-\infty, \min\left(\beta_z^{IV}, \beta_{v(1)}^{IV}\right)\right] & \textit{if } \sigma_{xu} > 0 \\ \left[\max\left(\beta_{v(1)}^{IV}, \beta_z^{IV}\right), \infty\right) & \textit{if } \sigma_{xu} < 0. \end{cases}$$

These bounds are sharp.

The bound $\beta_{v(1)}^{IV}$ has a particularly simple characterization. It 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 (2.2), i.e. $V(1) \equiv \sigma_z X - \sigma_x Z$. $\beta_{v(1)}^{IV}$ can also be written as a weighted average of β^{OLS} and β_{IV}, since

$$\begin{aligned} \beta_{v(1)}^{IV} &= \frac{\sigma_{v(1)y}}{\sigma_{xv(1)}} = \frac{1}{\sigma_{xv(1)}}\left(\sigma_z \sigma_x^2 \beta^{OLS} - \sigma_x \sigma_{xz} \beta_z^{IV}\right) \\ &= \frac{\sigma_z \sigma_x^2}{\sigma_x\left(\sigma_z \sigma_x - \sigma_{xz}\right)}\beta^{OLS} - \frac{\sigma_x \sigma_{xz}}{\sigma_x\left(\sigma_z \sigma_x - \sigma_{xz}\right)}\beta_z^{IV}. \end{aligned}$$

In cases where $\sigma_{xz} > 0$, Proposition 2 implies that imposing assumption A4 only slightly improves upon the bound of Lemma 1 (in the sense that the bound is still one sided.) However, when $\sigma_{xz} < 0$, Proposition 2 improves, in a somewhat more meaningful way, on the bounds given in Lemma 1. So while Assumption A3 and A4 jointly bound λ^* between zero and one, they do not in general provide two-sided bounds for the parameter of interest, β. To see why, note that if $\lambda^* = \rho_{xz}$ then $\beta_{v(\lambda^*)}^{IV}$ is not defined, so even though λ is bounded β might not be bounded. If $\sigma_{xz} > 0$, we cannot rule out this case and β is not bounded from both sides. If $\sigma_{xz} < 0$ we can rule out $\lambda^* = \rho_{xz}$ since $0 \leq \lambda^* \leq 1$, and β is bounded from above and below.

While Assumption A4 does not give us two-sided bounds where they did not already exist, it does help us tighten the bounds. Corollary 2, applies to the case where $\sigma_{xz} < 0$, and where Proposition 2 achieves two-sided bounds for beta. This corollary characterizes the degree to which use of $\beta_{v(1)}^{IV}$ improves upon the bounds provided by Lemma 1, which are just β^{OLS} and β_z^{IV}. The implication is that the greater the magnitude of the correlation between x and z, the tighter the bound achieved using $\beta_{v(1)}^{IV}$ instead of β^{OLS}, with maximal improvement obtained when $\rho_{xz} = -1$, in which case the size of the bounds is halved.

Corollary 2 *If $\sigma_{xz} < 0$, then*

$$\frac{\beta_{v(1)}^{IV} - \beta_z^{IV}}{\beta^{OLS} - \beta_z^{IV}} = \frac{1}{1 - \rho_{xz}}.$$

Our results are consistent with the finding of Hahn and Hausman (2003a) that if $\mathbb{E}[ZU] \neq 0$, it is possible for β^{OLS} to be closer to β than β_z^{IV}. However, Proposition 2 further shows that this is possible even if the

10

instrument Z is less correlated with U than is the endogenous regressor X. Consider, for example, the case where $\sigma_{xu} < 0$ and $\sigma_{xz} < 0$, so that $\beta \in \left[\beta_{v(1)}^{IV}, \beta_z^{IV} \right]$, and $\beta^{OLS} \leq \beta_{v(1)}^{IV}$. $\beta_{v(1)}^{IV}$ is a feasible value for β, and $\beta_{v(1)}^{IV}$ is in fact closer to β^{OLS} than is β_z^{IV} because $\beta_{v(1)}^{IV} - \beta^{OLS} = \frac{\sigma_{xz}}{\sigma_x \sigma_z} \left(\beta_{v(1)}^{IV} - \beta_z^{IV} \right)$. In this case, if β is sufficiently close to $\beta_{v(1)}^{IV}$, β_z^{IV} "over-corrects" for β^{OLS}. Thus, even if the instrument is "less endogenous" than X, it is possible that β_z^{IV} lead the researcher further astray than β^{OLS}, and it may not be the case that the instrumental variables estimator offers an improvement over β^{OLS}.

3.2 Additional Regressors

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

$$Y = X\beta + W\delta + U, \tag{3.8}$$

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 such that satisfies assumptions A3 and A4 with respect to the endogenous regressor X. The additional regressors W may include exogenous and endogenous components, but we assume that A2 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 $\mathbb{E}\left[(X, W)' Z^w\right]$ had full rank, $\theta = \left(\beta, \delta'\right)'$ 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

The remainder of this section first generalizes our results from the simple linear model to derive bounds on β. We then construct the identification interval for each individual component of δ. The identification results are constructive and the bounds take the form of expressions that are trivially estimable with standard regression software. We focus on the case where there is only one IIV, as the identified set for β or any component of δ when there are multiple IIVs is the intersection of the intervals derived from using each of the multiple IIVs individually, see section 3.3.

3.2.1 Bounds on β

As shown in section 2.2, assumptions A3 and A4 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 θ is given by the usual IV formula:

$$\theta = \mathbb{E}\left[(V(\lambda^*), Z^w)'(X, W)\right]^{-1} \mathbb{E}\left[(V(\lambda^*), Z^w)'Y\right] \tag{3.9}$$

The residuals of an IV regression of X and Y on W using Z^w as the instruments are given by

$$\tilde{X} \equiv X - W\mathbb{E}[Z^{w\prime}W]^{-1}\mathbb{E}[Z^{w\prime}X], \tag{3.10a}$$

$$\tilde{Y} \equiv Y - W\mathbb{E}[Z^{w\prime}W]^{-1}\mathbb{E}[Z^{w\prime}Y]. \tag{3.10b}$$

Isolating the first component of θ in the IV regression above gives

$$\beta = \mathbb{E}\left[V(\lambda^*)\tilde{X}\right]^{-1}\mathbb{E}\left[V(\lambda^*)\tilde{Y}\right],$$

or equivalently that

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

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 $\tilde{X}$, etc.

Proposition 3 *Let A1, A3, and A4 hold, and assume that $\mathbb{E}[Z^{w\prime}W]$ is nonsingular and $\mathbb{E}[Z^{w\prime}U] = 0$. Assume WLOG that $\rho_{xu} \geq 0$. If $(\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z)\sigma_{z\tilde{x}} > 0$, then B^* is a closed interval, while if $(\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z)\sigma_{z\tilde{x}} \leq 0$, B^* is an open interval Specifically, if $(\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z)\sigma_{z\tilde{x}} > 0$ then B^* is given by*

$$B^* = \left[\beta_{v(1)}^{IV}, \beta_z^{IV}\right] \text{ if } \sigma_{z\tilde{x}} < 0,$$

$$B^* = \left[\beta_z^{IV}, \beta_{v(1)}^{IV}\right] \text{ if } \sigma_{z\tilde{x}} > 0.$$

If $(\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z)\sigma_{z\tilde{x}} \leq 0$ then

$$B^* = \left[\max\left\{\beta_z^{IV}, \beta_{v(1)}^{IV}\right\}, \infty\right) \text{ if } \sigma_{z\tilde{x}} < 0,$$

$$B^* = \left(-\infty, \min\left\{\beta_z^{IV}, \beta_{v(1)}^{IV}\right\}\right] \text{ if } \sigma_{z\tilde{x}} > 0,$$

where

$$\beta_z^{IV} = \frac{\sigma_{z\tilde{y}}}{\sigma_{z\tilde{x}}},$$

$$\beta_{v(1)}^{IV} = \frac{\sigma_{v(1)\tilde{y}}}{\sigma_{v(1)\tilde{x}}} = \frac{\sigma_{z\tilde{y}}\sigma_x - \sigma_{x\tilde{y}}\sigma_z}{\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z}.$$

Proposition 3 is a straightforward generalization of Proposition 2 from the simple linear regression to the multiple linear regression model. The bound β_z^{IV} corresponds to the probability limit of β from an IV regression of Y on (X, W) using (Z, Z^w) as instruments, and the bound $\beta_{v(1)}^{IV}$ corresponds to the probability limit of β from an IV regression of Y on (X, W) using $(V(1), Z^w)$ as instruments, where as before $V(1) \equiv \sigma_z X - \sigma_x Z$. Consistent estimation of B^* can thus be achieved by employing standard linear IV regression.

3.2.2 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 or not 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

$$\delta = \mathbb{E}\left[Z^{w\prime}W\right]^{-1}\mathbb{E}\left[Z^{w\prime}Y^*\right].$$

Then the j-th component of δ is given by

$$\delta_j = \mathbb{E}\left[Z_j^{w\prime}\tilde{W}_j\right]^{-1}\mathbb{E}\left[Z_j^{w\prime}\tilde{Y}^*\right], \tag{3.11}$$

where

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

$$\tilde{Y}^* \equiv Y^* - W_{-j}\mathbb{E}\left[Z_{-j}^{w\prime}W_{-j}\right]^{-1}\mathbb{E}\left[Z_{-j}^{w\prime}Y^*\right],$$

13

and subscript $-j$ denotes a vector with its j-th 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 implied values of δ_j over admissible value of β. The formal result, given in the following proposition, is that the bounds for δ_j are given by evaluating (3.11) at the endpoints of the identification region for β.

Proposition 4 *Let the conditions of Proposition 3 hold. Then the identification region for any δ_j, $j = 1, ..., k_w$ is given by the interval ranging from δ_{j0} to δ_{j1} where*

$$
\begin{aligned}
\delta_{j0} &\equiv \mathbb{E}\left[Z_j^{w\prime}\tilde{W}_j\right]^{-1} \mathbb{E}\left[Z_j^{w\prime}\tilde{Y}\right] - \mathbb{E}\left[Z_j^{w\prime}\tilde{W}_j\right]^{-1} \mathbb{E}\left[Z_j^{w\prime}\tilde{X}\right]\beta_L, \\
\delta_{j1} &\equiv \mathbb{E}\left[Z_j^{w\prime}\tilde{W}_j\right]^{-1} \mathbb{E}\left[Z_j^{w\prime}\tilde{Y}\right] - \mathbb{E}\left[Z_j^{w\prime}\tilde{W}_j\right]^{-1} \mathbb{E}\left[Z_j^{w\prime}\tilde{X}\right]\beta_U,
\end{aligned}
$$

and where β_L and β_U are the extreme points of B^*.

3.3 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 could help tighten the identification region for β, serve as a specification test of sort, and with additional assumptions helps us get two sided bounded where they were previously unavailable.

When there are multiple imperfect instruments that satisfy A3 and A4, i.e. $k_z > 1$ and $Z = (Z_1, ..., Z_{k_z})$, then Proposition 3 can be used to derive bounds on β (one or two-sided, depending on the sign of σ_{xz_r}) for each Z_r, $r = 1, ..., k_z$. For each r, denote the bounds implied by Proposition 3 as $B_r^* = [\beta_{l,r}, \beta_{u,r}]$ (where one of the two is possibly $\pm\infty$). In addition, let $D_{j,r}^* = \left[\delta_{l,r}^j, \delta_{u,r}^j\right]$ denote the bounds on δ_j for each $j = 1, ..., k_w$ given by Proposition 4. It follows that the identification region for β is the intersection of all of the intervals B_r^*, and the identification region for each δ_j is the intersection over $r = 1, ..., k_z$ of the intervals $D_{j,r}^*$.

Proposition 5 *Assume AL and A1-A4. Then the identified set for β is $B^* = \left[\max_r \beta_{l,r}, \min_r \beta_{u,r}\right]$, and the identification region for δ_j, each $j = 1, ..., k_w$ is $D_j^* = \left[\max_r \delta_{l,r}^j, \min_r \delta_{u,r}^j\right]$. These bounds are sharp.*

This proposition is a result of applying Propositions 3 and 4 to each of the instruments Z_j, $j = 1, ..., J$. Furthermore, this exploits all the identifying power of the multiple IIVs, in the sense that there is no

additional identifying power from imposing A3 and A4 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 are a direct result of their characterization in Proposition 4 as functions of β. Note, that in principle B^* and the intervals D_j^* can be empty, which may be used to serve as a specification test.

A potential drawback is that when $(\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z)\sigma_{z\tilde{x}} < 0$, Proposition 5 will only provide a one-sided bound on β. 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 such cases, bringing this knowledge to bear can achieve tighter identification. In particular, in the case where the bounds of Proposition 5 are one-sided, such an assumption can provide two-sided bounds.

Let there be two instruments, Z_1 and Z_2, each satisfying assumptions A3 and A4, and $\sigma_{xz_1} > 0$, $\sigma_{xz_2} > 0$. Define the following weighted average of Z_1 and Z_2:

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

where $\gamma \in [0, 1]$. For the purpose of exposition in what follows we assume that $\sigma_{xu} > 0$, which is with out loss of generality, as long as we continue to assume A3.

Proposition 6 *Let A1, A2, and A3 hold for both Z_1 and Z_2, and in addition assume that there exists $\gamma^* \in [0, 1]$ such that $\sigma_{\omega(\gamma^*)u} \geq 0$ and $\left(\sigma_{\omega(\gamma^*)\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_{\omega(\gamma^*)}\right)\sigma_{\omega(\gamma^*)\tilde{x}} \geq 0$. This yields the following bounds for β.*

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

The proposition simply says that if there exists a γ^* such that the conditions on $\omega(\gamma^*)$ are met, then we can apply Proposition 5 to obtain a two-sided bound. The following Lemma provides more basic conditions that guarantee the required assumption, and allow one to test it with the data in the case where the additional regressors W are exogenous.

Lemma 3 *Let $W = Z^w$ and let A3 hold, for both Z_1 and Z_2. Then the following statements are equivalent: (i) there exists $\gamma^* \in [0, 1]$ such that $\sigma_{\omega(\gamma^*)u} \geq 0$ and $\sigma_{\omega(\gamma^*)\tilde{x}} < 0$; (ii) there exists a known $\gamma^* \in [0, 1]$, such that $\frac{\sigma_{\tilde{x}z_1}}{\sigma_{\tilde{x}z_2}} > \frac{\gamma^*}{1 - \gamma^*} > \frac{\sigma_{z_1 u}}{\sigma_{z_2 u}}$; (iii) $\sigma_{z_1\tilde{y}}\sigma_{\tilde{x}z_2} < \sigma_{z_2\tilde{y}}\sigma_{\tilde{x}z_1}$. Furthermore, a sufficient condition for the inequality $\left(\sigma_{\omega(\gamma^*)\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_{\omega(\gamma^*)}\right)\sigma_{\omega(\gamma^*)\tilde{x}} \geq 0$ is that the partial correlation between X and $\omega(\gamma^*)$ controlling for W, is negative.*

The Lemma shows the assumptions on the unobserved covariances that guarantee that the weighted average $\omega(\gamma^*)$ both satisfies A3 and the conditions in Proposition 2. More importantly it provides conditions to test the assumption in the data.

Note that while Proposition 6 provides two-sided bounds, it relies on knowing γ^*. Lemma 3 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 only reveals 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. Another natural choice is $\gamma^* = \sigma_{z_1} / (\sigma_{z_1} + \sigma_{z_2})$.

3.4 Relation to Manski and Pepper

Our approach for handling the problem of correlation between instruments and econometric error terms is in part motivated by the work of Manski and Pepper (2000) on monotone instrumental variables, or MIVs. They focus on models where the outcome variable is bounded between zero and one, and examine nonparametric models. We, on the other hand, derive most of our results for the linear regression model. While their results are nonparametric, an earlier version of their paper, Manski and Pepper (1998), available through the NBER, devotes a section to linear models as well. However, even in the context of the linear model, our modeling assumptions neither nest nor are nested by theirs. Here, we examine the implications of our assumptions relative to those of Manski and Pepper (1998) (henceforth MP).

In section 4 of their NBER paper, MP invoke the following assumptions:

Assumption MP1 (Linear Response Model)[1]:

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

$$E(U) = 0.$$

Assumption MP2 (MIV) for all pairs z_2, z_1 such that $z_2 \geq z_1$:

$$\forall x \in \text{supp}(X),\ E(\alpha + \beta x + U | Z = z_2) \geq E(\alpha + \beta x + U | Z = z_1)$$

The second assumption is equivalent to assuming $\mathbb{E}(U | Z = z_2) \geq \mathbb{E}(U | Z = z_1)$ for all z_1, z_2 pairs where

---

[1] Manski and Pepper in fact do not use an intercept and do not restrict U to have mean zero. This is only a difference in notation, as including the intercept makes the mean zero restriction on U meaningless.

$z_2 \geq z_1$. Under these restrictions, Manski and Pepper derive the following sharp bounds for β:

Proposition (Manski and Pepper (1998) Proposition 2) Assume (MP1) and (MP2). Then

$$\beta \leq \frac{\mathbb{E}\left[Y|Z=z_2\right] - \mathbb{E}\left[Y|Z=z_1\right]}{\mathbb{E}\left[X|Z=z_2\right] - \mathbb{E}\left[X|Z=z_1\right]} \text{ if } \mathbb{E}\left[X|Z=z_2\right] - \mathbb{E}\left[X|Z=z_1\right] > 0, \text{ and}$$

$$\beta \geq \frac{\mathbb{E}\left[Y|Z=z_2\right] - \mathbb{E}\left[Y|Z=z_1\right]}{\mathbb{E}\left[X|Z=z_2\right] - \mathbb{E}\left[X|Z=z_1\right]} \text{ if } \mathbb{E}\left[X|Z=z_2\right] - \mathbb{E}\left[X|Z=z_1\right] < 0.$$

These bounds are sharp.

In this paper, we mostly focus on a linear model, which is equivalent to MP1. Instead of MP2, we assume A3 and A4, and then assuming that the sign of ρ_{xu} is known, it is WLOG to assume that $\rho_{xu} \geq \rho_{zu} \geq 0$. The inequality $\rho_{zu} \geq 0$ is implied by MP1 and MP2, although their restriction of monotonicity of $\mathbb{E}\left(U|Z=z\right)$ is stronger. The case where the two assumptions coincide is that where $\mathbb{E}\left(U|Z=z\right)$ is linear in z, but otherwise $\mathbb{E}\left(U|Z=z\right)$ is a sharper restriction. The MP assumption is a restriction on the distribution of U conditional on Z for each value of z, while our restriction is a restriction on the average relation between U and Z over their entire support.

If $\mathbb{E}\left(X|Z=z\right)$ is monotone then MP get only one sided bounds, just like we would if $\sigma_{xz} > 0$. In order to get two sided-bounds they require $\mathbb{E}\left(X|Z=z\right)$ to be non-monotone, while we require that $\sigma_{xz} < 0$.

4   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. Consistent estimators of our bounds can thus be computed using standard regression software.

Regarding statistical inference, there are a variety of methods from the recent literature that appear applicable in the present context, including Pakes, Porter, Ho, and Ishii (2005), Chernozhukov, Hong, and Tamer (2007), Andrews and Guggenberger (2007), and the references cited therein. In Section 5, we employ a variant of the inferential procedure proposed by Chernozhukov, Lee, and Rosen (2008). Their method is specifically designed for settings where there is interval identification, where the identified set is the intersection of many intervals, which is precisely the case here.

Specifically, we use the method of Chernozhukov, Lee, and Rosen (2008) to construct 95% confidence intervals for each of the interval identified parameters in our regressions. In each case, we use a sample analog estimator for the identified set of the form of $[\max\left\{L_{n1}, ..., L_{nR}\right\}, \min\left\{U_{n1}, ..., U_{nS}\right\}]$, where each L_{nr}, U_{ns}

are consistent estimators 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 estimated. Those estimates whose standard errors are high require a larger adjustment than those whose standard errors are low. To be precise, the confidence intervals are given by $CI_\alpha = [L_\alpha, U_\alpha]$, where

$$
\begin{aligned}
L_\alpha &\equiv \max\left\{ L_{n1} - \hat{s}_{l1}^{-1} q^l\left(\alpha_n\right), ..., L_{nR} - \hat{s}_{lR} q^l\left(\alpha_n\right) \right\}, \\
U_\alpha &\equiv \min\left\{ U_{n1} + \hat{s}_{u1} q^u\left(\alpha_n\right), ..., U_{nS} + \hat{s}_{uR} q^u\left(\alpha_n\right) \right\},
\end{aligned}
$$

where $\hat{s}_{lr}, \hat{s}_{ur}$ are the standard errors of L_{nr}, U_{nr}, respectively. $q^l\left(\alpha\right)$ and $q^u\left(\alpha\right)$ are the α-quantiles of the maxima of mean zero multivariate normal random vectors with variance matrices equal to the estimated joint variance covariance matrices $[L_{n1}, ..., L_{nR}]$, $[U_{n1}, ..., U_{nS}]$, respectively, and are computed via simulation. The value of α_n is chosen to provide the desired nominal coverage for the object of interest, the identified set or the parameter of interest. For inference on the identified set we use $\alpha_n = 0.975$, for coverage for the parameter we use 0.95, and for uniform asymptotic coverage we use an appropriately defined intermediate value that depends on the size of the estimated identified set, in similar spirit to Imbens and Manski (2004) and Stoye (2007). A by-product is that this also provides a specification test: if the lower and upper bounds of the confidence set computed with $\alpha_n = 0.975$ cross, then the model is rejected at the 0.05 level. For further details we refer to Chernozhukov, Lee, and Rosen (2008).

5 Applications

In this section we provide two applications, both motivated by recent work in the Industrial Organization literature. First, we examine the estimation of a Cobb-Douglas production function. Next, we examine the estimation of a differentiated-products demand system.

5.1 Estimation of a Production Function

5.1.1 The model

Consider the following Cobb-Douglas production function:

$$Q_{it} = e^{\alpha} L_{it}^{\beta_l} K_{it}^{\beta_k} R_{it}^{\beta_r} e^{u_{it}},$$

where Q_{it} is the output of firm i at time t, L_{it} is labor, K_{it} is capital, R_{it} is R&D capital, u_{it} is an error term, α, β_l, β_k, β_r are parameters to be estimated. The error term, u_{it}, includes technology or management differences, measurement errors and variation in external factors.

Taking logs we obtain

$$y_{it} = \alpha + \beta_l l_{it} + \beta_k k_{it} + \beta_r r_{it} + u_{it},$$

where: $y_{it} \equiv \log(Q_{it})$, $l_{it} \equiv log(L_{it})$, $k_{it} \equiv log(K_{it})$, $r_{it} \equiv log(R_{it})$.

A major concern in the literature is that the variable input, l_{it}, is chosen after the error term is observed and is therefore correlated with it. Capital variables, both physical and R&D, on the other hand, are assumed to be either exogenous or pre-determined (conditional on a firm-fixed effect). Assume that $u_{it} = \alpha_i + \omega_{it} + \varepsilon_{it}$, where α_i and ω_{it} are "transmitted" (i.e., impact the choice of l_{it}), while ε_{it} is white noise that is uncorrelated with any of the variables. Different methods have been proposed in the literature to consistently estimate the parameters of the model. Each model imposes different restrictions on the error term. We discuss some of these methods below as we present estimates. For a more detailed discussion see, for example, Griliches and Mairesse (1998), Ackerberg, Caves, and Frazer (2006) and Bond and Soderbom (2005).

5.1.2 Results

To demonstrate our method we use the data from Griliches and Mairesse (1998) (see Hall (1990) for further information on the data). The sample includes U.S. R&D performing firms listed on the major stock exchanges during 1973-1988. The data are at 5 year intervals. Overall, the sample includes 2971 observations from 4 different periods for 1400 firms. Our analysis focuses on the firms that existed for at least 2 (consecutive) periods. There are 1502 such observations with 820 firms. The balanced panel, of firms that were present for the whole period, consists of 856 observations and 214 firms. Table 1 provides some summary statistics.

We provide estimates of the slope coefficients, using different assumptions. In all cases we focus on the

set of firms that were present in at least two cross sections. We ignore the issue of sample selection.[2] Various corrections for sample selection can easily be included in the analysis. The results are presented in Table 2.

The first column presents results using OLS. These results are biased if firm-specific effects, α_i, are present and correlated with the choice of labor. A standard correction is to difference the equation in order to get rid of the firm-specific effects. The second column in the table presents the results from first differencing. The change in both the labor and capital coefficients is significant, suggesting that indeed the firm-specific effects are present and correlated with both the fixed and variable inputs.

One of the problems with the first difference model is that it does not allow the transmitted productivity shock to vary over time. The results in the third column, are based on an Olley and Pakes (1996) specification.[3] This estimator sets α_i to 0, but allows for a transmitted time-varying productivity shock, ω_{it}, and lets this shock vary over time according to a first-order Markov process. Note that both the Olley-Pakes estimator and the fixed effect estimator allow for a non-transmitted time varying shock, ε_{it}. The first order Markov process does not allow for more than a one period persistence in the transmitted productivity shocks. Where this impacts the Olley-Pakes analysis is in the inversion from investment to productivity shock. If the two firms are observed to invest the same amounts in the same period (controlling for differences in stocks), then they are inferred to have the same productivity shock. In reality, however, the firms' expected shocks could be different, which would be the case if the process was more persistent, implying that even though their investment is the same their current productivity shock is different. Indeed, the estimates are much closer to the OLS estimates than to those of the first-difference estimator, which probably suggests that the first-order Markov process might not be persistent enough to capture the process in the transmitted productivity shock.

The dynamic panel literature (e.g., Arellano and Bond (1991) and Blundell and Bond (1998)) offers an alternative way to estimate the parameters. The results in columns 4 and 5 present the estimates from these methods. The results from the Arellano and Bond estimator are very noisy and the point estimates are not reasonable. This is a common problem and is usually attributed to weak instruments. The "system" estimator of Blundell and Bond aims to address this problem and indeed the results in column 5 are more reasonable. The estimators in columns 4 and 5 are not consistent if the instruments are not valid, which

---

[2]In this data set selection does not appear to impact the estimates, despite the patterns observed in Table 1. OLS results using all the observations, just those that are present 2 or more periods, or the balanced panel, are essentially identical. This might not be the case for other data sets.

[3]Specifically, the parameters are estimated in two steps. In the first step, the labor cofficient is estimated by regressing the log of output on the log of employment, and a second-order polynominal in the capital, R&D capital and investment. In a second step the coefficients on capital and R&D capital are estimated exploiting that the unexpected innovation in productivity is mean independent of lagged capital. See Ackerberg, Caves, and Frazer (2006).

would be the case if ω_{it} is autocorrelated. Indeed a test of over-identification of the instruments fails.

The last two columns in the table present estimates from our procedure. The estimated equation is in first difference. Our concern is that labor is not strictly exogenous, and therefore correlated with the transformed error term. Both economic reasoning and the previous results suggest that the correlation is positive. We explore two IIVs. In column 6 we use lagged R&D investment and in column 7 we use lagged capital investment. It is reasonable to assume that lagged investment, conditional on all the other variables and a fixed effect, is positively correlated with the productivity shock. The logic is similar to that motivating the Olley-Pakes estimator. Furthermore, both types of investment are negatively correlated with employment (conditional on the other variables). Regressing the first difference in log of employment on lagged R&D, controlling for the other variables, we get a negative coefficient with a t-value of -3.4. The same regression using lagged investment yields a negative coefficient with a t-stat of -5.9. Thus, both the variables satisfy the conditions of Proposition 2 and thus are valid IIVs.

The results in columns 6 and 7 are quite reasonable. The results on column 6, using lagged R&D as the IIV suggest that the labor coefficient is between 0.50 and 0.71. In itself this is not a very useful bound since it includes the OLS estimate and is very close to the first-difference estimates. The bounds for both the capital and R&D coefficients are slightly more informative. For example, the bounds on the capital coefficients rule out most of the point estimates suggested by alternative methods. The results using lagged capital investment yield similar results, although the bounds are tighter. For example, the point estimates of the labor coefficients do not include all but one of the previous estimates, the Blundell-Bond system estimate. The fact that the two IIV yield somewhat similar regions is assuring. The bounds can be further tightened by taking the intersection of the regions. The results are presented in the last column. Now the bounds are both tight and do not include any of the previous estimates.

The table reports standard errors in parentheses, for columns 6-8 we report 95% confidence intervals (CI) using the methods described in Section 4: the top number reports the CI for the identified set, the bottom number is the CI for the parameter and the middle number report the CI with uniform asymptotic coverage.

5.2 Differentiated-Products Demand

In this section, we apply our method to the estimation of the demand for differentiated products. We use the Logit model. Assume that the indirect utility for consumer i for 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 project j in market t. We assume that when making their purchases, each consumer chooses exactly one good, and also has the option to choose an "outside" good, i.e. not to buy any of the products. We normalize the mean value of the outside good to be zero so that $u_{i0t} = \epsilon_{i0t}$. Furthermore, ϵ_{ijt} is assumed to be distributed iid extreme value, from which it follows that each product j has market share s_{jt} in market t, where

$$s_{jt} = \frac{\exp\left(p_{jt}\beta + w'_{jt}\Gamma + \xi_{jt}\right)}{1 + \sum\limits_{k=1}^{J} \exp\left(p_{kt}\beta + w_{kt}\Gamma + \xi_{kt}\right)},$$

and

$$\log\left(s_{jt}\right) - \log\left(s_{0t}\right) = p_{jt}\beta + w'_{jt}\Gamma + \xi_{jt}. \tag{5.1}$$

If price p_{jt} and market characteristics w_{jt} are 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 a 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 20 quarters, and for this application focus attention on the top 25 brands (in terms of market share), and the San Francisco and Boston markets. The key variables observed for each product, market, quarter combination are market share, price[4], quantity sold during promotional periods, and brand-level advertising. For additional information on the data source and the details of the RTE cereal industry, we refer the reader to Nevo (2000) and Nevo (2001). Table 3 provides a descriptive summary of the observed 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 (e.g., Hausman, Leonard, and Zona (1994), Hausman (1997), and Nevo (2001)). The idea is 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

---

[4] In order to normalize per portion, we take price to be total revenue divided by quantity.

assumption has been challenged (e.g., 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 cereal are found to be healthy the preferences for these cereals could vary. Any of these stories would render the IV, and the implied estimates, not valid. There is evidence that despite these theoretical concerns the IV are valid (e.g., Nevo (2001)), nevertheless, some concerns linger over the validity of the estimates.

We use prices in other cities as an imperfect IVs. The examples above suggest that usually we worry about positive association between demand errors in different markets. Thus, it is natural to assume that correlation of prices in other cities with the error term is positive. Unfortunately, in our data it is also the case that prices in other markets are positively correlated with price. Therefore, these IIV will only yield one sided bounds. However, we can exploit the fact that we have multiple cities to generate a valid IIV, using the results in Section 3.3. 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 Appendix B 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 regions.

In the sample the correlations $\rho_{pz_1} = 0.81$ and $\rho_{pz_2} = 0.48$ satisfy the first of these assumptions. Furthermore, Lemma 3 allows 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. We explore two options. We explore $\gamma = 0.5$ and $\gamma = \sigma_{z_1}/(\sigma_{z_1} + \sigma_{z_2})$. For both we can verify that the correlation between price and $\omega(\gamma) = \gamma Z_2 - (1-\gamma)Z_1$ is negative.

The results are presented in Table 4. 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 city 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 one in absolute value. Once we use the average regional price as an IV the price coefficient becomes more negative, as expected.

The next four columns use weighted differences between the average regional price and the price in the other city as IIV. The difference between the columns is in the weight used. Columns (3) and (4) use a weight

23

of 0.5, while the last two columns use $\gamma = \sigma_{z_1}/(\sigma_{z_1} + \sigma_{z2})$.Columns (3) and (5) only impose assumption A3, on the differenced IV, while columns (4) and (6) also impose assumption A4.

The results yield a fairly consistent picture. If we do not impose assumption A4 then the price coefficient is between -4 and -8.7 (with a confidence interval of approximately -2.3 and -11.4). On the other hand, if we impose assumption A4 the price coefficient is between -6 and -8.7 (with a confidence interval of approximately -4 and -11.25). In all cases the OLS point estimate is outside the confidence interval. When we impose assumption A4, the IV estimate is very close to the boundary of the confidence interval.

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 increase the price elasticities by another factor of two. Generally, the results suggest that the price elasticties 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 (e.g., Nevo (2001)). Our results suggest that the estimates of PCM using the standard IV assumption might too high. Another use of demand estimates is for simulation of the effects of mergers (e.g., Hausman, Leonard, and Zona (1994); and Nevo (2000)). The results in Table 4 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   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 the endogenous regressor.

Under our assumptions, we first derive an abstract characterization of the identified set for all parameters, and then focus primarily on identification of the slope parameter on the endogenous regressor. Consistent estimates of these bounds can be computed by standard OLS and linear 2SLS regressions under the usual rank conditions (A5). We found that the bounds that can be obtained for the slope parameter, and in particular whether they form an open or closed interval, depend 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 A3 and A4 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, as discussed in section 3.4, 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 could have significant identifying power in a nonparametric model.

References

ACKERBERG, D. A., K. CAVES, AND G. FRAZER (2006): "Structural Identification of Production Functions," working paper, UCLA.

ANDERSON, T., AND H. RUBIN (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46–63.

——— (1950): "The Asymptotic Properties of Estimates of the Parameters of a Single Equation system in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 21(4), 570–582.

ANDREWS, D. W., AND J. H. STOCK (2007): "Inference with Weak Instruments," in *Advances in Economics and Econometrics: Theory and Applications, Econometric Society Ninth World Congress Proceedings Vol. III*, ed. by R. Blundell, W. Newey, and T. Persson. Cambridge University Press.

ANDREWS, D. W. K., AND P. GUGGENBERGER (2007): "Validity of Subsampling and Plug-In Asymptotic Inference for Parameters Defined by Moment Inequalities," working paper, UCLA.

ANDREWS, D. W. K., AND V. MARMER (2008): "Exactly Distribution-Free Inference in Instrumental Variables Regression with Possibly Weak Instruments," *Journal of Econometrics*, 142, 183–200.

ANDREWS, D. W. K., M. J. MOREIRA, AND J. H. STOCK (2006): "Optimal Two-sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74, 715–752.

ARELLANO, M., AND S. BOND (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297.

BLUNDELL, R., AND S. BOND (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115–143.

BOND, S., AND M. SODERBOM (2005): "Adjustment Costs and the Identification of Cobb-Douglas Production Functions," working paper No. 05/04, Institute for Fiscal Studies.

BONTEMPS, C., T. MAGNAC, AND E. MAURIN (2006): "Set Identified Linear Models," working paper, IDEI.

BOUND, J., D. A. JAEGER, AND R. M. BAKER (1995): "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, 90(430), 443–450.

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

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5).

CHERNOZHUKOV, V., S. LEE, AND A. ROSEN (2008): "Intersection Bounds, Estimation and Inference," working paper, MIT and CEMMAP.

CHIODA, L., AND M. JANSSON (2005): "Optimal Conditional Inference for Instrumental Variables Regression," working paper, U.C. Berkeley.

CONLEY, T., C. HANSEN, AND P. E. ROSSI (2006): "Plausibly Exogenous," working paper, Chicago Graduate School of Business.

DUFOUR, J.-M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65(6), 1365–1387.

——— (2003): "Identification, Weak Istruments, and Statistical Inference in Econometrics," *Canadian Journal of Economics*, 36(4), 767–808.

DUFOUR, J.-M., AND J. JASIAK (2001): "Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors," *International Economic Review*, 42(3), 815–843.

DUFOUR, J.-M., AND M. TAAMOUTI (2005): "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," *Econometrica*, 73(4), 1351–1365.

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

GRILICHES, Z., AND J. MAIRESSE (1998): "Production Functions: The Search for Identification," in *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*, ed. by S. Strom, pp. 169–203. Cambridge University Press.

GUGGENBERGER, P., AND R. J. SMITH (2005): "Generalized Empirical Likelihood Estimators and Tests Under Partial, Weak, and Strong Identification," *Econometric Theory*, 21, 667–709.

HAHN, J., AND J. HAUSMAN (2002): "A New Specification Test for the Validity of Instrumental Variables," *Econometrica*, 70(1), 163–189.

——— (2003a): "IV Estimation with Valid and Invalid Instruments," working paper, MIT.

——— (2003b): "Weak Instruments: Diagnosis and Cures in Empirical Econometrics," *American Economic Review*, 93(2), 118–125.

HALL, B. (1990): "The Manufacturing Sector Master File: 1959-1987," NBER working paper.

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

HAUSMAN, J. A., G. K. LEONARD, AND J. D. ZONA (1994): "Competitive Analysis with Differentiated Products," *Annales D'Economie et de Statistique*, (34), 159–180.

IMBENS, G., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

KLEIBERGEN, F. (2005): "Testing Parameters in GMM Without Assuming that They Are Identified," *Econometrica*, 74(4), 1103–1123.

KLEPPER, S., AND E. E. LEAMER (1984): "Consistent Sets of Estimates for Regressions with Errors in All Variables," *Econometrica*, 52(1), 163–184.

LEAMER, E. E. (1981): "Is It a Demand Curve, or Is It a Supply Curve? Partial Identification through Inequality Constraints," *Review of Economics and Statistics*, 63(3), 319–327.

MANSKI, C. F., AND J. PEPPER (1998): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *NBER working paper*.

MANSKI, C. F., AND J. V. PEPPER (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68(4), 997–1010.

NELSON, C. R., AND R. STARTZ (1990): "Some Further Results on the Small Sample Properties of the Instrumental Variable Estimator," *Econometrica*, 58(4), 967–976.

NEVO, A. (2000): "Mergers with Differentiated Products: the Case of the Ready-to-Eat Cereal Industry," *Rand Journal of Economics*, 31(3), 395–421.

——— (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69(2), 307–342.

OLLEY, S., AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Industry," *Econometrica*, 64, 1263–1295.

PAKES, A., J. PORTER, K. HO, AND J. ISHII (2005): "The Method of Moments with Inequality Constraints," working paper, Harvard University.

28

PHILLIPS, P. C. (1989): "Partially Identified Econometric Models," *Econometric Theory*, 5, 181–240.

ROTHENBERG, T. (1984): "Approximating the Distributions of Econometric Estimators and Test Statistics," in *The Handbook of Econometrics*, ed. by Z. Griliches, and M. Intriligator, vol. 2, pp. 881–935. Elsevier, New York: North Holland.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557–586.

STOCK, J. H., AND J. H. WRIGHT (2000): "GMM with Weak Identification," *Econometrica*, 68(5), 1055–1096.

STOCK, J. H., J. H. WRIGHT, AND M. YOGO (2002): "A Survery of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20(4), 518–529.

STOCK, J. H., AND M. YOGO (2005): "Testing For Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas J. Rothenberg*, ed. by D. W. Andrews, and J. H. Stock. Cambridge University Press, Cambridge, U.K.

STOYE, J. (2007): "More on Confidence Regions for Partially Identified Parameters," working paper, New York University.

WANG, J., AND E. ZIVOT (1998): "Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments," *Econometrica*, 66(6), 1389–1440.

ZIVOT, E., R. STARTZ, AND C. R. NELSON (1998): "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, 39(4), 1119–1144.

Appendix A: Proofs

Proposition 1

Proof. For any fixed value of θ, the left hand side of the system (2.3a), (2.3b) can be recovered from the data, under random sampling assumption A1. Under A2-A4, any value of θ that satisfies each of these restrictions as well as (2.3c) is a feasible value and all such values of θ are observationally equivalent. ■

Corollary 1

Proof. Let $\theta^* = (\beta^*, \Gamma^{*\prime})$ and $\theta^{**} = (\beta^{**}, \Gamma^{**\prime})$ such that both θ^*, θ^{**} are elements of Θ^*. Let $\alpha \in [0,1]$ and define $\theta_\alpha = \alpha \theta^* + (1 - \alpha) \theta^{**}$. Because the left hand sides of conditions (2.3a) and (2.3b) are linear in θ, $\theta_\alpha \in \Theta^*$. ∎

Lemma 1

Proof. The result follows directly from inspection of the expressions for β^{OLS} and β_z^{IV} given by (3.5) and (3.6), respectively. If $\sigma_{xu} > 0$ (< 0), then β^{OLS} is an upper (lower) bound for β. If $\sigma_{zu}/\sigma_{xz} > 0$ (< 0), then β_z^{IV} is an upper (lower) bound for β. ∎

Lemma 2

Proof. Using expressions (3.5) and (3.6),

$$\beta^{OLS} - \beta_z^{IV} = \frac{\sigma_{xu}}{\sigma_x^2} - \frac{\sigma_{zu}}{\sigma_{xz}}, \quad \beta^{OLS} - \beta = \frac{\sigma_{xu}}{\sigma_x^2}.$$

To show $sgn\left(\beta^{OLS} - \beta_z^{IV}\right) = sgn\left(\beta^{OLS} - \beta\right)$ it is equivalent to show that $\left(\beta^{OLS} - \beta_z^{IV}\right)\left(\beta^{OLS} - \beta\right) > 0$, i.e.

$$\frac{\sigma_{xu}}{\sigma_x^2}\left(\frac{\sigma_{xu}\sigma_{xz} - \sigma_x^2\sigma_{zu}}{\sigma_x^2\sigma_{xz}}\right) > 0.$$

Case 1 $\sigma_{xz} < 0$. *By A3 σ_{xu} and σ_{zu} are either both non-positive or non-negative. Therefore $\left(\sigma_{xu}\sigma_{xz} - \sigma_x^2\sigma_{zu}\right)/\sigma_{xz}$ has the same sign as σ_{xu}, and $\left(\beta^{OLS} - \beta_z^{IV}\right)\left(\beta^{OLS} - \beta\right) > 0$.*

Case 2 $\lambda^* < \rho_{xz}$, *and $\sigma_{xz} > 0$. Isolating σ_{zu} in the definition of λ^* (equation (2.1)) gives $\sigma_{zu} = \lambda^*\sigma_{xu}\sigma_z/\sigma_x$. Substituting into $\beta^{OLS} - \beta_z^{IV}$, it follows that*

$$\beta^{OLS} - \beta_z^{IV} = \frac{\sigma_{xu}}{\sigma_x}\left(\frac{1}{\sigma_x} - \frac{\lambda^*\sigma_z}{\sigma_{xz}}\right),$$

which has the same sign as σ_{xu} (given that $\sigma_{xz} > 0$) if $\sigma_{xz} - \sigma_x\sigma_z\lambda^* > 0$, or equivalently if $\lambda^* < \rho_{xz}$.

∎

Proposition 2

Proof. Suppose that $\rho_{xu} > 0$. Then A4 gives

$$\rho_{xu} \geq \rho_{zu} \geq 0$$

$$\Rightarrow \sigma_z \sigma_{xu} \geq \sigma_x \sigma_{zu}$$

$$\Leftrightarrow \sigma_z \left(\sigma_{xy} - \sigma_x^2 \beta \right) \geq \sigma_x \left(\sigma_{zy} - \sigma_{xz} \beta \right)$$

$$\Leftrightarrow \beta_{v(1)}^{IV} \geq \beta,$$

where as in (3.7),

$$
\begin{aligned}
\beta_{v(1)}^{IV} &= \frac{\sigma_z \sigma_{xy} - \sigma_x \sigma_{zy}}{\sigma_x \left(\sigma_z \sigma_x - \sigma_{xz} \right)} \\
&= \frac{\sigma_z \sigma_x}{\left(\sigma_z \sigma_x - \sigma_{xz} \right)} \beta^{OLS} - \frac{\sigma_{xz}}{\left(\sigma_z \sigma_x - \sigma_{xz} \right)} \beta_z^{IV}
\end{aligned}
$$

which makes use of the assumption that x and z are not perfectly correlated (implied by A5), as well as the Cauchy-Schwartz Inequality (i.e. $\sigma_z \sigma_x \geq \sigma_{xz}$). If instead $\rho_{xu} < 0$ then $\sigma_z \sigma_{xu} \leq \sigma_x \sigma_{zu}$ and $\beta_{v(1)}^{IV} \leq \beta$. Note that $\beta_{v(1)}^{IV} = \gamma \beta^{OLS} + (1 - \gamma) \beta_z^{IV}$, where $\gamma \equiv \sigma_z \sigma_x / \left(\sigma_z \sigma_x - \sigma_{xz} \right) = 1/ \left(1 - \rho_{xz} \right)$.

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 $\sigma_{xu} > 0$, Lemma 1 implies that $\beta \in \left[\beta_z^{IV}, \beta^{OLS} \right]$. 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, i.e. $\beta \in \left[\beta_z^{IV}, \beta_{v(1)}^{IV} \right]$. If instead $\sigma_{xu} < 0$, then similar logic leads to $\beta \in \left[\beta_{v(1)}^{IV}, \beta_z^{IV} \right]$.

Now suppose $\sigma_{xz} > 0$. Then if $\sigma_{xu} > 0$, $\beta_{v(1)}^{IV} \geq \beta$, and Lemma 1 gives $\beta \leq \min \left\{ \beta^{OLS}, \beta_z^{IV} \right\}$. An immediate implication is that $\beta \leq \min \left\{ \beta^{OLS}, \beta_z^{IV}, \beta_{v(1)}^{IV} \right\}$, but it turns out that β^{OLS} is redundant in that $\min \left\{ \beta^{OLS}, \beta_z^{IV}, \beta_{v(1)}^{IV} \right\} = \min \left\{ \beta_z^{IV}, \beta_{v(1)}^{IV} \right\}$. This is because if $\beta^{OLS} < \beta_z^{IV}$, then $\beta_{v(1)}^{IV} \leq \beta^{OLS}$, while clearly if instead $\beta^{OLS} < \beta_z^{IV}$, then $\min \left\{ \beta^{OLS}, \beta_z^{IV} \right\} = \beta_z^{IV}$. The claim that $\beta^{OLS} < \beta_z^{IV} \Rightarrow \beta_{v(1)}^{IV} \leq \beta^{OLS}$ holds because $\sigma_{xz} > 0$ implies that $1 - \gamma = -\sigma_{xz} / \left(\sigma_z \sigma_x - \sigma_{xz} \right) < 0$, 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 $\beta \geq \max \left\{ \beta_z^{IV}, \beta_{v(1)}^{IV} \right\}$. The sharpness of the bounds follows as an implication of Proposition 5, which covers the case of multiple imperfect instruments and is proven below. ∎

Corollary 2

Proof . Using the above notation,

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

Therefore

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

■

Proposition 3

Proof. A3, A4 and $\rho_{xu} \geq 0$ give

$$\rho_{xu} \geq \rho_{zu} \geq 0$$

$$\Leftrightarrow \sigma_{xu}\sigma_z \geq \sigma_{zu}\sigma_x \geq 0$$

$$\Leftrightarrow \sigma_{x\tilde{y}}\sigma_z - \sigma_{x\tilde{x}}\sigma_z\beta \geq \sigma_{z\tilde{y}}\sigma_x - \sigma_{z\tilde{x}}\sigma_x\beta \geq 0,$$

where the third line uses $\tilde{Y} = \beta\tilde{X} + U$, which is shown below in Lemma 4 below. The first inequality 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. The second inequality 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 gives the conclusion of the proposition. ■

Lemma 4 *Let the conditions of Proposition 3 hold. Then $\tilde{Y} = \beta\tilde{X} + U$.*

Proof. Subtracting $W E [Z^{w\prime}W]^{-1} E [Z^{w\prime}Y]$ from both sides of (3.8) gives

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

■

Proposition 4

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

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

where $\tilde{Y} = Y - W_{-j}E\left[Z_{-j}^{w\prime}W_{-j}\right]^{-1}E\left[Z_{-j}^{w\prime}Y\right]$. Now plugging this into (3.11) gives

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

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 5

To establish sharpness of the bounds in this proposition, we make use of the following corollary to Proposition 3.

Corollary 3 *If $\sigma_{xz} < 0$, then any $\lambda^* \in [0,1]$ is feasible. If $\sigma_{xz} > 0$, then if $\beta_{v(1)}^{IV} > \beta_z^{IV}$, $\lambda^* \in (\rho_{xz}, 1]$, while if $\sigma_{xz} > 0$ and $\beta_{v(1)}^{IV} < \beta_z^{IV}$, then $\lambda^* \in [0, \rho_{xz})$.*

Proof . First, write λ^* as a function of β:

$$
\lambda^* = \frac{\sigma_{zu}\sigma_x}{\sigma_{xu}\sigma_z} = \frac{\sigma_x\sigma_{\tilde{y}z} - \beta\sigma_x\sigma_{\tilde{x}z}}{\sigma_z\sigma_{x\tilde{y}} - \beta\sigma_z\sigma_{\tilde{x}x}}, \tag{6.1}
$$

where $\tilde{Y}$ and $\tilde{X}$ are as defined in Lemma 4, so that the above equality follows from $U = \tilde{Y} - \tilde{X}\beta$. The derivative of λ^* with respect to β is

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

from which we see that λ^* monotone in β in the direction of $\sigma_{\tilde{x}x}\sigma_{\tilde{y}z} - \sigma_{\tilde{x}z}\sigma_{x\tilde{y}}$.

First suppose the bounds are two-sided, in which case $(\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z)\sigma_{z\tilde{x}} > 0$. Then the bounds on β from Proposition 3 are $\dfrac{\sigma_{z\tilde{y}}}{\sigma_{z\tilde{x}}}$ and $\dfrac{\sigma_{z\tilde{y}}\sigma_x - \sigma_{x\tilde{y}}\sigma_z}{\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z}$. These correspond to values of λ^* of 0 and 1, respectively, and as β lies between these values we have that $\lambda^* \in [0,1]$. If instead $(\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z)\sigma_{z\tilde{x}} \le 0$, the bounds are one-sided, and we can again trace out the feasible values of λ^* as a function of β. We now further break this case into the following subcases:

Case 3 $\sigma_{z\tilde{x}} < 0$. Then $\beta \in \left[\max\left\{\dfrac{\sigma_{z\tilde{y}}}{\sigma_{z\tilde{x}}}, \dfrac{\sigma_{z\tilde{y}}\sigma_x - \sigma_{x\tilde{y}}\sigma_z}{\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z}\right\}, \infty\right)$ and $\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z \ge 0$.

1. If $\dfrac{\sigma_{z\tilde{y}}\sigma_x - \sigma_{x\tilde{y}}\sigma_z}{\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z} > \dfrac{\sigma_{z\tilde{y}}}{\sigma_{z\tilde{x}}}$, then $\sigma_{z\tilde{y}}\sigma_{x\tilde{x}} - \sigma_{z\tilde{x}}\sigma_{x\tilde{y}} < 0$, $\beta \in \left[\dfrac{\sigma_{z\tilde{y}}\sigma_x - \sigma_{x\tilde{y}}\sigma_z}{\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z}, \infty\right)$, and $\dfrac{d\lambda^*}{d\beta} < 0$. Thus the upper bound on λ^* is 1, when $\beta = \dfrac{\sigma_{z\tilde{y}}\sigma_x - \sigma_{x\tilde{y}}\sigma_z}{\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z}$. Using L'Hospital's rule, if follows that the limit of λ^* as $\beta \to \infty$ is $\lambda^* = \dfrac{\sigma_x\sigma_{\tilde{x}z}}{\sigma_z\sigma_{\tilde{x}x}}$, which is ≥ 0 from the inequalities $\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z \ge 0$ (since $(\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z)\sigma_{z\tilde{x}} \le 0$) and $\sigma_{z\tilde{x}} < 0$.

2. If $\dfrac{\sigma_{z\tilde{y}}\sigma_x - \sigma_{x\tilde{y}}\sigma_z}{\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z} \le \dfrac{\sigma_{z\tilde{y}}}{\sigma_{z\tilde{x}}}$, then $\sigma_{z\tilde{y}}\sigma_{x\tilde{x}} - \sigma_{z\tilde{x}}\sigma_{x\tilde{y}} \ge 0$, $\beta \in \left[\dfrac{\sigma_{z\tilde{y}}}{\sigma_{z\tilde{x}}}, \infty\right)$, and $\dfrac{d\lambda^*}{d\beta} \ge 0$. Thus the lower bound on λ^* is 0, when $\beta = \dfrac{\sigma_{z\tilde{y}}}{\sigma_{z\tilde{x}}}$. The upper bound on λ^* is the limit of expression (6.1) as $\beta \to \infty$, which is $\lambda^* = \dfrac{\sigma_x\sigma_{\tilde{x}z}}{\sigma_z\sigma_{\tilde{x}x}} \ge 0$, again since $\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z \ge 0$.

Case 4 $\sigma_{z\tilde{x}} > 0$. Then $\beta \in \left(-\infty, \min\left\{\dfrac{\sigma_{z\tilde{y}}}{\sigma_{z\tilde{x}}}, \dfrac{\sigma_{z\tilde{y}}\sigma_x - \sigma_{x\tilde{y}}\sigma_z}{\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z}\right\}\right]$ and $\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z \le 0$. Following the same logic as when $\sigma_{z\tilde{x}} < 0$, we have that when $\dfrac{\sigma_{z\tilde{y}}}{\sigma_{z\tilde{x}}} < \dfrac{\sigma_{z\tilde{y}}\sigma_x - \sigma_{x\tilde{y}}\sigma_z}{\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z}$, $\dfrac{d\lambda^*}{d\beta} < 0$, and λ^* as a function of β takes values on the interval $\left[0, \dfrac{\sigma_x\sigma_{\tilde{x}z}}{\sigma_z\sigma_{\tilde{x}x}}\right)$. When instead $\dfrac{\sigma_{z\tilde{y}}}{\sigma_{z\tilde{x}}} \ge \dfrac{\sigma_{z\tilde{y}}\sigma_x - \sigma_{x\tilde{y}}\sigma_z}{\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z}$, $\lambda^* \in \left(\dfrac{\sigma_x\sigma_{\tilde{x}z}}{\sigma_z\sigma_{\tilde{x}x}}, 1\right]$. In both cases, $\dfrac{\sigma_x\sigma_{\tilde{x}z}}{\sigma_z\sigma_{\tilde{x}x}} \in [0,1]$ follows from the inequalities $\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z \le 0$ and $\sigma_{z\tilde{x}} > 0$.

∎

Having established Corollary 3, we now provide the proof of Proposition 5.

Proof . $\beta \in B^*$ as defined in the statement of the proposition follows immediately from Proposition 2 in the simple linear model, and more generally from Proposition 3. This is because Proposition 3 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, i.e. that any value of $\beta \in B^*$ is feasible. To this end, we make use of Corollary 3, which gives upper and lower bounds on λ_j^* for each j. Consider any $\beta \in B^*$. Then by (2.3b) there exists for each j a λ_j^* contained in the bounds given by Corollary 3 such that

$$\beta = \frac{\sigma_{\tilde{y}z_j}\sigma_x - \sigma_{x\tilde{y}}\sigma_{z_j}\lambda_j^*}{\sigma_x\sigma_{\tilde{x}z_j} - \sigma_{\tilde{x}x}\sigma_{z_j}\lambda_j^*}, \tag{6.2}$$

34

where $\tilde{Y}$ and $\tilde{X}$ are as defined in (3.10a) and (3.10b). Define $U = \tilde{Y} - \tilde{X}\beta$. Then straightforward algebra shows that (6.2) $\Rightarrow \rho_{zu}/\rho_{xu} = \lambda_j^*$, for any j. Since by Corollary 3 $\lambda_j^* \in [0,1]$, it follows that A3 and A4 hold. Thus the joint distribution of (W, Y, X, Z, U) satisfies our modeling assumptions, verifying that the conjectured value of $\beta \in B^*$ is indeed feasible. $\blacksquare$

Lemma 3

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

$$\gamma^* \geq \frac{\sigma_{z_1 u}}{\sigma_{z_1 u} + \sigma_{z_2 u}} \geq 0, \tag{6.3}$$

and

$$\gamma^* < \frac{\sigma_{\tilde{x}z_1}}{\sigma_{\tilde{x}z_1} + \sigma_{\tilde{x}z_2}} \leq 1, \tag{6.4}$$

from which it follows that

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

Substituting $\sigma_{z_j u} = \sigma_{z_j \tilde{y}} - \sigma_{\tilde{x}z_j}\beta$ 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 (iii). From the inequalities $\sigma_{z_j u} \geq 0$ and $\sigma_{\tilde{x}z_j} > 0$ it also follows that (iii) $\Rightarrow$ (i). The equivalence of (i) and (ii) follows since (6.3) holds if and only if $\frac{\sigma_{\tilde{x}z_1}}{\sigma_{\tilde{x}z_2}} > \frac{\gamma^*}{1 - \gamma^*}$, and (6.4) holds if and only if $\frac{\gamma^*}{1 - \gamma^*} > \frac{\sigma_{z_1 u}}{\sigma_{z_2 u}}$. 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 (i)) 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 $\sigma_{\omega(\gamma^*)\tilde{x}} < 0 \Rightarrow \left(\sigma_{\omega(\gamma^*)\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_{\omega(\gamma^*)}\right) < 0$, and consequently that $\left(\sigma_{\omega(\gamma^*)\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_{\omega(\gamma^*)}\right)\sigma_{\omega(\gamma^*)\tilde{x}} > 0$. $\blacksquare$

Proposition 6

Proof . The proof follows by application of Proposition 5 with $\omega(\gamma^*)$ as an IIV for X. $\blacksquare$

Appendix B: A model of IIV in demand

In this appendix, we provide a model of demand in which the IIV assumptions hold. To make the example as transparent as possible, we consider linear demand for a single product sold by a monopolist. This

basic model captures the essential features that make our assumptions plausible. The example extends to differentiated product markets, where the residual demand curve for any single differentiated product plays the role of the demand for the homogenous product in our example here.

Consider the demand for the good in four distinct geographic markets, which we denote $A1$, $A2$, $B1$, and $B2$. The markets $A1$ and $A2$ are in region A and are geographically close to one another, while $B1$ and $B2$ are in region B. But region A and B are far away from each other.

Suppose demand in each market $m \in \{A1, A2, B1, B2\}$ at time t has the form

$$Q_{mt} = \alpha_m + \beta p_{mt} + \epsilon_{mt},$$

where Q_{mt} is quantity demanded, p_{mt} is price, and ϵ_{mt} is an unobservable (to the researcher) demand shifter, e.g. unobserved promotional activities. Quantity and price are observed for each market. Assume for simplicity that for each market, $Var(\epsilon_{mt}) = \sigma_\varepsilon$.

A monopolist maximizing one period profits, with constant marginal cost and facing this demand curve will set prices

$$p_{mt} = \frac{mc_{mt}}{2} - \frac{\alpha_m + \epsilon_{mt}}{2\beta},$$

where mc_{mt} denotes marginal cost. If $\beta < 0$ (downward sloping demand) then $Cov(p_{mt}, \epsilon_{mt}) > 0$, and least squares estimation of the demand equation will yield biased estimates.

Assume that $Cov(mc_{mt}, \epsilon_{nt}) = 0$ for all m and n, and that $Var(mc_{mt}) = \sigma_{mc}$. The covariance between the price in city m and the the demand error in city n is

$$Cov(p_{mt}, \epsilon_{nt}) = -\frac{1}{2\beta} Cov(\epsilon_{mt}, \epsilon_{nt}),$$

If $Cov(\epsilon_{mt}, \epsilon_{nt}) = 0$ then p_{mt} is a valid instrument for p_{nt} in estimating the demand equation. However, if $Cov(\epsilon_{mt}, \epsilon_{nt}) > 0$ then the price in the other city is not a valid IV. If $Corr(\epsilon_{mt}, \epsilon_{nt}) < 1$ then $Corr(p_{mt}, \epsilon_{mt}) > Corr(p_{mt}, \epsilon_{nt}) > 0$ and our assumptions A3 and A4 are satisfied.

In order to examine whether we get one- or two-sided bounds we examine the covariance of prices in any two markets m and n given by

$$cov(p_{mt}, p_{nt}) = \frac{1}{4}\left(cov(mc_{mt}, mc_{nt}) + \frac{1}{\beta^2} cov(\epsilon_{mt}, \epsilon_{nt})\right).$$

As long as $cov\left(mc_{mt}, mc_{nt}\right) \geq 0$ then $cov\left(p_{mt}, p_{nt}\right) > 0$, and using p_{mt} as an IIV will yield only a one-sided bound. To get a two-sided bound we exploit regional differences.

Assume that marginal costs are more correlated within a region than across regions. So, for example, $cov\left(mc_{A1,t}, mc_{A2,t}\right) > cov\left(mc_{A1,t}, mc_{B1,t}\right) \geq 0$. On the other hand assume that demand shocks are more correlated across similar markets than within regions, so that $cov\left(\epsilon_{A1,t}, \epsilon_{B1,t}\right) > cov\left(\epsilon_{A1,t}, \epsilon_{A2,t}\right) \geq 0$,[5] then the conditions of Proposition 6 are satisfied.

---

[5] As we noted in section 3.3, this is actually a stronger assumption than we need.

Table 1: Summary Statistics Production Data

	All firms		present in 2 periods		balanced panel	
Variable	mean	st dev	mean	st dev	mean	st dev
Log sales	5.67	1.96	6.20	1.86	6.91	1.84
Log employment	1.26	1.78	1.71	1.69	2.41	1.62
Log capital	4.47	2.22	5.09	2.11	5.92	2.06
Log R&D capital	3.40	2.03	4.00	1.98	4.89	1.93
Log capital investment	2.67	2.17	3.19	2.14	4.07	2.06
Log R&D	1.79	2.05	2.32	2.07	3.22	1.99
$N =$	2971		1502		856	

Source: date from Griliches and Mairesse (1998) (see Hall (1990) for further information on the data)

Table 2: Production Function Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	1st diff	OP	AB	BB	IIV1	IIV2	both IIV
Log employment	0.54	0.74	0.58	0.14	0.63	[0.50 0.71]	[0.62 0.72]	[0.62 0.71]
	(0.01)	(0.02)	(0.01)	(0.41)	(0.09)	(0.07 0.77)	(0.38 0.77)	(0.34 0.78)
						(0.10 0.77)	(0.38 0.77)	(0.35 0.78)
						(0.14 0.76)	(0.41 0.77)	(0.37 0.77)
Log of capital	0.39	0.12	0.34	0.80	0.31	[0.13 0.23]	[0.12 0.17]	[0.13 0.17]
	(0.01)	(0.02)	(0.01)	(0.31)	(0.06)	(0.09 0.44)	(0.09 0.29)	(0.09 0.47)
						(0.09 0.43)	(0.09 0.29)	(0.09 0.46)
						(0.10 0.41)	(0.10 0.27)	(0.09 0.44)
Log of R&D Capital	0.05	0.04	0.06	-0.29	0.01	[0.05 0.08]	[0.04 0.06]	[0.05 0.06]
	(0.01)	(0.02)	(0.01)	(0.23)	(0.05)	(0.01 0.16)	(0.01 0.15)	(0.01 0.18)
						(0.01 0.16)	(0.01 0.12)	(0.01 0.18)
						(0.02 0.15)	(0.02 0.11)	(0.02 0.17)

The dependent variable in all columns is the log of sales. The sample includes 1502 observations and includes only firms that were present for 2 or more periods. The regressions also include a dummy variable for the computer industry (SIC 357), and annual dummy variables interacted with the computer dummy. The column labeled OP presents results from an Olley-Pakes specification, AB from a Arellano and Bond type estimation (first differences instrumented with lagged values), and BB a Blundell-Bond "system" estimator (where we add moments that instrument for a levels equation using differences). Standard errors are reported in parentheses. In columns 6-8 we report 95% confidence intervals (CI) using the method described in Section 4: the top number reports the CI for the identified set, the bottom number is the CI for the parameter and the middle number report the CI with uniform asymptotic coverage.

Table 3: Summary Statistics Demand Data

Variable	Mean	Median	Standard Deviation	Min	Max	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%	N/A	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 4: Logit Demand Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	IIV	opt IIV	IIV1	opt IIV1
price	-2.21	-4.08	[-8.69 -4.08]	[-8.69 -5.99]	[-8.55 -4.08]	[-8.55 -5.94]
	(0.72)	(0.89)	(-11.44 -2.32)	(-11.44 -4.04)	(-11.25 -2.32)	(-11.25 -4.00)
			(-11.00 -2.61)	(-11.00 -4.35)	(-10.82 -2.61)	(-10.82 -4.32)
			(-11.00 -2.61)	(-11.00 -4.35)	(-10.82 -2.61)	(-10.82 -4.32)
advertising/10	0.31	0.30	[0.28 0.30]	[0.28 0.29]	[0.28 0.30]	[0.28 0.29]
	(0.01)	(0.01)	(0.16 0.41)	(0.16 0.41)	(0.16 0.41)	(0.16 0.41)
			(0.16 0.41)	(0.16 0.41)	(0.16 0.41)	(0.16 0.41)
			(0.18 0.39)	(0.18 0.39)	(0.18 0.39)	(0.18 0.39)
first stage t-stat:		42.42	-19.49	-33.40	-19.99	-34.01

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 20 quarters (one of the brands was only introduced after 5 quarters). The regressions also include brand fixed effects, a dummy variable for San Francisco and 20 quarterly time dummy variables. Column (2) reports results using the average price in the region as an IV. Columns (3) reports results using the difference between the average price in the other city and the average price in the region as an IIV. Column (4) optimally combines this IIV with the OLS to sharpen the bounds. Columns (5) and (6) repeat but put a weight, $\sigma_{z_1}/(\sigma_{z_1} + \sigma_{z_2})$, on the price in the other city, and $\sigma_{z_2}/(\sigma_{z_1} + \sigma_{z_2})$ on the average price in the region. Standard errors are reported in parentheses. In columns 3-6 we report 95% confidence intervals (CI) using the method described in Section 4: the top number reports the CI for the identified set, the bottom number is the CI for the parameter and the middle number report the CI with uniform asymptotic coverage.