Recovering Price Elasticities Using Instrumental Variables

Bryan S. Graham, UC - Berkeley & NBER

February 18, 2023

The invention, and first application, of the method of instrumental variables is due to Wright (1928, Appendix B). Wright was undertaking a tariff analysis and, in support of this, sought to estimate the price elasticities of demand and supply for flaxseed (used to make linseed oil, which has a number of industrial uses). His treatment was remarkably elegant for its time. Today the method of instrumental variables (IV) is widely-used in economics, across the other social sciences, in biostatistics and beyond. The method is both a major success story and an export of the field of econometrics (see Angrist & Krueger, 2001). Research on IV continues; the 2021 Nobel Prize in Economics was awarded to Joshua Angrist and Guido Imbens, for their research on the method of instrumental variables.

Consider the following log-linear schedules for quantity supplied and demanded as a function of price:

$$\ln Q^{S}\left(p,w_{1}\right)=\alpha^{S}+\epsilon^{S}\ln p+w_{1}'\gamma^{S}+U^{S} \tag{1}$$

$$\ln Q^{D}\left(p,w_{2}\right)=\alpha^{D}+\epsilon^{D}\ln p+w_{2}'\gamma^{D}+U^{D}, \tag{2}$$

with the $J_1 \times 1$ vector W_{1i} being a set of observed supply curve "shifters" and the the $J_2 \times 1$ vector W_{2i} being a set of observed demand curve shifters; both for market $i \in \{1, \ldots, N\}$. The term "curve shifter" is due to Wright (1928, p. 312) and I will adopt it here as well. Below we will attempt to connect Wright's notion of a curve shifter with the modern definition of an instrumental variable.

I will call U_i^S and U_i^D supply and demand "shocks". Like W_{1i} and W_{2i} they generate cross-market (intercept) heterogeneity in the (log) supply and demand schedules; unlike W_{1i} and W_{2i} they are unobserved by the econometrician.

It is helpful to be a bit more precise about the relationship between the observed supply shifters, W_{1i}, and the unobserved supply shock, U_i^S. Let $\tilde{U}_i^S$ be a composite variable capturing the effects of *all* drivers of quantity supplied in market i other than price. This gives a supply schedule of

$$\ln Q^S(p) = \alpha^S + \epsilon^S \ln p + \tilde{U}^S. \tag{3}$$

We can then define γ^S vis-a-vis the mean regression decomposition

$$\tilde{U}_i^S = W_{1i}'\gamma^S + U_i^S, \quad \mathbb{E}\left[\left. U_i^S \right| W_{1i}\right] = 0. \tag{4}$$

Implicit in (4) is a linear functional form assumption on $\mathbb{E}\left[\left.\tilde{U}_i^S\right| W_{1i}\right]$, but this is not the feature I wish to emphasize here. Instead I wish to emphasize that $W_{1i}'\gamma^S$ is simply the projection of all non-price drivers of quantity supplied onto the vector of observed supply shifters W_{1i}. Our unobserved supply "shock", U_i^S, is simply the associated projection error. As long as we don't wish to use the γ^S vector to predict the effects of policy manipulations of W_{1i}, this is without loss of generality (functional form assumptions aside).

We will work with an analogous decomposition for demand:

$$\ln Q^S(p) = \alpha^S + \epsilon^S \ln p + \tilde{U}^S \tag{5}$$

$$\tilde{U}_i^D = W_{2i}'\gamma^D + U_i^D, \quad \mathbb{E}\left[\left. U_i^D \right| W_{2i}\right] = 0. \tag{6}$$

For what follows a bit of additional notation will be helpful. Let $\beta^S = \left(\alpha^S, \epsilon^S, \gamma^{S\prime}\right)$ with $\dim\left(\beta^S\right) = 1 + 1 + J_1$ and $\beta^D = \left(\alpha^D, \epsilon^D, \gamma^{D\prime}\right)$ with $\dim\left(\beta^D\right) = 1 + 1 + J_2$. Here $\beta^S \in \mathbb{B}^S \subset \mathbb{R}^{1+1+J_1}$ is the vector of unknown parameters indexing the family of supply schedules allowed by (1) and $\beta^D \in \mathbb{B}^D \subset \mathbb{R}^{1+1+J_2}$ is the corresponding vector of unknown parameters indexing the family of demand schedules allowed by (2). A maintained assumption of our analysis is that there exist some $\beta_0^S \in \mathbb{B}^S$ and some $\beta_0^D \in \mathbb{B}^D$ that accurately characterize the supply and demand schedules governing price and quantity determination in the population of markets from which our sample was drawn. Our goal is to use the sample data to recover (noisy) estimates of these "true" or population values. Finally let $W_i = (W_{1i}', W_{2i}')'$ and $U_i = \left(U_i^S, U_i^S\right)'$.

We will make the following assumptions:

1. The supply and demand schedule coefficients, respectively β_0^S and β_0^D, are the same across all markets in the sampled "population" of markets.

2. For each randomly sampled market we observe the vector $Z_i = (P_i, Q_i, W_i')'$ where we assume that, using (1) and (2) above,

$$Q_i = \ln Q^S(P_i, W_{1i}) = \ln Q^D(P_i, W_{2i})$$

(i.e., observed prices and quantities are precisely those which "clear the market").

It is helpful to make a few additional observations:

1. Variation in P_i and Q_i across markets is driven by variation in (W_i', U_i');
2. Random sampling of markets implies that $(W_i', U_i') \overset{iid}{\sim} F_{W,U}$ for $i = 1, \ldots, N$; variation in W_i and U_i is not modeled; these variables are *exogenous* to our system, whereas P_i and Q_i are *endogenous* variables (i.e., determined by our model):
   (a) $(W_i', U_i') \overset{iid}{\sim} F_{W,U}$ + Form of Supply/Demand Schedules + Market Clearing Assumption $\rightarrow$ Observed Data, Z_i;
   (b) Note that in this set-up Z_i and Z_j for $i \neq j$ are also iid.
3. Our model is *semiparametric*.
   (a) The family of supply and demand schedules take a specific *parametric* form;
   (b) $F_{W,U}$ is *not* assumed to belong to a specific parametric family of distributions, it is left *nonparametric*.

Instrumental variables

We will make the following pair of assumptions of the supply and demand shifters.

1. (Instrument Relevance) At least one element of γ_0^S and one element of γ_0^D is non-zero.
2. (Exclusion Restriction/Conditional Independence). We assume that

$$\mathbb{E}\left[U^S \middle| W_1, W_2\right] = \mathbb{E}\left[U^S \middle| W_1\right] = 0. \tag{7}$$

$$\mathbb{E}\left[U^D \middle| W_1, W_2\right] = \mathbb{E}\left[U^D \middle| W_2\right] = 0. \tag{8}$$

The first assumption ensures that at least some elements of W_1 and W_2 actually shift, respectively, the supply and demand schedules. This guarantees that W_1 and W_2 will be predictive of price, P_i.

The second assumption says that our demand curve shifters, W_2, conditional W_1, do not move the supply curve. Similarly the supply curve shifters, W_1, conditional on W_2, do not generate variation in demand. Angrist et al. (2000), in their analysis of the Fulton Fish

Market, for example, assume that W_1 is an indicator for whether it is stormy at sea. This generates variation in the supply schedule for fish since its more dangerous to fish when ocean conditions are poor. They assume that conditions at sea do not affect the demand for fish. To make this assumption more credible, they set W_2 equal to an indicator variable for whether it is rainy on shore. It may be that on- and off-shore weather are correlated. By conditioning on on shore weather it is reasonable to assume no additional effect of off shore weather on demand.

The presentation will focus on estimating the parameters indexing the demand schedule. Adapting the reasoning which follows to estimate the parameters of supply schedule is left as an exercise.

At the market-clearing price and quantity pair $(P_i, Q_i)'$ we have

$$\ln Q_i = \alpha_0^D + \epsilon_0^D \ln P_i + W_{2i}'\gamma_0^D + U_i^D.$$

Recall that the "0" subscript on a parameter denotes the (true) population parameter. We can therefore write the unobserved demand shock U_i^D as a function of the observed data and the unknown parameter β_S^D

$$\begin{aligned} U_i^D = \rho\left(\beta_0^D, Z_i\right) &\stackrel{def}{\equiv} \ln Q_i - \alpha_0^D - \epsilon_0^D \ln P_i - W_{2i}'\gamma_0^D. \\ &= Y_i - R_i'\beta_0^D \end{aligned}$$

with $R_i \stackrel{def}{\equiv} (1, \ln P_i, W_{2i}')'$ and $Y_i \stackrel{def}{\equiv} \ln Q_i$.

Using the exclusion restriction (4) we have the *conditional moment restriction*

$$\mathbb{E}\left[\left. U_i^D \right| W_i\right] = \mathbb{E}\left[\left. \rho\left(\beta_0^D, Z_i\right) \right| W_i\right] = 0. \tag{9}$$

By the Law of Iterated Expectations we therefore have the unconditional moment restrictions

$$\begin{aligned} \mathbb{E}\left[\rho\left(\beta_0^D, Z_i\right) g\left(W_i\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[\left. \rho\left(\beta_0^D, Z_i\right) g\left(W_i\right) \right| W_i\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\left. \rho\left(\beta_0^D, Z_i\right) \right| W_i\right] g\left(W_i\right)\right] \\ &= \mathbb{E}\left[0 \cdot g\left(W_i\right)\right] \\ &= 0, \end{aligned}$$

for any $L \times 1$ vector of functions $g\left(W_i\right)$. Note that the "0" after the last equality is an $L \times 1$ vector of zeros.

Analogy principle

Condition (9) says that when $\beta^D = \beta_0^D$ (i.e., is equal to its true or population value), then $\rho\left(\beta_0^D, Z_i\right)$ is conditionally mean zero given W_i. This means $\rho\left(\beta_0^D, Z_i\right)$ will be uncorrelated with *any* function of W_i. Hence under (9) it will be true that the following $L \times 1$ vector of sample covariances are close to zero

$$\frac{1}{N}\sum_{i=1}^{N}\rho\left(\beta_0^D, Z_i\right) g\left(W_i\right) \approx 0,$$

when N, the number of sampled markets, is large enough.

We can use this observation to motivate an approach to estimation. We will start with the *just identified* case where $L = 1 + 1 + J_2$ (since $g\left(W_i\right)$ will include a constant, this generally corresponds to assuming that $J_1 = \dim\left(W_{1i}\right) = 1$ (i.e., we have a single supply shifter). If $\frac{1}{N}\sum_{i=1}^{N}\rho\left(\beta_0^D, Z_i\right) g\left(W_i\right) \approx 0$ when $\beta^D = \beta_0^D$, it stands to reason that choosing $\hat{\beta}^D$ to solve

$$\frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right)\rho\left(\hat{\beta}^D, Z_i\right) = 0$$

should result in a parameter estimate close to the truth ($\hat{\beta}^D \approx \beta_0^D$). This is the *analogy principle* in action Goldberger (1991). We have a condition which holds in the population at the true parameter. We choose our parameter estimate to impose that condition in the sample in hand. The above system is just-identified because we have $L = 1+1+J_2$ equations and $\dim\left(\beta^D\right) = 1 + 1 + J_2$ unknowns.

Using our definition $\rho\left(\beta^D, Z_i\right)$ we have

$$\frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right)\left(Y_i - R_i'\hat{\beta}^D\right) = 0.$$

This gives, after some manipulation,

$$\left[\frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right) R_i'\right]\hat{\beta}^D = \left[\frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right) Y_i\right]$$

and hence that

$$\hat{\beta}^D = \left[\frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right) R_i'\right]^{-1} \times \left[\frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right) Y_i\right].$$

Large sample theory

We have

$$
\begin{aligned}
\hat{\beta}^D &= \left[\frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right) R_i'\right]^{-1} \times \left[\frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right) Y_i\right] \\
&= \left[\frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right) R_i'\right]^{-1} \times \left[\frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right) \left(R_i'\beta_0^D + U_i^D\right)\right] \\
&= \left[\frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right) R_i'\right]^{-1} \times \left[\frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right) R_i'\right] \beta_0^D \\
&\quad + \left[\frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right) R_i'\right]^{-1} \times \left[\frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right) \rho\left(\beta_0^D, Z_i\right)\right] \\
&= \beta_0^D + \left[\frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right) R_i'\right]^{-1} \times \left[\frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right) \rho\left(\beta_0^D, Z_i\right)\right].
\end{aligned}
$$

This gives, after some re-arrangement,

$$
\sqrt{N}\left(\hat{\beta}^D - \beta_0^D\right) = \left[\frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right) R_i'\right]^{-1} \times \left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} g\left(W_i\right) \rho\left(\beta_0^D, Z_i\right)\right].
$$

A Law of Large Numbers (LLN) gives

$$
\frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right) R_i' \xrightarrow{p} \mathbb{E}\left[g\left(W_i\right) R_i'\right] \stackrel{def}{\equiv} \Gamma_0,
$$

and a Central Limit Theorem (CLT) gives

$$
\frac{1}{\sqrt{N}}\sum_{i=1}^{N} g\left(W_i\right) \rho\left(\beta_0^D, Z_i\right) \xrightarrow{D} N\left(0, \Omega_0\right), \quad \Omega_0 \stackrel{def}{\equiv} \mathbb{E}\left[g\left(W_i\right) \rho\left(\beta_0^D, Z_i\right) \rho\left(\beta_0^D, Z_i\right)' g\left(W_i\right)'\right].
$$

Note that Γ_0 is an $L \times L$ "Jacobian" matrix and Ω_0 is the $L \times L$ variance-covariance matrix of the moments. A Slutsky Theorem finally gives

$$
\sqrt{N}\left(\hat{\beta}^D - \beta_0^D\right) \xrightarrow{p} N\left(0, \Gamma_0^{-1}\Omega_0\Gamma_0^{-1\prime}\right).
$$

We can consistently estimate Γ_0 and Ω_0 by

$$\hat{\Gamma} = \frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right) R_i'$$
$$\hat{\Omega} = \frac{1}{N}\sum_{i=1}^{N} g\left(W_i\right) \rho\left(\hat{\beta}^D, Z_i\right) \rho\left(\hat{\beta}^D, Z_i\right)' g\left(W_i\right)'$$

and construct standard errors and test statistics in the usual way.

Question: When does $\mathbb{E}\left[g\left(W_i\right) R_i'\right]^{-1}$ exist? Why is this important?

The Wald estimator

Consider the special case with no demand shifters ($J_2 = 0$) and a single binary supply shifter ($J_1 = 1$, $W_{1i} \in \{0, 1\}$). We set $g\left(W_i\right) = \left(1, W_{1i}\right)'$. Our moment condition continues to be

$$\mathbb{E}\left[\left.\rho\left(\beta_0^D, Z_i\right)\right| W_i\right] = 0.$$

This condition implies that

$$\mathbb{E}\left[\left.\rho\left(\beta_0^D, Z_i\right)\right| W_{1i} = 0\right] = \mathbb{E}\left[\left.\rho\left(\beta_0^D, Z_i\right)\right| W_{1i} = 1\right] = 0,$$

and hence, using the definition of $\rho\left(\beta_0^D, Z_i\right)$, that

$$\mathbb{E}\left[\left.\rho\left(\beta_0^D, Z_i\right)\right| W_{1i} = 0\right] = \mathbb{E}\left[\left.\ln Q_i\right| W_{1i} = 0\right] - \alpha_0^D - \epsilon_0^D \mathbb{E}\left[\left.\ln P_i\right| W_{1i} = 0\right] = 0$$
$$\mathbb{E}\left[\left.\rho\left(\beta_0^D, Z_i\right)\right| W_{1i} = 1\right] = \mathbb{E}\left[\left.\ln Q_i\right| W_{1i} = 1\right] - \alpha_0^D - \epsilon_0^D \mathbb{E}\left[\left.\ln P_i\right| W_{1i} = 1\right] = 0.$$

Solving for the price elasticity of demand yields

$$\epsilon_0^D = \frac{\mathbb{E}\left[\left.\ln Q_i\right| W_{1i} = 1\right] - \mathbb{E}\left[\left.\ln Q_i\right| W_{1i} = 0\right]}{\mathbb{E}\left[\left.\ln P_i\right| W_{1i} = 1\right] - \mathbb{E}\left[\left.\ln P_i\right| W_{1i} = 0\right]}.$$

Question: What is required to ensure that the denominator in the above expression is non-zero? Can you provide a graphical exposition of this identification result?

Over-identification

When $L > L = 1 + 1 + J_2$ our model is over-identified (we have more moment conditions than unknown parameters). This would typically occur when we have multiple supply shifters (with $g(W_i) = (1, W_{1i}', W_{2i}')'$. When our system is over-identified we choose $\hat{\beta}^D$ to set a linear combination of the L sample moments equal to zero.

$$\frac{1}{N}\sum_{i=1}^{N} \underbrace{\hat{C}}_{1+1+J_2 \times L} \underbrace{g(W_i)\,\rho\left(\hat{\beta}^D, Z_i\right)}_{1+L \times 1} = 0.$$

Here $\hat{C}$ is a "weight matrix". Note we cannot set all the sample moment conditions exactly equal to zero. We only have $1+1+J_2$ parameters we can wiggle around, which is less than the L equations we want to set equal to zero. The degree of over-identification is $L-1-1-J_2$. Using the definition of $\rho\left(\beta^D, Z_i\right)$ we have

$$\frac{1}{N}\sum_{i=1}^{N} \hat{C} g(W_i)\left(Y_i - R_i'\beta^S\right) = 0,$$

and hence that

$$\hat{\beta}^D = \left[\frac{1}{N}\sum_{i=1}^{N} \hat{C} g(W_i)\, R_i'\right]^{-1} \times \left[\frac{1}{N}\sum_{i=1}^{N} \hat{C} g(W_i)\, Y_i\right].$$

Re-arrangement gives

$$\sqrt{N}\left(\hat{\beta}^D - \beta_0^D\right) = \left[\frac{1}{N}\sum_{i=1}^{N} \hat{C} g(W_i)\, R_i'\right]^{-1} \times \left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \hat{C} g(W_i)\,\rho\left(\beta_0^D, Z_i\right)\right],$$

which, after an argument similar to the one outlined above yields (when $\hat{C} \xrightarrow{p} C_0$),

$$\sqrt{N}\left(\hat{\beta}^D - \beta_0^D\right) \xrightarrow{D} \mathcal{N}\left(0, \left(C_0\Gamma_0'\right)^{-1} C_0\Omega_0 C_0' \left(C_0\Gamma_0'\right)^{-1\prime}\right).$$

Question: How might we choose $\hat{C}$ in practice?

References

Angrist, J. D., Graddy, K., & Imbens, G. W. (2000). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *Review of Economic Studies*, 67(3), 499 – 527.

Angrist, J. D. & Krueger, A. (2001). Instrumental variables and the search for identification:

from supply and demand to natural experiments. *Journal of Economic Perspectives*, 15(4), 69 – 85.

Goldberger, A. S. (1991). *A Course in Econometrics.* Cambridge, MA: Harvard University Press.

Wright, P. G. (1928). *The Tariff on Animal and Vegetable Oils.* New York: MacMillan.