University College London

Department of Economics

M.Sc. in Economics

MC3: Econometric Theory and Methods

Course Notes 6

Identification, endogeneity and instrumental variables

Andrew Chesher

9/11/2005

1. Identification

At the start of this course we considered what restrictions, embodied in an econometric *model*, could allow identification of interesting features of the data generating process. These two lectures return to this fundamental topic.

Consider some feature of a data generating process, for example the value of a coefficient in a linear equation of an econometric model. Different data generating processes can imply the same probability distribution for outcomes given covariates. Data generating processes with this property are said to be *observationally equivalent*.

Definition 1.1. *A feature of a data generating process, S, is identifiable if it takes the same value in all data generating processes that are observationally equivalent to S.*

If this identification property does not hold then even with complete knowledge of the probability distribution of outcomes given covariates we could not determine the value of the feature of interest of the data generating process. Of course finite amounts of data only ever give us incomplete knowledge of this probability distribution.

If only one value of a feature of a data generating process is concordant with a joint probability distribution of outcomes given covariates then the value of that feature in the data generating process is identified. Values of identifiable parameters can be estimated, values of non-identifiable parameters cannot.

We now investigate identification in a number of models.

1.1. The linear model

We saw that if a linear econometric model was specified,

$$Y = X'\beta + \varepsilon$$

for an outcome Y given k covariates (or "explanatory variables") $X = (X_1, \ldots, X_k)$ and if the restriction $E[\varepsilon|X = x] = 0$ was added to the restrictions of the model, then the vector of coefficients β could be identified as long as the values of X available showed sufficient variation. This result arises because in this model

$$E[Y|X = x] = x'\beta$$

and on the left hand side of this equation we have a feature of the conditional distribution of Y given $X = x$ about which data are informative. If we knew $E[Y|X = x]$ at k or more values of x then we could solve for β using any subset of k linearly independent x values.

In practice we do not know $E[Y|X = x]$ at any value of x, but we do have data on Y at each of n values of x. With y_n representing these data, and with X_n representing a $n \times k$ matrix containing the values of x, and with ε_n now representing the n unobserved values of ε we have

$$y_n = X_n\beta + \varepsilon_n$$

and the OLS estimator

$$\hat{\beta}_{OLS} = (X_n'X_n)^{-1}X_n'y_n$$

or the GLS estimator

$$\hat{\beta}_{GLS} = (X_n' \Sigma_n^{-1} X_n)^{-1} X_n' \Sigma_n^{-1} y_n$$

which are unbiased estimators of β under the restrictions of the model. Efficiency considerations lead to the choice of one estimator rather than another.

1.2. Nonlinear models

If instead we posit a *nonlinear* model

$$Y = g(X, \theta) + \varepsilon$$

with some particular function g specified, and the restriction:

$$E[\varepsilon | X = x] = 0$$

then

$$E[Y|X = x] = g(x, \theta) \tag{1.1}$$

and with sufficient values of x available and knowledge of $E[Y|X = x]$ at these values we may be able to solve for a unique value of θ in which case θ is identifiable. In practice we do not know $E[Y|X = x]$ at any value of x, but we do have data on Y at each of n values of x.

In this circumstance we can define the nonlinear least squares estimator

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{n} (y_i - g(x_i, \theta))^2$$

which, in well behaved problems is the solution to

$$\sum_{i=1}^{n} (y_i - g(x_i, \hat{\theta})) \frac{d}{d\theta} g(x_i, \hat{\theta}) = 0$$

which can be thought of as flowing from the following (but not the only) implication of (1.1).

$$E[(Y - g(X, \theta)) \frac{d}{d\theta} g(X, \theta) | X = x] = 0.$$

1.3. Models in which the probability distribution of outcomes is parametrically specified

In some models the *complete* probability distribution of outcomes given covariates is specified. We have studied maximum likelihood estimation of parameters of such models. But, when are values of parameters of such models identified?

Consider an example of such a model in which it is maintained that $Y_1, \ldots, Y_n$ are identically and independently distributed, each with the exponential probability density function

$$f(y; \lambda) = \lambda \exp(-\lambda y), \quad \lambda > 0$$

The joint probability density function of $Y_1, \ldots, Y_n$ is therefore

$$f(y_1, \ldots, y_n; \lambda) = \lambda^n \exp(-\lambda \sum_{i=1}^{n} y_i).$$

Suppose we were told that, for a certain data generating process which conforms to this model, the joint probability density function of $Y_1, \ldots, Y_n$ is

$$f(y_1, \ldots, y_n; \lambda) = 4^n \exp(-4 \sum_{i=1}^{n} y_i).$$

It is clear that in that process the value of λ, λ_0 say, must be $\lambda_0 = 4$ and, evidently, the value of λ is identified.

But suppose the maintained model is that $Y_1, \ldots, Y_n$ are identically and independently distributed, each with the exponential probability density function

$$f(y; \lambda_1, \lambda_2) = \left(\lambda^1 + \lambda^2\right) \exp(-\left(\lambda^1 + \lambda^2\right) y).$$

In this case, knowing that the joint probability density function of $Y_1, \ldots, Y_n$ is: $4^n \exp\left(-4 \sum_{i=1}^{n} y_i\right)$ does not allow us to deduce the data generating values: λ_0^1 and λ_0^2. We can identify the value of $\lambda_0^1 + \lambda_0^2$ as 4, but not the values λ_0^1 or λ_0^2 because, for example $\left(\lambda_0^1, \lambda_0^2\right) = (1, 3)$ and $\left(\lambda_0^1, \lambda_0^2\right) = (2, 2)$ both result in the same joint probability density function of $Y_1, \ldots, Y_n$.[1]

In practice we do not know the joint density function of $Y_1, \ldots, Y_n$, just that it has the form, in the first case considered, $\lambda_0^n \exp(-\lambda_0 \sum_{i=1}^{n} y_i)$ where λ_0 is the data generating value of λ. The maximum likelihood estimator for this problem solves the equation

$$\frac{n}{\hat{\lambda}} - \sum_{i=1}^{n} y_i = 0$$

which can be thought of as flowing from the result that the expectation of the derivative with respect to λ of the log likelihood function evaluated at the data generating value of λ is zero, that is that

$$E[\frac{n}{\lambda_0} - \sum_{i=1}^{n} Y_i] = 0$$

where λ_0 is the data generating value of λ.

2. Identification when latent variables are correlated with covariates

In many problems studied in econometrics it is *not* possible to maintain restrictions requiring that the expected value of the latent variable in an equation is zero given the values of the right hand side variables in the equation. Here is an example.

[1] Of course we can be sure that each of λ_0^1 and λ_0^2 are less than 4 since both must be positive. This is an example of identification within a *set* of values.

Consider a simple version of the Mincer model for returns to schooling with the following *structural equations.*

$$W = \alpha_0 + \alpha_1 S + \alpha_2 Z + \varepsilon_1 \tag{2.1}$$
$$S = \beta_0 + \beta_1 Z + \varepsilon_2 \tag{2.2}$$

Here W is the log wage, S is years of schooling, Z is some characteristic of the individual, and ε_1 and ε_2 are unobservable latent random variables. We might expect those who receive unusually high levels of schooling given Z to also receive unusually high wages given Z and S, a situation that would arise if ε_1 and ε_2 were affected positively by ability, a characteristic not completely captured by variation in Z.

2.1. Endogeneity

In this problem we might be prepared to impose the following restrictions.

$$E[\varepsilon_1|Z = z] = 0 \tag{2.3}$$
$$E[\varepsilon_2|Z = z] = 0 \tag{2.4}$$

We could not impose the restriction

$$E[\varepsilon_1|S = s, Z = z] = 0$$

unless ε_1 was believed to be uncorrelated with ε_2 and we have already argued that ability may positively influence ε_1 and ε_2 which results in them being positively correlated.

Considering just the first (W) equation,

$$E[W|S = s, Z = z] = \alpha_0 + \alpha_1 s + \alpha_2 z + E[\varepsilon_1|S = s, Z = z]$$

and the final term here will not in general be zero because, if ε_1 is positively correlated with ε_2, then ε_1 will tend to be large when s is large relative to $\beta_0 + \beta_1 z$, that is when ε_2 is large. A variable like S, appearing in a structural form equation and correlated with the latent variable in the equation, is called an *endogenous* variable.

2.2. Reduced form equations

What can be learned about the values of parameters given the restrictions (2.3) and (2.4)? We can see the impact of these restrictions by substituting for S in the equation (2.1) for W, giving the following.

$$W = (\alpha_0 + \alpha_1\beta_0) + (\alpha_1\beta_1 + \alpha_2) Z + \varepsilon_1 + \alpha_1\varepsilon_2 \tag{2.5}$$
$$S = \beta_0 + \beta_1 Z + \varepsilon_2 \tag{2.6}$$

Equations like this, in which each equation involves exactly one endogenous variable are called *reduced form* equations.

The restrictions (2.3) and (2.4) imply that

$$E[W|Z = z] = (\alpha_0 + \alpha_1\beta_0) + (\alpha_1\beta_1 + \alpha_2) z \tag{2.7}$$
$$E[S|Z = z] = \beta_0 + \beta_1 z \tag{2.8}$$

and given enough (at least 2) distinct values of z and knowledge of the left hand side quantities we can solve for $(\alpha_0 + \alpha_1\beta_0)$, $(\alpha_1\beta_1 + \alpha_2)$, β_0 and β_1. So, the values of these *functions* of parameters of the structural equations *can* be identified.

In practice we do not know the left hand side quantities but with enough data we can estimate the data generating values of $(\alpha_0 + \alpha_1\beta_0)$, $(\alpha_1\beta_1 + \alpha_2)$, β_0 and β_1, for example by OLS applied first to (W, Z) data and then to (S, Z) data.

The values of β_0 and β_1 *are* identified but the values of α_0, α_1 and α_2 are *not*, for without further restrictions their values cannot be deduced from knowledge of $(\alpha_0 + \alpha_1\beta_0)$, $(\alpha_1\beta_1 + \alpha_2)$, β_0 and β_1, and the restrictions (2.3) and (2.4) have no further implications.

2.3. Identification using an exclusion restriction

One restriction we might be prepared to add to the model is the restriction $\alpha_2 = 0$. Whether or not that is a reasonable restriction to maintain depends on the nature of the variable Z.

If Z were a measure of some characteristic of the environment of the person at the time that schooling decisions were made (for example the parents' income, or some measure of an event that perturbed the schooling choice) then we might be prepared to maintain the restriction that, given schooling achieved (S), Z does not affect W, i.e. that $\alpha_2 = 0$.

This restriction may be sufficient to identify the remaining parameters. If the restriction is true then the coefficients on Z in (2.5) and on z in (2.7) become $\alpha_1\beta_1$. We have already seen that (the value of) the coefficient β_1 is identified. If β_1 is not itself zero (that is Z does indeed affect years of schooling) then α_1 is identified as the ratio of the coefficients on Z in the regressions of W and S on Z. With α_1 identified and β_0 already identified, identification of α_0 follows directly.

2.4. Indirect least squares estimation

Estimation could proceed under the restriction $\alpha_2 = 0$ by calculating OLS (or GLS) estimates of the "reduced form" equations

$$\begin{aligned} W &= \pi_{01} + \pi_{11}Z + U_1 \\ S &= \pi_{02} + \pi_{12}Z + U_2 \end{aligned}$$

where

$$\begin{aligned} \pi_{01} &= \alpha_0 + \alpha_1\beta_0 & \pi_{11} &= \alpha_1\beta_1 \\ \pi_{02} &= \beta_0 & \pi_{12} &= \beta_1 \\ U_1 &= \varepsilon_1 + \alpha_1\varepsilon_2 & U_2 &= \varepsilon_2 \end{aligned}$$

and

$$E[U_1|Z = z] = 0 \quad E[U_2|Z = z] = 0$$

which follow from the restrictions (2.3) and (2.4), and solving the equations

$$\begin{aligned} \hat{\pi}_{01} &= \hat{\alpha}_0 + \hat{\alpha}_1\hat{\beta}_0 & \hat{\pi}_{11} &= \hat{\alpha}_1\hat{\beta}_1 \\ \hat{\pi}_{02} &= \hat{\beta}_0 & \hat{\pi}_{12} &= \hat{\beta}_1 \end{aligned}$$

given values of the $\hat{\pi}$'s for values of the $\hat{\alpha}$'s and $\hat{\beta}$'s, as follows.

$$\begin{array}{rclcrcl} \hat{\alpha}_0 & = & \hat{\pi}_{01} - \hat{\pi}_{02}\,(\hat{\pi}_{11}/\hat{\pi}_{12}) & \qquad & \hat{\alpha}_1 & = & \hat{\pi}_{11}/\hat{\pi}_{12} \\ \hat{\beta}_0 & = & \hat{\pi}_{02} & & \hat{\beta}_1 & = & \hat{\pi}_{12} \end{array}$$

Estimators obtained in this way, by solving the equations relating structural form parameters to reduced form parameters with OLS estimates replacing the reduced form parameters, are known as *Indirect Least Squares* estimators. They were first proposed by Jan Tinbergen in 1930.

2.5. Over identification

Suppose that there are *two* covariates, Z_1 and Z_2 whose impact on the structural equations we are prepared to restrict so that *both* affect schooling choice but *neither* affect the wage given the amount of schooling achieved.

The structural equations then take the following form

$$\begin{array}{rcl} W & = & \alpha_0 + \alpha_1 S + \varepsilon_1 \\ S & = & \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \varepsilon_2 \end{array}$$

and the reduced form equations are as follows

$$\begin{array}{rcl} W & = & \pi_{01} + \pi_{11} Z_1 + \pi_{21} Z_2 + U_1 \\ S & = & \pi_{02} + \pi_{12} Z_1 + \pi_{22} Z_2 + U_2 \end{array}$$

where

$$\begin{array}{rclcrclcrcl} \pi_{01} & = & \alpha_0 + \alpha_1\beta_0 & \qquad & \pi_{11} & = & \alpha_1\beta_1 & \qquad & \pi_{21} & = & \alpha_1\beta_2 \\ \pi_{02} & = & \beta_0 & & \pi_{12} & = & \beta_1 & & \pi_{22} & = & \beta_2 \end{array}$$

and

$$U_1 = \varepsilon_1 + \alpha_1\varepsilon_2 \qquad U_2 = \varepsilon_2.$$

The values of the reduced form equations' coefficients are identified under restrictions (2.3) and (2.4). Now, note, there are *two* ways in which the coefficient α_1 can be identified, as follows

$$\alpha_1 = \frac{\pi_{11}}{\pi_{12}} \quad \alpha_1 = \frac{\pi_{21}}{\pi_{22}}$$

In this situation we say that the value of the parameter α_1 is *over identified.* In the case we examined before, in which there is just one way of deducing the value of a structural form parameter from knowledge of the reduced form equations' coefficients, we say the structural form parameter is *just identified.*

Of course the over identification of α_1 is of no particular interest if we *know* the reduced form equations' coefficients. But in practice we do not know the value of these coefficients and when we come to compute Indirect Least Squares *estimates* of α_1 using *estimated* reduced form coefficients:

$$\hat{\alpha}_1^{Z_1} = \frac{\hat{\pi}_{11}}{\hat{\pi}_{12}} \quad \hat{\alpha}_1^{Z_2} = \frac{\hat{\pi}_{21}}{\hat{\pi}_{22}}$$

we will usually find that $\hat{\alpha}_1^{Z_1} \neq \hat{\alpha}_1^{Z_2}$ even though these are both estimates of the value of the same structural form parameter.

If the discrepancy was found to be very large then we might doubt whether the restrictions of the model are correct. This suggests that tests of over identifying restrictions can detect misspecification of the econometric model. If the discrepancy is not large then there is scope for combining the estimates to produce a single estimate that is more efficient than either taken alone.

2.6. Indirect least squares and instrumental variables

Return to the problem of estimation of structural form coefficients under the restrictions which cause α_1 to be just identified.

Another way to look at this problem is as follows. With the restriction $\alpha_2 = 0$ *imposed* the structural form equations (2.1) and (2.2) are as follows.

$$\begin{aligned} W &= \alpha_0 + \alpha_1 S + \varepsilon_1 \\ S &= \beta_0 + \beta_1 Z + \varepsilon_2 \end{aligned}$$

and we have the restriction $E[\varepsilon_1 | Z = z] = 0$ which implies that

$$E[W - \alpha_0 - \alpha_1 S | Z = z] = 0$$

and, from this

$$\begin{aligned} E[W - \alpha_0 - \alpha_1 S] &= 0 \\ E[(W - \alpha_0 - \alpha_1 S)\, Z] &= 0 \end{aligned}$$

equivalently

$$\begin{bmatrix} E[W] \\ E[WZ] \end{bmatrix} = \begin{bmatrix} 1 & E[S] \\ E[Z] & E[SZ] \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix}$$

and so

$$\begin{aligned} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} &= \begin{bmatrix} 1 & E[S] \\ E[Z] & E[SZ] \end{bmatrix}^{-1} \begin{bmatrix} E[W] \\ E[WZ] \end{bmatrix} \\ &= \frac{1}{E[SZ] - E[S]E[Z]} \begin{bmatrix} E[SZ] & -E[S] \\ -E[Z] & 1 \end{bmatrix} \begin{bmatrix} E[W] \\ E[WZ] \end{bmatrix} \end{aligned}$$

from which note that

$$\alpha_1 = \frac{E[WZ] - E[W]E[Z]}{E[SZ] - E[S]E[Z]} = \frac{Cov(W,Z)}{Cov(S,Z)}. \tag{2.9}$$

Replacing these moments by sample analogues (that is sample means and mean products) gives an estimator of α_1 which is identical to the Indirect Least Squares estimator. The estimators are identical because

$$\hat{\pi}_{11} = \frac{\widehat{Cov}(W,Z)}{\widehat{Var}(Z)} \quad \hat{\pi}_{12} = \frac{\widehat{Cov}(S,Z)}{\widehat{Var}(Z)}$$

and so

$$\hat{\alpha}_1^{ILS} = \frac{\hat{\pi}_{11}}{\hat{\pi}_{12}} = \frac{\widehat{Cov}(W, Z)}{\widehat{Cov}(S, Z)}$$

which is the same as the right hand side of (2.9) with moments replaced by sample based estimates.

3. Instrumental variables

Consider again the linear model for an outcome Y given covariates (k in number) X

$$Y = X'\beta + \varepsilon \tag{3.1}$$

and suppose that the restriction $E[\varepsilon|X = x] = 0$ cannot be maintained but that there exist m variables Z for which the restriction $E[\varepsilon|Z = z] = 0$ can be maintained.

This restriction implies that

$$E[Y - X'\beta|Z = z] = 0$$

and thus that

$$E[Z(Y - X'\beta)|Z = z] = 0$$

which implies that, unconditionally

$$E[Z(Y - X'\beta)] = 0.$$

and thus

$$E[ZY] = E[ZX']\beta. \tag{3.2}$$

First suppose $m = k$, and that $E[ZX']$ has rank k. Then β can be expressed in terms of moments of Y, X and Z as follows

$$\beta = E[ZX']^{-1}E[ZY].$$

and β is (just) identifiable. This leads directly to an analogue type estimator if we replace expectations by sample moments of realisations of ZX' and ZY, denoted here by $\overline{ZX'}$ and $\overline{ZY}$, as follows.

$$\hat{\beta} = (\overline{ZX'})^{-1}(\overline{ZY})$$

If X_n denotes a matrix of realisations of the vector random variable X, with realisation $i \in (1, \ldots, n)$ occupying row i, with Z_n similarly denoting a matrix of realisations of Z and if y_n denotes a vector of realisations of Y, then

$$\begin{aligned} (\overline{ZX'}) &= n^{-1}(Z_n'X_n) \\ (\overline{ZY}) &= n^{-1}(Z_n'y_n) \end{aligned}$$

and

$$\hat{\beta} = (Z_n'X_n)^{-1}Z_n'y_n.$$

In the context of the just identified returns to schooling model set out above, this is the Indirect Least Squares estimator.

Now suppose that $m > k$. We could try to define an estimator $\hat{\beta}$, as before, as the solution to

$$(\overline{ZX'})\hat{\beta} = (\overline{ZY}) \tag{3.3}$$

where bar indicates a sample average. But in any finite sized sample it is unlikely that we will find a solution since we have $m > k$ equations in k unknowns. If the sample moments were exactly equal to the moments $E[ZX']$ and $E[ZY]$ then an estimator could be produced, the same value being obtained using any subset of k of the m equations by virtue of (3.2). But in practice the sample moments will almost certainly not be realised in this fortuitous manner. In this case we could define estimators using any subset of size k of the m equations (3.3) involving the sample moments. But this seems an inefficient way to proceed, involving the discarding of information.

3.1. Generalised Method of Moments estimation

An alternative is to define an estimator which comes as close as possible (in some sense) to satisfying all of the equations (3.3). One way to do this is to define a family of estimators, $\hat{\beta}_W$ as

$$\hat{\beta}_W = \arg\min_{\beta} \left((\overline{ZY}) - (\overline{ZX'})\beta\right)' W \left((\overline{ZY}) - (\overline{ZX'})\beta\right)$$

where W is a $m \times m$ full rank, positive definite symmetric matrix. This M-estimator is an example of what is known as the *Generalised Method of Moments* (GMM) estimator.

Different choices of W lead to different estimators unless $m = k$ and the choice among these is commonly made by considering their accuracy. In most cases exact sampling variances are difficult to calculate and depend on fine details of the data generating process about which we are likely to be ignorant. So we consider the limiting distribution of the GMM estimator for alternative choices of W and choose W to minimise the variance of the limiting distribution of $n^{1/2}(\hat{\beta}_W - \beta_0)$. In standard cases this means choosing W to be proportional to a consistent estimator of the inverse of the variance of the limiting distribution of $n^{1/2}\left((\overline{ZY}) - (\overline{ZX'})\beta\right)$.

3.2. Generalised Instrumental Variables estimation

Now write $\hat{\beta}_W$ explicitly in terms of sample moments, as follows

$$\begin{aligned}
\hat{\beta}_W &= \arg\min_{\beta} \left(n^{-1}(Z_n'y_n) - n^{-1}(Z_n'X_n)\beta\right)' W \left(n^{-1}(Z_n'y_n) - n^{-1}(Z_n'X_n)\beta\right) &(3.4)\\
&= \arg\min_{\beta} \left(n^{-1/2}(Z_n'y_n) - n^{-1/2}(Z_n'X_n)\beta\right)' W \left(n^{-1/2}(Z_n'y_n) - n^{-1/2}(Z_n'X_n)\beta\right)
\end{aligned}$$

and consider what the (asymptotically) efficient choice of W is by examining the variance of $n^{-1/2}(Z_n'y_n) - n^{-1/2}(Z_n'X_n)\beta$.

We have, since $y_n = X_n\beta + \varepsilon_n$,

$$n^{-1/2}(Z_n'y_n) - n^{-1/2}(Z_n'X_n)\beta = n^{-1/2}(Z_n'\varepsilon_n)$$

and *if* we suppose that $Var(\varepsilon_n|Z_n) = \sigma^2 I_n$,

$$Var\left(n^{-1/2}(Z_n'\varepsilon_n)|Z_n\right) = \sigma^2(n^{-1}Z_n'Z_n).$$

This suggests choosing $W = (n^{-1}Z_n'Z_n)$ in the equation below (3.4) leading to the following minimisation problem

$$\begin{aligned}\hat{\beta}_n &= \arg\min_{\beta} \left(n^{-1/2}(Z_n'y_n) - n^{-1/2}(Z_n'X_n)\beta\right)' (n^{-1}Z_n'Z_n)^{-1} \left(n^{-1/2}(Z_n'y_n) - n^{-1/2}(Z_n'X_n)\beta\right) \\ &= \arg\min_{\beta} \left((Z_n'y_n) - (Z_n'X_n)\beta\right)' (Z_n'Z_n)^{-1} \left((Z_n'y_n) - (Z_n'X_n)\beta\right)\end{aligned}$$

where on the second line factors involving n have been multiplied through.

The first order conditions for this problem, satisfied by $\hat{\beta}_n$ are:

$$2\hat{\beta}_n'(X_n'Z_n)(Z_n'Z_n)^{-1}(Z_n'X_n) - 2(X_n'Z_n)(Z_n'Z_n)^{-1}(Z_n'y_n) = 0$$

leading to the following estimator.

$$\hat{\beta}_n = \left((X_n'Z_n)(Z_n'Z_n)^{-1}(Z_n'X_n)\right)^{-1}(X_n'Z_n)(Z_n'Z_n)^{-1}(Z_n'y_n)$$

This is known as the *generalised instrumental variable estimator* (GIVE)

The asymptotic properties of this estimator are obtained as follows.

Substituting $y_n = X_n\beta + \varepsilon_n$ gives

$$\begin{aligned}\hat{\beta}_n &= \beta + \left(X_n'Z_n(Z_n'Z_n)^{-1}Z_n'X_n\right)^{-1} X_n'Z_n(Z_n'Z_n)^{-1}Z_n'\varepsilon_n \\ &= \beta + \left(n^{-1}X_n'Z_n(n^{-1}Z_n'Z_n)^{-1}n^{-1}Z_n'X_n\right)^{-1} n^{-1}X_n'Z_n(n^{-1}Z_n'Z_n)^{-1}n^{-1}Z_n'\varepsilon_n\end{aligned}$$

and if

$$\begin{aligned}\operatorname*{plim}_{n\to\infty}(n^{-1}Z_n'Z_n) &= \Sigma_{ZZ} && (3.5)\\ \operatorname*{plim}_{n\to\infty}(n^{-1}X_n'Z_n) &= \Sigma_{XZ} && (3.6)\\ \operatorname*{plim}_{n\to\infty}(n^{-1}Z_n'\varepsilon_n) &= 0\end{aligned}$$

with Σ_{ZZ} having full rank (m) and Σ_{XZ} having full rank (k) then

$$\operatorname*{plim}_{n\to\infty}\hat{\beta}_n = \beta$$

and we have a *consistent* estimator.

To obtain the limiting distribution of $n^{1/2}(\hat{\beta} - \beta)$ note that

$$n^{1/2}(\hat{\beta} - \beta) = \left(n^{-1}X_n'Z_n(n^{-1}Z_n'Z_n)^{-1}n^{-1}Z_n'X_n\right)^{-1} n^{-1}X_n'Z_n(n^{-1}Z_n'Z_n)^{-1}n^{-1/2}Z_n'\varepsilon_n$$

Here $n^{-1/2}$ multiplying the final term arises because of the scaling, $n^{1/2}$, applied to $\hat{\beta} - \beta$.

Under the conditions above (3.5) and (3.6) we have the limiting distribution result:

$$\operatorname{plim} n^{1/2}(\hat{\beta} - \beta) = \left(\Sigma_{XZ}\Sigma_{ZZ}^{-1}\Sigma_{ZX}\right)^{-1}\Sigma_{XZ}\Sigma_{ZZ}^{-1}\operatorname{plim}\left(n^{-1/2}Z_n'\varepsilon_n\right)$$

where $\Sigma_{ZX} = \Sigma'_{XZ}$, and if a Central Limit Theorem applies to $n^{-1/2} Z'_n \varepsilon_n$ giving

$$\text{plim}\left(n^{-1/2} Z'_n \varepsilon_n\right) = N(0, \sigma^2 \Sigma_{ZZ})$$

then

$$\text{plim}\, n^{1/2}(\hat{\beta} - \beta) = N(0, V)$$

where

$$\begin{aligned} V &= \sigma^2 \left(\Sigma_{XZ}\Sigma_{ZZ}^{-1}\Sigma_{XZ}\right)^{-1} \Sigma_{XZ}\Sigma_{ZZ}^{-1}\Sigma_{ZZ}\Sigma_{ZZ}^{-1}\Sigma_{ZX}\left(\Sigma_{XZ}\Sigma_{ZZ}^{-1}\Sigma_{ZX}\right)^{-1} \\ &= \sigma^2 \left(\Sigma_{XZ}\Sigma_{ZZ}^{-1}\Sigma_{ZX}\right)^{-1} \end{aligned}$$

and so

$$\text{plim}\, n^{1/2}(\hat{\beta} - \beta) \simeq N(0, \sigma^2 \left(\Sigma_{XZ}\Sigma_{ZZ}^{-1}\Sigma_{ZX}\right)^{-1}).$$

There are three important points to note here.

3.2.1. GIVE in just identified models

When $m = k$ so that there are the same number of instrumental variables as elements in X, the formula for $\hat{\beta}_n$ simplifies (check that this is true) to

$$\hat{\beta}_n = \left(Z'_n X_n\right)^{-1} Z'_n y_n$$

because then $X'_n Z_n$ and its transpose are square and (by assumption non-singular) and so can be separately inverted.

3.2.2. Elements of X as instruments

If some elements of X, say X_1, are such that $E[\varepsilon|X_1 = x_1] = 0$ then these elements can serve as instruments, i.e. can be elements of Z. Check that if all elements of X have this property then, above, on choosing $Z_n = X_n$ we retrieve the conventional OLS estimator.

3.2.3. GIVE is equivalent to OLS using predicted endogenous variables

Suppose there is a model for X,

$$X = Z\Phi + V$$

where $E[V|Z] = 0$. The OLS estimator of Φ is

$$\hat{\Phi}_n = \left(Z'_n Z_n\right)^{-1} Z'_n X_n$$

and the "predicted value" of X for a given Z is

$$\hat{X}_n = Z_n \left(Z'_n Z_n\right)^{-1} Z'_n X_n.$$

Note that

$$\hat{X}'_n \hat{X}_n = X'_n Z_n (Z'_n Z_n)^{-1} Z'_n X_n$$

and

$$\hat{X}_n' y_n = X_n' Z_n (Z_n' Z_n)^{-1} Z_n' y_n.$$

So the Generalised Instrumental Variables Estimator can be written as

$$\hat{\beta}_n = \left(\hat{X}_n' \hat{X}_n\right)^{-1} \hat{X}_n' y_n.$$

that is, as the OLS estimator of the coefficients of a linear relationship between y_n and the *predicted values* of X_n got from OLS estimation of a linear relationship between X_n and the instrumental variables Z_n.

4. Another structural model: market demand and supply

Elementary demand theory leads to demand equations which express quantity demanded by a consumer as a function of prices and income and other variables which affect tastes. Aggregated across consumers this leads to a market demand function which we write here for the sake of this example as the linear function

$$q^D = \gamma_D p + x' \beta_D + \varepsilon_D \tag{4.1}$$

where x contains a measure of income and perhaps its distribution, prices other than that of the good under consideration and variables that capture variation in tastes[2]. We might be prepared to maintain the restriction $E[\varepsilon_D | p, x] = 0$, which implies that $\gamma_D p + x' \beta_D$ is the expected value of q^D on p and x.

The elementary theory of the firm suggests a supply equation giving the desired supply on the part of firms at any given price,

$$q^S = \gamma_S p + x' \beta_S + \varepsilon_S \tag{4.2}$$

and we might be prepared to maintain the restriction that $E[\varepsilon_S | p, x] = 0$.

We might expect different elements of x to be relevant in the determination of the two desired quantities (demanded by consumers and supplied by firms) and this can be captured by allowing particular (different) elements of β_D and β_S to be zero. For example desired demand is likely to be shifted by variations in consumers' incomes and by things that affect tastes, while desired supply will likely be shifted by variation in variables affecting firm's costs.

We focus on the price coefficients and note that these may be of direct interest. For example if the government is contemplating an indirect tax on a good (e.g. tobacco) or a subsidy on a good (e.g. children's clothing) it may wish to know how demand would alter purely as the price paid by consumers varies. In this case the government is interested in the coefficient γ_D above.

It is rarely the case that we have data that bears directly on the aggregate (desired) demand curve. Rather we see a common quantity supplied and consumed and a market clearing price at which these transactions take place. Formally, we observe realisations of q^e where $q^e = q^D = q^S$ and of p^e where p^e is the market clearing price at which the equalities amongst desired and transacted quantities hold. Data on q^e, p^e and x, perhaps obtained by observation over a period of time or across a series of regions, gives us information about the joint distribution of q^e and

[2] In this section the upper/lowercase convention for random variables and their realisations is not respected.

p^e given x. This enables us to, for example, estimate the regression of q^e on x and of p^e on x. These are *reduced form* equations. The original relationships (4.1) and (4.2) together with the price clearing condition are the *structural form equations*.

When prices clear, equilibrium quantity and price, q^e and p^e, are determined by the structural form equations

$$q^e = \gamma_D p^e + x'\beta_D + \varepsilon_D \tag{4.3}$$
$$q^e = \gamma_S p^e + x'\beta_S + \varepsilon_S \tag{4.4}$$

which can be written in matrix notation as follows,

$$\begin{bmatrix} 1 & -\gamma_D \\ 1 & -\gamma_S \end{bmatrix} \begin{bmatrix} q^e \\ p^e \end{bmatrix} = \begin{bmatrix} \beta_D' \\ \beta_S' \end{bmatrix} x + \begin{bmatrix} \varepsilon_D \\ \varepsilon_S \end{bmatrix},$$

a special case of the generic linear simultaneous structural equation form

$$\Gamma y = Bx + \varepsilon.$$

Note that the restrictions $E[\varepsilon_D|p,x] = 0$ and $E[\varepsilon_S|p,x] = 0$ do not in general imply that $E[\varepsilon_D|p^e,x] = 0$ and $E[\varepsilon_S|p^e,x] = 0$ because the market clearing price will be a function of ε_D and ε_S. The earlier discussion suggests that these conditions on the latent variables in the desired demand and supply equations may not be useful in identifying the coefficients of those equations when the data available are generated with markets clearing. Further, OLS, GLS and similar estimation procedures applied to (4.3) and (4.4) will generally produce inconsistent estimators of the coefficients in these equations.

However we might be prepared to maintain the restrictions

$$E[\varepsilon_D|x] = 0 \quad E[\varepsilon_S|x] = 0 \tag{4.5}$$

and these are sufficient restrictions to allow identification of the reduced form equations' coefficients.

We obtain the reduced form equations as

$$y = \Gamma^{-1}Bx + \Gamma^{-1}\varepsilon$$

or

$$y = \Pi x + U$$

where $\Pi = \Gamma^{-1}B$ and $U = \Gamma^{-1}\varepsilon$ and in this market model,

$$\Gamma^{-1} = \frac{1}{\gamma_D - \gamma_S} \begin{bmatrix} -\gamma_S & \gamma_D \\ -1 & 1 \end{bmatrix}.$$

The conditions (4.5) imply that $E[U|x] = 0$ which is, as we have seen before, sufficient (with sufficient variation in x) to identify the coefficients Π.

The reduced form equations for the market model are:

$$q^e = \frac{1}{\gamma_D - \gamma_S} x' \left(-\gamma_S\beta_D + \gamma_D\beta_S\right) + \frac{1}{\gamma_D - \gamma_S}\left(-\gamma_S\varepsilon_D + \gamma_D\varepsilon_S\right)$$
$$p^e = \frac{1}{\gamma_D - \gamma_S} x' \left(-\beta_D + \beta_S\right) + \frac{1}{\gamma_D - \gamma_S}\left(-\varepsilon_D + \varepsilon_S\right).$$

which we write as

$$\begin{aligned} q^e &= x'\pi_Q + u_Q \\ p^e &= x'\pi_P + u_P. \end{aligned}$$

Under the conditions stated OLS, GLS etc., will give well behaved estimators of the coefficients on x in these equations. Note that the government's coefficient of interest, γ_D, is entwined here with many other coefficients. It is not clear how one would retrieve an estimate of γ_D from estimates of these combinations of γ_D, γ_S, β_D, and β_S. Indeed it is not clear that it can be done at all.

To see the problem, note that all structural forms

$$A\Gamma y = ABx + A\varepsilon$$

with full rank A have identical reduced forms because $(A\Gamma)^{-1} AB = \Gamma^{-1}B = \Pi$, independent of A, and $(A\Gamma)^{-1} A\varepsilon = \Gamma^{-1}\varepsilon = U$. Without further restrictions we cannot identify Γ and B separately though we can identify $\Gamma^{-1}B = \Pi$.

If we could maintain the restriction that an element of x does not feature in the structural demand equation (perhaps because it only affects producers' costs) then, as in the returns to schooling model, the parameters of the structural demand equation could be identified. Similarly, if we could maintain the restriction that an element of x does not feature in the structural supply equation (perhaps because it only affects consumers' preferences) then the parameters of the structural supply equation could be identified.

These are, of course, the sorts of exclusion restrictions that we have already met when studying the returns to schooling model. There are general results on identification under exclusion restrictions, discussed now.

5. Identification in simultaneous equations models

The identifiability of structural form coefficients in linear simultaneous equations models with restrictions only on the conditional expectation of latent variables given covariates depends upon our ability to deduce structural form coefficients from knowledge of reduced form coefficients.

The main result that we need here is as follows. When the only restrictions available are exclusion restrictions, i.e. restrictions that set certain structural form coefficients to zero (e.g. factors that affect firm's costs do not appear in structural demand functions) then, in order to be able to identify the coefficients of a structural form equation (i) in which there are M_i endogenous variables with non-zero coefficients, a priori, there must be at least $M_i - 1$ exogenous variables that appear elsewhere in the system that are excluded from equation i. By "appear elsewhere" we mean feature, with non-zero coefficients, in at least one other structural form equation.

The proof of this result involves consideration of the conditions under which the part of the equation $\Gamma\Pi = B$ relating to equation i ($\Gamma_i\Pi = B_i$ where Γ_i and B_i are the ith columns of Γ and B) can be solved for unrestricted elements of Γ_i and B_i given values for Π. Here we do not give a proof but note that the IV method could only work when there are M_i endogenous variables in an equation (one normalised to have a coefficient equal to one and appearing on the left hand side of the equation) if we could find $M_i - 1$ exogenous variables, correlated with the endogenous variables, and not already appearing in the equation. This will be the case under the condition outlined in the previous paragraph.

The exogenous variables excluded from the equation of interest serve as instruments for the included endogenous variables. The IV method provides a route to estimation of just and over identified equations. In just identified equations the IV estimator is, as has been shown earlier, the Indirect Least Squares estimator.

In over identified equations the GIVE estimator produces an estimator known in the simultaneous equations context as the Two Stage Least Squares estimator - "two stage" because in the first stage one calculates predictions of the endogenous variables in the equation of interest applying OLS to the reduced form equations and in the second stage one estimates the structural form equation's coefficients by applying OLS using predicted values of right hand side endogenous variables in place of their observed values.

6. Concluding remarks

More detail on the topics covered here is available in the two recommended texts and you should read these.

Endogeneity and identification are very important, but subtle, topics lying at the core of econometrics - they are primary aspects of econometric method and approach that distinguish econometrics from statistics. They arise in econometrics because we develop models of behaviour and outcomes which are on occasions in some way removed from the probabilistic process about which the data we see are informative.