8-26

8.4.2 A TEST FOR OVERIDENTIFICATION

The motivation for choosing the IV estimator is not efficiency. The estimator is constructed to be consistent; efficiency is not a consideration. In Chapter 13, we will revisit the issue of efficient method of moments estimation. The observation that 2SLS represents the most efficient use of all L instruments establishes only the efficiency of the estimator in the class of estimators that use K linear combinations of the columns of **Z**. The IV estimator is developed around the **orthogonality conditions**

KT KT

$$E[\mathbf{z}_i\varepsilon_i] = \mathbf{0}. \tag{8-12}$$

The sample counterpart to this is the **moment equation**,

AU: Term "moment equation" already a KT in chap Here also?

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i\varepsilon_i = \mathbf{0}. \tag{8-13}$$

minus

The solution, when $L = K$, is $\mathbf{b}_{\mathrm{IV}} = (\mathbf{Z'X})^{-1}\mathbf{Z'y}$, as we have seen. If $L > K$, then there is no single solution, and we arrived at 2SLS as a strategy. Estimation is still based on (8-13). However, the sample counterpart is now L equations in K unknowns and (8-13) has no solution. Nonetheless, under the hypothesis of the model, (8-12) remains true. We can consider the additional restictions as a hypothesis that might or might not be supported by the sample evidence. The excess of moment equations provides a way to test the **overidentification** of the model. The test wil be based on (8-13), which, when evaluated at $\mathbf{b}_{\mathrm{IV}}$, will not equal zero when $L > K$, though the hypothesis in (8-12) might still be true.

KT

The test statistic will be a Wald statistic. (See Section 5.4.) The sample statistic, based on (8-13) and the IV estimator, is

Rom Rom

$$\bar{\mathbf{m}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i e_{IV,i} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i(y_i - \mathbf{x}_i'\mathbf{b}_{IV}).$$

The Wald statistic is

$$\chi^2[L-K] = \bar{\mathbf{m}}'[\mathrm{Var}(\bar{\mathbf{m}})]^{-1}\bar{\mathbf{m}}.$$

To complete the construction, we require an estimator of the variance. There are two ways to proceed. Under the assumption of the model,

Rom

$$Var[\bar{\mathbf{m}}] = \frac{\sigma^2}{n^2}\mathbf{Z'Z},$$

which can be estimated easily using the sample estimator of σ^2. Alternatively, we might base the estimator on (8-12), which would imply that an appropriate estimator would be

Rom Rom Rom Rom

$$Est.Var[\bar{\mathbf{m}}] = \frac{1}{n^2}\sum_{i=1} (\mathbf{z}_i e_{IV,i})(\mathbf{z}_i e_{IV,i})' = \frac{1}{n^2}\sum_{i=1} e_{IV,i}^2\mathbf{z}_i\mathbf{z}_i'.$$

These two estimators will be numerically different in a finite sample, but under the assumptions that we have made so far, both (multiplied by n) will converge to the same matrix, so the choice is immaterial. Current practice favors the second. The Wald statistic is, then

$$\left(\frac{1}{n}\sum_{i=1}^{n}\overset{\text{Rom}}{\mathbf{z}_i e_{IV,i}}\right)'\left[\frac{1}{n^2}\sum_{i=1}^{n}\overset{\text{Rom}}{e_{IV,i}^2\mathbf{z}_i\mathbf{z}_i'}\right]^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\overset{\text{Rom}}{\mathbf{z}_i e_{IV,i}}\right).$$

A remaining detail is the number of degrees of freedom. The test can only detect the failure of $L-K$ moment equations, so that is the rank of the quadratic form; the limiting distribution of the statistic is chi squared with $L-K$ degrees of freedom.

EXAMPLE 8.8 Overidentification of the Labor Supply Equation

In Example 8.5, we computed 2SLS estimates of the parameters of an equation for weeks worked. The estimator is based on

$\mathbf{x}$ = [1,ln*Wage*,*Education*,*Union*,*Female*]

and

$\mathbf{z}$ = [1,*Ind*, *Education*,*Union*,*Female*,*SMSA*].

There is one overidentifying restriction. The sample moment based on the 2SLS results in Table 8.1 is

$$(1/4165)\,\mathbf{Z}'\mathbf{e}_{2SLS} = [0, .03476, 0, 0, 0, \overset{\text{minus}}{-}.01543]'.$$

The chi squared statistic is 1.09399 with one degree of freedom. If the first suggested variance estimator is used, the statistic is 1.05241. Both are well under the 95% critical value of 3.84, so the hypothesis of overidentification is not rejected. percent

We note a final implication of the test. One might conclude, based on the underlying theory of the model, that the overidentification test relates to one particular instrumental variable and not another. For example, in our market equilibrium example with two instruments for the demand equation, *Rainfall* and *InputPrice*, rainfall is obviously exogenous, so a rejection of the overidentification restriction would eliminate *InputPrice* as a valid instrument. However, this conclusion would be inappropriate; the test suggests only that one or more of the elements in (8-12) are nonzero. It does not suggest which elements in particular these are.

8-28

~~possibility of bias due to correlation between Y_t and ε_t. Consider instrumental variables estimation using Y_{t-1} and C_{t-1} as the instruments for Y_t, and, of course, the constant term is its own instrument. One observation is lost because of the lagged values, so the results are based on 203 quarterly observations. The Hausman statistic can be computed in two ways:~~

1. ~~Use the Wald statistic for H with the Moore-Penrose generalized inverse. The common s^2 is the one computed by least squares under the null hypothesis of no correlation. With this computation, $H = 8.481$. There is $K^* = 1$ degree of freedom. The 95 percent critical value from the chi-squared table is 3.84. Therefore, we reject the null hypothesis of no correlation between Y_t and ε_t.~~
2. ~~Using the Wu statistic based on (12-10), we regress C_t on a constant, Y_t, and the predicted value in a regression of Y_t on a constant, Y_{t-1} and C_{t-1}. The t ratio on the prediction is 2.968, so the F statistic with 1 and 201 degrees of freedom is 8.809. The critical value for this F distribution is 4.15, so, again, the null hypothesis is rejected.~~

200

3.888

8 ~~12~~.5 MEASUREMENT ERROR

Thus far, it has been assumed (at least implicitly) that the data used to estimate the parameters of our models are true measurements on their theoretical counterparts. In practice, this situation happens only in the best of circumstances. All sorts of measurement problems creep into the data that must be used in our analyses. Even carefully constructed survey data do not always conform exactly to the variables the analysts have in mind for their regressions. Aggregate statistics such as GDP are only estimates of their theoretical counterparts, and some variables, such as depreciation, the services of capital, and "the interest rate," do not even exist in an agreed-upon theory. At worst, there may be no physical measure corresponding to the variable in our model; intelligence, education, and permanent income are but a few examples. Nonetheless, they all have appeared in very precisely defined regression models.

8 ~~12~~.5.1 LEAST SQUARES ATTENUATION

In this section, we examine some of the received results on regression analysis with badly measured data. The general assessment of the problem is not particularly optimistic. The biases introduced by measurement error can be rather severe. There are almost no known finite-sample results for the models of measurement error; nearly all the results that have been developed are asymptotic.~~[2]~~[3] The following presentation will use a few simple asymptotic results for the classical regression model.

FN 3

The simplest case to analyze is that of a regression model with a single regressor and no constant term. Although this case is admittedly unrealistic, it illustrates the essential concepts, and we shall generalize it presently. Assume that the model,

$$y^* = \beta x^* + \varepsilon, \qquad \text{8-14 } (\cancel{12\text{-}11})$$

conforms to all the assumptions of the classical normal regression model. If data on y^* and x^* were available, then β would be estimable by least squares. Suppose, however, that the observed data are only imperfectly measured versions of y^* and x^*. In the context of an example, suppose that y^* is ln(output/labor) and x^* is ln(capital/labor). Neither factor input can be measured with precision, so the observed y and x contain

[3] ~~[2]~~ See, for example, Imbens and Hyslop (2001).

8-29

errors of measurement. We assume that

$$y = y^* + v \quad \text{with } v \sim N[0, \sigma_v^2],$$

8-15a ~~(12-12a)~~

$$x = x^* + u \quad \text{with } u \sim N[0, \sigma_u^2].$$

~~(12-12b)~~ 8-15b

Assume, as well, that u and v are independent of each other and of y^* and x^*. (As we shall see, adding these restrictions is not sufficient to rescue a bad situation.)

As a first step, insert (~~12-12a~~) into (~~12-11~~), assuming for the moment that only y^* is measured with error: 8-15a 8-14

$$y = \beta x^* + \varepsilon + v = \beta x^* + \varepsilon'.$$

This result conforms to the assumptions of the classical regression model. As long as the regressor is measured properly, measurement error on the dependent variable can be absorbed in the disturbance of the regression and ignored. To save some cumbersome notation, therefore, we shall henceforth assume that the measurement error problems concern only the independent variables in the model. 8-15b

Consider, then, the regression of y on the observed x. By substituting (~~12-12b~~) into (~~12-11~~), we obtain
8-14

$$y = \beta x + [\varepsilon - \beta u] = \beta x + w.$$

8-16 ~~(12-13)~~

Because x equals $x^* + u$, the regressor in (~~12-13~~) 8-16 is correlated with the disturbance:

$$\text{Cov}[x, w] = \text{Cov}[x^* + u, \varepsilon - \beta u] = -\beta\sigma_u^2.$$

8-17 ~~(12-14)~~

This result violates one of the central assumptions of the classical model, so we can expect the least squares estimator,

$$b = \frac{(1/n)\sum_{i=1}^n x_i y_i}{(1/n)\sum_{i=1}^n x_i^2},$$

to be inconsistent. To find the probability limits, insert (~~12-11~~) 8-14 and (~~12-12b~~) 8-15b and use the Slutsky theorem:

$$\text{plim}\, b = \frac{\text{plim}(1/n)\sum_{i=1}^n (x_i^* + u_i)(\beta x_i^* + \varepsilon_i)}{\text{plim}(1/n)\sum_{i=1}^n (x_i^* + u_i)^2}.$$

Because x^*, ε, and u are mutually independent, this equation reduces to

$$\text{plim}\, b = \frac{\beta Q^*}{Q^* + \sigma_u^2} = \frac{\beta}{1 + \sigma_u^2/Q^*},$$

8-18 ~~(12-15)~~

where $Q^* = \text{plim}(1/n)\sum_i x_i^{*2}$. As long as σ_u^2 is positive, b is inconsistent, with a persistent bias toward zero. Clearly, the greater the variability in the measurement error, the worse the bias. The effect of biasing the coefficient toward zero is called **attenuation.** KT

AQ: KT "attenuation" not in chap. list

In a multiple regression model, matters only get worse. Suppose, to begin, we assume that $\mathbf{y} = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and $\mathbf{X} = \mathbf{X}^* + \mathbf{U}$, allowing every observation on every variable to be measured with error. The extension of the earlier result is

$$\text{plim}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right) = \mathbf{Q}^* + \boldsymbol{\Sigma}_{uu}, \quad \text{and} \quad \text{plim}\left(\frac{\mathbf{X}'\mathbf{y}}{n}\right) = \mathbf{Q}^*\boldsymbol{\beta}.$$

8-30

Hence,

$$\text{plim}\,\mathbf{b} = [\mathbf{Q}^* + \boldsymbol{\Sigma}_{uu}]^{-1}\mathbf{Q}^*\boldsymbol{\beta} = \boldsymbol{\beta} - [\mathbf{Q}^* + \boldsymbol{\Sigma}_{uu}]^{-1}\boldsymbol{\Sigma}_{uu}\boldsymbol{\beta}.$$ 8-19 ~~(12-16)~~

This probability limit is a mixture of all the parameters in the model. In the same fashion as before, bringing in outside information could lead to **identification.** KT The amount of information necessary is extremely large, however, and this approach is not particularly promising.

It is common for only a single variable to be measured with error. One might speculate that the problems would be isolated to the single coefficient. Unfortunately, this situation is not the case. For a single bad variable—assume that it is the first—the matrix $\boldsymbol{\Sigma}_{uu}$ is of the form

$$\boldsymbol{\Sigma}_{uu} = \begin{bmatrix} \sigma_u^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

It can be shown that for this special case,

$$\text{plim}\,b_1 = \frac{\beta_1}{1 + \sigma_u^2 q^{*11}}$$ 8-20a ~~(12-17a)~~

[note the similarity of this result to ~~(12-15)~~ 8-18], and, for $k \neq 1$,

$$\text{plim}\,b_k = \beta_k - \beta_1\left[\frac{\sigma_u^2 q^{*k1}}{1 + \sigma_u^2 q^{*11}}\right],$$ 8-20b ~~(12-17b)~~

FN 4

FN 5

FN 6

where q^{*k1} is the $(k, 1)$th element in $(\mathbf{Q}^*)^{-1}$.~~[3]~~[4] This result depends on several unknowns and cannot be estimated. The coefficient on the badly measured variable is still biased toward zero. The other coefficients are all biased as well, although in unknown directions. A badly measured variable contaminates all the least squares estimates.~~[4]~~[5] If more than one variable is measured with error, there is very little that can be said.~~[5]~~[6] Although expressions can be derived for the biases in a few of these cases, they generally depend on numerous parameters whose signs and magnitudes are unknown and, presumably, unknowable.

~~12~~8.5.2 INSTRUMENTAL VARIABLES ESTIMATION

An alternative set of results for estimation in this model (and numerous others) is built around the method of instrumental variables. Consider once again the errors in variables model in ~~(12-11)~~ 8-14 and ~~(12-12a,b)~~ 8-15a,b. The parameters, β, σ_ε^2, q^*, and σ_u^2 are not identified in terms of the moments of x and y. Suppose, however, that there exists a variable z such that z is correlated with x^* but not with u. For example, in surveys of families, income is notoriously badly reported, partly deliberately and partly because respondents often

~~[3]~~[4] Use (A-66) to invert $[\mathbf{Q}^* + \boldsymbol{\Sigma}_{uu}] = [\mathbf{Q}^* + (\sigma_u\mathbf{e}_1)(\sigma_u\mathbf{e}_1)']$, where $\mathbf{e}_1$ is the first column of a $K \times K$ identity matrix. The remaining results are then straightforward.

~~[4]~~[5] This point is important to remember when the presence of measurement error is suspected.

~~[5]~~[6] Some firm analytic results have been obtained by Levi (1973), Theil (1961), Klepper and Leamer (1983), Garber and Klepper (1980), Griliches (1986), and Cragg (1997).

8-31

neglect some minor sources. Suppose, however, that one could determine the total amount of checks written by the head(s) of the household. It is quite likely that this z would be highly correlated with income, but perhaps not significantly correlated with the errors of measurement. If Cov[x^*, z] is not zero, then the parameters of the model become estimable, as

$$\text{plim}\,\frac{(1/n)\sum_i y_i z_i}{(1/n)\sum_i x_i z_i} = \frac{\beta\,\text{Cov}[x^*, z]}{\text{Cov}[x^*, z]} = \beta. \qquad \begin{array}{c}8\text{-}21\\ \cancel{(12\text{-}18)}\end{array}$$

For the general case, $\mathbf{y} = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\mathbf{X} = \mathbf{X}^* + \mathbf{U}$, suppose that there exists a matrix of variables $\mathbf{Z}$ that is not correlated with the disturbances or the measurement error but is correlated with regressors, $\mathbf{X}$. Then the instrumental variables estimator based on $\mathbf{Z}$, $\mathbf{b}_{\text{IV}} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$, is consistent and asymptotically normally distributed with asymptotic covariance matrix that is estimated with

$$\text{Est. Asy. Var}[\mathbf{b}_{\text{IV}}] = \hat{\sigma}^2[\mathbf{Z}'\mathbf{X}]^{-1}[\mathbf{Z}'\mathbf{Z}][\mathbf{X}'\mathbf{Z}]^{-1}. \qquad \begin{array}{c}8\text{-}22\\ \cancel{(12\text{-}19)}\end{array}$$

For more general cases, Theorem ~~12.1~~ 8.1 and the results in Section ~~12.3~~ 8.3 apply.

~~12~~ 8.5.3 PROXY VARIABLES

In some situations, a variable in a model simply has no observable counterpart. Education, intelligence, ability, and like factors are perhaps the most common examples. In this instance, unless there is some observable indicator for the variable, the model will have to be treated in the framework of missing variables. Usually, however, such an indicator can be obtained; for the factors just given, years of schooling and test scores of various sorts are familiar examples. The usual treatment of such variables is in the measurement error framework. If, for example,

$$\text{income} = \beta_1 + \beta_2\ \text{education} + \varepsilon$$

and

$$\text{years of schooling} = \text{education} + u,$$

then the model of Section ~~12~~ 8.5.1 applies. The only difference here is that the true variable in the model is "latent." No amount of improvement in reporting or measurement would bring the proxy closer to the variable for which it is proxying.

KT

The preceding is a pessimistic assessment, perhaps more so than necessary. Consider a **structural model,**

AQ: KT "structural model" not in chap. list

$$\textit{Earnings} = \beta_1 + \beta_2\ \textit{Experience} + \beta_3\ \textit{Industry} + \beta_4\ \textit{Ability} + \varepsilon.$$

Ability is unobserved, but suppose that an indicator, say, *IQ*, is. If we suppose that *IQ* is related to *Ability* through a relationship such as

$$\textit{IQ} = \alpha_1 + \alpha_2\ \textit{Ability} + v,$$

8-32

AU: Term "reduced form equation" already a KT in chap. Here also?

then we may solve the second equation for *Ability* and insert it in the first to obtain the **reduced form equation**

$$Earnings = (\beta_1 - \beta_4\alpha_1/\alpha_2) + \beta_2\ Experience + \beta_3\ Industry + (\beta_4/\alpha_2)IQ + (\varepsilon - v\beta_4/\alpha_2).$$

This equation is intrinsically linear and can be estimated by least squares. We do not have consistent estimators of β_1 and β_4, but we do have them for the coefficients of interest, β_2 and β_3. This would appear to "solve" the problem. We should note the essential ingredients; we require that the **indicator,** *IQ*, not be related to the other variables in the model, and we also require that v not be correlated with any of the variables. In this instance, some of the parameters of the structural model are identified in terms of observable data. Note, though, that *IQ* is not a proxy variable; it is an indicator of the latent variable, *Ability*. This form of modeling has figured prominently in the education and educational psychology literature. Consider, in the preceding small model how one might proceed with not just a single indicator, but say with a battery of test scores, all of which are indicators of the same latent ability variable.

KT

It is to be emphasized that a proxy variable is not an instrument (or the reverse). Thus, in the instrumental variables framework, it is implied that we do not regress **y** on **Z** to obtain the estimates. To take an extreme example, suppose that the full model was

$$\mathbf{y} = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$
$$\mathbf{X} = \mathbf{X}^* + \mathbf{U},$$
$$\mathbf{Z} = \mathbf{X}^* + \mathbf{W}.$$

That is, we happen to have two badly measured estimates of $\mathbf{X}^*$. The parameters of this model can be estimated without difficulty if **W** is uncorrelated with **U** and $\mathbf{X}^*$, *but not by regressing* **y** *on* **Z**. The instrumental variables technique is called for.

When the model contains a variable such as education or ability, the question that naturally arises is; If interest centers on the other coefficients in the model, why not just discard the problem variable?~~[6]~~[7] This method produces the familiar problem of an omitted variable, compounded by the least squares estimator in the full model being inconsistent anyway. Which estimator is worse? McCallum (1972) and Wickens (1972) show that the asymptotic bias (actually, degree of inconsistency) is worse if the proxy is omitted, even if it is a bad one (has a high proportion of measurement error). This proposition neglects, however, the precision of the estimates. Aigner (1974) analyzed this aspect of the problem and found, as might be expected, that it could go either way. He concluded, however, that "there is evidence to broadly support use of the proxy."

FN 7 7→ ✓ ✓✓ ✓

Example ~~12.5~~ 8.9 Income and Education in a Study of Twins

The traditional model used in labor economics to study the effect of education on income is an equation of the form

$$y_i = \beta_1 + \beta_2\ \text{age}_i + \beta_3\ \text{age}_i^2 + \beta_4\ \text{education}_i + \mathbf{x}_i'\boldsymbol{\beta}_5 + \varepsilon_i,$$

where y_i is typically a wage or yearly income (perhaps in log form) and $\mathbf{x}_i$ contains other variables, such as an indicator for sex, region of the country, and industry. The literature

7 ✓ ~~[6]~~ This discussion applies to the measurement error and latent variable problems equally.

contains discussion of many possible problems in estimation of such an equation by least squares using measured data. Two of them are of interest here:

1. Although "education" is the variable that appears in the equation, the data available to researchers usually include only "years of schooling." This variable is a proxy for education, so an equation fit in this form will be tainted by this problem of measurement error. Perhaps surprisingly so, researchers also find that reported data on years of schooling are themselves subject to error, so there is a second source of measurement error. For the present, we will not consider the first (much more difficult) problem.
2. Other variables, such as "ability"—we denote these μ_i—will also affect income and are surely correlated with education. If the earnings equation is estimated in the form shown above, then the estimates will be further biased by the absence of this "omitted variable." For reasons we will explore in Chapter 24, this bias has been called the **selectivity effect** in recent studies.

KT

Simple cross-section studies will be considerably hampered by these problems. But, in a ~~recent study~~ study of twins, Ashenfelter and Kreuger (1994) analyzed a data set that allowed them, with a few simple assumptions, to ameliorate these problems.[8]

FN 8

Annual "twins festivals" are held at many places in the United States. The largest is held in Twinsburg, Ohio. The authors interviewed about 500 individuals over the age of 18 at the August 1991 festival. Using pairs of twins as their observations enabled them to modify their model as follows: Let (y_{ij}, A_{ij}) denote the earnings and age for twin j, $j = 1, 2$, for pair i. For the education variable, only self-reported "schooling" data, S_{ij}, are available. The authors approached the measurement problem in the schooling variable, S_{ij}, by asking each twin how much schooling they had and how much schooling their sibling had. Denote reported schooling *by* sibling *m of* sibling *j* by $S_{ij}(m)$. So, the self-reported years of schooling of twin 1 is $S_{i1}(1)$. When asked how much schooling twin 1 has, twin 2 reports $S_{i1}(2)$. The measurement error model for the schooling variable is

$$S_{ij}(m) = S_{ij} + u_{ij}(m), \quad j, m = 1, 2, \text{ where } S_{ij} = \text{“true” schooling for twin } j \text{ of pair } i.$$

We assume that the two sources of measurement error, $u_{ij}(m)$, are uncorrelated and they and S_{ij} have zero means. Now, consider a simple bivariate model such as the one in (12-11):

Au: Provide correct x-ref equation number

$$y_{ij} = \beta S_{ij} + \varepsilon_{ij}.$$

As we saw earlier, a least squares estimate of β using the reported data will be attenuated:

$$\text{plim } b = \frac{\beta \times \text{Var}[S_{ij}]}{\text{Var}[S_{ij}] + \text{Var}[u_{ij}(j)]} = \beta q.$$

(Because there is no natural distinction between twin 1 and twin 2, the assumption that the variances of the two measurement errors are equal is innocuous.) The factor q is sometimes called the reliability ratio. In this simple model, if the reliability ratio were known, then β could be consistently estimated. In fact, the construction of this model allows just that. Since the two measurement errors are uncorrelated,

$$\text{Corr}[S_{i1}(1), S_{i1}(2)] = \text{Corr}[S_{i2}(1), S_{i2}(2)]$$

$$= \frac{\text{Var}[S_{i1}]}{\{(\text{Var}[S_{i1}] + \text{Var}[u_{i1}(1)]\} \times \{\text{Var}[S_{i1}] + \text{Var}[u_{i1}(2)]\}\}^{1/2}} = q.$$

In words, the correlation between the two reported education attainments measures the reliability ratio. The authors obtained values of 0.920 and 0.877 for 298 pairs of identical twins and 0.869 and 0.951 for 92 pairs of fraternal twins, thus providing a quick assessment of the extent of measurement error in their schooling data.

The earnings equation is a multiple regression, so this result is useful for an overall assessment of the problem, but the numerical values are not sufficient to undo the overall biases in the least squares regression coefficients. An instrumental variables estimator was used

[8] Other studies of twins and siblings include Bound, Cherkas, Haskel, Hawkes and Spector (2003), Ashenfelter and Rouse (1998), Ashenfelter and Zimmerman (1997), Behrman and Rosenzweig (1999), Isaacson (1999), Miller, Mulvey and Martin (1995), Rouse (1999), and Taubman (1976).

8-34

for that purpose. The estimating equation for $y_{ij} = \ln Wage_{ij}$ with the least squares (LS) and instrumental variable (IV) estimates is as follows:

$$\begin{array}{llllllll}
y_{ij} = \beta_1 + & \beta_2\,\text{age}_i + & \beta_3\,\text{age}_i^2 + & \beta_4\,S_{ij}(j) + & \beta_5\,S_{im}(m) + & \beta_6\,\text{sex}_i + & \beta_7\,\text{race}_i + \varepsilon_{ij} \\
\text{LS} & (0.088) & (-0.087) & (0.084) & & (0.204) & (-0.410) \\
\text{IV} & (0.088) & (-0.087) & (0.116) & (-0.037) & (0.206) & (-0.428).
\end{array}$$

In the equation, $S_{ij}(j)$ is the person's report of his or her own years of schooling and $S_{im}(m)$ is the sibling's report of the sibling's own years of schooling. The problem variable is schooling. To obtain a consistent estimator, the method of instrumental variables was used, using each sibling's report of the other sibling's years of schooling as a pair of instrumental variables. The estimates reported by the authors are shown below the equation. (The constant term was not reported, and for reasons not given, the second schooling variable was not included in the equation when estimated by LS.) This preliminary set of results is presented to give a comparison to other results in the literature. The age, schooling, and gender effects are comparable with other received results, whereas the effect of race is vastly different, −40 percent here compared with a typical value of +9 percent in other studies. The effect of using the instrumental variable estimator on the estimates of β_4 is of particular interest. Recall that the reliability ratio was estimated at about 0.9, which suggests that the IV estimate would be roughly 11 percent higher (1/0.9). Because this result is a multiple regression, that estimate is only a crude guide. The estimated effect shown above is closer to 38 percent.

The authors also used a different estimation approach. Recall the issue of selection bias caused by unmeasured effects. The authors reformulated their model as

$$y_{ij} = \beta_1 + \beta_2\,\text{age}_i + \beta_3\,\text{age}_i^2 + \beta_4\,S_{ij}(j) + \beta_6\,\text{sex}_i + \beta_7\,\text{race}_i + \mu_i + \varepsilon_{ij}.$$

Unmeasured latent effects, such as "ability," are contained in μ_i. Because μ_i is not observable but is, it is assumed, correlated with other variables in the equation, the least squares regression of y_{ij} on the other variables produces a biased set of coefficient estimates. [This is a "fixed effects model—See Section 9.4. The assumption that the latent effect, "ability," is common between the twins and fully accounted for is a controversial assumption that ability is accounted for by "nature" rather than "nurture." See, e.g., Behrman and Taubman (1989). A search of the internet on the subject of the "nature versus nurture debate" will turn up millions of citations. We will not visit the subject here.] The difference between the two earnings equations is

$$y_{i1} - y_{i2} = \beta_4[S_{i1}(1) - S_{i2}(2)] + \varepsilon_{i1} - \varepsilon_{i2}.$$

This equation removes the latent effect but, it turns out, worsens the measurement error problem. As before, β_4 can be estimated by instrumental variables. There are two instrumental variables available, $S_{i2}(1)$ and $S_{i1}(2)$. (It is not clear in the paper whether the authors used the two separately or the difference of the two.) The least squares estimate is 0.092, which is comparable to the earlier estimate. The instrumental variable estimate is 0.167, which is nearly 82 percent higher. The two reported standard errors are 0.024 and 0.043, respectively. With these figures, it is possible to carry out Hausman's test;

$$H = \frac{(0.167 - 0.092)^2}{0.043^2 - 0.024^2} = 4.418.$$

The 95 percent critical value from the chi-squared distribution with one degree of freedom is 3.84, so the hypothesis that the LS estimator is consistent would be rejected. (The square root of H, 2.102, would be treated as a value from the standard normal distribution, from which the critical value would be 1.96. The authors reported a t statistic for this regression of 1.97. The source of the difference is unclear.)

8-35

Anderson (1971) or Amemiya (1985) that under very general conditions,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i\varepsilon_i \xrightarrow{d} \mathbf{N}\left[\mathbf{0}, \sigma^2 \text{plim}\left(\frac{1}{n}\mathbf{Z}'\mathbf{\Omega}\mathbf{Z}\right)\right].$$

With the other results already in hand, we now have the following.

THEOREM 12.2 Asymptotic Distribution of the IV Estimator in the Generalized Regression Model

If the regressors and the instrumental variables are well behaved in the fashions just discussed, then

$$\mathbf{b}_{\text{IV}} \overset{a}{\sim} N[\boldsymbol{\beta}, \mathbf{V}_{\text{IV}}],$$

where **(12-21)**

$$\mathbf{V}_{\text{IV}} = \frac{\sigma^2}{n}(\mathbf{Q}_{\mathbf{XX.Z}})\,\text{plim}\left(\frac{1}{n}\mathbf{Z}'\mathbf{\Omega}\mathbf{Z}\right)(\mathbf{Q}'_{\mathbf{XX.Z}}).$$

8-6 ~~12.7~~ NONLINEAR INSTRUMENTAL VARIABLES ESTIMATION

In Section 8.2 ~~12.2~~, we extended the linear regression model to allow for the possibility that the regressors might be correlated with the disturbances. The same problem can arise in nonlinear models. The consumption function estimated in Section 11.3.1 7.2.5 is almost surely a case in point, and we reestimated it using the instrumental variables technique for linear models in Example 12.4. 8.7 In this section, we will extend the method of instrumental variables to nonlinear regression models.

In the nonlinear model,

$$y_i = h(\mathbf{x}_i, \boldsymbol{\beta}) + \varepsilon_i,$$

the covariates $\mathbf{x}_i$ may be correlated with the disturbances. We would expect this effect to be transmitted to the pseudoregressors, $\mathbf{x}_i^0 = \partial h(\mathbf{x}_i, \boldsymbol{\beta})/\partial\boldsymbol{\beta}$. If so, then the results that we derived for the linearized regression would no longer hold. Suppose that there is a set of variables $[\mathbf{z}_1, \ldots, \mathbf{z}_L]$ such that

$$\text{plim}(1/n)\mathbf{Z}'\boldsymbol{\varepsilon} = \mathbf{0} \qquad \text{~~(12-22)~~ } 8\text{-}23$$

and

$$\text{plim}(1/n)\mathbf{Z}'\mathbf{X}^0 = \mathbf{Q}_{zx}^0 \neq \mathbf{0},$$

where $\mathbf{X}^0$ is the matrix of pseudoregressors in the linearized regression, evaluated at the true parameter values. If the analysis that we used for the linear model in Section ~~12.3~~ 8.3 can be applied to this set of variables, then we will be able to construct a consistent estimator for $\boldsymbol{\beta}$ using the instrumental variables. As a first step, we will attempt to replicate the approach that we used for the linear model. The linearized regression

8-36

7-30 model is given in (11-7),

$$\mathbf{y} = \mathbf{h}(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon} \approx \mathbf{h}^0 + \mathbf{X}^0(\boldsymbol{\beta} - \boldsymbol{\beta}^0) + \boldsymbol{\varepsilon}$$

or

$$\mathbf{y}^0 \approx \mathbf{X}^0\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where

$$\mathbf{y}^0 = \mathbf{y} - \mathbf{h}^0 + \mathbf{X}^0\boldsymbol{\beta}^0.$$

For the moment, we neglect the approximation error in linearizing the model. In (~~12-22~~ 8-23), we have assumed that

$$\text{plim}(1/n)\mathbf{Z}'\mathbf{y}^0 = \text{plim}\,(1/n)\mathbf{Z}'\mathbf{X}^0\boldsymbol{\beta}. \qquad \text{8-24 } (\sim\!\!\sim 12\text{-}23 \sim\!\!\sim)$$

Suppose, as we assumed before, that there are the same number of instrumental variables as there are parameters, that is, columns in $\mathbf{X}^0$. (Note: This number need not be the number of variables.) Then the "estimator" used before is suggested:

$$\mathbf{b}_{\text{IV}} = (\mathbf{Z}'\mathbf{X}^0)^{-1}\mathbf{Z}'\mathbf{y}^0. \qquad \text{8-25 } (\sim\!\!\sim 12\text{-}24 \sim\!\!\sim)$$

8-24 The logic is sound, but there is a problem with this estimator. The unknown parameter vector $\boldsymbol{\beta}$ appears on both sides of (~~12-23~~). We might consider the approach we used for our first solution to the nonlinear regression model. That is, with some initial estimator in hand, iterate back and forth between the instrumental variables regression and recomputing the pseudoregressors until the process converges to the fixed point that we seek. Once again, the logic is sound, and in principle, this method does produce the estimator we seek.

If we add to our preceding assumptions

$$\frac{1}{\sqrt{n}}\mathbf{Z}'\boldsymbol{\varepsilon} \xrightarrow{d} N[\mathbf{0}, \sigma^2\mathbf{Q}_{zz}],$$

then we will be able to use the same form of the asymptotic distribution for this estimator that we did for the linear case. Before doing so, we must fill in some gaps in the preceding. First, despite its intuitive appeal, the suggested procedure for finding the estimator is very unlikely to be a good algorithm for locating the estimates. Second, we do not wish to limit ourselves to the case in which we have the same number of instrumental variables as parameters. So, we will consider the problem in general terms. The estimation criterion for nonlinear instrumental variables is a quadratic form,

$$\begin{aligned}\text{Min}_{\boldsymbol{\beta}}\, S(\boldsymbol{\beta}) &= \tfrac{1}{2}\{[\mathbf{y} - \mathbf{h}(\mathbf{X}, \boldsymbol{\beta})]'\mathbf{Z}\}(\mathbf{Z}'\mathbf{Z})^{-1}\{\mathbf{Z}'[\mathbf{y} - \mathbf{h}(\mathbf{X}, \boldsymbol{\beta})]\}\\ &= \tfrac{1}{2}\boldsymbol{\varepsilon}(\boldsymbol{\beta})'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}(\boldsymbol{\beta}).^{\cancel{8}\,9}\end{aligned} \qquad \text{8-26 } (\sim\!\!\sim 12\text{-}25 \sim\!\!\sim)$$

FN 9

[9] ~~8~~ Perhaps the more natural point to begin the minimization would be $S^0(\boldsymbol{\beta}) = [\boldsymbol{\varepsilon}(\boldsymbol{\beta})'\mathbf{Z}][\mathbf{Z}'\boldsymbol{\varepsilon}(\boldsymbol{\beta})]$. We have bypassed this step because the criterion in (~~12-25~~ 8-26) and the estimator in (~~12-26~~ 8-27) will turn out (following and in Chapter ~~15~~ 13) to be a simple yet more efficient GMM estimator.

8 -37

The first-order conditions for minimization of this weighted sum of squares are

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\mathbf{X}^{0\prime}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}(\boldsymbol{\beta}) = \mathbf{0}.$$

8-27 ~~(12-26)~~

This result is the same one we had for the linear model with $\mathbf{X}^0$ in the role of $\mathbf{X}$. This problem, however, is highly nonlinear in most cases, and the repeated least squares approach is unlikely to be effective. But it is a straightforward minimization problem in the frameworks of Appendix E, and instead, we can just treat estimation here as a problem in nonlinear optimization.

We have approached the formulation of this instrumental variables estimator more or less strategically. However, there is a more structured approach. The **orthogonality condition**

$$\text{plim}(1/n)\mathbf{Z}'\boldsymbol{\varepsilon} = \mathbf{0}$$

defines a GMM estimator. With the homoscedasticity and nonautocorrelation assumption, the resultant **minimum distance estimator** produces precisely the criterion function suggested above. We will revisit this estimator in this context, in Chapter ~~15.~~ 13

AU: Term "orthogonality condition" already KT in chap. Here also?

KT

With well-behaved *pseudoregressors* and instrumental variables, we have the general result for the nonlinear instrumental variables estimator; this result is discussed at length in Davidson and MacKinnon (2004).

THEOREM ~~12.3~~ 8.2 Asymptotic Distribution of the Nonlinear Instrumental Variables Estimator

With well-behaved instrumental variables and pseudoregressors,

$$\mathbf{b}_{\text{IV}} \overset{a}{\sim} N\left[\boldsymbol{\beta}, (\sigma^2/n)\left(\mathbf{Q}^0_{\mathbf{xz}}(\mathbf{Q}_{\mathbf{zz}})^{-1}\mathbf{Q}^0_{\mathbf{zx}}\right)^{-1}\right].$$

We estimate the asymptotic covariance matrix with

$$\text{Est. Asy. Var}[\mathbf{b}_{\text{IV}}] = \hat{\sigma}^2[\hat{\mathbf{X}}^{0\prime}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{X}}^0]^{-1},$$

where $\hat{\mathbf{X}}^0$ *is* $\mathbf{X}^0$ *computed using* $\mathbf{b}_{\text{IV}}$.

As a final observation, note that the "two-stage least squares" interpretation of the instrumental variables estimator for the linear model still applies here, with respect to the IV estimator. That is, at the final estimates, the first-order conditions (normal equations) imply that

$$\mathbf{X}^{0\prime}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = \mathbf{X}^{0\prime}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}^0\boldsymbol{\beta},$$

which says that the estimates satisfy the normal equations for a linear regression of $\mathbf{y}$ (not $\mathbf{y}^0$) on the predictions obtained by regressing the columns of $\mathbf{X}^0$ on $\mathbf{Z}$. The interpretation is not quite the same here, because to compute the predictions of $\mathbf{X}^0$, we must have the estimate of $\boldsymbol{\beta}$ in hand. Thus, this two-stage least squares approach does not show *how to compute* $\mathbf{b}_{\text{IV}}$; it shows a characteristic of $\mathbf{b}_{\text{IV}}$.

8 - 38

8

TABLE ~~1~~2.2 Nonlinear Least Squares and Instrumental Variable Estimates

	Instrumental Variables		Least Squares	
Parameter	*Estimate*	*Standard Error*	*Estimate*	*Standard Error*
α	627.031	26.6063	468.215	22.788
β	0.040291	0.006050	0.0971598	0.01064
γ	1.34738	0.016816	1.24892	0.1220
σ	57.1681	—	49.87998	—
$\mathbf{e}'\mathbf{e}$	650,369.805	—	495,114.490	—

8.10

Example ~~12.6~~ Instrumental Variables Estimates of the ~~Consumption Function~~

7.2.5 The consumption function in Section 11.3.1 was estimated by nonlinear least squares without accounting for the nature of the data that would certainly induce correlation between $\mathbf{X}^0$ and ε. As we did earlier, we will reestimate this model using the technique of instrumental variables. For this application, we will use the one-period lagged value of consumption and one- and two-period lagged values of income as instrumental variables estimates. Table 12.2 reports the nonlinear least squares and instrumental variables estimates. Because we are using two periods of lagged values, two observations are lost. Thus, the least squares estimates are not the same as those reported earlier. 8.2

12
8.2

The instrumental variable estimates differ considerably from the least squares estimates. The differences can be deceiving, however. Recall that the MPC in the model is $\beta\gamma Y^{\gamma-1}$. The 2000.4 value for *DPI* that we examined earlier was 6634.9. At this value, the instrumental variables and least squares estimates of the MPC are 1.1543 with an estimated standard error of 0.01234 and 1.08406 with an estimated standard error of 0.008694, respectively. These values do differ a bit but less than the quite large differences in the parameters might have led one to expect. We do note that the IV estimate is considerably greater than the estimate in the linear model, 0.9217 (and greater than one, which seems a bit implausible).

~~12.8 PANEL DATA APPLICATIONS~~

~~Recent **panel data** applications have relied heavily on the methods of instrumental variables that we are developing here. We will develop this methodology in detail in Chapter 15 where we consider generalized method of moments (GMM) estimation. At this point, we can examine two major building blocks in this set of methods, Hausman and Taylor's (1981) estimator for the random effects model and Bhargava and Sargan's (1983) proposals for estimating a dynamic panel data model. These two tools play a significant role in the GMM estimators of dynamic panel models in Chapter 15.~~

~~12.8.1 INSTRUMENTAL VARIABLES ESTIMATION OF THE RANDOM EFFECTS MODEL—THE HAUSMAN AND TAYLOR ESTIMATOR~~

~~Recall the original specification of the linear model for panel data in (9-1):~~

$$y_{it} = \mathbf{x}_{it}'\beta + \mathbf{z}_i'\alpha + \varepsilon_{it}. \qquad \textbf{(12-27)}$$

~~The random effects model is based on the assumption that the unobserved person-specific effects, $\mathbf{z}_i$, are uncorrelated with the included variables, $\mathbf{x}_{it}$. This assumption is a major shortcoming of the model. However, the random effects treatment does allow the model to contain observed time invariant characteristics, such as demographic~~

8.7 ~~12.9~~ WEAK INSTRUMENTS

Our analysis thus far has focused on the "identification" condition for IV estimation, that is, the "exogeneity assumption," AI9, which produces

$$\text{plim}\,(1/n)\mathbf{Z}'\boldsymbol{\varepsilon} = \mathbf{0}. \qquad \text{8-28 } \sout{(12-46)}$$

Taking the "relevance" assumption,

$$\text{plim}\,(1/n)\mathbf{Z}'\mathbf{X} = \mathbf{Q}_{\mathbf{ZX}}, \text{ a finite, nonzero, } L \times K \text{ matrix with rank } K, \qquad \text{8-29 } \sout{(12-47)}$$

as given produces a consistent IV estimator. In absolute terms, with 8-28 ~~(12-46)~~ in place, ~~(12-47)~~ is sufficient to assert consistency. As such, researchers have focused on *exogeneity* as the defining problem to be solved in constructing the IV estimator. A growing literature has argued that greater attention needs to be given to the relevance condition. While strictly speaking, ~~(12-47)~~ is indeed sufficient for the asymptotic results we have claimed, the common case of "weak instruments," in which ~~(12-47)~~ is only barely true has attracted considerable scrutiny. In practical terms, instruments are "weak" when they are only slightly correlated with the right-hand-side variables, **X**; that is, $(1/n)\mathbf{Z}'\mathbf{X}$ is *close* to zero. (We will quantify this theoretically when we revisit the issue in Chapter ~~13~~.) Researchers have begun to examine these cases, finding in some an explanation for perverse and contradictory empirical results.~~11~~ 10

8-29

8-29

8-29

10

Superficially, the problem of weak instruments shows up in the asymptotic covariance matrix of the IV estimator,

$$\text{Asy. Var}[\mathbf{b}_{\text{IV}}] = \frac{\sigma_\varepsilon^2}{n}\left[\left(\frac{\mathbf{X}'\mathbf{Z}}{n}\right)\left(\frac{\mathbf{Z}'\mathbf{Z}}{n}\right)^{-1}\left(\frac{\mathbf{Z}'\mathbf{X}}{n}\right)\right]^{-1},$$

which will be "large" when the instruments are weak, and, other things equal, larger the weaker they are. However, the problems run deeper than that. Hahn and Hausman (2003) list two implications: (i) the two stage least squares estimator is badly biased toward the ordinary least squares estimator, which is known to be inconsistent, and (ii) the standard first order asymptotics (such as those we have used in the preceding) will not give an accurate framework for statistical inference. Thus, the problem is worse than simply lack of precision. There is also at least some evidence that the issue goes well beyond "small sample problems." [See Bound, Jaeger, and Baker (1995).]

Nelson and Startz (1990a, b) and

Current research offers several prescriptions for detecting weakness in instrumental variables. For a single endogenous variable (**x** that is correlated with $\boldsymbol{\varepsilon}$), the standard approach is based on the first step least squares regression of two-stage least squares. The conventional F statistic for testing the hypothesis that all the coefficients in the regression

$$x_i = \mathbf{z}_i'\boldsymbol{\pi} + v_i$$

Nelson and Startz (1990 b),

are zero is used to test the "hypothesis" that the instruments are weak. An F statistic less than 10 signals the problem. [See Staiger and Stock (1997), and Stock and Watson (2007, Chapter 12) for motivation of this specific test.] When there are more than one

[10] ~~11~~ Important references are Staiger and Stock (1997), Stock, Wright, and Yogo (2002), Hahn and Hausman (2002, 2003), Kleibergen (2002), Stock and Yogo (2005), and Hausman, Stock, and Yogo (2005).

Nelson and Startz (1990a, b),

8-40

endogenous variable in the model, testing each one separately using this test is not sufficient, since collinearity among the variables could impact the result, but would not show up in either test. Shea (1997) proposes a four step multivariate procedure that can be used. Godfrey (1999) derived a surprisingly simple alternative method of doing the computation. For endogenous variable k, the Godfrey statistic is the ratio of the estimated variances of the two estimators, OLS and 2SLS,

$$R_k^2 = \frac{v_k(OLS)/\mathbf{e}'\mathbf{e}(\text{OLS})}{v_k(2SLS)/\mathbf{e}'\mathbf{e}(\text{2SLS})}$$

where $v_k(\text{OLS})$ is the kth diagonal element of $[\mathbf{e}'\mathbf{e}(\text{OLS})/(n-K)](\mathbf{X}'\mathbf{X})^{-1}$ and $v_k(\text{2SLS})$ is defined likewise. With the scalings, the statistic reduces to

$$R_k^2 = \frac{(\mathbf{X}'\mathbf{X})^{kk}}{(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{kk}}$$

where the superscript indicates the element of the inverse matrix. The F statistic can then be based on this measure; $F = [R_k^2/(L-1)]/[(1-R_k^2)/(n-L)]$ assuming that $\mathbf{Z}$ contains a constant term.

It is worth noting that the test for weak instruments is not a specification test, nor is it a constructive test for building the model. Rather, it is a strategy for helping the researcher avoid basing inference on unreliable statistics whose properties are not well represented by the familiar asymptotic results, e.g., distributions under assumed null model specifications. Several extensions are of interest. Other statistical procedures are proposed in Hahn and Hausman (2002) and Kleibergen (2002). We are also interested in cases of more than a single endogenous variable. We will take another look at this issue in Chapter ~~13~~ 10, where we can cast the modeling framework as a simultaneous equations model.

for example/

AU: OK to spell out "e.g." in text?

The stark results of this section call the IV estimator into question. In a fairly narrow circumstance, an alternative estimator is the "moment"-free LIML estimator discussed in the next chapter. Another, perhaps somewhat unappealing, approach is to revert to least squares. The OLS estimator is not without virtue. The asymptotic variance of the OLS estimator

$$\text{Asy. Var}[\mathbf{b}_{\text{LS}}] = (\sigma^2/n)\mathbf{Q}_{\mathbf{XX}}^{-1}$$

is unambiguously smaller than the asymptotic variance of the IV estimator

$$\text{Asy. Var}[\mathbf{b}_{\text{IV}}] = (\sigma^2/n)\left(\mathbf{Q}_{\mathbf{XZ}}\mathbf{Q}_{\mathbf{ZZ}}^{-1}\mathbf{Q}_{\mathbf{ZX}}\right)^{-1}.$$

(The proof is considered in the exercises.) Given the preceding results, it could be far smaller. The OLS estimator is inconsistent, however,

$$\text{plim}\,\mathbf{b}_{\text{LS}} - \boldsymbol{\beta} = \mathbf{Q}_{\mathbf{XX}}^{-1}\boldsymbol{\gamma}$$

[see (~~12~~ 8-4)]. By a mean squared error comparison, it is unclear whether the OLS estimator with

$$M(\mathbf{b}_{\text{LS}}\,|\,\boldsymbol{\beta}) = (\sigma^2/n)\mathbf{Q}_{\mathbf{XX}}^{-1} + \mathbf{Q}_{\mathbf{XX}}^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}'\mathbf{Q}_{\mathbf{XX}}^{-1},$$

or the IV estimator, with

$$M(\mathbf{b}_{\text{IV}}\,|\,\boldsymbol{\beta}) = (\sigma^2/n)\left(\mathbf{Q}_{\mathbf{XZ}}\mathbf{Q}_{\mathbf{ZZ}}^{-1}\mathbf{Q}_{\mathbf{ZX}}\right)^{-1},$$

8 - 41

is more precise. The natural recourse in the face of weak instruments is to drop the endogenous variable from the model or improve the instrument set. Each of these is a specification issue. Strictly in terms of estimation strategy within the framework of the data and specification in hand, there is scope for OLS to be the preferred strategy.

12.10 SUMMARY AND CONCLUSIONS

The instrumental variable (IV) estimator, in various forms, is among the most fundamental tools in econometrics. Broadly interpreted, it encompasses most of the estimation methods that we will examine in this book. This chapter has developed the basic results for IV estimation of linear models. The essential departure point is the exogeneity and relevance assumptions that define an instrumental variable. We then analyzed linear IV estimation in the form of the two-stage least squares estimator. With only a few special exceptions related to simultaneous equations models with two variables, almost no finite sample properties have been established for the IV estimator. (We temper that, however, with the results in Section 12.9 on weak instruments, where we saw evidence that whatever the finite sample properties of the IV estimator might be, under some well-discernible circumstances, these properties are not attractive.) We then examined the asymptotic properties of the IV estimator for linear and nonlinear regression models. Two important applications of the IV estimator are Hausman and Taylor's (1981) method of fitting the random effects model with endogenous regressors and Anderson and Hsiao's (1981) and Arellano and Bond's (1991) strategies for fitting dynamic panel data models. Finally, some cautionary notes about using IV estimators when the instruments are only weakly relevant in the model are examined in Section 12.9.

Key Terms and Concepts

- Anderson and Hsiao estimator
- Arellano and Bond estimator
- Asymptotic distribution
- Asymptotic covariance matrix
- Attenuation
- Consistent estimator
- Dynamic panel data model
- Exogeneity
- Feasible GLS
- Generalized regression model
- Hausman and Taylor's estimator
- Hausman's specification test
- Identification
- Indicator
- Instrumental variables
- Instrumental variable estimator
- Limiting distribution
- Minimum distance estimator
- Orthogonality condition
- Panel data
- Proxy variable
- Relevance
- Reliability ratio
- Reduced form equation
- Selectivity effect
- Specification test
- Structural model
- Two-stage least squares
- Variable addition test
- Weak instruments
- Wu test

Exercises

1. In the discussion of the instrumental variable estimator, we showed that the least squares estimator, $\mathbf{b}_{LS}$, is biased and inconsistent. Nonetheless, $\mathbf{b}_{LS}$ does estimate something—see (12-4). Derive the asymptotic covariance matrix of $\mathbf{b}_{LS}$ and show that $\mathbf{b}_{LS}$ is asymptotically normally distributed.

8-42

8

8.9 NATURAL EXPERIMENTS AND THE SEARCH FOR CAUSAL EFFECTS

Econometrics and statistics have historically been taught, understood and operated under the credo that "correlation is not causation." But, much of the still growing field of microeconometrics, and some of what we have done in this chapter, have been advanced as "causal modeling."[11] In the contemporary literature on treatment effects and program evaluation, the point of the econometric exercise really is to establish more than mere statistical association – in short, the answer to the question "does the program *work*?" requires an econometric response more committed than "the data seem to be consistent with that hypothesis." A cautious approach to econometric modeling has nonetheless continued to base its view of "causality" essentially on statistical grounds.[12]

FN 11

FN 12

An example of the sort of causal model considered here is a structural equation such as Krueger and Dale's (1999) model for earnings attainment and elite college attendance,

$$\ln Earnings = \mathbf{x}'\beta + \delta T + \varepsilon,$$

in which δ is the "causal effect" of attendance at an elite college. In this model, T cannot vary autonomously, outside the model. Variation in T is determined partly by the same hidden influences that determine lifetime earnings. Though a causal effect can be attributed to T, measurement of that effect, δ, cannot be done with multiple linear regression. The technique of linear instrumental variables estimation has evolved as a mechanism for disentangling causal influences. As does least squares regression, the method of instrumental variables must be defended against the possibility that the underlying statistical relationships uncovered could be due to "something else." But, when the instrument is the outcome of a "natural experiment," true exogeneity is claimed. It is this purity of the result that has fueled the enthusiasm of the most strident advocates of this style of investigation. The power of the method lends an inevitability and stability to the findings. This has produced a willingness of contemporary researchers to step beyond their cautious roots.[13] Example 8.11 describes a recent, controversial contribution to this literature. On the basis of a natural experiment, the authors identify a cause and effect relationship that would have been viewed as beyond the reach of regression modeling under earlier paradigms.[14]

FN 13

FN 14

[11] See, for example, Chapter 2 of Cameron and Trivedi (2005), which is entitled "Causal and Noncausal Models" and, especially, Angrist and Pischke (2009, 2010).

[12] See, among many recent commentaries on this line of inquiry, Heckman and Vytlacil (2007).

[13] See, e.g., Angrist and Pischke (2009, 2010). In reply, Keane (2010, p. 48) opines "What has always bothered me about the "experimentalist" school is the false sense of certainty it conveys. The basic idea is that if we have a "really good instrument," we can come up with "convincing" estimates of "causal effects" that are not "too sensitive to assumptions.""

[14] See the symposium in the Spring, 2010 *Journal of Economic Perspectives*, Angrist and Pischke (2010), Leamer (2010), Sims (2010), Keane (2010), Stock (2010) and Nevo and Whinston (2010).

Note single quotes within double quotes

Angrist, Imbens and Rubin (1996)
Angrist and Krueger (2001) and

Example 8.11. Does Television Cause Autism?

The following is the abstract of economists Waldman, Nicholson and Adilov's (2008) study of autism.[15]

FN 15

> Autism is currently estimated to affect approximately one in every 166 children, yet the cause or causes of the condition are not well understood. One of the current theories concerning the condition is that among a set of children vulnerable to developing the condition because of their underlying genetics, the condition manifests itself when such a child is exposed to a (currently unknown) environmental trigger. In this paper we empirically investigate the hypothesis that early childhood television viewing serves as such a trigger. Using the Bureau of Labor Statistics' American Time Use Survey, we first establish that the amount of television a young child watches is positively related to the amount of precipitation in the child's community. This suggests that, if television is a trigger for autism, then autism should be more prevalent in communities that receive substantial precipitation. We then look at county-level autism data for three states – California, Oregon, and Washington – characterized by high precipitation variability. Employing a variety of tests, we show that in each of the three states (and across all three states when pooled) there is substantial evidence that county autism rates are indeed positively related to county-wide levels of precipitation. In our final set of tests we use California and Pennsylvania data on children born between 1972 and 1989 to show, again consistent with the television as trigger hypothesis, that county autism rates are also positively related to the percentage of households that subscribe to cable television. *Our precipitation tests indicate that just under forty percent of autism diagnoses in the three states studied is the result of television watching due to precipitation, while our cable tests indicate that approximately seventeen percent of the growth in autism in California and Pennsylvania during the 1970s and 1980s is due to the growth of cable television. These findings are consistent with early childhood television viewing being an important trigger for autism*. (Emphasis added.) We also discuss further tests that can be conducted to explore the hypothesis more directly.

The authors add (at page 3), "Although consistent with the hypothesis that early childhood television watching is an important trigger for autism, our first main finding is also consistent with another possibility. Specifically, since precipitation is likely correlated with young children spending more time indoors generally, not just young children watching more television, our first main finding could be due to any indoor toxin. *Therefore, we also employ a second instrumental variable or natural experiment, that is correlated with early childhood television watching but unlikely to be substantially correlated with time spent indoors*." (Emphasis added.) They conclude (on pages 39-40): "Using the results found in Table 3's pooled cross-sectional analysis of California, Oregon, and Washington's county-level autism rates, we find that if early childhood television watching is the sole trigger driving the positive correlation between autism and precipitation then thirty-eight percent of autism diagnoses are due to the incremental television watching due to precipitation."

URL

[15] Extracts from from http://www.johnson.cornell.edu/faculty.profiles/waldman/autism-waldman-nicholson-adilov.pdf.

8-44

FN 16

Waldman, Nicholson and Adilov's (2008)[16] study provoked an intense and widespread response among academics, autism researchers and the public. Whitehouse (2007) surveyed some of the discussion, which touches upon the methodological implications of the search for "causal effects" in econometric research:

EXT.

> "Prof. Waldman's willingness to hazard an opinion on a delicate matter of science reflects the growing ambition of economists -- and also their growing hubris, in the view of critics. Academic economists are increasingly venturing beyond their traditional stomping ground, a wanderlust that has produced some powerful results but also has raised concerns about whether they're sometimes going too far."
>
> "Such debates are likely to grow as economists delve into issues in education, politics, history and even epidemiology. Prof. Waldman's use of precipitation illustrates one of the tools that has emboldened them: the instrumental variable, a statistical method that, by introducing some random or natural influence, helps economists sort out questions of cause and effect. Using the technique, they can create "natural experiments" that seek to approximate the rigor of randomized trials -- the traditional gold standard of medical research.
>
> "Instrumental variables have helped prominent researchers shed light on sensitive topics. Joshua Angrist of the Massachusetts Institute of Technology has studied the cost of war, the University of Chicago's Steven Levitt has examined the effect of adding police on crime, and Harvard's Caroline Hoxby has studied school performance. Their work has played an important role in public-policy debates. But as enthusiasm for the approach has grown, so too have questions. One concern: When economists use one variable as a proxy for another -- rainfall patterns instead of TV viewing, for example -- it's not always clear what the results actually measure. Also, the experiments on their own offer little insight into why one thing affects another. "There's a saying that ignorance is bliss," says James Heckman, an economics professor at the University of Chicago who won a Nobel Prize in 2000 for his work on statistical methods. "I think that characterizes a lot of the enthusiasm for these instruments." Says MIT economist Jerry Hausman, "If your instruments aren't perfect, you could go seriously wrong."

[16] Published as NBER working paper 12632 in 2006.

Example 8.12 Is Season of Birth a Valid Instrument?

Buckles and Hungerman (BH, 2008) list more than 20 studies of long term economic outcomes that use season of birth as an instrumental variable, beginning with one of the earliest and best known papers in the "natural experiments" literature, Angrist and Krueger (1991). The assertion of the validity of season of birth as a proper instrument is that family background is unrelated to season of birth, but it is demonstrably related to long term outcomes such as income and education. The assertion justifies using dummy variables for season of birth as instrumental variables in outcome equations. If, on the other hand, season of birth is correlated with family background, then it will "fail the exclusion restriction in most IV settings where it has been used." (BH, page 2). According to the authors, the randomness of quarter of birth over the population [see, e.g., Kleibergen (2002)] has been taken as a given, without scientific investigation of the claim. Using data from live birth certificates and census data, BH found a numerically modest, but statistically significant relationship between birth dates and family background. They found "women giving birth in the winter look different from other women; they are younger, less educated, and less likely to be married. The fraction of children born to women without a high school degree is about 10 percent higher (2 percentages points) in January than in May... We also document a 10 percent decline in the fraction of children born to teenagers from January to May." Precisely why there should be such a relationship remains uncertain. Researchers differ (of course) on the numerical implications of BH's finding. [See Lahart (2009).] But, the methodological implication of their finding is consistent with Hausman's observation above.

8.9 SUMMARY AND CONCUSIONS

The instrumental variable (IV) estimator, in various forms, is among the most fundamental tools in econometrics. Broadly interpreted, it encompasses most of the estimation methods that we will examine in this book. This chapter has developed the basic results for IV estimation of linear models. The essential departure point is the exogeneity and relevance assumptions that define an instrumental variable. We then analyzed linear IV estimation in the form of the two-stage least squares estimator.With only a few special exceptions related to simultaneous equations models with two variables, almost no finite sample properties have been established for the IV estimator. (We temper that, however, with the results in Section 8.7 on weak instruments, where we saw evidence that whatever the finite sample properties of the IV estimator might be, under some well-discernible circumstances, these properties are not attractive.)We then examined the asymptotic properties of the IV estimator for linear and nonlinear regression models. Finally, some cautionary notes about using IV estimators when the instruments are only weakly relevant in the model are examined in Section 8.7.

Key Terms and Concepts

- Asymptotic covariance matrix
- Attenuation bias
- Attrition bias
- Dynamic panel data model
- Endogenous
- Exogenous
- Identification
- Instrumental variables
- Limiting distribution
- Minimum distance estimator
- Natural experiment
- Omitted parameter heterogeneity
- Omitted variable bias
- Overidentification
- Proxy variable
- Reduced form equation
- Reliability ratio
- Selectivity effect
- Simultaneous equations bias
- Specification test
- Structural equation system
- Survivorship bias
- Two stage least squares (2SLS)
- Weak instruments
- Wu test
- Asymptotic distribution
- Attrition
- Consistent estimator
- Effect of the treatment on the treated
- Endogenous treatment effect
- Hausman statistic
- Indicator
- Instrumental variable estimator
- Measurement error
- Moment equations
- Nonrandom sampling
- Omitted variables
- Orthogonality conditions
- Panel data
- Random effects
- Relevance
- Sample selection bias
- Simultaneous equations
- Smearing
- Strongly exogenous
- Structural specification
- Truncation bias
- Variable addition test
- Weakly exogenous

AO: The following KTs were not bold KTs in chapter: Attrition, Dynamic panel data model, Endogenous treatment effect, Measurement error, Natural experiment, Nonrandom sampling, Omitted parameter heterogeneity, Omitted variables, Panel data, Relevance, Reliability ratio, Simultaneous equations, Specification test, Strongly exogenous Weak instruments, Weakly exogenous Wu test

8-47/End 8

Exercises

1. In the discussion of the instrumental variable estimator, we showed that the least squares estimator, **b**LS, is biased and inconsistent. Nonetheless, **b**LS does estimate something—see (8-4). Derive the asymptotic covariance matrix of **b**LS and show that **b**LS is asymptotically normally distributed.
2. For the measurement error model in (8-14) and (8-15), prove that when only x is measured with error, the squared correlation between y and x is less than that between y^* and x^*. (Note the assumption that $y^* = y$.) Does the same hold true if y^* is also measured with error?
3. Derive the results in (8-20a) and (8-20b) for the measurement error model. Note the hint in footnote 4 in Section 8.5.1 that suggests you use result (A-66) when you need to invert

$$[\mathbf{Q}^* + \Sigma_{uu}] = [\mathbf{Q}^* + (\sigma_u \mathbf{e}_1)(\sigma_u \mathbf{e}_1)'].$$

4. At the end of Section 8.7, it is suggested that the OLS estimator could have a smaller mean squared error than the 2SLS estimator. Using (8-4), the results of Exercise 1, and Theorem 8.1, show that the result will be true if

$$\mathbf{Q}_{\mathbf{XX}} - \mathbf{Q}_{\mathbf{XZ}}\mathbf{Q}_{\mathbf{ZZ}}^{-1}\mathbf{Q}_{\mathbf{ZX}} >> \frac{1}{(\sigma^2/n) + \gamma'\mathbf{Q}_{\mathbf{XX}}^{\overset{\text{minus}}{-1}}\gamma}\gamma\gamma'.$$

How can you verify that this is at least possible? The right-hand-side is a rank one, nonnegative definite matrix.What can be said about the left-hand-side?

5. Consider the linear model $y_i = \alpha + \beta x_i + \varepsilon_i$ in which Cov$[x_i, \varepsilon_i] = \gamma \neq 0$. Let z be an exogenous, relevant instrumental variable for this model. Assume, as well, that z is binary—it takes only values 1 and 0. Show the algebraic forms of the LS estimator and the IV estimator for both α and β.
6. In the discussion of the instrumental variables estimator, we showed that the least squares estimator **b** is biased and inconsistent. Nonetheless, **b** does estimate something: plim $\mathbf{b} = \boldsymbol{\theta} = \boldsymbol{\beta} + \mathbf{Q}^{-1}\boldsymbol{\gamma}$. Derive the asymptotic covariance matrix of **b**, and show that **b** is asymptotically normally distributed.

Application

1. In Example 8.5, we have suggested a model of a labor market. From the "reduced form" equation given first, you can see the full set of variables that appears in the model—that is the "endogenous variables," ln $Wage_{it}$ and Wks_{it} and all other exogenous variables. The labor supply equation suggested next contains these two variables and three of the exogenous variables. From these facts, you can deduce what variables would appear in a labor "demand" equation for ln $Wage_{it}$. Assume (for purpose of our example) that ln $Wage_{it}$ is determined by Wks_{it} and the remaining appropriate exogenous variables. (We should emphasize that this exercise is purely to illustrate the computations—the structure here would not provide a theoretically sound model for labor market equilibrium.)
   a. What is the labor demand equation implied by the preceding?
   b. Estimate the parameters of this equation by OLS and by 2SLS and compare the results. (Ignore the panel nature of the data set. Just pool the data.)
   c. Are the instruments used in this equation relevant? How do you know?