MODULE TWO, PART ONE: ISSUES OF ENDOGENEITY AND INSTRUMENTAL VARIABLES IN ECONOMIC EDUCATION RESEARCH

William E. Becker
Professor of Economics, Indiana University, Bloomington, Indiana, USA
Adjunct Professor of Commerce, University of South Australia, Adelaide, Australia
Research Fellow, Institute for the Study of Labor (IZA), Bonn, Germany
Editor, *Journal of Economic Education*
Editor, *Social Scinece Reseach Network: Economic Research Network Educator*

Federal Reserve Chair Ben Bernanke said that being an economist is like being a mechanic working on an engine while it is running. Economists typically do not have the convenience of random assignment as in laboratory experiments. However, in some situations they can take advantage of random events such as lotteries or nature. In other circumstances, they might be able to produce variables that have desired random components. When this is possible, they can use instrumental variable techniques and two-stage least squares estimation, which is the focus of the Module Two. Part One of Module Two is devoted to the general theoretical issues associated with endogeneity. Module Two, Parts Two, Three and Four provide the methods of instrumental variable estimation using LIMDEP (NLOGIT), STATA and SAS. To get started consider three types of problems for which instrumental variables are employed.[1]

The first problem of concern is omitted variables. When presenting regression results someone invariably proposes that an explanatory variable that is alleged to be relevant but was omitted is correlated with the included regressors. This renders the coefficient estimators of the included but correlated regressors biased and inconsistent. As stated in the introduction to these modules, examples of this can be traced back over one hundred years to a debate between statistician George Yule and economist Arthur Pigou, see Stephen Stigler (1986, pp. 356-357). Recall that Pigou criticized Yule's multiple regression (aimed at explaining the percentage of persons in poverty with the change in the percentage of disabled relief recipients, the percentage change in the proportion of old people, and the percentage change in the population) because it omitted the most important influences: superior program management and restrictive practices, which cannot be measured quantitatively.

Pigou identified the most enduring criticism of regression analysis; namely, the possibility that an unmeasured but relevant variable has been omitted from the regression and that it is this variable that is giving the appearance of a causal relationship between the dependent variable and the included regressors. As described by Michael Finkelstein and Bruce Levin (1990, pp. 363-364 and pp. 409-415), for example, defense attorneys continue to argue that the plaintiff's experts omitted relevant market and productivity variables when they use regression analysis to demonstrate that women are paid less than men. Modern academic journals are packed with articles that argue for one specification of a regression equation versus another for everything from the demand for places in higher education to the learning of economics in the introductory courses.

The second problem is errors in variables. The late Milton Friedman was awarded the Nobel prize in Economics in part because of his path-breaking work in estimating the relationship between consumption and permanent income, which is an unobservable quantity. His work was later applied in unrelated areas such as education research where a student's grade is hypothesized to be a function of his or her effort and ability, which are both unobservable. As we will see, unobserved explanatory variables for which index variables are created give rise to errors-in-variables problems. As seen in the early work of Becker and Salemi (1979), an outstanding example of this in economic education research occurs when the pretest is used as a proxy for existing knowledge, ability or prior understanding.

The third problem is simultaneity. At the aggregate level, estimating a Keynesian consumption function (in which consumption is a function of income) has problems caused by a second equation involving an accounting identity in which aggregate income must equal personal consumption plus other forms of aggregate expenditures. That is, for the nation as a whole there is a simultaneous relationship between income and consumption: consumption is a function of income and income is a function of consumption. Harvard/Stanford University researcher Caroline Hoxby (2000) identified a similar reverse causality problem in her study of the effect of competition among school districts on student performance, as reported in the *Wall Street Journal* (Oct 24, 2005). She hypothesized that more school districts in a community implied more competition and better schools. She also recognized, however, that there could be reverse causality in that a poor school district that could not be closed (because of state regulations, for example) would force politicians (through parental pressure) to start another school district. In economic education, Becker and Johnston (1999) identified a simultaneity problem in trying to explain scores on one type of test (say multiple choice) with scores on another (essay or free response), where causality is bidirectional. Students who score high on either are likely to score high on the other. As we will see, these are problems of simultaneity that involve endogenous regressors.

Omitted variables that are correlated with included explanatory variables, simultaneity and errors in variables are all examples of endogeneity problems for which single equation estimation is not sufficient.

PROBLEMS OF ENDOGENEITY

Put simply, the problem of endogeneity occurs when an explanatory variable is related to the error term in the population model of the data generating process, which causes the ordinary least squares estimators of the relevant model parameters to be biased and inconsistent. More precisely, for the least squared $\mathbf{b}$ vector to be a consistent estimator of the $\boldsymbol{\beta}$ vector in the population data generating model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, the $\mathbf{X'X}$ matrix must be a positive definite matrix (defined by $\mathbf{Q}$, as the sample size n goes to infinity) and there can be no relationship between the vector of population error terms ($\boldsymbol{\varepsilon}$) and the regressors (explanatory variables) in **X**. Mathematically, if

$$\lim_{n\to\infty}\left(\frac{1}{n}\mathbf{X'X}\right)=\mathbf{Q} \text{ (a positive definite matrix) and } p\lim\left(\frac{1}{n}\mathbf{X}'\boldsymbol{\varepsilon}\right)=0,$$

then

$$p\lim\mathbf{b}=\boldsymbol{\beta}+(\mathbf{Q})^{-1}\,p\lim\left(\frac{1}{n}\mathbf{X}'\boldsymbol{\varepsilon}\right)=\boldsymbol{\beta}$$

In words, if observations on the explanatory variables (the *X*s) are unrelated to draws from the error terms (in vector $\boldsymbol{\varepsilon}$), then the sampling distribution of each of the coefficients (the *b*s in **b**) will appear to degenerate to a spike on the relevant Beta, as the sample size increases. In probability limit, a *b* is equal to its β: $p\lim b=\beta$.

But if there is strong correlation between the *X*s and εs, and this correlation does not deteriorate as the sample size goes to infinity, then the least squares estimators are not consistent estimator of Betas and $p\lim\mathbf{b}\neq\boldsymbol{\beta}$. The **b** vector is an inconsistent estimator because of endogenous regressors. That is, the sampling distribution of at least one of the coefficient (one of the *b*s in **b**) will not degenerate to a spike on the relevant Beta, as the sample size continues to increase.

OMITTED VARIABLE

If someone asserts that a regression has omitted variable bias, he or she is saying that the population disturbance is related to an included regressor because a relevant explanatory variable is missing in the estimated regression and its effects must be in the disturbance. This is also known as **unobserved heterogeneity** because the effect of the omitted variable also leads to population error term heterogeneity. The straightforward solution is to include that omitted variable as a regressor, but often data on the missing variable are unavailable. For example, as described in Becker (2004), the U.S. Congressional Advisory Committee on Student Financial Assistance is interested in the functional relationship between the effects of financial variables (e.g., family income, loan availability, and/or grants) and the college-going decision, called persistence and measured by the probability of attending a post-secondary institution, number of post-secondary terms attempted and the like, in linear form:

$$Persistence = f\,(finances, random\ perturbation)$$

The U.S. Department of Education is concerned about getting students “college ready,” as measured by an index reflecting the completion of high school college prep courses, high school grades, SAT scores and the like:

$$Persistence = h(college\,ready, random\,error)$$

Putting the two interests together, where epsilon is the disturbance term, suggests that the appropriate linear model is

$$Persistence = \beta_1 + \beta_2(college\,ready) + \beta_3(finances) + \varepsilon$$

Information on college readiness is obtainable from Department of Education records but matching financial information is more difficult to obtain; thus, a researcher might consider estimating the parameters in

$$Persistence = \lambda_1 + \lambda_2(college\,ready) + u$$

Finances are now in the error term *u*. But students from wealthier families are known to be more college ready than those from less well-off families. Thus, the explanatory variable *college ready* is related to the error term *u*. If estimation is by OLS, bias and inconsistent estimation of λ_2 result:

$$\begin{aligned} \mathrm{E}[(college\,ready)u] &= \beta_3\mathrm{E}[(college\,ready)(finances)] + \mathrm{E}[(college\,ready)\varepsilon] \\ &= \beta_3\mathrm{E}[(college\,ready)(finances)] = \beta_3\mathrm{cov}[(college\,ready)(finances)] \neq 0 \end{aligned}$$

SIMULTANEITY

A classic case of simultaneity can be found in the most basic idea from microeconomics: that the competitive market of supply and demand determines the equilibrium quantity. The market data generating process is thus written as a three equation system:

Supply: $Qs = m + nP + U$

Demand: $Qd = a + bP + cZ + V$

Equilibrium $Q = Qd = Qs$

where *m, n, a, b* and *c* are parameters to be estimated. *P* is price. *Qd* and *Qs* are quantities demanded and supplied, which in equilibrium are equal to *Q*. *Z* is an exogenous variable and *U* and *V* are errors such that

$$\begin{aligned} &E(U) = E(V) = 0,\ E(UV) = 0, \\ &E(U^2) = \sigma_u^2,\ E(Y^2) = \sigma_v^2, \text{ and} \\ &E(VZ) = E(UZ) = 0. \end{aligned}$$

Suppose the supply curve now is to be estimated by OLS from observable market data for which it must be the case that quantity demand equals quantity supplied in equilibrium:

$$Q = m + nP + U.$$

The estimation slope coefficients in the supply equation would obtained as

$$\begin{aligned}\hat{n} &= (P'P)^{-1}P'Q \\ &= n + (P'P)^{-1}P'U.\end{aligned}$$

But from the market structure assumed to be generating the data we know

$$P = \frac{a-m}{n-b} + \frac{c}{n-b}\,Z + \frac{v-u}{n-b} = \beta_0 + \beta_1\,Z + \varepsilon_2\,.$$

Thus, $\mathrm{E}(P'U) \neq 0$

$$E(PU) = E[(\frac{a-m}{n-b})U + (\frac{c}{n-b}\,Z)U + (\frac{V-U}{n-b})U] = E(-\frac{U^2}{n-b}) = -\frac{\sigma_u^2}{n-b}\ .$$

The OLS estimator ($\hat{n}$) is downward biased; that is, the true population parameter is expected to be underestimated by the least squares estimator:

$$E\,(\hat{n}) = n - \frac{\sigma_u^2}{n-b}\,.$$

Next consider an example from macroeconomics in which an aggregate Keynesian consumption function is to be estimated.

$$C\ = A\ +\ BX + U$$

where C is consumption (realized and planned consumption are equal in equilibrium), X is current income and U is the disturbance term. A and B are parameters to be estimated. From the national income accounting rules, we know that

$X = C + V$, where V is other exogenous expenditure .

Thus, $X = (1-B)^{-1}(A + V + U)$. A shock in U causes a shock in X, and U and X are related by the algebra of the data generating process. The B cannot be estimated without bias using least squares.

Consider a third example of simultaneity that is more subtle. Carolyn Hoxby's problem in estimating the relationship between student performance and school competition was algebraically similar to the classic simultaneous equation problem of the Keynesian consumption function but yet quite a bit different in its theoretical origins.

She hypothesized that cities with many school districts provided more opportunity for parents to switch their children in the pursuit of better schools; thus, competition among school districts should lead to better schools as reflected in higher student test scores. Allowing for other explanatory variables, the implied regression is

$$Test\ scores = \beta_1 + \beta_2(number\ \ of\ \ school\ \ districts) + \ldots + \varepsilon \quad .$$

The causal effect of more school districts in a metropolitan area, however, may not be clearly discerned from this regression of mean metropolitan test scores on the number of school districts. Hoxby had anecdotal evidence that economy of scale arguments might lead to two good school districts being merged. At the other extreme, when districts were really bad they could not be merged with others and yet poor performance did not imply that the district would be shut down (it might be taken over by the state) even though a totally new district might be formed. That is, there is reverse causality: bad test performance leads to more districts and good performance leads to fewer.

As a final example of simultaneity, consider the Becker and Johnston (1999) study of the relationship between multiple-choice test and free-response test scores of economics understanding. Although these two form of tests are alleged to measure many different skills, matched scores are known to be highly correlated. Becker and Johnston assert that in part this is because both forms are a function of an unobservable ability that is cause in the error terms u and v in the following system of equations:

$$Multiple\text{-}choice\, score = \beta_1 + \beta_2(Free\text{-}response\, score) + \ldots + u\ .$$

$$Free\text{-}resonse\, score = \lambda_1 + \lambda_2(Multiple\text{-}choice\, score) + \ldots + v\ .$$

This system of equations should make the simultaneity apparent. As discussed in more detail later, the existence of the second equation (where both *u* and *v* include the effect of unobservable ability) makes the free-response test score an endogenous regressor in the first equation. Similarly, the existence of the first equation makes multiple-choice an endogenous regressor in the second.

ERRORS IN VARIABLES

Next consider an "errors in variables" problem that leads to regressor and error term correlation. In particular consider the example in which a student's *grade* on an exam in economics is hypothesized to be a function of *effort* and a random disturbance (*u*):

$$grade = A + B(effort) + u.$$

But effort is not observable (as was also the case for Milton Friedmen's permanent income). What is observable is the number of homework assignments completed, which may be either indicative of or the result of the amount of effort:

$$homework = C(effort) + v\ .$$

The equation to be estimated is then

$$grade = A + (B/C)homework + u^*,\ \text{ where }\ u^* = u - (1/C)v\ .$$

But a shock to v causes a shock to *homework*; thus, *homework* and u^* are correlated and the slope coefficient (B/C) cannot be estimated without bias via least squares.

A SINGLE VARIABLE INSTRUMENT

So what is the solution to these three problems of endogeneity? The instrumental variable (IV) solution is to find something that is highly correlated with the offending regressor but that is not correlated with the error term. In the case of Carolyn Hoxby's problem in estimating the relationship between student performance and school competition,

$$Test\ scores = \beta_1 + \beta_2(number\ \ of\ \ school\ \ districts) + \varepsilon\ ,$$

she observed that areas with a lot of school districts also had a lot of streams, possibly because the streams made natural boundaries for the school districts. She had what is become known as a **natural experiment**.[2] The number of streams was a random event in nature that had nothing to do with the population error term (ε) in the student performance equation but yet was highly related to number of school districts.[3]

For simplicity, ignoring any other variables in the student performance equation and measuring test scores, number of school districts and number of streams in deviation from their respective means, a consistent estimate of the effect of the number of school districts on test scores can be obtained with the **instrumental variable estimator**

$$b_2 = \frac{\sum (dev.\ in\, test\, scores)(dev.\ in\, number\, of\ \ streams)}{\sum\ (dev.\ in\, number\, of\ \ school\, districts)(dev.\ in\ \ number\, of\ \ steams)}\ .$$

To appreciate why the instrumental estimator works, consider the expected value of the terms in the numerator:

$$\begin{aligned}
&E(\ deviations\, in\, test\, score)(deviations\, in\, number\, of\ \ streams)\\
&\quad = E\{[\beta_2(dev.\ \ in\, number\ \ of\ \ school\ \ districts) + \varepsilon\ ](dev.in\ number\ of\, streams)\}\\
&\qquad = \beta_2 \operatorname{cov}(number\, of\ \ school\, districts, number\, of\ \ streams)\ ,
\end{aligned}$$

because the number of streams in an area is a purely random variable unrelated to epsilon.

In this example, deviations in one exogenous variable ($z-\bar{z}$: deviation in number of streams) could be used as an instrument for deviations in an endogenous explanatory variable ($x-\bar{x}$: deviations in number of school districts):

$$b_{IV} = \frac{\sum_{i=1}^{n}(z_i-\bar{z})(y_i-\bar{y})}{\sum_{i=1}^{n}(z_i-\bar{z})(x_i-\bar{x})} \; .$$

As with the OLS estimator, the IV estimator has an asymptotically normal distribution. The IV large sample variance is obtained by

$$s_{b_{IV}}^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2/n}{r_{x,z}^2\sum_{i=1}^{n}(x_i-\bar{x})^2} \; ,$$

where $r_{x,z}^2$ is the coefficient of determination (square of correlation coefficient) for *x* and *z*. Notice, if the correlation between *x* and *z* were perfect, the IV and OLS variance estimators would be the same. On the other hand, if the linear relationship between the *x* and *z* is weak, then the IV variance will greatly exceed that calculated by OLS.

Important to recognize is that a poor instrument is one that has a low $r_{x,z}^2$, causing the standard error of the estimated slope coefficient to be overly large, or has E($Z\varepsilon$) $\neq 0$, implying the Z was in fact endogenous. Unlike OLS estimators, the desired properties of IV estimators are all asymptotic; thus, to refer to small sample statistics like the *t* ratio is not appropriate. The appropriate statistic for testing with b_{IV} is the standard normal:

$$Z \cong \frac{B_{IV}-\beta}{S_{IV}/\sqrt{n}}, \text{ for large } n.$$

It is important that this instrumental variable approach is not restricted to continuous endogenous variables. For example, Angrist (1990) was interested in the lifetime earnings effect of being a Vietnam War veteran. Measuring earnings in logarithmic form, Angrist's model was

$$Ln(earnings) = \beta_1 + \beta_2 veteran + ... + \varepsilon \; ,$$

where *veteran* is one if a veteran of the Vietnam War and zero otherwise. Angrist recognized that there was a sample selection problem (to be discussed in detail in a later module). It is likely that those who expected their earnings to be enhanced by the military experience are the ones who volunteer for service. That is, being a veteran is dependent on earning expectations at the time of joining. To the extent that all the factors that go into these earnings expectations and the decision to join are not captured in this single equation model they are in the epsilon error term. Thus, the error term must be correlated with being a veteran, $\mathrm{E}[(verteran)(\varepsilon)] \neq 0$.

For his instrument, Angrist observed that the lottery used to draft young men provided a natural experiment. Lottery numbers were assigned randomly; thus, the number received would not be correlated with ε. Men receiving lower numbers faced a higher probability of being drafted; thus, lottery numbers are correlated with being a Vietnam vet.

The use of these natural experiments has and likely will continue to be a source of instrumental variables for endogenous explanatory variables. Michael Murray (2006) provided a detailed but easily read review of natural experiments and the use of the IV estimator.

INSTRUMENTAL VARIABLE ESTIMATORS IN GENERAL

Often there are many exogenous variables that could be used as instruments for endogenous variables. Let matrix **Z** contain the set of all the endogenous variables that could serve as an instrument set of regressors. The instrumental variable estimator is now of the general form

$$\mathbf{b}_{IV} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$$
$$\mathbf{Var}(\mathbf{b}_{IV}) = \boldsymbol{\sigma}^2(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{X}'\mathbf{Z})^{-1} \ .$$

Unlike the selective replacement of a regressor with its instrument, for sets of regressors the typical estimation procedure involves the project of each of the columns of **X** in the column space of **Z**; at least conceptually we have

$$\hat{\mathbf{X}} = \mathbf{Z}[(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}] \ .$$

This projected $\hat{\mathbf{X}}$ matrix is then substituted for **Z**.

$$\begin{aligned}\mathbf{b}_{IV} &= (\hat{\mathbf{X}}'\mathbf{X})^{-1}\hat{\mathbf{X}}'\mathbf{y}\\ &= [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}\\ &= [\mathbf{X}'(\mathbf{I}-\mathbf{M}_2)\mathbf{X}]^{-1}\mathbf{X}'(\mathbf{I}-\mathbf{M}_2)\mathbf{y}\\ &= (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{y} \ ,\end{aligned}$$

which suggests a two step process: 1) regress the endogenous regressor(s) on all the exogenous variables; 2) use the predicted values from step 1 as replacement for the endogenous regressor in the original equation. This instrumental variable procedure is referred to as **Two-Stage Least Squares** (TSLS).

Unfortunately the standard errors associated with this TSLS estimation approach do not reflect the fact that the instrument is a combination of variables. That is, the standard errors obtained from the second step do not reflect the number of variables used in the first step predictions. In the case of a single instrument the difference between the variances of OLS and IV estimators was captured in the magnitude of $r^2_{x,z}$ and a similar adjustment must be made when multiple variables are used to form the instruments. Advanced econometrics programs like LIMDEP, SAS and SAS automatically do this in their TSLS programs.

The asymptotic variances correctly calculated can be extremely large if **Z** is not highly correlated with **X**; that is, $(\mathbf{Z'X})^{-1}$ is large if **X** and **Z** are not related. Also, for poor fitting instruments, it is possible to get negative R^2 when the typical computational formula [1 − (ResSS/TotalSS)] is used – recall that least squares minimized the ResSS so that it necessarily is less than or equal to TotalSS. But the IV estimator will have an ResSS greater than or equal to that of least squares. The fit of the IV can be so bad that its ResSS exceeds the Total SS. (For demonstration of this see Becker and Kennedy, 1992.)

DURBIN, HAUSMAN AND WU SPECIFICATION TEST APPLIED TO ENDOGENEITY

We wish to test $p\lim\left(\mathbf{X}'\boldsymbol{\varepsilon}/n\right)=0$, but cannot use the covariance between $n\times K$ matrix **X** and the *n* residuals ($e_i = y_i - \hat{y}_i$) in the $n\times 1$ vector **e** because $\mathbf{X'e}=0$ is a byproduct of least squares. Greene (2003, pp. 80-83) outlined the testing procedure originally proposed by Durbin (1954) and then extended by Wu (1973) and Hausman (1978). Davidson and MacKinnon (1993) are recognized for providing an algebraic demonstration of test statistic equivalence. Asymptotically, a **Wald (*W*) statistic** may be used in a Chi-square (χ^2) test with *K** degrees of freedom, or for smaller samples, an *F* statistic, with *K** and *n* − (*K* + *K**) degrees of freedom, can be used to test the joint significance of the contribution of the predicted values ($\hat{\mathbf{X}}^*$) of a regression of the *K** endogenous regressors, in matrix **X*,** on the exogenous variables (and a column of ones for the constant term) in matrix **Z:**

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \hat{\mathbf{X}}^*\boldsymbol{\gamma} + \boldsymbol{\varepsilon}^*,$$

where $\mathbf{X}^* = \mathbf{Z}\boldsymbol{\lambda} + \mathbf{u}$, $\hat{\mathbf{X}}^* = \mathbf{Z}\hat{\boldsymbol{\lambda}}$, and $\hat{\boldsymbol{\lambda}}$ is a least squares estimator of $\boldsymbol{\lambda}$.

H_o: $\boldsymbol{\gamma} = 0$, the variables in **Z** are exogenous.

H_A: $\boldsymbol{\gamma} \neq 0$, at least one of the variables in **Z** is endogenous.

As an example, consider the previously introduced economic exam grade equation that has the number of homework assignments as an explanatory variables:

$$grade = \beta_1 + \beta_2\, homework + \varepsilon\ .$$

The theoretical data generating process that gave rise to this model suggests that number of homeworks completed is an endogenous regressor. To test this we need truly exogenous variables – say x_2 and x_3 , which might represent student gender and race. The number of homeworks is then regressed on these two exogenous variables to get the least square equation

$$predicted\ homework = \hat{\boldsymbol{\lambda}}_1 + \hat{\boldsymbol{\lambda}}_2 x_2 + \hat{\boldsymbol{\lambda}}_3 x_3\ .$$

This predicted homework variable is then added to the exam grade equation to form the augmented regression

$$grade = \beta_1 + \beta_2\, homework + \boldsymbol{\gamma}\,(predicted\ homework) + \boldsymbol{\varepsilon}^*$$

In this example, $K = 2$ (for β_1 *and* β_2) and $K^* = 1$ (for γ); thus, the degrees of freedom for the F statistic are 1 and $n - (K + K^*)$, which is also the square of a t statistic with $n - (K + K^*)$ degrees of freedom. That is, with only one endogenous variable and relatively small sample n, the t statistic printed by a computer program is sufficient to do the test. (Recall that asymptotically the t goes to the standard normal, with no adjustment for degrees of freedom required.) As with any other F, χ^2 , t or z test, calculated statistics greater than their critical values lead to the rejection of the null hypothesis. Important to keep in mind, however, is that failure to reject the null hypothesis at a specific probability of a Type I error does not prove exogeneity. The null hypothesis can always be rejected at some Type I error level.

Some introductory econometrics textbooks such as Wooldridge (2009, pp. 527-528) specify that the residuals from the auxiliary $\hat{\mathbf{X}}^* = \mathbf{Z}\hat{\boldsymbol{\lambda}}$ regression should be used in the augmented regression $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{X}^* - \hat{\mathbf{X}}^*)\boldsymbol{\gamma} + \boldsymbol{\varepsilon}^*$. For example, in the case of the test scores model the augmented regression would be

$$grade = \beta_1 + \beta_2\, homework + \boldsymbol{\gamma}\,(homework - predicted\ homework) + \boldsymbol{\varepsilon}^*$$

The additional calculation of this residual for inclusion in the augmented regression is not necessary because the absolute value of the estimate of γ and its standard error are identical regardless of whether predicted homework or the residual (= homework − predicted homework) is used.

Finally, keep in mind that you can use all the exogenous variables in the system to predict the endogenous variable. Some of these exogenous variables can even be in the original equation of interest – in the grade example, the grade equation might have been

$$grade = \beta_1 + \beta_2\, homework + \beta_3 x_3 + \varepsilon \ .$$

The auxiliary equation would still be

$$predicted\ homework = \hat{\boldsymbol{\lambda}}_1 + \hat{\boldsymbol{\lambda}}_2 x_2 + \hat{\boldsymbol{\lambda}}_3 x_3 \ .$$

As will become clear in the next section, the auxiliary equation should always have at least one more exogenous variable than the initial equation of interest.

IDENTIFICATION CONDITIONS

Whenever an instrumental variable estimator or two-stage least squares (2SLS) routine is employed consideration must be given to the identification conditions. To understand identification, consider a set of matched price and quantity observations (Figure 1, panel a) for which quantity values tend to rise as prices rise, as seen in the fitted OLS regression (Figure 1, panel b) The question to be asked: is this a supply relationship? As seen in Figure 1, panel c, the OLS line is not a supply curve. It is tracing equilibrium points.[4]

If a supply curve is to be estimated, more information than the observations that the quantity and price are positively related is needed. We need to identify a supply curve. This can be done if there is an exogenous variable that affects demand but does not affect supply. For example, household income likely affects demand but does not affect supply. In our previous simultaneous equation market model, for example,

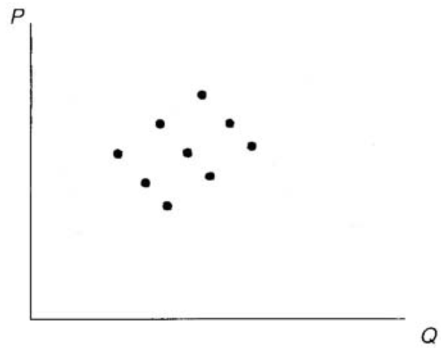
Supply in equilibrium: $Q = m + nP + U$

Demand in equilibrium: $Q = a + bP + cZ + V$

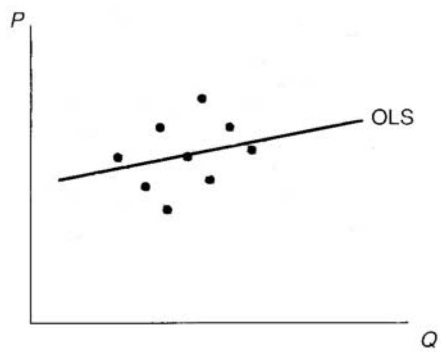
if Z is household income, then an increase in Z shifts the demand curve up, from D to D', but does not affect the supply curve (Figure 2); thus, the supply curve is identified by the change in equilibrium observations. Notice, however, that the demand curve is not identified because there is no unique exogenous variable in the supply equation.

Identification of this supply curve in this two endogenous variable system is achieved by an exclusionary or zero restriction -- the coefficient on income in the supply equation was restricted to zero**.** A necessary order condition for identification of any equation in a system is that the number of exogenous variables excluded from an equation must be at least as great as the number of endogenous variables less one. In this example, there were two endogenous variables (Q and P) and one exogenous variable (Z) excluded from the supply equation; thus, the necessary condition for identification was met: $2-1 \le 1$. This necessary condition for identification is called the **order condition**.

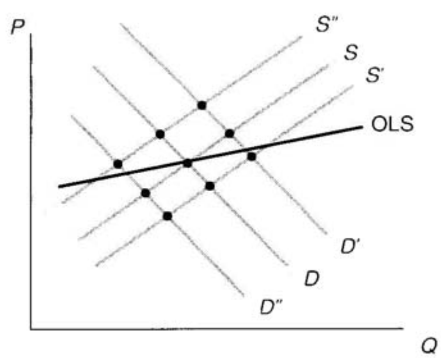
Figure 1. Market data.



Panel a: Scatter plot

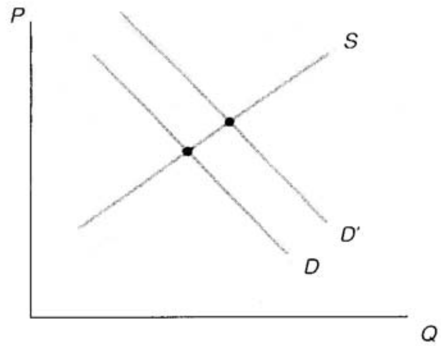


Panel b: OLS regression



Panel c: Demand and supply interaction

Figure 2. Supply curve is identified .



Any exogenous variable that is excluded from at least one equation in an equation system can be used as an instrumental variable. It can be used as an instrument in the equation from which it is excluded. For example, in the supply and demand equation system, the **reduced form** (no endogenous variables as explanatory variables) for P is

$$P = \frac{a-m}{n-b} + \frac{c}{n-b} Z + \frac{v-u}{n-b} = \beta_1 + \beta_2 Z + \varepsilon_2 \ .$$

And either the price predicted from this equation or Z itself can be used as the instrument for P in the supply equation. If there were more exogenous variables excluded from the supply equation then they could all be used to get predicted price from the reduced form equation.

Notice that the coefficient on Z in the reduced form equation for P must be nonzero for Z to be used as an instrument, which requires that $c \neq 0$ and $n - b \neq 0$. This requirement states that exogenous variable(s) excluded from the supply equation must have a nonzero population coefficient in the demand equation and that the effect of price cannot be the same in both demand and supply. This is known as the **rank condition**.

As an example of identification in economic education research consider the work of Becker and Johnston (1999). In addition to the multi-dimensional attributes of the Australian 12th grade test takers (captured in the explanatory X variables such as gender, age, English a second language, etc.), Becker and Johnston called attention to classroom and peer effects that might influence multiple-choice and essay type test taking skills in different ways. For example, if the student is in a classroom that emphasizes skills associated with multiple-choice testing (*e.g.*, risk-taking behavior, question analyzing skills, memorization, and keen sense of judging between close alternatives), then the student can be expected to do better on multiple-choice questions. By the same token, if placed in a classroom that emphasizes the skills of essay test question answering (*e.g.*, organization, good sentence and paragraph construction, obfuscation when uncertain, and logical argument), then the student can be expected to do better on the essay component. Thus, Becker and Johnston attempted to control for the type of class of which the student is a member. Their measure of "teaching to the multiple-choice questions" is the mean score on the multiple-choice questions for the school in which the i^{th} student took the 12^{th} grade economics course. Similarly, the mean school score on the essay questions is their measure of the i^{th} student's exposure to essay question writing skills.

In equation form, the two equations that summarize the influence of the various covariates on multiple-choice and essay test questions are written as the following **structural equations**:

$$M_i = \rho_{21} + \rho_{22} W_i + \rho_{23} \bar{M}_i + \sum_{j=4}^{J} \rho_{2j} X_{ij} + U_i^* \qquad .$$

$$W_i = \rho_{31} + \rho_{32} M_i + \rho_{33} \bar{W}_i + \sum_{j=4}^{J} \rho_{3j} X_{ij} + V_i^* \qquad .$$

M_i and W_i are the i^{th} student's respective scores on the multiple-choice test and essay test. $\bar{M}_i$ and $\bar{W}_i$ are the mean multiple-choice and essay test scores at the school where the i^{th} student took the twelfth grade economics course. The X_{ij} variables are the other exogenous variables used to explain the i^{th} student's multiple choice and essay marks, where the ρs are parameters to be estimated. U_I^* and V_i^* are assumed to be zero mean and constant variance error terms that may or may not each include an effect of unobservable ability.

Least squares estimation of the ρs will involve bias if the respective error terms U_i^* and V_i^* are related to regressors (W_i in the first equation, and M_i in second equation). Such relationships are seen in the **reduced form equations**, which are obtained by solving for M and W in terms of the exogenous variables and the error terms in these two equations:

$$M_i = \Gamma_{21} + \Gamma_{22} \bar{W}_i + \Gamma_{23} \bar{M}_i + \sum_{j=4}^{J} \Gamma_{2j} X_{ij} + U_i^{**} \qquad .$$

$$W_i = \Gamma_{31} + \Gamma_{32} \bar{M}_i + \Gamma_{33} \bar{W}_i + \sum_{j=4}^{J} \Gamma_{3j} X_{ij} + V_i^{**} \qquad .$$

The reduced form parameters (Γs) are functions of the ρs, and U^{**} and V^{**} are dependent on U^* and V^*:

$$U_i^{**} = \frac{U_i^* \;\; + \rho_{22} V_i^*}{1 - \rho_{22}\rho_{32}} \quad .$$

$$V_i^{**} = \frac{V_i^* \;\; + \rho_{32} U_i^*}{1 - \rho_{22}\rho_{32}} \quad .$$

In the reduced form error terms, it can be seen that a random shock in U^* causes a change in V^{**}, which causes a change in W in the reduced form. Thus, W and U^* are related in the essay structural equation, and consistent estimation of the parameters in this equation is not possible using least squares. Similarly, a shock in V^*, and a resulting change in U^{**}yields a change in M. Thus, M and V^* are dependent in the structural equation, and least squares estimators of the parameters in that equation are inconsistent.

The inclusion of $\bar{M}_i$ and $\bar{W}_i$ in their respective structural equations, and their exclusion from the other equation, enables both of the structural equations to be identified within the system. For example, if a student moves from a school with a low average multiple-choice test score to one with a higher average multiple-choice test score, then his or her multiple-choice score will rise via a shift in the *M-W* relationship in the first structural equation, but this shift is associated with a move along the *W-M* relationship in the second structural equation; thus, the second structural equation is identified. Similarly, if a student moves from a low average essay test score school to a higher one, then his or her essay test score will rise via a shift in the *W-M* relationship in second structural equation, but this shift implies a move along the *M-W* relationship in the first structural equation, and this first structural equation is thus identified. Most certainly, identification hinges critically on justifying the exclusionary rule employed.

To summarize, identification involved two conditions.

> **The order condition for identifying an equation in a model of *K* equations and *K* endogenous variables is that the equation exclude at least *K* – 1 variables that appear in the model. Alternatively, if the number of potential instruments (exogenous variables in the system but not in the equation) equals the number of endogenous regressors, the equation is exactly identified. If exactly *K* – 1 variables are excluded, then the equation is just identified. If more (less) than *K* – 1 variables are excluded, then the equation is over (under) identified.**
>
> **The order condition is a necessary condition, but not a sufficient condition for identification.**

The sufficient condition for identification is the rank condition. By the rank condition an equation is identified if and only if at least one nonzero determinant of order exists for the coefficients of the excluded variables that are included in the other equations of the model. This sufficient condition requires that variables excluded from the equation, but included in the other equations of the model, not be dependent. It ensures that the parameters can be estimated from the reduced form.

CONCLUDING COMMENTS

Eagerness to employ natural experiments and instrumental variables to address problems of endogeneity have exploded within economics, but along with that growth has come questions of validity, as seen most recently in criticism of the work of Waldman, Nicholson and Adilov (2006) that suggests that TV watching causes autism. Economist Waldman recognized that he could not simply run a regression of incidence of autism on amount of TV watched because autism might in some way influence the TV watching. He observed, however, that TV watching and precipitation were highly correlated. Because rainfall is a natural occurrence unrelated to the error term in the autism regression, he had his instrument for TV watching. As reported in the *Wall Street Journal*, Whitehouse (2007), those who specialize in the study of autism were not impressed, labeling Waldman's work "irresponsible" (because it shifts responsibilty to parents when experts claim that it is genetic and beyond the control of parent) and "junk science."

When instrumental variables are used, that which is measured is unclear. Unanswered in the Waldman, Nicholson and Adilov study is how TV watching influences autism. Arm-chair speculation that children are distracted by television is not convincing to those who have devoted their lives to studying autism. Joseph Piven, Director of the Neurodevelopment Disorder Research Center at the University of North Carolina, is quoted in the *WSJ* article stating that "it is just too much of a stretch to tie (autism) to television-watching. Why not tie it to carrying umbrellas?" More damning still are the quotes from Nobel Laureate in Economics James Heckman, "There's a saying that ignorance is bliss," and IV econometrician guru Jerry Hausman, "I think that characterizes a lot of the enthusiasm for these instruments. If your instruments aren't perfect, you could go seriously wrong."

REFERENCES

Angrist, Joshua D. (1990). "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records**, "***American Economic Review*," Vol. 80 (June): 313-336.

Angrist, Joshua D. and Alan B. Krueger (2001). "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *Journal of Economic Perspectives*, Vol. 15 (Fall),: 69-85.

Becker, William E. (2004). "Omitted Variables and Sample Selection Problems in Studies of College-Going Decisions," *Public Policy and College Access: Investigating the Federal and State Role in Equalizing Postsecondary Opportunity*, Edward St. John (ed), Vol. 19. NY: AMS Press: 65-86.

Becker, William E. and Carol Johnston (1999)."The Relationship Between Multiple Choice and Essay Response Questions in Assessing Economics Understanding," *Economic Record* (Economic Society of Australia), Vol. 75 (December): 348-357.

Becker, William E. and Peter Kennedy (1995). "A Lesson in Least Squares and R Squared," *American Statistician*, Vol. 55 (November): 282-283. Portions reprinted in Dale Poirier, *Intermediate Statistics and Econometrics* (Cambridge: MIT Press, 1995): 562-563.

Becker, William E. and Michael Salemi (1977). "The Learning and Cost Effectiveness of AVT Supplemented Instruction: Specification of Learning Models," *Journal of Economic Education* Vol. 8 (Spring) : 77-92.

Davidson, Russell and James G. MacKinnon (1993). *Estimation and Inference in Econometrics*. New York, Oxford University Press.

Durbin, James. (1954). "Errors in Variables," *Review of the International Statistical Institute*, Vol. 22(1): 23-32.

Finkelstein, Michael and Bruce Levin (1990). *Statistics for Lawyers*. New York, Springer-Verlag.

Greene, William (2003). *Econometric Analysis*. 5th Edition, New Jersey: Prentice Hall.

Hausman, Jerry (1978). "Specification Tests in Econometrics," *Econometrica*, Vol. 46 (November): 1251-1271.

Hilsenrath, Jon (2005). "Novel Way to Assess School Competition Stirs Academic Row," *Wall Street Journal* (October 24): A1 and A11

Hoxby, Caroline M. (2000). “Does Competition Among Public Schools Benefit Students and Taxpayers?’ *American Economic Review*. Vol. 90 (December): 1209-1238.

Kennedy, Peter (2003). *A Guide to Econometrics*, 5th Edition, United Kingdom: Blackwell.

Murray, Michael P. (2006). “Avoiding Invalid Instruments and Coping with Weak Instruments,” *Journal of Economic Perspectives*. Vol. 20 (Fall): 111-132.

Rosenzweig, Mark and Kenneth Wolpin (2000). "Natural 'Natural Experiments' in Economics," *Journal of Economic Literature*. Vol. 38 (December): 827-874.

Stigler, Stephen (1986). *The History of Statistics: The Measurement of Uncertainty Before 1900*. Cambridge: Harvard University Press.

Waldman, Michael, Sean Nicholson and Nodir Adilov (2006). “Does Television Cause Autism?” NBER Working Paper No. W12632 (October).

Whitehouse, Mark (2007). “Mind and Matter: Is an Economist Qualified to Solve Puzzle of Autism? Professor’s Hypothesis: Rainy Days and TV May Trigger Condition,” *Wall Street Journal*. February 27: A.1.

Wooldridge, Jeffrey M. (2009). *Introductory Econometrics: A Modern Approach.* 4th Edition Mason Oh; South-Western.

Working, E. J. (1927). “What Do Statistical ‘Demand Curves’ Show?” *Quarterly Journal of Economics*, 41(2) : 212-235.

Wu, De-Min. (1973). “Alternative Tests of Independence between Stochastic Regressors and Disturbances,” *Econometrica,* 41(July): 733-750.

ENDNOTES

[1] Conceptually there are more than three forms of endogeneity that could occur. For example, if there is a lagged dependent variable and the residuals are serially correlated, then the lagged dependent variable will be correlated with the error term. This is not a problem for the typical cross-section regressions considered by economic educators but does become a problem when time is introduced. To see this consider a data generating process in which knowledge of economics (Y_{it}) of the i^{th} student at time t is a linear function of the student's ability at time t (x_{it}) plus an error term (ε_{it}):

$$y_{it} = \beta_1 + \beta_2 x_{it} + \varepsilon_{it} \, .$$

At time t-1, knowledge is then given by

$$y_{it-1} = \beta_1 + \beta_2 x_{it-1} + \varepsilon_{it-1} \, .$$

If learning is assessed in the following equation, then the pretest y_{it-1} regressor is endogenous by construction:

$$y_{it} = \beta 1(1-\rho) + \beta_2 (x_{it} - \rho x_{it-1}) + \rho y_{it-1} + (\varepsilon_{it} - \rho \varepsilon_{it-1})$$

$$E[y_{it-1}(\varepsilon_{it} - \rho \varepsilon_{it-1})] = \rho E(y_{it-1}\varepsilon_{it-1}) \neq 0 \, .$$

As demonstrated in a later module, sample selection also leads to endogeneity problems. However, the sample selection form of endogeneity is typically associated with a truncation of the error term, which is a different problem than the three sources of endogeneity considered in the text of this module, where the error term is always assumed to be continuous.

[2] Natural experiments and instrumental variables are not synonymous but Rosenzweig and Wolpin (2000, pp.827-8) state "The most widely applied approach to identifying causal or treatment effects, which has a long history in economics, employs instrumental variable techniques . . .in standard instrumental variable studies, economists as well as researchers in other fields have sought out 'natural experiments,' random treatments that have arisen serendipitously . . ."

[3] Jon Hilsenrath reported in his *Wall Street Journal* (October 24, 2005, pp. A1 and A11) "Novel Way to Assess School Competition Stirs Academic Row," that Princeton University economist Jesse Rothstein questioned Hoxby's use of the instrumental variable technique because he could not replicate her count of streams, which aside from ethical questions posed by Hilsenrath introduces an added complication if her instrument has a measurement error problem.

[4] Working (1927) provided an early intuitive explanation of simultaneity and the identification problems that is still relevant today as seen in its modern rendition by Kennedy (2003).

MODULE TWO, PART TWO: ENDOGENEITY, INSTRUMENTAL VARIABLES AND TWO-STAGE LEAST SQUARES IN ECONOMIC EDUCATION RESEARCH USING LIMDEP

Part Two of Module Two provides a cookbook-type demonstration of the steps required to use LIMDEP to address problems of endogeneity using a two-stage least squares, instrumental variable estimator. The Durbin, Hausman and Wu specification test for endogeneity is also demonstrated. Users of this model need to have completed Module One, Parts One and Two, and Module Two, Part One. That is, from Module One, users are assumed to know how to get data into LIMDEP, recode and create variables within LIMDEP, and run and interpret regression results. From Module Two, Part One, they are expected to have an understanding of the problem of and source of endogeneity and the basic idea behind an instrumental variable approach and the two-stage least squares method. The Becker and Johnston (1999) data set is used throughout this module for demonstration purposes only. Module Two, Parts Three and Four demonstrate in STATA and SAS what is done here in LIMDEP.

THE CASE

As described in Module Two, Part One, Becker and Johnston (1999) called attention to classroom effects that might influence multiple-choice and essay type test taking skills in economics in different ways. For example, if the student is in a classroom that emphasizes skills associated with multiple choice testing (*e.g.*, risk-taking behavior, question analyzing skills, memorization, and keen sense of judging between close alternatives), then the student can be expected to do better on multiple-choice questions. By the same token, if placed in a classroom that emphasizes the skills of essay test question answering (*e.g.*, organization, good sentence and paragraph construction, obfuscation when uncertain, logical argument, and good penmanship), then the student can be expected to do better on the essay component. Thus, Becker and Johnston attempted to control for the type of class of which the student is a member. Their measure of “teaching to the multiple-choice questions” is the mean score or mark on the multiple-choice questions for the school in which the i^{th} student took the 12^{th} grade economics course. Similarly, the mean school mark or score on the essay questions is their measure of the i^{th} student’s exposure to essay question writing skills.

In equation form, the two equations that summarize the influence of the various covariates on multiple-choice and essay test questions are written as the follow structural equations:

$$M_i = \rho_{21} + \rho_{22} W_i + \rho_{23} \bar{M}_i + \sum_{j=4}^{J} \rho_{2j} X_{ij} + U_i^* .$$

$$W_i = \rho_{31} + \rho_{32} M_i + \rho_{33} \bar{W}_i + \sum_{j=4}^{J} \rho_{3j} X_{ij} + V_i^* .$$

M_i and W_i are the i^{th} student's respective scores on the multiple-choice test and essay test. $\bar{M}_i$ and $\bar{W}_i$ are the mean multiple-choice and essay test scores at the school where the i^{th} student took the 12th grade economics course. The X_{ij} variables are the other exogenous variables (such as gender, age, English a second language, etc.) used to explain the i^{th} student's multiple-choice and essay marks, where the ρs are parameters to be estimated. The inclusion of the mean multiple-choice and mean essay test scores in their respective structural equations, and their exclusion from the other equation, enables both of the structural equations to be identified within the system.

As shown in Module Two, Part One, the least squares estimation of the ρs involves bias because the error term U_i^* is related to W_i, in the first equation, and V_i^* is related to M_i, in second equation. Instruments for regressors W_i and M_i are needed. Because the reduced form equations express W_i and M_i solely in terms of exogenous variables, they can be used to generate the respective instruments:

$$M_i = \Gamma_{21} + \Gamma_{22}\bar{W}_i + \Gamma_{23}\bar{M}_i + \sum_{j=4}^{J}\Gamma_{2j}X_{ij} + U_i^{**}.$$

$$W_i = \Gamma_{31} + \Gamma_{32}\bar{M}_i + \Gamma_{33}\bar{W}_i + \sum_{j=4}^{J}\Gamma_{3j}X_{ij} + V_i^{**} .$$

The reduced form parameters (Γs) are functions of the ρs, and the reduced form error terms U^{**} and V^{**} are functions of U^* and V^*, which are not related to any of the regressors in the reduced form equations.

We could estimate the reduced form equations and get $\hat{M}_i$ and $\hat{W}_i$. We could then substitute $\hat{M}_i$ and $\hat{W}_i$ into the structural equations as proxy regressors (instruments) for M_i and W_i. The least squares regression of M_i on $\hat{W}_i$, $\bar{M}_i$ and the Xs and a least squares regression of W_i on $\hat{M}_i$, $\bar{W}_i$ and the Xs would yield consistent estimates of the respective ρs, but the standard errors would be incorrect. LIMDEP automatically does all the required estimations with the two-stage, least squares command:

```
2SLS; LHS=   ; RHS=      ; INST=     $
```

TWO-STAGE, LEAST SQUARES IN LIMDEP

The Becker and Johnston (1999) data are in the file named "Bill.CSV." Before reading these data into LIMDEP, however, the "Project Settings" must be increased from 200000 cells (222 rows and 900 columns) to accommodate the 4,178 observations. This can be done with a project setting of 4000000 cells (4444 rows and 900 columns), following the procedures described in Module One, Part Two. After increasing the project setting, file Bill.CSV can be read into LIMDEP with the following read command (the file may be located anywhere on your hard drive but here it is located on the e drive):

```
READ; NREC=4178; NVAR=44; FILE=e:\bill.csv; Names=
student,school,size,other,birthday,sex,eslflag,adultst,
mc1,mc2,mc3,mc4,mc5,mc6,mc7,mc8,mc9,mc10,mc11,mc12,mc13,
mc14,mc15,mc16,mc17,mc18,mc19,mc20,totalmc,avgmc,
essay1,essay2,essay3,essay4,totessay,avgessay,
totscore,avgscore,ma081,ma082,ec011,ec012,ma083,en093$
```

Using these recode and create commands, yields the following relevant variable definitions:

```
recode; size; 0/9=1; 10/19=2; 20/29=3; 30/39=4; 40/49=5;
50/100=6$
create; smallest=size=1; smaller=size=2; small=size=3;
large=size=4; larger=size=5; largest=size=6$
```

TOTALMC: Student's score on 12th grade economics multiple-choice exam (M_i).

AVGMC: Mean multiple-choice score for students at school ($\bar{M}_i$).

TOTESSAY: Student's score on 12th grade economics essay exam (W_i).

AVGESSAY: Mean essay score for students at school ($\bar{W}_i$).

ADULTST = 1, if a returning adult student, and 0 otherwise.
SEX = GENDER = 1 if student is female and 0 is male.
ESLFLAG = 1 if English is not student's first language and 0 if it is.
EC011 = 1 if student enrolled in first semester 11 grade economics course, 0 if not.
EN093 = 1 if student was enrolled in ESL English course, 0 if not
MA081 = 1 if student enrolled in the first semester 11 grade math course, 0 if not.
MA082 = 1 if student was enrolled in the second semester 11 grade math course, 0 if not.
MA083 = 1 if student was enrolled in the first semester 12 grade math course, 0 if not.
SMALLER = 1 if student from a school with 10 to 19 test takers, 0 if not.
SMALL = 1 if student from a school with 20 to 29 test takers, 0 if not.
LARGE = 1 if student from a school with 30 to 39 test takers, 0 if not.
LARGER = 1 if student from a school with 40 to 49 test takers, 0 if not.

In all of the regressions, the effect of being at a school with more than 49 test takers is captured in the constant term, against which the other dummy variables are compared. The smallest schools need to be rejected to treat the mean scores as exogenous and unaffected by any individual student's test performance, which is accomplished with the following command:

```
Reject; smallest = 1$
```

The descriptive statistics on the relevant variables are then obtained with the following command, yielding the LIMDEP output table shown:

```
Dstat;RHS=TOTALMC,AVGMC,TOTESSAY,AVGESSAY,ADULTST,SEX,ESLFLAG,
EC01,EN093,MA081,MA082,MA083,SMALLER,SMALL,LARGE,LARGER$
```

```
Descriptive Statistics
All results based on nonmissing observations.
==============================================================================
Variable        Mean          Std.Dev.        Minimum         Maximum      Cases
==============================================================================
------------------------------------------------------------------------------
All observations in current sample
------------------------------------------------------------------------------
TOTALMC   12.4355795      3.96194160      .000000000      20.0000000       3710
AVGMC     12.4355800      1.97263767      6.41666700      17.0714300       3710
TOTESSAY  18.1380054      9.21191366      .000000000      40.0000000       3710
AVGESSAY  18.1380059      4.66807071      5.70000000      29.7857100       3710
ADULTST   .512129380E-02  .713893539E-01  .000000000      1.00000000       3710
SEX       .390566038      .487943012      .000000000      1.00000000       3710
ESLFLAG   .641509434E-01  .245054660      .000000000      1.00000000       3710
EC011     .677088949      .467652064      .000000000      1.00000000       3710
EN093     .622641509E-01  .241667268      .000000000      1.00000000       3710
MA081     .591374663      .491646035      .000000000      1.00000000       3710
MA082     .548787062      .497681208      .000000000      1.00000000       3710
MA083     .420215633      .493659946      .000000000      1.00000000       3710
SMALLER   .462264151      .498641179      .000000000      1.00000000       3710
SMALL     .207277628      .405410797      .000000000      1.00000000       3710
LARGE     .106469003      .308478530      .000000000      1.00000000       3710
LARGER    .978436658E-01  .297143201      .000000000      1.00000000       3710
```

For comparison with the two-stage least squares results, we start with the least squares regressions shown after this paragraph. The least squares estimations are typical of those found in multiple-choice and essay score correlation studies, with correlation coefficients of 0.77 and 0.78. The essay mark or score, *W*, is the most significant variable in the multiple-choice score regression (first of the two tables) and the multiple-choice mark, *M*, is the most significant variable in the essay regression (second of the two tables). Results like these have led researchers to conclude that the essay and multiple-choice marks are good predictors of each other. Notice also that both the mean multiple-choice and mean essay marks are significant in their respective equations, suggesting that something in the classroom environment or group experience influences individual test scores. Finally, being female has a significant negative effect on the multiple choice-test score, but a significant positive effect on the essay score, as expected from the least squares results reported by others. We will see how these results hold up in the two-stage least squares regressions.

```
Regress;LHS=TOTALMC;RHS=TOTESSAY,ONE,ADULTST,SEX,AVGMC,
ESLFLAG,EC011,EN093,MA081,MA082,MA083,SMALLER,SMALL,LARGE,LARGER$
```

```
+----------------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = none          |
| Dep. var. = TOTALMC  Mean=   12.43557951    , S.D.=    3.961941603         |
| Model size: Observations =    3710, Parameters =  15, Deg.Fr.=   3695      |
| Residuals:  Sum of squares= 23835.89955    , Std.Dev.=        2.53985      |
| Fit:        R-squared=  .590590, Adjusted R-squared =          .58904      |
| Model test: F[ 14,    3695] =  380.73,    Prob value =         .00000      |
| Diagnostic: Log-L =  -8714.8606, Restricted(b=0) Log-L =  -10371.4459      |
|             LogAmemiyaPrCrt.=    1.868, Akaike Info. Crt.=      4.706      |
| Autocorrel: Durbin-Watson Statistic =   1.99019,   Rho =       .00490      |
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
TOTESSAY	.2707916883	.54725877E-02	49.481	.0000	18.138005
Constant	2.654956801	.33936151	7.823	.0000	
ADULTST	.4674947703	.59221296	.789	.4299	.51212938E-02
SEX	-.5259548390	.91287080E-01	-5.762	.0000	.39056604
AVGMC	.3793818833	.25290373E-01	15.001	.0000	12.435580
ESLFLAG	.3933259495	.85245570	.461	.6445	.64150943E-01
EC011	.1722643321E-01	.92648817E-01	.186	.8525	.67708895
EN093	-.3117337847	.86493864	-.360	.7185	.62264151E-01
MA081	-.1208070545	.18084020	-.668	.5041	.59137466
MA082	.3827058262	.19467371	1.966	.0493	.54878706
MA083	.3703758129	.11847674	3.126	.0018	.42021563
SMALLER	.6721051012E-01	.14743497	.456	.6485	.46226415
SMALL	-.5687831831E-02	.15706323	-.036	.9711	.20727763
LARGE	.6635816769E-01	.17852633	.372	.7101	.10646900
LARGER	.5654860817E-01	.18217561	.310	.7563	.97843666E-01

(Note: E+nn or E-nn means multiply by 10 to + or -nn power.)

```
Regress;LHS=TOTESSAY;RHS=TOTALMC,ONE, ADULTST,SEX,AVGESSAY,
ESLFLAG,EC011,EN093,MA081,MA082,MA083,SMALLER,SMALL,LARGE,LARGER$
```

```
+----------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = TOTESSAY Mean=   18.13800539    , S.D.=    9.211913659    |
| Model size: Observations =     3710, Parameters =  15, Deg.Fr.=   3695 |
| Residuals:  Sum of squares= 123011.3151    , Std.Dev.=        5.76986 |
| Fit:        R-squared=  .609169, Adjusted R-squared =          .60769 |
| Model test: F[ 14,    3695] =  411.37,    Prob value =          .00000 |
| Diagnostic: Log-L = -11759.0705, Restricted(b=0) Log-L =  -13501.8081 |
|             LogAmemiyaPrCrt.=    3.509, Akaike Info. Crt.=      6.347 |
| Autocorrel: Durbin-Watson Statistic =   2.03115,    Rho =      -.01557 |
+----------------------------------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
TOTALMC	1.408895961	.28223608E-01	49.919	.0000	12.435580
Constant	-8.948704180	.55427657	-16.145	.0000	
ADULTST	-.8291495512	1.3454556	-.616	.5377	.51212938E-02
SEX	1.239956900	.20801072	5.961	.0000	.39056604
AVGESSAY	.4000235352	.23711680E-01	16.870	.0000	18.138006
ESLFLAG	.4511403830	1.9369352	.233	.8158	.64150943E-01
EC011	.2985371912	.21044864	1.419	.1560	.67708895
EN093	-2.020881931	1.9647001	-1.029	.3037	.62264151E-01
MA081	.8495120566	.41061265	2.069	.0386	.59137466
MA082	.1590915478	.44249860	.360	.7192	.54878706
MA083	1.809541566	.26793945	6.754	.0000	.42021563
SMALLER	.6170663022	.33054246	1.867	.0619	.46226415
SMALL	.2693408755	.35476913	.759	.4477	.20727763
LARGE	.2646447973	.40526280	.653	.5137	.10646900
LARGER	.6150288712E-01	.41436703	.148	.8820	.97843666E-01

(Note: E+nn or E-nn means multiply by 10 to + or -nn power.)

Theoretical considerations discussed in Module Two, Part One, suggest that these least squares estimates involve a simultaneous equation bias that is brought about by an apparent reverse causality between the two forms of testing. Consistent estimation of the parameters in this simultaneous equation system is possible with two-stage least squares, where our instrument ($\hat{M}_i$) for M_i is obtained by a least squares regression of M_i on SEX, ADULTST, AVGMC, AVGESSAY, ESLFLAG,SMALLER,SMALL, LARGE, LARGER, EC011, EN093, MA081, MA082, and MA083. Our instrument ($\hat{W}_i$) for W_i is obtained by a least squares regression of

W_i on SEX, ADULTST, AVGMC, AVGESSAY, ESLFLAG, SMALLER, SMALL, LARGE, LARGER, EC011, EN093, MA081, MA082, and MA083. LIMDEP will do these regressions and the subsequent regressions for *M* and *W* employing these instruments via the following commands, which yield the subsequent output:[i]

```
2SLS; LHS = TOTALMC; RHS = TOTESSAY,ONE, ADULTST,SEX,AVGMC,
ESLFLAG,EC011,EN093,MA081,MA082,MA083,SMALLER,SMALL,LARGE,
LARGER; INST = ONE,SEX, ADULTST ,AVGMC,AVGESSAY,ESLFLAG,
SMALLER,SMALL,LARGE,LARGER,EC011,EN093,MA081,MA082,MA083$
```

```
+-------------------------------------------------------------------+
| Two stage   least squares regression    Weighting variable = none  |
| Dep. var. = TOTALMC  Mean=   12.43557951    , S.D.=   3.961941603  |
| Model size: Observations =    3710, Parameters =  15, Deg.Fr.=   3695 |
| Residuals:  Sum of squares= 46157.78754    , Std.Dev.=        3.53440 |
| Fit:        R-squared=  .203966, Adjusted R-squared =          .20095 |
|             (Note:  Not using OLS.  R-squared is not bounded in [0,1] |
| Model test: F[ 14,   3695] =   67.63,    Prob value =          .00000 |
| Diagnostic: Log-L =  -9940.7797, Restricted(b=0) Log-L =  -10371.4459 |
|             LogAmemiyaPrCrt.=    2.529, Akaike Info. Crt.=      5.367 |
| Autocorrel: Durbin-Watson Statistic =   2.07829,   Rho =      -.03914 |
+-------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 TOTESSAY -.5247790489E-01  .36407219E-01   -1.441   .1495     18.138005
 Constant    -.3038295700       .57375703    -.530   .5964
 ADULTST      .2533493567       .82444633     .307   .7586  .51212938E-02
 SEX      -.8971949978E-01      .13581404    -.661   .5089     .39056604
 AVGMC        .9748840572   .74429145E-01   13.098   .0000     12.435580
 ESLFLAG      .6744471036       1.1866603     .568   .5698  .64150943E-01
 EC011        .2925430155       .13244518    2.209   .0272     .67708895
 EN093      -1.588715660       1.2118154   -1.311   .1899  .62264151E-01
 MA081        .2995655100       .25587578    1.171   .2417     .59137466
 MA082        .8159710785       .27507326    2.966   .0030     .54878706
 MA083       1.635255739       .21583992    7.576   .0000     .42021563
 SMALLER      .2715919941       .20639788    1.316   .1882     .46226415
 SMALL     .4372991271E-01      .21863306     .200   .8415     .20727763
 LARGE        .1981182700       .24885626     .796   .4260     .10646900
 LARGER   -.8677104536E-01      .25400196    -.342   .7326  .97843666E-01
```

```
2SLS; LHS = TOTESSAY; RHS = TOTALMC,ONE, ADULTST,SEX,AVGE
SSAY,ESLFLAG,EC011,EN093,MA081,MA082,MA083,SMALLER,SMALL,
LARGE,LARGER; INST = ONE,SEX, ADULTST,AVGMC,AVGESSAY,ESLFLAG,
SMALLER,SMALL,LARGE,LARGER,EC011,EN093,MA081,MA082,MA083$
```

```
+-------------------------------------------------------------------+
| Two stage   least squares regression    Weighting variable = none  |
| Dep. var. = TOTESSAY Mean=   18.13800539    , S.D.=   9.211913659  |
| Model size: Observations =    3710, Parameters =  15, Deg.Fr.=   3695 |
| Residuals:  Sum of squares= 201898.9900    , Std.Dev.=        7.39196 |
| Fit:        R-squared=  .355924, Adjusted R-squared =          .35348 |
|             (Note:  Not using OLS.  R-squared is not bounded in [0,1] |
| Model test: F[ 14,   3695] =  145.85,    Prob value =          .00000 |
| Diagnostic: Log-L = -12678.2066, Restricted(b=0) Log-L =  -13501.8081 |
|             LogAmemiyaPrCrt.=    4.005, Akaike Info. Crt.=      6.843 |
| Autocorrel: Durbin-Watson Statistic =   2.10160,   Rho =      -.05080 |
+-------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 TOTALMC   .2788777265E-01      .15799711     .177   .8599     12.435580
 Constant   -1.179740796       1.1193206   -1.054   .2919
 ADULTST     -.1690793751       1.7252757    -.098   .9219  .51212938E-02
```

SEX	.6854633662	.27355130	2.506	.0122	.39056604
AVGESSAY	.8417622152	.57819872E-01	14.558	.0000	18.138006
ESLFLAG	1.723698602	2.4855173	.693	.4880	.64150943E-01
EC011	.7128702679	.27353325	2.606	.0092	.67708895
EN093	-3.983249481	2.5265144	-1.577	.1149	.62264151E-01
MA081	1.069628788	.52662071	2.031	.0422	.59137466
MA082	1.217026971	.57901457	2.102	.0356	.54878706
MA083	3.892551120	.41430603	9.395	.0000	.42021563
SMALLER	.3348223746	.42463421	.788	.4304	.46226415
SMALL	-.1364832691	.45674848	-.299	.7651	.20727763
LARGE	.3418924354	.51926721	.658	.5103	.10646900
LARGER	-.8251220287E-01	.53110191	-.155	.8765	.97843666E-01

The 2SLS results differ from the least squares results in many ways. The essay mark or score, *W*, is no longer a significant variable in the multiple-choice regression and the multiple-choice mark, *M*, is likewise insignificant in the essay regression. Each score appears to be measuring something different when the regressor and error-term-induced bias is eliminated by our instrumental variable estimators.

Both the mean multiple-choice and mean essay scores continue to be significant in their respective equations. But now being female is insignificant in explaining the multiple-choice test score. Being female continues to have a significant positive effect on the essay score.

DURBIN, HAUSMAN AND WU TEST FOR ENDOGENEITY

The theoretical argument is strong for treating multiple-choice and essay scores as endogenous when employed as regressors in the explanation of the other. Nevertheless, this endogeneity can be tested with the Durbin, Hausman and Wu specification test, which is a two-step procedure in LIMDEP versions prior to 9. 0.4.[ii]

Either a Wald statistic, in a Chi-square (χ^2) test with K^* degrees of freedom, or an F statistic with K^* and $n - (K + K^*)$ degrees of freedom, is used to test the joint significance of the contribution of the predicted values ($\hat{\mathbf{X}}^*$) of a regression of the K^* endogenous regressors, in matrix **X*,** on the exogenous variables (and column of ones for the constant term) in matrix **Z:**

$$\mathbf{y} = \mathbf{X\beta} + \hat{\mathbf{X}}^*\boldsymbol{\gamma} + \boldsymbol{\varepsilon}^*,$$

where $\mathbf{X}^* = \mathbf{Z\lambda} + \mathbf{u}$, $\hat{\mathbf{X}}^* = \mathbf{Z}\hat{\boldsymbol{\lambda}}$, and $\hat{\boldsymbol{\lambda}}$ is a least squares estimator of $\boldsymbol{\lambda}$.

H_o: $\boldsymbol{\gamma} = 0$, the variables in **Z** are exogenous

H_A: $\boldsymbol{\gamma} \neq 0$, at least one of the variables in **Z** is endogenous

In our case, $K^* = 1$ when the essay score is to be tested as an endogenous regressor in the multiple-choice equation and when the multiple-choice regressor is to be tested as endogenous in the essay equation. $\hat{\mathbf{X}}^*$ is an $n \times 1$ vector of predicted essay scores from a regression of essay scores on all the exogenous variables (for subsequent use in the multiple-choice equation) or an $n \times 1$ vector of predicted multiple-choice scores from a regression of multiple-choice scores on all the exogenous variables (for subsequent use in the essay equation). Because $K^* = 1$, the relevant test statistic is either the *t*, with $n - (K + K^*)$ degrees of freedom for small *n* or the standard normal, for large *n*.

In LIMDEP, the predicted essay score is obtained by the following command, where the specification ";keep=Essayhat" tells LIMDEP to predict the essay scores and keep them as a variable called "Essayhat":

```
Regres; lhs= TOTESSAY; RHS= ONE,ADULTST,SEX, AVGESSAY,AVGMC,
ESLFLAG,EC011,EN093,MA081,MA082,MA083,SMALLER,SMALL,LARGE,LARGER
;keep=Essayhat$
```

The predicted essay scores are then added as a regressor in the original multiple-choice regression:

```
Regress;LHS=TOTALMC;RHS=TOTESSAY,ONE,ADULTST,SEX,AVGMC,
ESLFLAG,EC011,EN093,MA081,MA082,MA083,SMALLER,SMALL,LARGE,
LARGER, Essayhat$
```

The test statistic for the Essayhat coefficient is then used in the test of endogeneity. In the below LIMDEP output, we see that the calculated standard normal test statistic *z* value is −12.916, which far exceeds the absolute value of the 0.05 percent Type I error critical 1.96 standard normal value. Thus, the null hypothesis of an exogenous essay score as an explanatory variable for the multiple-choice score is rejected. As theorized, the essay score is endogenous in an explanation of the multiple-choice score.

```
--> Regres; lhs= TOTESSAY; RHS= ONE,ADULTST,SEX, AVGESSAY,AVGMC,
   ESLFLAG,EC011,EN093,MA081,MA082,MA083,SMALLER,SMALL,LARGE,LARGER;keep=Esayhat$

+-----------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = TOTESSAY Mean=   18.13800539    , S.D.=   9.211913659     |
| Model size: Observations =    3710, Parameters =  15, Deg.Fr.=   3695 |
| Residuals:  Sum of squares= 205968.5911    , Std.Dev.=        7.46609 |
| Fit:        R-squared=  .345598, Adjusted R-squared =          .34312 |
| Model test: F[ 14,    3695] =  139.38,    Prob value =          .00000 |
| Diagnostic: Log-L = -12715.2253, Restricted(b=0) Log-L =  -13501.8081 |
|             LogAmemiyaPrCrt.=    4.025, Akaike Info. Crt.=      6.863 |
| Autocorrel: Durbin-Watson Statistic =   2.10143,   Rho =      -.05072 |
+-----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant    -1.186477526       1.1613927   -1.022   .3070
 ADULTST     -.1617772661       1.7412483    -.093   .9260  .51212938E-02
 SEX          .6819632415       .27234747    2.504   .0123     .39056604
 AVGESSAY     .8405321032    .64642750E-01  13.003   .0000     18.138006
 AVGMC      .2714761464E-01     .15534612     .175   .8613     12.435580
 ESLFLAG      1.739961011       2.5067133     .694   .4876  .64150943E-01
 EC011        .7199749635       .27219191    2.645   .0082     .67708895
 EN093       -4.021669541       2.5417647   -1.582   .1136  .62264151E-01
 MA081        1.076407689       .53146100    2.025   .0428     .59137466
 MA082        1.237970826       .57190601    2.165   .0304     .54878706
 MA083        3.932399725       .34253928   11.480   .0000     .42021563
 SMALLER      .3418961082       .43385196     .788   .4307     .46226415
 SMALL       -.1350660711       .46222353    -.292   .7701     .20727763
```

```
 LARGE        .3469098130       .52477155     .661   .5086     .10646900
 LARGER   -.8480793833E-01      .53618108    -.158   .8743  .97843666E-01
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```

--> Regress;LHS=TOTALMC;RHS=TOTESSAY,ONE,ADULTST,SEX,AVGMC,
ESLFLAG,EC011,EN093,MA081,MA082,MA083,SMALLER,SMALL,LARGE,LARGER,
Essayhat$

```
+----------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = none      |
| Dep. var. = TOTALMC  Mean=   12.43557951    , S.D.=    3.961941603     |
| Model size: Observations =    3710, Parameters =  16, Deg.Fr.=   3694  |
| Residuals:  Sum of squares= 22805.95017    , Std.Dev.=        2.48471  |
| Fit:        R-squared=  .608280, Adjusted R-squared =          .60669  |
| Model test: F[ 15,   3694] =  382.41,    Prob value =          .00000  |
| Diagnostic: Log-L =  -8632.9227, Restricted(b=0) Log-L =  -10371.4459  |
|             LogAmemiyaPrCrt.=    1.825, Akaike Info. Crt.=      4.662  |
| Autocorrel: Durbin-Watson Statistic =   2.07293,   Rho =      -.03647  |
+----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 TOTESSAY     .2855834321   .54748868E-02   52.162   .0000     18.138005
 Constant    -.3038295700       .40335588    -.753   .4513
 ADULTST      .2533493567       .57959250     .437   .6620  .51212938E-02
 SEX      -.8971949978E-01  .95478380E-01    -.940   .3474     .39056604
 AVGMC        .9748840572   .52324297E-01   18.632   .0000     12.435580
 ESLFLAG      .6744471036       .83423185     .808   .4188  .64150943E-01
 EC011        .2925430155   .93110045E-01    3.142   .0017     .67708895
 EN093       -1.588715660       .85191616   -1.865   .0622  .62264151E-01
 MA081        .2995655100       .17988277    1.665   .0958     .59137466
 MA082        .8159710785       .19337874    4.220   .0000     .54878706
 MA083        1.635255739       .15173722   10.777   .0000     .42021563
 SMALLER      .2715919941       .14509939    1.872   .0612     .46226415
 SMALL      .4372991270E-01     .15370083     .285   .7760     .20727763
 LARGE        .1981182700       .17494798    1.132   .2574     .10646900
 LARGER   -.8677104536E-01      .17856546    -.486   .6270  .97843666E-01
 ESSAYHAT    -.3380613370   .26173585E-01  -12.916   .0000     18.138005
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```

The similar estimation routine to test for the endogeneity of the multiple-choice test score in the essay equation yields a calculated *z* test statistic of -11.713, which far exceeds the absolute value of its 1.96 critical value. Thus, the null hypothesis of an exogenous multiple-choice score as an explanatory variable for the essay score is rejected. As theorized, the multiple-choice score is endogenous in an explanation of the essay score.

--> Regress;LHS=TOTALMC; RHS=ONE, ADULTST,SEX, AVGMC,AVGESSAY,
ESLFLAG,EC011,EN093,MA081,MA082,MA083,SMALLER,SMALL,LARGE,LARGER;
keep=MChat$

```
+----------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = none      |
| Dep. var. = TOTALMC  Mean=   12.43557951    , S.D.=    3.961941603     |
| Model size: Observations =    3710, Parameters =  15, Deg.Fr.=   3695  |
| Residuals:  Sum of squares= 39604.31525    , Std.Dev.=        3.27389  |
| Fit:        R-squared=  .319748, Adjusted R-squared =          .31717  |
| Model test: F[ 14,   3695] =  124.06,    Prob value =          .00000  |
| Diagnostic: Log-L =  -9656.7280, Restricted(b=0) Log-L =  -10371.4459  |
|             LogAmemiyaPrCrt.=    2.376, Akaike Info. Crt.=      5.214  |
| Autocorrel: Durbin-Watson Statistic =   2.07600,   Rho =      -.03800  |
+----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
```

```
 Constant    -.2415657153        .50927203    -.474   .6353
 ADULTST      .2618390887        .76353941     .343   .7317  .51212938E-02
 SEX         -.1255075019        .11942469   -1.051   .2933     .39056604
 AVGMC        .9734594072     .68119457E-01  14.290   .0000     12.435580
 AVGESSAY -.4410936377E-01  .28345921E-01   -1.556   .1197     18.138006
 ESLFLAG      .5831375952       1.0991967     .531   .5958  .64150943E-01
 EC011        .2547602379       .11935647    2.134   .0328     .67708895
 EN093       -1.377666868       1.1145668   -1.236   .2164  .62264151E-01
 MA081        .2430778897       .23304627    1.043   .2969     .59137466
 MA082        .7510049632       .25078145    2.995   .0027     .54878706
 MA083        1.428891640       .15020388    9.513   .0000     .42021563
 SMALLER      .2536500026       .19024459    1.333   .1824     .46226415
 SMALL     .5081789714E-01      .20268556     .251   .8020     .20727763
 LARGE        .1799131698       .23011294     .782   .4343     .10646900
 LARGER   -.8232050244E-01      .23511603    -.350   .7262  .97843666E-01
```

```
--> Regress;LHS=TOTESSAY;RHS=TOTALMC,ONE, ADULTST,SEX,AVGESSAY,
    ESLFLAG,EC011,EN093,MA081,MA082,MA083,SMALLER,SMALL,LARGE,LARGER,
    MChat$
```

```
+-----------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = TOTESSAY Mean=   18.13800539    , S.D.=   9.211913659     |
| Model size: Observations =    3710, Parameters =  16, Deg.Fr.=   3694 |
| Residuals:  Sum of squares= 118606.0003    , Std.Dev.=        5.66637 |
| Fit:        R-squared=  .623166, Adjusted R-squared =          .62164 |
| Model test: F[ 15,    3694] =  407.25,    Prob value =         .00000 |
| Diagnostic: Log-L = -11691.4200, Restricted(b=0) Log-L =  -13501.8081 |
|             LogAmemiyaPrCrt.=    3.473, Akaike Info. Crt.=      6.311 |
| Autocorrel: Durbin-Watson Statistic =   2.09836,   Rho =       -.04918 |
+-----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 TOTALMC      1.485222426   .28473026E-01   52.162   .0000     12.435580
 Constant    -1.179740796       .85802415   -1.375   .1691
 ADULTST     -.1690793751       1.3225239    -.128   .8983  .51212938E-02
 SEX          .6854633662       .20969294    3.269   .0011     .39056604
 AVGESSAY     .8417622152   .44322287E-01   18.992   .0000     18.138006
 ESLFLAG      1.723698602       1.9052933     .905   .3656  .64150943E-01
 EC011        .7128702679       .20967911    3.400   .0007     .67708895
 EN093       -3.983249481       1.9367199   -2.057   .0397  .62264151E-01
 MA081        1.069628788       .40368533    2.650   .0081     .59137466
 MA082        1.217026971       .44384827    2.742   .0061     .54878706
 MA083        3.892551120       .31758961   12.257   .0000     .42021563
 SMALLER      .3348223746       .32550676    1.029   .3037     .46226415
 SMALL       -.1364832691       .35012421    -.390   .6967     .20727763
 LARGE        .3418924354       .39804844     .859   .3904     .10646900
 LARGER   -.8251220288E-01      .40712043    -.203   .8394  .97843666E-01
 MCHAT       -1.457334653       .12441585  -11.713   .0000     12.435580
 (Note: E+nn or E-nn means multiply by 10 to + or -nn power.)
```

CONCLUDING COMMENTS

This cookbook-type introduction to the use of instrumental variables and two-stage least squares regression and testing for endogeneity has just scratched the surface of this controversial problem in statistical estimation and inference. It was intended to enable researchers to begin using instrumental variables in their work and to enable readers of that work to have an idea of what is being done. To learn more about these methods there is no substitute for a graduate level textbook treatment such as that found in William Greene's *Econometric Analysis*.

REFERENCES

Becker, William E. and Carol Johnston (1999).“The Relationship Between Multiple Choice and Essay Response Questions in Assessing Economics Understanding,” *Economic Record* (Economic Society of Australia), Vol. 75 (December): 348-357.

Greene, William (2003). *Econometric Analysis*. 5th Edition, New Jersey: Prentice Hall.

ENDNOTES

---

[i] In the default mode, relatively large samples are required for 2SLS in LIMDEP because a routine aimed at providing consistent estimators is employed; thus, for example, no degrees of freedom adjustment is made for variances; *i.e.,*

$$\hat{\sigma}^2 = (1/n)\sum(i^{th}\ prediction\ error)^2$$

As William Greene states, “this is consistent with most published sources, but (curiously enough) inconsistent with most other commercially available computer programs.” The degrees of freedom correction for small samples is obtainable by adding the following specification to the 2SLS command: `;DFC`

[ii] In Limdep version 9.0.4, the following command will automatically test *x3* for endogeneity:

```
Regress; lhs=y; rhs=one,x2,x3; inst=one,x2,x4; Wu test$
```

Because *x3* is not an instrument, LIMDEP knows the test for endogeneity is on this variable.

MODULE TWO, PART THREE: ENDOGENEITY, INSTRUMENTAL VARIABLES AND TWO-STAGE LEAST SQUARES IN ECONOMIC EDUCATION RESEARCH USING STATA

Part Three of Module Two demonstrates how to address problems of endogeneity using STATA's two-stage least squares instrumental variable estimator, as well as how to perform and interpret the Durbin, Hausman and Wu specification test for endogeneity. Users of this model need to have completed Module One, Parts One and Three, and Module Two, Part One. That is, from Module One, users are assumed to know how to get data into STATA, recode and create variables within STATA, and run and interpret regression results. From Module Two, Part One, they are expected to have an understanding of the problem of and source of endogeneity and the basic idea behind an instrumental variable approach and the two-stage least squares method. The Becker and Johnston (1999) data set is used throughout this module.

THE CASE

As described in Module Two, Part One, Becker and Johnston (1999) called attention to classroom effects that might influence multiple-choice and essay type test taking skills in economics in different ways. For example, if the student is in a classroom that emphasizes skills associated with multiple-choice testing (*e.g.*, risk-taking behavior, question analyzing skills, memorization, and keen sense of judging between close alternatives), then the student can be expected to do better on multiple-choice questions. By the same token, if placed in a classroom that emphasizes the skills of essay test question answering (*e.g.*, organization, good sentence and paragraph construction, obfuscation when uncertain, logical argument, and good penmanship), then the student can be expected to do better on the essay component. Thus, Becker and Johnston attempted to control for the type of class of which the student is a member. Their measure of "teaching to the multiple-choice questions" is the mean score or mark on the multiple-choice questions for the school in which the i^{th} student took the 12^{th} grade economics course. Similarly, the mean school mark or score on the essay questions is their measure of the i^{th} student's exposure to essay question writing skills.

In equation form, the two equations that summarize the influence of the various covariates on multiple-choice and essay test questions are written as the following structural equations:

$$M_i = \rho_{21} + \rho_{22} W_i + \rho_{23}\,\bar{M}_i + \sum_{j=4}^{J} \rho_{2j} X_{ij} + U_i^*.$$

$$W_i = \rho_{31} + \rho_{32} M_i + \rho_{33}\,\bar{W}_i + \sum_{j=4}^{J} \rho_{3j} X_{ij} + V_i^* \quad .$$

M_i and W_i are the i^{th} student's respective scores on the multiple-choice test and essay test. $\bar{M}_i$ and $\bar{W}_i$ are the mean multiple-choice and essay test scores at the school where the i^{th} student took the 12th grade economics course. The X_{ij} variables are the other exogenous variables (such as gender, age, English a second language, etc.) used to explain the i^{th} student's multiple-choice and essay marks, where the ρs are parameters to be estimated. The inclusion of the mean multiple-choice and mean essay test scores in their respective structural equations, and their exclusion from the other equation, enables both of the structural equations to be identified within the system.

As shown in Module Two, Part One, the least squares estimators of the ρs are biased because the error term U_i^* is related to W_i, in the first equation, and V_i^* is related to M_i in second equation. Instruments for regressors W_i and M_i are needed. Because the reduced form equations express W_i and M_i solely in terms of exogenous variables, they can be used to generate the respective instruments:

$$M_i = \Gamma_{21} + \Gamma_{22}\bar{W}_i + \Gamma_{23}\bar{M}_i + \sum_{j=4}^{J}\Gamma_{2j}X_{ij} + U_i^{**}.$$

$$W_i = \Gamma_{31} + \Gamma_{32}\bar{M}_i + \Gamma_{33}\bar{W}_i + \sum_{j=4}^{J}\Gamma_{3j}X_{ij} + V_i^{**} \quad .$$

The reduced form parameters (Γs) are functions of the ρs, and the reduced form error terms U^{**} and V^{**} are functions of U^* and V^*, which are not related to any of the regressors in the reduced form equations.

We could estimate the reduced form equations and get $\hat{M}_i$ and $\hat{W}_i$. We could then substitute $\hat{M}_i$ and $\hat{W}_i$ into the structural equations as proxy regressors (instruments) for M_i and W_i. The least squares regression of M_i on $\hat{W}_i$, $\bar{M}_i$ and the Xs and a least squares regression of W_i on $\hat{M}_i$, $\bar{W}_i$ and the Xs would yield consistent estimates of the respective ρs, but the standard errors would be incorrect. STATA automatically performs the required estimations with the instrumental variables command:[i]

> ivreg *dependent_variable independent_variables*
> *(engoenous_var_name=instruments)*

Here, *independent_variables* should be all of your included, exogenous variables, and in the parentheses, we must specify the endogenous variable as a function of its instruments.

TWO-STAGE, LEAST SQUARES IN STATA

The Becker and Johnston (1999) data are in the file named "Bill.CSV." Since this is a large dataset, users may need to increase the size of STATA following the procedures described in Module One, Part Three. For the version used in this Module (Intercooled STATA), the default memory is sufficient. Now, the file Bill.CSV can be read into STATA with the following insheet command. Note that, in this case, the directory has been changed beforehand so that we need only specify the file BILL.csv. For instance, say the file is located in the folder, "C:\Documents and Settings\My Documents\BILL.csv". Then users can change the directory with the command, *cd "C:\Documents and Settings\My Documents"*, in which case the file may be accessed simply by specifying the actual file name, BILL.csv, as in the following:

```
insheet student school size other birthday sex eslflag ///
adultst mc1 mc2 mc3 mc4 mc5 mc6 mc7 mc8 mc9 mc10 mc11  ///
mc12 mc13 mc14 mc15 mc16 mc17 mc18 mc19 mc20 totalmc   ///
avgmc essay1 essay2 essay3 essay4 totessay avgessay    ///
totscore avgscore ma081 ma082 ec011 ec012 ma083 en093  ///
using "BILL.csv", comma
```

Using these recode and generate commands yields the following relevant variable definitions:

```
recode size (0/9=1) (10/19=2) (20/29=3) (30/39=4) ///
(40/49=5) (50/100=6)
gen smallest=(size==1)
gen smaller=(size==2)
gen small=(size==3)
gen large=(size==4)
gen larger=(size==5)
gen largest=(size==6)
```

TOTALMC: Student's score on 12th grade economics multiple-choice exam (M_i).

AVGMC: Mean multiple-choice score for students at school ($\bar{M}_i$).

TOTESSAY: Student's score on 12th grade economics essay exam (W_i).

AVGESSAY: Mean essay score for students at school ($\bar{W}_i$).

ADULTST = 1, if a returning adult student, and 0 otherwise.
SEX = GENDER = 1 if student is female and 0 is male.
ESLFLAG = 1 if English is not student's first language and 0 if it is.
EC011 = 1 if student enrolled in first semester 11 grade economics course, 0 if not.
EN093 = 1 if student was enrolled in ESL English course, 0 if not
MA081 = 1 if student enrolled in the first semester 11 grade math course, 0 if not.
MA082 = 1 if student was enrolled in the second semester 11 grade math course, 0 if not.
MA083 = 1 if student was enrolled in the first semester 12 grade math course, 0 if not.
SMALLER = 1 if student from a school with 10 to 19 test takers, 0 if not.
SMALL = 1 if student from a school with 20 to 29 test takers, 0 if not.
LARGE = 1 if student from a school with 30 to 39 test takers, 0 if not.

LARGER = 1 if student from a school with 40 to 49 test takers, 0 if not.

In all of the regressions, the effect of being at a school with more than 49 test takers is captured in the constant term, against which the other dummy variables are compared. The smallest schools should not be included so that we can treat the mean scores as exogenous and unaffected by any individual student's test performance, which is accomplished by adding the following command to the end of the summary statistics and regression commands:

```
if smallest!=1
```

This command is added to the end of our regression and summary statistics commands as an option, and it says to only perform the desired command if smallest is not equal to 1. We could also completely remove these observations with the command:

```
drop if smallest==1
```

The problem with this approach, however, is that we cannot retrieve observations once they've been dropped (at least not easily), so it's generally sound practice to follow the first approach.

The descriptive statistics on the relevant variables are then obtained with the following command, yielding the STATA output shown:[ii]

```
sum totalmc avgmc totessay avgessay adultst sex eslflag ///
ec011 en093 ma081 ma082 ma083 smaller small large ///
larger if smallest!=1
```

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     totalmc |      3710    12.43558    3.961942          0         20
       avgmc |      3710    12.43558    1.972638   6.416667   17.07143
    totessay |      3710    18.13801    9.211914          0         40
    avgessay |      3710    18.13801    4.668071        5.7   29.78571
     adultst |      3710    .0051213    .0713894          0          1
-------------+--------------------------------------------------------
         sex |      3710     .390566     .487943          0          1
     eslflag |      3710    .0641509    .2450547          0          1
       ec011 |      3710    .6770889    .4676521          0          1
       en093 |      3710    .0622642    .2416673          0          1
       ma081 |      3710    .5913747     .491646          0          1
-------------+--------------------------------------------------------
       ma082 |      3710    .5487871    .4976812          0          1
       ma083 |      3710    .4202156    .4936599          0          1
     smaller |      3711    .4621396    .4986317          0          1
       small |      3711    .2072218    .4053704          0          1
       large |      3711    .1064403    .3084419          0          1
-------------+--------------------------------------------------------
      larger |      3711    .0978173    .2971075          0          1
```

For comparison with the two-stage least squares results, we start with the least squares regressions shown after this paragraph. The least squares estimations are typical of those found in multiple-choice and essay score correlation studies, with correlation coefficients of 0.77 and 0.78. The essay mark or score, *W*, is the most significant variable in the multiple-choice score regression (first of the two tables) and the multiple-choice mark, *M*, is the most significant variable in the essay regression (second of the two tables). Results like these have led researchers to conclude that the essay and multiple-choice marks are good predictors of each other. Notice also that both the mean multiple-choice and mean essay marks are significant in their respective equations, suggesting that something in the classroom environment or group experience influences individual test scores. Finally, being female has a significant negative effect on the multiple choice-test score, but a significant positive effect on the essay score, as expected from the least squares results reported by others. We will see how these results hold up in the two-stage least squares regressions.

```
regress totalmc totessay adultst sex avgmc eslflag ec011 ///
en093 ma081 ma082 ma083 smaller small large larger if ///
smallest!=1
```

```
      Source |       SS       df       MS              Number of obs =    3710
-------------+------------------------------           F( 14,  3695) =  380.73
       Model |  34384.2039    14  2456.01457           Prob > F      =  0.0000
    Residual |  23835.8996  3695  6.45085239           R-squared     =  0.5906
-------------+------------------------------           Adj R-squared =  0.5890
       Total |  58220.1035  3709  15.6969813           Root MSE      =  2.5399

------------------------------------------------------------------------------
     totalmc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    totessay |   .2707917   .0054726    49.48   0.000     .2600621    .2815213
     adultst |   .4674948    .592213     0.79   0.430    -.6936016    1.628591
         sex |  -.5259549   .0912871    -5.76   0.000    -.7049329   -.3469768
       avgmc |   .3793819   .0252904    15.00   0.000     .3297974    .4289663
     eslflag |   .3933259   .8524557     0.46   0.645    -1.278004    2.064656
       ec011 |   .0172264   .0926488     0.19   0.853    -.1644214    .1988743
       en093 |  -.3117338   .8649386    -0.36   0.719    -2.007538     1.38407
       ma081 |   -.120807   .1808402    -0.67   0.504    -.4753635    .2337494
       ma082 |   .3827058   .1946737     1.97   0.049     .0010273    .7643843
       ma083 |   .3703758   .1184767     3.13   0.002     .1380896     .602662
     smaller |   .0672105    .147435     0.46   0.649    -.2218514    .3562725
       small |  -.0056878   .1570632    -0.04   0.971    -.3136269    .3022513
       large |   .0663582   .1785263     0.37   0.710    -.2836616    .4163781
      larger |   .0565486   .1821756     0.31   0.756     -.300626    .4137232
       _cons |   2.654957   .3393615     7.82   0.000     1.989603    3.320311
------------------------------------------------------------------------------
```

```
regress totessay totalmc adultst sex avgessay eslflag ec011 ///
en093 ma081 ma082 ma083 smaller small large larger if ///
smallest!=1
```

```
      Source |       SS       df       MS              Number of obs =    3710
-------------+------------------------------           F( 14,  3695) =  411.37
       Model |  191732.026    14  13695.1447           Prob > F      =  0.0000
    Residual |  123011.315  3695  33.2912896           R-squared     =  0.6092
-------------+------------------------------           Adj R-squared =  0.6077
       Total |  314743.341  3709  84.8593533           Root MSE      =  5.7699
```

totessay	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
totalmc	1.408896	.0282236	49.92	0.000	1.353561	1.464231
adultst	-.8291495	1.345456	-0.62	0.538	-3.467058	1.808759
sex	1.239957	.2080107	5.96	0.000	.8321298	1.647784
avgessay	.4000235	.0237117	16.87	0.000	.3535343	.4465128
eslflag	.4511402	1.936935	0.23	0.816	-3.346427	4.248707
ec011	.2985372	.2104486	1.42	0.156	-.1140697	.7111441
en093	-2.020882	1.9647	-1.03	0.304	-5.872885	1.831122
ma081	.8495121	.4106127	2.07	0.039	.0444623	1.654562
ma082	.1590916	.4424986	0.36	0.719	-.7084739	1.026657
ma083	1.809542	.2679394	6.75	0.000	1.284218	2.334865
smaller	.6170663	.3305425	1.87	0.062	-.0309973	1.26513
small	.2693409	.3547691	0.76	0.448	-.4262217	.9649034
large	.2646448	.4052628	0.65	0.514	-.5299159	1.059206
larger	.0615029	.414367	0.15	0.882	-.7509076	.8739135
_cons	-8.948704	.5542766	-16.14	0.000	-10.03542	-7.861986

Theoretical considerations discussed in Module Two, Part One, suggest that these least squares estimates involve a simultaneous equation bias that is brought about by an apparent reverse causality between the two forms of testing. Consistent estimation of the parameters in this simultaneous equation system is possible with two-stage least squares, where our instrument $\hat{M}_i$ for M_i is obtained by a least squares regression of M_i on SEX, ADULTST, AVGMC, AVGESSAY, ESLFLAG,SMALLER,SMALL, LARGE, LARGER, EC011, EN093, MA081, MA082, and MA083. Our instrument for $\hat{W}_i$ for W_i is obtained by a least squares regression of W_i on SEX, ADULTST, AVGMC, AVGESSAY, ESLFLAG, SMALLER, SMALL, LARGE, LARGER, EC011, EN093, MA081, MA082, and MA083. STATA will do these regressions and the subsequent regressions for *M* and *W* employing these instruments via the following commands, which yield the subsequent output. Note that we should only specify as instruments variables that we are not including as independent variables in the full regression. As seen in the output tables, STATA correctly includes all of the exogenous variables as instruments in the two-stage least squares estimation:

```
ivreg totalmc adultst sex avgmc eslflag ec011 en093 ///
ma081 ma082 ma083 smaller small large larger ///
(totessay=avgessay) if smallest!=1
```

Instrumental variables (2SLS) regression

Source	SS	df	MS
Model	11874.9352	14	848.209657
Residual	46345.1683	3695	12.5427707
Total	58220.1035	3709	15.6969813

Number of obs =	3710
F(14, 3695) =	106.01
Prob > F =	0.0000
R-squared =	0.2040
Adj R-squared =	0.2010
Root MSE =	3.5416

totalmc	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
totessay	-.0524779	.036481	-1.44	0.150	-.1240029	.019047
adultst	.2533495	.8261181	0.31	0.759	-1.366343	1.873042
sex	-.0897195	.1360894	-0.66	0.510	-.3565373	.1770983
avgmc	.9748841	.0745801	13.07	0.000	.8286619	1.121106
eslflag	.6744471	1.189066	0.57	0.571	-1.656844	3.005738
ec011	.2925431	.1327137	2.20	0.028	.0323437	.5527425
en093	-1.588716	1.214273	-1.31	0.191	-3.969426	.7919948
ma081	.2995655	.2563946	1.17	0.243	-.2031234	.8022545

```
     ma082 |   .8159711    .275631     2.96   0.003     .2755672    1.356375
     ma083 |   1.635256   .2162776     7.56   0.000     1.211221    2.059291
   smaller |   .2715921   .2068164     1.31   0.189    -.1338934    .6770776
     small |     .04373   .2190764     0.20   0.842    -.3857925    .4732525
     large |   .1981185   .2493609     0.79   0.427      -.29078     .687017
    larger |   -.086771    .254517    -0.34   0.733    -.5857787    .4122366
     _cons |  -.3038295   .5749204    -0.53   0.597    -1.431022    .8233631
------------------------------------------------------------------------------
Instrumented:  totessay
Instruments:   adultst sex avgmc eslflag ec011 en093 ma081 ma082 ma083
               smaller small large larger avgessay
------------------------------------------------------------------------------
```

```
ivreg totessay adultst sex avgessay eslflag ec011 ///
en093 ma081 ma082 ma083 smaller small large larger ///
(totalmc=avgmc) if smallest!=1
```

```
Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =    3710
-------------+------------------------------           F( 14,  3695) =  141.62
       Model |  112024.731    14  8001.76652           Prob > F      =  0.0000
    Residual |   202718.61  3695  54.8629526           R-squared     =  0.3559
-------------+------------------------------           Adj R-squared =  0.3535
       Total |  314743.341  3709  84.8593533           Root MSE      =   7.407

------------------------------------------------------------------------------
    totessay |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     totalmc |   .0278877   .1583175     0.18   0.860    -.2825105     .338286
     adultst |  -.1690792   1.728774    -0.10   0.922    -3.558524    3.220366
         sex |   .6854633    .274106     2.50   0.012     .1480494    1.222877
    avgessay |   .8417622   .0579371    14.53   0.000     .7281703    .9553541
     eslflag |   1.723698   2.490557     0.69   0.489    -3.159304      6.6067
       ec011 |   .7128703   .2740879     2.60   0.009     .1754919    1.250249
       en093 |  -3.983249   2.531637    -1.57   0.116    -8.946793     .980295
       ma081 |   1.069629   .5276886     2.03   0.043     .0350393    2.104218
       ma082 |   1.217027   .5801887     2.10   0.036     .0795055    2.354548
       ma083 |   3.892551   .4151461     9.38   0.000     3.078613    4.706489
     smaller |   .3348224   .4254953     0.79   0.431    -.4994062    1.169051
       small |  -.1364833   .4576746    -0.30   0.766    -1.033803    .7608364
       large |   .3418925   .5203201     0.66   0.511    -.6782504    1.362035
      larger |  -.0825121   .5321788    -0.16   0.877    -1.125905     .960881
       _cons |  -1.179741    1.12159    -1.05   0.293    -3.378737    1.019256
------------------------------------------------------------------------------
Instrumented:  totalmc
Instruments:   adultst sex avgessay eslflag ec011 en093 ma081 ma082 ma083
               smaller small large larger avgmc
------------------------------------------------------------------------------
```

The 2SLS results differ from the least squares results in many ways. The essay mark or score, *W*, is no longer a significant variable in the multiple-choice regression and the multiple-choice mark, *M*, is likewise insignificant in the essay regression. Each score appears to be measuring something different when the regressor and error-term-induced bias is eliminated by our instrumental variable estimators.

Both the mean multiple-choice and mean essay scores continue to be significant in their respective equations. But now being female is insignificant in explaining the multiple-choice test score. Being female continues to have a significant positive effect on the essay score.

DURBIN, HAUSMAN AND WU TEST FOR ENDOGENEITY

The theoretical argument is strong for treating multiple-choice and essay scores as endogenous when employed as regressors in the explanation of the other. Nevertheless, this endogeneity can be tested with the Durbin, Hausman and Wu specification test. There are at least two ways to perform this test in STATA: One is with the auxiliary regression as done with LIMDEP in Module Two, Part Two, and the other is with the general Hausman command. For consistency across Parts Two and Three, we use the auxiliary regression method here; however, for those interested in the more general Hausman command, type *help hausman* in the command window for a brief description.

Either a Wald statistic, in a Chi-square (χ^2) test with K* degrees of freedom, or an F statistic with K* and $n - (K + K^*)$ degrees of freedom, is used to test the joint significance of the contribution of the predicted values ($\hat{\mathbf{X}}^*$) of a regression of the K* endogenous regressors, in matrix **X*,** on the exogenous variables (and column of ones for the constant term) in matrix **Z:**

$$\mathbf{y} = \mathbf{X\beta} + \hat{\mathbf{X}}^*\boldsymbol{\gamma} + \boldsymbol{\varepsilon}^*,$$

$$\text{where } \mathbf{X}^* = \mathbf{Z\lambda} + \mathbf{u},\ \hat{\mathbf{X}}^* = \mathbf{Z}\hat{\boldsymbol{\lambda}}, \text{ and } \hat{\boldsymbol{\lambda}} \text{ is a least squares estimator of } \boldsymbol{\lambda}.$$

$H_o:\ \boldsymbol{\gamma} = 0$, the variables in **Z** are exogenous

$H_A:\ \boldsymbol{\gamma} \neq 0$, at least one of the variables in **Z** is endogenous

In our case, $K^* = 1$ when the essay score is to be tested as an endogenous regressor in the multiple-choice equation and when the multiple-choice regressor is to be tested as endogenous in the essay equation. $\hat{\mathbf{X}}^*$ is an $n \times 1$ vector of predicted essay scores from a regression of essay scores on all the exogenous variables (for subsequent use in the multiple-choice equation) or an $n \times 1$ vector of predicted multiple-choice scores from a regression of multiple-choice scores on all the exogenous variables (for subsequent use in the essay equation). Because $K^* = 1$, the relevant test statistic is either the t, with $n - (K + K^*)$ degrees of freedom for small n or the standard normal, for large n.

In STATA, the predicted essay score is obtained by the following command, where the specification "predict totesshat, xb" tells STATA to predict the essay scores and keep them as a variable called "totesshat":

```
regress totessay adultst sex avgmc avgessay eslflag ///
ec011 en093 ma081 ma082 ma083 smaller small large  ///
larger if smallest!=1

predict totesshat, xb
```

The predicted essay scores are then added as a regressor in the original multiple-choice regression:

```
regress totalmc totessay adultst sex avgmc eslflag ///
```

```
ec011 en093 ma081 ma082 ma083 smaller small large ///
larger totesshat if smallest!=1
```

The test statistic for the totesshat coefficient is then used in the test of endogeneity. In the below STATA output, we see that the calculated standard normal test statistic z value is -12.92, which far exceeds the absolute value of the 0.05 percent Type I error critical 1.96 standard normal value. Thus, the null hypothesis of an exogenous essay score as an explanatory variable for the multiple-choice score is rejected. As theorized, the essay score is endogenous in an explanation of the multiple-choice score.

```
. regress totessay adultst sex avgmc avgessay eslflag ec011 en093 ma081 ma082 ///
ma083 smaller small large larger if smallest!=1

      Source |       SS       df       MS              Number of obs =    3710
-------------+------------------------------           F( 14,  3695) =  139.38
       Model |  108774.75    14  7769.62501           Prob > F      =  0.0000
    Residual | 205968.591  3695  55.7425145           R-squared     =  0.3456
-------------+------------------------------           Adj R-squared =  0.3431
       Total | 314743.341  3709  84.8593533           Root MSE      =  7.4661

------------------------------------------------------------------------------
    totessay |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     adultst |  -.1617771   1.741248    -0.09   0.926    -3.575679    3.252125
         sex |   .6819632   .2723475     2.50   0.012     .1479971    1.215929
       avgmc |   .0271476   .1553461     0.17   0.861     -.277425    .3317202
    avgessay |   .8405321   .0646428    13.00   0.000     .7137931    .9672711
     eslflag |   1.739961   2.506713     0.69   0.488    -3.174717    6.654638
       ec011 |    .719975   .2721919     2.65   0.008     .1863138    1.253636
       en093 |  -4.021669   2.541765    -1.58   0.114    -9.005069    .9617305
       ma081 |   1.076408    .531461     2.03   0.043     .0344219    2.118393
       ma082 |   1.237971    .571906     2.16   0.030     .1166884    2.359253
       ma083 |     3.9324   .3425393    11.48   0.000     3.260815    4.603984
     smaller |   .3418961    .433852     0.79   0.431    -.5087167    1.192509
       small |  -.1350661   .4622235    -0.29   0.770    -1.041304    .7711722
       large |   .3469099   .5247716     0.66   0.509    -.6819605     1.37578
      larger |  -.0848079   .5361811    -0.16   0.874    -1.136048    .9664321
       _cons |  -1.186477   1.161393    -1.02   0.307    -3.463511    1.090556
------------------------------------------------------------------------------

. predict totesshat, xb
(1 missing value generated)

.
. regress totalmc totessay adultst sex avgmc eslflag ec011 en093 ma081 ma082 ma083 ///
>     smaller small large larger totesshat if smallest!=1

      Source |       SS       df       MS              Number of obs =    3710
-------------+------------------------------           F( 15,  3694) =  382.41
       Model | 35414.1533    15  2360.94355           Prob > F      =  0.0000
    Residual | 22805.9502  3694  6.17378186           R-squared     =  0.6083
-------------+------------------------------           Adj R-squared =  0.6067
       Total | 58220.1035  3709  15.6969813           Root MSE      =  2.4847

------------------------------------------------------------------------------
     totalmc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    totessay |   .2855834   .0054749    52.16   0.000     .2748493    .2963175
     adultst |   .2533495   .5795925     0.44   0.662    -.8830033    1.389702
         sex |  -.0897195   .0954784    -0.94   0.347    -.2769151     .097476
       avgmc |    .974884   .0523243    18.63   0.000     .8722967    1.077471
     eslflag |   .6744471   .8342319     0.81   0.419    -.9611532    2.310047
       ec011 |   .2925431     .09311     3.14   0.002     .1099909    .4750952
       en093 |  -1.588716   .8519162    -1.86   0.062    -3.258988    .0815565
       ma081 |   .2995655   .1798828     1.67   0.096    -.0531138    .6522448
```

```
    ma082 |   .8159711   .1933787     4.22   0.000     .4368315    1.195111
    ma083 |   1.635256   .1517372    10.78   0.000     1.337759    1.932753
  smaller |   .2715921   .1450994     1.87   0.061    -.0128907    .5560749
    small |     .04373   .1537008     0.28   0.776    -.2576168    .3450769
    large |   .1981185    .174948     1.13   0.258    -.1448856    .5411227
   larger |   -.086771   .1785655    -0.49   0.627    -.4368675    .2633256
totesshat |  -.3380613   .0261736   -12.92   0.000    -.3893774   -.2867452
    _cons |  -.3038295   .4033559    -0.75   0.451    -1.094652    .4869926
------------------------------------------------------------------------------
```

The similar estimation routine to test for the endogeneity of the multiple-choice test score in the essay equation yields a calculated z test statistic of -11.71, which far exceeds the absolute value of its 1.96 critical value. Thus, the null hypothesis of an exogenous multiple-choice score as an explanatory variable for the essay score is rejected. As theorized, the multiple-choice score is endogenous in an explanation of the essay score.

```
. regress totalmc avgessay adultst sex avgmc eslflag ec011 en093 ma081 ma082 ma083 ///
>    smaller small large larger if smallest!=1

      Source |       SS       df       MS              Number of obs =    3710
-------------+------------------------------           F( 14,  3695) =  124.06
       Model |  18615.7882    14  1329.69916           Prob > F      =  0.0000
    Residual |  39604.3153  3695  10.7183533           R-squared     =  0.3197
-------------+------------------------------           Adj R-squared =  0.3172
       Total |  58220.1035  3709  15.6969813           Root MSE      =  3.2739

------------------------------------------------------------------------------
     totalmc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    avgessay |  -.0441094   .0283459    -1.56   0.120    -.0996846    .0114658
     adultst |   .2618392   .7635394     0.34   0.732    -1.235161    1.758839
         sex |  -.1255075   .1194247    -1.05   0.293    -.3596523    .1086372
       avgmc |   .9734594   .0681195    14.29   0.000      .839904    1.107015
     eslflag |   .5831376   1.099197     0.53   0.596    -1.571954    2.738229
       ec011 |   .2547603   .1193565     2.13   0.033     .0207492    .4887713
       en093 |  -1.377667   1.114567    -1.24   0.217    -3.562894    .8075597
       ma081 |   .2430779   .2330463     1.04   0.297    -.2138341    .6999899
       ma082 |   .7510049   .2507815     2.99   0.003     .2593213    1.242689
       ma083 |   1.428892   .1502039     9.51   0.000     1.134401    1.723382
     smaller |   .2536501   .1902446     1.33   0.183    -.1193446    .6266448
       small |    .050818   .2026856     0.25   0.802    -.3465686    .4482046
       large |   .1799134   .2301129     0.78   0.434    -.2712475    .6310742
      larger |  -.0823205    .235116    -0.35   0.726    -.5432904    .3786495
       _cons |  -.2415657    .509272    -0.47   0.635    -1.240048    .7569162
------------------------------------------------------------------------------

. predict totmchat, xb
(1 missing value generated)

.
. regress totessay totalmc adultst sex avgessay eslflag ec011 en093 ma081 ma082 ///
ma083 smaller small large larger totmchat if smallest!=1

      Source |       SS       df       MS              Number of obs =    3710
-------------+------------------------------           F( 15,  3694) =  407.25
       Model |  196137.341    15  13075.8227           Prob > F      =  0.0000
    Residual |      118606  3694  32.1077423           R-squared     =  0.6232
-------------+------------------------------           Adj R-squared =  0.6216
       Total |  314743.341  3709  84.8593533           Root MSE      =  5.6664

------------------------------------------------------------------------------
    totessay |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     totalmc |   1.485222    .028473    52.16   0.000     1.429398    1.541047
     adultst |  -.1690792   1.322524    -0.13   0.898    -2.762028     2.42387
         sex |   .6854633   .2096929     3.27   0.001      .274338    1.096589
    avgessay |   .8417622   .0443223    18.99   0.000     .7548637    .9286608
     eslflag |   1.723699   1.905293     0.90   0.366    -2.011832    5.459229
```

```
         ec011 |   .7128703   .2096791     3.40   0.001     .3017721    1.123969
         en093 |  -3.983249    1.93672    -2.06   0.040    -7.780395    -.186104
         ma081 |   1.069629   .4036853     2.65   0.008     .2781608    1.861097
         ma082 |   1.217027   .4438483     2.74   0.006     .3468152    2.087239
         ma083 |   3.892551   .3175896    12.26   0.000     3.269883    4.515219
       smaller |   .3348226   .3255068     1.03   0.304     -.303368    .9730132
         small |  -.1364831   .3501242    -0.39   0.697    -.8229389    .5499726
         large |   .3418926   .3980484     0.86   0.390    -.4385237    1.122309
        larger |  -.0825121   .4071204    -0.20   0.839     -.880715    .7156909
      totmchat |  -1.457335   .1244158   -11.71   0.000    -1.701265   -1.213404
         _cons |  -1.179741   .8580241    -1.37   0.169    -2.861988    .5025071
------------------------------------------------------------------------------
```

CONCLUDING COMMENTS

This cookbook-type introduction to the use of instrumental variables and two-stage least squares regression and testing for endogeneity has just scratched the surface of this controversial problem in statistical estimation and inference. It was intended to enable researchers to begin using instrumental variables in their work and to enable readers of that work to have an idea of what is being done. To learn more about these methods there is no substitute for a graduate level textbook treatment such as that found in William Greene's *Econometric Analysis*.

REFERENCES

Becker, William E. and Carol Johnston (1999).“The Relationship Between Multiple Choice and Essay Response Questions in Assessing Economics Understanding,” *Economic Record* (Economic Society of Australia), Vol. 75 (December): 348-357.

Greene, William (2003). *Econometric Analysis*. 5th Edition, New Jersey: Prentice Hall.

ENDNOTES

[i] As stated in Module Two, Part One, the 2SLS coefficients are consistent but not necessarily unbiased. Consistency is an asymptotic property for which there are no adjustments for degrees of freedom. Nevertheless, the default in the standard STATA routine for 2SLS, “ivreg,” adjusts standard errors for the degrees of freedom. As an alternative, if your institution permits downloads from STATA's user-written routines, then the “ivreg2” command rather than “ivreg” can be employed. The “ivreg2” command makes no adjustment for degrees of freedom.

To use ivreg2, type *findit ivreg2* into the STATA command window, where a list of information and links to download this routine appears. Click on one of the download links and STATA automatically downloads and installs the routine for use. Users can then access the documentation for this routine by typing *help ivreg2*.

If users do not have access to ivreg2 or are not permitted to download user-written routines on the machine in use, the following code provides the unadjusted standard errors after running a model using ivreg:

```
matrix large_sample_se=e(b)
matrix large_sample_var=e(V)*e(df_r)/e(N)
      local ncol=colsof(large_sample_se)
      forvalues i=1/`ncol' {
  matrix large_sample_se[1,`i']=sqrt(large_sample_var[`i',`i'])
}
matrix list large_sample_se
```

[ii] Notice that the size variables (smaller to larger) show 3711 observations but the others show the correct 3710. This was an artifact of the way the size variables were created in STATA. The extra blank space has no relevance and is ignored in the calculations that are all based on the original 3710 observations.

MODULE TWO, PART FOUR: ENDOGENEITY, INSTRUMENTAL VARIABLES AND TWO-STAGE LEAST SQUARES IN ECONOMIC EDUCATION RESEARCH USING SAS

Part Four of Module Two provides a cookbook-type demonstration of the steps required to use SAS to address problems of endogeneity using a two-stage least squares, instrumental variable estimator. The Durbin, Hausman and Wu specification test for endogeneity is also demonstrated. Users of this model need to have completed Module One, Parts One and Four, and Module Two, Part One. That is, from Module One, users are assumed to know how to get data into SAS, recode and create variables within SAS, and run and interpret regression results. From Module Two, Part One, they are expected to have an understanding of the problem of and source of endogeneity and the basic idea behind an instrumental variable approach and the two-stage least squares method. The Becker and Johnston (1999) data set is used throughout this module for demonstration purposes only. Module Two, Parts Two and Three demonstrate in LIMDEP and STATA what is done here in SAS.

THE CASE

As described in Module Two, Part One, Becker and Johnston (1999) called attention to classroom effects that might influence multiple-choice and essay type test taking skills in economics in different ways. For example, if the student is in a classroom that emphasizes skills associated with multiple choice testing (*e.g.*, risk-taking behavior, question analyzing skills, memorization, and keen sense of judging between close alternatives), then the student can be expected to do better on multiple-choice questions. By the same token, if placed in a classroom that emphasizes the skills of essay test question answering (*e.g.*, organization, good sentence and paragraph construction, obfuscation when uncertain, logical argument, and good penmanship), then the student can be expected to do better on the essay component. Thus, Becker and Johnston attempted to control for the type of class of which the student is a member. Their measure of "teaching to the multiple-choice questions" is the mean score or mark on the multiple-choice questions for the school in which the i^{th} student took the 12^{th} grade economics course. Similarly, the mean school mark or score on the essay questions is their measure of the i^{th} student's exposure to essay question writing skills.

In equation form, the two equations that summarize the influence of the various covariates on multiple-choice and essay test questions are written as the follow structural equations:

$$M_i = \rho_{21} + \rho_{22} W_i + \rho_{23} \bar{M}_i + \sum_{j=4}^{J} \rho_{2j} X_{ij} + U_i^* .$$

$$W_i = \rho_{31} + \rho_{32} M_i + \rho_{33} \bar{W}_i + \sum_{j=4}^{J} \rho_{3j} X_{ij} + V_i^* .$$

M_i and W_i are the i^{th} student's respective scores on the multiple-choice test and essay test. $\bar{M}_i$ and $\bar{W}_i$ are the mean multiple-choice and essay test scores at the school where the i^{th} student took the 12[th] grade economics course. The X_{ij} variables are the other exogenous variables (such as gender, age, English a second language, etc.) used to explain the i^{th} student's multiple-choice and essay marks, where the ρs are parameters to be estimated. The inclusion of the mean multiple-choice and mean essay test scores in their respective structural equations, and their exclusion from the other equation, enables both of the structural equations to be identified within the system.

As shown in Module Two, Part One, the least squares estimation of the ρs involves bias because the error term U_i^* is related to W_i, in the first equation, and V_i^* is related to M_i, in second equation. Instruments for regressors W_i and M_i are needed. Because the reduced form equations express W_i and M_i solely in terms of exogenous variables, they can be used to generate the respective instruments:

$$M_i = \Gamma_{21} + \Gamma_{22}\bar{W}_i + \Gamma_{23}\bar{M}_i + \sum_{j=4}^{J}\Gamma_{2j}X_{ij} + U_i^{**}.$$

$$W_i = \Gamma_{31} + \Gamma_{32}\bar{M}_i + \Gamma_{33}\bar{W}_i + \sum_{j=4}^{J}\Gamma_{3j}X_{ij} + V_i^{**}.$$

The reduced form parameters (Γs) are functions of the ρs, and the reduced form error terms U^{**} and V^{**} are functions of U^* and V^*, which are not related to any of the regressors in the reduced form equations.

We could estimate the reduced form equations and get $\hat{M}_i$ and $\hat{W}_i$. We could then substitute $\hat{M}_i$ and $\hat{W}_i$ into the structural equations as proxy regressors (instruments) for M_i and W_i. The least squares regression of M_i on $\hat{W}_i$, $\bar{M}_i$ and the Xs and a least squares regression of W_i on $\hat{M}_i$, $\bar{W}_i$ and the Xs would yield consistent estimates of the respective ρs, but the standard errors would be incorrect. SAS can automatically do all the required estimations with the two-stage, least squares command:

```
proc syslin data= dataset_name 2sls;
        endogenous  p;
        instruments y u s;
        Equation 1: model q = p y s;
        Equation 2: model q = p u;
run;
```

This ‘proc syslin’ package command, however, inappropriate adjustment for degrees of freedom in calculating the standard errors. What follows is the correct two-step procedure for the asymptotic efficient estimators, which involve no adjustment for degrees of freedom.

TWO-STAGE, LEAST SQUARES IN SAS

The Becker and Johnston (1999) data are in the file named “Bill.CSV.” The file Bill.CSV can be read into SAS with the following read command (the file may be located anywhere on your hard drive but here it is located on the e drive):

```
data work.bill; infile 'e:\ BILL.CSV' delimiter = ',' missover dsd;

informat student best32.;     informat school best32.;     informat size best32.;
informat other best32.;       informat birthday best32.;   informat sex best32.;
informat eslflag best32.;     informat adultst best32.;    informat mc1 best32.;
informat mc2 best32.;         informat mc3 best32.;        informat mc4 best32.;
informat mc5 best32.;         informat mc6 best32.;        informat mc7 best32.;
informat mc8 best32.;         informat mc9 best32.;        informat mc10 best32.;
informat mc11 best32.;        informat mc12 best32.;       informat mc13 best32.;
informat mc14 best32.;        informat mc15 best32.;       informat mc16 best32.;
informat mc17 best32.;        informat mc18 best32.;       informat mc19 best32.;
informat mc20 best32.;        informat totalmc best32.;    informat avgmc best32.;
informat essay1 best32.;      informat essay2 best32.;     informat essay3 best32.;
informat essay4 best32.;      informat totessay best32.;   informat avgessay best32.;
informat totscore best32.;    informat avgscore best32.;   informat ma081 best32.;
informat ma082 best32.;       informat ec011 best32.;      informat ec012 best32.;
informat ma083 best32.;       informat en093 best32.;

format student best12.;       format school best12.;       format size best12.;
format other best12.;         format birthday best12.;     format sex best12.;
format eslflag best12.;       format adultst best12.;      format mc1 best12.;
format mc2 best12.;           format mc3 best12.;          format mc4 best12.;
format mc5 best12.;           format mc6 best12.;          format mc7 best12.;
format mc8 best12.;           format mc9 best12.;          format mc10 best12.;
format mc11 best12.;          format mc12 best12.;         format mc13 best12.;
format mc14 best12.;          format mc15 best12.;         format mc16 best12.;
format mc17 best12.;          format mc18 best12.;         format mc19 best12.;
format mc20 best12.;          format totalmc best12.;      format avgmc best12.;
format essay1 best12.;        format essay2 best12.;       format essay3 best12.;
format essay4 best12.;        format totessay best12.;     format avgessay best12.;
format totscore best12.;      format avgscore best12.;     format ma081 best12.;
format ma082 best12.;         format ec011 best12.;        format ec012 best12.;
format ma083 best12.;         format en093 best12.;
```

input student school size other birthday sex eslflag adultst mc1 mc2 mc3 mc4 mc5 mc6 mc7 mc8 mc9 mc10 mc11 mc12 mc13 mc14 mc15 mc16 mc17 mc18 mc19 mc20 totalmc avgmc essay1 essay2 essay3 essay4 totessay avgessay totscore avgscore ma081 ma082 ec011 ec012 ma083 en093; run;

Using these recode and create commands, yields the following relevant variable definitions:

```
data Bill;
    set M2P4.Bill;
        smallest = 0;
        smaller = 0;
        small = 0;
        large = 0;
        larger = 0;
        largest = 0;
        if size > 0  & size < 10 then smallest = 1;
        if size > 9  & size < 20 then smaller = 1;
        if size > 19 & size < 30 then small = 1;
        if size > 29 & size < 40 then large = 1;
        if size > 39 & size < 50 then larger = 1;
        if size > 49 then largest = 1;
run;
```

TOTALMC: Student's score on 12th grade economics multiple-choice exam (M_i).

AVGMC: Mean multiple-choice score for students at school ($\bar{M}_i$).

TOTESSAY: Student's score on 12th grade economics essay exam (W_i).

AVGESSAY: Mean essay score for students at school ($\bar{W}_i$).

ADULTST = 1, if a returning adult student, and 0 otherwise.
SEX = GENDER = 1 if student is female and 0 is male.
ESLFLAG = 1 if English is not student's first language and 0 if it is.
EC011 = 1 if student enrolled in first semester 11 grade economics course, 0 if not.
EN093 = 1 if student was enrolled in ESL English course, 0 if not
MA081 = 1 if student enrolled in the first semester 11 grade math course, 0 if not.
MA082 = 1 if student was enrolled in the second semester 11 grade math course, 0 if not.
MA083 = 1 if student was enrolled in the first semester 12 grade math course, 0 if not.
SMALLER = 1 if student from a school with 10 to 19 test takers, 0 if not.
SMALL = 1 if student from a school with 20 to 29 test takers, 0 if not.
LARGE = 1 if student from a school with 30 to 39 test takers, 0 if not.
LARGER = 1 if student from a school with 40 to 49 test takers, 0 if not.

In all of the regressions, the effect of being at a school with more than 49 test takers is captured in the constant term, against which the other dummy variables are compared. The smallest schools need to be rejected to treat the mean scores as exogenous and unaffected by any individual student's test performance, which is accomplished with the following command:

```
data Bill; set Bill;
        if smallest = 1 then delete;
        if student = . then delete;
 run;
```

The descriptive statistics on the relevant variables are then obtained with the following command, yielding the SAS output table shown:

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
totalmc	3710	12.4355795	3.9619416	0	20.0000000
avgmc	3710	12.4355800	1.9726377	6.4166670	17.0714300
totessay	3710	18.1380054	9.2119137	0	40.0000000
avgessay	3710	18.1380059	4.6680707	5.7000000	29.7857100
adultst	3710	0.0051213	0.0713894	0	1.0000000
sex	3710	0.3905660	0.4879430	0	1.0000000
eslflag	3710	0.0641509	0.2450547	0	1.0000000
ec011	3710	0.6770889	0.4676521	0	1.0000000
en093	3710	0.0622642	0.2416673	0	1.0000000
ma081	3710	0.5913747	0.4916460	0	1.0000000
ma082	3710	0.5487871	0.4976812	0	1.0000000
ma083	3710	0.4202156	0.4936599	0	1.0000000
smaller	3710	0.4622642	0.4986412	0	1.0000000
small	3710	0.2072776	0.4054108	0	1.0000000
large	3710	0.1064690	0.3084785	0	1.0000000
larger	3710	0.0978437	0.2971432	0	1.0000000

For comparison with the two-stage least squares results, we start with the least squares regressions shown after this paragraph. The least squares estimations are typical of those found in multiple-choice and essay score correlation studies, with correlation coefficients of 0.77 and 0.78. The essay mark or score, *W*, is the most significant variable in the multiple-choice score regression (first of the two tables) and the multiple-choice mark, *M*, is the most significant variable in the essay regression (second of the two tables). Results like these have led researchers to conclude that the essay and multiple-choice marks are good predictors of each other. Notice also that both the mean multiple-choice and mean essay marks are significant in their respective equations, suggesting that something in the classroom environment or group experience influences individual test scores. Finally, being female has a significant negative effect on the multiple choice-test score, but a significant positive effect on the essay score, as expected from the least squares results reported by others. We will see how these results hold up in the two-stage least squares regressions.

```
proc reg data = Bill; model totalmc totessay adultst sex avgmc eslflag
ec011 en093 ma081 ma082 ma083 smaller small large larger; quit;
```

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	14	34384	2456.01457	380.73	<.0001
Error	3695	23836	6.45085		
Corrected Total	3709	58220			

Root MSE	2.53985	R-Square	0.5906
Dependent Mean	12.43558	Adj R-Sq	0.5890
Coeff Var	20.42408		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.65496	0.33936	7.82	<.0001
totessay	1	0.27079	0.00547	49.48	<.0001
adultst	1	0.46749	0.59221	0.79	0.4299
sex	1	-0.52595	0.09129	-5.76	<.0001
avgmc	1	0.37938	0.02529	15.00	<.0001
eslflag	1	0.39333	0.85246	0.46	0.6445
ec011	1	0.01723	0.09265	0.19	0.8525
en093	1	-0.31173	0.86494	-0.36	0.7186
ma081	1	-0.12081	0.18084	-0.67	0.5042
ma082	1	0.38271	0.19467	1.97	0.0494
ma083	1	0.37038	0.11848	3.13	0.0018
smaller	1	0.06721	0.14743	0.46	0.6485
small	1	-0.00569	0.15706	-0.04	0.9711
large	1	0.06636	0.17853	0.37	0.7101
larger	1	0.05655	0.18218	0.31	0.7563

proc reg data = Bill; model totessay = totalmc adultst sex avgessay eslflag ec011 en093 ma081 ma082 ma083 smaller small large larger; quit;

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	14	191732	13695	411.37	<.0001
Error	3695	123011	33.29129		
Corrected Total	3709	314743			

Root MSE	5.76986	R-Square	0.6092
Dependent Mean	18.13801	Adj R-Sq	0.6077
Coeff Var	31.81089		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-8.94870	0.55428	-16.14	<.0001
totalmc	1	1.40890	0.02822	49.92	<.0001
adultst	1	-0.82915	1.34546	-0.62	0.5378
sex	1	1.23996	0.20801	5.96	<.0001
avgessay	1	0.40002	0.02371	16.87	<.0001
eslflag	1	0.45114	1.93694	0.23	0.8158
ec011	1	0.29854	0.21045	1.42	0.1561
en093	1	-2.02088	1.96470	-1.03	0.3037
ma081	1	0.84951	0.41061	2.07	0.0386
ma082	1	0.15909	0.44250	0.36	0.7192
ma083	1	1.80954	0.26794	6.75	<.0001
smaller	1	0.61707	0.33054	1.87	0.0620
small	1	0.26934	0.35477	0.76	0.4478
large	1	0.26464	0.40526	0.65	0.5138
larger	1	0.06150	0.41437	0.15	0.8820

Theoretical considerations discussed in Module Two, Part One, suggest that these least squares estimates involve a simultaneous equation bias that is brought about by an apparent reverse causality between the two forms of testing. Consistent estimation of the parameters in this simultaneous equation system is possible with two-stage least squares, where our instrument ($\hat{M}_i$) for M_i is obtained by a least squares regression of M_i on SEX, ADULTST, AVGMC, AVGESSAY, ESLFLAG,SMALLER,SMALL, LARGE, LARGER, EC011, EN093, MA081, MA082, and MA083. Our instrument ($\hat{W}_i$) for W_i is obtained by a least squares regression of W_i on SEX, ADULTST, AVGMC, AVGESSAY, ESLFLAG, SMALLER, SMALL, LARGE, LARGER, EC011, EN093, MA081, MA082, and MA083. SAS will do these regressions and the subsequent regressions for M and W employing these instruments via the following commands, which yield the subsequent outputs:

```
proc model data = bill;
    instruments sex adultst avgmc avgessay eslflag smaller
        small large larger ec011 en093 ma081 ma082 ma083;
    totalmc = constant_1 + totessay_1*totessay + adultst_1*adultst +
        sex_1*sex + avgmc_1*avgmc + eslflag_1*eslflag +
        eco11_1*ec011 + en093_1*en093 + ma081_1*ma081 +
        ma082_1*ma082 + ma083_1*ma083 + smaller_1*smaller +
        small_1*small + large_1*large + larger_1*larger;

    totessay = constant_2 + totalm_2*totalmc + adultst_2*adultst +
        sex_2*sex + avgessay_2*avgessay + eslflag_2*eslflag +
        ec011_2*ec011 + en093_2*en093 + ma081_2*ma081 +
        a082_2*ma082+ ma083_2*ma083 + smaller_2*smaller +
        small_2*small + large_2*large + larger_2*larger;

     fit totalmc totessay / 2sls outv=vdata vardef=N;
quit;
```

Nonlinear 2SLS Summary of Residual Errors

Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
totalmc	15	3695	46345.2	12.4920	3.5344	0.2040	0.2010
totessay	15	3695	202719	54.6411	7.3920	0.3559	0.3535

\- _

Number of Observations		Statistics for System	
Used	3710	Objective	4.76E-23
Missing	0	Objective*N	1.766E-19

Nonlinear 2SLS Parameter Estimates

Parameter	Estimate	Approx Std Err	t Value	Approx Pr > |t|
constant_1	-0.30383	0.5738	-0.53	0.5965
totessay_1	-0.05248	0.0364	-1.44	0.1496
adultst_1	0.253349	0.8244	0.31	0.7586
sex_1	-0.08972	0.1358	-0.66	0.5089
avgmc_1	0.974884	0.0744	13.10	<.0001
eslflag_1	0.674447	1.1867	0.57	0.5698
ecoll_1	0.292543	0.1324	2.21	0.0273
en093_1	-1.58872	1.2118	-1.31	0.1899
ma081_1	0.299566	0.2559	1.17	0.2418
ma082_1	0.815971	0.2751	2.97	0.0030
ma083_1	1.635256	0.2158	7.58	<.0001
smaller_1	0.271592	0.2064	1.32	0.1883
small_1	0.04373	0.2186	0.20	0.8415
large_1	0.198118	0.2489	0.80	0.4260
larger_1	-0.08677	0.2540	-0.34	0.7327
constant_2	-1.17974	1.1193	-1.05	0.2920
totalm_2	0.027888	0.1580	0.18	0.8599
adultst_2	-0.16908	1.7253	-0.10	0.9219
sex_2	0.685463	0.2736	2.51	0.0123
avgessay_2	0.841762	0.0578	14.56	<.0001
eslflag_2	1.723699	2.4855	0.69	0.4880
ec011_2	0.71287	0.2735	2.61	0.0092
en093_2	-3.98325	2.5265	-1.58	0.1150
ma081_2	1.069629	0.5266	2.03	0.0423
ma082_2	1.217027	0.5790	2.10	0.0356
ma083_2	3.892551	0.4143	9.40	<.0001
smaller_2	0.334822	0.4246	0.79	0.4305
small_2	-0.13648	0.4567	-0.30	0.7651
large_2	0.341892	0.5193	0.66	0.5103
larger_2	-0.08251	0.5311	-0.16	0.8765

The 2SLS results differ from the least squares results in many ways. The essay mark or score, *W*, is no longer a significant variable in the multiple-choice regression and the multiple-choice mark, *M*, is likewise insignificant in the essay regression. Each score appears to be measuring something different when the regressor and error-term-induced bias is eliminated by our instrumental variable estimators.

Both the mean multiple-choice and mean essay scores continue to be significant in their respective equations. But now being female is insignificant in explaining the multiple-choice test score. Being female continues to have a significant positive effect on the essay score.

DURBIN, HAUSMAN AND WU TEST FOR ENDOGENEITY

The theoretical argument is strong for treating multiple-choice and essay scores as endogenous when employed as regressors in the explanation of the other. Nevertheless, this endogeneity can be tested with the Durbin, Hausman and Wu specification test, which is a two-step procedure in SAS.

Either a Wald statistic, in a Chi-square (χ^2) test with K^* degrees of freedom, or an F statistic with K^* and $n - (K + K^*)$ degrees of freedom, is used to test the joint significance of the contribution of the predicted values ($\hat{\mathbf{X}}^*$) of a regression of the K^* endogenous regressors, in matrix $\mathbf{X^*}$**,** on the exogenous variables (and column of ones for the constant term) in matrix $\mathbf{Z}$**:**

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \hat{\mathbf{X}}^* \boldsymbol{\gamma} + \boldsymbol{\varepsilon}^*,$$

where $\mathbf{X}^* = \mathbf{Z}\boldsymbol{\lambda} + \mathbf{u}$, $\hat{\mathbf{X}}^* = \mathbf{Z}\hat{\boldsymbol{\lambda}}$, and $\hat{\boldsymbol{\lambda}}$ is a least squares estimator of $\boldsymbol{\lambda}$.

$H_o: \boldsymbol{\gamma} = 0$, the variables in $\mathbf{Z}$ are exogenous

$H_A: \boldsymbol{\gamma} \neq 0$, at least one of the variables in $\mathbf{Z}$ is endogenous

In our case, $K^* = 1$ when the essay score is to be tested as an endogenous regressor in the multiple-choice equation and when the multiple-choice regressor is to be tested as endogenous in the essay equation. $\hat{\mathbf{X}}^*$ is an $n \times 1$ vector of predicted essay scores from a regression of essay scores on all the exogenous variables (for subsequent use in the multiple-choice equation) or an $n \times 1$ vector of predicted multiple-choice scores from a regression of multiple-choice scores on all the exogenous variables (for subsequent use in the essay equation). Because $K^* = 1$, the relevant test statistic is either the t, with $n - (K + K^*)$ degrees of freedom for small n or the standard normal, for large n.

In SAS, the predicted essay score is obtained by the following command, where the specification "output out=essaypredict p=Essayhat;" tells SAS to predict the essay scores and keep them as a variable called "Essayhat":

```
proc reg data = bill; model totessay = adultst sex avgessay avgmc eslflag
        ec011 en093 ma081 ma082 ma083 smaller small large larger;
output out=essaypredict p=Essayhat; quit;
```

The predicted essay scores are then added as a regressor in the original multiple-choice regression:

```
proc reg data = essaypredict; model totalmc = totessay adultst sex avgmc eslflag
        ec011 en093 ma081 ma082 ma083 smaller small large larger Essayhat;
quit;
```

The test statistic for the Essayhat coefficient is then used in the test of endogeneity. In the below SAS output, we see that the calculated standard normal test statistic z value is −12.916, which far exceeds the absolute value of the 0.05 percent Type I error critical 1.96 standard normal value. Thus, the null hypothesis of an exogenous essay score as an explanatory variable for the multiple-choice score is rejected. As theorized, the essay score is endogenous in an explanation of the multiple-choice score.

proc reg data = bill; model totessay = adultst sex avgessay avgmc eslflag ec011 en093 ma081 ma082 ma083 smaller small large larger; output out=essaypredict p=Essayhat; quit;

-

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	14	108775	7769.62501	139.38	<.0001
Error	3695	205969	55.74251		
Corrected Total	3709	314743			

Root MSE	7.46609	R-Square	0.3456
Dependent Mean	18.13801	Adj R-Sq	0.3431
Coeff Var	41.16269		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-1.18648	1.16139	-1.02	0.3070
adultst	1	-0.16178	1.74125	-0.09	0.9260
sex	1	0.68196	0.27235	2.50	0.0123
avgessay	1	0.84053	0.06464	13.00	<.0001
avgmc	1	0.02715	0.15535	0.17	0.8613
eslflag	1	1.73996	2.50671	0.69	0.4877
ec011	1	0.71997	0.27219	2.65	0.0082
en093	1	-4.02167	2.54176	-1.58	0.1137
ma081	1	1.07641	0.53146	2.03	0.0429
ma082	1	1.23797	0.57191	2.16	0.0305
ma083	1	3.93240	0.34254	11.48	<.0001
smaller	1	0.34190	0.43385	0.79	0.4307
small	1	-0.13507	0.46222	-0.29	0.7701
large	1	0.34691	0.52477	0.66	0.5086
larger	1	-0.08481	0.53618	-0.16	0.8743

proc reg data = essaypredict; model totalmc = totessay adultst sex avgmc eslflag ec011 en093 ma081 ma082 ma083 smaller small large larger Essayhat; quit;

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-0.30383	0.40336	-0.75	0.4513
totessay		1	0.28558	0.00547	52.16	<.0001
adultst		1	0.25335	0.57959	0.44	0.6621
sex		1	-0.08972	0.09548	-0.94	0.3474
avgmc		1	0.97488	0.05232	18.63	<.0001
eslflag		1	0.67445	0.83423	0.81	0.4189
ec011		1	0.29254	0.09311	3.14	0.0017
en093		1	-1.58872	0.85192	-1.86	0.0623
ma081		1	0.29957	0.17988	1.67	0.0959
ma082		1	0.81597	0.19338	4.22	<.0001
ma083		1	1.63526	0.15174	10.78	<.0001
smaller		1	0.27159	0.14510	1.87	0.0613
small		1	0.04373	0.15370	0.28	0.7760
large		1	0.19812	0.17495	1.13	0.2575
larger		1	-0.08677	0.17857	-0.49	0.6270
essayhat	Predicted Value of totessay	1	-0.33806	0.02617	-12.92	<.0001

The similar estimation routine to test for the endogeneity of the multiple-choice test score in the essay equation yields a calculated *z* test statistic of −11.713, which far exceeds the absolute value of its 1.96 critical value. Thus, the null hypothesis of an exogenous multiple-choice score as an explanatory variable for the essay score is rejected. As theorized, the multiple-choice score is endogenous in an explanation of the essay score.

proc reg data = bill; model totalmc = adultst sex avgmc avgessay eslflag ec011 en093 ma081 ma082 ma083 smaller small large larger; output out=mcpredict p=MChat; quit;

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	14	18616	1329.69916	124.06	<.0001
Error	3695	39604	10.71835		
Corrected Total	3709	58220			

Root MSE	3.27389	R-Square	0.3197
Dependent Mean	12.43558	Adj R-Sq	0.3172
Coeff Var	26.32680		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.24157	0.50927	-0.47	0.6353
adultst	1	0.26184	0.76354	0.34	0.7317
sex	1	-0.12551	0.11942	-1.05	0.2934
avgmc	1	0.97346	0.06812	14.29	<.0001
avgessay	1	-0.04411	0.02835	-1.56	0.1198
eslflag	1	0.58314	1.09920	0.53	0.5958
ec011	1	0.25476	0.11936	2.13	0.0329
en093	1	-1.37767	1.11457	-1.24	0.2165
ma081	1	0.24308	0.23305	1.04	0.2970
ma082	1	0.75100	0.25078	2.99	0.0028
ma083	1	1.42889	0.15020	9.51	<.0001
smaller	1	0.25365	0.19024	1.33	0.1825
small	1	0.05082	0.20269	0.25	0.8020
large	1	0.17991	0.23011	0.78	0.4344
larger	1	-0.08232	0.23512	-0.35	0.7263

proc reg data = mcpredict; model totessay = totalmc adultst sex avgessay eslflag ec011 en093 ma081 ma082 ma083 smaller small large larger MChat; quit;

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-1.17974	0.85802	-1.37	0.1692
totalmc		1	1.48522	0.02847	52.16	<.0001
adultst		1	-0.16908	1.32252	-0.13	0.8983
sex		1	0.68546	0.20969	3.27	0.0011
avgessay		1	0.84176	0.04432	18.99	<.0001
eslflag		1	1.72370	1.90529	0.90	0.3657
ec011		1	0.71287	0.20968	3.40	0.0007
en093		1	-3.98325	1.93672	-2.06	0.0398
ma081		1	1.06963	0.40369	2.65	0.0081
ma082		1	1.21703	0.44385	2.74	0.0061
ma083		1	3.89255	0.31759	12.26	<.0001
smaller		1	0.33482	0.32551	1.03	0.3037
small		1	-0.13648	0.35012	-0.39	0.6967
large		1	0.34189	0.39805	0.86	0.3904
larger		1	-0.08251	0.40712	-0.20	0.8394
MChat	Predicted Value of totalmc	1	-1.45733	0.12442	-11.71	<.0001

CONCLUDING COMMENTS

This cookbook-type introduction to the use of instrumental variables and two-stage least squares regression and testing for endogeneity has just scratched the surface of this controversial problem in statistical estimation and inference. It was intended to enable researchers to begin using instrumental variables in their work and to enable readers of that work to have an idea of what is being done. To learn more about these methods there is no substitute for a graduate level textbook treatment such as that found in William Greene's *Econometric Analysis*.

REFERENCES

Becker, William E. and Carol Johnston (1999).“The Relationship Between Multiple Choice and Essay Response Questions in Assessing Economics Understanding,” *Economic Record* (Economic Society of Australia), Vol. 75 (December): 348-357.

Greene, William (2003). *Econometric Analysis*. 5th Edition, New Jersey: Prentice Hall.