978–0–19–928654–6 17-Box-c17 OUP249-Box (Typeset by SPI, Delhi) 404 of 431 March 14, 2008 20:26

CHAPTER 17

ENDOGENEITY AND STRUCTURAL EQUATION ESTIMATION IN POLITICAL SCIENCE

JOHN E. JACKSON

This chapter discusses a topic—endogeneity—where the importation of econometric methods substantially advanced the practice of empirical work in political science, enabling important substantive findings. The first section examines why econometrics, and particularly the treatment of endogeneity, proved so valuable for political scientists. The chapter discusses and applies the most prominent methods used to deal with endogeneity. We also include a critique of the critical conditions needed to support the use and interpretation of this method and discuss frequently used diagnostics developed to assess the severity of the problems. The estimation method and diagnostics are illustrated with a running example with data from a previous study of US congressional politics (Jackson and King 1989).

1. Econometrics and Political Science

The application and refinement of econometric techniques is one of the dominant themes of political methodology for the past forty years. The early expansion of these techniques is well documented by King (1991), Bartels and Brady (1993), and Jackson (1996). The reasons for the rapid and extensive spread of these techniques is fairly easy to understand, in retrospect. Econometric techniques shift attention from the observed data to a model of the behavior being investigated and of the process that likely generated the observed data, referred to as the data-generating process (DGP). This attention is present in the linear model but takes center stage with the subsequent interest in maximum likelihood estimation (MLE). (See King 1989 and econometrics texts such as Greene 2003, ch. 17.) Other chapters discuss MLE and its many extensions in detail and it occupies only a minor part of this chapter. This chapter focuses on one particular but vitally important part of the data-generating model.

1.1 The Linear Model and Observational Data

The experimental paradigm remains the object of emulation based on its ability to vary exogenously the magnitude of treatments among treatment and control groups and to completely randomize the assignment of subjects among the treatment and control groups. See Freedman, Pisani, and Purves (1998, chs. 1 and 2) for an excellent discussion of the requirements and advantages of good experimental design and the perils of observational studies. In many substantively important contexts, however, the experimentalists' manipulations and randomization are impossible. Social scientists are no more likely to be given the license to manipulate an economy to ascertain its impacts on voting or to initiate conflicts to study their duration than medical researchers are to force people to engage in risky behaviors. This leaves many empirical researchers with only observational data where there may not even be any well-defined and distinct treatment and control groups but only variations in the variables of interest.

Enter the linear model, $Y_i = X_i\beta + U_i$. The standard interpretation is that the values of the X_i variables can be treated as fixed and exogenously given, analogously to the treatments in an experimental study. Further, the values of U_i, representing omitted factors, are drawn from distributions with a zero mean for all values of X_i, i.e. $E(U_i) = 0$ for all i, analogously to randomization.[1] When these conditions are satisfied one has the equivalent of a well-designed experiment and the estimated regression coefficients, denoted as b are unbiased, i.e. $E(b) = \beta$. Additional conditions that further emulate the experimental setting, that the U_i's are iid, establish the linear estimator as the best, meaning minimum variance, unbiased linear estimator.

[1] The assumption of a zero mean can be relaxed to be a constant value for all i, which only alters the constant term in the model.

Finally, if the U's are normally distributed these estimates are normally distributed, permitting classical inference tests. With normally distributed U's the coefficients are the maximum likelihood estimates.

This chapter addresses a very specific and central aspect of the DGP assumed in the linear model—the assumption (a fiction for some) that X is equivalent to a fixed set of exogenously determined "treatments" that could be duplicated in an arbitrary number of replications. If the X's are not fixed exogenous treatments the condition that $E(U_i) = 0$ is hard to justify, which means the linear estimator is biased, $E(b - \beta) \neq 0$ and inconsistent, $plim(b - \beta) \neq 0$ negating the justifications for and desirability of the linear model.

1.2 Fixed Treatments and Endogeneity

There are classic situations that render the idea of exogenous and fixed X's questionable if not perfectly ludicrous. One is where variations in the left-hand side variable, the outcome variable, are likely causing changes in one of more of the right-hand side (RHS) variables, a condition referred to as simultaneity. The canonical economic example is the relationship between price and quantity. The demand equation predicts a negative relationship between p and q as larger quantities force lower prices in order to clear the market. The supply equation, on the other hand, predicts a positive association between p and q as higher prices induce more production. Given that one only observes actual quantities sold and the associated prices without direct manipulation of one but not the other it is impossible to estimate either relationship with the linear model described above.

Political science examples are abundant but three examples illustrate the problems well. Jacobson (1978) in a classic article explores how challenger and incumbent campaign expenditures affect the challenger's vote share in the election. He appropriately points out that candidate fundraising, hence expenditures, are not exogenous "treatments" but may be influenced by the expected vote shares. "OLS regression models presuppose . . . that spending produces votes." But, "The *expectation* [author's emphasis] that a candidate will do well may bring campaign contributions" (Jacobson 1978, 470), Bartels (1991) offers an excellent critique of the methods chosen by Jacobson and others to confront this endogeneity problem.

A second example is political economists' interest in the effect of institutions on economic outcomes. For example, does a PR form of parliamentary government result in more social welfare spending? In examining this and similar propositions Persson and Tabellini (2003, 114) say, "our inference becomes biased if the variation in constitutional rules used to explain performance is related to the random (unexplained) component of performance. Simultaneity problems can take the form of *reverse causation*, different forms of *selection bias*, and *measurement error* [authors' emphasis]." Their empirical strategy to overcome these biases includes but is not limited to the methods surveyed in this chapter. Acemoglu (2005) has an excellent review and assessment of their methods and models.

The last example is the effort by international relations scholars to estimate models of rivalry and reciprocity, such as arms races and tit-for-tat diplomacy, where the actions of country Y are modeled as a function of the actions of a competitor country, X. But there must be a tit for every tat, meaning the actions of X are not exogenous to the actions of country Y. As Dixon (1986, 434) says, "Although the reciprocity models . . . depict only the behavior of nation Y, each equation is implicitly paired with a complementary specification for nation X. . . . The obvious mutual dependency between endogenous variables y_{it} and x_{it} in this two-equation system signals a violation of the basic regression model." Dixon and others then proceed with some of the methods discussed below.

For the purposes of this chapter consider an example drawn from legislative politics. A theory of representation is that representatives' votes should be responsive to the mean preferences within their constituency, $V = \alpha + \beta \bar{P}$ (Achen 1978). Elections are the enforcement mechanism in this argument as representatives who deviate from the mean preference should lose votes and ultimately the election itself. Let *Marg* be representatives' electoral margins and *Dev* measure how much their votes deviates from constituency preferences, $Dev = |V - \bar{P}|$. The punishment model is

$$Marg_i = \gamma_1 Dev_i + X_1 \beta_1 + U_i, \tag{1}$$

with $\gamma_1 < 0$, meaning that increased deviations from constituency preferences lead to smaller electoral margins. X_1 represents other factors that may influence incumbents' vote margins, such as the partisan composition of the district. The difficulty in estimating equation (1) is that *Dev* is unlikely to be exogenously determined. Some Congressional scholars (e.g. Kingdon 1973, ch. 2) propose that members with larger electoral margins are freer to deviate from constituency preferences, and may do so. This proposition implies that

$$Dev_i = \gamma_2 Marg_i + X_2 \beta_2 + V_i, \tag{2}$$

with $\gamma_2 > 0$. X_2 describes pressures that lead members to vote differently from their constituents' preferences, such as pressure from party leaders. So, *Marg* and *Dev* in any observational study are simultaneously related, thus violating the linear model's basic assumption. Even if X_1 and X_2 contain all the variables that systematically affect representatives' deviations from constituency preferences and their electoral margins so that $E(U_i) = E(V_i) = 0$ the basic regression model gives biased and inconsistent estimates for γ_1 and γ_2. The simple correlation between *Marg* and *Dev* or a regression with either variable on the RHS and the other on the LHS will not reveal anything about the propensity of elections to function as the reward/punishment mechanism in Achen's model of representation.

The likely endogeneity of an explanatory variable is not limited to cases of simultaneity. There are situations with observational data where it is difficult to believe that one of the X's is fixed and would have the same values in subsequent replications of the "experiment." One must consider carefully the context of the quasi or natural experiment and decide if every X is really determined outside the process being studied.

This is a difficult requirement in many political science studies where some of the right-hand side variables themselves are measures of political behavior or outcomes.

As we will illustrate with our legislative example, representatives' partisanship may be a right-hand side variable in the equation for deviations, as party leaders are likely to be effective at getting members to vote with the party rather than with their constituents. If party is a right-hand side variable in the deviation equation it is likely endogenous. Each district's representative's party is the result of a series of political factors that may not be fixed, as required by the experimental paradigm being emulated by the linear model. Think about whether representatives' party affiliations are assigned randomly among districts! (See Jackson and King 1989.) Any explanatory variable, such as member partisanship, that cannot be justified as exogenous is implicitly the left-hand side variable in another equation. OLS estimates of an equation with this endogenous variables on the RHS will be consistent only if the stochastic term in this implicit equation is uncorrelated with the stochastic term in the equation being estimated.[2]

The condition of fixed and exogenous X's also fails if any of the right-hand side variables are measured with systematic or random error. The case of random measurement errors is the, relatively, easier one to confront and is the focus of this discussion.[3] Random measurement error does not create endogeneity in the causal way discussed above. It does, however, present identical problems and is approached with the same estimation strategy. Basic texts show that if the correct variables are X_i, but instead one observes $Z_i = X_i + \epsilon_i$ and estimates the model $Y_i = Z_i B + U_i$, the resulting coefficients are biased and inconsistent. Even if only one of K RHS variables is measured with error all the coefficients, not just the one attached to the erroneous variable, may be biased and inconsistent (Achen 1988).

These situations all lead to conditions that violate the basic assumptions of the linear model.[4] The explanatory, or treatment, variables are now correlated with the stochastic term implicit in the DGP and thus in the measures for Y, which yields biased and inconsistent estimates for the model's coefficients. The textbook treatment of this situation is as Greene (2003) describes, $E[\epsilon_i|X_i] \neq 0$ which means that $E[X_i\epsilon_i] = \eta$ and plim $(1/n)(X'\epsilon) = \eta$, which has the following implications for the estimated coefficients,

$$E[b|X] = \beta + (X'X)^{-1}X'\eta \neq \beta \quad \text{and}$$

$$\text{plim } b = \beta + \text{plim}\left(\frac{X'X}{n}\right)^{-1} \text{plim}\left(\frac{X'\epsilon}{n}\right) = \beta + \Sigma_x^{-1}\eta \neq \beta.$$

This is clearly a failure of the classic paradigm and requires new estimation strategies.

[2] These conditions, if met, would imply a hierarchical *and* recursive system in which case OLS estimates are consistent.

[3] Systematic measurement errors present much greater difficulties and require specialized estimators (see Berinsky 1999; Brady 1988; Jackson 1993) and will not be discussed here.

[4] Other situations may also lead to similar violations. These are often encountered in dynamic models that include lagged values for Y on the right-hand side of the equation and an autocorrelated stochastic term. This chapter does not address these situations, though they can be dealt with in ways similar to the subsequent discussions here.

978–0–19–928654–6 17-Box-c17 OUP249-Box (Typeset by SPI, Delhi) 409 of 431 March 14, 2008 20:26

1.3 OLS Estimates of the Margin and Deviation Relationship

We can illustrate this failure with estimates of the model of legislative margins and deviations. Data for estimating the model come from Jackson and King (1989) and Barone, Ujifusa, and Matthews (1979) and relate to the 1978 session of Congress. Representatives' voting behavior in 1978 is assessed using their ADA and ACA scores, with $V = .5^*(ADA - ACA + 100)$. The Jackson and King data contain an estimate of mean constituency preference on income redistribution, $\bar{P}$. Members' deviations are the absolute diffence between their vote score and the proportion of the constituency supporting income redistribution, $Dev = |V - \bar{P}|$. The assumption here is that the mean constituency preference for redistribution assesses the liberalness of the constituency. The absolute difference measures the extent to which the members' votes deviate from this constituency position. The variable *Marg* is incumbents' vote share in the 1978 election (the closer of the primary or general elections). The proposition implicit in Achen's and other representation models is that members who deviate more will have smaller re-election margins.

The model and data are being used to illustrate the methodological problems and estimation procedures associated with endogenous variables. This is not intended as a subtantive study of Congressional elections or of representation. Obvious variables are omitted from the analysis that would be required for the latter to be true—such as candidate quality, campaign expenditures, the representatives' characteristics, etc. But, in most cases addition of these variables complicates rather than resolves questions of specification and endogeniety.

The simple correlation between *Marg* and *Dev* is 0.13, a very slight positive association. Do members who deviate more do slightly better in the next election, contradicting theories of representation? Or do larger margins embolden members of Congress, leading to increases in deviations from constituency preferences? Or are *Marg* and *Dev* really uncorrelated? Table 17.1 shows the OLS regressions with margin and deviation as the left-hand side variables. The regressions include additional variables hypothesized to relate to margin and deviation respectively. Partisan advantage is defined as the 1976 Ford vote in the district for incumbent Republicans and (1 − Ford) for Democrats, $Partadv = Repub * Ford + (1 - Repub) * (1 - Ford)$. A south region variable and the gap between the representative's age and the mean age of district voters are included in the vote margin equation along with partisan advantage and deviations. Given a partisan advantage, seats may be safer in the south, and the larger the age gap between representatives and their constituents the more competitive the seat. The deviation equation includes the member's party and population change and income growth. The expectation is that the more a district's demographic composition changed between 1970 and 1978 the greater the possibility that the member's votes became out of synch with district preferences. Party is included as Republicans exhibited higher unity scores than Democrats (CQ 1983), suggesting the Republican leadership may have been more successful at getting members to deviate from constituency preferences. Representatives who did not run for re-election are omitted.

Table 17.1. Ordinary least squares estimates

Variable	Equation			
	Margin		Deviation	
	Coeff.	St. err.	Coeff.	St. err.
Deviation	0.063	*0.046*		
Margin			0.159	*0.042*
Part. adv.	0.697	*0.077*		
Republican			0.152	*0.014*
South	0.102	*0.016*		
Age gap	−0.120	*0.071*		
ΔPop			0.178	*0.040*
$\text{Inc}_{78}/\text{Inc}_{70}$			0.035	*0.025*
Constant	0.310	*0.044*	−0.194	*0.068*
R^2	0.268		0.327	
N	376		376	

Contrary to expectations, deviations from constituency opinion appear to have a positive relationship with electoral margin, suggesting that the more representatives deviate from constituents' preferences the larger their vote margin. The deviation equation suggests that larger electoral margins are associated with more deviation, as expected. But, the hypothesized endogeneity between the two variables makes both results, and the other coefficients, suspect. What to do?

2. Instrumental Variables: Theory

Econometricians working with the linear model recognized the problem posed by endogenous regressors fairly early, which created an extended and creative search for alternative methods and assumptions. The traditional method for dealing with endogenous explanatory variables, and the one of concern in this chapter, is called instrumental variable, or IV, estimation. The method also goes by the name two-stage least squares as that describes how the estimation was originally done. The statistics are relatively simple in theory, but are quite difficult in practice.

2.1 The Instrumental Variables Estimator

The classic model with endogenous RHS variables is,

$$y = X\beta + Y\gamma + U = W\delta + U, \qquad (3)$$

where y is the outcome variable; X is the matrix of observations on the K truly exogenous variables; Y is the matrix of observations on the M variables suspected to be endogenous for any of the reasons cited above; $W = (X, Y)$; and $\delta' = (\beta', \gamma')$ is the vector of coefficients of interest. The endogeneity problem arises because $\text{plim}\frac{1}{n}(Y'U) \neq 0$.

The theory of IV estimation depends on the existence of a second set of variables, which we denote by Z. Z must contain at least M variables with finite variances and covariances, be correlated with Y, but be uncorrelated with U, i.e.,

$$\text{plim } \frac{1}{n} Z'Z = \Sigma_{zz}, \quad \text{an L} \times \text{L finite, positive definite matrix}$$

$$\text{plim } \frac{1}{n} Z'Y = \Sigma_{zy}, \quad \text{an L} \times \text{M matrix with rank M}$$

$$\text{plim } \frac{1}{n} Z'U = 0.$$

In the simplest case, $L = K$, meaning that there is one "instrument" or Z for each Y. Let $Z^* = (X, Z)$, which is equivalent to letting each of the exogenous variables in X function as its own instrument. In this case, the IV estimator is,

$$d_{IV} = (Z^{*'} W)^{-1}(Z^{*'} y). \tag{4}$$

Greene (2003, 76, 77) shows that this estimator is consistent with an asymptotic normal distribution.

The situation where one has more instruments than original right-hand side endogenous variables, $L > M$, is handled somewhat differently, though the theory is the same. Since plim $(Z'U)/n = 0$ any set of M variables selected from Z will yield consistent estimates for γ. Further, all M linear combinations of the variables in Z also give consistent estimates. It turns out that the linear combination obtained by regressing each variable in Y on all the variables in Z and using the predicted values for Y from these regressions, denoted as $\hat{Y}$, as the right-hand side variables in the regression with y as the left-hand side variable has the lowest asymptotic variance of any linear combination of the Z's (Brundy and Jorgenson 1971; Goldberger 1973).

This common reference to this method as two-stage least squares, or 2SLS, derives from this process. More formally, denote by a the matrix of coefficients obtained from regressing each variable in Y on the variables in Z and by $\hat{W}$ the matrix $\hat{W} = (X, \hat{Y}) = (X, Za) = [X, Z(Z'Z)^{-1}Z'Y] = Z^*(Z^{*'}Z^*)^{-1}Z^{*'}W$:

$$\begin{aligned}
\text{Step One:} \quad \hat{Y} &= Za = Z(Z'Z)^{-1}(Z'Y) \\
\text{Step Two:} \quad d_{2sls} &= (\hat{W}'\hat{W})^{-1}(\hat{W}'Y) \\
&= [(W'Z^*)(Z^{*'}Z^*)^{-1}(Z^{*'}W)]^{-1}[(W'Z^*)(Z^{*'}Z^*)^{-1}(Z^{*'}y)] \\
&= [W'P_{z^*}W]^{-1}W'P_{z^*}y,
\end{aligned}$$

where $P_{z^*} = Z^*(Z^{*'}Z^*)^{-1}Z^{*'}$. Step two is just estimating the equation

$$y = X\beta + \hat{Y}\gamma + \epsilon. \tag{5}$$

This expression for d_{2sls} reduces to equation (4) if there are exactly as many instruments as variables in Y, i.e. $K = L$.

Some intuition into IV estimation would be helpful before discussing alternative versions. Consider a companion equation to equation (3) that relates the included endogenous variables, Y, to the set of instruments, Z,

$$Y = Z\pi + \epsilon, \tag{6}$$

with $E(Z'\epsilon) = 0$. In the full model discussed in section 4 these instruments will include the exogenous variables in the structural equations for Y but excluded from the equation for y, equation (3), as well as the exogenous variables in equation (3).[5] For now, we only need the condition that Z is uncorrelated with ϵ plus the previous conditions for instrumental variables that Z is related to Y, meaning that $\pi \neq 0$, and Z is uncorrelated with U.

Substituting equation (6) relating Y to Z into equation (3) for y gives,

$$y = X\beta + (Z\pi)\gamma + U + \epsilon\gamma. \tag{7}$$

Were π known we could get unbiased estimates for β and γ in equation (7) by regressing y on X and the product $Z\pi$, assuming the various conditions making U and ϵ uncorrelated with X and Z are satisfied. But, alas, π is not known but only estimated in the first stage regression of Y on Z, $p = (Z'Z)^{-1}Z'Y = \pi + (Z'Z)^{-1}Z'\epsilon$. Substituting $\hat{Y} = Zp$ for $Z\pi$ in equation (7) is the second stage of the 2SLS estimator, shown in equation (5). With this substitution the second stage is estimating the equation,

$$y = X\beta + (Zp)\gamma + u + \epsilon\gamma. \tag{8}$$

Equation (8) shows both the attractiveness and limitations of the IV estimator. With an infinite sample the distribution of p collapses at π, or formally plim $p = \pi$, which means that the estimates for β and γ will also collapse about their true values. Hence the consistency result. The IV estimator is biased for finite samples, however, because p is a function of ϵ from the first stage, meaning that Zp and the error term in equation (8) are correlated. These conditions are what makes the IV estimator a biased but consistent estimator for equation (3).

A second important IV estimator is the limited information maximum likelihood estimator, or LIML. Following the notation in Hansen, Hausman, and Newey (2006), the LIML estimator is,

$$d_{liml} = (W' P_{z^*} W - \tilde{\alpha} W' W)^{-1}(W' P_{z^*} y - \tilde{\alpha} W' y), \tag{9}$$

where $\tilde{\alpha}$ is the smallest eigenvalue of the matrix $(\tilde{W}'\tilde{W})^{-1}(\tilde{W}' P_{z^*} \tilde{W})$ with $\tilde{W} = (y, W) = (y, X, Y)$.[6] This is the maximum likelihood estimator for d if the stochastic terms are normally distributed. The LIML estimator also has more desirable small

[5] In this model equation (6) is referred to as the reduced form equations for Y.

[6] The 2SLS and LIML estimator are special cases of a general estimator referred to as a k-class estimator, which is $d_k = (W' P_{z^*} W - \hat{\alpha} W' W)^{-1}(W' P_{z^*} y - \hat{\alpha} W' y)$. For the LIML estimator $\hat{\alpha}$ is defined above and for 2SLS $\hat{\alpha} = 0$.

sample properties, as measured by a mean squared error criteria, than 2SLS under many conditions (Donald and Newey 2001). We will consider both the 2SLS and LIML estimators.

There is a further consideration in choosing instruments. The justification for an IV estimator is its consistency, but it is biased in finite samples, as noted previously. Donald and Newey (2001, 1165) give expressions for the mean squared error of the 2SLS and LIML estimators and Hahn and Hausman (2002) give an expression showing the bias of the 2SLS estimator with one endogenous variable. These errors are proportional to the number of instruments and to the correlation between the error terms in the reduced form, ϵ, and the structural equation, U, and inversely related to the variance of the reduced form error term. There does not seem to be a generalization to equations with more than one endogenous variable, but these proportional relationships should still hold. The implication here is that one should be cautious about selecting both the number and the composition of the instruments. (See section 3 for a discussion of these concerns.)

2.2 An Example: IV

We can illustrate the use of IV estimation with the model relating to representatives' deviations and electoral margins. Jackson and King propose that representatives' party should be treated as an endogenous variable. Doing so here enables us to illustrate IV estimation with an endogenous right-hand side variable that is not simultaneously related to other endogenous variables. Accordingly we treat members' party as endogenous and use a polynomial with Ford's vote to the third, fourth, and fifth power as instruments,[7]

$$Repub = B_{11} + B_{21}Ford^3 + B_{31}Ford^4 + B_{41}Ford^5 + U_1. \tag{10}$$

This equation can be estimated with OLS as the 1976 Ford vote variables are considered to be exogenous and uncorrelated with the stochastic term in the party equation.

The designation of members' party as an endogenous variable implies that the party electoral advantage variable, which includes members' party, is also endogenous but in a non-linear manner. This specification permits this example to demonstrate how IV estimation can be used to accommodate non-linear relationships among the endogenous and exogenous variables. The important works on structural models with non-linear specifications are Goldfeld and Quandt (1972), Jorgenson and Laffont (1974), Kelejian (1971), and Newey (1990), and see Achen (1986) and Wooldridge (2003) for good discussions of how to apply these results. The key element is to expand the non-linear parts of the model to find an expression that is linear in the parameters and that includes enough functions of the exogenous variables to

[7] In this model *Repub* is being treated as a continuous interval variable, which it is not. The model with polynomial Ford variables gives predicted probabilities that range from −0.02 to 0.97 and whose correlation with the predicted probabilities from a probit model is 0.9988, suggesting this equation very closely approximates the better specified but non-linear probit model.

function as instruments for the non-linear term. In the example here the expansion of the party advantage variable is readily done and leads to a clear specification of instruments and estimation strategy for the first stage of the IV procedure. Recall that *Partadv* = *Repub* ∗ *Ford* + (1 − *Repub*) ∗ (1 − *Ford*). Substituting the expression for *Repub* in equation (10) gives,

$$Partadv = (1 - B_{11}) + (2B_{11} - 1)Ford - B_{21}Ford^3 + (2B_{21} - B_{31})Ford^4 + (2B_{31} - B_{41})Ford^5 + 2B_{41}Ford^6 + (2 * Ford - 1) * U_1. \quad (11)$$

The coefficients in equation (11) are linear combinations of those in equation (10). The best way to estimate such a system is as a seemingly unrelated regression model (SUR) that incorporates these constraints and allows for the correlation between the error terms in the two equations.[8] The exogenous variables in the *Marg* and *Dev* equations—*South*, *Age Gap*, *ΔPop*, and *Δ Inc*—constitute the remaining instruments.

The first-stage equations for *Repub* and *Partadv* are estimated with the SUR model, but with two different specifications. The first specification excludes the exogenous variables from the *Marg* and *Dev* equations, as they are not included in the reduced-form equations for *Repub* and *Partadv*. Achen (1986) suggests that for hierarchical models such as this exogenous variables from higher-order equations should be excluded for asymptotic efficiency reasons. Most standard procedures, and programs such as Stata, included all the exogenous variables as instruments in the first-stage equations for all included endogenous variables. Our examples will use both specifications, as the results illustrate questions related to the selection of instruments.

The estimated equations for electoral margin and voting deviations using the IV procedures are shown in Table 17.2. The estimation is done using the 2SLS and LIML estimators. The OLS estimates from Table 17.1 are shown for comparisons. The asymptotic coefficient standard errors (Greene 2003, 400) and the corrected standard errors derived in Bekker (1994) and discussed in Hansen, Hausman, and Newey (2006) are presented.[9]

All the IV estimation results for the coefficients relating *Dev* and *Marg*, shown in boldface, differ substantially from the OLS estimates and are in line with the initial propositions. Larger deviations from constituency opinion are associated with decreased electoral margins, with the coefficients ranging from −0.24 to −0.3. Conversely, members with larger margins, and presumably safer seats, deviate more from constituency opinion, with the coefficient likely between 0.4 and 0.5. In all but one set of estimations the 2SLS and LIML estimates are very similar. The exception is

[8] The stochastic term in equation (11) is heteroskedastic because of the $(2 * Ford - 1)$ term. Proper GLS estimation requires weighting by the reciprocal of this term, which is not defined for $Ford = 0.5$. This heteroskedasticity is ignored in subsequent estimations as it only affects the efficiency of the first-stage estimations. The IV estimator remains consistent.

[9] The calculations are done in Stata following the expressions given in Hansen et al. 2006, 4. The corrected standard errors do not include the terms related to the third and fourth moments that Hansen et al. say, "… are present with some forms of nonnormality."

Table 17.2. Instrumental variable estimates

Variable	Estimation Method						
	OLS	Two-stage least squares			Limited information ML		
	Coeff	Coeff	s_b(cse)	s_b(asy)	Coeff	s_b(cse)	s_b(asy)
Margin equation—limited instruments							
Deviation	**0.063**	**−0.280**	0.133	0.133	**−0.314**	0.147	0.142
Part. Adv.	0.697	0.673	0.098	0.098	0.668	0.101	0.101
South	0.102	0.123	0.019	0.019	0.125	0.020	0.019
Age Gap	−0.120	−0.071	0.075	0.076	−0.071	0.077	0.077
Constant	0.310	0.396	0.070	0.069	0.408	0.075	0.072
R^2	0.268	0.156			0.132		
Margin equation—all instruments							
Deviation	**0.063**	**−0.261**	0.134	0.133	**−0.244**	0.126	0.128
Part. Adv.	0.697	0.679	0.099	0.098	0.670	0.096	0.097
South	0.102	0.122	0.019	0.019	0.121	0.018	0.019
Age Gap	−0.120	−0.131	0.076	0.076	−0.130	0.075	0.075
Constant	0.310	0.398	0.069	0.068	0.392	0.066	0.067
R^2	0.268	0.169			0.108		
Deviation equation—limited instruments							
Margin	**0.159**	**0.496**	0.131	0.128	**0.835**	0.243	0.237
Republican	0.152	0.227	0.045	0.044	0.338	0.083	0.081
ΔPop	0.178	0.184	0.044	0.045	0.165	0.062	0.062
$\text{Inc}_{78}/\text{Inc}_{70}$	0.035	0.046	0.028	0.028	0.048	0.037	0.037
Constant	−0.194	−0.494	0.123	0.120	−0.757	0.208	0.204
R^2	0.327	0.164			−0.428		
Deviation equation—all instruments							
Margin	**0.159**	**0.411**	0.112	0.113	**0.426**	0.119	0.118
Republican	0.152	0.213	0.040	0.040	0.218	0.042	0.041
ΔPop	0.178	0.174	0.044	0.044	0.173	0.044	0.045
$\text{Inc}_{78}/\text{Inc}_{70}$	0.035	0.044	0.027	0.027	0.044	0.027	0.027
Constant	−0.194	−0.411	0.108	0.108	−0.423	0.113	0.111
R^2	0.327	0.232			0.219		

the LIML estimate of the *Dev* equation with the limited set of instruments for the *Republican* variable. We discuss this discrepancy in the next section.

The coefficients on partisan advantage and the exogenous variables are quite consistent across all three estimation methods and match expectations, with one exception. The larger the members' partisan advantage the larger their electoral margin, and southern representatives had safer seats than representatives from outside the south. The OLS estimation and the IV estimations using the full set of instruments suggest that the larger the age gap between representative and constituency the smaller the electoral margin. The IV estimations with only the Ford vote variables as instruments for *Repub* have much smaller and statistically insignificant coefficients for *Age Gap* though the differences are less than the standard errors of the coefficients.

978–0–19–928654–6 17-Box-c17 OUP249-Box (Typeset by SPI, Delhi) 416 of 431 March 14, 2008 20:26

In the deviation equations, Republicans and members from districts with large population changes and income growth deviated more from constituency opinion than did Democrats and those from stable districts.

The R^2 are lower in the IV models but this a not particularly meaningful statistic. The residuals in the estimated equation are calculated using the observed values of the RHS endogenous variables, not their predicted values from the first stage. Subsequent sample statistics—the estimate for σ_u^2, the estimated coefficient standard errors, and the R^2—all depend on these calculations. By construction the R^2 of the IV estimates will be smaller than that of the OLS estimates because the latter are chosen to minimize the sum of squared errors using the observed values of the included endogenous variables. The R^2's in the OLS estimations, however, do not have much meaning as we are starting with the premiss that because of the endogenity these OLS coefficients are biased and inconsistent.

These results suggest that one's underlying assumptions about the data-generating process and the choice of method may have profound substantive consequences. In the presence of endogeneity, which may be likely in observational studies, if the conditions for IV estimation are met it is possible, in theory, to overcome these confounding effects. But, theory and practice are not always the same thing.

3. Instrumental Variables in Practice

The application of instrumental variables poses several serious, possibly daunting, issues. The most critical is that the instruments must be independent of the stochastic term in the equation of interest. The second condition is that the instrument must be correlated with the variable for which it it being used as the instrument. Bartels (1991), in a classic discussion of these problems, shows how they affect the asymptotic mean squared error, AMSE, of the IV estimator. (Also, see Bound, Jaeger, and Baker 1995.) Bartels shows how the AMSE of the IV estimator is related to these two correlations.

$$AMSE\,(b^{IV}) \propto \frac{\rho^2_{ZU|X} + 1/N}{\rho^2_{ZY|X}}, \tag{12}$$

where $\rho^2_{ZU|X}$ is the squared population partial correlation between the instruments and the stochastic term in the equation being estimated and $\rho^2_{ZY|X}$ is the squared population partial correlation between the instruments and the endogenous variables for which they serve as instruments. Bartels makes it clear that it is the partial correlations holding the included exogenous variables, X, constant that are critical. This requirement is also evident in equation (5). As with any OLS regression the greater the variance in $\hat{Y}$ and the lower the correlations of $\hat{Y}$ with the variables in X the more reliable the estimates of γ in the second-stage estimation of equation (5).

978–0–19–928654–6 17-Box-c17 OUP249-Box (Typeset by SPI, Delhi) 417 of 431 March 14, 2008 20:26

Bartels argues that these conditions are problematic in practice, hence the name "quasi-instruments," so users must consider what happens in practice, how to diagnose potential problems, and how best to manage the tradeoffs among second-best estimators. The discussion and assessment of weak instruments is also the subject of important work in econometrics over the past decade. These conditions and how one might assess their fit, or lack of, to specific data are discussed and illustrated in this section.

3.1 Correlations of Instruments and Endogenous Variables

We begin the discussion of the practice of IV estimation with the denominator in equation (12) as it is easier to observe and to estimate with available data. The term $\rho^2_{ZY|X}$, referring to the squared population partial correlation, can be estimated by the partial correlation between Z and Y in the sample data. The most direct way to make this assessment is to regress Y on X and Z and then conduct the F-test for the null hypothesis that the coefficients on Z equal zero. There are two weaknesses to this test, however. The first is that the null hypothesis being tested by the F-statistic is not the null hypothesis of interest and one should be most concerned about a type II not a type I error. Rejecting the null hypothesis that the coefficients are zero is not identical to saying they are not zero with some probability of error. Baum, Schaffer, and Stillman (2003) suggest that with one included endogenous variable an F-statistic greater than ten is an appropriate rule of thumb of accepting the alternative hypothesis that the coefficients are not zero.

A second difficulty arises when there is more than one RHS endogenous variable, as in this example. The previous test is designed to ensure that the included endogenous variables are strongly correlated with the excluded exogenous variables, Z, holding constant the included exogenous variables, X. But, what if the same variables in Z account for the independent variation in all the included Y's? This effectively means that the coefficients on these included endogenous variables are not really identified despite appearing to meet the technical criteria.

Shea (1997) presents a test for the strength of the partial relationship between each included endogenous variable and the excluded instruments controlling for the included instruments and the other included endogenous variables. Shea presents a four-step procedure for calculating these statistics, usually referred to as Shea's partial R^2 and denoted as R^2_p. Godfrey (1999) presents a simplified calculation based on the ratios of the coefficients' variances and the R^2 statistics from the OLS and 2SLS estimations respectively. This expression is:

$$R^2_p = \left(\frac{\sigma_b^{ols}}{\sigma_b^{2sls}}\right)^2 \left(\frac{\sigma_u^{2sls}}{\sigma_u^{ols}}\right)^2 = \left(\frac{\sigma_b^{ols}}{\sigma_b^{2sls}}\right)^2 \left(\frac{1 - R^2_{2sls}}{1 - R^2_{ols}}\right),$$

where σ_b is the coefficient standard error, σ_u is the standard error of the estimate, and R^2 is the unadjusted centered R-squared from the respective estimated equations. Unfortunately neither Shea nor Godfrey report a test statistic for this partial R-squared

Table 17.3. Tests for instrument relevance

Endogenous Variable		Deviation	Part advant.	Margin	Republican
1^{st} Stage	R^2	0.165	0.721	0.236	0.255
	$F(K_1, K_2)^a$	8.02	135.59	12.59	17.97
Partial	R^2	0.141	0.713	0.233	0.226
	$F(K_1, K_2)^b$	8.55	182.85	15.87	21.43
Shea's R^2		0.135	0.698	0.156	0.149

[a] $K_1 = 9$ *and* $K_2 = 366$ for *Dev* and *Marg* and $K_1 = 7$ *and* $K_2 = 368$ for *Repub* and *Part Adv.*

[b] $K_1 = 7$ *and* $K_2 = 366$ for *Dev* and *Marg* and $K_1 = 5$ *and* $K_2 = 368$ for *Repub* and *Part Adv.*

so we cannot test the null hypothesis that it is zero, and thus that the instruments for that particular coefficient are irrelevant. In performing these computations one must assure that all estimated standard errors are calculated with the same degrees of freedom.[10]

Table 17.3 reports the R^2 and associated F-statistics for the first-stage regressions, the partial R^2 and associated F-statistics and the Shea's partial R^2 for the estimations just reported, using the full set of instruments.[11] None of these statistics suggests a problem with how well the necessary instruments correlate with the included endogenous variables. The F-statistic for the partial R^2 for *Dev* is only 8.55, but the P-value for this F with the appropriate degrees of freedom is approximately $1.X10^{-09}$, which suggests a very low likelihood of a type II error if the null hypothesis of no relationship is rejected. The other F-statistics are even larger. The Shea's partial R^2 statistics range from 0.135 to 0.156, with the value for Partisan Advantage equal to 0.70. On the basis of these statistics we conclude that these instruments adequately meet the standards for relevance, as discussed by Bartels, Bound, and subsequent authors.

3.2 Independence of Instruments and Stochastic Terms

The second critical requirement for consistent IV estimators is independence between the instruments and the stochastic term in the equation being estimated. This is the numerator in Bartels's assessment of the asyptotic mean squared error of the IV estimator, equation (12). This condition is difficult to test in practice as it requires information on the unobservable stochastic term in the equation being estimated.

[10] Some 2SLS estimations do not use a degrees of freedom correction in calculating standard errors under the assumption that only the asymptotic properties are of interest. In Stata, for example, the "small" option adjusts for the degrees of freedom in the equation being estimated.

[11] The equations for *Repub* and *Part Adv* have fewer degrees of freedom because following equations (10) and (11) there are only nine coefficients being estimated. The equations for *Dev* and *Marg*, because *Ford* and *Ford*[6] are included and none of the coefficients are constrained estimate eleven coefficients.

978–0–19–928654–6 17-Box-c17 OUP249-Box (Typeset by SPI, Delhi) 419 of 431 March 14, 2008 20:26

There are several approaches to approximating this test discussed in this section. All, however, require that the equation being estimated is overidentified, meaning that there are more instruments than included endogenous variables, $L > K$. The first statistic is attributed to Sargan (1958) and is reported as the Sargan statistic. This statistic is the proportion of the sum of squared residuals from the estimated equation that can be "explained" by the full set of instruments,

$$\text{Sargan statistic } = \frac{\hat{u}' Z^*(Z^{*'} Z^*)^{-1} Z^{*'} \hat{u}}{\hat{u}'\hat{u}/n}, \tag{13}$$

where $Z^* = (X, Z)$ and $\hat{u}$ denotes the residuals from the estimated equation using the observed values for the included endogenous variables. Sargan shows that asymptotically this ratio is distributed as a Chi-squared statistic with $(L - K)$ degrees of freedom. If we knew the true values of the stochastic term and if the instruments are truly independent of the stochastic term this statistic should be zero, meaning that the instruments do not explain any of the residual variance. The Sargan statistic approximates this ratio by using the residuals calculated from the estimated equation. There is also a version of the Sargan statistic that multiplies the ratio by $(n - L)$ rather than n, which is also distributed as a Chi-squared statistic with $(L - K)$ degrees of freedom.

Basmann (1960) proposes a second statistic that compares the same fitted values of the residuals in the Sargan statistic to the unexplained variance in the residuals rather than the total variance of the residuals,

$$\text{Basmann's statistic} = \frac{\hat{u}' Z^*(Z^{*'} Z^*)^{-1} Z^{*'}\hat{u}/(L - K)}{[\hat{u}'\hat{u} - \hat{u}' Z^*(Z^{*'} Z^*)^{-1} Z^{*'}\hat{u}]/(n - L)}. \tag{14}$$

Basmann shows that asymptotically this statistic has an F-distribution with $(L - K)$ and $(n - L)$ degrees of freedom. Basmann and Sargan present other variations on these statistics but the core comparison is the relationship between the residuals in the estimated structural equation and the instrumental variables.

One weakness of these tests is the same as with the tests of instrument relevance. The null hypothesis being tested is that the instruments and stochastic term are uncorrelated, plim $Z'u/n = 0$. If this null is rejected then it is clear that the variables are not adequate instruments. But, not rejecting the null is not equivalent to accepting it, which is the critical issue. The smaller the Sargan and/or Basmann statistics the higher the likelihood of getting that value by chance if the null hypothesis is true, and thus the lower the probability of a type II error if the null is accepted. But, at the end of the analysis the best that can be said is that these statistics are useful for rejecting the null hypothesis and thus putting the instruments in doubt but are not strong enough for establishing that these instruments are valid.

Table 17.4 shows the Sargan and Basmann statistics for the estimated equations, including the estimations with the limited set of instruments for the *Repub* and *Part Adv* variables. Recall from Table 17.2 that the LIML estimates for the *Dev* equation with the limited set of instruments produced questionable results. The 2SLS and LIML estimations using the full set of instruments in the first-stage estimates for *Repub* and

Table 17.4. Tests for instrument independence

Test (DOF)	Marg		Dev	
	2SLS	LIML	2SLS	LIML
	Full instruments			
Sargan $\chi^2(5)$	1.143	1.213	2.114	2.098
P-Value	0.950	0.944	0.833	0.835
Basmann F(5,366)	0.187	0.198	0.346	0.343
P-Value	0.967	0.963	0.885	0.887
	Limited instruments			
Sargan $\chi^2(5)$	1.844	1.872	2.769	9.963
P-Value	0.870	0.867	0.736	0.076
Basmann F(5,366)	0.301	0.306	0.454	1.665
P-Value	0.912	0.909	0.810	0.142

Part Adv have very low Sargan and Basmann statistics with very high probabilities of occurrence if the null hypothesis of independence is true. The estimates of the *Marg* equation with only the *Ford* variables as instruments for *Part Adv* have similarly low Sargan and Basmann statistics, again suggesting instrument independence.

The questionable results are the limited instrument LIML estimation of the *Dev* equation. We noted earlier that the coefficient on *Marg* was inconsistent with the other estimations and that the estimates fit the observed data very badly. The Sargan and Basmann statistics indicate there is a problem with the assumption that the instruments are uncorrelated with the stochastic term. It is hard to explain these results, as all of the other results with similar specifications and even the 2SLS estimates with the same specifications give more consistent results and have more acceptable summary statistics. There are two things one may want to conclude from these results. Results are sensitive to the choice of instruments, as is also seen in the estimates for the coefficients on *Age gap* in Table 17.2, and estimation method in unexpected ways. This makes the exploration of the robustness of the results to these choices important. It is also the case that some of the diagnostic statistics, such as the assessments of instrument relevance and independence, may reveal problems and should be reported in all studies.

4. Full Information Estimation

The previous estimators are limited information estimators because the estimation proceeds one equation at a time. A second estimation strategy, referred to as full information estimation, estimates the entire structural model. This section talks briefly about this strategy and applies it to the running example. The reference to the entire structure means there is an equation of the form of equation (3) for each endogenous variable in the model,

$$\begin{aligned} y_1 &= X_1\beta_1 + Y_1\gamma_1 + U_1 = W_1\delta_1 + U_1 \\ &\vdots \qquad\qquad \vdots \\ y_m &= X_m\beta_m + Y_m\gamma_m + U_m = W_m\delta_m + U_m \\ &\vdots \qquad\qquad \vdots \\ y_M &= X_M\beta_M + Y_M\gamma_M + U_M = W_M\delta_M + U_M. \end{aligned} \tag{15}$$

Assume that $E(X'U) = 0$ and that the U_i's are iid, so that,

$$E(U_{is}U'_{jt}) = \Sigma_u, \text{ for } i = j \ \&\ s = t \tag{16}$$

$$= 0, \text{ otherwise.} \tag{17}$$

We also assume that the specification of the exogenous and endogenous variables excluded from each equation plus any other a priori constraints are sufficient to meet the identification requirements. (Identification of each equation is required for estimation and is a complex topic that is not addressed in this chapter. The classic work here is Fisher 1966, but see any text, such as Greene 2003, 385–95 or Wooldridge 2003, 211–30.) We discuss two different full information estimation strategies and both parallel those examined above. The first is an extension of the 2SLS estimator to the full model, referred to as Three-Stage Least Squares, or 3SLS. The second is a maximum likelihood, or FIML, estimator.

4.1 Three-Stage Least Squares

Three-Stage Least Squares (Zellner and Theil 1962) stacks each of the M equations in equation (16) into one large $TM \times 1$ system, as shown in equation (18),

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \\ \vdots \\ y_M \end{pmatrix} = \begin{pmatrix} W_1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & W_m & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & W_M \end{pmatrix} \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_m \\ \vdots \\ \delta_M \end{pmatrix} + \begin{pmatrix} U_1 \\ \vdots \\ U_m \\ \vdots \\ U_M \end{pmatrix}. \tag{18}$$

This expression can be summarized as $Y = W\delta + U$.

Each equation in this system could be estimated by 2SLS using the same IV procedures and instruments discussed above. In this estimation the instruments for each equation are all the exogenous variables in the system, X, so that $\hat{W}_m = (X_m, \hat{Y}_m) = [X_m, X(X'X)^{-1}X'Y_m]$. The 2SLS estimator for each equation is,

$\hat{d}_{2sls} = (\hat{W}'\hat{W})^{-1}\hat{W}'Y$. Unless Σ_u is diagonal, meaning all the stochastic terms are independent of each other, this method loses asymptotic efficiency by not considering these stochastic term covariances. (See Judge et al. 1988, 646–51.) The standard 3SLS estimator uses the residuals from the 2SLS estimation of each equation to form an estimate for the elements in Σ_u, which is the second stage in the 3SLS procedure. The inverse of this estimated variance-covariance matrix is used in a feasible GLS estimation of the full system. Let $\hat{\Sigma}_u^{-1}$ denote this inverted estimated variance-covariance matrix, which following the conditions in equation (17) has the following structure,

$$\hat{\Sigma}_u^{-1} = \begin{pmatrix} \hat{\sigma}^{11}I & \cdots & \hat{\sigma}^{1M}I \\ \vdots & \vdots & \vdots \\ \hat{\sigma}^{1M}I & \cdots & \hat{\sigma}^{MM}I, \end{pmatrix} \tag{19}$$

where the I are $T \times T$ identity matrices. The 3SLS estimator is,

$$\hat{d}_{3sls} = (\hat{W}'\hat{\Sigma}_u^{-1}\hat{W})^{-1}(\hat{W}'\hat{\Sigma}_u^{-1}Y). \tag{20}$$

This is a specific application of Zellner's (1962) seemingly unrelated regression. Often this process is iterated using the residuals from successive 3SLS estimations to get $\hat{\Sigma}_u$ until convergence is achieved, which is usually fairly rapidly.

The justification for 3SLS depends heavily on its asymptotic properties, even more so than 2SLS. In addition to the conditions required for consistency and the factors that make 2SLS biased in finite samples, with 3SLS the user is relying on the asymptotic condition that the variances and covariances of the stochastic terms in the equations being esimated are given by Σ_u and that the variances and covariances of the residuals from the estimated equations are an adequate approximation to this matrix. But, this ignores the fact that the error term in the equation being estimated consists of U_m *and* a term involving the deviations of the estimated reduced form coefficients from their true values. Asymptotically this term goes to zero, but in practice for finite samples it is not zero and thus will contribute to the variance of the error term in the equation being estimated. Whether including the FGLS approximation to Σ_u actually improves efficiency depends on the actual sample and might be open to conjecture.

4.2 Full Information Maximum Likelihood

The full information analog to the LIML estimator uses the assumption that the stochastic terms are normally distributed, i.e. U_i is $N(0, \Sigma_u)$. To develop the FIML estimator equation (16) is rearranged to give,

$$(y_1, \ldots, y_M) = X(\beta_1, \ldots, \beta_M) + Y(\gamma_1, \ldots, \gamma_M) + (U_1, \ldots, U_M), \tag{21}$$

where Y is a $(T \times M)$ matrix of observations on all the endogenous variables and X is a $(T \times K)$ matrix of observations on all the exogenous variables in the model. β_m is a $(K \times 1)$ vector of coefficients relating y_m to the exogenous variables, which implies that any exogenous variable excluded from this equation has a value of zero in this vector. γ_m is an $(M \times 1)$ vector of coefficients relating y_m to the other endogenous variables. Any endogenous variable omitted from the RHS of this equation has a zero coefficient in this vector, including the m^{th} entry, which corresponds to y_m. Equation (22) summarizes this expression,

$$Y\Gamma + XB + U = 0. \tag{22}$$

The matrix Γ has values of −1 on its main diagonal corresponding to the implicit coefficient on y_m in equation (21).

The FIML estimator rearranges equation (22) to isolate the stochastic component,

$$V = -U\Gamma^{-1} = Y + XB\Gamma^{-1}. \tag{23}$$

If the U_i's are normally distributed then V is $N(0, \Gamma^{-1'}\Sigma_u\Gamma^{-1})$. From this, the log likelihood function for the observed data is,

$$\begin{aligned}
\log L &= -\frac{T}{2}[M\log(2\pi) + \log|\Sigma_v| + tr(\frac{1}{T}\Sigma_v^{-1}V'V)] \\
&= -\frac{T}{2}\left\{\kappa + \log|\Gamma^{-1'}\Sigma_u\Gamma^{-1}| + tr\left[\frac{1}{T}(\Gamma'\Sigma_u^{-1}\Gamma)(Y + XB\Gamma^{-1})'(Y + XB\Gamma^{-1})\right]\right\} \\
&= -\frac{T}{2}[\kappa + \log|\Sigma_u| - 2\log|\Gamma| + tr(\Sigma_u^{-1}S)],
\end{aligned} \tag{24}$$

where $s_{ij} = \frac{1}{T}(Y\Gamma_i + XB_i)'(Y\Gamma_j + XB_j)$. The inclusion of $\log|\Gamma|$ in the likelihood function imposes one more constraint on the model in order to compute the FIML estimator. As the value of this determinant approaches zero the likelihood function approaches minus infinity, effectively ruling out that particular set of parameter values.

The FIML estimator provides a test of model specification not present in the other estimators if the model is overidentified. The estimated model's fit is based on how well the estimated model predicts the variance-covariance matrix of observed variables. In an exactly identified model the number of parameters being estimated, including the variances and covariances of the stochastic terms, exactly equals the number of entries in the observed variance-covariance matrix. In this case the estimated model will fit the observed matrix perfectly. As additional restrictions are added to the model by specifying that specific variables are excluded from certain equations, i.e. that certain elements of B and Γ are zero, the fit decreases as does the log likelihood function. The test of the overidentifying restrictions then is a function of the difference in the log likelihood functions of the just identified model and the estimated model. Asymptotically minus twice this difference is distributed as a χ^2 variable with degrees of freedom equal to the number of overidentifying restrictions (Jöreskog 1973). A poor fit and accompanying large value for the χ^2 statistic leads to rejection of the null hypothesis that the overidentifying restrictions are satisfied. The functional difference with the Sargan statistic is that this statistic is testing the fit of the whole model, not just the restrictions in an individual equation.

978–0–19–928654–6 17-Box-c17 OUP249-Box (Typeset by SPI, Delhi) 424 of 431 March 14, 2008 20:26

Estimation proceeds by maximizing the expression in equation (24) subject to the constraints and restrictions placed on the model to obtain identification and to conform to any other a priori information about the model. This produces a set of non-linear simultaneous equations that requires a numerical analysis for solution that is challenging to solve in practice and the discussion of which is beyond the scope of this chapter. The possible presence of sets of parameter values that lead to the condition that $\log|\Gamma| = 0$ further complicates the computational task. These singularities create distinct regions that must be searched separately in order to find the global maximum. The complexity of the computational tasks has led most analysts to use 3SLS more often than FIML. The 3SLS computational advantage actually comes at little cost asymptotically if the stochastic terms are normally distributed. Greene (2003, 409) shows that FIML is also an IV estimator. He then goes on to state that with normally distributed stochastic terms 3SLS and FIML have identical asymptotic distributions. He also says that small sample properties are "ambiguous" and may differ between the two methods.

4.3 Example: Full Information Estimation

The estimates reported in Table 17.2 were redone using both the 3SLS and FIML procedures.[12] The results are shown in Table 17.5, which include the estimates for the *Repub* equation. (Coefficients for the *Part Adv* equation are not reported as they are just linear functions of the coefficients in the *Repub* equation. See equation (11)). The table also shows the Sargan statistic calculated for the Margin and Deviation equations. The results from the two methods are remarkably similar, differing by at most 0.012 and in most cases by considerably less and the estimated asymptotic standard errors are nearly identical as well. A likely explanation for the similarities in results are that the estimated covariances among the stochastic term are quite small, on the order of 0.13 to 0.23. The Sargan statistics are quite low and have a high probability of occurring by chance if the instruments and stochastic terms are independent. The χ^2 test of the whole model with the FIML estimation is 22.61 with 26 degrees of freedom, which has a P-value of 0.665, again suggesting it is probably safe to accept the null hypothesis that the identifying restrictions are consistent with the data and justify the IV variables.

Full information estimators are more susceptible to specification problems, though this sensitivity was not evidenced in this example. Because 2SLS and LIML estimate each equation separately a misspecification in one equation in the structural system does not influence the estimation of other equations, subject to the requirements that the instruments for the equation being estimated are both relevant and uncorrelated with the stochastic terms in the estimated equation. With the full information methods a misspecification in one equation can affect the estimates of other equations because the 3SLS and FIML procedures try to fit the entire structure. This means that the

[12] The FIML estimation was done using the package LISREL for Windows.

Table 17.5. Full information results

	3SLS		FIML	
	Coeff	s_b(asy)	Coeff	s_b(asy)
Repubican				
Ford^3/100	−0.136	0.031	−0.135	0.031
Ford^4/100	0.543	0.102	0.542	0.102
Ford^5/100	−0.443	0.084	−0.443	0.084
Constant	0.023	0.022	0.023	0.022
R^2	0.250		0.250	
Margin				
Dev	−0.296	0.132	−0.308	0.139
Part Adv	0.685	0.098	0.687	0.097
South	0.119	0.019	0.119	0.018
Age Gap	−0.144	0.075	−0.145	0.075
Constant	0.406	0.067	0.409	0.069
R^2	0.147		0.139	
Sargan $\chi^2(5)$	1.208		1.241	
P-Value	0.944		0.941	
Deviation				
Marg	0.451	0.113	0.464	0.123
Repub	0.218	0.039	0.223	0.042
ΔPop	0.171	0.043	0.170	0.043
ΔInc	0.037	0.026	0.037	0.026
Constant	−0.423	0.104	−0.433	0.113
R^2	0.205		0.191	
Sargan $\chi^2(5)$	1.788		1.844	
P-Value	0.878		0.870	

conditions of relevance and independent required of the instrumental variables must be satisfied in every equation (Wooldrige 2003, 199). For these and computational reasons, one of the k-class of limited information estimators seem to be preferred in practice and are more frequently encountered in the literature.

5. Comparisons of Estimators

This section briefly compares and summarizes the results obtained with the different estimators. The discussion focuses on the relationships between *Dev* and *Marg* estimated in each of their respective equations. Table 17.6 shows the five different estimated relationships between these variables, where first *Marg* and *Dev* are the left-hand side variables. There is little substantive difference among the IV estimations, though the full information estimates are somewhat larger than the limited

Table 17.6. Coefficients with different estimation methods

LHS Var.	RHS Var.	Estimation Method				
		OLS	2SLS	LIML	3SLS	FIML
Marg	*Dev*	0.063	−0.261	−0.244	−0.296	−0.308
Dev	*Marg*	0.159	0.411	0.426	0.451	0.464

information estimates. There is even less statistical difference as even the largest difference between the estimates is less than half the magnitude of the standard errors.

The substantial difference is between the OLS and IV estimators, as one would expect if *Marg*, *Dev*, and *Repub* are not exogenous variables, and if the instruments are appropriate. The IV estimates for the effect of voting deviations from constituency preferences on electoral margin range from −0.24 to −0.30, compared to 0.06 with OLS estimation. Conversely, the estimated effect of electoral margin on deviations ranges from 0.41 to 0.46 with the IV estimators, compared to 0.16 for the OLS model. The IV results are more consistent with expectations, particularly in the equations testing the proposition that deviations from constituency preferences are punished in the next election. The final topic is a test of whether RHS variables can be treated as exogenous, and thus which of these estimators is preferred.

If the RHS variables are fully exogenous then the OLS estimator is preferred because it is unbiased, consistent, and efficient and it is the MLE estimator. This gives it a great advantage over less efficient estimators, such as IV. If the RHS variables are not exogenous, then the IV estimators are biased and less efficient but consistent. Specification tests proposed by Durbin (1954), Wu (1973), and Hausman (1978), often referred to as DWH tests, provide information for this choice.[13] Excellent discussions of the versions of these tests are in Baum, Schaffer, and Stillman (2003), Davidson and MacKinnon (1993, 237–42), Greene (2003, 80–3), and Nakamura and Nakamura (1981). The basis for the test is a comparison of the differences in the OLS and IV coefficient values and variances. Denote by b_{iv} the consistent but inefficient IV estimates and by S_{iv} the matrix of estimated asymptotic variances and covariances of these estimates. The null hypothesis is no endogeneity, in which case the OLS estimated coefficients, denoted as b_{ols}, are also consistent but more efficient than the IV estimator. Denote the estimated variances and covariances of these efficient estimates as S_{ols}. The test statistic is,

$$\begin{aligned} H &= (b_{iv} - b_{ols})'[S_{iv} - S_{ols}]^{-1}(b_{iv} - b_{ols}) \\ &= n(b_{iv} - b_{ols})'\left[\frac{(\hat{W}'\hat{W})^{-1}}{s_u^2} - \frac{(W'W)^{-1}}{s_u^2}\right]^{-1}(b_{iv} - b_{ols}), \end{aligned}$$

13 Hausman's paper and statistic is developed for the general case where one wants to compare a consistent with an efficient but possibly inconsistent estimator, which includes a number of important contexts and is not limited to the case of possible endogeneity and IV estimation.

Table 17.7. Endogeneity tests

	Equation			
	Margin		Deviation	
	Statistic	P-Value	Statistic	P-Value
$\chi^2(2)$				
2SLS	8.51	0.014	6.88	0.032
3SLS	11.39	0.003	9.47	0.009
F(2,369)				
2SLS	4.85	0.008	3.78	0.024

where following the notation in equation (16) $W = (X_m, Y_m)$ and $\hat{W} = (X_m, \hat{Y}_m)$ and s_u^2 is the estimated variance of the stochastic term. This statistic has a χ-squared distribution with M_m degrees of freedom where M_m denotes the number of LHS endogenous variables.

The variations in the calculations of the H test statistic arise from one's choice of the estimate for s_u^2. One choice is the standard error of the estimate from the OLS estimation, which was Durbin's proposal. A second choice is the estimate from the IV estimation and a third option is to use each of the estimates in their respective estimates for S.[14] Asymptotically these three ratios should be equal, but in finite samples they will not be. Baum et al., among others, suggest using the OLS estimate as it is the most efficient estimate.

Both Davidson and MacKinnon (1993) and Greene (2003) describe a computationally easier way to test for endogeneity in the IV estimation context. Recall from equation (16) the equation being estimated, $y_m = X_m\beta_m + Y_m\gamma_m + U_m$. Now estimate the equation,

$$y_m = X_m\beta_m + Y_m\gamma_m + \hat{Y}_m\alpha + U_m. \tag{25}$$

Wu (1973) and Davidson and MacKinnon (1993) show that the conventional F-test for the null hypothesis $H_0 : \alpha = 0$ is asymptotically equivalent to the DWH test above. This F-statistic has M_m and $(n - K_m - 2 * M_m)$ degrees of freedom where K_m and M_m are the number of included exogenous and endogenous variables. Again, with finite samples this F-statistic will not be identical in its P-value to the versions of the H χ^2 statistics.

Table 17.7 shows the DWH χ^2 and F statistics for the 2SLS equations reported in Table 17.2 and the χ^2 statistic for the 3SLS equations reported in Table 17.5. The χ^2 statistics are computed using the OLS estimates for s_u^2. All results quite clearly reject the null hypothesis that OLS produces asymptotically equivalent but more

[14] If the third option is chosen the difference in the matrices of asymptotic variances and covariances may not be positive definite meaning the generalized inverse must be used in the calculations.

efficient estimates than the different IV estimators at conventional confidence levels. This means that the IV estimators are probably "better" estimates of the joint relationships between *Dev* and *Marg*, being biased and inefficient, but consistent. The DWH statistics for the 3SLS estimation are larger than for the 2SLS, which should be expected as the difference between these estimates for the relationships between *Dev* and *Marg* and the OLS estimates are larger than the differences with the 2SLS estimator. In this particular example the F-tests had smaller P-values than the DWH test. In this example both tests reject the null, but it is possible the DWH test is more conservative on this point than the F-test. Staiger and Stock (1997, 568) argue that the DWH test using the OLS estimate for s_u^2 has more power than the other forms of the test and is recommended if one suspects weak instruments. There seems to be an absence of work comparing the χ^2 and F-test versions of the test, possibly because of Hausman's demonstration that the former can be applied in a number of important estimation contexts.

6. Concluding Remarks

The continuing need to use observational data routinely raises the problem of endogeneity of RHS variables and thus of covariance between these variables and the equation's stochastic term, in violation of a key condition justifying the OLS or GLS estimators. This has and continues to be a daunting problem for empirical social science researchers. At one point it was hoped that instrumental variables and the associated estimators discussed here constituted a possible and adequate remedy. (See Goldberger and Duncan 1973 as one example.) The difficulty of matching real data to the conditions justifying IV estimation and the estimators' potential small sample biases despite the asymptotic properties has reduced some of this optimism and enthusiasm. These concerns have been augmented by various studies showing the weakness of the IV estimators even with very large data sets (Bound, Jaeger, and Baker 1995).

Well-founded reservations about the robustness and reliability of the IV estimators are leading some researchers to undertake very creative studies that follow the experimental paradigm. (On experiments see Kinder and Palfrey 1993; and Druckman et al. 2006.) There is also considerable interest in building on the work of statisticians such as Rubin and Rosenbaum to develop techniques to mimic the experimental conditions of treatment and control groups by matching subjects in observational studies. (See Rosenbaum 2001 for an overview and Arceneaux, Gerber, and Green 2006 for a critique of matching methods.) The hope is that these efforts will lead to better tests of causal arguments by avoiding or at least reducing possible endogeneity biases. And one can only applaud and encourage work that expands the methodologists, ability to generate and analyze data in wider contexts.

It is not, however, a situation of replacing one method with another. There will continue to be important substantive questions engaging political science researchers that are not amenable to some of these other methods. Consider the examples given at the beginning of this chapter. Propositions about inter-state rivalry, or the effects of institutions on economic performance, or the consequences of campaign spending are unlikely to be manipulable in an experimental setting. There is also a need to examine how well results obtained in well-controlled laboratory settings apply in non-laboratory political settings. Thus, in many instances our empirical evidence will continue to be the classic observational study, replete with endogeneity problems. Rather than foregoing analyses of these data and questions because of the reservations about a particular method it is better to use and to improve those methods in the context of these studies and data. This is surely what stimulates the ongoing work in econometrics to understand better the properties of IV estimators, to develop and use better the diagnostics that help inform judgments about the statistical results, and to encourage the reporting of these newer statistics.

References

Acemoglu, D. 2005. Constitutions, politics, and economics: a review essay on Persson and Tabellini's *The Economic Effects of Constitutions*. *Journal of Economic Literature*, 43 (4): 1025–48.

Achen, C. H. 1978. Measuring representation. *American Journal of Political Science*, 22 (3): 475–510.

—— 1986. *The Statistical Analysis of Quasi-Experiments*. Berkeley, and Los Angeles: University of California Press.

—— 1988. Proxy variables and incorrect signs on regression coefficients. *Political Methodology*, 11: 299–316.

Arceneaux, K., Gerber, A. S., and Green, D. P. 2006. Comparing experimental and matching methods using a large-scale vote mobilization experiment. *Political Analysis*, 14 (1): 37–62.

Barone, M., Ujifusa, G., and Matthews, D. 1979. *Almanac of American Politics.* New York: E. P. Dutton.

Bartels, L. M. 1991. Instrumental and quasi-instrumental variables. *American Journal of Political Science*, 35 (3): 777–800.

—— and Brady, H. E. 1993. The state of quantitative political methodology. In *Political Science: The State of the Discipline II*, ed. A. Finifter. Washington, DC: American Political Science Association.

Basmann, R. 1960. On finite sampling distributions of generalized classical linear identifiability test statistics. *Journal of the American Statistical Association*, 55 (292): 650–9.

Baum, C. F., Schaffer, M. E., and Stillman, S. 2003. Instrumental variables and GMM: estimation and testing. *Stata Journal*, 3 (1): 1–31.

Bekker, P. 1994. Alternative approximations to the distribution of instrumental variable estimators. *Econometrica*, 62 (3): 657–81.

Berinsky, A. 1999. The two faces of public opinion. *American Journal of Political Science*, 43 (4): 1209–30.

Bound, J., Jaeger, D. J., and Baker, R. M. 1995. Problems with instrumental variables estimation when the correlations between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90: 443–50.

Brady, H. 1988. The perils of survey research: interpersonally incomparable responses. *Political Methodology*, 11: 269–91.

Brundy, J., and Jorgenson, D. 1971. Consistent and efficient estimation of systems of simultaneous equations by means of instrumental variables. *Review of Economics and Statistics*, 53: 207–24.

Congressional Quarterly. 1983. *CQ Weekly Report*. 41 (52): 2789.

Davidson, R., and MacKinnon, J. G. 1993. *Estimation and Inference in Econometrics*. New York: Oxford University Press.

Dixon, W. J. 1986. Reciprocity in unites states–soviet relations: multiple symmetry of issue linkage. *American Journal of Political Science*, 30 (2): 421–45.

Donald, S. G., and Newey, W. K. 2001. Choosing the number of instruments. *Econometrica*, 69 (5): 1161–91.

Druckman, J. N., Green, D. P., Kuklinski, J. H., and Lupia, A. 2006. The growth and development of experimental research in political science. *American Political Science Review.* 100 (4): 627–37.

Durbin, J. 1954. Errors in variables. *Review of the International Statistical Institute*, 22: 23–32.

Fisher, F. M. 1966. *The Identification Problem in Econometrics.* New York: McGraw-Hill.

Freedman, D., Pisani, R., and Purves, R. 1998. *Statistics*. 3rd edn. New York: W. W. Norton.

Godfrey, L. G. 1999. Instrument relevance in multivariate linear models. *Review of Economics and Statistics*, 81 (3): 550–52.

Goldberger, A. S. 1973. Efficient estimation in overidentified models. Pp. 131–52 in *Structural Equation Models in the Social Sciences*, ed. A. S. Goldberger and O. D. Duncan. New York: Seminar Press.

—— and Duncan, O. D. 1973. *Structural Equation Models in the Social Sciences*. New York: Seminar Press.

Goldfeld, S. M., and Quandt, R. E. 1972. *Nonlinear Methods in Econometrics.* Amsterdam: North Holland.

Greene, W. H. 2003. *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice Hall.

Hahn, J., and Hausman, J. 2002. New specification test for the validity of instrumental variables. *Econometrica*, 70 (1): 163–89.

Hansen, C., Hausman, J., and Newey, W. K. 2006. Estimation with many instrumental variables. Xerox, Dept. of Economics, MIT.

Hausman, J. A. 1978. Specification tests in econometrics. *Econometrica*, 46 (6): 1251–71.

Jackson, J. E. 1993. Attitudes, no opinions, and guesses. *Political Analysis*, 5: 39–60.

—— 1996. Political methodology: an overview. In *New Handbook of Political Science*, ed. R. Goodin and H.-D. Klingemann. Oxford: Oxford University Press.

—— and King, D. C. 1989. Public goods, private interests, and representation. *American Political Science Review*, 83 (4): 1143–64.

Jacobson, G. C. 1978. The effects of campaign spending in congressional elections. *American Political Science Review*, 72 (2): 469–91.

Jöreskog, K. 1973. A general method for estimating a linear structural equation system. In *Structural Equation Models in the Social Sciences*, ed. A. S. Goldberger and O. D. Duncan. New York: Seminar Press.

Jorgenson, D. W., and Laffont, J. 1974. Efficient estimation of nonlinear simultaneous equations with additive disturbances. *Annals of Economic and Social Measurement*, 3: 615–40.

Judge, G., Hill, G. R. C., Griffiths, W. E., Lütkepohl, H., and Lee, T.-C. 1988. *Introduction to the Theory and Practice of Econometrics*. New York: John Wiley & Sons.

Kelejian, H. H. 1971. Two stage least squares and econometric systems linear in the parameters but nonlinear in the endogenous variables. *Journal of the American Statistical Association*, 66: 373–4.

Kinder, D. R., and Palfrey, T. R. 1993. *Experimental Foundations of Political Science.* Ann Arbor: University of Michigan Press.

King, G. 1989. *Unifying Political Methodology*. Cambridge: Cambridge University Press.

—— 1991. On political methodology. *Political Analysis*, 2: 1–29.

Kingdon, J. W. 1973. *Congressman's Voting Decisions*. New York: Harper & Row.

Nakamura, A., and Nakamura, M. 1981. On the relationship among several specification tests presented by Durbin, Wu, and Hausman. *Econometrica*, 49 (6): 1583–88.

Newey, W. K. 1990. Efficient instrumental variable estimation of nonlinear models. *Econometrica*, 58 (4): 809–37.

Persson, T., and Tabellini, G. 2003. *The Economic Effects of Constitutions*. Cambridge, MA: MIT Press.

Rosenbaum, P. R. 2001. Observational studies: overview. Pp. 10810–15 in *International Encyclopedia of the Social and Behavioral Sciences*, ed. N. Smelser and P. Baltes. New York: Elsevier.

Sargan, J. 1958. The estimation of economic relationships using instrumental variables. *Econometrica*, 26 (3): 393–415.

Shea, J. 1997. Instrument relevance in multivariate linear models: a simple measure. *Review of Economics and Statistics*, 79 (2): 348–52.

Staiger, D., and Stock, J. H. 1997. Instrumental variables regression with weak instruments. *Econometrica*, 65 (3): 557–86.

Wooldridge, J. M. 2003. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.

Wu, D.-M. 1973. Alternative tests of independence between stochastic regressors and disturbances. *Econometrica*, 41 (4): 733–50.

Zellner, A. 1962. An efficient method of estimating seemingly unrelated regressions, and tests for aggregation bias. *Journal of the American Statistical Association*, 57: 348–68.

—— and Theil, H. 1962. Three-stage least squares: simultaneous estimation of simultaneous equations. *Econometrica*, 30: 54–78.