Instrumental Variables Estimation Using Quasi-Instrumental Variables, with an Application to Campaign Spending[1]

Jay Goodliffe

Department of Political Science
Brigham Young University
Provo, Utah 84602-5545
goodliffe@byu.edu

Abstract

Quasi-instrumental variables are instruments that are not perfectly exogenous (Bartels 1991). In this paper, I examine how different instrumental variable estimators are affected by using quasi-instruments instead of true instruments. Using Monte Carlo methods, I explore the properties of 2SLS, LIML, and Jackknife estimators. I find that all estimators are seriously biased and inconsistent. I then use these methods to estimate the effect of spending on electoral success in U.S. Senate elections, using data from Gerber (1998).

---

[1]I began this paper while visiting at Washington University in St. Louis. I am particularly thankful to Steve Smith for providing me space there. I am grateful to Charles Franklin and Walter Mebane for their helpful comments; and Alan Gerber for providing his data set. My intellectual debt to Larry Bartels is obvious.

1. Introduction

As Bartels (1985, 1991) and Leamer (1978) have noted, all econometric models are misspecified. Consider the case of instrumental variables. Suppose we have the following model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$
$$X_i = \pi_0 + \pi_1 Z_i + \eta_i$$

If *X* is correlated with ε as a result of omitted variables or endogeneity, then β_1 will be biased and inconsistent. A common remedy is to find an instrument for *X*, in this case *Z*. The assumptions required for *Z* to be a valid instrument for *X* are that *X* and *Z* are correlated, and *Z* and ε are not. There has been substantial work examining the case when the correlation between *X* and *Z* (or a set of *Z*'s) is weak (see, for example, Bound, Jaeger, and Baker 1995).[2] Less attention has been given to the case of misspecification when *Z* and ε are correlated. Bartels (1991) calls instruments that are correlated with the error term "quasi-instruments." This paper examines various instrumental variable (IV) estimators when the instruments are not truly exogenous, i.e. they are quasi-instruments. To isolate this type of misspecification, I investigate the effect of quasi-instruments on estimation when the instruments are strong.

Using Monte Carlo techniques, I investigate how Two Stage Least Squares (2SLS) and its recent extensions (Jackknife Instrumental Variable Estimators) and the Limited Information

[2] In fact, the main problem with weak instruments comes when the instrument is correlated with the error term. This correlation is always non-zero in finite samples, but even with very small correlation, sufficiently weak instruments can cause problems.

Maximum Likelihood (LIML) estimator and its variations fare when the instruments are not truly exogenous. The results are not encouraging. Although all of the IV estimators are better than Ordinary Least Squares (OLS), they are severely biased, and none of them are particularly better than any of the others.

The paper proceeds as follows: I briefly review the different estimators. Then I describe the Monte Carlo study and present its results. Finally, I apply the estimators to a recent study on the effects of spending in U.S. Senate races.

2. The Estimators

When faced with an endogenous regressor, the most common approach (in political science) is to use 2SLS. The basic approach is that various instruments are used to predict the endogenous regressor's value, and these predicted values are used instead of the actual regressor in the main regression. As the instruments are (by assumption) uncorrelated with the main equation error, the predicted value of the regressor will also be uncorrelated (in the limit). However, in finite samples, it is well-established that 2SLS is biased, e.g. Nagar (1959). This bias depends specifically on the number of instruments and the correlation between the errors in the two equations (ε, η).

Although invented before 2SLS, LIML is not used as frequently in econometrics, and hardly at all in political science. However, with cheap computational power, there is no reason not to use LIML.[3] In fact, Davidson and MacKinnon (1993) state that LIML is preferable to 2SLS when sample size is not large and with a large number of instruments (overidentifying restrictions). LIML also has the nice (maximum likelihood) property that the estimator is invariant to various transformations.[4] LIML is known as a *k*-class estimator, where *k* is estimated (in contrast with OLS, where *k* is set to 0; and 2SLS, where *k* = 1). Although LIML often has desirable small-sample properties (e.g. it is approximately median unbiased; see Anderson, Kunitomo, and Sawa 1982), it is particularly vulnerable to misspecification. Fisher (1967) suggests several circumstances where 2SLS may be less biased than LIML under small specification errors. Several authors have suggested ways of modifying LIML to reduce bias.[5] I examine the simplest of these proposed by Fuller (1977). He suggests that the investigator use $k = \hat{k} - \alpha/(n - K)$, where $\hat{k}$ is the *k* used by LIML, n is the number of observations, and K is the number of RHS variables in the main equation. α is chosen by the investigator. Fuller suggests $\alpha = 1$ for unbiasedness and $\alpha = 4$ for minimum mean squared error. Besides LIML, I will use both of these recommended modifications to LIML, and call them LIMLF1 and LIMLF4.

Recently, several jackknifed modifications of 2SLS have been proposed (Angrist and Krueger 1995, Angrist, Imbens and Krueger 1999). In this paper, I implement both of the Jackknife Instrumental Variable Estimators proposed by Angrist, Imbens, and Krueger (1999),

---

[3]I followed the LIML implementation of Greene (2000), adapting the corresponding LIMDEP program on his web site.

[4]This is particularly desirable in the area of campaign finance, where there is no agreement about the functional form of the spending-vote relationship.

[5]There have also been estimators proposed to reduce bias in 2SLS, which I leave for a future day.

called JIVE1 and JIVE2.[6] The intuition behind JIVE is that the bias in 2SLS is (partially) caused by correlation between the error terms in the two equations (ε, η), specifically between the error term of the *i*th observation in first equation and the error term of the *i*th observation in the second equation. For the *i*th observation, JIVE uses all but the *i*th observation in the instruments to obtain a predicted value. Angrist, Imbens and Krueger claim that JIVE has the desirable properties of both LIML and 2SLS.[7]

In addition to the estimators discussed above, I will also use OLS, which is generally more biased than 2SLS, but with less variance. Thus, the estimators I will examine in this paper are OLS, 2SLS, LIML, LIMLF1, LIMLF4, JIVE1 and JIVE2.

3. The Monte Carlo Study

This section reports on the finite-sample properties of different instrumental variables estimators when the instruments are not exogenous. I report quantiles of the Monte Carlo distribution, root mean squared error (RMSE),[8] median absolute error (MAE), and coverage rates for the 95% confidence interval computed using the usual *t*-statistics.[9]

---

[6]I adapted a SAS program written by Alan Krueger, found on his web site.

[7]In a Monte Carlo investigation, Blomquist and Dahlberg (1999) find that the JIVE estimators are no better (and no worse) than 2SLS or LIML with weak instruments and small sample sizes.

[8]Although LIML (and JIVE1 and JIVE2) has no finite moments (in the limiting distribution), the Monte Carlo distribution does have a mean and variance.

[9]The Monte Carlo program is written in STATA (7.0).

In the first model, there is a single overidentifying restriction. In the second model, there are multiple overidentifying restrictions. In both models, there are 200 observations (which will correspond approximately with the application that follows).

Model 1

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + \varepsilon_i \\ X_i &= \pi_0 + \pi_1 Z_{1i} + \pi_2 Z_{2i} + \eta_i \\ Z_{ji} &= \gamma_{j0} + \gamma_{j1} W_{ji} + \nu_{ji} ; j = 1,2 \end{aligned}$$

with $\beta_0 = 1$, $\beta_1 = 0.3$, $\pi_0 = 0$, $\pi_1 = 0.3$, $\pi_2 = 0.3$, $\gamma_0 = 0$, $\gamma_{j1} = 0.3$, and

$$\begin{pmatrix} \varepsilon_i \\ \eta_i \\ \nu_{ji} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.25 & & \\ 0.10 & 0.25 & \\ 0.04 & 0.10 & 0.25 \end{pmatrix} \right)$$

All W_{ji} are independent, normally distributed random variables with zero mean and unit variance. The corresponding correlation matrix is:

corr	ε	η	ν
ε	1		
η	0.40	1	
ν	0.16	0.40	1

In addition, this induces the following correlations between ε and X and Z and $\hat{X}$ (the predicted X from a regression of X on Z_1 and Z_2):

correlation	ε
X	0.38
Z_j	0.14
$\hat{X}$	0.20

Finally, the auxiliary regression of X on Z_1 and Z_2 yields an R^2 of 0.56. Adapting a formula in Bound, Jaeger, and Baker (1995), 2SLS is superior to OLS (in the probability limit) if $\text{cov}(\hat{X};\varepsilon) < \text{cov}(X;\varepsilon)\times R^2(X;Z_1,Z_2)$. In this case, $0.045 < 0.122 \times 0.56 = 0.068$, so 2SLS dominates OLS from an asymptotic mean squared error criterion in the models.

Table 1 presents the results of Model 1.

Table 1: One Overidentifying Restriction, n = 200, 10000 replications

	quantiles ($\beta_1 = 0.3$)							
	0.10	0.25	0.50	0.75	0.90	RMSE	MAE	95% CI Coverage
OLS	0.524	0.555	0.588	0.621	0.652	0.292	0.288	0
2SLS	0.411	0.453	0.498	0.543	0.582	0.208	0.198	0.185
LIML	0.410	0.452	0.497	0.542	0.582	0.208	0.197	0.188
LIMLF1	0.411	0.453	0.498	0.543	0.582	0.208	0.198	0.183
LIMLF4	0.419	0.461	0.506	0.551	0.591	0.216	0.206	0.148
JIVE1	0.407	0.450	0.495	0.541	0.581	0.206	0.195	0.204
JIVE2	0.407	0.450	0.496	0.541	0.581	0.206	0.196	0.205

With the exception of OLS, the instrumental variable estimators have a very similar distribution. Although OLS is severely biased (with β_1 estimated at about twice its value), all of the other estimators are biased as well. Most likely, the endogeneity of the instruments causes the estimators to be biased. The LIMLF4 estimator is marginally worse than the other IV estimators, and JIVE1 and JIVE2 appear to have the best properties, from a RMSE, MAE or coverage perspective, though not by much.

Model 2

Model 2 adds 19 worthless instruments to Model 1. In general, as the number of instruments increases, one expects 2SLS to deteriorate. Since the instruments as a whole are not weak, however, this may not be the case.

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + \varepsilon_i \\ X_i &= \pi_0 + \pi_1 Z_{1i} + \pi_2 Z_{2i} + \eta_i \\ Z_{ji} &= \gamma_{j0} + \gamma_{j1} W_{ji} + \nu_{ji} \end{aligned}$$

with $\beta_0 = 1$, $\beta_1 = 0.3$, $\pi_0 = 0$, $\pi_1 = 0.3$, $\pi_2 = 0.3$, $\pi_3 = \pi_4 = \cdots = \pi_{21} = 0$, $\gamma_{j0} = 0$, $\gamma_{j1} = 0.3$, and

$$\begin{pmatrix} \varepsilon_i \\ \eta_i \\ \nu_{ji} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.25 & & \\ 0.10 & 0.25 & \\ 0.04 & 0.10 & 0.25 \end{pmatrix} \right)$$

All W_{ji} are independent, normally distributed random variables with zero mean and unit variance. As the instruments added are worthless, the covariance and correlation matrices (and R^2 of the auxiliary regression) are the same as Model 1. Table 2 has the results of Model 2.

Table 2: Twenty Overidentifying Restrictions, n = 200, 10000 replications

	quantiles (β_1 = 0.3)							
	0.10	0.25	0.50	0.75	0.90	RMSE	MAE	95% CI Coverage
OLS	0.525	0.555	0.589	0.622	0.652	0.293	0.289	0.0002
2SLS	0.430	0.471	0.515	0.557	0.596	0.224	0.215	0.109
LIML	0.408	0.452	0.500	0.546	0.588	0.211	0.200	0.190
LIMLF1	0.410	0.453	0.500	0.546	0.588	0.212	0.200	0.185
LIMLF4	0.417	0.461	0.508	0.554	0.596	0.219	0.208	0.155
JIVE1	0.404	0.450	0.498	0.545	0.587	0.210	0.198	0.228
JIVE2	0.404	0.450	0.498	0.545	0.587	0.210	0.198	0.232

Compared to Model 1, the OLS estimates of Model 2 are very similar (and still very biased). As expected, the 2SLS estimator suffers as a result of the added (weak) instrumental variables. In contrast, the other estimators are about the same as Model 1. (In fact, the coverage improves, but this is most likely a result of added worthless estimators which increase the variance of the estimator.) But similar to Model 1, all of the estimators are biased; most of the time, the 95% confidence interval does not cover the true parameter.

I also ran both Models 1 and 2 using 30 observations and 800 observations (results not shown). With fewer observations, standard errors increase, and coverage increases, so ironically, with fewer observations, researchers may incorrectly reject the null less often. With more observations, standard errors decrease, coverage decreases so much that no confidence interval covers the true parameter. Also as expected, with fewer observations, RMSE increases, and with more observations, RMSE and MAE decrease. Adding more observations also removes the comparative disadvantage that 2SLS (and LIMLF4) have.

From the Monte Carlo studies, it appears that with few overidentifying restrictions, either 2SLS or LIML or JIVE will perform similarly, although with endogenous instruments, none of the estimators are particularly good (though better than OLS). In contrast, with many overidentifying restrictions (and especially in small samples), LIML or JIVE outperform 2SLS, which is similar to what other studies have found when including exogenous instruments.

4. Empirical Application

In this section, I apply some of the estimators examined above to the case of campaign spending in U.S. Senate races. As LIMLF1 and LIMLF4 had similar (and at times inferior) properties to LIML, I only examine LIML. And since JIVE1 and JIVE2 were very similar, I will only examine JIVE1.

I use data from and extend the analysis of Gerber (1998), applying OLS, 2SLS, LIML, and JIVE1. The question Gerber addresses is the relative efficacy of spending by incumbents and challengers in U.S. Senate races (from 1974 to 1992). A well-known problem in this area is that spending is an endogenous variable in the prediction of incumbent vote. Gerber carefully selects instrumental variables that he argues are correlated with spending but not the incumbent vote. These variables are challenger wealth, state (voting age) population, and lagged spending—taken from the state's most recent Senate seat election (which is not the current seat/election).

As Milyo (1998) has argued, it is very difficult to overcome two problems that induce correlation between proposed instruments and the error term. The first problem is that it is difficult to find variables that are correlated with spending, but not the vote. In this example, challenger wealth may cause a problem: if voters are drawn to wealthy candidates not because of the money they spend, but because they are drawn to, say, successful entrepreneurs, then challenger spending is not a true instrument. The second problem is that the instrumental variables "must not be correlated with omitted causes of electoral success" (Milyo 1998, 3). The standard unobserved (and omitted) variable in this literature is candidate charisma or personality. It may be that more charismatic candidates appear in states with larger populations (Gray Davis, notwithstanding). It may be that charismatic challengers are also, on average, more wealthy. And finally, it may be that a certain state has more charismatic candidates in reserve (due to, say, the state legislature), who happen to raise and spend more money. Thus, even the other race's spending may be a quasi-instrument.

This is not to say that a researcher should throw up her arms and give up. This is merely to point out that there is always a chance of misspecification, and to examine its likely consequences.

I examine three models (two used by Gerber) of the effect of spending on the vote. Model A only includes controls for year and party as exogenous variables and uses the variables discussed above as instruments.[10] There is one overidentifying restriction as challenger spending and incumbent spending are endogenous RHS variables, and the three variables discussed used as instruments. As variable controls are intentionally omitted, there is the potential for omitted variable bias. Indeed, a Hausman test of exogeneity of the instruments rejects the null of exogeneity ($p = 0.02$).

[10]Model A corresponds to Column 3 in Table 3 of Gerber (1998).

Model B includes 12 (more) variables controlling for such things as challenger background (e.g. officeholder), incumbent vulnerability (e.g. scandal), and state-level variables (e.g. partisanship).[11] These variables are posited to have a direct effect on the vote, and thus are included in the main equation. Since these added variables are not in the auxiliary equation, there is still only one overidentifying restriction.[12]

Model C makes the 12 variables added in Model B instruments, i.e., they no longer affect the incumbent vote directly, but instead work through incumbent and challenger spending. This yields a model with 13 overidentifying restrictions. Although this model is almost certainly misspecified, I include it for comparison. Indeed, a Hausman test of the overidentifying restrictions rejects the null that the restrictions are valid ($p < 0.05$).[13]

For simplicity, I only report the coefficients (and standard errors) of the spending variables;[14] I do not report the coefficients for the control variables (or instruments). There 156 observations in the data set.[15] The results are in Table 3.

---

[11]Model B corresponds to Column 4 in Table 3 of Gerber (1998).

[12]I am grateful to Alan Gerber for providing his data. I was able to replicate all tables in his paper.

[13]Following Bartels (1991), I calculated the strength of the instruments, after controlling for the other exogenous variables. The instruments are weakest in Model A (0.13 and 0.23), moderate in Model B (0.34 and 0.37), and strongest in Model C (0.49 and 0.50).

[14]The functional form of spending is ln(real spending/voter + 0.01)

[15]Gerber (1998) also considers the case where lagged spending is not used as an instrument. However, in this case the model is just identified, and thus 2SLS is equivalent to LIML.

Table 3: Estimates of the Effects of Campaign Spending on Incumbent Vote

	Model A		Model B		Model C	
	chal spend	inc spend	chal spend	inc spend	chal spend	inc spend
OLS b	−5.31	2.78	−3.94	1.99	−5.31	2.78
(se)	(0.57)	(0.92)	(0.58)	(0.58)	(0.57)	(0.92)
2SLS b	−6.92	8.78	−5.57	6.06	−8.29	7.85
(se)	(2.95)	(3.30)	(2.68)	(2.90)	(1.09)	(1.84)
LIML b	−6.96	8.84	−5.65	6.15	−10.48	11.65
(se)	(2.79)	(3.12)	(2.45)	(2.64)	(1.45)	(2.47)
JIVE1 b	−5.53	11.01	−3.57	6.80	−11.96	14.51
(se)	(2.76)	(4.13)	(1.85)	(3.17)	(2.85)	(5.15)

The first thing to notice about these results is that they are much more unstable than the simulated Monte Carlo results. That is, the results depend on the estimator used as well as the model. (Since there is only one data set, not 10,000 simulated data sets, this is not that surprising.) As Model A has few overidentifying restrictions, but likely has endogenous instruments, it roughly corresponds to Model 1 in the Monte Carlo study. Like Model 1, the LIML and 2SLS estimates are very similar. Surprisingly, the JIVE estimator shows a much stronger effect for incumbent spending.

Model C has several (rejected) overidentifying conditions, which cause misspecification (see Fisher 1967). It roughly corresponds to Model 2 in the Monte Carlo study. In that Model, LIML and JIVE performed better than 2SLS (and OLS), though all were biased. According to Angrist, Imbens, and Krueger’s (1995) results, both 2SLS and JIVE should be consistent, while LIML should have difficulty. But in the results shown here, both LIML and JIVE coefficients are much larger, most likely because the control variables are constrained to act through the spending variables, making their effect seem larger.

Model B is probably the best specification. Even in this specification (with few overidentifying restrictions and relatively strong instruments), there are some differences between the estimators. In particular the JIVE coefficient of challenger spending is lower than the other estimates (although their confidence intervals overlap). If anything the results show that incumbents are more effective with their money than challengers (a finding similar to Erikson and Palfrey 1998 in U.S. House races). Angrist, Imbens, and Krueger suggest that when different estimators yield different results, that there is misspecification in the model.

5. Discussion

There is no magic solution to endogenous instruments. When the instruments are strongly endogenous, no estimator works well. An alternative is to use other instruments, but any instruments that are more exogenous are usually weaker. If one chooses this path, then LIML or JIVE are better estimators (particularly with a large number of instruments and smaller sample size).

The next step in this program is to extend the Monte Carlo studies where the relationship between error terms is weaker and/or the instruments are weaker. A political science application of such a situation is Bartels (1994). I will then evaluate how much endogeneity could eliminate (or induce) the findings of previous campaign spending studies (following Wand (2002)).

If there are no alternative instruments, then one should consider whether OLS would work just as well (following Bartels 1985). And alternative estimators should be used to ensure that the results are robust.

References

Anderson, T.W., N. Kunitomo, and T. Sawa. 1982. "Evaluation of the distriubtion function of the limited information maximum likelihood estimator." *Econometrica* 50:1008-1027.

Angrist, J.D., G.W. Imbens, and A.B. Krueger. 1999. "Jackknife Instrumental Variables Estimation." *Journal of Applied Econometrics* 14:57-67.

Angrist, J.D., and A.B. Krueger. 1995. "Split Sample Instrumental Variables Estimates of the Return to Schooling." *Journal of Business and Economic Statistics* 13:225-235.

Bartels, Larry M. 1985. "Alternative Misspecifications in Simultaneous-Equation Models." *Political Methodology* 11:181-199.

Bartels, Larry M. 1991. "Instrumental and 'Quasi-Instrumental' Variables." *American Journal of Political Science* 35:777-800.

Bartels, Larry M. 1994. "The American Public's Defense Spending Preferences in the Post-Cold War Era." *Public Opinion Quarterly* 58:479-508.

Blomquist, Sören, and Matz Dahlberg. 1999. "Small Sample Properties of LIML and Jackknife IV Estimators: Experiments with Weak Instruments." *Journal of Applied Econometrics* 14:69-88.

Bound, John, David A. Jaeger, and Regina M. Baker. 1995. "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association* 90:443-450.

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

Erikson, Robert S., and Thomas R. Palfrey. 1998. "Campaign Spending and Incumbency: An Alternative Simultaneous Equations Approach." *Journal of Politics* 60:355-373.

Fisher, Franklin M. 1967. "Approximate Specification and the Choice of a *k*-Class Estimator." *Journal of the American Statistical Association* 62:1265-1276.

Fuller, W.A. 1977. "Some Properties of a Modification of the Limited Information Estimator." *Econometrica* 45:939-953.

Gerber, Alan. 1998. "Estimating the Effect of Campaign Spending on Senate Election Outcomes Using Instrumental Variables." *American Political Science Review* 92:401-411.

Greene, William H. 2000. *Econometric Analysis*, 4th ed. Upper Saddle River, New Jersey: Prentice Hall.

Leamer, Edward E. 1978. *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York: John Wiley & Sons.

Milyo, Jeffrey. 1998. *The Electoral Effects of Campaign Spending in House Elections*. Citizens' Research Foundation.

Nagar, A.L. 1959. "The Bias and Moment Matrix of the General *k*-Class Estimators of the Parameters in Simultaneous Equations." *Econometrica* 27:575-595.

Wand, Jonathan. 2002. "Evaluating the Consequences of Assumptions Using Simulations." *The Political Methodologist* 11(1):21-25.