SMALL SAMPLE PERFORMANCE OF INSTRUMENTAL VARIABLES PROBIT ESTIMATORS: A MONTE CARLO INVESTIGATION

LEE C. ADKINS

Presented at the Joint Statistical Meetings, Denver, Colorado, August 2008.

Abstract. In this paper I revisit the question of how several estimator of an endogenous probit regression model perform in small samples. Modern software usually contains two estimator that can be used to estimate such a model. A simple generalized least squares estimator suggested by Amemiya and explored by Newey is computationally simple, though not necessarily efficient. A maximum likelihood estimator is also used, though its properties are less apparent in small samples. The paper uses a simple experimental design employed by Rivers and Vuong (1988) to estimate the parameters of an endogenous probit model and conduct subsequent tests of parameter significance. Although Rivers and Vuong (1988) find that their two-stage conditional maximum likelihood (2SCML) performs well in terms of bias and mean square error, and similarly to other consistent alternatives, they did not examine how well the estimators perform in significance tests. In the probit model it is not altogether clear what the magnitude of the parameters actually mean; however, getting the correct signs and being able to test for parameter significance is important. So, this paper can be seen as an important extension of their work.

I add to the list of estimators compared, increase the dimension of the experimental design, and explore the size of significance tests based on these estimators. Most importantly, the effect of instrument strength is explored. Other dimensions that affect the performance of the estimators are modeled, including sample size, proportion of observations equal to 1, correlation between instruments and endogenous variables, correlation between endogenous regressor and equation error, and overidentification. Finally, the estimators are used in an example to examine the effect of managerial incentives on the use of foreign-exchange derivatives.

Date: September 3, 2008.

Key words and phrases. Probit, Instrumental Variables, Monte Carlo, significance tests.

1. Introduction

Yatchew and Griliches (1985) analyze the effects of various kinds of misspecification on the probit model. Among the problems explored was that of errors-in-variables. In linear regression, a regressor measured with error causes least squares to be inconsistent and Yatchew and Griliches find similar results for probit. Rivers and Vuong (1988) and Smith and Blundell (1985) suggest two-stage estimators for probit and tobit, respectively. The strategy is to model a continuous endogenous regressor as a linear function of the exogenous regressors and some instruments. Predicted values from this regression are then used in the second stage probit or tobit. These two-step methods are not efficient, but are consistent. Consistent estimation of the standard errors is not specifically considered and these estimators are used mainly to test for endogeneity of the regressors–not their statistical significance.

Newey (1987) explores the more generic problem of endogeneity in limited dependent variable models (which include probit and tobit). He proposes what is sometimes called Amemiya's Generalized Least Squares (AGLS) estimator as a way to efficiently estimate the parameters of probit or tobit when they include a continuous endogenous regressor. This has become a standard way to estimate these models and is an option in Stata 10.0 when the MLE is difficult to obtain. The main benefit of using this estimator is that it produces a consistent estimator of the standard errors and can easily be used for subsequent hypothesis testing of the parameters.

More recent papers have explored limited dependent variable models that have discrete endogenous regressors. Nicoletti and Peracchi (2001) look at binary response models with sample selection, Kan and Kao (2005) consider a simulation approach to modeling discrete endogenous regressors, and Arendt and Holm (2006) extends Nicoletti and Peracchi (2001) to include multiple endogenous discrete variables.

Iwata (2001) uses a very simple approach to dealing with errors-in-variables for probit and tobit. He shows that simple recentering and rescaling of the observed dependent variable may restore consistency of the standard IV estimator if the true dependent variable and the IVs are jointly normally distributed. His Monte Carlo simulation shows evidence that the joint normality may not be necessary to obtain improved results. However, the results for tobit were quite a bit better than those for probit. He compares this estimator to a linear instrumental variable estimator that uses a consistent estimator of standard errors. This estimator is used below.

A paper by Blundell and Powell develops and implements "semiparametric methods for estimating binary response (binary choice) models with continuous endogenous regressors. It extends existing results on semiparametric estimation in single-index binary response models to the case of endogenous regressors. It develops an approach to account for endogeneity in triangular and fully simultaneous binary response models." (Blundell and Powell, 2004, p. 655)

In this paper I compare the AGLS estimator to several of these alternatives. The AGLS estimator is useful because it is simple to compute and yields consistent estimators of standard error that can be used for significance tests of the model's parameters. The other plug-in estimators (like 2SCML) are consistent for the parameters but not the standard errors, making it unlikely that they will perform satisfactorily in hypothesis testing.

The Monte Carlo design is based on that of Rivers and Vuong (1988), which gives us a way to calibrate results. Rivers and Vuong (1988) compare several limited information estimators for simultaneous probit models. The comparison includes three different 2-step estimators and a limited information maximum likelihood estimator (ML). They are compared based on computation ease, bias and MSE, asymptotic efficiency, and as the basis for an exogeneity test. In these limited dimensions, the 2SCML actually performs reasonably well compared to the ML estimator.

Still there are a few issues left unresolved by Rivers and Vuong (1988). First, the instruments they use are very strong and variation in instrument strength is not part of their design. In light of the what we now know about the effect of weak instrument in linear regression (Staiger and Stock (1997); Stock and Yogo (2005) we need to know how these estimators perform when the instruments are weak. Second, they examine the performance of several estimators only in the case where the proportion of 1s and 0s in the binary dependent variable is balanced. Following Zuehlke and Zeman (1991), we need to know whether this proportion affects bias and testing in these models. Third, Rivers and Vuong only examine the bias of the estimators; but, how do they measure up as a means of testing parameter significance? Finally, a few alternative estimators are added to the simulation 'derby', principally a pretest estimator, to determine whether the poor properties of 2SCML or AGLS can be improved when endogeneity is not a problem.

2. Linear model

Following the notation in Newey (1987), consider a linear statistical model in which the continuous dependent variable will be called y_t^* but it is not directly observed. Instead, we observe y_t in only one of two possible states. So,

$$y_t^* = Y_t\beta + X_{1t}\gamma + u_t = Z_t\delta + u_t, \quad t = 1, \ldots, N \tag{2.1}$$

where $Z_t = [Y_t, X_{1t}], \delta' = [\beta', \gamma']$, Y_t is the tth observation on an endogenous explanatory variable, X_{1t} is a 1xs vector of exogenous explanatory variables, and δ is the $qx1$ vector of regression parameters.

The endogenous variable is related to a $1XK$ vector of instrumental variables X_t by the equation

$$Y_t = X_{1t}\Pi_1 + X_{2t}\Pi_2 + V_t = X_t\Pi + V_t \tag{2.2}$$

where V_t is a disturbance. The $K - s$ variables in X_{2t} are additional exogenous explanatory variables. Equation (2.2) is the reduced form equation for the endogenous explanatory variable. Without loss of generality only one endogenous explanatory variable is considered below. See Newey (1987) for notation extending this to additional endogenous variables.

When the continuous variable y_t^* is observed, then one could use either least squares or instrumental variable estimator to estimate δ. Collecting the n observations into matrices y^*, X, and Z of which the t^{th} row is y_t^*, X_t, and Z_t, respectively we have the least squares estimator of δ, $\hat{\delta}_{ols} = (Z^T Z)^{-1} Z^T y^*$, which is biased and inconsistent.

The instrumental variable estimator uses the orthogonal projection of Z onto the column space of X, i.e., $P_X Z$ where $P_X = X(X^T X)^{-1} X^T$. The IV estimator is

$$\delta_{liv} = (Z^T P_X Z)^{-1} Z^T P_X y^*. \tag{2.3}$$

The (linear) instrumental variable estimator is biased in finite samples, but consistent. The heteroskedasticity robust estimator of covariance (Davidson and MacKinnon, 2004, p. 335) is

$$\hat{\Sigma}_{HCCME} = (Z^T P_X Z)^{-1} Z^T P_X \hat{\Phi} P_X Z (Z^T P_X Z)^{-1} \tag{2.4}$$

where $\hat{\Phi}$ is an nxn diagonal matrix with the t^{th} diagonal element equal to $\hat{u}_t^2$, the squared IV residual.

3. Binary Choice

In some cases, y_t^* is not directly observed. Instead, we observe

$$y_t = \begin{cases} 1 & y_t^* > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

Assuming the errors of the model (2.1) are normally distributed leads to the probit model.

There are several estimators of this model that will be considered, some consistent for δ and others not. The first is least squares. The least squares estimator $\hat{\delta}_{ols} = (Z^TZ)^{-1}Z^Ty^*$ is consistent if Z is exogenous. If any of the elements of Z are endogenous then it is not. Still, it is easy to compute and the degree of inconsistency may be small in certain circumstances.

The linear instrumental variable estimator (2.3) is also inconsistent and heteroscedastic. Iwata (2001) suggests a means of rescaling and recentering (RR) the data that can bring about consistency in this case. However, in his Monte Carlo the RR versions of OLS and IV estimation don't perform particularly well for probit (although much better for tobit).

Next, the usual probit mle can be estimated. However, if the regressors are endogenous, then this estimator is also inconsistent (Yatchew and Griliches (1985)). To develop the notation, let the probability that y_t is equal one be denoted

$$pr(y_t = 1) = \Phi(y_t, Y_t\beta + X_{1t}\gamma) = \Phi(y_t, Z_t\delta) \tag{3.2}$$

where Φ is the normal cumulative density, y_t is the observed binary dependent variable, and $Y_t\beta + X_{1t}\gamma$ is the (unnormalized) index function. As usual, the model is normalized assuming $\sigma^2 = 1$. Basically, this equation implies that Y_t, and X_{1t} be included as regressors in the probit model and the log likelihood function is maximized with respect to $\delta^T = [\beta^T, \gamma^T]$. Since the endogeneity of Y_t is ignored, the mle is inconsistent.

The next estimator used predicted values of Y_t from a first stage least squares estimation of equation (2.2). Denote the first stage as $\hat{Y}_t = X_{1t}\hat{\Pi}_1 + X_{2t}\hat{\Pi}_2 = X_t\hat{\Pi}$ where $X_t = [X_{1t} \vdots X_{2t}]$ and $\hat{\Pi}^T = [\hat{\Pi}_1^T \vdots \hat{\Pi}_2^T]$. Then the conditional probability

$$pr(y_t = 1) = \Phi(y_t, \hat{Z}_t\delta) \tag{3.3}$$

with $\hat{Z}_t = [\hat{Y}_t \vdots X_{1t}]$. The parameters are found by maximizing the conditional likelihood. This is referred to here as IV probit (IVP).

Another estimator adds the least squares residuals from equation (2.2), $\hat{V}_t = Y_t - X_t\hat{\Pi}$ to (3.3). This brings

$$pr(y_t = 1) = \Phi(y_t, \hat{Y}_t\beta + X_{1t}\gamma + \hat{V}_t\lambda) = \Phi(y_t, \hat{Z}_t\delta + \hat{V}_t\lambda) \tag{3.4}$$

which is estimated by maximum likelihood, again conditional on $\hat{\Pi}$. This is similar to an estimator used by Rivers and Vuong (1988) which takes the form

$$pr(y_t = 1) = \Phi(y_t, Z_t\delta + \hat{V}_t\rho) \tag{3.5}$$

The parameter ρ is related to λ in (3.4) by $\lambda = \rho + \beta$. This follows because $Z_t\delta = \hat{Z}_t\delta + \hat{V}_t\beta$.

An efficient alternative to (3.4) is Amemiya's generalized least squares (AGLS) estimator as proposed by Newey (1987). The AGLS estimator of the endogenous probit model is fairly easy to compute, though there are several steps. The basic algorithm proceeds as follows:

(1) Estimate the reduced form (2.2), saving the estimated residuals, $\hat{V}_t$ and predicted values $\hat{Y}_t$.
(2) Estimate the parameters of a reduced form equation for the probit model using the mle. In this case,

$$pr(y_t = 1) = \Phi(y_t, X_t\alpha + \hat{V}_t\lambda) \tag{3.6}$$

Note that all exogenous variables, X_{1t} and instruments X_{2t} are used in the probit reduced form and the parameters on these variables is labeled α. Let the mle be denoted $\hat{\alpha}$. Also, save the portion of the estimated covariance matrix that corresponds to $\hat{\alpha}$, calling it $\hat{J}^{-1}_{\alpha\alpha}$.
(3) Another probit model is estimated by maximum likelihood. In this case it is the 2SIV estimator of equation (3.4). Save $\hat{\rho} = \hat{\lambda} - \hat{\beta}$ which is the coefficient of $\hat{V}_t$ minus that of $\hat{Y}_t$.
(4) Multiply $\hat{\rho}Y_t$ and regress this on X_t using least squares. Save the estimated covariance matrix from this, calling it $Hat\Sigma$.
(5) Combine the last two steps into a matrix, $\Omega = \hat{J}^{-1}_{\alpha\alpha} + \hat{\Sigma}$.
(6) Then, the AGLS estimator is

$$\delta_A = [D(\hat{\Pi})^T\Omega^{-1}D(\hat{\Pi})]^{-1}D(\hat{\Pi})^T\Omega^{-1}\hat{\alpha} \tag{3.7}$$

The estimated variance covariance is $[D(\hat{\Pi})^T\Omega^{-1}D(\hat{\Pi})]^{-1}$ and $D(\hat{\Pi}) = [\hat{\Pi}\vdots I_1]$ where I_1 is a Kxs selection matrix such that $X_{1t} = X_tI_1$.

Below is a summary of the estimators used in the simulation:

Estimator	Variables	Parameters	Equation
OLS	Z_t	δ, σ^2	$(Z^TZ)^{-1}Z^Ty$
Linear IV	$\hat{Z}_t$	δ, σ^2	$(Z^TP_XZ)^{-1}Z^TP_Xy$
Probit mle	Z_t	δ	(3.2)
Probit mle (IVP)	$\hat{Z}_t$	δ	(3.3)
Probit mle	$\hat{Z}_t$, $\hat{V}_t$	δ, λ	(3.4)
Probit mle (2SCML)	Z_t, $\hat{V}_t$	δ, ρ	(3.5)
Probit mle (RF)	X_t, $\hat{V}_t$	α, λ	(3.6)
AGLS	$D(\hat{\Pi})$, $\hat{\alpha}$	δ	(3.7)

Rivers and Vuong (1988) compare several of these. They compare (3.3), (3.4), and (3.7).

One thing that complicates comparison of these estimators is that some use a different normalization. One alternative is to compare marginal effects. This is the approach taken by Arendt (2001). This choice is appealing since this is the quantity that interests many. However, I have chosen to compare coefficients, which will require some adjustments to the design to eliminate any differences that would be due to the different normalizations.

Rivers and Vuong also consider several tests of exogeneity. A simple Wald test was shown to be useful in testing exogeneity of the regressor and it will be used as the basis of a pretest estimator that will also be considered. The pretest estimator can be written

$$\delta_{pt} = I(t)_{[0,c_\alpha)}\delta_{mle} + I(t)_{[c_\alpha,\infty)}\delta_{iv} \tag{3.8}$$

where $I(t)_{a,b}$ is an indicator function that takes the value of 1 if t falls within the $[a, b)$ interval and is zero otherwise. In our example, t will be the test statistic associated with the exogeneity null hypothesis, c_α is the α level critical value from the sampling distribution of t, δ_{mle} is the usual probit mle and δ_{iv} is one of the instrumental variables probit estimators.

Since none of the IV probit estimators perform very well when the regressor is exogenous, one usually tests this proposition first to determine which estimator to use. Below a pretest is conducted and IV or probit is estimated based on the outcome of this test.

4. Simulation

The statistical properties of the various estimators of an endogenous probit model will be studied using simulation. Bias, standard error, and the size of a test of significance on the endogenous variable will be studied. There are various dimensions that can affect the performance of estimators of this model. Sample size, proportion of observations where $y_t = 1$, correlation between instruments and the endogenous variable, the correlation between the endogenous variable and the equation's error, and the relative variability of the endogenous regressor and the equation's error. Other dimensions could be examined. These include the effects of overidentification and of other estimators of the model.

Once a successful estimator is found, additional simulations are conducted to determine if marginal improvements can be obtained by using ML estimators and tests. These simulations are much more limited in scope due to their computational complexity. Still, based on the performance of the simple estimators, interesting design points can be explored and one can gain a good idea of how the ML (or bootstrapping) may perform in practice.

4.1. **Design.** A simple model is considered that has a single, possibly endogenous, regressor. The Monte Carlo design shares some similarity to that of Hill et al. (2003) which is based on Zuehlke and Zeman (1991), and modified by Nawata and Nagase (1996). To make comparisons with prior research easier to make, the design of Rivers and Vuong (1988) is incorporated as well and their notation will be adopted with some minor modifications. The performances of the estimators are examined under various circumstances. This list may grow as the research proceeds.

The vector of endogenous explanatory variables contains a constant and one continuous explanatory variable, y_{2i}, and an exogenous regressor, x_{2i}.

$$y_{1i}^* = \gamma y_{2i} + \beta_1 + \beta_2 x_{2i} + u_i \tag{4.1}$$

In the just identified case

$$y_{2i} = \pi_1 + \pi_2 x_{2i} + \pi_3 x_{3i} + \nu_i \tag{4.2}$$

and the over-identified case,

$$y_{2i} = \pi_1 + \pi_2 x_{2i} + \pi_3 x_{3i} + \pi_4 x_{4i} + \nu_i \tag{4.3}$$

The exogenous variables (x_{2i}, x_{3i}, x_{4i}) are drawn from multivariate normal distribution with zero means, variances equal 1 and covariances of .5. The disturbances are creates using

$$u_i = \lambda \nu_i + \eta_i \tag{4.4}$$

where ν_i and η_i standard normals and the parameter λ is varied on the interval $[-2, 2]$ to generate correlation between the endogenous explanatory variable and the regression's error. The parameters of the reduced form are $\theta\pi$ where $\pi = \pi_1 = 0, \pi_2 = 1, \pi_3 = 1, \pi_4 = -1$ and θ is varied on the interval $[.05, \quad 1]$. This allows us to vary the strength of the instruments, a design element not considered in Rivers and Vuong (1988).

In the probit regression, $\beta_2 = -1$. The intercept takes the value $-2, 0, 2$, which corresponds roughly to expected proportions of $y_{1i} = 1$ of 25%, 50%, and 75%, respectively. In terms of the notation developed in the preceding section $\delta = \gamma, \beta_1, \beta_2$. For the simulation, $\gamma = 0$. This will make it possible to compare test sizes without adopting different normalizations for the various models. Other simulations were conducted with $\gamma = 1$ and no substantive differences were noted. When $\gamma = 0$, the endogenous regressor is still correlated with the probit equation's error even though it has no direct effect on y_{1i}. This allows us to measure the size of a t-test on the endogenous variable without having to worry about differences in scaling under different parameterizations of the model (Rivers and Vuong, 1988, p. 361).

Two sample sizes are considered, 200 and 1000. One thousand Monte Carlo samples are generated for each combination of parameters. Several statistics are computed at each round of the simulation. These include the estimator of $\delta = [\gamma, \beta_1, \beta_2]$, an estimate of their standard errors, a t-ratio of the hypothesis that $\gamma = 0$ (for size). Power will be examined separately and only indirectly when a comparison is made with the ML estimator. A direct comparison is difficult due to the implied differences in scaling.

Below you will find a summary of the design characteristics of the Monte Carlo experiments. The first design element is variation of the parameter λ. This parameter controls the degree of correlation between the endogenous explanatory variable and the probit's error. When $\lambda = 0$, the regressor is exogenous and the usual probit (or least squares/linear probability model) should perform satisfactorily. The correlations associated with each value of λ are given below. Also, I have included the parameter ω, which measures the standard error of the probit's reduced form error[1]. Notice that higher values of correlation increase the standard error of the reduced form. Also, these values differ a bit from Rivers and Vuong (1988) since I have let $\lambda = 0$.

	λ						
	2	1	0.5	0	-0.5	-1	-2
corr(u,v)	0.894	0.707	0.447	0	-0.447	-0.707	-0.894
ω	2.236	1.414	1.118	1	1.118	1.414	2.236

Instrument strength is varied in the experiments. Below you will find a table showing the relationship between the design parameter θ and more conventional measures of the fit provided by the reduced form equations. For each of the design points, the R^2 and overall F-statistic of regression significance were computed. The average values for each design are included in the table.

[1] $\sqrt{(1 + (\gamma + \lambda)^2)\sigma_v^2)}$ and $\lambda = 0$ and $\sigma_v^2 = 1$

One thing is obvious. The fit is not being held constant in the experiments. By using the same value of θ in each of the four sets of experiments, the R^2 and overall-F statistic of regression significance vary. In general, adding observations reduces R^2 and increases the overall F. Adding regressors (overidentification) reduces both. As will be seen, the resulting biases are reasonably controlled when the overall F statistic is above 10. This is consistent with the results of Stock and Yogo (2005).

	θ				
	0.05	0.1	0.25	0.5	1
	n=200; just identified				
R^2	.023	.048	.19	.47	.78
Overall-F	1.8	4.5	22.5	88	348
	n=1000; just identified				
R^2	.011	.032	.165	.44	.76
Overall-F	4.8	16.3	97	389	1558
	n=200; over identified				
R^2	.024	.037	.120	.50	.65
Overall-F	1.3	2.2	8.6	31	121
	n=1000; over identified				
R^2	.009	.022	.109	.32	.65
Overall-F	2.6	7.1	40	157	625

5. Results

Initial computations indicated that the proportion of 1's in the sample have no systematic effect on the magnitude of bias. This may be more important in other uses, e.g., sample selectivity models (see Hill et al. (2003)) and the results include below exclude these cases.

Below you will find eight tables. Table 1 includes bias for each design point based on 1000 Monte Carlo samples. It is broken into subtables a, b, c, and d, reflecting differences in sample size and overidentification of the model. Table 1a is based on samples of size 200 for a just identified model. Table 1b is just identified with 1000 observations. Tables 1c and 1d are for overidentified models, with c based on 200 observations and d on 1000.

The parameter labeled 'theta' controls the strength of the instruments. As θ increases, instruments become stronger. Rivers and Vuong (1988) implicitly used only $\theta = 1$, which implies very strong instruments.

Before presenting more detailed results, here is a summary of the main findings.

(1) When there is no endogeneity, OLS and Probit work well (as expected).
(2) It is clear that OLS and Probit should be avoided when you have an endogenous regressor.
(3) Weak instruments increase the bias of AGLS. The bias increases as the correlation between the endogenous regressor and the equation's error increases.
(4) The actual size of a parameter significance test based on the instrumental variable probit is reasonably close to the nominal level in nearly every instance. This is surprising for at least two reasons. 1) The bias of IVP is substantial when instruments are weak. 2) The test statistic is based on an inconsistent estimator of the standard error. No attempt was made to estimate the covariance of this estimator consistently, as is done in Limdep 9 Greene (2007). It is not often when one can use a biased estimator with inconsistent standard errors to construct a t-test that has desirable levels of type I error. This is the case here.
(5) The size of the significance tests based on the AGLS estimator is also reasonable, but the actual size is larger than the nominal size–a situation that gets worse as severity of the endogeneity problem increases. When instruments are very weak, the actual test rejects a true null hypothesis twice as often as it should.
(6) Linear instrumental variables estimators that use consistent estimators of standard errors can be used for this purpose (significance testing) though their performance is not quite up to that of the AGLS estimator. The Linear IV estimator performs better when the model is just identified.
(7) There is an improvement in bias and the size of the significance test when samples are larger. Mainly, smaller samples require stronger instruments in order for bias to be small and tests to work properly (other than IVP, which as mentioned above, works fairly well all the time).
(8) There is little to be gained by pretesting for endogeneity. When instruments are extremely weak it is outperformed by the other estimators considered, except when the no endogeneity hypothesis is true (and probit should be used). Bias is reduced by small amounts, but it is uncertain what one would use as an estimator of standard errors for a subsequent t-test.
(9) When instruments are weak, t-tests based on ML are no better than ones based on AGLS (in fact, one could argue that they are worse). Significance testing based on the ML estimator is much more reliable in large samples.

5.1. **Bias.** In table 1a the bias of each estimator is given for each of the design points considered. There is one endogenous variable and one instrument; the model is just identified. If there is any correlation between the regressor and the regression error, then weak instruments create considerable bias. It is actually not clear whether these instrumental variables estimators even have a mean in this case since subsequent simulations yielded quite different numerical results (though they were just as biased). When $\theta = .25$ the corresponding $F - statistic$ is 22, indicating that the instruments are reasonably strong. This is well beyond the oft recommended threshold of 10 Staiger and Stock (1997). Bias of the AGLS estimator is in the $\pm.5$ range. Once $\lambda = .5$ bias is actually quite small. Notice that the AGLS and 2SCML results are identical; this is as expected. The benefit of using AGLS in our case is that it yields a consistent estimator of its standard error. This is not the case for 2SCML. This will become apparent when looking at the actual sizes of the 10% t-tests.

In table 1b the sample size is increased to 1000. The main difference is that biases are smaller and the results for $\theta = .25$ are now quite good; the average value of the F-statistic is 97. IVP is erratic when instruments are weakest, and in this case the AGLS and 2SCML are better choices. When the instruments are very strong ($\theta >= .5$), all perform well in terms of bias.

In table 1c you will find the results for samples of size 200 for a model that has 2 instruments (overidentified). Overidentification appears to have reduced bias somewhat. Certainly, bias figures for $\theta = .25$ in samples of 200 are quite good. There is some small deviation between AGLS and 2SCML now. This is expected in an overidentified model. When instruments are only moderately strong, AGLS actually performs slightly better; the biases are less than .01 in absolute value except the case when $\lambda = -2$. For stronger instruments, like those considered by Rivers and Vuong, there is no difference.

Increasing the sample size to 1000 in the overidentified case (table 1d) improves things further. Only under severe correlation among errors does the bias of AGLS rise above .17 when instruments are very weak ($\theta = .1$).

The bottom line is, if your sample is small and instruments weak, don't expect very reliable estimates of the IV probit model's parameters. They are quite erratic (see Tables 3a and 3b for Monte Carlo standard errors) and the bias can be substantial. If instruments are strong and correlation low, then the two-step AGLS estimator performs about as well as can be expected and is a reasonable choice. This may account for it's inclusion as an option in Stata.

5.2. **Size.** In table 2a the actual size of a nominal 10% significance test on γ is measured. Again, there is one endogenous variable and one instrument; the model is just identified. The first thing to notice is that the actual size of the test based on the IVP estimator is very close to the nominal 0.1 level. This is a surprise for two reasons. Even though it is consistent, the estimator is seldom used. Limdep 9 is an exception. Second, no effort was made to estimate the standard errors consistently. Simply using the outcome of the Newton-Raphson algorithm in the second step is known to be inconsistent. Limdep uses a Murphy-Topel Murphy and Topel (1985) sandwich covariance estimator to obtain consistent estimates of standard error. Bootstrapping would be another way of obtaining consistent standard errors. Given the relatively good performance of this estimator, a more careful comparison of AGLS and Limdep's approach would be useful. It remains to be seen whether these alternatives could be used to further improve performance here.

The linear IV estimator also performs reasonably well in all circumstances. The largest size distortion is .03 ($\theta = .05$, $\lambda = 2$). In only three instances is the size larger than .11. This is very respectable. The AGLS estimator is not quite as good. When the correlation increases, the actual size of the test increases with it. The worst case is when the instruments are very weak. The 2SCML results can be safely disregarded since no effort was made to estimate its standard error consistently.[2] When the instruments are strong, the set of consistent tests perform quite well even with a relatively small sample of size 200.

In table 2b the sample is increased to 1000. Predictably the results improve for most cases. Unfortunately, the size distortion of the AGLS estimator seems to be getting worse as the correlation between the errors increases. At $\theta = 1$, $\lambda = -2$ a nominal 10% test is rejecting a true null hypothesis 14% of the time. This is not terrible, but I expected better.

In table 2c we examine the overidentified case using 200 observations. Overidentification is not improving things here at all. The magic IVP estimator is now experiencing some small size distortion when instruments are weak and the size distortion of the AGLS estimator is becoming quite large (.198) at some points. In table 2d the larger sample reduces the size distortion of AGLS, but it is still rejecting a true null hypothesis at higher rates than we'd like (.175). The t-test does not perform poorly, but its behavior is a little puzzling. Overidentification does not improve its performance and you are probably better off discarding excess instruments, even if they are relatively strong.

[2] I used the $-H^{-1}$ where H is the Hessian matrix evaluated at the parameter estimates (i.e., Newton-Raphson see Griffiths et al. (1987) for details.

In tables 3a and 3b the Monte Carlo standard errors of the estimated coefficient on the endogenous variable are given. In table 3a the estimator is based on a sample of 200; in table 3b the sample size is 1000. When instruments are weak, the variation in instrumental variables estimators is very large, especially when correlation between errors is zero (or very large). The former result is expected.

When the instruments are relatively strong ($\theta \geq .5$ for n=200 or $\theta \geq .25$ for n=1000) the variation is small and in most cases the biases of the IV estimators (tables 1a-1d) are not significantly different from zero. The erratic behavior of these estimators when instruments are weak should be apparent when instruments are weak, though.

5.3. **ML vs AGLS.** The last comparison is between ML (maximum likelihood) and Amemiya's GLS estimator. These are the two options available in Stata 10, which make them popular choices in applied work. A more thorough analysis of ML was not conducted because of computational difficulties.[3] There are many circumstances when the ML estimator refuses to converge and this makes analysis of its properties in Stata very difficult.

To get some idea of how these two estimators compare, I chose 4 designs and examined the summary statistics associated with the coefficient estimates and the p-values for the t-test. This was repeated for samples of 200 and 1000. The four designs consist of combinations of strong/weak instruments and high/low correlation among errors. Accordingly, the four combinations of $\lambda = .5, 2$ and $\theta = .1, 1$ were examined. The results for n=200 appear in table 4a. For the sample size of 1000, ML refused to converge for many designs when $\theta = .1$ so I strengthened the instruments in this case by letting $\theta = .25$; these results appear in table 4b.

Looking at table 4a, you'll notice that the AGLS estimator is very imprecise when instruments are weak and correlation low ($\theta = .1, \quad \lambda = .5$). The 1% estimate is -44.751 and the 99% is 48. Compare that to ML which ranged from -1.021 to 1.253. By this measure ML looks like a much better choice for estimating the parameters. The AGLS estimator is slightly more biased (negatively) due to the large (negative) skewness. Kurtosis of this (asymptotically) normal statistic is 255!

From a testing standpoint, however, ML doesn't look so good. The 5% and 10% p-values are essentially zero when the instruments are weak. The truth (no effect) is always rejected. On the other hand, those of AGLS are substantially greater than nominal values when endogeneity problem is not

[3]The MLE is badly behaved when instruments are weak. It is prone to no converge when the model's parameters are not strongly identified via the data.

as severe. As the endogeneity worsens the AGLS actually performs close to the desired level.

Increasing the correlation between errors ($\theta = .1, \quad \lambda = 2$) makes parameter estimation worse. The problem here though is that the t-ratios tend to be too large, by a substantial amount. The 10% test rejects only 1% of the time. The AGLS based test now has about the right size (10.5%).

When the instruments are stronger, the choice is much clearer. Both AGLS and ML perform similarly in testing; ML is still much more precise and we would expect tests based on it to be more powerful.

6. Example

I this section a brief example from Adkins et al. (2007). The main goal of that paper was to determine whether managerial incentives affect the us of foreign exchange derivatives by bank holding companies (BHC). There was some speculation that several of the variables in the model were endogenous. The dependent variable of interest is an indicator variable that takes the value 1 if the BHC uses foreign exchange derivative. The independent variables are as follows:

Ownership by Insiders. When managers have a higher ownership position in the bank, their incentives are more closely aligned with shareholders so they have an incentive to take risk to increase the value of the call option associated with equity ownership. This suggests that a higher ownership position by insiders (officers and directors) results in less hedging. The natural logarithm of the percentage of the total shares outstanding that are owned by officers and directors is used as the independent variable.

Ownership by Institutional Blockholders. Institutional blockholders have incentive to monitor the firm's management due to the large ownership stake they have in the firm (Shleifer and Vishny (1986)). Whidbee and Wohar (1999) argue that these investors will have imperfect information and will most likely be concerned about the bottom line performance of the firm. The natural logarithm of the percentage of the total shares outstanding that are owned by all institutional investors is included as an independent variable and predict that the sign will be positive, with respect to the likelihood of hedging.

CEO Compensation. CEO compensation also provides its own incentives with respect to risk management. In particular, compensation with more option-like features induces management to take on more risk to increase the value of the option (Smith and Blundell (1985); Tufano (1996)). Thus, higher options compensation for managers results in less hedging. Two measures of CEO compensation are used: 1) annual cash bonus and 2) value of option awards.

There is a possibility that CEO compensation is endogenous in that successful hedging activity could in turn lead to higher executive compensation. The instruments used for the compensation variables are based on the executive's human capital (age and experience), and the size and scope of the firm (number of employees, number of offices and subsidiaries). These are expected to be correlated with the CEOs compensation and be predetermined with respect to the BHCs foreign exchange hedging activities.

BHC Size. The natural logarithm of total assets is used to control for the size of the BHC.

Capital. The ratio of equity capital to total assets is included as a control variable. The variable for dividends paid measures the amount of earnings that are paid out to shareholders. The higher the variable, the lower the capital position of the BHC. The dividends paid variable is expected to have a sign opposite that of the leverage ratio.

Like the compensation variables, leverage should be endogenously determined. Firms that hedge can take on more debt and thus have higher leverage, other things equal.

Foreign Exchange Risk. A bank's use of currency derivatives should be related to its exposure to foreign exchange rate fluctuations. The ratio of interest income from foreign sources to total interest income measures foreign exchange exposure. Greater exposure, as represented by a larger proportion of income being derived from foreign sources, should be positively related to both the likelihood and extent of currency derivative use.

Profitability. The return on equity is included to represent the profitability of the BHCs. It is used as a control.

6.1. **Results.** In this section the results of estimation are reported. Table 5 contains some important results from the reduced form equations. Due to the endogeneity of leverage and the CEO compensation variables, instrumental variables estimation is used to estimate the probability equations.

Table 6 reports the coefficient estimates for the instrumental variable estimation of the probability that a BHC will use foreign exchange derivatives for hedging. The first column of results correspond to the AGLS estimator and the second column, ML.

In Table 5 summary results from the reduced form are presented. The columns contain p-values associated with the null hypothesis that the indicated instrument's coefficient is zero in each of the four reduced form equations. The instruments include number of employees, number of subsidiaries, number of offices, CEO's age–which proxies for his or her experience, the 12 month maturity mismatch, and the ratio of cash flows to total assets (CFA). The p-values associated with the other variables have been suppressed to conserve space.

Each of the instruments appears to be relevant in that each is significantly different from zero at the 10% (p-value < 0.1) in at least one equation; the number of employees, number of subsidiaries, and CEO age and CFA are significant in one equation; the number of offices, employees, subsidiaries are significant in two equations.

The overall strength of the instruments can be roughly gauged by looking at the overall fit of the equations. The R^2 in the leverage equation is the smallest (0.29), but is still high relative to the results of the Monte Carlo simulation. The instruments, other than the 12 month maturity mismatch, appear to be strong and we have no reason to expect poor performance from either estimator in terms of bias.

Given the simulations suggest discarding extra instruments, this would be recommended here. Which to drop, other than the mismatch variable is unclear. CFA, Age, and subsidiaries are all strongly correlated with leverage; office and employees with options; and, employees, subsidiaries, and offices with bonuses. The fit in the leverage equation is weakest, yet the p-values for each individual variable is relatively high. For illustrative purposes, we plow forward with the current specification.

Table 5. **Summary Results from Reduced-form Equations**. The table contains p-values for the instruments and R^2 for each reduced form regression. The data are taken from the Federal Reserve System's Consolidated Financial Statements for Bank Holding Companies (FR Y-9C), the *SNL Executive Compensation Review*, and the *SNL Quarterly Bank Digest*, compiled by SNL Securities.

	Reduced Form Equation		
	Leverage	Options	Bonus
Instruments	Coefficient P-values		
Number of Employees	0.182	0.000	0.000
Number of Subsidiaries	0.000	0.164	0.008
Number of Offices	0.248	0.000	0.000
CEO Age	0.026	0.764	0.572
12 Month Maturity Mismatch	0.353	0.280	0.575
CFA	0.000	0.826	0.368
R-Square	0.296	0.698	0.606

Table 6: **IV Probit Estimates of the Probability of Foreign-Exchange Derivatives Use By Large U.S. Bank Holding Companies (1996-2000)**. This table contains estimates for the probability of foreign-exchange derivative use by U.S. bank holding companies over the period of 1996-2000. To control for endogeneity with respect to compensation and leverage, we use an instrumental variable probit estimation procedure. The dependent variable in the probit estimations (i.e., probability of use) is coded as 1 if the bank reports the use of foreign-exchange derivatives for purposes other than trading. The data are taken from the Federal Reserve System's Consolidated Financial Statements for Bank Holding Companies (FR Y-9C), the *SNL Executive Compensation Review*, and the *SNL Quarterly Bank Digest*, compiled by SNL Securities. Approximate p-values based on the asymptotic distribution of the estimators are reported in parentheses beneath the parameter estimates.

	Instrumental Variables Probit	
	AGLS	ML
Leverage	21.775	12.490
	(0.104)	(0.021)
Option Awards	-8.79E-08	-5.11E-08
	(0.098)	(0.002)
Bonus	1.76E-06	1.02E-06
	(0.048)	(<0.001)

Continued from preceding page		
	Instrumental Variables Probit	
	AGLS	ML
Total Assets	0.365	0.190
	(0.032)	(0.183)
Insider Ownership %	0.259	0.145
	(0.026)	(0.016)
Institutional Ownership %	0.370	0.201
	(0.006)	(0.041)
Return on Equity	-0.034	-0.020
	(0.230)	(0.083)
Market-to-Book ratio	-0.002	-0.001
	(0.132)	(0.098)
Foreign to Total Interest Income Ratio	-3.547	-2.177
	(0.356)	(0.127)
Derivative Dealer Activity Dummy	-0.280	-0.154
	(0.257)	(0.288)
Dividends Paid	-8.43E-07	-4.84E-07
	(0.134)	(0.044)
D=1 if 1997	-0.024	-0.016
	(0.930)	(0.914)
D=1 if 1998	-0.244	-0.133
	(0.352)	(0.383)
D=1 if 1999	-0.242	-0.134
	(0.391)	(0.395)
D=1 if 2000	-0.128	-0.065
	(0.643)	(0.685)
Constant	-9.673	-5.188
	(<0.001)	(4.40E-02)
Sample size	794	794

In light of the results from the Monte Carlo the significance tests based on ML may be misleading in this instance. The results correspond the closest to those in Table 2d. The model is overidentified, sample is large (700+), and the instruments are very strong ($\theta = .5$ or $\theta = 1$). Leverage is significant in ML at the 10% level, but not with AGLS. Similarly, return-on-equity, market-to-book, and dividends paid are all significant in the ML regression but not AGLS. This divergence of results is a little troubling. In the simulations, ML p-values were too small when instruments were mildly strong and correlation low. If the endogeneity problem is not severe, then the ML estimation and AGLS results tended to diverge. If this is the case, then AGLS estimates may be more reliable for testing purposes. In the case of very strong instruments, the AGLS estimator tended to be insignificant too often. In this example, we fell right in the middle and no strong recommendation can be made.

7. Conclusion

The bottom line is this: if you are stuck with weak instruments, and your goal is to test the significance of a variable in an endogenous probit model, be careful. None of the estimators considered do this very well, but a small nod goes to IVP and AGLS. When instruments are strong, by all means use the more efficient ML estimator, especially if it is computationally feasible. Also, overidentification should be avoided if possible. It doesn't appear to help the performance of the IV estimators either in terms of bias or testing. Certainly, a more thorough examination of this would need to be made before passing final judgement.

As final note, Limdep's approach deserves some study. Limdep uses the IVP estimator along with consistent estimators of the standard errors using Murphy-Topel approach. This is easy to implement in other software and a comparison of this to the AGLS and ML estimator looks like a good way to proceed.

8. Gauss Code

```
retcode=0;
nn=200;
nmc=1000;
state = 1890780797;
state = 189078079741;
state1=890809;
qlevel = { .025, .975 };

/* Generate X for the reduced form equation */
t=100;
cov=.5*ones(3,3);
cov[1,1]=1;
cov[2,2]=1;
cov[3,3]=1;
x=rndn(nn,3)*chol(cov);
xm=ones(nn,1)~x;
pie = {0,1,1,-1};
```

```
/* Coefficients for the RF  */
pies = {0,1,1,-1};

/*  X = regressors
**  W = Probit regressors
**  Z = instruments
**  y = regression dep var
**  v = probit dep var
*/

/* Create some storage matricies */
bmat=zeros(nmc,8);
smat1=zeros(nmc,7);        /* storage for probit IV */
tmat0=zeros(nmc,7);
tmat1=zeros(nmc,7);
otherstat=zeros(nmc,4);
xbmat=zeros(1,12);
xbiasmat=zeros(1,12);
xsmat=zeros(1,11);
t0mat=zeros(1,11);
t1mat=zeros(1,11);
mcsemat=zeros(1,12);
othermat = zeros(1,4);
crlmat=zeros(1,11);
crumat=zeros(1,11);

/* First Loop:
** Correlation between u and v
*/
/* Instrument strength  */

rhoxiter = 1;
do while (rhoxiter <= 5);
/* Instrument correlation */
if rhoxiter == 1; theta=.05;  endif;
if rhoxiter == 2; theta=.1;  endif;
if rhoxiter == 3; theta=.25;  endif;
if rhoxiter == 4; theta=.5;  endif;
if rhoxiter == 5; theta=1;  endif;

rhoiter = 1;
do while (rhoiter <= 7);

if rhoiter == 1; lambda=2; endif;
if rhoiter == 2; lambda=1; endif;
```

```
if rhoiter == 3; lambda=.5; endif;
if rhoiter == 4; lambda=0; endif;
if rhoiter == 5; lambda=-.5; endif;
if rhoiter == 6; lambda=-1; endif;
if rhoiter == 7; lambda=-2; endif;

/* Proportion */
giter = 2;
do while (giter <= 2);
/* Conrols censoring */
if giter == 1; gamm = { -1 , 1 }; endif;
if giter == 2; gamm = { 0 , 1 };    endif;
if giter == 3; gamm = { 1 , 1 };  endif;

/* stuff to make the rf computation faster */
/*  Here is where I overidentify if needed */
x13=xm[.,1:3];
xtxinv=invpd(x13'*x13);
px=x13*xtxinv*x13';
mx=eye(nn)-px;

x1=xm[.,1];
p1=x1*x1'./rows(x1);
m1=eye(nn)-p1;

x12=xm[.,1:2];

/* Sample size, k, degrees of freedom */
k=cols(x12)+1;
df=rows(x12)-k;

omeg = sqrt(1+(1+lambda)^2);

/* Begin Monte Carlo */
ii=1; o1=0;
do while ii le nmc;
over:
if retcode == 1; o1=o1+1; endif;
retcode=0;

/* Generate Errors  */
{errors,state} = rndKMn(nn,2,state);
v = errors[.,1];
u = lambda * v + errors[.,2] ;
```

```
/* Generate endogenous variable */
y2 = theta.*(x13*pies[1:3])+v;

/* Generate latent variable */
y1star = x12*gamm+u;
y1 = (y1star .>= 0);                              /* censoring vector;*/
n = sumc(y1);
p = meanc(y1);
mfx  = pdfn(cdfni(p));
/* Generate variables for Probit */
w=y2~x12;
wtwinv=invpd(w'w);

/* Probit */
{gam1,vgam,retcode}=nr(w,y1);
if retcode ne 0; goto over; endif;
bmat[ii,2]=gam1[1];
se1=(sqrt(diag(vgam)))';
smat1[ii,2]=(se1[1]);
tmat0[ii,2]=gam1[1]/se1[1];

/* some stats */
R2=y2'*px*y2/y2'y2;
sseu=y2'*mx*y2;
sser=y2'*m1*y2;
F=(sser-sseu)*(nn-cols(x13))/((sseu)*(cols(x13)-1));

/* IV Probit */
what=px*w;

{gam2,vgam2,retcode}=nr(what,y1);
if retcode ne 0; goto over; endif;
bmat[ii,3]=gam2[1]*omeg;
se2=(sqrt(diag(vgam2)))';
smat1[ii,3]=se2[1];
tmat0[ii,3]=gam2[1]/se2[1];

/* 2SIV Probit
{gam3,vgam3,retcode}=nr(what~mx*y2,y1);
if retcode ne 0; goto over; endif;
bmat[ii,5]=gam3[1];
se3=(sqrt(diag(vgam3)))';
smat1[ii,5]=se3[1];
tmat0[ii,5]=gam3[1]/se3[1];
*/
```

```
/* 2SCML Probit */
/* eq. (3.5)  */
{gam7,vgam7,retcode}=nr(w~mx*y2,y1);
bmat[ii,7]=gam7[1];
se7=(sqrt(diag(vgam7)))';
smat1[ii,7]=se7[1];
tmat0[ii,7]=gam7[1]/se7[1];

/* Iwata VI 1 */ ;
invwpzw=invpd(w'px*w);
gam4 = invwpzw*w'*px*y1;
bmat[ii,4]=gam4[1]./mfx;
eivhat = y1-w*gam4;
cov = invwpzw*w'*px*diagrv(eye(rows(eivhat)),eivhat.^2)*px*w*invwpzw;
se4=(sqrt(diag(cov)))';
smat1[ii,4]=se4[1];
tmat0[ii,4]=gam4[1]/se4[1];

/* AGLS  */
{gam5,segam5,gam3a} = agls(y1,w,x13,1);
bmat[ii,6]=gam5[1];
smat1[ii,6]=segam5[1];
tmat0[ii,6]=gam5[1]/segam5[1];


/* Pretest  */
k7=rows(gam7);
tv = gam7[k7]/se7[k7];
if abs(tv) > 1.645;
gam8=gam7[1];
else;
gam8=gam1[1];
endif;
bmat[ii,8]=gam8;
/* Estimate the Probit model using LS */
b=wtwinv*w'y1;
ehat=y1-w*b;
sig2=ehat'ehat/df;
covb=sig2*wtwinv;
seb=sqrt(diag(covb))';
smat1[ii,1]=seb[1];
tmat0[ii,1]=(b[1])/seb[1];


/* Save LS and the tratio */
```

```
bmat[ii,1]=b[1]./mfx;

/* Save other statistics */
otherstat[ii,1]=n;
otherstat[ii,2]=p;
otherstat[ii,3]=R2;
otherstat[ii,4]=F;

ii=ii+1;
endo;
"over " o1;
/* Bias of LS and IV estimators */
meanb=meanc(bmat);
means=meanc(smat1);
meant0=meanc(tmat0);

/* Critical Values */
xc = msort(tmat1);
cl = critl(xc,.05);
cu = critu(xc,.05);

/*  */
xb=gamm[1]~theta~lambda~omeg~meanb';
xs=gamm[1]~theta~lambda~omeg~means';
v0 = abs(tmat0) .> 1.645;
t0=gamm[1]~theta~lambda~omeg~meanc(v0)';
MCse=gamm[1]~theta~lambda~omeg~stdc(bmat)'./sqrt(nmc);
crl=gamm[1]~theta~lambda~omeg~cl;
cru=gamm[1]~theta~lambda~omeg~cu;

xbmat=xbmat|xb;
semat=xsmat|xs;
t0mat=t0mat|t0;
mcsemat=mcsemat|mcse;
othermat=othermat|(meanc(otherstat))';
crlmat=crlmat|crl;
crumat=crumat|cru;


giter = giter + 1;
endo;
rhoiter = rhoiter + 1;
endo;
rhoxiter = rhoxiter + 1;
endo;
```

```
format /m1 /rd 16,3;
"      Censoring         theta          lambda          Omega             ols

"coeffs ";
xbmat;

"      Censoring         theta          lambda          Omega             ols

"SE ";
semat;

"      Censoring         theta          lambda          Omega             ols

"size ";
t0mat;

"      Censoring         theta          lambda          Omega             ols

"MC SE ";
mcsemat;

"      Censoring         theta          lambda          Omega             ols

"crl and cru ";
crlmat;
crumat;

"Some Stats";
t0mat[.,1:4]~Othermat;


proc errn(xm,corrmat);
local cd, nn, xx, xe, er;
cd = chol(corrmat);
nn=rows(xm);
{xx,state} = rndKMn(nn,1,state);
xx = xm~xx;
xe=xx*cd;
xe=xe[.,4];
retp(xe);
endp;

proc errn2(xm,corrmat);
local cd, nn, xx, xe, er;
```

```
cd = chol(corrmat);
nn=rows(xm);
{xx,state} = rndKMn(nn,2,state);
xx = xm~xx;
xe=xx*cd;
xe=xe[.,5:6];
retp(xe);
endp;

proc(3) = agls(y,x,z,c);
local pz, mz, b, xhat, rho, zhat, alpha, xs, gam2siv, vgamsiv, retcode,
      t1, xstar, lam, pihat, s2, v2, omega, jinv, k, d, delt, cov, se, J,vhat;
/* v is 0,1
** z are instruments
** x is the endog regressor (w in the probit model)
*/
/* Step 1:  ls reduced form */
k=cols(z);
pz = z*invpd(z'z)*z';
mz = eye(rows(z))-pz;
d=invpd(z'z)*z'x;
xhat=pz*x;
vhat = mz*x[.,c];

/* Step 2:  probit rf */
{alpha,Jinv,retcode}=nr(z~vhat,y);
jinv=jinv[1:k,1:k];
lam = alpha[k+1];
alpha = alpha[1:k,.];


/* Step 3:  2siv  */
xs = xhat~vhat;
{gam2siv,vgamsiv,retcode}=nr(xs,y);
rho = lam-gam2siv[2];

/* Step 4:  v2 inv(x'x/n)  */
xstar = x[.,c]*rho;
pihat = invpd(z'z)*z'*xstar;
s2=(xstar-z*pihat)'(xstar-z*pihat)/(rows(z)-cols(z));
v2=s2*invpd(z'z);
@ other[ii,.]=rho~s2; @

/* Step 5  Omega */
```

```
omega = jinv + v2;

/* Step 6 AGLS */
delt =invpd(d'*invpd(omega)*d)*d'*invpd(omega)*alpha;
cov=invpd(d'invpd(omega)*d);
se=sqrt(diag(cov));
retp(delt,se,gam2siv);
endp;

/*  This computes the observation by observation derivatives.
**  Column summation yields the function below.
*/

proc (1) = Zgrad(beta,x,y);
local zhat, pdf, cdf, del;
zhat = x*beta;
        pdf = pdfn(zhat);
        cdf = cdfn(zhat);
        del=(y-cdf).*pdf./(cdf.*(1-cdf));
        del = (del.*.ones(1,cols(x)).*x);
        retp(del);
 endp;

/*  Calculate gradient */

fn grad(beta,x,y)= sumc(zgrad(beta,x,y));

/*  Calculate Hessian     */

  proc (1) = hessi(beta,x,y);
  local k, zhat, pdf, cdf, d, H;
        k = cols(x);
        zhat = x*beta;
        pdf = pdfn(zhat);
        cdf = cdfn(zhat);
        d = y.*((pdf+zhat.*cdf)/cdf.^2) + (1-y).*((pdf - zhat.*(1-cdf))/(1-cdf).^2);
        d = pdf.*d;
        H = -x'*((ones(1,k).*.d).*x);
        retp(H);
  endp;

/*  Calculate Outer Product of the Gradients matrix (BHHH)   */

  proc (1) = opg(beta,x,y);
  local k, zhat, pdf, cdf, d, H;
```

```
        h=zgrad(beta,x,y)'*zgrad(beta,x,y);
        retp(H);
  endp;

proc (3) = NR(x,y);
local jj, b, retcode, vgam ;
retcode=0;
jj=1;
b=invpd(x'x)*x'y;
do while abs(maxc(grad(b,x,y))) > .00001 ;
b=b+invpd(-hessi(b,x,y))*grad(b,x,y);
if jj ge 100;  b=0; vgam=eye(cols(x)); retcode=1; goto bomb; endif;
jj=jj+1;
endo;

vgam=invpd(-hessi(b,x,y));
bomb:
retp(b,vgam,retcode);
endp;

/*  Critl returns the lower pc percentile of the matrix, MAT */

 proc (1) = critl(mat,pc);
 local n,row,mm;
        N=ROWs(MAT);
        ROW=ceil(PC*n);
        MM=mat[ROW,.];
        retp(mm);
 endp;

/*  Critu returns the upper pc percentile of the matrix, MAT */

 proc (1) = critu(mat,pc);
 local n, row, mm;
        N=ROWs(MAT);
        ROW=ceil((1-PC)*n)+1;
        MM=mat[ROW,.];
        retp(mm);
 endp;

 proc (1) = msort(mat);
 local n,m,boot,b,s,ii;
        n=rows(mat);
        m=cols(mat);
        boot = zeros(n,m);
```

```
        ii = 1;
        do while ii le  m;
        b = mat[.,ii];
        s = sortc(b,1);
        boot[.,ii]=s;
        ii = ii + 1;
        endo;
        retp(boot);
  endp;
```

9. Stata Code

```
clear
set seed 18079741
set more off
set mem 100m
set obs 1000
set matsize 1000
local B = 1000
matrix Bvals = J(‘B’, 2, 0)
matrix pvals = J(‘B’, 2, 0)
matrix pwvals1 = J(‘B’, 6, 0)
matrix pwvals2 = J(‘B’, 6, 0)

matrix covx = J(3,3,.5)
matrix covx[1,1]=1
matrix covx[2,2]=1
matrix covx[3,3]=1
corr2data x1 x2 x3, corr(covx) cstorage(full)

mkmat x1 x2 x3

forvalues b = 1/‘B’ {
 drop _all
 quietly set obs 1000
 gen cons = 1
 gen e = invnormal(uniform())
 gen v = invnormal(uniform())
 gen u = .5*v + e

 svmat x1
 svmat x2
 svmat x3
```

```
 gen y2 = .1*(x1 + x2 - x3) + v
 gen y1 = x1 + u
 gen y = 0
 qui replace y =1 if y1>0

 qui ivprobit y (y2=x2) x1, twostep
 matrix betas1 = e(b)
 matrix Bvals[‘b’,1] = betas1[1,1]

 qui testparm y2
 matrix pvals[‘b’,1] = r(p)

 qui ivprobit y (y2=x2) x1, difficult
 matrix betas2 = e(b)
 matrix Bvals[‘b’,2] = betas2[1,1]

 qui testparm y2
 matrix pvals[‘b’,2] = r(p)
 }
drop _all
svmat Bvals
svmat pvals
summ *, det
```

References

Adkins, Lee C., David A. Carter and W. Gary Simpson (2007), ‘Managerial incentives and the use of foreign-exchange derivatives by banks’, *Journal of Financial Research* **15**, 399–413.

Arendt, Jacob N. (2001), Endogeneity and heterogeneity in ldv panel data models, Technical report, Institute of Local Government Studies, Nyropsgade 37, 1602 Copenhagen V, Denmark.

Arendt, Jacob Nielsen and Anders Holm (2006), Probit models with binary endogenous regressors, Discussion Papers on Business and Economics 4/2006, Department of Business and Economics Faculty of Social Sciences University of Southern Denmark.

Blundell, Richard W. and James L. Powell (2004), ‘Endogeneity in semiparametric binary response models’, *Review of Economic Studies* **71**, 655–679. available at http://ideas.repec.org/a/bla/restud/v71y2004ip655-679.html.

Davidson, Russell and James G. MacKinnon (2004), *Econometric Theory and Methods*, Oxford University Press, Inc., New York.

Greene, William H. (2007), *LIMDEP Version 9.0 Econometric Modeling Guide, Volume 1*, Econometrics Software, Inc., 15 Gloria Place, Plainview, NY.

Griffiths, W. E., R. C. Hill and P. Pope (1987), 'Small sample properties of probit model estimators', *Journal of the American Statistical Association* **82**, 929–937.

Hill, R. Carter, Lee C. Adkins and Keith Bender (2003), Test statistics and critical values in selectivity models, *in* R. C.Hill and T.Fomby, eds, 'Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later', Vol. 17 of *Advances in Econometrics*, Elsevier Science, pp. 75–105.

Iwata, Shigeru (2001), 'Recentered and rescaled instrumental variable estimation of tobit and probit models with errors in variables', *Econometric Reviews* **24**(3), 319–335.

Kan, Kamhon and Chihwa Kao (2005), Simulation-based two-step estimation with endogenous regressors, Center for Policy Research Working Papers 76, Center for Policy Research, Maxwell School, Syracuse University. available at http://ideas.repec.org/p/max/cprwps/76.html.

Murphy, Kevin M. and Robert H. Topel (1985), 'Estimation and inference in two-step econometric models', *Journal of Business and Economic Statistics* **3**(4), 370–379.

Nawata, Kazumitsu and Nobuko Nagase (1996), 'Estimation of sample selection bias models', *Econometric Reviews* **15**(4), 387–400.

Newey, Whitney (1987), 'Efficient estimation of limited dependent variable models with endogenous explanatory variables', *Journal of Econometrics* **36**, 231–250.

Nicoletti, Cheti and Franco Peracchi (2001), Two-step estimation of binary response models with sample selection, Technical report, Faculty of Economics, Tor Vergata University, I-00133 Rome, Italy. Please do not quote.

Rivers, D. and Q. H. Vuong (1988), 'Limited information estimators and exogeneity tests for simultaneous probit models', *Journal of Econometrics* **39**(3), 347–366.

Shleifer, A. and R. W. Vishny (1986), 'Large shareholders and corporate control', *Journal of Political Economy* **94**, 461–488.

Smith, Richard J. and Richard W. Blundell (1985), 'An exogeneity test for a simultaneous equation tobit model with an application to labor supply', *Econometrica* **54**(3), 679–685.

Staiger, D. and J. H. Stock (1997), 'Instrumental variables regression with weak instruments', *Econometrica* **65**, 557586.

Stock, J. H. and M. Yogo (2005), Testing for weak instruments in linear iv regression, *in* D. W. K.Andrews and J. H.Stock, eds, 'Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg', Cambridge University Press, Cambridge, pp. 80–108.

Tufano, P. (1996), 'Who manages risk? an empirical examination of risk management practices in the gold mining industry', *Journal of Finance* **51**, 1097–1137.

Whidbee, D. A. and M. Wohar (1999), 'Derivative activities and managerial incentives in the banking industry', *Journal of Corporate Finance* **5**, 251–276.

Yatchew, Adonis and Zvi Griliches (1985), 'Specification error in probit models', *The Review of Economics and Statistics* **67**(1), 134–139.

Zuehlke, Thomas W. and Allen R. Zeman (1991), 'A comparison of two-stage estimators of censored regression models', *Review of Economics and Statistics* **73**(1), 185–188.

Lee C. Adkins, Professor of Economics, 303 College of Business Administration, Oklahoma State University, Stillwater OK 74078

E-mail address: lee.adkins@okstate.edu

Table 1a Bias of each estimator based on samples of size 200. Monte Carlo used 1000 samples. The model is just identified. The approximate proportion of 1's in each sample is .5.

Design		Estimator						
θ	λ	ols	probit	IV probit	Linear IV	agls	tscml	pretest
0.05	2	0.818	2.103	-6.807	-1.533	-1.858	-1.858	0.699
0.05	1	0.575	1.034	2.934	1.005	1.572	1.572	1.082
0.05	0.5	0.326	0.510	-6.885	-3.057	-3.717	-3.717	-0.600
0.05	0	0.004	0.006	-12.681	-7.284	-8.732	-8.732	0.105
0.05	-0.5	-0.330	-0.515	-5.085	-2.915	-4.721	-4.721	-0.210
0.05	-1	-0.573	-1.028	-0.853	-0.834	-0.302	-0.302	-0.700
0.05	-2	-0.817	-2.078	-1.478	-0.972	-2.429	-2.429	-1.980
0.1	2	0.813	2.043	22.393	6.184	7.702	7.702	8.046
0.1	1	0.572	1.023	3.000	0.041	-0.423	-0.423	0.446
0.1	0.5	0.324	0.509	1.580	0.473	0.960	0.960	0.628
0.1	0	-0.001	-0.001	12.316	6.766	8.767	8.767	0.007
0.1	-0.5	-0.328	-0.510	-0.196	-0.182	-0.405	-0.405	-0.324
0.1	-1	-0.570	-1.020	0.251	0.095	0.221	0.221	-0.217
0.1	-2	-0.813	-2.037	-0.069	-0.052	-0.285	-0.285	-1.023
0.25	2	0.785	1.848	-0.625	-0.188	-0.508	-0.508	-0.482
0.25	1	0.547	0.966	-0.286	-0.137	-0.199	-0.199	-0.010
0.25	0.5	0.312	0.488	-0.127	-0.104	-0.075	-0.075	0.189
0.25	0	-0.005	-0.004	0.027	-0.057	0.018	0.018	-0.016
0.25	-0.5	-0.317	-0.487	0.150	0.040	0.143	0.143	-0.111
0.25	-1	-0.550	-0.965	0.183	0.111	0.273	0.273	0.049
0.25	-2	-0.782	-1.840	0.288	0.175	0.456	0.456	0.400
0.5	2	0.694	1.390	-0.086	-0.030	-0.053	-0.053	-0.053
0.5	1	0.485	0.809	-0.065	-0.039	-0.040	-0.040	-0.031
0.5	0.5	0.274	0.425	-0.045	-0.041	-0.029	-0.029	0.055
0.5	0	-0.005	-0.002	-0.005	-0.031	-0.004	-0.004	-0.006
0.5	-0.5	-0.283	-0.427	0.014	-0.014	0.013	0.013	-0.070
0.5	-1	-0.487	-0.807	0.036	0.015	0.049	0.049	0.040
0.5	-2	-0.696	-1.385	0.030	0.013	0.056	0.056	0.056
1	2	0.478	0.738	0.005	-0.001	0.004	0.004	0.004
1	1	0.335	0.505	-0.003	-0.008	-0.002	-0.002	-0.002
1	0.5	0.186	0.280	0.001	-0.011	0.001	0.001	0.010
1	0	-0.004	0.002	0.009	-0.010	0.006	0.006	0.004
1	-0.5	-0.198	-0.285	0.007	-0.006	0.007	0.007	-0.001
1	-1	-0.338	-0.498	0.011	0.001	0.016	0.016	0.016
1	-2	-0.480	-0.730	0.014	0.006	0.028	0.028	0.028

Table 1b Bias of each estimator based on samples of size 1000. Monte Carlo used 1000 samples.
The model is just identified. The approximate proportion of 1's in each sample is .5.

Design		Estimator						
θ	λ	ols	probit	IV probit	Linear IV	agls	tscml	pretest
0.05	2	0.811	2.008	1.397	0.382	0.551	0.551	0.551
0.05	1	0.572	1.008	0.474	0.089	0.212	0.212	0.212
0.05	0.5	0.327	0.501	-0.158	-0.056	-0.310	-0.310	-0.310
0.05	0	0.000	0.000	1.266	0.204	0.895	0.895	0.895
0.05	-0.5	-0.328	-0.501	-1.216	-0.770	-1.386	-1.386	-1.386
0.05	-1	-0.569	-1.001	-10.904	-7.669	-14.615	-14.615	-14.615
0.05	-2	-0.811	-2.011	-1.135	-0.761	-1.850	-1.850	-1.850
0.1	2	0.808	1.982	-0.229	-0.087	-0.196	-0.196	-0.196
0.1	1	0.568	0.997	-3.672	-1.381	-1.869	-1.869	-1.869
0.1	0.5	0.326	0.499	-0.923	-0.448	-0.549	-0.549	-0.549
0.1	0	-0.002	-0.002	-0.092	-0.112	-0.065	-0.065	-0.065
0.1	-0.5	-0.328	-0.501	-0.072	-0.075	-0.095	-0.095	-0.095
0.1	-1	-0.567	-0.993	0.136	0.072	0.184	0.184	0.184
0.1	-2	-0.809	-1.981	-0.208	-0.137	-0.227	-0.227	-0.227
0.25	2	0.778	1.782	-0.040	-0.017	-0.029	-0.029	-0.029
0.25	1	0.547	0.946	-0.023	-0.022	-0.017	-0.017	-0.017
0.25	0.5	0.314	0.481	-0.026	-0.030	-0.016	-0.016	-0.016
0.25	0	-0.002	-0.001	0.001	-0.021	0.001	0.001	0.001
0.25	-0.5	-0.316	-0.481	0.023	-0.004	0.023	0.023	0.023
0.25	-1	-0.547	-0.944	0.015	-0.001	0.021	0.021	0.021
0.25	-2	-0.779	-1.779	0.039	0.019	0.058	0.058	0.058
0.5	2	0.690	1.352	0.003	-0.002	0.002	0.002	0.002
0.5	1	0.484	0.795	-0.002	-0.007	0.000	0.000	0.000
0.5	0.5	0.278	0.418	-0.001	-0.010	-0.001	-0.001	-0.001
0.5	0	-0.002	0.000	-0.003	-0.012	-0.002	-0.002	-0.002
0.5	-0.5	-0.279	-0.417	0.005	-0.005	0.005	0.005	0.005
0.5	-1	-0.486	-0.796	-0.003	-0.009	-0.003	-0.003	-0.003
0.5	-2	-0.689	-1.344	0.010	0.004	0.014	0.014	0.014
1	2	0.474	0.719	-0.002	-0.002	-0.004	-0.004	-0.004
1	1	0.331	0.491	-0.002	-0.004	0.000	0.000	0.000
1	0.5	0.190	0.279	-0.002	-0.005	-0.001	-0.001	-0.001
1	0	-0.001	0.002	0.004	-0.004	0.003	0.003	0.003
1	-0.5	-0.193	-0.277	0.000	-0.005	0.000	0.000	0.000
1	-1	-0.334	-0.492	0.002	-0.002	0.003	0.003	0.003
1	-2	-0.475	-0.721	0.000	-0.002	-0.001	-0.001	-0.001

Table 1c Bias of each estimator based on samples of size 200. Monte Carlo used 1000 samples.
The model is overidentified. The approximate proportion of 1's in each sample is .5.

Design		Estimator						
θ	λ	ols	probit	IV probit	Linear IV	agls	tscml	pretest
0.050	2.000	0.830	2.078	2.376	0.668	1.707	1.692	1.789
0.050	1.000	0.592	1.030	0.989	0.302	0.642	0.650	0.803
0.050	0.500	0.342	0.515	0.613	0.222	0.353	0.352	0.388
0.050	0.000	-0.002	-0.003	0.039	-0.023	0.027	0.029	-0.008
0.050	-0.500	-0.342	-0.511	-0.428	-0.322	-0.431	-0.434	-0.484
0.050	-1.000	-0.591	-1.033	-0.525	-0.427	-0.776	-0.767	-0.787
0.050	-2.000	-0.828	-2.072	-0.996	-0.649	-1.701	-1.694	-1.931
0.100	2.000	0.823	2.047	1.227	0.333	0.946	0.938	1.164
0.100	1.000	0.587	1.018	0.598	0.176	0.374	0.374	0.564
0.100	0.500	0.339	0.508	0.287	0.069	0.163	0.163	0.316
0.100	0.000	0.000	0.001	-0.015	-0.073	-0.010	-0.011	-0.034
0.100	-0.500	-0.340	-0.504	-0.167	-0.161	-0.155	-0.156	-0.376
0.100	-1.000	-0.587	-1.016	-0.255	-0.222	-0.396	-0.395	-0.683
0.100	-2.000	-0.823	-2.034	-0.456	-0.315	-0.755	-0.740	-0.951
0.250	2.000	0.781	1.762	0.007	-0.007	-0.006	-0.008	0.003
0.250	1.000	0.557	0.951	0.008	-0.018	0.007	0.007	0.128
0.250	0.500	0.321	0.480	0.009	-0.030	0.003	0.004	0.173
0.250	0.000	-0.003	0.000	0.010	-0.036	0.006	0.007	-0.004
0.250	-0.500	-0.325	-0.482	-0.008	-0.038	-0.010	-0.010	-0.190
0.250	-1.000	-0.559	-0.944	0.005	-0.020	0.008	0.009	-0.120
0.250	-2.000	-0.780	-1.768	0.038	0.015	0.039	0.041	0.032
0.500	2.000	0.666	1.240	0.000	-0.004	-0.002	-0.004	-0.004
0.500	1.000	0.471	0.752	-0.003	-0.013	-0.003	-0.003	0.000
0.500	0.500	0.269	0.400	-0.005	-0.019	-0.005	-0.004	0.056
0.500	0.000	-0.005	0.000	-0.004	-0.022	-0.004	-0.003	0.002
0.500	-0.500	-0.281	-0.410	-0.007	-0.023	-0.010	-0.009	-0.072
0.500	-1.000	-0.478	-0.759	0.010	-0.004	0.017	0.017	0.014
0.500	-2.000	-0.664	-1.239	0.010	0.001	0.009	0.009	0.009
1.000	2.000	0.414	0.592	0.002	-0.002	-0.001	-0.001	-0.001
1.000	1.000	0.293	0.421	0.000	-0.006	-0.002	-0.002	-0.002
1.000	0.500	0.168	0.245	-0.001	-0.009	-0.001	-0.001	0.003
1.000	0.000	-0.006	-0.002	-0.002	-0.011	-0.002	-0.002	-0.002
1.000	-0.500	-0.177	-0.246	0.001	-0.008	0.001	0.001	-0.003
1.000	-1.000	-0.301	-0.431	-0.007	-0.011	-0.011	-0.011	-0.011
1.000	-2.000	-0.417	-0.601	0.000	-0.002	0.003	0.003	0.003

Table 1d Bias of each estimator based on samples of size 1000. Monte Carlo used 1000 samples.
The model is overidentified. The approximate proportion of 1's in each sample is .5.

Design		Estimator						
θ	λ	ols	probit	IV probit	Linear IV	agls	tscml	pretest
0.05	2	0.817	2.007	0.873	0.276	0.649	0.650	0.953
0.05	1	0.578	1.005	0.415	0.220	0.274	0.275	0.515
0.05	0.5	0.333	0.500	0.214	0.172	0.116	0.117	0.327
0.05	0	0.000	0.000	-0.077	0.073	-0.054	-0.054	0.005
0.05	-0.5	-0.333	-0.502	-0.086	0.044	-0.088	-0.088	-0.255
0.05	-1	-0.578	-1.003	-0.282	-0.171	-0.400	-0.401	-0.684
0.05	-2	-0.815	-2.002	-0.413	-0.243	-0.694	-0.695	-0.930
0.1	2	0.811	1.966	0.270	0.094	0.171	0.171	0.208
0.1	1	0.574	0.994	0.028	0.059	0.009	0.010	0.211
0.1	0.5	0.332	0.499	-0.019	0.062	-0.007	-0.007	0.216
0.1	0	0.001	-0.001	-0.006	0.080	-0.004	-0.004	-0.007
0.1	-0.5	-0.329	-0.496	0.016	0.079	0.023	0.023	-0.198
0.1	-1	-0.572	-0.990	-0.001	0.045	0.006	0.005	-0.171
0.1	-2	-0.811	-1.968	0.041	0.044	0.075	0.074	0.040
0.25	2	0.775	1.739	0.008	0.009	0.009	0.010	0.010
0.25	1	0.548	0.927	-0.033	0.007	-0.018	-0.018	-0.017
0.25	0.5	0.319	0.476	-0.008	0.025	-0.005	-0.005	0.035
0.25	0	0.000	-0.002	0.000	0.034	0.000	0.000	0.001
0.25	-0.5	-0.315	-0.473	-0.001	0.027	-0.001	-0.001	-0.044
0.25	-1	-0.546	-0.928	-0.001	0.018	-0.001	-0.001	-0.001
0.25	-2	-0.774	-1.730	0.002	0.008	0.002	0.002	0.002
0.5	2	0.667	1.248	0.015	0.008	0.011	0.011	0.011
0.5	1	0.473	0.753	0.000	0.009	-0.001	-0.001	-0.001
0.5	0.5	0.274	0.399	0.000	0.014	0.001	0.001	0.001
0.5	0	0.003	-0.001	0.003	0.018	0.002	0.002	-0.001
0.5	-0.5	-0.269	-0.398	0.002	0.015	0.002	0.002	0.002
0.5	-1	-0.469	-0.752	-0.002	0.007	-0.004	-0.004	-0.004
0.5	-2	-0.667	-1.243	0.000	0.004	0.000	0.000	0.000
1	2	0.429	0.617	-0.004	0.001	-0.003	-0.003	-0.003
1	1	0.305	0.433	0.002	0.005	0.002	0.002	0.002
1	0.5	0.178	0.249	0.001	0.008	0.001	0.001	0.001
1	0	0.003	-0.001	-0.004	0.006	-0.003	-0.003	-0.001
1	-0.5	-0.171	-0.248	0.001	0.008	0.000	0.000	0.000
1	-1	-0.300	-0.432	0.001	0.006	0.002	0.002	0.002
1	-2	-0.428	-0.617	-0.002	0.000	-0.003	-0.003	-0.003

Table 2a The size of 10% nominal tests. Only Linear IV and agls use consistent standard errors. N=200, mc=1000, just identified.

Design		Estimator					
θ	λ	ols	probit	IV probit	Linear IV	agls	tscml
0.05	2	1.000	1.000	0.099	0.130	0.141	0.379
0.05	1	1.000	1.000	0.096	0.046	0.110	0.197
0.05	0.5	0.996	0.998	0.097	0.011	0.086	0.124
0.05	0	0.099	0.099	0.104	0.002	0.092	0.107
0.05	-0.5	0.998	0.997	0.092	0.025	0.086	0.123
0.05	-1	1.000	1.000	0.082	0.049	0.108	0.194
0.05	-2	1.000	1.000	0.096	0.115	0.121	0.365
0.1	2	1.000	1.000	0.089	0.108	0.114	0.339
0.1	1	1.000	1.000	0.092	0.045	0.102	0.193
0.1	0.5	0.999	0.999	0.103	0.032	0.105	0.137
0.1	0	0.099	0.088	0.110	0.008	0.102	0.111
0.1	-0.5	0.997	0.998	0.087	0.022	0.090	0.114
0.1	-1	1.000	1.000	0.091	0.067	0.110	0.192
0.1	-2	1.000	1.000	0.108	0.111	0.124	0.355
0.25	2	1.000	1.000	0.112	0.084	0.139	0.343
0.25	1	1.000	1.000	0.104	0.084	0.141	0.216
0.25	0.5	0.999	0.999	0.091	0.049	0.090	0.118
0.25	0	0.105	0.106	0.092	0.052	0.089	0.094
0.25	-0.5	0.999	0.999	0.089	0.060	0.098	0.125
0.25	-1	1.000	1.000	0.085	0.083	0.117	0.188
0.25	-2	1.000	1.000	0.088	0.105	0.127	0.369
0.5	2	1.000	1.000	0.085	0.085	0.114	0.348
0.5	1	1.000	1.000	0.093	0.084	0.114	0.192
0.5	0.5	0.994	0.995	0.115	0.097	0.127	0.156
0.5	0	0.097	0.101	0.113	0.094	0.111	0.114
0.5	-0.5	0.998	0.995	0.090	0.106	0.099	0.116
0.5	-1	1.000	1.000	0.099	0.098	0.122	0.193
0.5	-2	1.000	1.000	0.086	0.105	0.129	0.386
1	2	1.000	1.000	0.086	0.102	0.139	0.370
1	1	1.000	1.000	0.087	0.095	0.114	0.200
1	0.5	0.953	0.957	0.091	0.094	0.102	0.123
1	0	0.108	0.101	0.103	0.101	0.098	0.105
1	-0.5	0.976	0.966	0.095	0.111	0.104	0.132
1	-1	1.000	1.000	0.089	0.104	0.115	0.202
1	-2	1.000	1.000	0.073	0.092	0.112	0.379

Table 2b Compute rejection rate for 10% nominal t-tests. Standard errors for agls and Linear IV are consistent. N=1000, mc=1000, model is just identified.

Design		Estimator					
θ	λ	ols	probit	IV probit	Linear IV	agls	tscml
0.05	2	1.000	1.000	0.106	0.102	0.116	0.364
0.05	1	1.000	1.000	0.086	0.051	0.103	0.180
0.05	0.5	1.000	1.000	0.097	0.024	0.108	0.132
0.05	0	0.107	0.108	0.102	0.005	0.098	0.103
0.05	-0.5	1.000	1.000	0.100	0.036	0.107	0.134
0.05	-1	1.000	1.000	0.079	0.062	0.101	0.178
0.05	-2	1.000	1.000	0.085	0.110	0.124	0.348
0.1	2	1.000	1.000	0.090	0.090	0.121	0.359
0.1	1	1.000	1.000	0.080	0.062	0.101	0.173
0.1	0.5	1.000	1.000	0.091	0.044	0.096	0.115
0.1	0	0.092	0.101	0.122	0.043	0.120	0.121
0.1	-0.5	1.000	1.000	0.105	0.057	0.104	0.131
0.1	-1	1.000	1.000	0.098	0.084	0.119	0.192
0.1	-2	1.000	1.000	0.089	0.088	0.129	0.345
0.25	2	1.000	1.000	0.082	0.086	0.122	0.339
0.25	1	1.000	1.000	0.078	0.070	0.113	0.184
0.25	0.5	1.000	1.000	0.103	0.076	0.118	0.137
0.25	0	0.101	0.112	0.111	0.091	0.111	0.111
0.25	-0.5	1.000	1.000	0.095	0.089	0.112	0.130
0.25	-1	1.000	1.000	0.086	0.089	0.112	0.190
0.25	-2	1.000	1.000	0.080	0.077	0.116	0.327
0.5	2	1.000	1.000	0.077	0.086	0.130	0.343
0.5	1	1.000	1.000	0.069	0.071	0.102	0.172
0.5	0.5	1.000	1.000	0.110	0.091	0.121	0.139
0.5	0	0.094	0.099	0.106	0.097	0.104	0.106
0.5	-0.5	1.000	1.000	0.092	0.092	0.096	0.116
0.5	-1	1.000	1.000	0.087	0.102	0.110	0.198
0.5	-2	1.000	1.000	0.089	0.089	0.118	0.351
1	2	1.000	1.000	0.087	0.096	0.131	0.351
1	1	1.000	1.000	0.079	0.080	0.108	0.177
1	0.5	1.000	1.000	0.089	0.093	0.107	0.124
1	0	0.099	0.102	0.097	0.090	0.096	0.096
1	-0.5	1.000	1.000	0.098	0.092	0.107	0.134
1	-1	1.000	1.000	0.090	0.104	0.122	0.203
1	-2	1.000	1.000	0.093	0.110	0.141	0.382

Table 2c The size of 10% nominal tests. Only Linear IV and agls use consistent standard errors. N=200, mc=1000, model is overidentified.

Design		Estimator					
θ	λ	ols	probit	IV probit	Linear IV	agls	tscml
0.050	2.000	1.000	1.000	0.143	0.235	0.198	0.460
0.050	1.000	1.000	1.000	0.129	0.107	0.156	0.258
0.050	0.500	1.000	1.000	0.123	0.047	0.137	0.163
0.050	0.000	0.098	0.086	0.111	0.007	0.102	0.113
0.050	-0.500	1.000	0.999	0.122	0.052	0.125	0.159
0.050	-1.000	1.000	1.000	0.113	0.124	0.140	0.238
0.050	-2.000	1.000	1.000	0.137	0.232	0.195	0.442
0.100	2.000	1.000	1.000	0.134	0.238	0.198	0.451
0.100	1.000	1.000	1.000	0.111	0.099	0.129	0.223
0.100	0.500	0.999	0.998	0.100	0.046	0.099	0.122
0.100	0.000	0.105	0.111	0.106	0.020	0.099	0.111
0.100	-0.500	0.997	0.997	0.096	0.063	0.099	0.117
0.100	-1.000	1.000	1.000	0.095	0.118	0.124	0.204
0.100	-2.000	1.000	1.000	0.111	0.209	0.156	0.395
0.250	2.000	1.000	1.000	0.087	0.118	0.128	0.370
0.250	1.000	1.000	1.000	0.115	0.121	0.132	0.221
0.250	0.500	1.000	0.999	0.103	0.085	0.108	0.133
0.250	0.000	0.108	0.115	0.113	0.076	0.110	0.115
0.250	-0.500	0.999	0.999	0.090	0.096	0.100	0.127
0.250	-1.000	1.000	1.000	0.088	0.123	0.112	0.209
0.250	-2.000	1.000	1.000	0.092	0.144	0.132	0.361
0.500	2.000	1.000	1.000	0.090	0.098	0.124	0.370
0.500	1.000	1.000	1.000	0.094	0.091	0.108	0.188
0.500	0.500	0.994	0.996	0.106	0.098	0.111	0.134
0.500	0.000	0.124	0.117	0.096	0.110	0.097	0.101
0.500	-0.500	0.997	0.994	0.110	0.109	0.111	0.141
0.500	-1.000	1.000	1.000	0.082	0.096	0.108	0.190
0.500	-2.000	1.000	1.000	0.091	0.119	0.129	0.365
1.000	2.000	1.000	1.000	0.085	0.100	0.122	0.351
1.000	1.000	1.000	1.000	0.101	0.115	0.118	0.191
1.000	0.500	0.931	0.946	0.108	0.113	0.115	0.139
1.000	0.000	0.115	0.122	0.093	0.098	0.092	0.095
1.000	-0.500	0.955	0.951	0.089	0.100	0.095	0.121
1.000	-1.000	1.000	1.000	0.094	0.122	0.113	0.196
1.000	-2.000	1.000	1.000	0.084	0.095	0.125	0.357

Table 2d The size of 10% nominal tests. Standard errors of agls and Linear IV are consistent. N=1000, mc=1000, model is overidentified.

Design		Estimator					
θ	λ	ols	probit	IV probit	Linear IV	agls	tscml
0.05	2	1.000	1.000	0.122	0.206	0.147	0.415
0.05	1	1.000	1.000	0.108	0.133	0.117	0.184
0.05	0.5	1.000	1.000	0.096	0.054	0.110	0.130
0.05	0	0.086	0.084	0.099	0.023	0.100	0.099
0.05	-0.5	1.000	1.000	0.106	0.036	0.112	0.135
0.05	-1	1.000	1.000	0.085	0.090	0.115	0.195
0.05	-2	1.000	1.000	0.135	0.201	0.175	0.398
0.1	2	1.000	1.000	0.100	0.153	0.120	0.341
0.1	1	1.000	1.000	0.091	0.138	0.123	0.199
0.1	0.5	1.000	1.000	0.085	0.083	0.096	0.110
0.1	0	0.111	0.109	0.109	0.065	0.109	0.109
0.1	-0.5	1.000	1.000	0.099	0.042	0.104	0.119
0.1	-1	1.000	1.000	0.093	0.076	0.131	0.192
0.1	-2	1.000	1.000	0.073	0.111	0.123	0.332
0.25	2	1.000	1.000	0.095	0.116	0.155	0.378
0.25	1	1.000	1.000	0.098	0.108	0.126	0.201
0.25	0.5	1.000	1.000	0.097	0.104	0.101	0.128
0.25	0	0.102	0.109	0.095	0.100	0.095	0.095
0.25	-0.5	1.000	1.000	0.097	0.089	0.110	0.128
0.25	-1	1.000	1.000	0.108	0.112	0.125	0.207
0.25	-2	1.000	1.000	0.098	0.095	0.130	0.365
0.5	2	1.000	1.000	0.089	0.106	0.119	0.344
0.5	1	1.000	1.000	0.085	0.104	0.107	0.179
0.5	0.5	1.000	1.000	0.086	0.101	0.091	0.111
0.5	0	0.089	0.093	0.109	0.106	0.106	0.108
0.5	-0.5	1.000	1.000	0.122	0.120	0.121	0.151
0.5	-1	1.000	1.000	0.087	0.095	0.112	0.195
0.5	-2	1.000	1.000	0.060	0.071	0.094	0.311
1	2	1.000	1.000	0.081	0.097	0.128	0.335
1	1	1.000	1.000	0.095	0.108	0.116	0.187
1	0.5	1.000	1.000	0.114	0.126	0.124	0.148
1	0	0.103	0.107	0.122	0.117	0.120	0.121
1	-0.5	1.000	1.000	0.106	0.108	0.122	0.146
1	-1	1.000	1.000	0.088	0.102	0.114	0.201
1	-2	1.000	1.000	0.096	0.111	0.149	0.372

Table 3a Monte Carlo standard error each estimator based on samples of size 200, 1000 samples. The model is just identified. The approximate proportion of 1's in each sample is .5.

Design		Estimator						
θ	λ	ols	probit	IV probit	Linear IV	agls	tscml	pretest
0.05	2	0.002	0.010	7.894	1.865	2.939	2.939	1.060
0.05	1	0.002	0.005	2.063	0.715	1.086	1.086	0.712
0.05	0.5	0.002	0.004	3.382	1.599	1.876	1.876	1.116
0.05	0	0.002	0.003	12.405	7.046	8.544	8.544	0.378
0.05	-0.5	0.002	0.004	3.882	2.047	3.876	3.876	0.662
0.05	-1	0.002	0.005	1.773	1.389	3.186	3.186	0.434
0.05	-2	0.002	0.010	0.463	0.292	0.744	0.744	0.559
0.1	2	0.002	0.009	22.052	6.168	8.284	8.284	8.241
0.1	1	0.002	0.005	3.107	0.440	0.918	0.918	0.646
0.1	0.5	0.002	0.004	0.736	0.267	0.452	0.452	0.222
0.1	0	0.002	0.003	12.608	7.070	8.960	8.960	0.108
0.1	-0.5	0.002	0.004	0.214	0.113	0.284	0.284	0.086
0.1	-1	0.002	0.005	0.755	0.551	1.002	1.002	0.981
0.1	-2	0.002	0.009	0.382	0.233	0.625	0.625	0.511
0.25	2	0.002	0.008	0.154	0.044	0.138	0.138	0.139
0.25	1	0.002	0.005	0.075	0.028	0.050	0.050	0.052
0.25	0.5	0.002	0.004	0.063	0.028	0.037	0.037	0.031
0.25	0	0.002	0.003	0.064	0.027	0.045	0.045	0.033
0.25	-0.5	0.002	0.004	0.033	0.020	0.033	0.033	0.026
0.25	-1	0.002	0.005	0.057	0.043	0.085	0.085	0.087
0.25	-2	0.002	0.008	0.072	0.046	0.109	0.109	0.107
0.5	2	0.002	0.006	0.024	0.007	0.017	0.017	0.017
0.5	1	0.002	0.004	0.018	0.006	0.011	0.011	0.012
0.5	0.5	0.002	0.003	0.015	0.006	0.010	0.010	0.012
0.5	0	0.002	0.003	0.012	0.006	0.009	0.009	0.006
0.5	-0.5	0.002	0.003	0.009	0.006	0.009	0.009	0.011
0.5	-1	0.002	0.004	0.008	0.006	0.011	0.011	0.012
0.5	-2	0.002	0.006	0.011	0.007	0.017	0.017	0.017
1	2	0.001	0.003	0.011	0.003	0.008	0.008	0.008
1	1	0.002	0.003	0.008	0.003	0.005	0.005	0.005
1	0.5	0.002	0.003	0.007	0.003	0.004	0.004	0.005
1	0	0.002	0.003	0.006	0.003	0.004	0.004	0.003
1	-0.5	0.002	0.003	0.004	0.003	0.004	0.004	0.005
1	-1	0.002	0.003	0.004	0.003	0.005	0.005	0.005
1	-2	0.001	0.003	0.005	0.003	0.008	0.008	0.008

Table 3b Monte Carlo standard error each estimator based on samples of size 1000, 1000 samples. The model is just identified. The approximate proportion of 1's in each sample is .5.

Design		Estimator						
θ	λ	ols	probit	IV probit	Linear IV	agls	tscml	pretest
0.05	2	0.001	0.004	1.31	0.377	0.751	0.751	0.712
0.05	1	0.001	0.002	0.821	0.297	0.49	0.49	0.304
0.05	0.5	0.001	0.002	2.168	0.879	1.349	1.349	0.16
0.05	0	0.001	0.001	2.438	1.193	1.724	1.724	1.551
0.05	-0.5	0.001	0.002	2.122	1.279	2.089	2.089	1.981
0.05	-1	0.001	0.002	8.888	6.092	11.608	11.608	11.607
0.05	-2	0.001	0.004	1.256	0.771	1.487	1.487	1.378
0.1	2	0.001	0.004	0.368	0.1	0.243	0.243	0.243
0.1	1	0.001	0.002	3.428	1.253	1.714	1.714	0.056
0.1	0.5	0.001	0.002	0.682	0.297	0.401	0.401	0.053
0.1	0	0.001	0.001	0.195	0.099	0.138	0.138	0.129
0.1	-0.5	0.001	0.002	0.207	0.123	0.222	0.222	0.204
0.1	-1	0.001	0.002	0.038	0.029	0.051	0.051	0.049
0.1	-2	0.001	0.004	0.501	0.311	0.623	0.623	0.623
0.25	2	0.001	0.003	0.02	0.006	0.014	0.014	0.014
0.25	1	0.001	0.002	0.015	0.005	0.009	0.009	0.01
0.25	0.5	0.001	0.002	0.013	0.005	0.008	0.008	0.01
0.25	0	0.001	0.001	0.01	0.005	0.007	0.007	0.005
0.25	-0.5	0.001	0.002	0.008	0.005	0.008	0.008	0.01
0.25	-1	0.001	0.002	0.007	0.005	0.009	0.009	0.009
0.25	-2	0.001	0.003	0.009	0.006	0.014	0.014	0.014
0.5	2	0.001	0.003	0.01	0.003	0.007	0.007	0.007
0.5	1	0.001	0.002	0.007	0.003	0.004	0.004	0.004
0.5	0.5	0.001	0.001	0.006	0.003	0.004	0.004	0.004
0.5	0	0.001	0.001	0.005	0.002	0.004	0.004	0.003
0.5	-0.5	0.001	0.001	0.004	0.003	0.004	0.004	0.004
0.5	-1	0.001	0.002	0.003	0.003	0.004	0.004	0.004
0.5	-2	0.001	0.002	0.004	0.003	0.006	0.006	0.006
1	2	0.001	0.001	0.005	0.001	0.003	0.003	0.003
1	1	0.001	0.001	0.003	0.001	0.002	0.002	0.002
1	0.5	0.001	0.001	0.003	0.001	0.002	0.002	0.002
1	0	0.001	0.001	0.002	0.001	0.002	0.002	0.001
1	-0.5	0.001	0.001	0.002	0.001	0.002	0.002	0.002
1	-1	0.001	0.001	0.002	0.001	0.002	0.002	0.002
1	-2	0.001	0.001	0.002	0.001	0.003	0.003	0.003

Table 4a Comparison of agls and LIML. Sample size = 200, model just identified.
Upper panel compars the coefficient on the endogenous variable (γ=0)
Lower panel compares the percentiles to the pvalue of the corresponding t-ratio.

	λ	0.5		2		0.5		2	
	θ	0.1		0.1		1		1	
		agls	LIML	agls	LIML	agls	LIML	agls	LIML
Coefficient	1%	-44.751	-1.021	-45.860	-0.96689	-0.563	-0.371	-0.720	-0.325
	5%	-7.270	-0.947	-10.488	-0.85039	-0.347	-0.271	-0.425	-0.235
	10%	-3.649	-0.864	-5.034	-0.70906	-0.271	-0.221	-0.328	-0.195
	25%	-0.790	-0.489	-0.842	-0.27075	-0.137	-0.118	-0.173	-0.114
	50%	0.300	0.293	1.117	0.888625	-0.008	-0.008	-0.009	-0.006
	75%	1.462	1.003	2.994	1.557343	0.113	0.109	0.136	0.108
	90%	3.645	1.111	8.057	2.068173	0.221	0.219	0.246	0.212
	95%	8.198	1.166	12.735	2.246212	0.270	0.269	0.318	0.272
	99%	48.105	1.253	64.591	2.512663	0.420	0.417	0.433	0.384
	Mean	-0.368	0.235	3.462	0.703199	-0.020	-0.005	-0.029	0.001
	Std. Dev.	31.512	0.756	87.029	1.033331	0.193	0.167	0.233	0.158
	Variance	992.991	0.571	7574.060	1.067773	0.037	0.028	0.055	0.025
	Skewness	-10.139	-0.216	19.665	-0.0193	-0.341	0.155	-0.502	0.395
	Kurtosis	255.376	1.546	497.026	1.71487	3.670	3.050	3.758	3.495
p-values	1%	0.077	0.00E+00	0.004	7.46E-17	0.019	0.001	0.017	0.004
	5%	0.222	1.78E-38	0.037	1.33E-06	0.079	0.027	0.075	0.045
	10%	0.299	2.60E-16	0.105	0.001	0.129	0.083	0.126	0.097
	25%	0.479	3.92E-04	0.329	0.076	0.265	0.228	0.277	0.245
	50%	0.697	0.222	0.660	0.393	0.517	0.517	0.499	0.489
	75%	0.868	0.696	0.856	0.720	0.773	0.775	0.753	0.755
	90%	0.952	0.915	0.934	0.884	0.905	0.905	0.903	0.903
	95%	0.976	0.958	0.965	0.938	0.957	0.958	0.954	0.954
	99%	0.996	0.995	0.994	0.987	0.995	0.995	0.984	0.983

Table 4b Comparison of agls and LIML. Sample size = 1000, model just identified.
Upper panel compars the coefficient on the endogenous variable (γ=0)
Lower panel compares the percentiles to the pvalue of the corresponding t-ratio.

	λ	0.5		2		0.5		2	
	θ	0.25		0.25		1		1	
		agls	LIML	agls	LIML	agls	LIML	agls	LIML
Coefficient	1%	-1.379	-0.646	-2.295	-0.548	-0.222	-0.183	-0.261	-0.160
	5%	-0.709	-0.454	-1.212	-0.370	-0.154	-0.133	-0.168	-0.109
	10%	-0.532	-0.376	-0.901	-0.307	-0.115	-0.104	-0.128	-0.086
	25%	-0.247	-0.199	-0.439	-0.177	-0.060	-0.054	-0.074	-0.050
	50%	-0.013	-0.012	-0.006	-0.003	-0.005	-0.005	-0.001	0.000
	75%	0.218	0.210	0.338	0.187	0.051	0.049	0.063	0.048
	90%	0.411	0.410	0.601	0.388	0.102	0.099	0.125	0.096
	95%	0.534	0.533	0.736	0.505	0.130	0.128	0.158	0.127
	99%	0.787	0.748	0.961	0.731	0.201	0.199	0.220	0.177
	Mean	-0.042	0.009	-0.101	0.021	-0.005	-0.002	-0.004	0.002
	Std. Dev.	0.397	0.300	0.643	0.273	0.087	0.080	0.100	0.072
	Variance	0.158	0.090	0.414	0.075	0.007	0.006	0.010	0.005
	Skewness	-0.845	0.257	-1.243	0.455	-0.104	0.112	-0.141	0.210
	Kurtosis	5.384	2.832	6.080	3.172	3.182	3.099	2.937	2.877
p-values	1%	0.010	7.38E-05	0.004	0.004	0.006	0.003	0.009	0.008
	5%	0.069	0.006	0.050	0.050	0.040	0.031	0.042	0.042
	10%	0.114	0.037	0.129	0.108	0.090	0.079	0.094	0.091
	25%	0.255	0.215	0.288	0.261	0.232	0.234	0.245	0.236
	50%	0.506	0.498	0.509	0.494	0.505	0.501	0.488	0.484
	75%	0.757	0.760	0.736	0.734	0.753	0.754	0.724	0.724
	90%	0.907	0.907	0.896	0.895	0.910	0.910	0.886	0.887
	95%	0.959	0.959	0.946	0.946	0.955	0.955	0.941	0.941
	99%	0.995	0.995	0.989	0.989	0.988	0.988	0.992	0.992