TESTING PARAMETER SIGNIFICANCE IN INSTRUMENTAL VARIABLES PROBIT ESTIMATORS: SOME SIMULATION RESULTS

LEE C. ADKINS

Abstract. Binary choice models that contain endogenous regressors can now be estimated routinely using modern software. Two packages, Stata 10 (StataCorp, 2009) and Limdep 9 (Greene, 2007), each contain two estimators that can be used to estimate such a model. Both contain maximum likelihood estimators, though they differ slightly in their computations details and yield marginally different results. Stata also includes a simple generalized least squares estimator suggested by Amemiya and explored by Newey that is computationally simple, though not necessarily efficient. Limdep also allows a user to use a plug-in estimator in conjunction with a 'robust' variance-covariance estimator. This choice is available, though, when there is only one endogenous regressor.

This paper compares the performance of these and three other estimators in samples of size 200 and 1000 using simulation. Specifically, the paper focuses on tests of parameter significance under various degrees of instrument strength and severity of endogeneity. Although the MLE performs well in large samples, there is some evidence that the more computationally robust AGLS estimator may perform better in smaller samples when instruments are weak. It also appears that instruments in endogenous probit estimation need to be even stronger than when used in linear IV estimation.

Keywords: Probit, GMM, Instrumental Variables, Monte Carlo, significance tests, dichotomous choice, endogenous regressors, generalized linear model, binary response

Classcode: 62J02, 62P20

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

Date: July 5, 2010.

Professor of Economics, Oklahoma State University, Stillwater OK 74078. Email: lee.adkins@okstate.edu.

My thanks to Randy Campbell for his careful reading of the manuscript. All remaining errors are mine alone.

regression, a regressor measured with error causes least squares to be inconsistent and similar results are found in binary choice models (Yatchew and Griliches, 1985). Rivers and Vuong (1988) and Smith and Stulz (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 to test the statistical significance of the model's 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. The Iwata estimator is reconsidered below in the context of endogenous probit model.

Blundell and Powell (2004) develop and implement what they refer to as "semiparametric methods for estimating binary response (binary choice) models with continuous endogenous regressors." Their approach enables one 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 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 (for example, 2SCML considered in Rivers and Vuong, 1988) are consistent for the parameters but not the standard errors, making it unlikely that they will perform satisfactorily in hypothesis testing. This was a preliminary finding of Adkins (2008).

The Monte Carlo design is based on that of Rivers and Vuong (1988), which gives us a way to calibrate results. Their purpose was different from ours, but the set of estimators they examined is at least partially relevant. They compared three different 2-step estimators and a limited information maximum likelihood estimator (ML) 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.

The instruments used in Rivers and Vuong (1988) were very highly correlated with the endogenous variable; in effect the instruments they used would be classified as being very strong; they did not assess the behavior of the estimators when instruments are weak. The other major departure is that the emphasis here is on hypothesis testing, rather than bias and mse. Given that the actual value of the location parameters in the probit model have little inherent meaning (since, under the usual normalization, the scale parameter is not identified) the magnitude of bias is not that meaningful–it lacks a scale; what matters is whether the variable(s) in question affect(s) the probability of observing the event and to a lesser extent a comparison of the magnitudes of significant marginal effects. Consequently, this paper measures the size distortions of the various t-ratios based on the asymptotic normality of distributions.

2. Statistical 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,

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

where $Z_t = [Y_t, X_{1t}], \delta^T = [\beta^T, \gamma^T]$, Y_t is the t^{th} 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

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

where V_t is a disturbance. The $K - s$ variables in X_{2t} are additional exogenous explanatory variables. Equation (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.

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

$$y_t = \begin{cases} 1 & y_t^* > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Assuming the errors of the model (1), u_t, are normally distributed leads to the probit model.

3. Estimators

One is certainly free to use simple linear estimators of this model 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, least squares estimator is $\hat{\delta}_{ols} = (Z^T Z)^{-1} Z^T y$. The least squares estimator is only consistent if Z is exogenous or predetermined. Still, it is easy to compute and the degree of inconsistency may be small in certain circumstances.

Iwata (2001) suggests a means of rescaling and recentering (RR) y_t that may improve the performance of least squares. The transformation of y_t is straightforward:

$$\tilde{y}_t = (y_t - \hat{\psi})/\hat{\phi} \tag{4}$$

and $\hat{\delta} = \Phi^{-1}(\bar{y})$, $\hat{\phi} = \phi(\hat{\delta})$, and $\hat{\psi} = \bar{y} - \hat{\phi}\hat{\delta}$; ϕ and Φ are the standard normal pdf and cdf, respectively. Collecting all observations into the vector $\tilde{y}$ and replace y in the least squares estimator yields:

$$\hat{\delta}_{rrols} = (Z^T Z)^{-1} Z^T \tilde{y} \tag{5}$$

The dependent variable, y_t, is heteroskedastic so a sandwich covariance (Davidson and MacKinnon, 2004, p. 199) is often recommended to obtain consistent standard errors.

$$\hat{\Sigma}_{HC0} = (Z^T Z)^{-1} Z^T \hat{\Omega} Z (Z^T Z)^{-1} \tag{6}$$

where $\hat{\Omega}$ is an nxn diagonal matrix with the t^{th} diagonal element equal to $\hat{u}_t^2$, the squared RROLS residual. The endogeneity of elements of Z ruins the proposed consistency of HC0, but this estimator is compared to other consistent ones, both by Iwata (2001) and below.

The usual linear instrumental variable estimator is also inconsistent and heteroscedastic. Iwata (2001) again suggests rescaling and recentering (RR) the data that can bring about consistency in this case. Iwata's rescaled and recentered generalized method of moments estimator (RRGMM) estimator is

$$\delta_{rrgmm} = (Z^T X \hat{H} X^T Z)^{-1} Z^T X \hat{H} X^T \tilde{y} \tag{7}$$

where $\tilde{y}$ is the rescaled and recentered binanry dependent variable, $\hat{H} = X^T \hat{\Omega} X$ and $\hat{\Omega}$ is an nxn diagonal matrix with the t^{th} diagonal element equal to $\hat{u}_t^2$, the squared IV residual. The variance-covariance is simply estimated using $(Z^T X \hat{H} X^T Z)^{-1}$.[1]

The usual probit mle can be used. However, if the regressors are endogenous, then this estimator is also inconsistent, Yatchew and Griliches (1985).

To develop the notation a bit further, let the probability that y_t is equal one be denoted

$$pr(y_t = 1) = \Phi(y_t, Y_t\beta + X_{1t}\gamma) = \Phi(y_t, Z_t\delta) \tag{8}$$

where once again 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. The usual normalization sets $\sigma^2 = 1$. 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.

Another approach is to use predicted values of Y_t from a first stage least squares estimation of equation (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

$$pr(y_t = 1) = \Phi(y_t, \hat{Z}_t\delta) \tag{9}$$

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) or simply the 'plug-in' estimator.

The major problem with the plug-in estimator is that the usual variance-covariance estimator yielded from maximizing the conditional likelihood is inconsistent. Murphy and Topel

[1]This is verified using the `ivregress gmm` command in Stata 10. See section 8 below for example code.

(1985) consider a relatively simple solution to this problem that can be composed after estimation of the first and second stage regressions. The MT computation is fairly simple when one has a single endogenous explanatory variable, but not when there are more than one.

Using the result in (Greene, 2008, p. 507), the Murphy-Topel (MT) estimator of the variance-covariance is

$$V = V_\delta + V_\delta[CV_\pi C^T - RV_\pi C^T - CV_\pi R^T]V_\delta \tag{10}$$

where V_π is the estimated covariance of the least squares estimated reduced form, V_δ is the estimated covariance from the second stage probit, C is the sum of the product of the probit loglikelihood and the partial derivatives of the probit loglikelihood with respect to the parameters of the reduced form, and R is the sum of the product of the probit logliklihood equations and the partial derivatives of the probit loglikelihood and least squares normal equations from the first stage regression. See Greene (2008) for details. It is not possible to use (10) when one has two or more endogenous explanatory variables. Although this limits the usefulness of this approach, as shown below it performs reasonably well in small samples.

Rivers and Vuong (1988) two stage conditional ML estimator (2SCML) adds the least squares residuals from equation (2), $\hat{V}_t = Y_t - X_t\hat{\Pi}$ to (9). This brings

$$pr(y_t = 1) = \Phi(y_t, \hat{Y}_t\beta + X_{1t}\gamma + \hat{V}_t\lambda) = \Phi(y_t, \hat{Z}_t\delta + \hat{V}_t\lambda) \tag{11}$$

which is estimated by maximum likelihood, again conditional on $\hat{\Pi}$. This takes the form

$$pr(y_t = 1) = \Phi(y_t, Z_t\delta + \hat{V}_t\rho) \tag{12}$$

The parameter ρ is related to λ in (11) by $\lambda = \rho + \beta$. This follows because $Z_t\delta = \hat{Z}_t\delta + \hat{V}_t\beta$.

The 2SCML estimator is not used to estimate the model's parameters but for testing the exogeneity of Y_t. A simple Wald test based on the regression in (12) was shown by Rivers and Vuong (1988) to perform reasonably well and it will be used as the basis of a pretest estimator that will also be considered. The pretest estimator is written

$$\delta_{pt} = I(t)_{[0,c_\alpha)}\delta_{mle} + I(t)_{[c_\alpha,\infty)}\delta_{iv} \tag{13}$$

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 an instrumental variables probit estimator–specifically, AGLS.

An efficient alternative to (11) that also yields a consistent estimator of the precision is Amemiya's generalized least squares (AGLS) estimator as proposed by Newey (1987). The AGLS estimator of the endogenous probit model is easy to compute, though there are several steps. The basic algorithm from Adkins (2008) is:

(1) Estimate the reduced form (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. The exogenous variables are augmented by the residuals obtained in step 1. Hence,

$$pr(y_t = 1) = \Phi(y_t, X_t\alpha + \hat{V}_t\lambda) \tag{14}$$

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 (11). 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

$$\delta_A = [D(\hat{\Pi})^T\Omega^{-1}D(\hat{\Pi})]^{-1}D(\hat{\Pi})^T\Omega^{-1}\hat{\alpha} \tag{15}$$

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	Equation
RROLS	Z_t	(5)
RRGMM	$\hat{Z}_t$	(7)
Probit mle	Z_t	(8)
Probit mle (plug-in)	$\hat{Z}_t$	(9)
AGLS	$D(\hat{\Pi})$, $\hat{\alpha}$	(15)
PT	$\hat{V}_t$, Z_t, $D(\hat{\Pi})$, $\hat{\alpha}$	(13)

One thing that complicates comparison of these estimators is that they do not all use the same normalization. One alternative is to compare marginal effects. This is the approach taken by Arendt and Holm (2006). This choice is appealing since this is the quantity that interests many. In principle, the marginal effects shouldn't be sensitive to normalization, although the analytical computation does depend on the normalization used.

Another alternative is to compare a pivotal statistic (asymptotically) like the t-ratio. This is attractive because the model is commonly used to test parameter significance. Testing whether a coefficient is zero should not be materially affected by normalization and this is what I have chosen to investigate. This simplifies the design of the Monte Carlo simulations without sacrificing generality.

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 is studied using simulation. The main simulations were conducted in Gauss 7.0 using code written by the author. These basic results were confirmed using Stata 10.0, based on an additional simulation that compares the AGLS and Maximum Likelihood estimators. This latter simulation was necessarily more limited in scope due to computational limitations of mle estimation caused by weak instruments.

Bias and the size of a test of a significance on the endogenous variable are compared. 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, the relative variability of the endogenous regressor and the equation's error, and the effects of overidentification.

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 used in Rivers and Vuong (1988) is incorporated as well and their notation will be adopted with some minor modifications.

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

$$y_{1i}^* = \gamma y_{2i} + \beta_1 + \beta_2 x_{2i} + u_i \tag{16}$$

In the just identified case

$$y_{2i} = \pi_1 + \pi_2 x_{2i} + \pi_3 x_{3i} + \nu_i \tag{17}$$

and the over-identified case,

$$y_{2i} = \pi_1 + \pi_2 x_{2i} + \pi_3 x_{3i} + \pi_4 x_{4i} + \nu_i \tag{18}$$

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

$$u_i = \lambda \nu_i + \eta_i \tag{19}$$

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, an important design element not considered by 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 compare the actual size of a t-test on the endogenous variable to its nominal level 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 aforementioned 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[2]. 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.

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).

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

	θ					
	0.05	0.1	0.15	0.25	0.5	1
	n=200; just identified					
R^2	.022	.045	.079	.175	.446	.76
Overall-F	1.8	4.2	8.0	20.5	79	313
	n=1000; just identified					
R^2	.011	.030	.061	.150	.410	.735
Overall-F	4.4	14 .8	32.1	87.5	346	1383
	n=200; over identified					
R^2	.024	.038	.059	.122	.330	.658
Overall-F	1.3	2.3	3.8	8.8	32.3	126
	n=1000; over identified					
R^2	.009	.024	.049	.120	.348	.680
Overall-F	2.8	8.0	16.9	45	177	706

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 a series of tables. Table 1 includes bias for each design point based on 1000 Monte Carlo samples. Tables 2 and 3 contain the sizes of 10% nominal tests and the Monte Carlo standard errors, respectively. Tables 1 and 2 are broken into sub-tables a, b, c, and d, reflecting differences in sample size and identification of the model. Tables 1a, 2a, and 3a are based on samples of size 200 for a just identified model. Tables 1b, 2b, and 3b are for just identified models having 1000 observations. Tables labeled c and d are for overidentified models with 200 and 1000 observations, respectively. The Monte Carlo standard errors for the overidentified models are omitted, but are essentially the same as those for the just identified models.

Tables 4a and 4b compare the AGLS estimator to the maximum likelihood estimator for a limited number of design. For computational reasons, the scope of the comparison is limited; designs based on weak instruments and overidentified models posed convergence problems for the mle. This illustrates the fragility of the mle when parameters are poorly identified.

In all of the tables, the parameter labeled θ controls the strength of the instruments. As θ increases, instruments become stronger. It should be noted that Rivers and Vuong (1988) only considered $\theta = 1$, which implies very strong instruments. The parameter labeled λ controls the strength of the endogeneity. When $\lambda = 0$ the regressors are exogenous. As $|\lambda|$ increases, the correlation between the errors of the model and the endogenous regressor increase.

5.1. **Bias.** In tables 1a-1d the bias of each estimator is given for each of the design points considered. For the results in table 1a the design included one endogenous variable and one instrument; the model is just identified. When the instrument is weak (e.g. $\theta = 0.05$) and there is any correlation between the regressor and the regression error ($\lambda \neq 0$), then weak instruments create considerable 'bias.' It is unlikely that the instrumental variables estimators have a mean in this case since subsequent simulations yielded quite different numerical results (though the performance is always very poor). When $\theta = .15$ the corresponding $F-$statistic is 8.0, indicating that the instruments are nearing the usual threshold of 10 suggested for linear models by Staiger and Stock (1997). Bias is substantial at this point and continues to exceed .5 until instruments become quite strong ($\theta = .5$ where the corresponding $F-$statistic is 79).

The pretest estimator actually performs quite well. When the instruments are very weak, the pretest picks the probit mle (exogenous regressors) often. As the instruments gain strength the pretest picks the consistent estimators with high frequency. On balance then, the pretest estimator is relatively effective in estimating the parameter of interest (at least compared to the competitors). AGLS and IVP have smaller biased than RRGMM when the instruments are strong. 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 87.5; this is large by conventional standards. RRGMM, IVP (plug-in), AGLS, and pretest estimators are all erratic when instruments are weakest. When the instruments are very strong ($\theta >= .5$), all of the IV estimators perform reasonably 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 the plug-in estimator now. Both outperform RRGMM by a small amount.

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 .1 when instruments are very, very weak ($\theta = .05$).

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 justifies it's inclusion as an option in Stata. RRGMM is not far behind in terms of bias. Clearly, when the regressors are endogenous, RROLS and the usual mle are not recommended, except when instruments are barely correlated with the endogenous variable(s).

Since the scale parameter is not identified, the magnitude of the coefficients is not very important in the probit model. More importantly, one is usually interested in testing the statistical significance of one or more variables in the model. For this, comparing the sizes of 10% nominal tests, which are asymptotically pivotal, can be much more revealing about the performance of the various estimators considered.

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 AGLS estimator is very close to the nominal 0.1 level when the endogeneity problem is most severe and instruments very weak. This is a somewhat of a surprise, given the large biases recorded in table 1a.

As the instruments gain strength, RRGMM and the plug-in estimator begin to dominate the AGLS. The AGLS estimator performs at the desired level when endogeneity weak, but suffers from size distortions as λ becomes large. Overall, the RMSE of the AGLS estimator is significantly smaller (.015) than the others, helped by its rather good performance with weak instruments.

In table 2b the sample is increased to 1000. Predictably, the results improve for most cases and the gap between AGLS, RRGMM, and the plug-in estimators narrows. AGLS still exhibits some size distortion when $\lambda =$ is large. for instance, when $\lambda = 2$ and $\theta = .25$ a nominal 10% test is rejecting a true null hypothesis 14% of the time. This is not terrible, but there appears to be little improvement from increasing the sample size in this case.

In table 2c we examine the overidentified case using 200 observations. Overidentification is not improving things here at all. The plug-in and RRGMM estimators are now experiencing larger size distortion when instruments are weak and the size distortion of the AGLS estimator is becoming quite large (.19) 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 (¿.14) when instruments are weak and endogeneity severe. Overidentification does improve its performance once instruments gain some strength.

In table 2d, which corresponds to overidentified models with samples of size 1000, the size of distortions drop further. The overall RMSE for AGLS is now .019. That of the RRGMM estimator is just slightly larger at .021. The plug-in estimator actually wins this derby by a small amount. All of the estimators struggle the most when instruments are weakest.

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 Newey's AGLS estimator. These are the two options available in Stata 10, which make them popular choices in applied work. Although a more thorough analysis of the maximum likelihood estimator (mle) would be welcome, one could not be conducted because of computational difficulties. When instruments are weak, the mle is prone to not converge. In the designs considered within this paper, there were far too many circumstances when the ML estimator failed to converge and this makes a proper analysis of its properties in Stata impossible. One could reasonably draw the conclusion from this that if the mle fails to converge to anything reasonable with a particular dataset, then perhaps the model itself needs to be rethought.

To get some idea of how these two estimators compare, I chose 4 designs for which both AGLS and the mle would converge for each of the 1000 samples generated. I examined the summary statistics associated with the t-ratio and the 5% and 10% 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 $\lambda = -0.25, -2$ and $\theta = .15, 1$ were examined. The results for n=200 appear in table 4a.

Looking at table 4a, notice that the t-ratios for the AGLS estimator are actually more precise than those for the mle; variance of the AGLS estimator is everywhere less than one. The AGLS estimator consistently outperforms the mle by getting closer to the 5% and 10% nominal test size (lower panel). The 5% and 95% pecentiles of the t-ratio should be close to -1.645 and 1.645 if the ratio is nearing its asymptotic distribution. Both estimators are skewed and the critical values are not symmetric. Overall, the AGLS estimator performs better, but note that for one design ($\lambda = -2$ and $\theta = .15$ the 95% percentile is 0.512, which is quite a distance from its theoretical limit. In this case we would never find a positive coefficient different from zero. The mle performs dreadfully, though with the t-ratios in the rejection region of the test far more frequently than they should be. One thought was to try using the outer product of the gradient to compute standard errors, but this had no appreciable effect on the statistic. Normality of the t-ratio was tested using a Shapiro-Wilk W statistic and normality was rejected in each instance. Still, for a two-sided test, the AGLS estimator actually gets quite close to the desired rejection rates, skewness notwithstanding.

For samples of 1000, the mle refused to converge for many designs when $\theta = .15$ so stronger instruments had to be used, in this case by letting $\theta = .25$; these results appear in table 4b. The only designs where AGLS outperforms ML is when instruments are relatively weak. Otherwise it is essentially a draw, at least in terms of rejection rates for the test. On the other hand, the mle is approximately normal when the instruments are very strong. In all cases, the mle now demonstrates less skewness. For large samples with strong instruments, the near normality of the mle makes it the one to use.

The absence of results for overidentified models deserves mention. Stata, despite its top notch algorithms, fails to converge for many of the samples for the designs considered and hence yielded no usable results.

These simulations were repeated using positive correlation between errors ($\lambda = 0.25, 2$ and the results were roughly similar.

6. Example

In this section the differences between ML and AGLS estimators is demonstrated using data and a model similar to one used by Adkins et al. (2007). The main goal of that paper was to determine whether managerial incentives affect the use 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 (**?**). 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.

Table 5. **Summary Results from Reduced-form Equations**. The table contains p-values for the instruments and R^2 for each reduced form regression which is estimated using least squares. 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

Each of the instruments (except the 12 month maturity mismatch) appears to be relevant in that each is significantly different from zero at the 10% (p-value < 0.1) in at least one equation; CEO age and CFA are significant in one equation; the number of offices, employees, and 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.

The simulation results suggest there may be some small benefit to be had from discarding extra instruments. 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.

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. Approximate p-values based on the asymptotic distribution of the estimators are reported in parentheses beneath the parameter estimates. Significant parameters are typeset in bold.

	Instrumental Variables Probit		
	RRGMM	AGLS	ML
Leverage	11.108	21.775	**12.490**
	(0.134)	(0.104)	(0.021)
Option Awards	-5.23E-08	**-8.79E-08**	**-5.11E-08**
	(0.113)	(0.098)	(0.002)
Bonus	8.44E-07	**1.76E-06**	**1.02E-06**
	(0.155)	(0.048)	(<0.001)
Total Assets	**0.361**	**0.365**	0.190
	(0.003)	(0.032)	(0.183)
Insider Ownership %	0.095	**0.259**	**0.145**
	(0.156)	(0.026)	(0.016)
Institutional Ownership %	**0.100**	**0.370**	**0.201**
	(0.069)	(0.006)	(0.041)
Return on Equity	-0.015	-0.034	**-0.020**
	(0.395)	(0.230)	(0.083)
Market-to-Book ratio	-0.001	-0.002	**-0.001**
	(0.306)	(0.132)	(0.098)
Foreign to Total Interest Income Ratio	0.135	-3.547	-2.177
	(0.958)	(0.356)	(0.127)

Continued from preceding page			
	Instrumental Variables Probit		
	RRGMM	AGLS	ML
Derivative Dealer Activity Dummy	-0.057	-0.280	-0.154
	(0.727)	(0.257)	(0.288)
Dividends Paid	-2.37E-07	-8.43E-07	**-4.84E-07**
	(0.519)	(0.134)	(0.044)
D=1 if 1997	-0.005	-0.024	-0.016
	(0.979)	(0.930)	(0.914)
D=1 if 1998	-0.170	-0.244	-0.133
	(0.283)	(0.352)	(0.383)
D=1 if 1999	-0.135	-0.242	-0.134
	(0.446)	(0.391)	(0.395)
D=1 if 2000	-0.087	-0.128	-0.065
	(0.634)	(0.643)	(0.685)
Constant	-7.553	-9.673	-5.188
	(<0.001)	(<0.001)	(4.40E-02)
Sample size	794	794	794

Only two variables are significantly different from zero at the 10% level in the model estimated by RRGMM: Total assets and Institutional ownership percentage.[3] Leverage is significant in the ML estimation 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, but as the simulations show, the ML estimator tends to find significance when there isn't any, especially if the sample size is not large and the instruments on the weak side or in larger samples when endogeneity is not severe.

The results correspond the closest to those in tables 2d and 4b. The model is overidentified, sample is relatively large (700+), and the instruments are very strong ($\theta = .5$ or $\theta = 1$). The degree of endogeneity is unknown. In these designs, AGLS performs well and the actual size of the nominal 10% tests varies within an acceptable range for all levels of endogeneity (though is exceeds 13% for severe endogeneity).

[3]The RRGMM was estimated using gretl 1.8.2–Stata 10 could not estimate the model using RRGMM.

Given the overall strength of the instruments, I see little reason not to use the mle in this case. In the simulations it was more likely to be normally distributed in large samples with strong instruments; furthermore, in this case the sizes of the tests were close to the nominal level, although slightly prone to over-reject the zero null hypothesis.

7. Conclusion

Based on the results from the simulations the following general conclusions can be made.

(1) When there is no endogeneity, RROLS and Probit work well (as expected) but RROLS and Probit should be avoided when you have an endogenous regressor.
(2) When instruments are very weak, it is unlikely that the estimators converge to the mean unless the sample is very large. As sample size increases and instruments become stronger, the instrumental variables probit estimators considered become essentially unbiased.
(3) The size of the significance tests based on the AGLS estimator is 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 nearly twice as often as it should. The RRGMM and plug-in estimators perform relatively better when endogeneity is severe.
(4) RRGMM estimator that use consistent estimators of standard errors can be used for significance testing. It actually outperforms AGLS in smaller samples when instruments are moderately strong. In larger samples the size distortions are much more similar.
(5) 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 the plug-in estimator, which as mentioned above, works fairly well most of the time). The AGLS estimator is prone to very high variance when samples are small and instruments weak (comparing variance results in table 4a).
(6) For point estimation pretesting for endogeneity is useful when the sample is very small and the available instruments weak.
(7) In small samples the AGLS estimator outperforms mle when it comes to testing for the significance of a parameter in the model. When instruments are weak, it also outperforms the mle in larger samples. As instruments get stronger, the mle – at

least in large sample – is faster to converge to its asymptotic distribution and is in that case recommended.

(8) For point estimation, there is no question that mle is more precise. Though not reported in any of the tables, there is much smaller variation in the parameter estimates themselves with the mle. It's poor relative performance is due to underestimation of standard error which in turn leads to landing in the rejection region of a 10% test far too often.

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 does this very well, but a small nod goes to AGLS when endogeneity is not extreme. The ML estimator led to unacceptably high levels of type one error in small samples with weak instruments. It performs much better when sample size increases and as endogeneity worsens. RRGMM actually performs well relative to AGLS as endogeneity worsens and instruments are strong.

8. Appendix: Stata and Gretl code to RRgmm

The following code examples show how simple the rrgmm estimator is to compute. Note, gretl code to compute AGLS can be found in **?**.

```
y = binary dependent variable
xi = exogenous regressors (i=1,2,3)
wi = exogenous instruments (i=1,2)
y1 = endogenous regressor
```

8.1. **Stata 10.**

```
egen ybar = mean(y)
scalar delta = invnormal(ybar)
scalar phi = normalden(delta)
scalar phi2=ybar-phi*delta
scalar phi1=phi
gen ytil = (y-phi2)/phi1
ivregress gmm ytil x1 x2 x3 (y1 = w1 w2), wmatrix(robust) vce(robust)
```

8.2. **gretl 1.8.2.**

```
ybar=mean(y)
delta=invcdf(N,ybar)
phi = pdf(N,delta)
phi2 = ybar-delta*phi
genr ytil = (y-phi2)/phi

tsls ytil const x1 y1 x2 x3 ; const x1 x2 x3 \
  w1 w2 --gmm --iterate
```

References

Adkins, Lee C. (2008), Small sample performance of instrumental variables probit estimators: A monte carlo investigation.

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 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 9.0*, Econometric Software, Inc., Plainview, NY.

Greene, William H. (2008), *Econometric Analysis*, 6th edn, Prentice-Hall, Upper Saddle River, NJ.

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, Jr., C. W. and R. M. Stulz (1985), 'The determinants of firms' hedging policies', *Journal of Financial and Quantitative Analysis* **20**, 391–405.

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

StataCorp (2009), *Stata Statistical Software: Release 11*, StataCorp LP, College Station, TX.

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.

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.

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	RRrols	Probit	RRgmm	IVP	AGLS	Pretest
0.05	2	2.047	2.066	-4.771	-5.443	-9.613	1.492
0.05	1	1.438	1.029	0.404	1.478	1.192	1.551
0.05	0.5	0.818	0.513	20.418	-2.381	-1.660	0.793
0.05	0	-0.006	-0.002	-4.048	-2.467	-1.684	0.049
0.05	-0.5	-0.817	-0.512	37.586	18.977	19.152	-0.482
0.05	-1	-1.438	-1.031	-2.344	-1.060	-1.652	-1.228
0.05	-2	-2.052	-2.075	-1.552	-0.831	-0.870	-2.649
0.1	2	2.037	2.058	0.315	0.511	0.545	0.673
0.1	1	1.429	1.024	-1.513	-1.393	-0.914	0.780
0.1	0.5	0.818	0.512	0.046	0.023	-0.017	0.395
0.1	0	-0.001	0.002	-2.918	-1.341	-0.930	0.388
0.1	-0.5	-0.824	-0.512	6.176	4.307	4.193	3.909
0.1	-1	-1.432	-1.022	-1.892	-1.527	-0.346	-0.101
0.1	-2	-2.040	-2.034	0.987	0.557	2.044	2.072
0.15	2	2.023	1.989	2.214	2.859	2.042	0.595
0.15	1	1.407	1.006	-0.823	-0.755	-0.294	-0.051
0.15	0.5	0.816	0.510	-1.911	-1.388	-0.909	-0.612
0.15	0	-0.023	-0.009	-0.936	-0.365	-0.260	-0.032
0.15	-0.5	-0.820	-0.510	-0.530	-0.001	-0.074	-0.351
0.15	-1	-1.416	-1.006	0.112	0.141	0.208	0.027
0.15	-2	-2.019	-1.997	-0.679	-0.154	2.286	1.828
0.25	2	1.967	1.831	-0.226	-0.257	-0.211	-0.193
0.25	1	1.370	0.966	-0.356	-0.265	-0.166	0.037
0.25	0.5	0.784	0.492	-0.626	-0.384	-0.207	0.002
0.25	0	-0.012	-0.002	-0.283	-0.023	-0.017	-0.001
0.25	-0.5	-0.804	-0.493	0.013	0.129	0.123	-0.118
0.25	-1	-1.375	-0.968	0.110	0.109	0.160	-0.004
0.25	-2	-1.971	-1.845	0.390	0.270	0.442	0.386
0.5	2	1.723	1.361	-0.075	-0.081	-0.065	-0.065
0.5	1	1.198	0.803	-0.077	-0.029	-0.018	-0.012
0.5	0.5	0.669	0.416	-0.111	-0.027	-0.013	0.079
0.5	0	-0.020	-0.002	-0.099	0.005	0.003	0.004
0.5	-0.5	-0.710	-0.427	-0.073	0.007	0.007	-0.067
0.5	-1	-1.213	-0.794	0.004	0.027	0.039	0.033
0.5	-2	-1.732	-1.363	0.074	0.058	0.091	0.091
1	2	1.168	0.713	-0.018	-0.012	-0.007	-0.007
1	1	0.808	0.489	-0.034	-0.011	-0.008	-0.008
1	0.5	0.452	0.277	-0.049	-0.007	-0.005	0.002
1	0	-0.019	0.002	-0.055	-0.003	-0.002	0.002
1	-0.5	-0.485	-0.275	-0.030	0.007	0.007	-0.001
1	-1	-0.829	-0.486	-0.006	0.009	0.011	0.011
1	-2	-1.180	-0.713	-0.002	0.004	0.002	0.002
RMSE		1.100	0.860	2.259	1.183	1.250	0.504

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	RRrols	Probit	RRgmm	IVP	AGLS	Pretest
0.05	2	2.057	2.004	-1.543	-1.919	-1.514	-1.173
0.05	1	1.456	1.006	0.605	0.902	0.569	0.707
0.05	0.5	0.837	0.500	0.465	0.821	0.629	0.559
0.05	0	0.000	0.000	-0.910	0.143	0.129	-0.102
0.05	-0.5	-0.840	-0.502	5.346	3.119	2.598	-0.160
0.05	-1	-1.457	-1.005	4.075	2.447	3.859	1.344
0.05	-2	-2.054	-2.013	46.372	29.716	39.307	38.596
0.1	2	2.041	1.974	-0.775	-0.993	-0.742	-0.698
0.1	1	1.445	0.994	1.018	1.208	0.798	1.051
0.1	0.5	0.837	0.499	-0.352	-0.110	-0.071	0.163
0.1	0	-0.004	-0.001	-0.288	0.013	0.009	0.042
0.1	-0.5	-0.838	-0.499	0.807	0.873	0.917	-0.160
0.1	-1	-1.448	-0.993	0.051	0.098	0.124	-0.104
0.1	-2	-2.043	-1.979	0.361	0.262	0.445	0.430
0.15	2	2.023	1.930	-0.183	-0.204	-0.144	-0.143
0.15	1	1.433	0.986	-0.223	-0.140	-0.085	-0.018
0.15	0.5	0.830	0.497	-0.234	-0.081	-0.051	0.115
0.15	0	-0.003	0.001	-0.194	-0.007	-0.005	0.003
0.15	-0.5	-0.834	-0.497	-0.098	0.037	0.038	-0.146
0.15	-1	-1.437	-0.984	-0.013	0.044	0.069	0.000
0.15	-2	-2.027	-1.922	0.061	0.059	0.093	0.091
0.25	2	1.963	1.777	-0.061	-0.056	-0.040	-0.040
0.25	1	1.389	0.941	-0.118	-0.068	-0.043	-0.042
0.25	0.5	0.800	0.477	-0.102	-0.015	-0.010	0.052
0.25	0	-0.003	0.001	-0.094	0.008	0.006	-0.002
0.25	-0.5	-0.812	-0.480	-0.070	0.011	0.012	-0.053
0.25	-1	-1.399	-0.943	-0.017	0.019	0.025	0.024
0.25	-2	-1.965	-1.774	0.006	0.016	0.028	0.028
0.5	2	1.731	1.336	-0.029	-0.027	-0.014	-0.014
0.5	1	1.225	0.792	-0.042	-0.015	-0.009	-0.009
0.5	0.5	0.703	0.417	-0.035	0.007	0.004	0.005
0.5	0	-0.010	-0.001	-0.053	-0.001	-0.001	-0.004
0.5	-0.5	-0.713	-0.415	-0.035	0.005	0.005	0.004
0.5	-1	-1.232	-0.791	-0.018	0.005	0.005	0.005
0.5	-2	-1.737	-1.334	-0.010	0.000	0.004	0.004
1	2	1.177	0.706	-0.007	-0.002	-0.001	-0.001
1	1	0.832	0.491	-0.008	0.006	0.004	0.004
1	0.5	0.472	0.277	-0.028	-0.007	-0.003	-0.003
1	0	-0.009	0.001	-0.026	-0.001	0.000	0.001
1	-0.5	-0.488	-0.275	-0.015	0.004	0.003	0.003
1	-1	-0.840	-0.484	-0.010	0.002	0.003	0.003
1	-2	-1.180	-0.700	-0.007	-0.001	-0.002	-0.002
RMSE		1.110	0.838	1.542	1.035	1.248	1.098

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

Design		Estimator					
θ	λ	RRols	Probit	RRgmm	IVP	AGLS	Pretest
0.05	2	2.088	2.078	1.679	2.348	1.589	1.816
0.05	1	1.503	1.035	1.158	1.272	0.851	1.025
0.05	0.5	0.867	0.512	0.453	0.495	0.306	0.550
0.05	0	-0.004	-0.002	-0.226	-0.129	-0.092	-0.012
0.05	-0.5	-0.872	-0.513	-0.419	-0.218	-0.212	-0.357
0.05	-1	-1.497	-1.028	-1.393	-0.677	-0.930	-0.907
0.05	-2	-2.092	-2.069	-1.520	-0.937	-1.486	-1.720
0.1	2	2.071	2.035	0.728	1.024	0.720	0.952
0.1	1	1.486	1.023	0.426	0.504	0.323	0.658
0.1	0.5	0.863	0.508	0.387	0.414	0.261	0.363
0.1	0	-0.001	0.000	-0.119	-0.061	-0.041	-0.002
0.1	-0.5	-0.862	-0.508	-0.783	-0.425	-0.452	-0.332
0.1	-1	-1.488	-1.017	-0.438	-0.186	-0.273	-0.630
0.1	-2	-2.067	-2.021	-0.655	-0.393	-0.644	-0.850
0.15	2	2.045	1.924	0.146	0.219	0.130	0.285
0.15	1	1.463	0.999	0.224	0.293	0.188	0.436
0.15	0.5	0.848	0.499	0.076	0.098	0.059	0.231
0.15	0	-0.014	-0.008	-0.042	0.003	0.002	-0.012
0.15	-0.5	-0.852	-0.498	-0.177	-0.081	-0.081	-0.284
0.15	-1	-1.467	-0.998	-0.181	-0.068	-0.101	-0.374
0.15	-2	-2.046	-1.954	-0.212	-0.116	-0.176	-0.357
0.25	2	1.972	1.751	0.073	0.109	0.049	0.061
0.25	1	1.405	0.942	-0.002	0.019	0.015	0.119
0.25	0.5	0.821	0.480	-0.001	0.017	0.010	0.188
0.25	0	-0.008	-0.003	-0.014	0.007	0.005	-0.004
0.25	-0.5	-0.827	-0.482	-0.018	0.006	0.007	-0.187
0.25	-1	-1.411	-0.943	-0.036	-0.003	-0.008	-0.133
0.25	-2	-1.971	-1.757	0.005	0.012	0.021	0.009
0.5	2	1.667	1.220	-0.007	-0.005	-0.003	-0.003
0.5	1	1.189	0.741	-0.002	0.010	0.006	0.007
0.5	0.5	0.688	0.395	-0.011	0.001	0.002	0.059
0.5	0	-0.007	-0.003	-0.024	-0.009	-0.006	-0.004
0.5	-0.5	-0.701	-0.399	-0.016	0.001	0.001	-0.056
0.5	-1	-1.191	-0.741	-0.001	0.006	0.007	0.004
0.5	-2	-1.674	-1.225	-0.004	0.002	-0.001	-0.001
1	2	1.037	0.579	-0.004	-0.003	-0.003	-0.003
1	1	0.741	0.410	0.000	0.005	0.003	0.003
1	0.5	0.430	0.238	-0.006	0.000	0.000	0.003
1	0	-0.011	-0.004	-0.017	-0.007	-0.005	-0.004
1	-0.5	-0.440	-0.238	-0.013	-0.004	-0.004	-0.007
1	-1	-0.748	-0.412	0.000	0.003	0.005	0.005
1	-2	-1.042	-0.584	-0.009	-0.004	-0.007	-0.007
RMSE		1.107	0.828	0.279	0.243	0.216	0.31

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

Design		Estimator					
θ	λ	RRrols	Probit	RRgmm	IVP	AGLS	Pretest
0.05	2	2.028	1.998	0.597	0.834	0.535	0.752
0.05	1	1.428	1.004	0.350	0.369	0.236	0.522
0.05	0.5	0.823	0.501	0.001	-0.023	-0.012	0.171
0.05	0	0.004	0.002	0.111	0.061	0.042	0.009
0.05	-0.5	-0.821	-0.500	-0.127	-0.102	-0.101	-0.328
0.05	-1	-1.428	-1.003	-0.589	-0.306	-0.466	-0.652
0.05	-2	-2.032	-2.008	-0.494	-0.308	-0.521	-0.844
0.1	2	2.016	1.964	0.073	0.094	0.081	0.117
0.1	1	1.421	0.995	0.007	-0.011	-0.001	0.179
0.1	0.5	0.822	0.501	0.096	0.075	0.043	0.248
0.1	0	0.000	0.000	0.079	0.046	0.032	0.010
0.1	-0.5	-0.818	-0.498	0.032	0.009	0.011	-0.211
0.1	-1	-1.416	-0.990	-0.060	-0.038	-0.049	-0.218
0.1	-2	-2.017	-1.968	-0.010	-0.009	-0.013	-0.041
0.15	2	1.992	1.896	0.031	0.034	0.022	0.022
0.15	1	1.404	0.980	0.013	0.002	-0.002	0.038
0.15	0.5	0.808	0.492	-0.018	-0.030	-0.016	0.105
0.15	0	-0.001	-0.001	-0.009	-0.018	-0.013	-0.005
0.15	-0.5	-0.807	-0.492	-0.007	-0.011	-0.008	-0.141
0.15	-1	-1.402	-0.976	-0.018	-0.013	-0.018	-0.052
0.15	-2	-1.992	-1.899	0.026	0.015	0.018	0.018
0.25	2	1.918	1.717	0.009	0.010	0.008	0.008
0.25	1	1.347	0.923	-0.002	-0.010	-0.006	-0.005
0.25	0.5	0.776	0.470	0.003	-0.005	-0.002	0.035
0.25	0	0.002	0.001	0.030	0.018	0.013	0.008
0.25	-0.5	-0.778	-0.471	0.023	0.009	0.010	-0.025
0.25	-1	-1.349	-0.922	0.000	-0.002	-0.003	-0.003
0.25	-2	-1.918	-1.712	0.002	0.001	0.000	0.000
0.5	2	1.640	1.219	0.004	0.002	0.006	0.006
0.5	1	1.153	0.742	0.011	0.008	0.006	0.006
0.5	0.5	0.669	0.397	0.013	0.008	0.005	0.005
0.5	0	0.003	0.001	0.006	0.002	0.001	0.002
0.5	-0.5	-0.665	-0.395	-0.002	-0.004	-0.004	-0.004
0.5	-1	-1.155	-0.742	0.010	0.004	0.004	0.004
0.5	-2	-1.644	-1.222	0.005	0.003	0.003	0.003
1	2	1.037	0.596	0.002	0.002	0.001	0.001
1	1	0.731	0.420	-0.001	-0.003	-0.002	-0.002
1	0.5	0.422	0.242	0.001	-0.001	0.000	0.000
1	0	0.001	0.000	0.002	0.000	0.000	0.000
1	-0.5	-0.420	-0.241	0.003	0.000	0.001	0.001
1	-1	-0.728	-0.418	0.003	0.001	0.001	0.001
1	-2	-1.039	-0.595	-0.003	-0.002	-0.004	-0.004
RMSE		1.068	0.812	0.069	0.060	0.055	0.114

Table 2a Computed rejection rate for 10% nominal t-tests. Sample size is 200.
The model is just identified. The approximate proportion of 1's in each sample is .5.

Design		Estimator					
θ	λ	RRrols	Probit	RRgmm	IVP	AGLS	Pretest
0.05	2	1.000	1.000	0.092	0.087	0.117	0.805
0.05	1	1.000	1.000	0.034	0.042	0.080	0.904
0.05	0.5	0.999	0.999	0.016	0.018	0.094	0.957
0.05	0	0.104	0.108	0.007	0.009	0.103	0.197
0.05	-0.5	1.000	1.000	0.025	0.021	0.103	0.939
0.05	-1	1.000	1.000	0.046	0.032	0.094	0.903
0.05	-2	1.000	1.000	0.122	0.121	0.134	0.782
0.1	2	1.000	1.000	0.109	0.122	0.133	0.574
0.1	1	1.000	1.000	0.040	0.045	0.090	0.805
0.1	0.5	0.999	0.999	0.020	0.023	0.105	0.911
0.1	0	0.106	0.105	0.011	0.009	0.083	0.175
0.1	-0.5	0.996	0.997	0.036	0.026	0.105	0.922
0.1	-1	1.000	1.000	0.061	0.043	0.102	0.807
0.1	-2	1.000	1.000	0.125	0.119	0.120	0.591
0.15	2	1.000	1.000	0.111	0.112	0.131	0.357
0.15	1	1.000	1.000	0.056	0.061	0.125	0.718
0.15	0.5	0.999	0.998	0.028	0.042	0.104	0.876
0.15	0	0.097	0.116	0.021	0.018	0.098	0.186
0.15	-0.5	0.999	1.000	0.053	0.027	0.091	0.893
0.15	-1	1.000	1.000	0.099	0.078	0.122	0.679
0.15	-2	1.000	1.000	0.127	0.104	0.118	0.362
0.25	2	1.000	1.000	0.090	0.093	0.139	0.157
0.25	1	1.000	1.000	0.072	0.075	0.124	0.453
0.25	0.5	0.999	0.999	0.051	0.057	0.101	0.792
0.25	0	0.135	0.124	0.064	0.057	0.103	0.194
0.25	-0.5	1.000	1.000	0.070	0.048	0.105	0.790
0.25	-1	1.000	1.000	0.095	0.072	0.114	0.421
0.25	-2	1.000	1.000	0.096	0.085	0.121	0.144
0.5	2	1.000	1.000	0.108	0.096	0.150	0.150
0.5	1	1.000	1.000	0.091	0.085	0.114	0.122
0.5	0.5	0.994	0.997	0.084	0.077	0.103	0.467
0.5	0	0.102	0.099	0.089	0.088	0.095	0.163
0.5	-0.5	0.997	0.994	0.107	0.092	0.111	0.427
0.5	-1	1.000	1.000	0.117	0.090	0.103	0.106
0.5	-2	1.000	1.000	0.097	0.080	0.137	0.137
1	2	1.000	1.000	0.115	0.091	0.125	0.125
1	1	1.000	1.000	0.094	0.078	0.119	0.119
1	0.5	0.941	0.962	0.118	0.108	0.125	0.156
1	0	0.111	0.109	0.105	0.095	0.100	0.148
1	-0.5	0.974	0.963	0.110	0.101	0.119	0.170
1	-1	1.000	1.000	0.103	0.088	0.111	0.111
1	-2	1.000	1.000	0.108	0.086	0.125	0.125
RMSE		0.770	0.771	0.032	0.035	0.015	0.372

Table 2b Computed rejection rate for 10% nominal t-tests. Sample size is 1000.
The model is just identified. The approximate proportion of 1's in each sample is .5.

Design		Estimator					
θ	λ	RRrols	Probit	RRgmm	IVP	AGLS	Pretest
0.05	2	1.000	1.000	0.091	0.102	0.111	0.528
0.05	1	1.000	1.000	0.054	0.066	0.129	0.803
0.05	0.5	1.000	1.000	0.016	0.021	0.102	0.929
0.05	0	0.096	0.102	0.004	0.007	0.095	0.182
0.05	-0.5	1.000	1.000	0.036	0.021	0.105	0.929
0.05	-1	1.000	1.000	0.080	0.064	0.128	0.818
0.05	-2	1.000	1.000	0.105	0.101	0.122	0.511
0.1	2	1.000	1.000	0.059	0.065	0.101	0.136
0.1	1	1.000	1.000	0.059	0.079	0.117	0.496
0.1	0.5	1.000	1.000	0.037	0.042	0.106	0.815
0.1	0	0.094	0.095	0.055	0.038	0.107	0.183
0.1	-0.5	1.000	1.000	0.067	0.044	0.115	0.809
0.1	-1	1.000	1.000	0.097	0.074	0.101	0.521
0.1	-2	1.000	1.000	0.103	0.097	0.134	0.185
0.15	2	1.000	1.000	0.091	0.092	0.122	0.123
0.15	1	1.000	1.000	0.059	0.070	0.122	0.228
0.15	0.5	1.000	1.000	0.069	0.071	0.108	0.629
0.15	0	0.109	0.109	0.087	0.073	0.105	0.193
0.15	-0.5	1.000	1.000	0.087	0.058	0.094	0.666
0.15	-1	1.000	1.000	0.097	0.067	0.092	0.210
0.15	-2	1.000	1.000	0.103	0.095	0.120	0.120
0.25	2	1.000	1.000	0.093	0.093	0.140	0.140
0.25	1	1.000	1.000	0.090	0.086	0.127	0.127
0.25	0.5	1.000	1.000	0.088	0.078	0.113	0.348
0.25	0	0.098	0.091	0.093	0.081	0.099	0.165
0.25	-0.5	1.000	1.000	0.104	0.096	0.111	0.348
0.25	-1	1.000	1.000	0.091	0.071	0.112	0.112
0.25	-2	1.000	1.000	0.104	0.084	0.130	0.130
0.5	2	1.000	1.000	0.101	0.080	0.127	0.127
0.5	1	1.000	1.000	0.094	0.078	0.104	0.104
0.5	0.5	1.000	1.000	0.092	0.084	0.095	0.095
0.5	0	0.119	0.116	0.109	0.106	0.107	0.179
0.5	-0.5	1.000	1.000	0.099	0.087	0.102	0.102
0.5	-1	1.000	1.000	0.112	0.091	0.117	0.117
0.5	-2	1.000	1.000	0.103	0.090	0.127	0.127
1	2	1.000	1.000	0.103	0.094	0.127	0.127
1	1	1.000	1.000	0.115	0.107	0.129	0.129
1	0.5	1.000	1.000	0.110	0.099	0.112	0.112
1	0	0.103	0.091	0.093	0.083	0.083	0.124
1	-0.5	1.000	1.000	0.117	0.101	0.108	0.108
1	-1	1.000	1.000	0.113	0.093	0.126	0.126
1	-2	1.000	1.000	0.090	0.079	0.120	0.120
RMSE		0.772	0.773	0.020	0.024	0.015	0.212

Table 2c Computed rejection rate for 10% nominal t-tests. Sample size is 200.
The model is over identified. The approximate proportion of 1's in each sample is .5.

Design		Estimator					
θ	λ	RRrols	Probit	RRgmm	IVP	AGLS	Pretest
0.05	2	1.000	1.000	0.242	0.235	0.199	0.818
0.05	1	1.000	1.000	0.112	0.106	0.156	0.920
0.05	0.5	0.999	0.999	0.036	0.036	0.113	0.954
0.05	0	0.110	0.104	0.005	0.006	0.088	0.173
0.05	-0.5	1.000	1.000	0.046	0.038	0.114	0.945
0.05	-1	1.000	1.000	0.129	0.111	0.160	0.919
0.05	-2	1.000	1.000	0.242	0.221	0.185	0.821
0.1	2	1.000	1.000	0.212	0.211	0.159	0.569
0.1	1	1.000	1.000	0.125	0.121	0.135	0.823
0.1	0.5	1.000	1.000	0.065	0.066	0.119	0.925
0.1	0	0.099	0.094	0.020	0.018	0.095	0.171
0.1	-0.5	0.999	0.999	0.066	0.060	0.122	0.924
0.1	-1	1.000	1.000	0.134	0.119	0.136	0.800
0.1	-2	1.000	1.000	0.221	0.206	0.162	0.589
0.15	2	1.000	1.000	0.180	0.167	0.133	0.327
0.15	1	1.000	1.000	0.120	0.117	0.127	0.662
0.15	0.5	1.000	0.999	0.058	0.058	0.098	0.854
0.15	0	0.119	0.109	0.046	0.039	0.103	0.185
0.15	-0.5	0.999	0.999	0.083	0.068	0.110	0.861
0.15	-1	1.000	1.000	0.150	0.126	0.128	0.674
0.15	-2	1.000	1.000	0.168	0.163	0.125	0.317
0.25	2	1.000	1.000	0.138	0.137	0.117	0.125
0.25	1	1.000	1.000	0.099	0.093	0.117	0.314
0.25	0.5	0.996	0.996	0.083	0.074	0.103	0.726
0.25	0	0.128	0.121	0.071	0.072	0.097	0.187
0.25	-0.5	0.997	0.998	0.078	0.067	0.090	0.707
0.25	-1	1.000	1.000	0.105	0.084	0.095	0.326
0.25	-2	1.000	1.000	0.107	0.095	0.097	0.107
0.5	2	1.000	1.000	0.092	0.084	0.107	0.107
0.5	1	1.000	1.000	0.111	0.092	0.114	0.114
0.5	0.5	0.991	0.988	0.102	0.096	0.101	0.357
0.5	0	0.115	0.110	0.098	0.091	0.101	0.163
0.5	-0.5	0.993	0.994	0.104	0.087	0.100	0.344
0.5	-1	1.000	1.000	0.097	0.079	0.110	0.114
0.5	-2	1.000	1.000	0.130	0.105	0.142	0.142
1	2	1.000	1.000	0.103	0.083	0.109	0.109
1	1	1.000	1.000	0.111	0.099	0.114	0.114
1	0.5	0.945	0.947	0.083	0.070	0.084	0.103
1	0	0.111	0.109	0.117	0.108	0.112	0.147
1	-0.5	0.946	0.940	0.105	0.090	0.094	0.116
1	-1	1.000	1.000	0.118	0.091	0.126	0.126
1	-2	1.000	1.000	0.121	0.104	0.127	0.127
RMSE		0.77	0.769	0.038	0.036	0.023	0.35

Table 2d Computed rejection rate for 10% nominal t-tests. Sample size is 1000.
The model is over identified. The approximate proportion of 1's in each sample is .5.

Design		Estimator					
θ	λ	RRrols	Probit	RRgmm	IVP	AGLS	Pretest
0.05	2	1.000	1.000	0.204	0.197	0.154	0.546
0.05	1	1.000	1.000	0.111	0.106	0.133	0.798
0.05	0.5	1.000	1.000	0.047	0.046	0.112	0.916
0.05	0	0.121	0.119	0.016	0.017	0.119	0.217
0.05	-0.5	1.000	1.000	0.047	0.050	0.110	0.920
0.05	-1	1.000	1.000	0.112	0.106	0.128	0.792
0.05	-2	1.000	1.000	0.180	0.173	0.140	0.527
0.1	2	1.000	1.000	0.123	0.119	0.113	0.141
0.1	1	1.000	1.000	0.104	0.097	0.123	0.442
0.1	0.5	1.000	1.000	0.072	0.074	0.101	0.776
0.1	0	0.111	0.109	0.045	0.051	0.096	0.184
0.1	-0.5	1.000	1.000	0.061	0.062	0.101	0.775
0.1	-1	1.000	1.000	0.111	0.114	0.139	0.437
0.1	-2	1.000	1.000	0.114	0.110	0.105	0.126
0.15	2	1.000	1.000	0.119	0.115	0.118	0.118
0.15	1	1.000	1.000	0.097	0.093	0.134	0.184
0.15	0.5	1.000	1.000	0.074	0.074	0.108	0.551
0.15	0	0.096	0.101	0.061	0.074	0.100	0.174
0.15	-0.5	1.000	1.000	0.092	0.089	0.115	0.592
0.15	-1	1.000	1.000	0.085	0.082	0.098	0.149
0.15	-2	1.000	1.000	0.094	0.088	0.121	0.121
0.25	2	1.000	1.000	0.110	0.097	0.121	0.121
0.25	1	1.000	1.000	0.110	0.103	0.128	0.128
0.25	0.5	1.000	1.000	0.102	0.105	0.118	0.243
0.25	0	0.090	0.084	0.095	0.099	0.106	0.167
0.25	-0.5	1.000	1.000	0.100	0.104	0.109	0.234
0.25	-1	1.000	1.000	0.087	0.081	0.096	0.096
0.25	-2	1.000	1.000	0.100	0.097	0.141	0.141
0.5	2	1.000	1.000	0.112	0.100	0.143	0.143
0.5	1	1.000	1.000	0.092	0.088	0.117	0.117
0.5	0.5	1.000	1.000	0.088	0.095	0.096	0.096
0.5	0	0.093	0.099	0.078	0.088	0.091	0.147
0.5	-0.5	1.000	1.000	0.104	0.106	0.114	0.114
0.5	-1	1.000	1.000	0.112	0.107	0.135	0.135
0.5	-2	1.000	1.000	0.113	0.100	0.148	0.148
1	2	1.000	1.000	0.100	0.093	0.123	0.123
1	1	1.000	1.000	0.095	0.085	0.110	0.110
1	0.5	1.000	1.000	0.097	0.106	0.109	0.109
1	0	0.092	0.097	0.087	0.101	0.099	0.118
1	-0.5	1.000	1.000	0.109	0.106	0.108	0.108
1	-1	1.000	1.000	0.112	0.108	0.122	0.122
1	-2	1.000	1.000	0.116	0.100	0.132	0.132
RMSE		0.773	0.773	0.021	0.018	0.019	0.194

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

Design		Estimator					
θ	λ	RRrols	Probit	RRgmm	IVP	AGLS	Pretest
0.05	2	0.004	0.009	7.498	9.167	11.443	0.635
0.05	1	0.005	0.005	1.713	2.096	1.526	1.261
0.05	0.5	0.005	0.004	20.472	4.215	2.695	0.271
0.05	0	0.006	0.004	2.695	1.961	1.379	0.184
0.05	-0.5	0.005	0.004	36.744	18.718	18.955	0.220
0.05	-1	0.005	0.005	1.131	0.619	0.892	0.308
0.05	-2	0.004	0.010	1.660	1.078	1.429	0.937
0.1	2	0.004	0.009	0.706	0.959	0.831	0.812
0.1	1	0.005	0.005	1.793	1.886	1.138	0.748
0.1	0.5	0.006	0.004	1.261	0.988	0.521	0.174
0.1	0	0.006	0.003	3.244	2.133	1.509	0.309
0.1	-0.5	0.005	0.004	5.805	4.045	3.874	3.846
0.1	-1	0.005	0.005	2.590	2.020	1.187	0.278
0.1	-2	0.004	0.010	2.418	1.423	3.250	3.232
0.15	2	0.004	0.009	2.039	2.578	1.702	0.589
0.15	1	0.005	0.005	0.400	0.420	0.302	0.175
0.15	0.5	0.005	0.004	1.236	1.041	0.697	0.694
0.15	0	0.006	0.003	0.362	0.252	0.178	0.079
0.15	-0.5	0.005	0.004	0.480	0.259	0.244	0.201
0.15	-1	0.005	0.005	0.850	0.449	0.572	0.536
0.15	-2	0.004	0.010	1.900	1.165	2.058	2.041
0.25	2	0.004	0.008	0.107	0.155	0.098	0.098
0.25	1	0.004	0.005	0.041	0.045	0.028	0.032
0.25	0.5	0.005	0.004	0.396	0.351	0.188	0.187
0.25	0	0.006	0.004	0.036	0.030	0.022	0.013
0.25	-0.5	0.005	0.004	0.065	0.043	0.043	0.043
0.25	-1	0.005	0.005	0.092	0.048	0.074	0.075
0.25	-2	0.004	0.009	0.065	0.042	0.068	0.065
0.5	2	0.004	0.006	0.018	0.025	0.017	0.017
0.5	1	0.004	0.004	0.016	0.018	0.011	0.012
0.5	0.5	0.005	0.004	0.014	0.014	0.009	0.011
0.5	0	0.005	0.003	0.014	0.011	0.008	0.006
0.5	-0.5	0.005	0.003	0.014	0.009	0.009	0.011
0.5	-1	0.004	0.004	0.015	0.008	0.011	0.011
0.5	-2	0.004	0.006	0.017	0.011	0.018	0.018
1	2	0.003	0.003	0.008	0.011	0.008	0.008
1	1	0.004	0.003	0.007	0.008	0.005	0.005
1	0.5	0.004	0.003	0.007	0.007	0.004	0.005
1	0	0.004	0.003	0.007	0.006	0.004	0.003
1	-0.5	0.004	0.003	0.007	0.004	0.004	0.005
1	-1	0.004	0.003	0.007	0.004	0.005	0.005
1	-2	0.003	0.003	0.008	0.005	0.008	0.008

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

Design		Estimator					
θ	λ	RRrols	Probit	RRgmm	IVP	AGLS	Pretest
0.05	2	0.002	0.004	0.867	1.165	0.913	0.898
0.05	1	0.002	0.002	1.196	1.216	0.740	0.622
0.05	0.5	0.002	0.002	2.509	1.744	1.129	0.360
0.05	0	0.003	0.001	1.350	0.909	0.633	0.180
0.05	-0.5	0.002	0.002	6.289	3.392	3.044	0.124
0.05	-1	0.002	0.002	4.013	2.267	2.949	1.456
0.05	-2	0.002	0.004	45.241	28.925	38.530	38.521
0.1	2	0.002	0.004	0.144	0.198	0.186	0.186
0.1	1	0.002	0.002	1.195	1.243	0.817	0.816
0.1	0.5	0.002	0.002	0.050	0.047	0.029	0.027
0.1	0	0.003	0.001	0.082	0.059	0.042	0.037
0.1	-0.5	0.002	0.002	0.981	0.817	0.853	0.038
0.1	-1	0.002	0.002	0.065	0.037	0.050	0.037
0.1	-2	0.002	0.004	0.144	0.094	0.138	0.132
0.15	2	0.002	0.004	0.031	0.043	0.028	0.028
0.15	1	0.002	0.002	0.027	0.029	0.018	0.021
0.15	0.5	0.002	0.002	0.023	0.023	0.014	0.016
0.15	0	0.003	0.001	0.024	0.019	0.014	0.010
0.15	-0.5	0.002	0.002	0.022	0.014	0.014	0.016
0.15	-1	0.002	0.002	0.024	0.012	0.017	0.020
0.15	-2	0.002	0.004	0.026	0.017	0.025	0.025
0.25	2	0.002	0.003	0.015	0.021	0.014	0.014
0.25	1	0.002	0.002	0.014	0.016	0.010	0.010
0.25	0.5	0.002	0.002	0.013	0.013	0.008	0.010
0.25	0	0.003	0.001	0.012	0.010	0.007	0.005
0.25	-0.5	0.002	0.002	0.012	0.008	0.008	0.010
0.25	-1	0.002	0.002	0.013	0.007	0.009	0.009
0.25	-2	0.002	0.003	0.014	0.009	0.014	0.014
0.5	2	0.002	0.002	0.007	0.010	0.007	0.007
0.5	1	0.002	0.002	0.006	0.007	0.004	0.004
0.5	0.5	0.002	0.001	0.006	0.006	0.004	0.004
0.5	0	0.003	0.001	0.006	0.005	0.003	0.003
0.5	-0.5	0.002	0.001	0.006	0.004	0.004	0.004
0.5	-1	0.002	0.002	0.006	0.003	0.004	0.004
0.5	-2	0.002	0.002	0.007	0.004	0.007	0.007
1	2	0.001	0.001	0.003	0.005	0.003	0.003
1	1	0.002	0.001	0.003	0.004	0.002	0.002
1	0.5	0.002	0.001	0.003	0.003	0.002	0.002
1	0	0.002	0.001	0.003	0.002	0.002	0.001
1	-0.5	0.002	0.001	0.003	0.002	0.002	0.002
1	-1	0.002	0.001	0.003	0.002	0.002	0.002
1	-2	0.001	0.001	0.003	0.002	0.003	0.003

Table 3c Monte Carlo Standard Errors of each estimator based on samples of size 200.
The model is over identified. The approximate proportion of 1's in each sample is .5.

Design		Estimator					
θ	λ	RRrols	Probit	RRgmm	IVP	AGLS	Pretest
0.05	2	0.004	0.010	0.113	0.155	0.102	0.089
0.05	1	0.005	0.005	0.122	0.134	0.079	0.061
0.05	0.5	0.005	0.004	0.179	0.161	0.100	0.064
0.05	0	0.006	0.003	0.120	0.091	0.065	0.048
0.05	-0.5	0.005	0.004	0.173	0.097	0.100	0.051
0.05	-1	0.005	0.005	0.200	0.104	0.130	0.080
0.05	-2	0.004	0.009	0.108	0.067	0.105	0.086
0.1	2	0.004	0.009	0.100	0.144	0.104	0.102
0.1	1	0.005	0.005	0.178	0.192	0.098	0.082
0.1	0.5	0.005	0.004	0.084	0.076	0.048	0.034
0.1	0	0.006	0.003	0.092	0.070	0.049	0.027
0.1	-0.5	0.005	0.004	0.483	0.276	0.297	0.028
0.1	-1	0.005	0.005	0.102	0.049	0.066	0.054
0.1	-2	0.004	0.010	0.129	0.078	0.118	0.114
0.15	2	0.004	0.008	0.067	0.090	0.064	0.067
0.15	1	0.005	0.005	0.069	0.072	0.044	0.035
0.15	0.5	0.005	0.004	0.072	0.068	0.044	0.026
0.15	0	0.006	0.003	0.059	0.044	0.032	0.021
0.15	-0.5	0.005	0.004	0.056	0.032	0.032	0.025
0.15	-1	0.005	0.005	0.059	0.028	0.039	0.038
0.15	-2	0.004	0.009	0.066	0.042	0.067	0.066
0.25	2	0.004	0.008	0.030	0.040	0.028	0.028
0.25	1	0.005	0.005	0.027	0.029	0.018	0.021
0.25	0.5	0.006	0.004	0.027	0.025	0.015	0.016
0.25	0	0.006	0.003	0.027	0.020	0.014	0.010
0.25	-0.5	0.005	0.004	0.026	0.015	0.015	0.016
0.25	-1	0.005	0.005	0.028	0.013	0.018	0.022
0.25	-2	0.004	0.008	0.032	0.020	0.030	0.030
0.5	2	0.004	0.005	0.014	0.019	0.013	0.013
0.5	1	0.004	0.004	0.013	0.014	0.009	0.009
0.5	0.5	0.005	0.003	0.013	0.012	0.007	0.009
0.5	0	0.005	0.003	0.012	0.009	0.007	0.005
0.5	-0.5	0.005	0.003	0.013	0.007	0.007	0.009
0.5	-1	0.004	0.004	0.013	0.006	0.009	0.009
0.5	-2	0.004	0.005	0.015	0.009	0.014	0.014
1	2	0.003	0.003	0.007	0.009	0.006	0.006
1	1	0.004	0.002	0.006	0.007	0.004	0.004
1	0.5	0.004	0.002	0.006	0.005	0.003	0.004
1	0	0.004	0.002	0.006	0.005	0.003	0.003
1	-0.5	0.004	0.002	0.006	0.003	0.003	0.004
1	-1	0.003	0.002	0.007	0.003	0.004	0.004
1	-2	0.003	0.003	0.007	0.004	0.007	0.007

Table 3d Monte Carlo Standard Errors of each estimator based on samples of size 1000.
The model is over identified. The approximate proportion of 1's in each sample is .5.

Design		Estimator					
θ	λ	RRrols	Probit	RRgmm	IVP	AGLS	Pretest
0.05	2	0.002	0.004	0.106	0.150	0.114	0.108
0.05	1	0.002	0.002	0.074	0.084	0.052	0.047
0.05	0.5	0.002	0.002	0.247	0.238	0.142	0.136
0.05	0	0.003	0.002	0.075	0.062	0.044	0.029
0.05	-0.5	0.002	0.002	0.074	0.045	0.045	0.033
0.05	-1	0.002	0.002	0.099	0.049	0.062	0.052
0.05	-2	0.002	0.004	0.090	0.056	0.079	0.075
0.1	2	0.002	0.004	0.031	0.043	0.029	0.031
0.1	1	0.002	0.002	0.043	0.048	0.029	0.032
0.1	0.5	0.002	0.002	0.031	0.031	0.019	0.018
0.1	0	0.003	0.001	0.031	0.025	0.018	0.011
0.1	-0.5	0.002	0.002	0.031	0.019	0.018	0.018
0.1	-1	0.002	0.002	0.031	0.015	0.022	0.025
0.1	-2	0.002	0.004	0.038	0.023	0.035	0.036
0.15	2	0.002	0.004	0.021	0.029	0.020	0.020
0.15	1	0.002	0.002	0.021	0.023	0.014	0.016
0.15	0.5	0.002	0.002	0.019	0.019	0.012	0.014
0.15	0	0.003	0.001	0.018	0.015	0.010	0.007
0.15	-0.5	0.002	0.002	0.019	0.012	0.012	0.014
0.15	-1	0.002	0.002	0.020	0.010	0.013	0.015
0.15	-2	0.002	0.004	0.022	0.014	0.020	0.020
0.25	2	0.002	0.003	0.012	0.017	0.011	0.011
0.25	1	0.002	0.002	0.012	0.013	0.008	0.008
0.25	0.5	0.002	0.002	0.011	0.011	0.007	0.008
0.25	0	0.002	0.001	0.011	0.009	0.006	0.004
0.25	-0.5	0.002	0.002	0.011	0.007	0.007	0.008
0.25	-1	0.002	0.002	0.011	0.005	0.008	0.008
0.25	-2	0.002	0.003	0.013	0.008	0.012	0.012
0.5	2	0.002	0.002	0.006	0.009	0.006	0.006
0.5	1	0.002	0.002	0.006	0.006	0.004	0.004
0.5	0.5	0.002	0.001	0.005	0.005	0.003	0.003
0.5	0	0.002	0.001	0.005	0.004	0.003	0.002
0.5	-0.5	0.002	0.001	0.005	0.003	0.003	0.003
0.5	-1	0.002	0.002	0.006	0.003	0.004	0.004
0.5	-2	0.002	0.002	0.006	0.004	0.006	0.006
1	2	0.001	0.001	0.003	0.004	0.003	0.003
1	1	0.002	0.001	0.003	0.003	0.002	0.002
1	0.5	0.002	0.001	0.003	0.003	0.002	0.002
1	0	0.002	0.001	0.002	0.002	0.001	0.001
1	-0.5	0.002	0.001	0.003	0.002	0.002	0.002
1	-1	0.002	0.001	0.003	0.001	0.002	0.002
1	-2	0.001	0.001	0.003	0.002	0.003	0.003

Table 4a: Comparison of AGLS and ML. Sample size = 200, model just identified.
Upper panel compares the percentiles of the computed t-ratio and its summary statistics.
Lower panel compares the percentiles to the p-value of the corresponding t-ratio.

λ	-0.25		-2		-0.25		-2	
θ	0.15		0.15		1		1	
	AGLS	ML	AGLS	ML	AGLS	ML	AGLS	ML
1%	-1.854	-1.78E+01	-2.984	-7.583	-2.233	-2.666	-2.489	-2.217
5%	-1.329	-6.453	-2.189	-3.227	-1.566	-1.677	-1.686	-1.441
10%	-1.074	-2.880	-1.724	-2.203	-1.244	-1.284	-1.265	-1.108
25%	-0.534	-0.817	-0.873	-0.920	-0.599	-0.601	-0.543	-0.509
50%	0.032	0.042	-0.130	-0.157	0.098	0.099	0.163	0.168
75%	0.562	1.117	0.233	0.516	0.810	0.877	0.708	0.800
90%	0.901	2.561	0.429	1.267	1.279	1.500	1.199	1.535
95%	1.061	3.797	0.512	1.769	1.603	1.958	1.432	1.918
99%	1.425	7.130	0.688	2.583	2.166	3.173	1.792	2.801
	Summary Statistics for the t-ratio and p-value for a test for normality							
Variance	0.567	13.113	0.736	3.057	0.964	1.325	0.910	1.092
Skewness	-0.304	-2.375	-1.202	-1.801	-0.123	0.204	-0.584	0.200
Kurtosis	2.578	12.801	4.013	12.317	2.606	3.755	3.326	3.584
W (p-value	<.0001	<.0001	<.0001	<.0001	0.00432	0.00012	0.0001	0.0029
	5% and 10% percentiles of the p-value for the two-sided t-test							
5%	0.154	0.000	0.029	3.74E-04	0.058	0.019	0.064	0.034
10%	0.227	0.000	0.085	1.42E-02	0.114	0.067	0.119	0.084

Table 4b: Comparison of AGLS and ML. Sample size = 1000, model just identified.
Upper panel compares the percentiles of the computed t-ratio and its summary statistics.
Lower panel compares the percentiles to the p-value of the corresponding t-ratio.

λ	-0.25		-2		-0.25		-2	
θ	0.25		0.25		1		1	
	AGLS	ML	AGLS	ML	AGLS	ML	AGLS	ML
1%	-2.327	-2.81E+00	-3.055	-2.105	-2.367	-2.413	-2.507	-2.233
5%	-1.666	-1.782	-1.922	-1.401	-1.670	-1.679	-1.619	-1.495
10%	-1.326	-1.366	-1.536	-1.190	-1.321	-1.319	-1.263	-1.183
25%	-0.631	-0.634	-0.672	-0.587	-0.599	-0.596	-0.628	-0.606
50%	0.013	0.001	0.018	0.019	0.024	0.024	0.104	0.104
75%	0.716	0.761	0.598	0.719	0.754	0.769	0.702	0.734
90%	1.216	1.434	0.979	1.423	1.317	1.380	1.227	1.337
95%	1.438	1.840	1.183	1.923	1.635	1.739	1.602	1.802
99%	1.941	2.759	1.387	2.741	2.242	2.467	2.145	2.541
	Summary Statistics for the t-ratio and p-value for a test for normality							
Variance	0.926	1.266	0.968	1.032	1.041	1.107	0.963	0.996
Skewness	-0.237	0.050	-0.910	0.381	-0.104	0.009	-0.316	0.062
Kurtosis	2.732	3.665	3.956	3.386	2.957	3.160	3.202	3.139
W (p-value	0.02905	0.0067	<.0001	<.0001	0.5446	0.8239	0.0001	0.3145
	5% and 10% percentiles of the p-value for the two-sided t-test							
5%	0.055	0.023	0.055	0.037	0.046	0.040	0.058	0.048
10%	0.100	0.070	0.123	0.090	0.098	0.088	0.107	0.094