Political Analysis, 8:2

Two-Stage Estimation of Nonrecursive Choice Models

R. Michael Alvarez and Garrett Glasgow
California Institute of Technology

Questions of causation are important issues in empirical research on political behavior. Most of the discussion of the econometric problems associated with multiequation models with reciprocal causation has focused on models with continuous dependent variables (e.g., Markus and Converse 1979; Page and Jones 1979). Yet, many models of political behavior involve discrete or dichotomous dependent variables; this paper describes two techniques which can consistently estimate reciprocal relationships between dichotomous and continuous dependent variables. The first, two-stage probit least squares (2SPLS), is very similar to two-stage instrumental variable techniques. The second, two-stage conditional maximum likelihood (2SCML), may overcome problems associated with 2SPLS, but has not been used in the political science literature. We demonstrate the potential pitfalls of ignoring the problems of reciprocal causation in nonrecursive choice models and examine the properties of both techniques using Monte Carlo simulations: we find that 2SPLS slightly outperforms 2SCML in terms of bias but that 2SCML produces more accurate standard errors. However, the 2SCML model offers an explicit statistical test for endogeneity. The results from our simulations, and the statistical test for exogeneity, lead us to advocate the use of 2SCML for estimation of this class of causal models. We then apply both of these techniques to an empirical example focusing on the relationship between voter preferences in a presidential election and the voter's uncertainty about the policy positions taken by the candidates. This example demonstrates the importance of these techniques for political science research.

1 Introduction

MANY INTERESTING ASPECTS of political behavior involve dichotomous decisions. For example, a potential voter decides whether to go to the polls on election day (Wolfinger and Rosenstone 1980), activists decide to donate time or resources to a campaign (Verba and Nie 1972), and candidates decide to enter particular races in certain political contexts (Jacobson and Kernell 1981). One of the most studied dichotomous choices, though, occurs once a citizen has entered the voting booth, since in American national elections voters essentially have two ways to cast their ballot—for Democratic or Republican candidates. While sometimes there are other viable choices in each of these examples, much of the empirical research in political behavior has examined binary choices.

Authors' note: John Aldrich, John Brehm, Tara Butterfield, Jonathan Katz, Eric Lawrence, and Jonathan Nagler provided important comments and advice. Abby Delman and Gail Nash gave invaluable assistance. The John M. Olin Foundation provided support for part of this research, as did the National Science Foundation through Grant SBR-9709372.

Copyright 1999 by the Society for Political Methodology

The practical econometric difficulties associated with dichotomous dependent variables are now well known in political research. Given that ordinary least squares does not perform well when the dependent variable is binary, researchers now turn to linear probability, logit, or probit models (Achen 1986; Aldrich and Nelson 1984). In either framework, under certain assumptions, the dichotomous nature of the dependent variable is not an obstacle to unbiased estimation of model coefficients.

However, researchers using these dichotomous dependent variable models have not incorporated them adequately into larger nonrecursive models of political behavior. In fact, the prominent examples of nonrecursive models in the literature either have introduced surrogate variables for binary candidate choices (e.g., Page and Jones 1979) or have resorted to least squares estimation of a discrete choice model (e.g., Markus and Converse 1979). In only a few instances have researchers dealt with the problems of endogeneity in discrete choice models in political science, with the most notable being models of party identification (Franklin and Jackson 1983; Fiorina 1981), interest group lobbying (Caldeira and Wright 1998), and tax compliance (Schotz and Lubell 1998), all estimated using some type of two-stage technique. But little is known about the statistical properties of two-stage estimation of models involving discrete and continuous dependent variables, making it difficult to assess the reliability of the published empirical results using two-stage estimation techniques.

We begin the next section of this paper with an examination of two techniques which can be used to estimate reciprocal relationships between dichotomous and continuous dependent variables. One technique, two-stage probit least squares (2SPLS), is similar to two-stage instrumental variable techniques. The second technique, two-stage conditional maximum likelihood (2SCML), might alleviate some of the shortcomings of 2SPLS, but has not seen widespread use in the political science literature. We then examine the properties of both 2SPLS and 2SCML through Monte Carlo simulations and compare these results to those obtained with a model that does not correct for endogeneity. Finally, we show the applicability of both models to a problem of contemporary interest.

2 Nonrecursive Two-Stage Choice Models

We begin this discussion with a simple two-variable nonrecursive system:[1]

$$y_1^* = \gamma_1 y_2 + \beta_1 X_1 + \mu_1 \tag{1}$$
$$y_2 = \gamma_2 y_1^* + \beta_2 X_2 + \mu_2 \tag{2}$$

where y_2 is a continuous variable, X_1 and X_2 are independent variables, μ_1 and μ_2 are error terms, and γ and β are parameters to be estimated. We do not directly observe the value of y_1^*, instead observing:

$$y_1 = \begin{cases} 1 & \text{if} \quad y_1^* > 0 \\ 0 & \text{if} \quad y_1^* \leq 0 \end{cases}$$

[1] The variables in these and the following equations should properly be subscripted by i. To simplify notation we drop the subscript i without loss of generality.

From these, the analogues of the reduced-form equations are[2]

$$y_1 = \pi_1 X_1 + \pi_2 X_2 + \nu_1 \tag{3}$$
$$y_2 = \pi_3 X_1 + \pi_4 X_2 + \nu_2 \tag{4}$$

With this simple model, if the usual assumptions for the probit model held for Eq. (1) and the usual ordinary least-squares (OLS) assumptions held for Eq. (2), independent estimation of each equation would produce consistent estimates. However, this implies the following restrictions on the model:

$$E\left(r_{y_2\mu_1}\right) = \frac{1}{n} E\left[y_2\mu_1\right] = 0$$
$$E\left(r_{y_1^*\mu_2}\right) = \frac{1}{n} E[y_1^*\mu_2] = 0$$

Only if the endogenous variable on the right-hand side of each equation is uncorrelated with the error term in that equation might OLS or probit produce consistent estimates of the coefficients of interest in either equation.

In practice it will be difficult for these assumptions to be met. If the model in Eqs. (1) and (2) is *fully recursive* (meaning that both γ_1 and γ_2 are nonzero), these assumptions will never be met, even if the errors across equations are uncorrelated [i.e., $E(\mu_1\mu_2) = 0$]. This is easily demonstrated by simply substituting for the endogenous variable on the right-hand side of either equation; the dependent variable of that equation will always be a function of the error term of the other equation.

However, even if the model is assumed to be hierarchical (either γ_1 or γ_2 is zero), it is still unlikely that these assumptions will be met. First, if common factors are left out of the specification of the model, and these factors influence each dependent variable, then these restrictions will be violated. Note that a hierarchical model still requires great confidence in the "correct" specification of both equations; if even one variable is left out of the right-hand side of each equation, estimation of these equations by OLS or probit will yield biased coefficients. Second, if the endogenous variables are not correctly measured, that measurement error can itself lead to the violation of these assumptions. Thus even in a hierarchical model there is good reason to be concerned about the violation of these assumptions.

Therefore, application of probit or OLS models to Eqs. (1) and (2) without correcting for endogeneity will most likely result in biased coefficients and incorrect inferences about the actual political phenomenon under analysis. Fortunately, two-stage estimation of models which deal with endogeneity have been discussed in the literature, with two techniques advocated (Achen 1986; Amemiya 1978; Maddala 1983). In the first approach, two-stage probit least squares (2SPLS), reduced-form equations for each endogenous variable are estimated initially. This method can be applied to either a binary dependent variable with a continuous endogenous regressor on the right-hand side or a continuous dependent variable with a binary endogenous regressor on the right-hand side. The reduced-form equation for the continuous variable (Eq. 4) is estimated in the usual fashion, using ordinary least squares, while the reduced-form equation for the binary choice variable (Eq. 3) is estimated

[2] Note that Eqs. (3) and (4) are not reduced form equations in the usual sense, since they cannot be directly derived from Eqs. (1) and (2).

via probit analysis. The parameters from the reduced-form equations are then used to generate a predicted value for each endogenous variable, and these predicted values are then substituted for each endogenous variable as they appear on the right-hand side of the respective equation (i.e., Eqs. (1) and (2)).[3] Then the equations are estimated, with the predicted values from the reduced-forms serving as instruments on the right-hand sides of the equations. It has been shown that the estimates obtained in this second stage are consistent (Achen 1986; Amemiya 1978).[4]

But there is a potential problem with the 2SPLS approach—the estimated standard errors are likely to be biased. If the second-stage estimation has a continuous dependent variable, the standard errors can be easily corrected by multiplying the estimated standard errors by an appropriate weighting factor, as summarized by Achen (1986, p. 43). This weighting correction is simple to implement. Call the variance of the residuals from the second-stage continuous variable regression $\sigma^2_{\varepsilon_P}$. Then compute the variance of a slightly different set of residuals using the continuous variable coefficients estimated in the second stage, but after substituting the actual value of the endogeneous right-hand side variable for the values calculated from the instrumental variable regression; call this residual variance $\sigma^2_{\varepsilon_U}$. Then each standard error in the second-stage continuous equation should be multiplied by $\sqrt{\sigma^2_{\varepsilon_P}/\sigma^2_{\varepsilon_U}}$. These standard errors are superior to the uncorrected standard errors (Achen 1986).

Unfortunately there is no simple correction for the coefficient standard errors when the second stage estimation involves a binary choice equation (Achen 1986, p. 49). The asymptotic covariance matrix of the probit estimates has been derived by Amemiya (1978), but is exceptionally complex and computationally difficult. Indeed, those in the political science literature who have utilized the 2SPLS methodology have been willing to settle with consistent estimates and possibly incorrect standard errors, due to this computational difficulty (Fiorina 1981; Franklin and Jackson 1983).

Yet it is important to estimate reliable standard errors for statistical inference about the results. The second estimation technique advanced in the literature may mitigate the problems with incorrect standard errors, so that corrections to the coefficient standard errors may not be necessary. Rivers and Vuong (1988) developed what they term the two-stage conditional maximum-likelihood (2SCML) approach to obtaining consistent and asymptotically efficient estimates for the probit equation. Unlike 2SPLS, which allows the dependent variable in the second-stage equation to be either binary or continuous, the 2SCML approach assumes interest in only the structural parameters of the probit equations.[5] To estimate the probit coefficients and their variances in the 2SCML method, first estimate the reduced form for the continuous variable equation, obtain the residuals from the reduced form regression,

[3] Note that the predicted value from the probit reduced-form is the linear predictor, $\mathbf{X}_i\beta$, not a transformed probability for each voter.

[4] The use of two-stage, or limited-information models, instead of full-information models, can be justified on two grounds. First, limited-information models are easier to estimate and interpret than their full-information counterparts. Derivation of a full-information likelihood function for the model presented later in this paper yielded an exceptionally complex function, which made estimation computationally difficult [an example of the complexity of the FIML case is given by King (1989, Section 8.2)]. Second, full-information models, while theoretically more efficient since they utilize information in the data more fully, can be quite problematic if even one of the equations in the model is misspecified since the biases associated with specification errors will be distributed throughout the model. Limited-information models are not problematic in this regard, since they ignore information about the joint distribution of the error terms across the equations, which leads to a loss of potential efficiency.

[5] This limitation is not problematic, since it is possible using the usual 2SPLS method to obtain consistent and efficient estimates of the coefficients in the continuous variable equations.

and add these residuals to the probit equation for the binary choice variable as an additional variable with a corresponding parameter to be estimated.

Rivers and Vuong demonstrate a number of useful properties of the 2SCML approach, properties which might make it the preferred estimator for this class of models. First, they show that 2SCML produces consistent and asymptotically efficient estimates. While there is no clear efficiency ordering among the various two-stage estimators they examine, the evidence they provide on the finite sample properties of 2SCML indicates that it is at least as efficient as other simultaneous probit estimators under general conditions. Second, Rivers and Vuong discuss an extremely useful property of the 2SCML model—it provides a practical means of testing the hypothesis of exogeneity.

They show that tests analogous to the usual Wald, likelihood-ratio, and Lagrange multiplier tests can be constructed for exogeneity in the 2SCML model. In particular, the likelihood-ratio test is easy to implement and is perhaps the best known of the tests to the political science community. It is computed as

$$\text{LR} = -2(\ln \hat{L}_R - \ln \hat{L}_U) \tag{5}$$

where $\hat{L}_R$ is the log-likelihood function evaluated at the restricted estimates (probit without the regression reduced-form residuals on the right-hand side) and $\hat{L}_U$ is the log-likelihood at the unrestricted estimates (probit with the regression reduced-form residuals on the right-hand side). Rivers and Vuong show that this test has a chi-square distribution with degrees of freedom equal to the number of endogenous variables in the probit equation.

The utility of this test for exogeneity cannot be overstated. First, it is simple to estimate; only the log-likelihoods from two probit models are necessary. Second, this test is not available for the other estimators which have been suggested for these models, including 2SPLS. Finally, remember that models with binary dependent variables do not have "residuals" like models with continuous dependent variables. Without residuals, diagnosing violations of assumptions such as autocorrelation, heteroskedasticity, and endogeneity are extremely difficult; hence the utility of this likelihood-ratio test statistic.

Thus, there are two different techniques which can be used to consistently estimate the coefficients for the binary choice equation—2SPLS and 2SCML. The major difference between the two techniques is that the 2SCML technique may produce standard error estimates which are more accurate than 2SPLS; both should provide consistent coefficient estimates. Also, 2SCML provides a test for exogeneity. So, by estimating the binary choice equation using both techniques, greater confidence in the second-stage probit estimates of parameters and standard errors is possible, especially if the two methodologies produce similar results.

While these techniques for nonrecursive choice models should be employed frequently in political science research, they are not. The 2SPLS estimator has seen limited applications in political science, while the 2SCML estimator has appeared only once to date in published political science work (Alvarez 1998). Also, little is known about the performance of these estimation techniques, other than the Monte Carlo work by Rivers and Vuong (1988). The following section evaluates the performance of these two estimation techniques for nonrecursive choice models with results from Monte Carlo simulations and compares them to the performance of a probit model that does not correct for endogeneity. Then we present a substantive example of a problem in which endogeneity is suspected in a system of equations. The application of both techniques to this problem, focusing on the relationship between voter choice and voter uncertainty about candidate policy stances, underscores the importance of these techniques for political science research.

3 Monte Carlo Results

In order to examine the potential bias in our coefficient estimates when endogeneity is not corrected for, and to examine the relative properties of those estimation techniques which correct for endogeneity, we performed a set of Monte Carlo simulations. The simulations were based on 1000 replications of 700 and 10,000 observation datasets, with this first sample size (700) chosen to approximate the sample size of datasets often employed in political science research and the second (10,000) to probe the large-sample properties of these models. The "true model" in each Monte Carlo simulation was

$$y_1^* = 2y_2 + 1.5x_3 \tag{6}$$

$$y_2 = x_1 + 2x_2 \tag{7}$$

where each X is randomly drawn from a normal distribution.[6] Then, an error term for the first equation (ε_1) was drawn from a normal distribution (mean zero, unit standard deviation). An error term for the second equation was constructed by drawing an identical normal variate and transforming it:

$$\varepsilon_2 = \lambda(\varepsilon_1) + N(1, 0) \tag{8}$$

where $N(1, 0)$ is the newly drawn random variate, ε_1 is the error for the first equation, and ε_2 is the error term for the second equation. By changing the values of λ we simulate different degrees of correlation between these two error terms and, thus, examine the effects of different error correlations on the distribution of model estimates.

In these Monte Carlo simulations we focus on the estimation of a binary probit model when there is a continuous endogenous variable on the right-hand side of the model. We coded y_1 as 1 if y_1^* was above the mean value of y_1^*, and 0 otherwise; y_2 is a continuous variable. Three techniques were used to estimate Eq. (6): probit, 2SPLS, and 2SCML. The estimation of the probit model for Eq. (6) was performed without considering possible endogeneity; we refer to this model as the "naive" probit model. For the 2SPLS and 2SCML estimators an instrumental variables regression for y_2 was first estimated, using the three independent variables, and predicted values for y_2 were calculated from the reduced-form estimates. 2SPLS and 2SCML equations were then estimated for y_1, with the appropriate instrument for y_2 from the reduced-form equations substituted on the right-hand side of the equation for 2SPLS or with the reduced-form residuals added to the right-hand side for 2SCML. This procedure, starting from the drawing of the first error term, was then replicated 1000 times for 13 values of λ (producing error correlations ranging from $-.95$ to .95), and summary statistics for the coefficients of these replications were calculated. The exact procedures used in the Monte Carlo simulations are described in the expanded version of this paper, which is available at the Political Analysis website.

The results of these Monte Carlo simulations allow us to compare the relative coefficient bias due to endogeneity in a model that does not correct for endogeneity (probit) with two models that do correct for endogeneity (2SPLS and 2SCML). This comparison is slightly complicated by the different normalization of the 2SPLS models relative to the probit and 2SCML models. The 2SCML models include the residual from the first stage regression as a variable in the second stage probit estimation. This variable is ε_1, while the coefficient

[6] The three normal variates used on the right-hand side of this model were drawn from a distribution with a zero mean and a unit standard deviation. All of the Monte Carlo simulations reported in this paper were performed using STATA 5.0.

on this variable is an estimate of λ. Thus, the probit model in the second stage of 2SPLS is normalized with respect to a different error term than the probit or 2SCML models, which include $\lambda(\varepsilon_1)$ in the error term. To make the results comparable, we use the following normalization for the 2SPLS estimates:

$$\hat{\omega} = \left(1 + (\hat{\gamma} + \hat{\lambda})^2 s_v^2\right)^{\frac{1}{2}} \tag{9}$$

where $\hat{\gamma}$ is the estimate of the endogenous parameter from the 2SCML estimates, λ is the error covariance parameter, $\hat{v}_i$ is the error from the reduced form regression, and $s_v^2 = (1/n)\sum_{i=1}^{n} \hat{v}_i^2$. The value of $\hat{\omega}$ is used to weight each estimated parameter. Thus, the values we compare the 2SPLS estimates to are $2/\hat{\omega}$ and $1.5/\hat{\omega}$.[7]

In Table 1 we present results from the Monte Carlo simulations foreach estimator on Eq. (6) using both 700 and 10,000 observation samples. The table is organized with the

Table 1 Bias of estimators[a]

	Probit		*2SPLS*		*2SCML*	
$r_{\epsilon 1,\epsilon 2}$	γ	β	γ	β	γ	β
$N = 700$						
−0.95	−1.48	−0.98	0.03	0.02	0.04	0.03
−0.70	−0.62	−0.37	0.04	0.02	0.06	0.04
−0.60	−0.42	−0.24	0.03	0.02	0.06	0.04
−0.45	−0.22	−0.11	0.03	0.02	0.06	0.04
−0.20	−0.04	−0.001	0.03	0.02	0.06	0.05
−0.10	0.03	0.03	0.03	0.02	0.06	0.05
0	0.05	0.04	0.02	0.02	0.06	0.04
0.10	0.06	0.03	0.02	0.01	0.06	0.04
0.20	0.04	−0.01	0.02	0.01	0.06	0.04
0.45	−0.05	−0.10	0.02	0.01	0.07	0.05
0.60	−0.20	−0.24	0.02	0.01	0.07	0.05
0.70	−0.36	−0.37	0.02	0.01	0.07	0.05
0.95	−1.13	−0.98	0.01	0.01	0.07	0.06
$N = 10{,}000$						
−0.95	−1.49	−0.99	0.001	0.001	0.002	0.002
−0.70	−0.64	−0.39	0.01	0.004	0.01	0.01
−0.60	−0.45	−0.26	0.004	0.003	0.01	0.004
−0.45	−0.25	−0.13	0.004	0.003	0.01	0.005
−0.20	−0.09	−0.03	0.003	0.003	0.01	0.01
−0.10	−0.02	−0.002	0.002	0.002	0.01	0.004
0	0.005	0.004	0.002	0.002	0.01	0.004
0.10	0.01	−0.003	0.002	0.002	0.005	0.004
0.20	−0.01	−0.04	0.001	0.002	0.004	0.003
0.45	−0.10	−0.13	0.001	0.001	0.003	0.002
0.60	−0.24	−0.26	0.001	0.001	0.004	0.003
0.70	−0.40	−0.39	0.002	0.002	0.01	0.01
0.95	−1.14	−0.98	0.001	0.001	0.005	0.004

[a]Entries are estimates of bias in the two parameters of the simulation model. Each pair of columns represents bias estimates for one of the estimation approaches.

[7]The normalized true parameters are given in the Appendix of the expanded version of this paper, which is available at the Political Analysis website.

error correlation for the particular Monte Carlo simulation given in the left-hand column; the next three pairs of columns give the bias for both coefficients in the "naive" probit model (columns 2 and 3), 2SPLS (columns 4 and 5), and 2SCML (columns 6 and 7).[8] For each model, γ is the coefficient on the endogenous right-hand side variable and β is the coefficient on the exogenous right-hand side variable. Results for the smaller dataset ($N = 700$) are in the top half of the table, and results for the large dataset ($N = 10{,}000$) are in the bottom half of the table.

Examination of Table 1 reveals that the "naive" probit model, which ignores the presence of endogeneity, performs quite poorly in a sample of 700 observations. The estimate of γ exhibits a great deal of bias. Unfortunately, the bias is also present in the estimates of β as well. The bias is much greater for γ than for β, and the magnitude of the bias increases as the error correlation increases in absolute magnitude. The same result holds in the large sample of 10,000 observations. Bias in all of the coefficients persists when the two equation error terms are highly correlated, especially for γ. However, when the error correlations approach zero, note that the bias in these coefficients is extremely low; this is what we would expect to see, given that under the appropriate assumptions in a large sample (no endogeneity, in particular) the probit estimator is unbiased. In general, the results in Table 1 show that endogeneity in a simple discrete choice model can have very serious effects on model estimates.

The two-stage probit least squares model produces results which show much less bias than the probit estimates in Table 1. Even in the 700 observation simulations the reduction in bias is clearly seen by comparing the results of the two models; there is no question that even in the smaller sample used in this Monte Carlo analysis, the 2SPLS estimator produces coefficient estimates which are much superior to those of the "naive" probit model, no matter what the level of error correlation or sample size.

The two-stage conditional maximum-likelihood results show this model to be superior in terms of bias to the "naive" probit model but not to two-stage probit least squares. In both the 700 and the 10,000 observation simulations, note that the level of bias for 2SCML is slightly higher relative to that for 2SPLS and is roughly comparable to that seen in the "naive" probit results when there is low error correlation. Thus, in terms of coefficient bias, two-stage conditional maximum-likelihood outperforms the "naive" probit model when the level of error correlation is high but does not necessarily outperform two-stage probit least squares.

However, although 2SCML may not be superior to 2SPLS in terms of coefficient bias, it may still produce standard errors which are more accurate. Obviously we want to know which estimator has the more accurate standard errors, since biased standard errors can lead to incorrect inferences. In order to compare the relative accuracy of the standard errors produced by 2SPLS and 2SCML, we compute an "overconfidence" statistic as described by Beck and Katz (1995). We determine the relative bias of each of the estimated standard errors from our three models by comparing the root mean square average of the estimated standard errors to the corresponding standard deviation of the coefficient estimates. Thus, for each coefficient from each estimator for each simulation we calculate

$$\text{Overconfidence} = 100 \frac{\sqrt{\sum_{i=1}^{1000} (\theta_i - \bar{\theta})^2}}{\sqrt{\sum_{i=1}^{1000} (\text{SE}(\theta_i)^2)}} \tag{10}$$

[8] Bias is defined as $E(\hat{\theta} - \theta)$, which is the average difference between the estimated coefficient in a particular Monte Carlo simulation and the actual coefficient in the true model.

Table 2 Overconfidence (%) of estimators[a]

	Probit		*2SPLS*		*2SCML*	
$r_{\epsilon 1,\epsilon 2}$	γ	β	γ	β	γ	β
$N = 700$						
−0.95	96.89	100.96	104.66	123.11	104.49	123.59
−0.70	101.87	99.25	104.94	102.46	103.47	103.88
−0.60	102.93	100.76	103.41	99.00	102.07	99.86
−0.45	101.24	99.98	103.99	97.70	102.36	100.28
−0.20	101.18	100.18	103.42	94.17	100.50	99.79
−0.10	102.67	98.59	103.34	92.94	102.38	98.91
0	102.44	99.47	102.51	91.82	102.13	99.87
0.10	99.92	98.55	100.21	90.68	100.71	99.56
0.20	102.81	99.87	100.05	90.07	105.81	102.63
0.45	101.60	100.59	98.34	89.89	104.96	103.65
0.60	102.29	101.74	96.44	89.73	104.55	103.57
0.70	100.73	98.91	93.76	88.04	105.86	105.00
0.95	101.46	99.85	91.27	86.78	104.88	115.13
$N = 10{,}000$						
−0.95	98.93	100.08	102.80	119.72	105.83	122.11
−0.70	96.85	100.23	98.80	100.94	98.32	103.42
−0.60	99.97	100.03	98.86	99.55	100.50	103.77
−0.45	99.70	99.04	99.44	96.30	99.11	101.26
−0.20	100.34	98.78	98.87	92.32	99.82	99.79
−0.10	100.31	100.38	96.39	89.42	99.72	100.36
0	100.16	101.52	97.49	90.40	100.03	101.51
0.10	99.74	102.35	95.99	89.59	100.25	102.67
0.20	98.71	102.20	94.10	88.14	100.07	103.61
0.45	99.90	101.97	93.40	88.15	102.80	105.70
0.60	100.04	99.60	94.46	86.76	104.11	105.76
0.70	99.77	95.80	94.02	84.00	104.72	105.77
0.95	98.83	98.93	90.39	83.62	105.43	120.89

[a]Entries are estimates of overconfidence in the two parameters of the simulation model. Each pair of columns represents overconfidence estimates for one of the estimation approaches.

This statistic is a measure of the percentage by which the estimated standard errors of a particular estimator under- or overstate variability. An estimate of 100% indicates that the estimated standard error is exactly equal to the true sampling variability. Numbers higher than this indicate estimated standard errors that are too small (overconfident), and numbers lower than this indicate estimated standard errors that are too large (inefficient). The overconfidence results are presented in Table 2. This table is organized like Table 1, except that the overconfidence results take the place of the bias results.

Note that the overconfidence estimates for the probit model are always close to 100% in both the 700 and the 10,000 observation simulations. This indicates that regardless of the degree of endogeneity, the probit model is producing standard errors that closely approximate the true sampling variability. This is *not* the case for 2SPLS, since most of the overconfidence statistics are below 100%. The estimated standard errors of 2SPLS show an upward bias, indicating that while 2SPLS is unbiased, it does not accurately estimate the

true standard errors, since the estimated standard errors are larger than the true sampling variability. However, 2SCML produces standard errors with an accuracy comparable to those of the probit model for γ, and comparable for β except at extremely high levels of error correlation (-0.95 or 0.95). The relative accuracy of the standard errors estimated by 2SCML over 2SPLS is especially apparent in the lower section of Table 2, when we look at the simulations for the 10,000 observation sample size. There it is clear that in extremely large samples, 2SCML produces more accurate standard errors than 2SPLS. Thus, even though 2SPLS outperforms 2SCML in terms of coefficient bias, 2SCML produces more accurate standard errors.

So what conclusions can be drawn from these Monte Carlo simulations? First, ignoring endogeneity in most cases will lead to biased and high variance estimates. The only exception is when the equation error terms have little or no correlation, a situation which cannot be guaranteed and which will be difficult to diagnose in real data. This means that ignoring endogeneity in these models will lead to incorrect conclusions about political phenomena. Second, two-stage models which deal with endogeneity are much better alternatives, since they generally produce estimates which are closer to the true parameters which generated the data. However, the simulation results here showed that two-stage probit least squares slightly outperforms two-stage conditional maximum likelihood in both small and large samples in terms of bias, while 2SCML produces more accurate standard errors.

4 Application to Real Data: Modeling Votes and Uncertainty

Thus far, using simulated data we have shown both the properties of these two-stage estimators and how they perform relative to "naive" models. Simulations are ideal for making these points, since the "true" underlying model is known, and the simulation results can then be compared directly with the known model. However, simulations do not speak to how these two-stage models perform with real-world data, so in this section we apply each model to a nonrecursive model of presidential election voting.

This example concerns the relationship between voter choice in a presidential election and voter uncertainty about the positions of the candidates on various issues. First, as has been shown in the positive theory of voter decision making under uncertainty, given the assumption that voters are risk averse (implied by the assumption that voter utility functions for candidates are single-peaked and concave), uncertainty about the positions of candidates on policy issues should depress a voter's support for a candidate (Alvarez 1998; Bartels 1986; Enelow and Hinich 1984; Franklin 1991).

The second component of the model focuses on the determinants of a voter's uncertainty about a candidate's policy positions. Three factors account for voter uncertainty of the candidate's policy positions: their personal information costs, their exposure to information, and the flow of information during the campaign. The more costly it is for a voter to obtain, process, and store information, the more uncertain he/she should be about the candidate's position; the less exposed to information and the less attentive and interested the voter is, the greater his/her uncertainty about the position of the candidate; and the greater the amount of information available about the candidates, the less the uncertainty a voter will have regarding the positions of the candidates (Alvarez 1998). With these variables a model of uncertainty can be constructed under certain assumptions about the relationship between these independent effects and voter uncertainty.

This uncertainty should directly influence the voter's evaluation of the candidate, controlling for other policy and nonpolicy factors relevant to the voter's calculus. The uncertainty measure is thus an important explanatory variable in the determinants of candidate

evaluation and choice, as well as an important endogenous variable. This causal process relating uncertainty and candidate evaluation and choice is usually depicted in the literature as a hierarchical model (Bartels 1986; Franklin 1991). This hierarchical model can be shown as two equations:

$$v_{iJ} = \beta_1 + \beta_{11}X_{1i} + \beta_{12}X_{2i} + \tau_{11}u_{iJ} + \xi_{1i} \tag{11}$$

$$u_{iJ} = \beta_2 + \beta_{21}X_{1i} + \beta_{23}X_{3iJ} + \tau_{21}v_{iJ} + \xi_{2i} \tag{12}$$

where v_{iJ} is voter i's uncertainty about candidate J, u_{iJ} is the utility or evaluation of the voter for candidate J, X_{1i} are demographic variables, X_{2i} are variables measuring voter i's information costs and exposure to political information, X_{3iJ} are variables relating to i's evaluation of candidate J's policy and nonpolicy attributes, the β's and τ's are parameters to be estimated, and the ξ's are error terms in each model.[9] Past research regarding voter uncertainty of candidate policy positions has assumed that τ_{11} is zero, implying that a voter's evaluation of the candidate does not influence their uncertainty of the candidate (Bartels 1986). Under this assumption this model is hierarchical, implying that there is no reason to suspect correlation between ξ_{1i} and ξ_{2i}. But if τ_{11} is *not zero*, then the two error terms are likely to be correlated, and the error term in each of the equations is likely to be correlated with the right-hand side variables in each equation. As a consequence of this endogeneity, the estimates of the parameters in this model are likely to be biased.

Since we have reason to suspect that τ_{11} might be nonzero, and that a simultaneous relationship exists between candidate evaluations and voter uncertainty, independent estimation of these uncertainty equations is inappropriate. Rather, the endogeneity between candidate evaluations and voter uncertainty must be appropriately modeled so that consistent empirical results can be obtained.

The terms in the model can be expanded to give a general statistical model of uncertainty and evaluations in a two-candidate election:

$$v_{iJ} = \beta_1 + \beta_{11}X_{1i} + \beta_{12}X_{2i} + \tau_{11}u_i + \xi_{1i} \tag{13}$$

$$v_{iG} = \beta_2 + \beta_{21}X_{1i} + \beta_{22}X_{2i} + \tau_{21}u_i + \xi_{2i} \tag{14}$$

$$u_i^* = \beta_3 + \beta_{31}X_{1i} + \beta_{33}X_{3iJ} + \beta_{34}X_{4iJ} + \tau_{31}v_{iJ} + \beta_{35}X_{3iG} + \beta_{36}X_{4iG} + \tau_{32}v_{iG} + \xi_{3i} \tag{15}$$

where v_{iJ} is the voter's uncertainty for candidate J, v_{iG} is his/her uncertainty for candidate G, u_i^* denotes a binary preference variable expressing which candidate i prefers (J or G), X_{1i} are demographic variables, X_{2i} are variables measuring voter i's information costs and exposure to political information, X_{3iJ} and X_{3iG} are vectors of variables for policy-specific information about each candidate, and X_{4iJ} and X_{4iG} are variables for nonpolicy information about the candidates. The β's and τ's are parameters to be estimated, and the ξ's are error terms.

This nonrecursive model provides an excellent case for the comparison of the two-stage estimators discussed above, and how they perform relative to "naive" models which do

[9] The demographic variables (X_{1i}) are in the first equation since we expect that two demographic groups, minorities and females, might be more uncertain than others. It is possible that other demographic groupings might be useful, such as income and socioeconomic status, but these are concepts which surveys are not well suited to measure. These same variables are in the second equation to control for nonpolicy and candidate variations across individuals in their candidate preference.

not address endogeneity. Our approach is to estimate the parameters of the model in two ways—first with "naive" OLS or probit and then with both two-stage models. Comparison between all of the models will show the performance of these estimators in real-world data and will shed light on a research problem of substantial interest.

Here we use data from the 1976 presidential election. This is a particularly interesting election to examine, for methodological and substantive reasons. Methodologically, this allows the use of the 1976 panel study of the presidential election campaign conducted by Thomas E. Patterson, which provides an excellent vehicle for studying uncertainty and voter decision making. Substantively, recall that Carter had been a virtually invisible governor before the spring months of 1976; and before being appointed to serve the remainder of Nixon's term, Ford was a low-key House Minority Leader. Neither candidate was a long-established national figure, and neither was very well-known at the beginning of the campaign. But while the statements of the candidates were covered widely in the press (Patterson 1980; Witcover 1977), Ford and Carter were extremely moderate in most of their positions, and there were few issues of public policy on which Ford and Carter took distinct stands. Therefore, this could be an election in which voter uncertainty over the candidates' policy positions was extreme, given two moderate and relatively poorly known candidates. Thus it is reasonable to expect that this uncertainty may have strongly influenced voter evaluations of the candidates.

Below we present results for the uncertainty equations (Table 3) and the vote choice equation (Table 4). Details about the operationalization of the variables in the model, descriptive statistics for the independent variables, and the reduced form estimates are available in the expanded version of this paper, which is available at the Political Analysis website. Note that candidate preference is coded 1 for Carter and 0 for Ford and that the voter's uncertainty is coded so that high values show a high uncertainty about the candidate's issue positions.

The uncertainty models in Table 3 give estimates for each candidate's uncertainty from OLS (second and fourth columns) and by 2SPLS (third and fifth columns).[10] The first thing to note in Table 3 is that the models fit the data reasonably well and that most variables have the theoretically expected signs and are statistically significant.[11]

[10] No corrections to the estimated coefficient variances are used here. The correction suggested by Achen (1986, p. 43) produced weights of 1.00 for both uncertainty equations, which leaves the results in Table 3 unchanged.

[11] The indicators of voter information costs in these models, the first four variables in the table (education, political information, gender, and race) all influence uncertainty in the expected direction. That is, better educated and informed voters were less uncertain of the policy positions of Carter and Ford in 1976, while both women and racial minorities were more uncertain of the positions of these two candidates. And only the estimate for racial minorities fails to reach statistical significance in these models. The rest of the variables in these two models, besides the relative candidate preference indicator, measure various dimensions of voter attachment to the political world and their exposure to political information. Unlike the National Election Studies, however, the 1976 Patterson data contained a useful set of questions which allowed us to incorporate three additional variables into these two models, which indicated whether the respondent watched either or both of the televised debates or recalled seeing some advertisement from the particular candidate's paid media campaign. So the estimates for watching the debates or campaign advertisements allows examination of two specific types of exposure to campaign information. Of all of these variables, only the estimates for partisan strength and the first debate indicators are not signed in accordance with our theory; however, these variables are not statistically significant. All of the rest have the expected sign, and most do reach reasonable levels of statistical significance. The first two indicators—media exposure and political efficacy—both show that the more exposed and efficacious voters are statistically more certain of the policy positions of the two candidates in this election. Also, voters who watched the first debate were more certain of Ford's policy positions, while those who watched the second debate were more certain of Carter's positions. What is fascinating about these results is that they comport with both the information made available by these two debates in 1976 and popular perceptions concerning which of the two candidates had more effectively presented himself and his campaign positions in each debate. Most observers concluded that in the first debate, Ford had articulated his positions on unemployment and the economy

Table 3 Two-stage uncertainty results, 1976 election: Carter and Ford uncertainty[a]

Independent variable	*Carter uncertainty*		*Ford uncertainty*	
	OLS	*IV*	*OLS*	*IV*
Constant	6.0**	5.5**	5.28**	5.1**
	0.89	0.89	0.88	0.87
Education	−0.28**	−0.28**	−0.32**	−0.29**
	0.14	0.14	0.14	0.14
Political information	−0.28**	−0.28**	−0.24**	−0.25**
	0.04	0.04	0.04	0.04
Gender	0.60**	0.65**	0.73**	0.78**
	0.24	0.24	0.24	0.24
Race	0.05	0.24	0.68*	0.46
	0.46	0.49	0.46	0.49
Partisan strength	0.12	0.12	0.11	0.06
	0.15	0.15	0.15	0.15
Media exposure	−0.14**	−0.14**	−0.21**	−0.22**
	0.06	0.06	0.06	0.06
Political efficacy	−0.07**	−0.06**	−0.05**	−0.03*
	0.02	0.03	0.02	0.03
First debate	0.01	0.004	−0.26*	−0.25*
	0.17	0.17	0.17	0.17
Second debate	−0.21*	−0.22*	−0.11	−0.13
	0.16	0.16	0.16	0.16
Candidate advertising	−0.16	−0.17	−0.50*	−0.49*
	0.37	0.37	0.34	0.35
Candidate preference	−1.04**	−0.08**	−0.34*	0.02
	0.25	0.03	0.25	0.03
Adjusted R^2	0.25	0.23	0.22	0.22
Model SE	2.5	2.5	2.5	2.5
Uncertainty mean	3.9	3.9	3.8	3.8
Number of cases	464	464	464	464

[a]Entries are two-stage least-squares estimates and their associated adjusted standard errors. *A $p = 0.10$ level of statistical significance; **a $p = 0.05$ level (both one-tailed tests).

The estimated effects of candidate preference on candidate uncertainty deserve close scrutiny. Had we proceeded without recognition of the possibility of endogeneity in this model, we would have estimated strong and inconsistently signed effects of candidate preferences on candidate policy uncertainty. The coefficient of −1.04 in the Carter OLS

quite forcefully, and had put Carter on the defensive. Then, while Carter began to respond, the debates were interrupted by technical difficulties which most believed damaged Carter's ability to get his arguments across (Witcover 1977). During the second debate, both candidates attacked their opponent's foreign policy positions, which might account for the negative effect watching this debate appears to have had on the policy uncertainty for both candidates in the models. But the second debate was marred by Ford's "no Soviet domination of Eastern Europe" comment, retracted within the next 5 days—which might account for the only marginal reduction in Ford uncertainty for voters who watched the second debate.

Table 4 Two-stage voting models, 1976 election: Probability of Carter support[a]

Independent variable	*Probit estimate*	*2SCML estimate*	*2SPLS estimate*
Constant	1.58**	1.58**	1.71**
	0.75	0.91	0.91
Ford issue	0.16**	0.18**	0.18**
distance	0.07	0.08	0.08
Carter issue	−0.10*	−0.10*	−0.11*
distance	0.07	0.07	0.07
Ford	0.06	0.66*	0.68*
uncertainty	0.08	0.51	0.51
Carter	−0.11*	−0.75*	−0.77*
uncertainty	0.08	0.55	0.55
Ford	−0.41**	−0.40**	−0.39**
traits	0.05	0.05	0.05
Carter	0.37**	0.33**	0.33**
traits	0.05	0.06	0.06
Party	−0.33**	−0.29**	−0.29**
identification	0.07	0.07	0.07
Education	−0.33**	−0.40**	−0.39**
	0.16	0.18	0.18
Gender	−0.16	−0.35	−0.34
	0.26	0.31	0.30
Ford error		−0.61	
		0.53	
Carter error		0.64	
		0.56	
% correct	96.0	95.3	95.7
Number of cases	464	464	464

[a]Entries are maximum-likelihood estimates and their associated asymptotic standard errors. *A $p = 0.10$ level of statistical significance; **a $p = 0.05$ level (both one-tailed tests). 2SCML is the River–Vuong conditional maximum-likelihood model; 2SPLS is the limited-information probit and least-squares model. Candidate uncertainties in the 2SPLS model are instruments from a reduced-form regression; the Ford and Carter error indicators are the error terms from the same regression.

uncertainty regression implies that the *greater* the probability of Carter support, the *lower* their uncertainty about Carter's positions—a result in accord with the expectations discussed earlier. But the negative and statistically significant estimate of candidate preference in the Ford OLS regression indicates that the *greater* the probability of Carter support, the *lower* the uncertainty of Ford's positions as well; certainly this is a counterintuitive result.

But, when these models are estimated with 2SPLS, note the dramatic effects on the results. Here, the candidate preference indicator has the expected sign for both candidates but is statistically significant now only in the Carter equation. These results are interesting substantively as well as methodologically, since the candidate preference indicator is significant only for the challenging candidate. This suggests that while voters do engage in selective information processing about presidential candidates, such strategies may not be

necessary for incumbent candidates: voters may have already obtained enough information about incumbents to make selective processing unnecessary.

Next, Table 4 gives the results from the voter choice equations. Recall that the dichotomous dependent variable here is coded so that support for Carter is the high category, and Ford is the low category. Thus the parameter estimates express the relative effect of the particular variable on the probability of Carter support. Note that both models fit the data very well; each correctly predicts slightly over 95% of the cases in the sample. Furthermore, the variables of interest have the theoretically expected signs and are statistically significant at reasonable levels.[12]

Beginning with the "naive" probit estimates of the uncertainty coefficients, note that they have the theoretically expected signs, and that the estimated effect of Carter uncertainty is statistically significant. But, compared with the estimated effects of the uncertainty terms in the other models, the estimates in the second column ("naive" probit) seem to be dramatically underestimated. For, in both the 2SPLS and the 2SCML models the estimated effects of both candidates' uncertainty terms have the expected signs, are statistically significant, and have a much larger impact on candidate preference than was estimated with "naive" probit.

But is there evidence of endogeneity in this model? Recall from above that Rivers and Vuong (1988) demonstrate that the two estimated parameters for the reduced form errors in 2SCML give a test of exogeneity. The likelihood-ratio test for the 2SCML model versus a similar model without these two parameters yields a χ^2 of 10.88, which is larger than the critical value of 5.99 at 2 degrees of freedom, showing the joint significance of these parameters. Therefore, endogeneity between candidate evaluation and voter policy uncertainty is evident, and needs to be accounted for in these models.

Of course, as is usually the case in nonlinear probit models, the parameters in the evaluation models cannot be interpreted directly, since the models are nonlinear and the effect of any particular variable on the probability of supporting one candidate is dependent on the values of the other variables and parameters in the model. To give a more intuitive feel for the magnitude of the effect of candidate uncertainty upon candidate support, we utilize graphical methods (King 1989; McCullagh and Nelder 1991).[13] The results for

[12] Both squared issue distance terms have the expected sign, yet the estimate for Carter issue distance is statistically significant at only the $p = 0.10$ level, while the similar estimate for Ford issue distance is estimated more precisely ($p = 0.05$). But the effect of issues is greater for Ford than for Carter, as witnessed by the relative sizes of the coefficients for Ford issue distance. With more uncertainty about Carter's positions, and with Carter uncertainty having more of an effect on candidate support, it is not surprising that Carter's positions on the issues had less of an influence on voter evaluations of the candidates in 1976. The other two sets of parameters of interest—the nonpolicy dimensions of candidate evaluations—all have the expected sign and are precisely estimated. That is, the partisan affiliation of the voter clearly influenced their evaluations of Ford and Carter. Also, their assessments of the characters of the candidates influenced their evaluations, with higher evaluations of either candidate's personal and professional characteristics leading to greater support for the candidate. But it is interesting to note that, here again, the effect of candidate characteristics on candidate support is greater for Ford than for Carter. Perhaps the decision by Ford and his advisors to focus on the "character issue" and to employ the "Rose Garden strategy" had some effect on the electorate, leading to more positive assessments of his character than for Carter.

[13] Graphical methods involve simple simulations using the parameters of the model (given in Table 4) and some combination of values of the independent variables. Here we set all but one of the independent variables to their mean values in the sample of voters used to estimate the model. Then, we change the one variable of interest (here, either Carter or Ford uncertainty) across the range of values the variable takes in the actual data. This produces an estimate of the linear predictor for each value of the variable of interest (which is either Carter or Ford uncertainty), which is then transformed into an estimate of the hypothetical individual's probability of voting for one candidate (here Carter). In the "naive" probit and 2SCML simulations, we use the mean values of the uncertainty variables (and in the latter we use the mean values of the reduced-form errors). In the 2SPLS simulations, we use mean values of the uncertainty instruments.

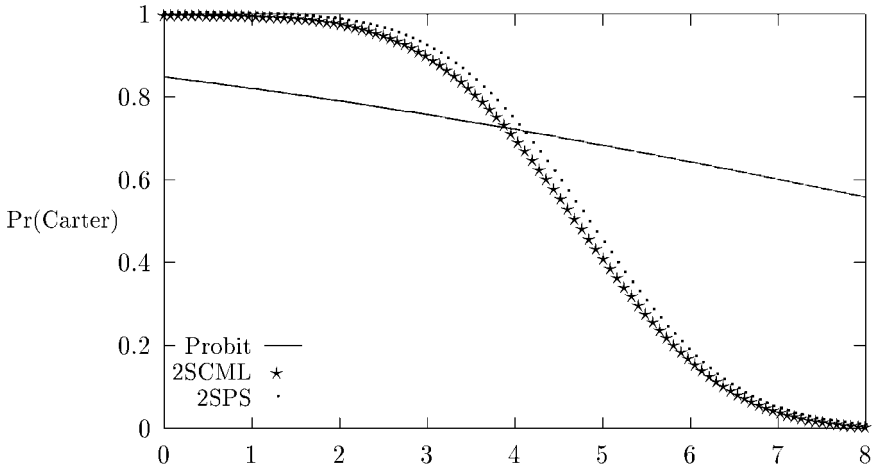

Fig. 1 Comparison of estimators. This gives the simulated probabilities of Carter support across the ranges of uncertainty about Carter in the 1976 election. The coefficients these calculations are based upon are listed in Table 4.

the candidate policy uncertainty parameters from the three models are graphed in Figs. 1 (Carter uncertainty) and 2 (Ford uncertainty).

In both figures the voter's uncertainty about the candidate's policy positions is graphed along the x-axis, and the probability that the voter would support Carter on the y-axis. The dark line gives the effect of candidate policy uncertainty on probability of Carter support as estimated by the "naive" probit model, the stars give the effect estimated by 2SCML, and the dots as estimated by 2SPLS (holding the other variables constant at their sample mean values). The strong effect of uncertainty on candidate evaluation as estimated by the two-stage models is clear in these graphs. Take two identical voters, with "mean" values

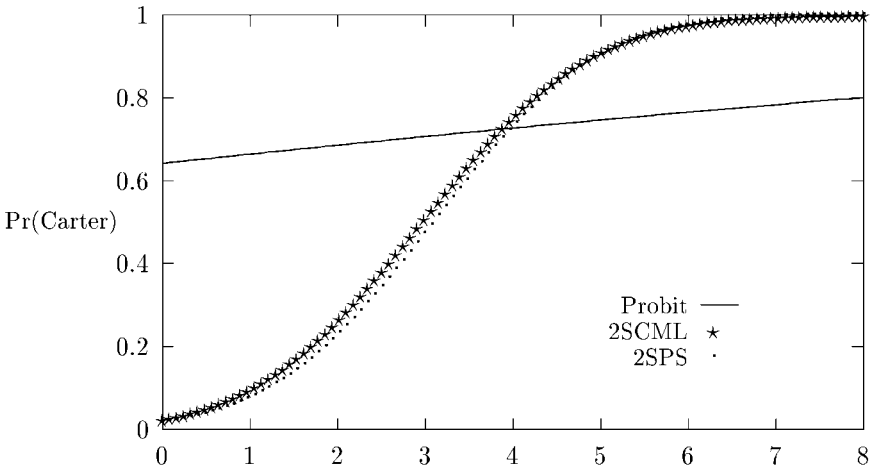

Fig. 2 Comparison of estimators. This gives the simulated probabilities of Carter support across the ranges of uncertainty about Ford in the 1976 election. The coefficients these calculations are based upon are listed in Table 4.

on all the variables, but one who is very certain of Carter's position on the issues (1) and the other who is relatively uncertain of Carter's positions (5). The graph indicates that the former voter—the relatively certain voter—has a very high probability of supporting Carter, while the uncertain voter is slightly less than 50% likely to support Carter. Thus, by changing the relative uncertainty of the voter from very certain to relatively uncertain, the probability of supporting Carter in 1976 changes by over 50%. But the effects of uncertainty as estimated by "naive" probit are quite slight. To take the same hypothetical voters, the difference in probability of Carter support between a voter certain of Carter's positions (1) and a voter uncertain of Carter's positions (5) is not great, just over 10 points. This shows the substantial importance of accounting for endogeneity in this particular model of voter choice—had the endogeneity not been modeled, the effects of policy uncertainty would have been underestimated by a large margin.[14]

5 Discussion

This substantive example should make clear the importance of nonrecursive choice models in political science. The theoretical model connecting voter preferences for candidates with their uncertainty about the candidates shows the endogeneity of these two factors, a point which has not been taken into consideration in past research on either topic. The empirical estimation of this nonrecursive model demonstrates the endogeneity of preferences and uncertainty and, in so doing, produces interesting insights into the politics of presidential campaigns (Alvarez 1998). Also, this example demonstrates the pitfalls associated with not accounting for endogeneity in a discrete choice model. Recall that the effects of candidate preferences on candidate policy uncertainty would have been vastly overestimated (and in the wrong direction in the case of Ford), and the reciprocal effects of uncertainty on preferences dramatically underestimated.

As a research methodology, then, these two-stage models of binary and continuous variables have great promise. Both two-stage models discussed here are easy to estimate, and one (2SCML) produces a simple test for exogeneity. Not only did the substantive example point to their superior performance, but also the comparisons made between the performance of both two-stage models relative to "naive" models in simulated data showed the pitfalls of assuming exogeneity. The Monte Carlo evidence presented earlier in this paper showed that 2SPLS slightly outperforms 2SCML in terms of bias but that the 2SCML model outperforms 2SPLS in terms of accurate estimation of the standard errors.

Given these results, what practical advice can be given to researchers faced with substantive problems involving a nonrecursive set of binary and continuous variable equations? First, if there are theoretical reasons to suspect endogeneity in the model, it must be modeled. Ignoring endogeneity will lead to biased estimates, and the severity of this bias was shown in both the simulated and the real data earlier in this paper.

Second, researchers have two methods available with which to estimate models involving both nonrecursive continuous and binary variables: 2SPLS and 2SCML. Which is the correct methodology to apply? Our Monte Carlo simulations showed that 2SPLS is slightly superior in terms of bias but that 2SCML is superior in terms of accurately estimating

[14] Worthy of additional note, though, is the observation that the effects of policy uncertainty were greater for Carter uncertainty than for Ford uncertainty. In other words, the voter's uncertainty for Carter's positions had a greater effect on which candidate the voter supported than the voter's uncertainty of Ford's positions. These differential effects are probably the result of greater uncertainty about the positions of Carter, who was somewhat less visible before the general election began and who did not have the tools of an incumbent president to make his policy positions known to the electorate.

the standard errors. Thus, in our opinion, the balance of evidence from our simulations indicates to us a preference for the use of the 2SCML model for this class of causal models—nonrecursive models with continuous and binary dependent variables. This is especially true given the fact that the 2SCML model provides a simple and convenient statistical test for exogeneity.

However, more work on estimation techniques for other types of nonrecursive models, involving different types of discrete dependent variables is needed. Unfortunately, little is known about the finite sample properties of nonrecursive choice models with discrete dependent variables. Examples of empirical applications of such choice models are appearing in the literature: Alvarez and Butterfield (1999a, b) have used two-stage models to estimate choice models involving sets of binary choices, Alvarez (1998) has used a two-stage approach to estimate models with both continuous and unordered choice dependent variables variables, and Schotz and Lubell (1998) have used a two-stage approach where they have ordinal and continuous dependent variables. Given that discrete choice models are appearing at a high frequency in the political science literature, understanding the properties of different nonrecursive discrete choice models in finite samples will become increasingly important in coming years.

References

Achen, C. H. 1986. *The Statistical Analysis of Quasi-Experiments*. Berkeley: University of California Press.

Aldrich, J. H., and F. D. Nelson. 1984. *Linear Probability, Logit, and Probit Models*. Beverly Hills, CA: Sage.

Aldrich, J. H., R. G. Niemi, G. Rabinowitz, and D. W. Rohde. 1982. "The Measurement of Public Opinion About Public Policy: A Report on Some New Issue Question Formats." *American Journal of Political Science* 26:391–414.

Alvarez, R. M. 1998. *Information and Elections, Revised to Include the 1998 Presidential Election*. Ann Arbor: University of Michigan Press.

Alvarez, R. M., and T. L. Butterfield. 1999a. "The Resurgence of Nativism in California? The Case of Proposition 187 and Illegal Immigration," manuscript. Pasadena: California Institute of Technology.

Alvarez, R. M., and T. L. Butterfield. 1999b. "The Revolution Against Affirmative Action in California: Politics, Economics and Proposition 209," manuscript. Pasadena: California Institute of Technology.

Alvarez, R. M., and C. H. Franklin. 1994. "Uncertainty and Political Perceptions." *Journal of Politics* 56:671–688.

Amemiya, T. 1978. "The Estimation of a Simultaneous Equation Generalized Probit Model." *Econometrica* 46:1193–1205.

Bartels, L. M. 1986. "Issue Voting Under Uncertainty: An Empirical Test." *American Journal of Political Science* 30:709–728.

Beck, N., and J. N. Katz. 1995. "What to Do (and Not to Do) With Time-Series Cross-Section Data." *American Political Science Review* 89:634–647.

Brehm, J. 1993. *The Phantom Respondents*. Ann Arbor: University of Michigan Press.

Caldeira, G. A., and J. R. Wright. 1998. "Lobbying for Justice: Organized Interests, Supreme Court Nominations, and the United States Senate." *American Journal of Political Science* 42:499–523.

Campbell, J. E. 1983. "Ambiguity in the Issue Positions of Presidential Candidates: A Causal Analysis." *American Journal of Political Science* 27:284–293.

Downs, A. 1957. *An Economic Theory of Democracy*. New York: Harper and Row.

Enelow, J. M., and M. J. Hinich. 1984. *The Spatial Theory of Voting*. New York: Cambridge University Press.

Fiorina, M. P. 1981. *Retrospective Voting in American National Elections*. New Haven, CT: Yale University Press.

Fiske, S. T., and M. A. Pavelchak. 1986. "Category-Based versus Piecemeal-Based Affective Responses: Developments in Schema-Triggered Affect." In *The Handbook of Motivation and Cognition*, R. M. Sorrentino and E. T. Higgins, eds. New York: Guilford Press.

Franklin, C. H. 1991. "Eschewing Obfuscation? Campaigns and the Perceptions of U.S. Senate Incumbents." *American Political Science Review* 85:1193–1214.

Franklin, C. H., and J. E. Jackson. 1983. "The Dynamics of Party Identification." *American Political Science Review* 77:957–973.

Graber, D. A. 1988. *Processing the News*. White Plains, NY: Longman.

Hanushek, E. A., and J. E. Jackson. 1977. *Statistical Methods for Social Scientists*. New York: Academic Press.

Heckman, J. J. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models." *Annals of Economic and Social Measurement* 5:475–492.

Heckman, J. J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47:153–161.

Jacobson, G. C., and S. Kernell. 1981. *Strategy and Choice in Congressional Elections*. New Haven, CT: Yale University Press.

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

Lazarsfeld, P., B. Berelson, and H. Gaudet. 1944. *The People's Choice*. New York: Columbia University Press.

Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Markus, G. B., and P. E. Converse. 1979. "A Dynamic Simultaneous Equation Model of Electoral Choice." *American Political Science Review* 73:1055–1070.

McCullagh, P., and J. A. Nelder. 1983. *Generalized Linear Models*, 2nd ed. London: Chapman and Hall.

Page, B. I., and C. Jones. 1979. "Reciprocal Effects of Policy Preferences, Party Loyalties and the Vote." *American Political Science Review* 73:1071–1089.

Patterson, T. C. 1980. *The Mass Media Election*. New York: Praeger.

Rahn, W. M. 1990. *Perception and Evaluation of Political Candidates: A Social-Cognitive Perspective*, Ph.D. dissertation. Minneapolis: University of Minnesota.

Rivers, D., and Q. H. Vuong. 1988. "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models." *Journal of Econometrics* 39:347–366.

Schotz, J. T., and M. Lubell. 1998. "Trust and Taxpaying: Testing the Heuristic Approach to Collective Action." *American Journal of Political Science* 42:398–417.

Verba, S., and N. H. Nie. 1972. *Participation in America*. New York: Harper and Row.

Witcover, J. 1977. *Marathon*. New York: Viking Press.

Wolfinger, R. E., and S. J. Rosenstone. 1980. *Who Votes?* New Haven, CT: Yale University Press.