Statistical Backwards Induction:
A Simple Method for Estimating Strategic Models

Muhammet Ali Bas[*]
Department of Political Science
University of Rochester

Curtis S. Signorino
Department of Political Science
University of Rochester

Robert W. Walker
Department of Political Science
Washington University in Saint Louis

September 13, 2006

[*]Our thanks to Kevin Clarke, John Londregan, Jeff Ritter, Ahmer Tarar and Kuzey Yilmaz for helpful discussions concerning this paper. Support from the National Science Foundation (Grants SES-9817947, SES-0213771, and SES-0413381) and from the Peter D. Watson Center for Conflict and Cooperation is gratefully acknowledged. Paper previously presented at the 2002 Political Methodology Summer Meeting. Updates of this paper may be found at www.rochester.edu/College/PSC/signorino. Email: curt.signorino@rochester.edu, rww@artsci.wustl.edu, mbas@mail.rochester.edu.

Abstract

We present a simple method for estimating regressions based on extensive-form games. Our procedure, which can be implemented in most standard statistical packages, involves sequentially estimating standard logits (or probits) in a manner analogous to backwards induction. We demonstrate that the technique produces consistent parameter estimates and show how to calculate consistent standard errors. To illustrate the method, we replicate Leblang's (2003) study of speculative attacks by financial markets and government responses to these attacks.

1 INTRODUCTION

Strategic interaction is a fundamental consideration in the study of political choice. The choices of members of Congress to seek reelection or retire is almost certainly related to the decisions of potential challengers (Carson 2003). Executives must decide whether or not to veto legislation in light of the potential for a legislative override (Carson and Marshall 2004). In international politics, Signorino and Tarar (2006) examine the strategic sources of extended deterrence and Leblang (2003) shows that the strategic interaction between governments and markets is a critical element in understanding currency crises. All of the aforementioned theoretical insights require empirical tests that roughly conform to the important structural elements of the decision making environment.

Despite the increased use of choice (or game) theoretic explanations in modern political science research, there is often a subtle disconnect between theories and the empirical tests of these theories. "Indirect" statistical tests of formal models generally fail to properly characterize the hypothesized relationships in statistical testing.[1] Techniques for ameliorating this problem have been previously presented (Signorino 1999, Signorino 2003, Signorino 2002, Lewis and Schultz 2003, Signorino and Tarar 2006). However, these methods generally require researchers to program and then optimize a frequently complex, problem-specific model.

In this paper, we simplify the estimation of statistical strategic models in an effort to bridge the disconnect between theory and empirical analysis. Our motivation is primarily practical with respect to most contemporary research in political science. Signorino (1999, 2002, 2003) provides a general framework for estimating statistical strategic models, along with different ways of conceptualizing the uncertainty that can make strategic

[1] Signorino and Yilmaz (2003) focus on this problem.

models "statistical." Based on this, the program STRAT was developed, allowing scholars to estimate strategic models relevant to their substantive research.[2] The types of games that can be estimated in STRAT, however, are limited to six or seven relatively simple — though quite common — decision structures. Scholars desiring to estimate parameters in a strategic model that is not included in STRAT must therefore construct and estimate their own maximum likelihood models.

In the following sections, we first demonstrate a useful simplification that allows recursive statistical strategic models to be estimated with simple variants of logit and probit — and in a way that is consistent with the underlying theory. The method is very similar to the game-theoretic notion of backwards induction. We also show how to correctly estimate the standard errors using this simplified technique. Finally, using the proposed method, we replicate Leblang's (2003) study of currency crises.

2 REFERENT EXAMPLE

It will be helpful to employ an example throughout our analysis. Figure 1 displays the structure of our referent model. In this game, player A must choose between playing Right (R) or Left (L). If A chooses Right, player B must then choose whether to play right (r) or left (ℓ). Playing left results in Outcome $R\ell$. If B chooses right, (Rr) is the outcome. The players utilities are shown at the terminal nodes — e.g., $U_A(L)$ is state A's utility for the L outcome, and $U_B(L\ell)$ is state B's utility for the $L\ell$ outcome.

Structurally, the model is quite simple. The fact that players do not make their decisions simultaneously (or without knowledge of other players' moves) makes this a fully *recursive*

[2] "STRAT: A Program for Estimating Statistical Strategic Models" is available at www.rochester.edu/College/PSC/signorino.

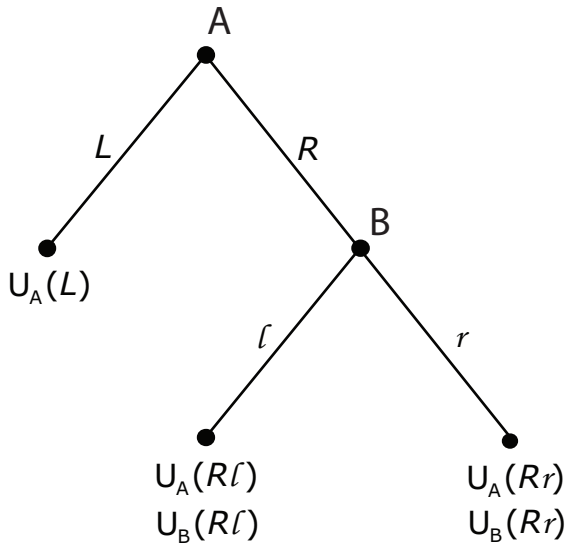


Figure 1: Two-Player Game. *Player A must choose whether to play Left (L) or Right (R). If Player A plays R, Player B must then choose whether to play Left (ℓ) or Right (r).*

model. Assuming the players and we as analysts have complete information concerning the utilities, then for any specification of the utilities, the model can be solved via backwards induction for the subgame perfect equilibrium.

The logic of backwards induction is relatively straightforward. We start at the end nodes of the game and then ask what choice the player will make at the preceding decision node. With perfect and complete information we can determine which option will be selected. Given that knowledge, we can step back to the next preceding decision node and determine what that player will do. For example, consider Figure 1. Let us first suppose that A's preferences are $Rr \succ L \succ R\ell$ and B's are $Rr \succ R\ell$.[3] If we start with B's decision, we can see that, if A chooses R, B will choose r, leading to Rr. Player A knows this, so her decision between L and R is really a decision between outcomes L and Rr. Given A's preferences above, A will choose R, leading to the subgame perfect equilibrium outcome

[3] $x \succ y \succ z$ means that option x is preferred to y, that y is preferred to z, and that (via transitivity) x is preferred to z. We omit L in B's preferences because it is not relevant for this game.

Rr. Now suppose A's preferences are as stated above, but B's are exactly the opposite: $R\ell \succ Rr$. In this case, if A chooses R, then B will choose ℓ. Because A prefers L to $R\ell$, the equilibrium outcome will be L. It is important to note that it is this ability to work from the end of the game back towards its beginning that allows us to employ the statistical technique we will later present.

In its present form, the model is not statistical. To construct a statistical strategic model, we need to assume some form of uncertainty on the part of the players and/or analyst. There are a number of ways we might do so: agent error, private information by the players, or complete information by the players but imperfect measurement (or specification of regressors) by the analyst (Signorino 2003). Of the three, the one that is consistent with our simplified estimation technique is the agent error specification.[4] We now give Figure 1 an agent error representation.

2.1 The Statistical Strategic Model

To fully specify the statistical model, we assume there is a shock to the players' expected utilities for their actions. Consider B's decision. B's utilities will be specified as

$$U_B^*(\ell) = U_B(\ell) + \alpha_\ell = U_B(R\ell) + \alpha_\ell \quad (1)$$

$$U_B^*(r) = U_B(r) + \alpha_r = U_B(Rr) + \alpha_r \quad (2)$$

where $U_B^*(\cdot)$ is considered the "true" utility, $U_B(\cdot)$ is the component of the utility that is observable to the other players and to the analyst, and α is a random private component,

[4]Fortunately, for those who do not like the Non-Nash interpretation of the agent error specification, Signorino (2003) shows that, for simple models such as that in Figure 1, it will yield similar results as the (strictly Nash) private information version.

observable only to player B. We assume B maximizes her (true) utility.

Because we as the analysts do not observe the α terms, we can only make probabilistic statements about whether B is likely to play right or left. Once a density is assumed for α, derivation of the choice probabilities is straightforward, following traditional random utility models, as demonstrated in a strategic context in Signorino (1999, 2003).

We will assume throughout this paper that the α are distributed Type I Extreme Value, leading to logit probabilities.[5] B's probabilities, p_r and p_ℓ, of playing right and left, respectively, are

$$p_r = \frac{e^{U_B(Rr)}}{e^{U_B(R\ell)} + e^{U_B(Rr)}} \tag{3}$$

$$p_\ell = \frac{e^{U_B(R\ell)}}{e^{U_B(R\ell)} + e^{U_B(Rr)}} \tag{4}$$

Now consider A's choice between playing Right and Left. If A plays Left, the game ends. However, A's decision to pick Right depends on what she thinks B will do. Because the α's are private information, A is uncertain of B's action and, therefore, must estimate the probability that B will play right or left. Therefore, A's utility for playing Right is an expected utility over the lottery consisting of B's choice, with probabilities as above: p_r, and p_ℓ. Player A's utilities are thus

$$U_A^*(L) = U_A(L) + \alpha_L \tag{5}$$

$$U_A^*(R) = EU_A(R) + \alpha_R \tag{6}$$

$$= p_\ell U_A(R\ell) + p_r U_A(Rr) + \alpha_R, \tag{7}$$

[5]The techniques demonstrated here can be implemented in exactly the same fashion if one assumes the α are Normally distributed, resulting in probit probabilities.

We will assume that the analyst shares the same uncertainty as the players. With the same assumptions as before, the probabilities of A playing Right versus Left are logit probabilities. The twist here is that they are based on an expected utility calculation and do not take the same form as logit probabilities in typical statistical analyses. A's probabilities, p_R and p_L, of picking Right and Left, respectively, are

$$p_L = \frac{e^{U_A(L)}}{e^{U_A(L)} + e^{EU_A(R)}} = \frac{e^{U_A(L)}}{e^{U_A(L)} + e^{p_\ell U_A(R\ell) + p_r U_A(Rr)}} \tag{8}$$

$$p_R = \frac{e^{EU_A(R)}}{e^{U_A(L)} + e^{EU_A(R)}} = \frac{e^{p_\ell U_A(R\ell) + p_r U_A(Rr)}}{e^{U_A(L)} + e^{p_\ell U_A(R\ell) + p_r U_A(Rr))}} \tag{9}$$

For any specification of the utilities, (p_ℓ, p_r, p_L, p_R) are the equilibrium probabilities of the statistical strategic model. Since we have assumed that the uncertainty enters as Type I extreme value perturbations to the action utilities, (p_ℓ, p_r, p_L, p_R) is a Logit Quantal Response Equilibrium (see McKelvey and Palfrey (1998) and Signorino (1999)). Because the α terms are assumed to be independently distributed, the equilibrium probabilities for the outcomes are just the product of the choice probabilities along each outcome's path:

$$\Pr[L] = p_L \tag{10}$$

$$\Pr[R\ell] = p_R \cdot p_\ell \tag{11}$$

$$\Pr[Rr] = p_R \cdot p_r \tag{12}$$

Having specified the statistical strategic model, a natural question concerns how the equilibria of the statistical model compare to the subgame perfect equilibria of the model with complete information — i.e., without the shocks to players' expected utilities. The intuition for this is best explained with reference to the variance of the error terms, assuming

the observed utilities are held constant. As the variance of the error terms goes to zero, the model becomes one of players with perfect and complete information, who maximize their utility at each decision point. Therefore, in the limit, the equilibria are subgame perfect. When the variance of the error terms is non-zero but extremely small relative to the observed utilities, the statistical equilibria often resemble a smoothed version of the subgame perfect equilibria. As the variance of the error terms increase, the equilibria may at times resemble smoothed versions of the subgame perfect equilibria, but may at other times look very different from the subgame perfect equilibria because the choice probabilities affect expected utility calculations.

The last step in specifying the model is to assign regressors to the utilities and then estimate the parameters accompanying those regressors. The statistical strategic model is just a strategic, random utility model. Therefore, the equilibrium choice and outcome probabilities provide a probability model that can be used in maximum likelihood estimation. The method of estimation — and, specifically, a new, simpler method for doing so — is the subject to which we now turn.

3 STATISTICAL BACKWARDS INDUCTION

We provide a technique in this section that allows practitioners to estimate recursive strategic models (of any depth or breadth) requiring only variations of logit or probit in standard statistical packages. This technique has a number of nice properties in relation to the system estimator currently used. Before turning to the new procedure, we first specify the system estimator.

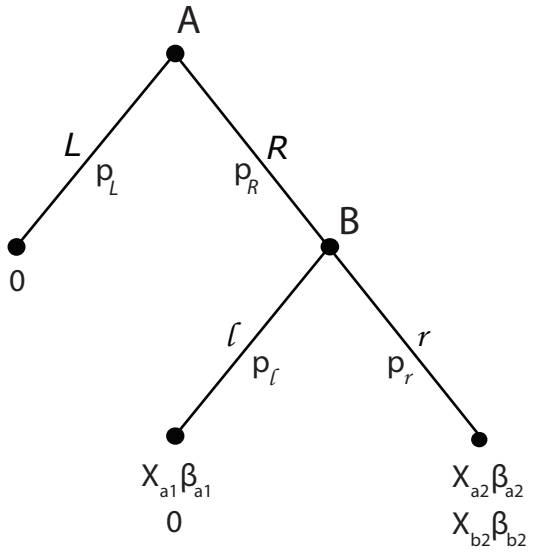


Figure 2: Two-Player Model with Binary Choices.

3.1 The System Estimator

To date, empirical analyses employing statistical strategic models have employed a "system" approach to estimation — where all parameters are estimated simultaneously for the entire model (see, for example, Signorino (1999, 2003, 2002); Signorino and Tarar (2006); Signorino and Yilmaz (2003); Guo (2002); Carter (2005); Quackenbush (2005); Carson (2003, 2005), Carson and Marshall (2004)).[6] To compare the system estimator to Statistical Backwards Induction [SBI], it will be helpful to change our notation slightly.

Consider the game in Figure 2, which is a more fully specified version of the model in Figure 1. Let us now assume that our data is coded as

$$y_A = \begin{cases} 1 & \text{if } U_A^*(R) \geq U_A^*(L) \\ 0 & \text{if } U_A^*(L) > U_A^*(R) \end{cases} \tag{13}$$

[6]STRAT currently implements a systems approach to maximum likelihood estimation of the parameters. This is also referred to as a "full information maximum likelihood" estimator.

and

$$y_B = \begin{cases} 1 & \text{if } U_B^*(r) \geq U_B^*(\ell) \\ 0 & \text{if } U_B^*(\ell) > U_B^*(r) \end{cases} \tag{14}$$

where $y_A = 1$ and $y_B = 1$ correspond to A and B choosing R and r, respectively. The equilibrium action probabilities are the same as before. According to the new coding, they are $p_R = \Pr(y_A = 1)$ and $p_r = \Pr(y_B = 1)$.

Figure 2 also displays a very simple specification of regressors for the utilities. Here, we have normalized $U_A(y_A = 0) = 0$ and $U_B(y_A = 1, y_B = 0) = 0$. The remaining utilities are specified with a single regressor and parameter:

$$\begin{aligned} U_A(y_A = 1, y_B = 0) &= X_{a1}\beta_{a1} \\ U_A(y_A = 1, y_B = 1) &= X_{a2}\beta_{a2} \\ U_B(y_A = 1, y_B = 1) &= X_{b2}\beta_{b2} \end{aligned}$$

Assuming we have data for all outcomes and regressors, then the system estimator would maximize the log of the following likelihood with respect to the regression coefficients β_{a1}, β_{a2}, and β_{b2}:

$$L = \prod^{n} p_L^{(1-y_A)} \cdot (p_R\, p_\ell)^{y_A(1-y_B)} \cdot (p_R\, p_r)^{y_A\, y_B}$$

where the observation index i has been dropped. In sum, for the system estimator, we construct a probability model for all of the actions and outcomes, and then use that to form the likelihood to be maximized. All parameters for all of the players are then estimated simultaneously.

3.2 Statistical Backwards Induction Estimator

In order to illustrate estimation by Statistical Backwards Induction (SBI), it is helpful to think of the strategic model as a recursive system of equations (see also Signorino (2002) and Signorino and Yilmaz (2003)). The main insight is that a recursive system of equations can be estimated equation-by-equation. Therefore, a recursive statistical strategic model can be estimated by univariate analogs of backwards induction.

3.2.1 The Strategic Model as a Recursive System of Equations

Consider the strategic model from the previous section and the data coding in equations 13 and 14. These imply the following system of latent variable equations:

$$y_A^* = U_A^*(R) - U_A^*(L) \tag{15}$$

$$y_B^* = U_B^*(r) - U_B^*(\ell) \tag{16}$$

where the data are coded as $y_j = 1$ if $y_j^* \geq 0$, and $y_j = 0$ if $y_j^* < 0$, for $j \in \{A, B\}$. Given the regressor specification in Figure 2, the system can be written as

$$y_A^* = p_\ell \, X_{a1} \, \beta_{a1} + p_r \, X_{a2} \, \beta_{a2} + \epsilon_A \tag{17}$$

$$y_B^* = X_{b2} \, \beta_{b2} + \epsilon_B \tag{18}$$

If we assume the α perturbations to the expected utilities are distributed Type I Extreme Value, then the ϵ_j terms are distributed logistic (McFadden (1974)). The resulting choice

probabilities $p_r = \Pr[y_B^* > 0]$ and $p_R = \Pr[y_A^* > 0]$ will be logit probabilities.[7]

It bears reiterating that this system of latent variable equations is completely consistent with the statistical game with agent error — it is just a different way of writing the same model. Moreover, the probabilities are all logit probabilities, but, again, where the expected utility calculations are explicitly modeled.

Because the α shocks are assumed independent, the ϵ terms are also independent. Therefore, Equations 17–18 are not "linked" through their error terms. Moreover, the system of equations is recursive, which, in theory, would imply that we can "start from the bottom and work up," estimating first the equation for y_B^*, and then using that information in the estimation of y_A^*. Indeed, it is not coincidental that the game-theoretic model can be solved via backwards induction and that this statistical model can be estimated in an analogous fashion. The only problem with this setup so far is that, although the equations for y_B^* and y_A^* result in logit probabilities, the systematic component of y_A^* does not take the same functional form as in typical logit models. This is, of course, due to the expected utility calculation. If Figure 2 generated the data and we were to use logit with the typical first-order linear $X\beta$ specification for each equation, then our parameter estimates would be biased and inconsistent (see Signorino and Yilmaz (2003)). However, as we will show, this problem can be easily overcome.

3.2.2 The Basic SBI Procedure

Statistical Backwards Induction requires three components common to empirical political science: (1) standard logit or probit regression (e.g., in Stata, SPSS, etc), (2) the ability to calculate predicted probabilities from the aforementioned regression, and (3) the ability to generate new variables by multiplying two existing variables. The basic idea for the

[7] Probit probabilities result from an assumption of Normally-distributed error terms.

SBI is to estimate the system equation-by-equation using standard logit, but transforming regressors in expected utility calculations into "new" regressors that can be used in standard logit estimation. Consider our strategic system of equations in Equations 17–18 describing Figure 2. For our referent example, the basic SBI estimation procedure is the following:

1. **B's Choice**. Since $y_B^* = X_{b2}\beta_{b2} + \epsilon_B$ does not require information concerning y_A^*, logit may be used to estimate $\hat{\beta}_{b2}$. Once $\hat{\beta}_{b2}$ is obtained, calculate $\hat{p}_r$ and $\hat{p}_\ell = 1 - \hat{p}_r$.

2. **A's Choice**. Substitute $\hat{p}_\ell$ and $\hat{p}_r$ into A's equation, giving

$$y_A^* = \hat{p}_\ell X_{a1}\beta_{a1} + \hat{p}_r X_{a2}\beta_{a2} + \epsilon_A$$

   To use standard statistical packages, we must first create the transformed regressors $Z_{a1} = \hat{p}_\ell X_{a1}$ and $Z_{a2} = \hat{p}_r X_{a2}$. Logit may then be used to estimate the parameters in the equation

$$y_A^* = Z_{a1}\beta_{a1} + Z_{a2}\beta_{a2} + \epsilon_A$$

It is important to note that in each stage, the transformed latent variable regression equation is exactly the same as the original, just written in a different way. Therefore, although the parameters in the transformed equations are associated with transformed regressors, the interpretation of the parameters is the same as in the original equation. For example, $\hat{\beta}_{a2}$ is still the estimated effect of X_{a2} on A's utility for outcome $(y_A = 1, y_B = 1)$. The transformation of the regressors is only done to allow for estimation in standard computer packages, and it does not affect the interpretation of the parameter estimates.

Statistical Backwards Induction implies a number of nice properties above and beyond the ease of implementation. Logit estimation in each stage will yield consistent estimates of the parameters and functions of the parameter estimates, such as the equilibrium choice

and outcome probabilities. The consistency of the parameter estimates derives from their status as maximum likelihood estimators. Obviously, system estimators would also yield consistent estimates. Indeed, because the equations are recursive and because the error terms are uncorrelated, we would expect both methods (system and SBI) to yield virtually identical results. However, they may not — and there are additional benefits to using SBI.

First, SBI often provides a feasible estimator, when the system estimator might be much more difficult to estimate. The likelihood function for the system estimator is not guaranteed to be globally concave. That is not to say that there are multiple extremum, but rather that the likelihood function may only be quasi-concave. It is easy to show just from plotting the likelihood that as the number of observations becomes small and as the variance of the error term becomes small, the likelihood function becomes increasingly "step-like," reflecting the fact that the equilibrium probabilities closely approximate the subgame perfect step response. The effect of this is that system estimation may have problems with weak (or fragile) identification. In contrast, logit and probit have globally concave likelihood functions, resulting in much easier and more rapid numerical optimization.[8]

Second, SBI is fast. Optimized algorithms exist for implementing logit (or probit and their variants) in a host of software packages, while maximizing a user-specified likelihood is often a slow and painstaking process that heavily depends on the quality of starting values. Across a host of tests and implementations, the SBI strategic model iterates and converges more rapidly and requires fewer iterations than an identical systems approach.

Finally, SBI retains more data for estimation. In practice, the system method requires all of the data appearing anywhere in the game to be sent simultaneously to a procedure that calculates the log-likelihood of each observation. Although a number of techniques exist

[8]Monte carlo simulations suggest a rather rapid decay in the convergence properties of system estimators (in Stata) that posed no problem for SBI.

for dealing with missing data, listwise deletion is still often the default method. In this case, if any observation contains a missing data point, the entire row is generally deleted. So for example, if a variable in A's decision contains missing data, then the entire row will be deleted, even if that variable does not appear in B's decision. This produces a needless loss of data in estimating B's parameters. Because the system is recursive, and because the SBI approach estimates each equation individually, missing data outside a given subgame will not affect the sample for the equation associated with that subgame. Therefore, the iterative approach is able to use as much data as possible in estimating the parameters.

Finally, we show in the appendix that the SBI estimator produces consistent estimates of β_{a1}, β_{a2}, and β_{b2}. It also produces consistent estimates of the standard error of β_{b2}. However, because $\hat{p}_\ell$ and $\hat{p}_r$ are substituted into A's equation and the resulting $Z_{a1} = \hat{p}_\ell X_{a1}$ and $Z_{a2} = \hat{p}_r X_{a2}$ variables are treated as data (without estimation uncertainty), SBI will yield biased estimates of the standard errors of β_{a1} and β_{a2}. Fortunately, this is easily remedied by a nonparametric bootstrap.

3.3 SBI with Bootstrapped Standard Errors

To retain the simplicity of Statistical Backwards Induction, we rely on a simulation method for calculating the standard errors — one that adds only a single step to the procedure outlined in the previous section. The approach we use is a form of the nonparametric bootstrap. The bootstrap procedure is quite simple. M iterations of the bootstrap are run. During each iteration $m = 1, 2, \ldots, M$ of the bootstrap, a sample is randomly drawn with replacement from the original data. SBI is then used to estimate parameters governing B's choice, generate new regressors, and estimate A's parameters (which are saved). After the M iterations of the bootstrap procedure, the standard errors of A's parameter estimates are calculated by simply taking the standard deviations of the saved estimates.

As mentioned previously, the bootstrap correction is only necessary for the estimation of A's parameters. Because B's actions do not depend on any auxiliary parameters, the first-stage maximum likelihood standard errors are consistent. It is only when we turn to choices that depend on the expected choices of others that a correction for the presence of a random action probability is necessary. With this need for correction in mind, we compare the system and SBI approaches for estimating statistical strategic models in a Monte Carlo experiment.

3.4 Monte Carlo Analysis

For the Monte Carlo analysis, we assumed a similar regressor specification as in Figure 2. We have normalized $U_A(y_A = 0) = 0$ and $U_B(y_A = 1, y_B = 0) = 0$. The remaining utilities are specified as follows:

$$\begin{aligned} U_A(y_A = 1, y_B = 0) &= X_{a2}\beta_{a2} \\ U_A(y_A = 1, y_B = 1) &= X_{a3}\beta_{a3} + X_c\beta_{a3c} \\ \\ U_B(y_A = 1, y_B = 1) &= \beta_{b0} + X_{b3}\beta_{b3} + X_c\beta_{b3c}. \end{aligned}$$

where $\beta_{a2} = \beta_{a3} = \beta_{a3c} = \beta_{b3} = 1$, $\beta_{b3c} = -1$ and $\beta_{b0} = \pi$. X_c is a common regressor that appears in both players' utilities. Data were generated based on the behavioral assumptions of the game, with X uniformly distributed over $[-2\pi, 2\pi]$, ϵ distributed logistic, with $V(\epsilon) = \pi^2/3$. Simulations were run for sample sizes N=500 and N=5000. Once data were generated, the parameters were estimated using both the system and SBI methods, and correct standard errors for SBI are calculated using the bootstrap with 1000 bootstrap iterations. Simulations were repeated 5000 times for N=500 and 2000 times for N=5000 to

form the densities of the estimators.

	N=500		N=5000	
Parameter Estimates	$\hat{\beta}_{a3}$	$\hat{\beta}_{a3c}$	$\hat{\beta}_{a3}$	$\hat{\beta}_{a3c}$
System	1.096	1.090	1.037	1.034
	(.224)	(.223)	(.061)	(.058)
SBI	1.047	1.043	1.024	1.021
	(.227)	(.226)	(.065)	(.063)
Standard Errors	$\widehat{se(\hat{\beta}_{a3})}$	$\widehat{se(\hat{\beta}_{a3c})}$	$\widehat{se(\hat{\beta}_{a3})}$	$\widehat{se(\hat{\beta}_{a3c})}$
System	.215	.211	.062	.061
	(.054)	(.055)	(.004)	(.004)
Bootstrapped SBI	.255	.252	.067	.067
	(.083)	(.084)	(.006)	(.006)

Table 1: Monte Carlo Results. *The first number in each cell is the mean of the monte carlo density. The number in parentheses is the standard deviation.*

Table 1 summarizes the results of the Monte Carlo analysis for β_{a3} and β_{a3c}. The first number in each cell is the mean of the monte carlo density. The number in parentheses is the standard deviation. As we previously noted, the SBI estimator will produce consistent estimates for the parameters (i.e., the β's) for all players in the game, as well as for the standard errors for B. The upper section of Table 1 displays the monte carlo sampling distributions of $\hat{\beta}_{a3}$ and $\hat{\beta}_{a3c}$ for the system and SBI estimators. First, note that both the SBI and system estimators are consistent; they both recover the true parameter values on average.[9] Second, although the SBI estimator uses less information than the system estimator,

[9] Although not shown here, the results for β_{a2} are similar to those reported in Table 1.

it is nearly as efficient for samples of at least 500. Indeed, it is also useful to note that the standard deviation of the system estimator's density is the gold standard for the standard error estimators. So, for samples of 500, we would like the average estimated standard error to be close to .22, while for N=5000, it should be close to .06.

For the monte carlo, A's estimated standard errors were calculated in two ways. The "system" standard errors are based on the (estimated) asymptotic variance from systems estimation. The "bootstrapped SBI" standard errors are produced by running the bootstrap procedure on the SBI estimator and then taking the standard deviation of the saved parameter estimates.

The lower section of Table 1 displays the average estimated standard error based on system and bootstrapped SBI estimation. The system standard errors are obviously very close to the standard deviations of the sampling distributions in the upper section. More importantly for our purposes here, the bootstrapped SBI standard errors are also quite close. As is to be expected, the bootstrapped SBI standard errors are slightly larger than the system standard errors. This is because the SBI method is itself less efficient than the system method. Moreover, the bootstap method will induce some inefficiency. Nevertheless, the sampling properties of the SBI method and of bootstapping imply that the boostrapped SBI standard errors will be consistent. In fact, we can see that as the sample increases to 5000, the difference in the standard errors is quite small.

Although not reported here, the SBI standard errors for B's utilities will be larger than the corresponding system standard errors. SBI uses a smaller sample to estimate the coefficients at "lower" nodes, so this again comes as little surprise. Because the system is recursive, SBI standard errors for this step are consistent. However, SBI is less efficient than the systems estimator, which simultaneously utilizes all available information.

With a theoretical model and its statistical counterpart in hand, we illustrate the power

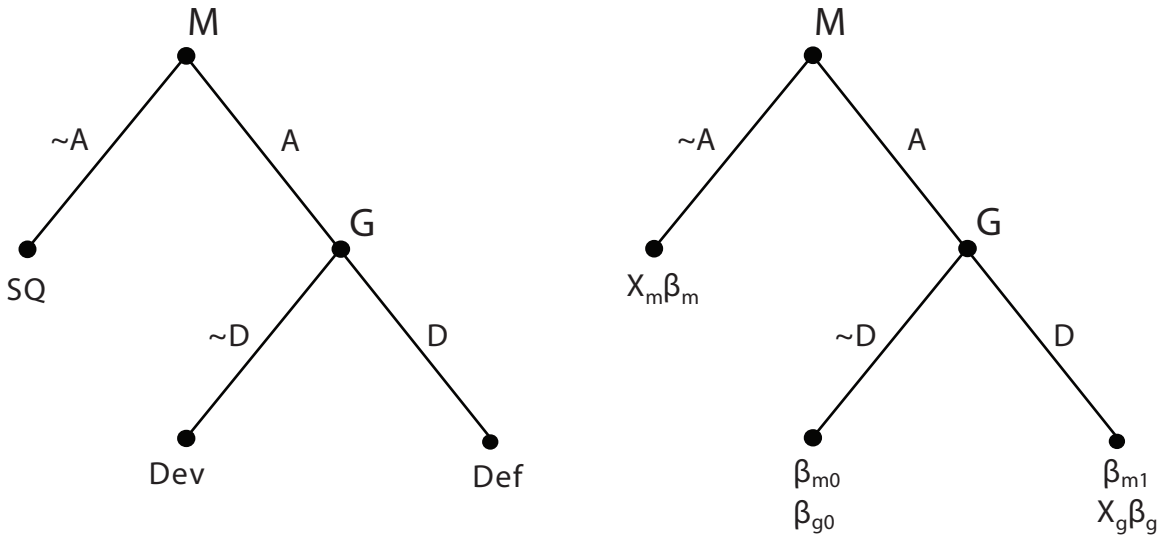
of this technique with real data in the next section.

4 DATA ANALYSIS

For most countries, exchange rate policy is the single most important macroeconomic policy decision to be made (Cooper 1999). Because there are costs to exchange rate variability, many states choose to fix their currencies to some external anchor, e.g. the US dollar or the Euro. When fixed exchange rates are chosen governments must control fiscal and monetary policy to maintain a par with a low inflation currency. As the ill effects of fiscal and monetary excess compound, arbitrage opportunities arise that may lead individual investors to take financial positions that combine to "attack" the declared par value of a currency (in terms of some other currency). Governments now confront a difficult choice, to expend resources in defense of the par value or to allow the currency to devalue, with all the concomitant distributional issues accompanying devaluation. A game theoretic rendition is presented in Figure 3. It is this basic decision structure – market decisions to engage in speculative attacks and governmental responses to these attacks – that we use to compare system estimation to Statistical Backwards Induction.

Leblang (2003) analyzes the strategic aspect of speculative attacks on currencies in international political economy using the model in Figure 3. Using Signorino's (1999) ("system") method, he estimates a model of strategic interaction between speculators in currency markets and policymakers in governments. To be clear, he analyzes a sample of monthly data from states that are nominally democratic[10] with pegged exchange rates. Because we are merely interested in demonstrating the similarities among the estimation methods, we necessarily give short shrift to the substantive implications of the results.

[10]Countries with Polity scores greater than 5.



(a) Speculative Attack Game

(b) With Regressors

Figure 3: Leblang's Model of Speculative Attacks on Currencies.

The variables that measure market utilities (U_M) for the status quo, defense, and devaluation, and government utility (U_G) for defense and devaluation are listed in Table 2. In general terms, markets are argued to attack currencies when there is a disconnect between the fixed exchange rate and the equilibrium rate of exchange [proxied by Reserves, Real Exchange Rate Overvaluation, Credit Growth], a history of questionable fixes [proxied by Prior Attacks], and/or greater structural incentives to challenge a government's resolve [proxied by Contagion, US Interest Rates, and Debt Service]. Similarly, Governments are argued to defend pegged exchange rates when it is politically expedient [proxied by Unified Government, Campaign/Election. Right Government, Export Sector, and Post Election]. At the same time, Governments must consider the resources at their disposal with which to combat speculative attacks [proxied by Interest Rates, Capital Controls, and Reserves]. With general concepts and measured factors determining the market's utility for specula-

tive attack and government responses in mind, we provide estimates of statistical strategic models in Table 3.

Variables	Utility	Sign	Measures
Unified Government	G	+	Binary: Party that controls the executive also controls the lower house of the legislature[a]
Export Sector$_{t-1}$	G	-	Log($\frac{Exports}{GDP}$)$_{t-1}$: Size of the export sector relative to GDP[b]
Campaign/Election	G	+	Binary: Three month campaigns and the month of elections[a]
Post Election	G	-	Binary: Three months postdating every election month[a]
Right Government	G	+	Binary: Governments classified Left, Center, and Right according to Left-Right positions on state control of the economy[a]
Interest Rates$_{t-1}$	G	+	Deflated discount rates, money market rates, or deposit rates depending on availability[c]
Capital Controls$_{t-1}$	G,M	+,+	Binary: Controls on the capital account[d]
Reserves	G,M	+,+	Log($\frac{Reserves}{BaseMoney}$)$_{t-1}$: Total reserves minus gold to base money [M0][c]
RER Overvaluation	M	-	Hodrick-Prescott residuals of the monthly real exchange rate[c]
Credit Growth$_{t-1}$	M	-	Rate of growth in domestic credit[c]
US Interest Rates$_{t-1}$	M	-	Interest rate on 90 day US deposits[c]
Debt Service$_{t-1}$	M	-	Total interest and debt repayment in foreign currencies[c] – IMF, short, and long term loans – as a percentage of exports[b]
Contagion	M	-	Number of speculative attacks outside country i in month t.
Prior Attacks	M	-	Number of speculative attacks for country i in all months preceding t.

[a] Database of Political Institutions, Beck, Clarke, Groff, Keefer and Walsh (2001) updated by Leblang (2003).

[b] World Development Indicators, (World Bank N. d.).

[c] International Financial Statistics, (International Monetary Fund N. d.*b*).

[d] Annual Report on Exchange Arrangements and Exchange Restrictions, International Monetary Fund (N. d.*a*).

Table 2: Variables and Measures from Leblang (2003)

4.1 Results

The replication results using system and SBI estimators are presented in Table 3. The table is divided into an upper and lower section. In the upper section, the parameter estimates for the constant terms ($\beta_{m0}, \beta_{g0}, \beta_{g1}$) are presented. In the lower section, the estimates associated with regressors in X_m and X_g are presented. The column labeled *System* replicates the system estimator results in Leblang's (2003) Table 2. The columns labeled *SBI* replicate Leblang's results, but using Statistical Backwards Induction. Bootstrapping is used to correct the bias in standard errors for Markets for reasons described in Section A.2. The far left column presents the name of the variables while the remaining columns reflect the estimation method and the estimated coefficient relating that variable to the utility for the outcome described by the column heading.

The most prominent result is that reported statistical significance does not differ across the estimation techniques. Put simply, Statistical Backwards Induction results in inferences that do not at all diverge from the systems estimator. The magnitudes and signs of the estimates also do not substantially differ between the systems and SBI estimators providing further evidence of the equation-by-equation technique's value. The standard errors are also quite similar. This replication exercise demonstrates a simplified technique for the analysis of currency crises that is both useful and closely mirrors its system equivalent.

Assessing the fit of the general model, the bottom row of Table 3 shows a very small difference between the log-likelihoods of the system and SBI approaches to estimating this statistical strategic model. Comparing the sum of the log-likelihoods from the SBI estimator to the system log-likelihood, the resulting difference is .0466.[11] The tiny differences in realized values confirm that this simplified method is very similar to the full systems estimator. With these similarities in mind, we turn to the evidence regarding government

[11]This result is obtained by subtracting the sum of SBI log-likelihoods from the system log-likelihood.

Constants	System		SBI	
	Market	*Government*	*Market*	*Government*
β_{g0}		-.43		.20
		(.78)		(.78)
β_{m0}	-3.66*		-4.05*	
	(.30)		(.47)	
β_{m1}	-3.14*		-3.41*	
	(.29)		(.50)	

Variable	System		SBI	
	Market β_m	*Government* β_g	*Market* β_m	*Government* β_g
Unified Government		-.35		-.07
		(.35)		(.45)
Export Sector$_{t-1}$		-.20		-.29
		(.17)		(.23)
Campaign/Election		1.66*		2.23*
		(.75)		(.93)
Post Election		1.06		1.10
		(.59)		(.74)
Right Government		-.94*		-1.55*
		(.45)		(.65)
Interest Rates$_{t-1}$		1.93*		1.33*
		(.64)		(.69)
Capital Controls$_{t-1}$	-.45	.07	-.42	.67
	(.25)	(.75)	(.47)	(.79)
Reserves$_{t-1}$	.23*	.31*	.29*	.59*
	(.06)	(.17)	(.07)	(.21)
RER Overvaluation	-.44*		-.46*	
	(.09)		(.18)	
Credit Growth$_{t-1}$	-.06*		-.07*	
	(.03)		(.04)	
US Interest Rates$_{t-1}$	-.05		-.05	
	(.06)		(.06)	
Debt Service$_{t-1}$	-.03		-.03	
	(.05)		(.05)	
Contagion	-.12*		-.13*	
	(.05)		(.05)	
Prior Attacks	-.12*		-.12*	
	(.05)		(.05)	
N	7240	7240	7240	88
Log-likelihood	-482.02		-432.27	-49.79

Table 3: Replication of Leblang's Results.

decisions to defend or devalue before assessing the determinants of speculative attacks.

4.1.1 To Defend or Devalue? (G)

We first assess government decisions to defend or devalue. *Unified Government* fails to reject the null hypothesis using both systems and Statistical Backwards Induction estimation. Furthermore, the astute reader will note that the estimates have overlapping 95% confidence intervals. Similarly, the *Export Section* is statistically insignificant without regard to the estimation technique and the 95% confidence intervals across the two estimation methods contain both estimates. The same holds for the *Post Election* period. The two estimates are very similar numerically and the confidence intervals have considerable overlap. *Capital Controls* fail to reject the null hypothesis of no effect and the parameter is sufficiently variable that the 95% confidence intervals overlap. With the support for the null hypothesis explored, we now complete the results for Government decisions.

Systems and SBI techniques provide identical inferences and similar estimates of the effects of exogenous factors on Government (G) utilities. For example, *Campaign/Election* periods increase the likelihood of defenses of par values without regard to the estimation technique. Furthermore, the systems and SBI estimates are separated by less than a single standard deviation. *Right Governments* are less likely to defend pegged exchange rates, all other things equal and this result does not depend on the estimation strategy. Though it might appear that the two estimates are quite different, each estimate is contained in the 95% confidence interval of the other. As *Interest Rates* increase, defenses become more likely and the 95% confidence intervals of both estimates show considerable overlap. Lastly, foreign exchange *Reserves* are liquid assets with which to defend an exchange rate; it is not surprising that *Reserves* increase the likelihood of defenses. Government utilities showcase minimal differences between systems estimation and SBI. We now turn to the

determinants of market decisions.

4.1.2 Speculative Attacks and the Status Quo (M)

Just as before, there is a strong congruence between the results of the two approaches to estimating statistical strategic models. For example, *Capital Controls* have no discernible effect on the utility of speculative attacks and the systems and SBI estimates are nearly numerically identical. Similarly, *Reserves*, as expected, deter speculation and the result does not depend on the estimator used. The estimated effects and their associated standard errors differ ever so slightly, but each estimate is within one standard deviation of the other suggesting significant overlap. Market utility for the status quo is increasing in the size of foreign exchange reserves.

Real Exchange Rate Overvaluation and *Credit Growth* both decrease market utility for the status quo with estimates that are nearly numerically identical. Without regard to the estimation technique employed, overvalued real exchange rates and high levels of domestic credit growth encourage speculative attacks. Turning to the effects of *US Interest Rates*, we see that the estimates are statistically insignificant but numerically identical. Similarly, *Debt Service* has no statistically discernible influence on market utilities for the status quo, but the results to two decimal places are identical for both the effect and its standard error. Continuing the theme of identical estimates, *Contagion* and *Prior Attacks* both decrease the utility that markets derive from the status quo; the estimated effects are almost identical while the associated standard errors are identical. To conclude the discussion of results, we briefly turn our attention to the remaining parameters.

There is no statistically significant difference between the market utilities for defense and devaluation in either the systems or SBI estimates as represented by the constant. Substantively, this implies that there is no evidence that markets necessarily prefer devaluations

to defenses given a speculative attack.

Two general themes emerge from this reanalysis of Leblang's (2003) work on currency crises. First, there are no important differences between Statistical Backwards Induction and systems methods of estimating statistical strategic models. Second, there is strong evidence of strategic interaction between states and markets in interactions that lead to non-events, currency crises and devaluations, and defenses of exchange rate pegs.

5 CONCLUSION

Our objective in this paper has been to provide a simple method for estimating recursive, statistical strategic models. Researchers can now use common commands (e.g., logit, probit, bootstrap) in their favorite statistical package to estimate these models. Moreover, the technique is flexible, allowing researchers to estimate models corresponding to a wide array of strategic situations.

As we have shown, this user-friendliness does come with a small cost: the SBI estimator is less efficient than the full information system estimator. However, for most data sets, the difference should be negligible. Indeed, we found no substantive differences between the system and SBI replications of the Leblang (2003) analysis.

The above cost, we believe, is well offset by the introduction of a large tool set for political science scholars. Deriving statistical models that are consistent with theoretical models is difficult enough. The programming requirements generally associated with these specialized models are often a roadblock to their implementation. By removing that roadblock, researchers can get on with the business of substantive research concerning strategic political behavior.

REFERENCES

Beck, Thorsten, George Clarke, Alberto Groff, Philip Keefer and Patrick Walsh. 2001. "New tools in comparative political economy: The Database of Political Institutions." *World Bank Economic Review* 15(1):165–76.

Carson, Jamie L. 2003. "Strategic Interaction and Candidate Competition in U.S. House Elections: Empirical Applications of Probit and Strategic Probit Models." *Political Analysis* 11(4):368–80.

Carson, Jamie L. 2005. "Strategy, Selection, and Candidate Competition in U.S. House and Senate Elections." *Journal of Politics* 67(1):1–28.

Carson, Jamie L. and Bryan W. Marshall. 2004. "Checking Power with Power: A Strategic Choice Analysis of Presidential Vetoes and Congressional Overrides." American Political Science Association Annual Meetings, September 2004.

Carter, Timothy A. 2005. United Nations Peacekeeping: United Nations Peacekeeping: Intervention Decisions and Efficacy PhD thesis Department of Political Science, University of Rochester.

Cooper, Richard N. 1999. Exchange Rate Choices. In *Rethinking the International Monetary System*, ed. Jane Sneddon Little and Giovanni P. Olivei. Boston, MA: Federal Reserve Bank of Boston, Conference Series No. 43 pp. 99–123.

Greene, William H. 2000. *Econometric Analysis, Fourth Edition*. New York, NY: Prentice-Hall, Inc.

Guo, Gang. 2002. Party Recruitment and Political Participation in Mainland China PhD thesis Department of Political Science, University of Rochester.

International Monetary Fund. N. d.*a*. *Annual Report on (Exchange Arrangements and) Exchange Restrictions*. Vol. Various Years Washington D. C.: International Monetary Fund.

International Monetary Fund. N. d.*b*. *International Financial Statistics*. Vol. Various Years International Monetary Fund.

Karaca-Mandic, Pinar and Kenneth Train. 2003. "Standard error correction in two-stage estimation with nested samples." *Econometrics Journal* 6:401–07.

Leblang, David A. 2003. "To Defend or to Devalue: The Political Economy of Exchange Rate Policy." *International Studies Quarterly* 47:533–59.

Lewis, Jeffrey B. and Kenneth A. Schultz. 2003. "Revealing Preferences: Empirical Estimation of a Crisis Bargaining Game with Incomplete Information." *Political Analysis* 11:345–67.

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

McFadden, Daniel L. 1974. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*, ed. Paul Zarembka. New York, NY: Academic Press pp. 105–142.

McKelvey, Richard D. and Thomas R. Palfrey. 1998. "Quantal Response Equilibria in Extensive Form Games." *Experimental Economics* 1:9–41.

Murphy, Kevin M. and Robert H. Topel. 1985. "Estimation and Inference in 'Two-Step' Econometric Models." *Journal of Business and Economic Statistics* 3(October):370–80.

Newey, Whitney K. 1984. "A Method of Moments Interpretation of Sequential Estimators." *Economic Letters* 14:201–6.

Newey, Whitney K. and Daniel L. McFadden. 1994. Large Sample Estimation and Hypothesis Testing. In *Handbook of Econometrics, Volume 4*, ed. Robert F. Engle and Daniel L. McFadden. Amsterdam: Elsevier Science pp. 2111–2245.

Quackenbush, Stephen L. 2005. "Strategic Interaction and General Deterrence Deterrence." Department of Political Science, University of Missouri.

Signorino, Curtis S. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *American Political Science Review* 93(2):279–97.

Signorino, Curtis S. 2002. "Strategy and Selection in International Relations." *International Interactions* 28(1):93–115.

Signorino, Curtis S. 2003. "Structure and Uncertainty in Discrete Choice Models." *Political Analysis* 11(4):316–44.

Signorino, Curtis S. and Ahmer R. Tarar. 2006. "A Unified Theory and Test of Extended Immediate Deterrence." *American Journal of Political Science* July(Forthcoming).

Signorino, Curtis S. and K. Yilmaz. 2003. "Strategic Misspecification in Regression Models." *American Journal of Political Science* 47(3):551–66.

World Bank. N. d. *World Development Indicators*. Vol. Various Years Washington, DC: World Bank Development Data Group.

A PROPERTIES OF THE SBI ESTIMATOR

The general problem of multi-step estimation is considered in a broad econometric literature. Consistency is almost certain, because maximum likelihood estimators are consistent, as are functions of them. That said, it is instructive to investigate the equation-by-equation technique and to provide a general analysis of its properties. We first ask, is SBI estimation consistent?

A.1 Consistency

The SBI strategic estimator is a special case of the two-step maximum likelihood estimators described in Murphy and Topel (1985). Thus, consistency of the SBI estimator follows from the consistency of two step maximum likelihood estimation procedures. For simplicity, consider a simple case with two models, in which one model is embedded in the other:

$$y_1 = f_1(x_1, \theta_1) \tag{19}$$
$$y_2 = f_2(x_1, x_2, \theta_1, \theta_2) \tag{20}$$

Two-step maximum likelihood first estimates the parameter vector θ_1 by maximum likelihood. θ_2 is then estimated by maximum likelihood with $\hat{\theta}_1$ inserted in place of θ_1 as if it were known (Greene 2000). The consistency of the first step follows from the consistency of MLE estimators. Murphy and Topel (1985) show that estimating θ_2 from $f_2(y_2 \mid x_1, x_2, \hat{\theta}_1, \theta_2)$ is asymptotically equivalent to estimating it from $f_2(y_2 \mid x_1, x_2, \theta_1, \theta_2)$, therefore the second step is also consistent (Murphy and Topel 1985).

In our equation-by-equation strategic method, the first step consists of estimating B's choice probabilities by logit (or probit), and then using these estimates as if they were known in estimating A's choice probabilities with logit (or probit). Thus, it follows that both steps of the SBI estimator provide consistent estimates of players' choice probabilities.

A.2 Efficiency

Our analysis of efficiency is based on the considerable econometric literature on multi-step maximum likelihood estimators, see Murphy and Topel (1985) and Newey and McFadden (1994). We first examine the properties of multi-step estimators.

Newey and McFadden (1994, Theorem 6.2, p. 2180) characterize the general conditions where the presence of a first-step estimator influences asymptotic inference.[12] Subject

[12]Their analysis is of the general class of method of moments estimators.

to certain regularity conditions, the key to valid asymptotic inference relies on the consequences of consistency in the first-stage estimates. If consistency of the first-stage estimates is critical for the consistency of subsequent estimates, Newey and McFadden (1994) prove that the (estimated asymptotic) standard errors for the second stage are inconsistent. With reference to SBI estimation, Newey and McFadden show that the standard errors are generally incorrect because consistency of the estimated action probabilities "below" is critical for consistency of the parameters. Upon reflection, this may not be critical. Given the common practice of testing zero null hypotheses, we are more concerned with falsely rejecting true null hypotheses, indeed the power of tests is seldom if ever discussed in applied research. Unfortunately, a further difficulty for the equation-by-equation technique is suggested by Newey (1984).

Newey (1984) argues that second stage standard errors of equation-by-equation estimators will be incorrect and too small (in a positive semi-definite sense). An intuition arises from Karaca-Mandic and Train's (2003, p. 401) remark that "the covariance matrix of the second-stage estimator includes noise introduced by the first-stage estimates." Newey (1984) demonstrates that sequential generalized method of moments estimators will have undersized asymptotic standard errors when the covariances between the stages are zero and the (estimated) quantities are assumed to be known. Recall that the SBI estimation technique begins by calculating B's action probabilities and employing predicted values $\hat{p}_r$ as the expected actions of B to form new regressors $Z_A = [\hat{p}_r X_{a2}, \hat{p}_\ell X_{a1}]$. Employing the theory of the partitioned inverse, Newey partitions the estimation problem into first and second stages to demonstrate that a zero covariance among the stages (which we assume to be true because the players choices are independent) and treatment of the predictions as data requires that the estimated (asymptotic) standard errors will be too small because they fail to account for the sampling variance of the predicted probabilities. We first turn to an analytic characterization of the problem.

A.2.1 Efficiency: Analytics

For a two-player model as in Figure 2, we can derive the standard error correction for the second stage of the SBI estimator.[13] As a special case of maximum likelihood estimation, the first step of the equation-by-equation estimator finds $\hat{\beta}_B$ that solves the moment condition $g_B(\beta_B) = 0$ and then the second stage finds $\hat{\beta}_A$ that satisfies $g_A(, \beta_A, \hat{\beta}_B) = 0$, where

[13]The derivation follows closely Karaca-Mandic and Train (2003)

$$g_B(\beta_B) = \frac{1}{N_B}\sum_{i=1}^{N_B}\frac{\delta L_B(y_{Bi}\,\hat{\beta}_B)}{\delta\beta_B} \tag{21}$$

$$g_A(\beta_A,\hat{\beta}_B) = \frac{1}{N}\sum_{i=1}^{N}\frac{\delta L_A(y_{Ai};\hat{\beta}_A,\hat{\beta}_B)}{\delta\beta_A} \tag{22}$$

Taking a first order Taylor's expansion of first and second stage moment conditions around the true parameter values β_B^* and β_A^* we get

$$g_B(\beta_B^*) - R_1(\hat{\beta_B} - \beta_B^*) \stackrel{A}{=} 0 \tag{23}$$

$$g_A(\beta_A^*,\beta_B^*) - R_2(\hat{\beta_B} - \beta_B^*) - R_3(\hat{\beta_A} - \beta_A^*) \stackrel{A}{=} 0 \tag{24}$$

where $\stackrel{A}{=}$ denotes asymptotic equality and

$$R_1 = -plim\frac{\delta g_B(\beta_B^*)}{\delta\beta_B}$$

$$R_2 = -plim\frac{\delta g_A(\beta_A^*,\beta_B^*)}{\delta\beta_B}$$

$$R_3 = -plim\frac{\delta g_A(\beta_A^*,\beta_B^*)}{\delta\beta_A}$$

Let us further assume that R_1, R_2, and R_3 are nonsingular square matrices.

A.2.2 B's Decisions

We can use (23) to construct

$$\sqrt{N_B}(\hat{\beta_B} - \beta_B^*) \stackrel{A}{=} \frac{1}{\sqrt{N_B}}\sum_{i=1}^{N_B} R_1^{-1}\frac{\delta L_B(y_{Bi}\,\hat{\beta}_B)}{\delta\beta_B}$$

$$\sqrt{N_B}(\hat{\beta_B} - \beta_B^*) \stackrel{A}{=} \frac{1}{\sqrt{N_B}}\sum_{i=1}^{N_B}\Theta_B$$

where $\Theta_B = R_1^{-1}\frac{\delta L_B(y_{Bi}\ \hat{\beta}_B)}{\delta\beta_B}$. Using Θ_B as defined and the information matrix equality, we can write

$$\hat{\beta}_B \overset{A}{\sim} N(\beta_B^*, Var(\Theta_B)/N_B)$$

to characterize the asymptotic distribution of $\hat{\beta}_B$, a standard MLE. Of particular interest, the "bottom" of the tree contains all of the relevant information with which to assess the final node of a recursive decision problem. As a result, estimates obtained from the terminal node of an extensive form set of choices have all of the desirable properties of a univariate maximum likelihood estimator. *The SBI estimator is consistent and efficient at the terminal node of a recursive decision problem.*

A.2.3 A's Decisions

Unfortunately, the same cannot be said for "upper" decisions precisely because of the sampling distribution of the parameters $\hat{\beta}_B$ and the associated sampling distributions of functions of these parameters – $\hat{p}_\ell, \hat{p}_r$. Turning our attention to (24) and the parameters governing A's choices, we have

$$\begin{aligned}
(\hat{\beta_A} - \beta_A^*) &\overset{A}{=} R_3^{-1}(g_A(\beta_A^*, \beta_B^*) - R_2(\hat{\beta_B} - \beta_B^*)) \\
&\overset{A}{=} R_3^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\delta L_A(y_{Ai};\hat{\beta}_A,\hat{\beta}_B)}{\delta\beta_A} - R_2R_1^{-1}\frac{1}{N_B}\sum_{i=1}^{N_B}\frac{\delta L_B(y_{Bi}\ \hat{\beta}_B)}{\delta\beta_B}\right) \\
&\overset{A}{=} \frac{1}{N}R_3^{-1}\left(\sum_{i=1}^{N}\frac{\delta L_A(y_{Ai};\hat{\beta}_A,\hat{\beta}_B)}{\delta\beta_A} - R_2R_1^{-1}\frac{N}{N_B}\sum_{i=1}^{N}y_{Ai}\frac{\delta L_B(y_{Bi}\ \hat{\beta}_B)}{\delta\beta_B}\right) \\
\sqrt{N}(\hat{\beta_A} - \beta_A^*) &\overset{A}{=} \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Theta_{Ai}
\end{aligned}$$

where

$$\Theta_{Ai} = R_3^{-1}\frac{\delta L_A(y_{Ai};\hat{\beta}_A,\hat{\beta}_B)}{\delta\beta_A} - R_3^{-1}R_2R_1^{-1}\frac{N}{N_B}y_{Ai}\frac{\delta L_B(y_{Bi}\ \hat{\beta}_B)}{\delta\beta_B}.$$

This allows us to write the asymptotic distribution of $\hat{\beta}_A$ as

$$\hat{\beta}_A \overset{A}{\sim} N(\beta_A^*, Var(\Theta_A)/N).$$

In implementation, Θ_A, R_1,R_2, R_3 are calculated using the estimated parameters, and

the *plim* evaluations are replaced with their sample approximations.[14] If we assume that N/N_B is finite, $Var(\Theta_A)$ is given by

$$Var(\Theta_A) = \frac{1}{N}\sum_{i=1}^{N} \hat{\Theta}_{Ai}\hat{\Theta}'_{Ai}. \tag{25}$$

The standard error correction described above applies only for the two-stage SBI estimator, or when there are only two decision nodes to estimate. As the number of decision nodes increases, the number of iterations of the SBI estimator also increases, and deriving the exact standard error correction becomes practically impossible. A more practical approach that goes well with the ease of the SBI estimator is bootstrapping, a nonparametric method for correctly estimating upper stage standard errors. This approach uses simulation methods to recover the sampling distribution of the probabilities.

[14] Readers familiar with Murphy and Topel (1985) and Maddala (1983) will note the similarities.

B R AND STATA CODE FOR THE SBI ESTIMATOR AND BOOTSTRAPPING STANDARD ERRORS

In this section, we present example R and Stata codes for running the SBI estimator for the model in Figure 2 that we used for our Monte Carlo experiment. As we describe in Section 3 in more detail, $U_A(y_A = 0)$ and $U_B(y_A = 1, y_B = 0)$ are normalized to zero. The remaining utilities are specified as follows:

$$\begin{aligned} U_A(y_A = 1, y_B = 0) &= X_{a2}\beta_{a2} \\ U_A(y_A = 1, y_B = 1) &= X_{a3}\beta_{a3} + X_c\beta_{a3c} \\ \\ U_B(y_A = 1, y_B = 1) &= \beta_{b0} + X_{b3}\beta_{b3} + X_c\beta_{b3c}. \end{aligned}$$

In the following R code, we define a simple function `SBI` that runs two consecutive logit regressions to estimate A and B's utilities. We then use R's canned bootstrapping function `boot` from the **boot** library to calculate the bootstrapped standard errors.

```
SBI ←function(Data, i){

       BS ← Data[i,]                # Draw the bootstrap sample
       IVB ← BS[,3:4][BS[,1]==1,]      # B's regressors

# Step 1:  Run a logit to estimate B's utilities and calculate p̂_B

       logitB ← glm(BS[,2][BS[,1]==1]∼ IVB,
              family=binomial(link="logit"))

       bB ← logitB$coefficients
       pB ← exp(cbind(BS[,3:4],1) %*% b2)/(1+exp(cbind(BS[,3:4],
                 1) %*% b2))

# Step 2:  Transform A's regressors with pB

       IVA ← cbind((1-pB)*BS[, 5], vecmat(pB, BS[,6:7]))

# Step 3:  Logit for A's utilities

       logitA ← glm(BS[,1] ∼ IVA-1,family=binomial(link="logit"))

logitA$coefficients}


# Step 4:  Calculate standard errors with bootstrapping

boot(McData, SBI, 500)
```

The above SBI model can also be estimated in STATA very easily. In order to do so, we first define a function called `sbi` that runs two logits and estimates A and B's utilities. We then use STATA's canned `bs` function to calculate the standard errors.

```
capture program drop sbi
program sbi, rclass
   version 8.0
   syntax, ua2(varlist num) ua3(varlist num) ub3(varlist num) y1(varname)
          y2(varname)

* Step 1:  Run a logit to estimate B's utilities and calculate p̂_B
       logit 'y2' 'ub3' if 'y1'
       predict p2def

* Step 2:  Transform A's regressors with p2def

       foreach x of varlist 'ua3'{
       gen 'x'_t3 = 'x'*p2def
       }
       foreach x of varlist 'ua2'{
       gen 'x'_t2 = 'x'*(1-p2def)
       }

* Step 3:  Logit for A's utilities

       logit 'y1' *_t2 *_t3, nocons
       matrix b = e(b)

       local co = 1
       foreach x of varlist 'ua2' 'ua3'{
       ret scalar 'x' = b[1, 'co']
       local co = 'co'+1
       }
end

* Step 4:  Calculate standard errors with bootstrapping

bootstrap "sbi, ua2(x12) ua3(x13 xc) ub3(x23 xc) y1(y1) y2(y2)" r(x12)
       r(x13) r(xc), reps(500) dots
```