Supplementary Materials for: Detecting and Correcting for Separation in Strategic Choice Models (Not for publication)

Casey Crisman-Cox* Olga Gasparyan† Curtis S. Signorino‡

October 24, 2022

Contents

*Texas A&M University. Email: c.crisman-cox@tamu.edu

†Hertie School. Email: o.gasparyan@hertie-school.org

‡University of Rochester. Email: curt.signorino@rochester.edu

A The linear programming diagnostic

Following Konis (2007) we begin with a $D \times k$ design matrix W and a binary outcome y. When either complete or quasi-complete separation exists, there is a separating hyperplane through $\mathbb{R}^k$ that describes the separation problem (i.e., all values of y_d are 1 above the plane and 0 below it). For each observation $d = 1, \ldots, D$ we are interested in the total unscaled distance between w_d and the hyperplane is given by $\sum_{d=1}^{D} w_d' \gamma$. If there is separation, then $w_d' \gamma \geq 0$ when $y_d = 1$ and $w_d' \gamma \leq 0$ when $y_d = 0$. Let $\bar{W} = [(2y_d - 1)w_{d,1}, \ldots, (2y_d - 1)w_{d,k}]_{d=1}^{D}$ to make all the distances positive, and consider the following constrained optimization problem:

$$
\begin{aligned}
&\text{maximize } \mathbf{1}_D' \bar{W} \gamma \\
&\quad\text{s.t. } \bar{W}\gamma \geq 0 \\
&\qquad -\infty < \gamma_j < \infty \text{ for } j = 1, \ldots, k,
\end{aligned}
$$

where $\mathbf{1}_D$ is a length-D column vector of 1s.

If there is an overlap in the data, then the only solution to this problem is $\gamma = 0$ (a separating hyperplane does not exist). However, when there is either complete or quasi-complete separation, the problem is unbounded (Konis 2007, 79, Remark 1). To see this result, consider the case of (quasi-complete) separation. In this case, a (semi-)separating hyperplane will exist. Suppose that $\tilde{\gamma}$ is feasible for the linear program, with $\bar{W}\tilde{\gamma} \geq 0$ and $\mathbf{1}_D' \bar{W} \tilde{\gamma} > 0$. If $\tilde{\gamma}$ is a solution, then $c\tilde{\gamma}$ will also be a solution for any $c > 0$. As c grows, the objective function grows with it to infinity while still satisfying the constraint. Identifying separation then becomes as easy as checking the convergence of the above lp. The R function `detectseparation::detect_separation` conducts the diagnostic.

B Additional simulations and analysis

B.1 Effect of estimating p_B on estimating α_1

In the main Monte Carlo results presented in Table 1, we find that both the ordinary SBI and BR-SBI struggled in estimating α_1. In the case of the former, this is unsurprising as the estimate of p_B for the ordinary SBI is tainted by separation induced inflation in $\hat{\beta}_1$. In the case of the latter, this is surprising as the BR-SBI does reasonably well at estimating β_1. Additional analysis reveals that the distribution of BR-SBI estimates $\hat{\alpha}_1$ has an unusually long tail away from zero. This finding prompts the question: how much of the trouble here is about the two-step nature of the SBI? Put another way, how much of bias/variance in the SBI estimates of α_1 results from estimating p_B separately and in the presence of separation?

The results of this simulation are reported in Table B.1. Here, we see the ordinary SBI is greatly affected in its ability to estimate the parameter α_1 and that nearly all of this difficulty is due to how it estimates p_B. Notably, we see that the standard deviation over simulations decreases by several orders of magnitude, suggesting that even small issues with $\hat{p}_B$ can spill into and cause cascading bias in the second stage estimates. This trend is also true for the BR-SBI, where the standard deviation over simulations is much closer to the true standard error and the average standard error over simulations. Even though the BR-SBI still exhibits some sensitivity to estimating p_B it appears to be more robust to these issues than the ordinary SBI.

B.2 Separation in both utilities

Our second set of Monte Carlos looks at the problem when there is separation in both equations. In this case B's choice is given by

$$y_B = \mathbb{I}\left[0.25 + 4X_B + \varepsilon_B(1) - \varepsilon_B(0) > 0\right],$$

Table B.1: Analysis of $\hat{\alpha}$ when SBI uses the true p_B

Estimator	Statistic	α_0	α_1
Ordinary SBI	Expected value	1.50	-2.80
	St. Dev.	0.13	1.04
	St. Err.	0.13	1.58
	Coverage	0.95	0.96
	Power	1.00	0.99
BR-SBI	Expected value	1.50	-2.52
	St. Dev.	0.13	0.81
	St. Err.	0.13	0.73
	Coverage	0.95	0.93
	Power	1.00	1.00
Truth	Parameters	1.50	-2.50
	St. Err.	0.13	0.71

and $p_B = \Phi\left(\frac{0.25+4X_B}{\sqrt{2}}\right)$. Likewise, player A's choice is given as

$$y_A = \mathbb{I}\left[1.5 + 4(X_A p_B) + \varepsilon_A(1) - \varepsilon_A(0) > 0\right].$$

In terms of Figure 2, we have $\alpha_0 = -1.5$, $\alpha_1 = 4$, $\beta_0 = 0.25$, $\beta_1 = 4$. Here SF is the most likely outcome (about 3/4 of observations), with BD and SQ representing about 15% and 10% of outcomes, respectively. We keep results where the lp-diagnostic finds separation between either covariate and the outcomes BD and either (SQ or SF).[1]

The Monte Carlo results are reported in Table B.2. As before, there is striking improvement in bias, variance, RMSE and power when employing a BR estimator. Additionally, we again see that the average standard errors for the BR estimates where separation is not a problem are very close to the simulated standard deviation and average true standard errors, while the BR standard errors for the corrected estimates remain conservative relative to the simulated standard deviation of the sampling distribution.

One interesting point about the results in Table B.2 is the difference between ordinary SBI and FIML. Specifically, in estimating α_1 the ordinary SBI produces an average esti-

[1] Again, when the lp-diagnostic does not detect a problem, the six estimators produce nearly identical results.

Table B.2: Monte Carlo results when separation is present at both decision nodes.

Estimator	Quantity	α_0	α_1	β_0	β_1	RMSE
Ordinary SBI	Est	−1.50	8.57	0.25	8.10	7.48
	St.Dev.	0.14	4.26	0.12	0.13	
	St.Err.	0.14	469.59	0.12	370.52	
	Power	1.00	0.39	0.57	0.00	
	Coverage	0.95	0.97	0.95	1.00	
BR-SBI	Est	−1.50	3.66	0.25	3.87	0.77
	St.Dev.	0.14	0.64	0.12	0.12	
	St.Err.	0.14	1.03	0.12	0.70	
	Power	1.00	1.00	0.57	1.00	
	Coverage	0.95	0.92	0.95	1.00	
Ordinary FIML	Est	−1.50	6.16	0.25	6.75	4.26
	St.Dev.	0.14	2.36	0.12	0.60	
	St.Err.	0.14	11.10	0.12	71.71	
	Power	1.00	0.39	0.57	0.00	
	Coverage	0.95	0.97	0.95	1.00	
BR-FIML (Firth)	Est	−1.50	3.65	0.26	3.87	0.76
	St.Dev.	0.14	0.63	0.12	0.12	
	St.Err.	0.14	1.08	0.12	0.73	
	Power	1.00	1.00	0.60	1.00	
	Coverage	0.95	0.93	0.95	1.00	
BR-FIML (Cauchy)	Est	−1.51	3.88	0.26	4.31	0.86
	St.Dev.	0.14	0.76	0.12	0.13	
	St.Err.	0.14	1.27	0.12	1.13	
	Power	1.00	1.00	0.61	1.00	
	Coverage	0.95	0.94	0.95	1.00	
BR-FIML (log-F)	Est	−1.51	3.69	0.27	4.16	0.75
	St.Dev.	0.14	0.62	0.12	0.12	
	St.Err.	0.14	1.12	0.12	0.97	
	Power	1.00	1.00	0.63	1.00	
	Coverage	0.95	0.94	0.95	1.00	
Truth	Parameters	−1.50	4.00	0.25	4.00	
	St. Err. (SBI)	0.14	1.20	0.12	0.78	
	St. Err. (FIML)	0.14	1.22	0.12	0.82	

mate of nearly 9 while the ordinary FIML is close to 6. Indeed, the ordinary SBI tends to inflate the estimates associated with separation more than the ordinary FIML across simulations. The reason this difference emerges has to do with the underlying default optimization software. Ordinary probits are typically fit using iterative reweighted least squares (IRLS) which closely approximates the Newton-Raphson method for optimization, while the ordinary FIML is fit using the quasi-Newton BFGS algorithm by default. If we switch the FIML to Newton-Raphson or the SBI probit to BFGS the results tend to align more closely. However, these differences further demonstrate the numerical difficulties induced by separation. Two standard optimization algorithms can produce wildly different results as they try to find the maximum of a monotonic likelihood function and neither is the true maximum likelihood estimate of infinity. Note that these differences do not emerge once a penalty is added. The BR-SBI and BR-Firth tend to be very similar despite using the same differing optimization routines as their uncorrected counterparts. Likewise, changing the optimization routine for either of the density penalties does very little to the results.

B.3 Small sample size

The first additional simulation we consider is: how well do the estimators perform with a smaller sample size. This is an important check as separation problems are more common in smaller datasets, and the biases introduced by small samples can compound separation problems. Additionally, the Signorino and Tarar (2006) replication has a small sample, and we want to make sure that our proposed tools will work as expected in such a situation.

In the first exercise, we set $D = 50$ with parameters equal to the true values indicated in Table B.3. In the second one, we still have $D = 50$, but the parameters now set to the true values presented in Table B.4. Note that in both experiments we tweak the parameters compared to the large-sample simulations. This is done to accommodate the fact that it takes less extreme parameter values to induce separation with a small sample. As noted, separation is essentially a small-sample problem. These adjustments have the additional value of demonstrating that our main conclusions are not driven by specific choices in parameter values. In the first experiment we still only consider cases where the lp-diagnostic finds that the outcome BD is perfectly predicted, while in the second case we look at cases where the lp-diagnostic finds separation associated with the outcomes BD and either SF or SQ.

Table B.3: Coefficient estimates, standard errors, and root mean-squared error when separation is present in Player B's decision and $D = 50$

Estimator	Quantity	α_0	α_1	β_0	β_1	RMSE
Ordinary SBI	Est	−0.48	−2.17	−1.29	9.51	7.15
	St.Dev.	0.35	1.70	1.49	1.85	
	St.Err.	0.33	59.06	558.84	2311.15	
	Power	0.29	0.79	0.49	0.01	
	Coverage	0.94	0.95	0.98	1.00	
BR-SBI	Est	−0.47	−1.95	−1.03	3.57	1.34
	St.Dev.	0.34	0.83	0.56	0.60	
	St.Err.	0.33	0.78	0.55	1.24	
	Power	0.28	0.78	0.49	0.97	
	Coverage	0.95	0.96	0.98	1.00	
Ordinary FIML	Est	−0.49	−2.21	−1.18	7.23	4.82
	St.Dev.	0.35	1.62	0.94	1.27	
	St.Err.	0.33	1.24	6.56	4228.03	
	Power	0.29	0.77	0.54	0.01	
	Coverage	0.95	0.96	0.97	1.00	
BR-FIML (Firth)	Est	−0.48	−1.93	−1.02	3.59	1.41
	St.Dev.	0.35	0.91	0.54	0.63	
	St.Err.	0.34	0.81	0.53	1.31	
	Power	0.28	0.75	0.52	0.96	
	Coverage	0.94	0.95	0.97	1.00	
BR-FIML (Cauchy)	Est	−0.45	−1.88	−0.99	3.85	1.51
	St.Dev.	0.34	0.79	0.58	0.67	
	St.Err.	0.33	0.79	0.53	1.46	
	Power	0.25	0.74	0.48	0.97	
	Coverage	0.95	0.96	0.96	1.00	
BR-FIML (log-F)	Est	−0.44	−1.86	−0.90	3.58	1.24
	St.Dev.	0.33	0.73	0.50	0.52	
	St.Err.	0.34	0.79	0.51	1.33	
	Power	0.24	0.74	0.42	0.97	
	Coverage	0.96	0.96	0.96	1.00	
Truth	Parameters	−0.50	−2.00	−1.00	3.00	
	St. Err. (SBI)	0.33	0.78	0.52	1.02	
	St. Err. (FIML)	0.34	0.83	0.51	1.09	

In both small-sample experiments we see that the main results still mostly hold. The BR-SBI narrowly edges out the BR-FIML with Jeffreys prior or the Cauchy in one experiment, but these differences are fairly minor compared to the gaps between all BR and the two ordinary estimators. The log-F BR-FIML estimator continues to do very well. Overall,

Table B.4: Coefficient estimates, standard errors, and root mean-squared error when separation is present at both decision nodes decision and $D = 50$

Estimator	Quantity	α_0	α_1	β_0	β_1	RMSE
Ordinary SBI	Est	−0.20	−9.94	−0.78	8.83	13.79
	St.Dev.	0.38	9.93	1.74	2.39	
	St.Err.	0.36	19665.98	879.46	2845.98	
	Power	0.08	0.68	0.06	0.01	
	Coverage	0.94	0.94	0.98	0.99	
BR-SBI	Est	−0.16	−4.80	−0.49	2.82	3.38
	St.Dev.	0.38	2.63	0.63	0.73	
	St.Err.	0.35	1.78	0.67	1.37	
	Power	0.07	0.96	0.04	0.61	
	Coverage	0.94	0.94	0.98	0.99	
Ordinary FIML	Est	−0.21	−9.31	−0.57	6.51	15.77
	St.Dev.	0.37	13.91	0.73	1.46	
	St.Err.	0.36	20.35	0.88	11472.00	
	Power	0.09	0.33	0.08	0.01	
	Coverage	0.95	1.00	0.97	0.99	
BR-FIML (Firth)	Est	−0.18	−4.07	−0.29	2.53	2.59
	St.Dev.	0.39	2.09	0.50	0.72	
	St.Err.	0.36	1.76	0.55	1.33	
	Power	0.08	0.87	0.05	0.54	
	Coverage	0.94	0.96	0.98	0.97	
BR-FIML (Cauchy)	Est	−0.13	−3.93	−0.17	2.73	1.83
	St.Dev.	0.37	1.24	0.54	0.67	
	St.Err.	0.35	1.64	0.56	1.45	
	Power	0.07	0.91	0.02	0.45	
	Coverage	0.95	1.00	0.99	0.98	
BR-FIML (log-F)	Est	−0.11	−3.58	−0.11	2.63	1.34
	St.Dev.	0.35	0.81	0.48	0.55	
	St.Err.	0.35	1.39	0.55	1.39	
	Power	0.05	0.96	0.01	0.44	
	Coverage	0.95	1.00	1.00	0.98	
Truth	Parameters	−0.20	−3.00	−0.15	3.00	
	St. Err. (SBI)	0.35	0.98	0.63	1.70	
	St. Err. (FIML)	0.35	1.06	0.52	1.55	

these results provide additional evidence that separation can be very problematic for ordinary estimators and that any type of BR should be preferred in terms of RMSE.

B.4 Joint separation induced by a single variable

Next, we consider the effect of having a single variable that induces separation into both actors' utility calculations. Specifically, let B's choice be given by

$$y_B = \mathbb{I}\left[-0.5 + 0.25X_B + 5X + \varepsilon_B(1) - \varepsilon_B(0) > 0\right],$$

and A's decision is

$$y_A = \mathbb{I}\left[2 + (0.25X_A + 2.5X)p_B + \varepsilon_A(1) - \varepsilon_A(0) > 0\right].$$

Note that the variable X enters both the utility function for both A and B, while X_A and X_B are unique to their respective actors. Additionally, X_A and X_B are now distributed i.i.d standard normal, while X is distributed Bernoulli with mean 0.5. As in the main simulations we use a sample size of 500 and we only consider the cases where the lp-diagnostic finds separation associated with outcomes SQ and BD. The results are reported in Table B.5 and are largely consistent with the results in the main text. All the BR estimators perform better than their ordinary counterparts, and the BR-FIML estimators (with any correction) outperforms the BR-SBI.

B.4.1 Dropping offending variables

Some statistical software/routines (e.g., Stata's logit and probit) drop variables that lead to in-sample separation. On the one hand, this solution allows analysts to ignore the separation problem and continue their analysis. On the other hand, this strategy risks inducing omitted variable bias. In the strategic setting, dropping variables in B's choice equation may lead to not only omitted variable bias as it is commonly understood, but it may also lead to cascading bias through the endogenous quantity p_B. What are the consequences of dropping variables that lead to separation problems in this context? To assess this, we consider two additional simulations. In both cases, we tweak the DGP from

Table B.5: Coefficient estimates, standard errors, and root mean-squared error when a single variable produces separation in both players' utilities

Estimator	Quantity	α_0	α_1	α_2	β_0	β_1	β_2	RMSE
Ordinary SBI	Est	−2.03	0.26	5.96	−0.51	0.25	9.01	5.74
	St.Dev.	0.18	0.46	2.14	0.12	0.12	0.28	
	St.Err.	0.17	0.44	417.20	0.12	0.12	358.34	
	Power	1.00	0.09	0.20	0.99	0.54	0.00	
	Coverage	0.95	0.95	0.98	0.95	0.95	1.00	
BR-SBI	Est	−2.01	0.23	2.12	−0.50	0.25	4.72	0.79
	St.Dev.	0.17	0.40	0.41	0.12	0.12	0.15	
	St.Err.	0.17	0.36	0.67	0.12	0.12	0.70	
	Power	1.00	0.10	1.00	0.99	0.54	1.00	
	Coverage	0.95	0.94	0.90	0.95	0.95	1.00	
Ordinary FIML	Est	−2.03	0.26	4.52	−0.51	0.25	8.14	4.06
	St.Dev.	0.18	0.46	1.43	0.12	0.12	0.48	
	St.Err.	0.17	0.44	17.74	0.12	0.12	96.47	
	Power	1.00	0.09	0.20	0.99	0.55	0.00	
	Coverage	0.95	0.95	0.98	0.95	0.95	1.00	
BR-FIML (Firth)	Est	−2.01	0.23	2.13	−0.50	0.25	4.73	0.79
	St.Dev.	0.17	0.40	0.41	0.12	0.12	0.15	
	St.Err.	0.17	0.37	0.70	0.12	0.12	0.74	
	Power	1.00	0.09	1.00	0.99	0.55	1.00	
	Coverage	0.95	0.95	0.92	0.95	0.95	1.00	
BR-FIML (Cauchy)	Est	−2.04	0.22	2.51	−0.50	0.25	5.24	0.72
	St.Dev.	0.18	0.37	0.48	0.12	0.12	0.16	
	St.Err.	0.17	0.41	1.01	0.12	0.12	1.19	
	Power	1.00	0.05	0.99	0.99	0.54	1.00	
	Coverage	0.95	0.97	0.97	0.95	0.96	1.00	
BR-FIML (log-F)	Est	−2.03	0.22	2.48	−0.50	0.25	5.03	0.64
	St.Dev.	0.17	0.37	0.44	0.12	0.12	0.15	
	St.Err.	0.17	0.40	0.98	0.12	0.12	0.97	
	Power	1.00	0.05	0.99	0.99	0.54	1.00	
	Coverage	0.95	0.97	0.97	0.95	0.96	1.00	
Truth	Parameters	−2.00	0.25	2.50	−0.50	0.25	5.00	
	St. Err. (SBI)	0.16	0.40	0.94	0.12	0.12	0.92	
	St. Err. (FIML)	0.16	0.41	0.95	0.12	0.12	0.96	

the experiment in Appendix B.4, above, but we drop the X as it induces the monotonic likelihood problem.

In the first experiment (Table B.6), the regressors are correlated such that

$$\begin{pmatrix} X^* \\ X_A \\ X_B \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & 0.7 & -0.7 \\ 0.7 & 1 & -.1 \\ -0.7 & -.1 & 1 \end{bmatrix}\right)$$

$$x_d = \mathbb{I}(x_d^* > 0).$$

In the second experiment (Table B.7), the regressors are independent of each other:

$$X \sim \text{Bernoulli}(0.5)$$
$$X_A \sim N(0,1)$$
$$X_B \sim N(0,1).$$

The only specification concern here is the effect on $\hat{p}_B$. Here, we want to see the problems with dropping regressors that are associated with the in-sample issues even under the ideal condition of regressor independence.

In the case of correlated variables being dropped (Table B.6), we see both omitted variable and broader misspecification bias from dropping a variable that is relevant to p_B. Overall, all six estimators produce nearly identical results, but unfortunately, they are all quite bad. Notably, two estimates (α_1 and β_0) now have the wrong sign. Likewise, even though there is reasonable power against the null hypothesis that $\alpha_1 = 0$, the estimates are in the wrong direction leading to a mistaken directional inference. Even more shocking, the RMSE from all six estimators are effectively identical and uniformly worse than even the worst RMSE in Table B.5. This commonly used solution can be worse than the original problem.

In order to be sure that these results are not just driven by the omitted variable side of things, we repeat this experiment with independent regressors (Table B.7). The only issue here is misspecification in p_B. As above, all the estimators largely agree with each other here, and the RMSE is still just as bad as using the correctly specified, but uncorrected SBI.

Table B.6: Monte Carlo results when we drop variables that induce in-sample separation and the omitted variable is correlated with other covariates

Estimator	Quantity	α_0	α_1	α_2	β_0	β_1	β_2	RMSE
Ordinary SBI	Est	−2.60	−0.47	0.00	0.78	1.02	0.00	5.87
	St.Dev.	0.16	0.18	0.00	0.09	0.11	0.00	
	St.Err.	0.16	0.22		0.09	0.11		
	Power	1.00	0.56		1.00	1.00		
	Coverage	0.01	0.06		0.00	0.00		
BR-SBI	Est	−2.57	−0.45	0.00	0.78	1.01	0.00	5.86
	St.Dev.	0.15	0.18	0.00	0.09	0.11	0.00	
	St.Err.	0.16	0.22		0.09	0.11		
	Power	1.00	0.55		1.00	1.00		
	Coverage	0.01	0.06		0.00	0.00		
BR-FIML (Firth)	Est	−2.60	−0.47	0.00	0.78	1.02	0.00	5.87
	St.Dev.	0.16	0.18	0.00	0.09	0.11	0.00	
	St.Err.	0.16	0.23		0.09	0.11		
	Power	1.00	0.53		1.00	1.00		
	Coverage	0.01	0.07		0.00	0.00		
BR-FIML (Firth)	Est	−2.60	−0.47	0.00	0.78	1.02	0.00	5.87
	St.Dev.	0.16	0.18	0.00	0.09	0.11	0.00	
	St.Err.	0.16	0.23		0.09	0.11		
	Power	1.00	0.53		1.00	1.00		
	Coverage	0.01	0.07		0.00	0.00		
BR-FIML (Cauchy)	Est	−2.60	−0.47	0.00	0.78	1.02	0.00	5.87
	St.Dev.	0.16	0.18	0.00	0.09	0.11	0.00	
	St.Err.	0.16	0.23		0.09	0.11		
	Power	1.00	0.53		1.00	1.00		
	Coverage	0.01	0.07		0.00	0.00		
BR-FIML (log-F)	Est	−2.60	−0.47	0.00	0.78	1.02	0.00	5.87
	St.Dev.	0.16	0.18	0.00	0.09	0.11	0.00	
	St.Err.	0.16	0.23		0.09	0.11		
	Power	1.00	0.53		1.00	1.00		
	Coverage	0.01	0.07		0.00	0.00		
Truth	Parameters	−2.00	0.25	2.50	−0.50	0.25	5.00	
	St. Err. (SBI)	0.19	0.51	0.99	0.14	0.15	1.08	
	St. Err. (FIML)	0.19	0.52	1.00	0.14	0.15	1.12	

Table B.7: Monte Carlo results when we drop variables that induce in-sample separation and the omitted variable is independent of the other covariates

Estimator	Quantity	α_0	α_1	α_2	β_0	β_1	β_2	RMSE
Ordinary SBI	Est	−2.51	0.12	0.00	0.72	0.13	0.00	5.76
	St.Dev.	0.16	0.22	0.00	0.09	0.09	0.00	
	St.Err.	0.15	0.21		0.09	0.09		
	Power	1.00	0.08		1.00	0.34		
	Coverage	0.04	0.90		0.00	0.70		
BR-SBI	Est	−2.48	0.12	0.00	0.72	0.13	0.00	5.75
	St.Dev.	0.15	0.21	0.00	0.08	0.09	0.00	
	St.Err.	0.15	0.21		0.09	0.09		
	Power	1.00	0.08		1.00	0.34		
	Coverage	0.05	0.90		0.00	0.70		
Ordinary FIML	Est	−2.51	0.12	0.00	0.72	0.13	0.00	5.76
	St.Dev.	0.16	0.22	0.00	0.09	0.09	0.00	
	St.Err.	0.15	0.21		0.09	0.09		
	Power	1.00	0.08		1.00	0.34		
	Coverage	0.04	0.90		0.00	0.70		
BR-FIML (Firth)	Est	−2.51	0.12	0.00	0.72	0.13	0.00	5.76
	St.Dev.	0.16	0.22	0.00	0.09	0.09	0.00	
	St.Err.	0.15	0.21		0.09	0.09		
	Power	1.00	0.08		1.00	0.34		
	Coverage	0.04	0.90		0.00	0.70		
BR-FIML (Cauchy)	Est	−2.51	0.12	0.00	0.72	0.13	0.00	5.76
	St.Dev.	0.16	0.22	0.00	0.09	0.09	0.00	
	St.Err.	0.15	0.21		0.09	0.09		
	Power	1.00	0.08		1.00	0.34		
	Coverage	0.04	0.90		0.00	0.70		
BR-FIML (log-F)	Est	−2.51	0.12	0.00	0.72	0.13	0.00	5.76
	St.Dev.	0.16	0.22	0.00	0.09	0.09	0.00	
	St.Err.	0.15	0.21		0.09	0.09		
	Power	1.00	0.08		1.00	0.34		
	Coverage	0.04	0.90		0.00	0.70		
Truth	Parameters	−2.00	0.25	2.50	−0.50	0.25	5.00	
	St. Err. (SBI)	0.16	0.40	0.94	0.12	0.12	0.92	
	St. Err. (FIML)	0.16	0.41	0.95	0.12	0.12	0.96	

For the other estimators, this solution is still worse than the original separation problem.

B.5 Monte Carlo with real and many covariates

Finally, we consider a more sophisticated set of experiments. Here we use independent variables from Signorino and Tarar (2006) to conduct a Monte Carlo experiment based on real-world data. In this simulation we use the BR-FIML estimates from Table 4 to generate new outcome data using the independent variables from Signorino and Tarar (2006). For each Monte Carlo iteration we generate new values of the dependent variable and refit the model using five of the above estimators. As in the main text, we ignore the BR-FIML with Jeffreys prior, as this particular example has such a very poorly behaved objective function that the logged Jeffreys prior penalty is frequently undefined. We repeat this simulation 500 times.

Note that this experiment provides two important extensions over the previous simulations. First, as mentioned, it uses real-world data with many covariates, which gives the most realistic setup of any experiment we consider. Second, it uses a data generating process that is based on private information over outcomes rather than actions. This setup shows us how the solutions work with this different information structure.

The results of this simulation are reported in Table B.8, where each cell indicates the RMSE of the proposed estimator relative to the RMSE of the BR-FIML (log-F). Cases where this ratio is less than 1 are bolded and indicate that the alternative estimator does a better job than the BR-FIML (log-F) at estimating this quantity. Note that this occurs only in one of the cases considered. The final row presents the multivariate RMSE of each estimator relative to the BR-FIML. As expected, the BR-SBI and BR-FIML (Cauchy) perform reasonably well and much better than the ordinary estimators, but overall the BR-FIML (log-F) is the best in this experiment.

As a final simulation, we want to show the possibilities for separation-induced type-1 and type-2 errors with this more realistic data. To do this, we re-evaluate the previous simulation but we fix $\beta_{\text{Military Alliance}}$ to 0. We focus then on the null hypotheses that $\beta_{\text{Military Alliance}} = 0$

Table B.8: Relative RMSE of Estimates Compared to BR-FIML (log-F)

	Ordinary SBI	BR-SBI	Ordinary FIML	BR-FIML (Cauchy)
U_A(SQ): Const.	36.43	2.30	123.24	2.13
U_A(SQ): Tit-for-Tat	106.12	1.43	211.62	1.30
U_A(SQ): Firm-Flex	73.48	1.20	124.95	1.38
U_A(SQ): Democratic Attacker	37.95	1.36	111.11	1.03
U_A(SQ): Year	73.50	1.11	170.48	1.30
U_A(BD): Const.	47.18	2.72	138.75	3.64
U_A(War): Nuclear	44.94	1.60	101.04	1.19
U_A(War): Immediate Balance	10.95	1.53	149.58	1.32
U_A(War): Short-term Balance	71.78	2.19	157.53	1.31
U_A(War): Long-term Balance	25.04	1.10	74.50	1.16
U_A(War): Military Alliance	54.22	1.39	157.13	1.43
U_A(War): Arms Transfers	25.13	**0.70**	66.86	1.19
U_B(War): Const.	166.72	1.70	156.45	1.08
U_B(War): Nuclear	75.57	1.69	164.59	1.25
U_B(War): Immediate Balance	124.35	2.17	125.59	1.21
U_B(War): Short-term Balance	119.90	1.82	142.71	1.20
U_B(War): Military Alliance	55.71	1.40	79.26	1.13
U_B(War): Arms Transfers	49.30	1.52	107.78	1.19
U_B(War): Foreign Trade	153.26	1.23	159.47	1.27
U_B(War): Stalemate	81.14	1.65	113.17	1.07
U_B(War): Democratic Defender	83.28	1.54	111.75	1.11
Multivariate RMSE	77.33	1.83	135.13	1.78

Bold values indicate estimates where the estimator outperforms the BR-FIML (log-F) estimator

and $\beta_{\text{Nuclear}} = 0$, the former is true in this simulation and rejecting it is an example of a type-1 error, the latter is false and rejecting it reflects the power of the study.[2] We choose these two parameters as we believe that military alliance is at least one variable that induces the separation problem in this sample (see Figure 3), while the BR-FIML estimate for the effect of nuclear weapons is the only BR-FIML estimate in $B's$ utility function that has statistical significance in the Table 4.

Table B.9 presents the results of this simulation. Here we see that the ordinary SBI commits no type-1 errors, but also has no power against the null that nuclear weapons have no effect. The BR-SBI has the second-lowest type-1 error rate, but also the second-highest type-2 error rate. The ordinary FIML has a type-1 error rate that is nearly twice what we

[2]If an estimator returns a missing value for the standard error we say that it failed the test (committed a type-1 or type-2 error depending on the situation).

want given a $p < 0.05$ rejection rule, but it has the most power. The two BR-FIMLs have very good type-1 error rates that are smaller than expected at $p < 0.05$, which is consistent with their conservative standard errors. Likewise, both improve these type-1 error rates without sacrificing much power over the ordinary FIML in identifying an effect of nuclear weapons.

There are three major takeaways here. First, we again show that separation can lead to problems with power, particularly in the SBI context. Second, this simulation indicates that ordinary-FIML can be prone to type-1 errors when separation is present. Third, the BR-FIMLs have a low type-1 error rate and reasonable power, making it a good choice when separation is present in the sample.

Table B.9: Power and type-1 errors using Signorino and Tarar data

	SBI	BR-SBI	FIML	BR-FIML (Cauchy)	BR-FIML (log-F)
U_B(War): Nuclear (Power)	0.00	0.10	0.25	0.22	0.25
U_B(War): Military Alliance (Type-1 error)	0.00	0.01	0.08	0.02	0.03

C Signorino and Tarar specification and variable description

Signorino and Tarar (2006) use data on 58 interstate crises between 1885 and 1983 that was originally collected by Huth (1988). In this application, A's status quo utility is given by a constant term along with a time trend and indicators for whether: (i) past military buildups and preparations between A and B have been proportional or "tit-for-tat" rather than one-sided, (ii) past negotiations and relations between A and B were "firm-but-flexible" as opposed to hostile or conciliatory, and (iii) A is a democracy (polity score of at least 5).

When B stands firm and fights, A's utility is specified without a constant and includes: whether B has nuclear weapons (binary), the immediate balance of power (ratio of B's ground troops to A's), the short-term balance of power (ratio of B's ground forces, air forces, and first-level reserves to A's), the long-term balance of power (a more complicated ratio of B's ability to build and mobilize to A's),[3] whether there is a military alliance between B and the protégé state (binary), and the percentage of the protégé's arms imports that come from B. Finally, A's utility when B backs down is given by only a constant term.

Turning to B, the utility for war is based on all the factors that affect A's war payoff plus an indicator for whether the last crisis between A and B ended in a stalemate (i.e., they avoided armed conflict but failed to resolve any of the underlying issues between them), an indicator for whether B is a democracy, and a measure of the trade dependence between the protégé and B.

The authors use standard normal, outcome-specific shocks, so the choice probabilities are

$$\Pr(y_B = 1) = p_B = \Phi\left(\frac{U_B(SF)}{\sqrt{2}}\right)$$
$$\Pr(y_A = 1) = p_A = \Phi\left(\frac{(1-p_B)U_A(BD) + p_B U_A(SF) - U_A(SQ)}{\sqrt{(1-p_B)^2 + p_B^2 + 1}}\right).$$

For a more detailed discussion on how these choice probabilities are derived see Signorino 2003. Note that, as in the Monte Carlo simulations, we adjust the SBI results to account for the information structure and make them comparable to the FIML estimates. Player B's regressors are all multiplied by $1/\sqrt{2}$, while player A's regressors are multiplied by $1/\sqrt{1 + (1-\hat{p}_B)^2 + \hat{p}_B^2}$

D Replication study: Currency crises

Leblang 2003 considers an interaction between bondholders and states. Player A is a credit holder who has the choice of initiating a speculative attack against country B; B then defends or devalues its currency. The data record 90 developing democracies (polity score ≥ 5) from 1985-1998 and are organized into country-months. The three possible outcomes (status quo, attack-devalue, attack-defend) form the dependent variable. The status quo

[3] See Huth (1988, 61-2) for more details on long-term balance.

outcome is recorded in any month when no speculative attack occurs. An attack-devalue outcome is recorded when a speculative attack occurs in month t and the country devalues its currency in $t+1$. Finally, an attack-defend outcome is recorded when a currency attack occurs in period t and the state still has its currency pegged to a fixed exchange rate at $t+1$. See Leblang (2003, 544-5) for more details on how he determines whether a speculative attack has occurred.

Table D.1: Checking for separation in Leblang (2003)

Regressors	Outcome	Result
X_B	$y_B \mid y_A = 1$	No
Z^{SBI}	y_A	No
$[Z^{FIML} \;\; X_B]$	$\mathbb{I}(y_A = 0)$	No
$[Z^{FIML} \;\; X_B]$	$\mathbb{I}(y_A = 1)\mathbb{I}(y_B = 0)$	No
$[Z^{FIML} \;\; X_B]$	$\mathbb{I}(y_A = 1)\mathbb{I}(y_B = 1)$	No

Note: The Z variables are transformed using estimates of p_B from the unpenalized estimators.

The independent variables that parameterize the A's value of the status quo include an indicator for whether state B has capital controls in place, the logged value of B's foreign exchange reserves, an indicator for whether B's pegged exchange rate overvalues B's currency, the rate of domestic credit growth in B, the current interest rate in the United States, the ratio of debt service to GDP in B, a measure of contagion (nearby currency crises), and the number of previous attacks against B. Additionally, A's value for outcomes attack-defend and attack-devalue are specified as constants.

The specification of B's utility for defending its pegged exchange rate, includes B's foreign exchange reserves, the logged value of exports as a percentage of GDP, the real interest rate within B, and indicators for capital controls, whether an election is coming in the next three months, whether an election happened in the last three months, whether a right-wing party leads the government, and whether the party in power has a majority in the legislature. All non-indicator variables are standardized.

We fit the model with four of the above estimators. The results are presented in Table D.2. Note that there are slight differences between the results presented here and those printed in Leblang (2003). However they are relatively minor and likely result from slight differences in software. Additionally, we checked the results against the BR-FIML estimators with the density penalties and found no notable differences.

What we can see from the table is that the corrections do not change the results in any noticeable way. Additionally, there are no appreciable differences between the SBI and FIML estimates. These results are consistent with our belief that these data do not have a separation problem. Furthermore, this example reassures us that the correction does not affect point estimates when the optimization problem is already numerically stable.

References

Huth, P. K. 1988. *Extended Deterrence and the Prevention of War*. New Haven: Yale University Press.

Konis, K. 2007. "Linear Programming Algorithms for Detecting Separated Data in Binary Logistic Regression Models." PhD thesis. University of Oxford.

Leblang, D. 2003. "To Devalue or to Defend? the Political Economy of Exchange Rate Policy." *International Studies Quarterly* 47(4): 533–560.

Signorino, C. S. 2003. "Stucture and Uncertainty in Discrete Choice Models." *Political Analysis* 11(4): 316–344.

Signorino, C. S. and A. Tarar. 2006. "A Unified Theory and Test of Extended Immediate Deterrence." *American Journal of Political Science* 50(3): 586–605.

Table D.2: Leblang Replication

	FMLE	SBI	BR-FMLE	BR-SBI
U_A(SQ): Capital Controls	−0.45	−0.36	−0.39	−0.35
	(0.25)	(0.25)	(0.24)	(0.24)
U_A(SQ): Log(Reserves)	0.23	0.25	0.23	0.24
	(0.06)	(0.07)	(0.06)	(0.07)
U_A(SQ): Overvalued	−0.44	−0.42	−0.44	−0.39
	(0.09)	(0.22)	(0.09)	(0.15)
U_A(SQ): Credit Growth	−0.06	−0.07	−0.06	−0.07
	(0.03)	(0.03)	(0.03)	(0.03)
U_A(SQ): U.S. Interest	−0.05	−0.05	−0.05	−0.05
	(0.06)	(0.06)	(0.06)	(0.06)
U_A(SQ): Service	−0.03	−0.03	−0.03	−0.03
	(0.05)	(0.05)	(0.05)	(0.05)
U_A(SQ): Contagion	−0.12	−0.12	−0.11	−0.12
	(0.05)	(0.05)	(0.05)	(0.05)
U_A(SQ): Prior Attack	−0.12	−0.12	−0.12	−0.12
	(0.05)	(0.05)	(0.05)	(0.05)
U_A(Devalue): Devaluation	−3.66	−3.59	−3.62	−3.55
	(0.30)	(0.26)	(0.29)	(0.28)
U_A(Defend): Defense	−3.14	−3.08	−3.05	−2.94
	(0.29)	(0.31)	(0.30)	(0.33)
U_B(Defend): Const.	0.43	−0.20	0.29	−0.16
	(0.78)	(0.78)	(0.77)	(0.77)
U_B(Defend): Unified Gov't	−0.36	−0.06	−0.35	−0.04
	(0.36)	(0.45)	(0.35)	(0.44)
U_B(Defend): Log(Exports)	−0.20	−0.29	−0.18	−0.26
	(0.17)	(0.23)	(0.17)	(0.22)
U_B(Defend): Pre-election	1.66	2.23	1.42	1.90
	(0.75)	(0.93)	(0.69)	(0.86)
U_B(Defend): Post-election	1.06	1.10	0.93	0.96
	(0.59)	(0.72)	(0.56)	(0.70)
U_B(Defend): Right Gov't	−0.94	−1.54	−0.82	−1.34
	(0.45)	(0.66)	(0.43)	(0.62)
U_B(Defend): Real Interest	1.80	1.24	1.63	0.89
	(0.60)	(0.64)	(0.59)	(0.52)
U_B(Defend): Capital Control	0.07	0.66	0.20	0.57
	(0.76)	(0.80)	(0.74)	(0.79)
U_B(Defend): Log(Reserves)	0.31	0.59	0.29	0.51
	(0.17)	(0.20)	(0.16)	(0.19)
Observations	7240	7240	7240	7240

Notes: Standard errors in parenthesis