Finding and Accounting for Separation Bias in Strategic Choice Models

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

February 1, 2022

Abstract

Separation bias or "perfect prediction" is a common problem in discrete choice models that leads to inflated estimates. Standard statistical packages do not provide clear advice on how to deal with these problems. Furthermore, separation can go completely undiagnosed in advanced models that optimize a user-supplied log-likelihood rather than relying on pre-programmed estimation procedures. In this paper, we both describe the problems that separation bias causes and address the issue of detecting it in empirical models of strategic interaction by introducing and adapting a linear programming diagnostic from Konis (2007). We then consider several solutions based on penalized maximum likelihood estimation. Using both Monte Carlo experiments and replication studies we demonstrate that when separation is detected, the penalized methods we consider are superior to ordinary maximum likelihood estimators.

WORD COUNT: 6,857

[*]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

1 Introduction

Separation bias is a common problem in dealing with a categorical dependent variable wherein a linear combination of one or more exploratory variables perfectly predicts values of the outcome variable.[1] When present, separation will bias estimates, sometime heavily, toward positive or negative infinity. In the case of binary outcome models, solutions to the separation problem for logit and probit models have been proposed and examined by Cook, Hays and Franzese (2020), Gelman et al. (2008), Rainey (2016), Zorn (2005), and others. This line of inquiry has been invaluable for applied researchers. However, these binary cases are only one type of categorical choice model in political science, and separation problem can also plague more advanced or complicated models.[2]

Specifically, no one has approached the separation problem within the context of more sophisticated discrete choice models that account for strategic interaction (e.g., Signorino 1999). By considering this issue, we make three specific contributions. First, using Monte Carlo simulations and two replication studies we demonstrate that solutions based on penalized likelihood (PL) estimation apply to this class of theoretically interesting models by deriving bias-reduced (BR) strategic estimators. Second, we introduce political scientists to tools for diagnosing separation from Konis (2007) and discuss how they apply to strategic models. Third, we provide software for researchers to easily fit the BR strategic estimators that builds on the existing `games` package for `R`.

Throughout, we focus on separation problems in the two-player, extensive-form deterrence game, which is the standard workhorse model for political scientists interested in the empirical implications of theoretical models (EITM). This model and extensions to it

[1]Sometimes a distinction is drawn between quasi-complete separation and complete separation, we differentiate when relevant but frequently use the term separation to refer to ether situation. Practically, neither the problem and solution depend on this distinction.

[2]For example, Cook, Niehaus and Zuhlke (2018) discuss the separation problem within the context of the multinomial logit model.

have been used to study many central questions across political science subfields, including, but not limited to, interstate conflict (Nieman 2016; Signorino and Tarar 2006), currency crises (Leblang 2003), candidate entry into U.S. Senate races (Carson 2005), and compliance within the EU (König and Mäder 2014). Deriving and fitting theoretically-informed empirical models offers researchers flexibility in analyzing and testing formal models. In many cases, scholars derive their own empirical model from a formal theory and then feed a self-coded, log-likelihood function to a numeric optimizer (e.g., `optim` in R). This approach is extremely useful for fitting advanced or complicated models to data. However, separation bias becomes more difficult to diagnose in these setting, as optimizers will issue successful convergence codes without raising any warnings about the underlying numerical instability caused by separation.

2 Separation Bias: Problems

Separation bias occurs in multinomial choice models when a linear combination of one or more independent variables perfectly predicts values of the dependent variable. It is perhaps easiest to see the issues in a single binary outcome model; for exposition, consider data of the form $\{y_d, x_d\}_{d=1}^{D}$ where y_d is a binary outcome and x_d is a single independent variable. Within this context, separation means that there exists some value x_0 such that for all $x_d < x_0$, (almost) all $y_d = c$, and for all $x_d > x_0$, (almost) all $y_d = 1 - c$, where $c \in \{0, 1\}$. When these perfect predictors are included in a logit or probit regression the log-likelihood function can become monotonic, and estimates of the coefficients associated with x are biased, sometimes heavily, toward positive or negative infinity (Albert and Anderson 1984; Zorn 2005).[3]

In Figure 1, we illustrate this problem by considering a single variable x_d, $d = 1, 2, 3, 4$ between –2 and 2 to illustrate the problem. Let $\beta_x = 0$ describe the true relationship between x and y. By considering a few different but equally plausible, samples we can see

[3]We assume that true coefficients are always finite, but briefly discuss the alternative below.

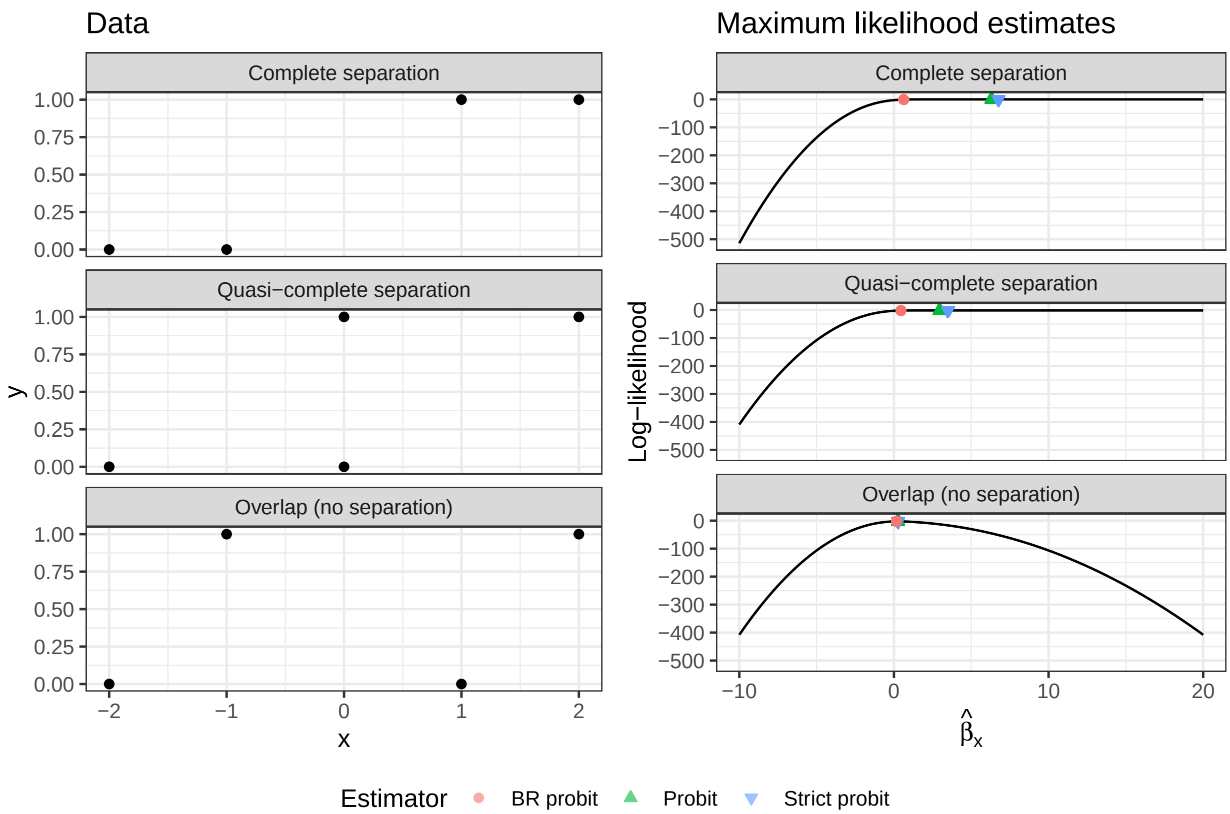

Figure 1: The effect of separation in binary choice

the effect that separation has on the log-likelihood and estimates. In the right column, we plot the probit log-likelihood against estimates of β_x. Here, the pointed-up triangle denotes the probit estimate under the default convergence criterion (1×10^{-8}), the pointed-down triangle represents probit with a stricter criterion (1×10^{-64}), and the circle represents the BR-probit estimate based on work by Firth (1993), which we describe in more detail below.[4]

In the bottom row, we have the ideal case for the probit estimator: overlap between x and y. Without an obvious cutpoint in x that allows us to always correctly classify y, we observe the typical upside-down-U shaped log-likelihood, and all three estimators find a maximum likelihood (ML) estimate near the true value of 0. In the top two rows, we see complete and quasi-complete separation: when $x > 0$, y is always 1. In both cases, the data produce a monotonically increasing log-likelihood and the ML estimate is ∞, even though the true value is still 0. The probit estimates are pulled towards infinity and the stricter convergence tolerance makes this inflation worse. In contrast, the BR probit estimate is still close to the true value of 0. The extent of the separation bias, while present in both cases, is notably more severe in when the separation is complete.

This example illustrates three points. First, when separation exists, point estimates from standard estimators are pushed away from zero, and the problem worsens when the separation is complete. Second, efforts to help a probit estimator by imposing a stricter convergence criteria can make this bias worse as they allow the model to move closer the actual ML estimate of $\pm\infty$. Third, penalization methods can be used to induce numerical stability on the optimization problem and lead to better estimates that are closer to the true parameter value.

Before expanding our discussion to strategic statistical models, there is one additional point we should make: there are two main reasons why separation appears. The first is if we include covariates that perfectly predict an outcome by definition. For example, including a covariate that records whether someone is infertile in a pregnancy study will perfectly predict

[4]The BR-probits are fit using the `brglm::brglm` function in R.

failed outcomes. In these cases, the true parameter may well be positive or negative infinity. We do not directly consider these situations as they seem unlikely to arise in most political science applications, and to the extent that they do, we encourage analysts to restrict their samples and questions to only include variables and observations where they expect the true parameters of interest to be finite. This leads us to the second situation where separation occurs: the true parameter of interest is finite, but separation appears as an artifact of the specific sample we observe. In this case, separation can be thought of as a finite-sample problem, where if enough additional data can be collected, the problem disappears. However, in many situations additional data may not exist or is costly to collect. In these situations we want to know if separation problems can be diagnosed and solved in strategic choice estimators.

2.1 Solutions

With logits and probits, the primary existing solutions to the separation problem are PL estimators (Rainey 2016; Zorn 2005). Penalization requires the analyst to impart some extra-empirical information (i.e., information from outside the data) to induce numerical stability into the optimization routine. Specifically, we want to choose information that reflects our belief that the coefficient estimates should not be too large. From a Bayesian perspective, penalization is equivalent to a prior belief that the true parameters are finite. In most cases, this information takes the form a Jeffreys prior penalty term that is maximized when the parameters are all zero, although others have found that a penalty term based on the Cauchy distribution works well (Gelman et al. 2008). Both penalties pull the estimates away from $\pm\infty$ and towards 0. Because the Jeffreys prior penalty is much more commonly used in political science (e.g., Cook, Hays and Franzese 2020; Zorn 2005) we focus on that in the exposition here. We consider Cauchy penalty in the online appendix and return to it when we replicate Signorino and Tarar (2006).

Consider the extensive-form deterrence game in Figure 2. There are two actors A and B, each of whom has two actions $y_i \in \{0, 1\}$ for $i \in \{A, B\}$. At the start of the game, each

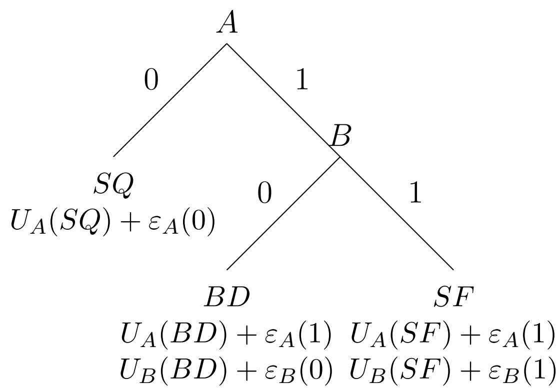

Figure 2: Standard two-player deterrence game

player receives private information in the form of an action-specific shock $\varepsilon_i(y_i)$.[5] Each shock reflects private information that i has regarding her payoff for taking action y_i.

After receiving her information, A selects an action. If A chooses $y_A = 0$, the game ends at the status quo (SQ). However, if A challenges B by taking action $y_A = 1$, then B responds. Player B can either back down to A's challenge by taking action $y_B = 0$ (ending the game at BD), or B can stand firm against A by taking action $y_B = 1$ (ending the game at SF). When the game ends at outcome $o \in \{SQ, BD, SF\}$, players receive a playoff equal to $U_i(o) + \varepsilon_i(y_i)$. This payoff contains a deterministic component: $U_i(o)$ representing a commonly known and observable payoff to player i and a stochastic component: $\varepsilon_i(y_i)$, which is the privately known cost/benefit to player i for taking action y_i.

The solution concept for this game is quantal response equilibrium (QRE) and a unique QRE exists for any set of parameter values (Signorino 1999). At the QRE, B chooses 1 if $U_B(SF) + \varepsilon_B(1) > U_B(BD) + \varepsilon_B(0)$, which can be described as

$$y_B = \mathbb{I}\left(U_B(SF) - U_B(BD) + \varepsilon_B(1) - \varepsilon_B(0) > 0\right),$$

[5]Other versions of this model use an outcome-specific shock rather than an action-specific shock. This results in slightly different choice probabilities, which we return to in the applications, but otherwise everything we derive here carries applies to both setups.

where $\mathbb{I}(\cdot)$ is the indicator function. Likewise, A chooses 1 if

$$U_A(SQ) + \varepsilon_A(0) < (1 - \Pr(y_B = 1))U_A(BD) + \Pr(y_B = 1)U_A(SF) + \varepsilon_A(1),$$

which can be described as

$$y_A = \mathbb{I}\left((1 - \Pr(y_B = 1))U_A(BD) + \Pr(y_B = 1)U_A(SF) - U_A(SQ) + \varepsilon_A(1) - \varepsilon_A(0) > 0\right).$$

To transform this game into an empirical model we need to specify two things. First, the deterministic portion of the utilities need to be rewritten in terms of observed data. Second, we need to assume a distribution for the action-specific shocks. For the purpose of exposition consider the problem of estimating a model with following utilities and choice probabilities.

$$\begin{aligned}
U_B(SF) &= X_B\beta \\
U_B(BD) &= 0 \\
U_A(SQ) &= X_{SQ}\,\alpha_{SQ} \\
U_A(BD) &= X_{BD}\,\alpha_{BD} \\
U_A(SF) &= X_{SF}\,\alpha_{SF} \\
\Pr(y_B = 1) = p_B &= F_B\left(U_B(SF)\right) \\
\Pr(y_A = 1) = p_A &= F_A\left((1 - p_B)U_A(BD) + p_B U_A(SF) - U_A(SQ)\right),
\end{aligned}$$

where F_i is the distribution that describes $\varepsilon_i(1) - \varepsilon_i(0)$. Our goal is to estimate the parameters $\theta = (\alpha, \beta)$ using D observations of actors playing this game.

Standard practice is to estimate the parameters in one of two ways. The first approach is to follow Signorino (1999) and derive a full information maximum likelihood (FIML) estimator. The other, from Bas, Signorino and Walker (2008), uses a two-step procedure known as statistical backwards induction (SBI).

2.1.1 Statistical Backwards Induction

SBI procedure for this model is as follows:

1. Using only observations where $y_A = 1$, regress y_B on X_B using a logit or probit (depending on F_B) to produce $\hat{\beta}^{SBI}$. Estimate $\hat{p}_B^{SBI} = F_B(X_B\hat{\beta}^{SBI})$.

2. Regress y_A on $Z^{SBI} = \begin{bmatrix} X_{BD}(1-\hat{p}_B^{SBI}) & X_{SF}(\hat{p}_B^{SBI}) & -X_{SQ} \end{bmatrix}$ using a logit or probit (depending on F_A) to produce $\hat{\alpha}^{SBI}$.

Note that because each step in this process is a binary outcome model, $\hat{\theta}$ solves

$$\begin{aligned}
\hat{\beta}^{SBI} &= \underset{\beta}{\operatorname{argmax}} \sum_{d:y_{A,d}=1} \mathbb{I}(y_{B,d}=1)\log(F_B(x'_{B,d}\beta)) + \mathbb{I}(y_{B,d}=0)\log(1-F_B(x'_{B,d}\beta)) \qquad (1)\\
\hat{\alpha}^{SBI} &= \underset{\alpha}{\operatorname{argmax}} \sum_{d=1}^{D} \mathbb{I}(y_{A,d}=1)\log(F_A(z'_d\alpha)) + \mathbb{I}(y_{A,d}=0)\log(1-F_A(z'_d\alpha)),
\end{aligned}$$

where $d = 1, \ldots, D$ indexes each observed play of this game.

Because this approach relies on two distinct binary outcome models, standard solutions to separation bias based on PL estimation with Jeffreys prior can be applied. Let $L_B(\beta)$ and $L_A(\alpha)$ be the objective functions in Eq. 1, then the bias-reduced SBI (BR-SBI) are given as

$$\begin{aligned}
\hat{\beta}^{BR-SBI} &= \underset{\beta}{\operatorname{argmax}}\, L_B(\beta) + \frac{1}{2}\log(\det(I(\beta))) \qquad (2)\\
\hat{\alpha}^{BR-SBI} &= \underset{\alpha}{\operatorname{argmax}}\, L_A(\alpha) + \frac{1}{2}\log(\det(I(\alpha))),
\end{aligned}$$

where I is the estimated Fisher-information matrix. Following standard practice, I is found using the Hessian of the uncorrected model, such that

$$I(\beta) = -\frac{\partial^2 L_B(\beta)}{\partial\beta\partial\beta'},$$

and likewise for $I(\alpha)$. Note, Firth (1993) points out that standard errors for $\hat{\beta}^{BR-SBI}$ can be estimated by evaluating $I(\beta)^{-1}$ at the BR estimates. This means that standard errors for $\hat{\alpha}^{BR-SBI}$ can be estimated using ordinary two-step maximum likelihood results (e.g., Murphy and Topel 1985).

2.1.2 Full Information ML

The SBI estimator is easily implemented, but this ease comes at the cost of statistical efficiency and perhaps increased concerns about separation if the sample used to estimate β is small. A more efficient alternative comes in the form of maximizing a single log-likelihood function that re-computes the choice probabilities at every step in optimization process. When the underlying strategic model has a unique equilibrium, as is the case here, the FIML is consistent and asymptotically efficient (Rust 1987).

Using the above parameterization, the FIML estimates are given as

$$\begin{aligned}\hat{\theta}^{FIML} = \underset{\theta}{\text{argmax}} \sum_{d=1}^{D} & \mathbb{I}(y_{A,d} = 0) \log(1 - p_{A,d}) \\ & + \mathbb{I}(y_{A,d} = 1)\mathbb{I}(y_{B,d} = 0) \log(p_{A,d} \cdot (1 - p_{B,d})) \\ & + \mathbb{I}(y_{A,d} = 1)\mathbb{I}(y_{B,d} = 1) \log(p_{A,d} \cdot p_{B,d}).\end{aligned} \tag{3}$$

Let $L(\theta)$ be the objective function in Eq. 3, we can construct the same penalized solution using where the (full) information matrix is now

$$I(\theta) = -\frac{\partial^2 L(\theta)}{\partial\theta\partial\theta'}.$$

Our bias-reduced FIML estimates are

$$\begin{aligned}\hat{\theta}^{BR-FIML} = \underset{\beta}{\text{argmax}} \sum_{d=1}^{D} & \mathbb{I}(y_{A,d} = 0) \log(1 - p_{A,d}) \\ & + \mathbb{I}(y_{A,d} = 1)\mathbb{I}(y_{B,d} = 0) \log(p_{A,d} \cdot (1 - p_{B,d})) \\ & + \mathbb{I}(y_{A,d} = 1)\mathbb{I}(y_{B,d} = 1) \log(p_{A,d} \cdot p_{B,d}) + \frac{1}{2}\log(\det(I(\theta))).\end{aligned} \tag{4}$$

Computing the information matrix for this likelihood at each step in the numeric optimization process is a non-trivial computational task; as such we provide an extension to R's `games` package that fits the penalized BR model.

As mentioned, the Jeffreys prior penalty introduced by Firth is not the only possible solution. Another common approach is a Cauchy penalty, where instead of adding on the logged Jeffreys prior we add on the logged density function for the Cauchy distribution with location 0 and scale 2.5. Implementing this approach is straightforward, and we consider its performance in the online materials. Overall, both approaches work well and both result in improved estimates when separation is present. We are overall agnostic to which approach analysts use, except to note that while Jeffreys prior is more common, it requires that the Hessian be positive definite at every guess of the parameter values. This requirement will always be met in logit and probit models, but may fail with more complicated likelihood functions. When the logged Jeffreys prior does not exist, the Cauchy provides an easy-to-use alternative.

2.2 Diagnosis

Having considered a solution to the separation problem, we are left with the task of diagnosing it. Current best practices in political science are based on rules of thumb. Specifically, common advice is to simply look for point estimates and standard errors that are so large as to strain credibility and if separation is suspected, an analyst can look for monotonicities in the fitting criteria (i.e., a log-likelihood, deviance, or other loss function). While these approaches are important tools for analysts, more principled alternatives exist. Specifically, diagnostics based on linear programming (lp) have been proposed by Albert and Anderson (1984) and others. One easy-to-implement approach comes from Konis (2007); we introduce his lp-diagnostic to political scientists.

The lp setup considers 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 $d = 1, \ldots, D$, the unscaled distance between w_d and the hyperplane is given by $s(\gamma) = \sum_{d=1}^{D} w_d'\gamma$. If there is separation, then $w_d'\gamma \geq 0$ when $y_d = 1$ and $w_d'\gamma \leq 0$

when $y_d = 0$. Consider the following constrained optimization problem:

$$\begin{aligned} &\text{maximize } s(\gamma) \\ &\text{s.t. } w_d'\gamma \geq 0 \text{ if } y_d = 1 \\ &\qquad w_d'\gamma \leq 0 \text{ if } y_d = 0 \\ &\qquad -\infty < \gamma_j < \infty \text{ for } j = 1, \ldots, k. \end{aligned}$$

The above can rewritten into the simpler form of:

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

where $\mathbf{1}_D$ is a length-D column vector of 1s and $\bar{W} = [(2y_d - 1)w_{d,1}, \ldots, (2y_d - 1)w_{d,k}]_{d=1}^{D}$.

If there is 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. Here, 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. Diagnosing separation, then becomes as easy as checking the convergence of the above lp. The R function `detectseparation::detect_separation` conducts this diagnostic.

Directly generalizing this diagnostic to the full information strategic setting is infeasible, because the full design matrix contains the endogenous quantity p_B. As a result of this endogenous quantity, we cannot know *a priori* if separation exists between the covariates describing A's decision-making and the three outcomes of the strategic model. However, the diagnostic can be applied both before and after estimation. We recommend the following work flow when analysts are concerned about the possibility of separation:

1. Using `detect_separtion`, check for separation in X_B and y_B using the observations where $y_A = 1$.

2. If separation is not detected in (1) generate $\hat{p}_B^{SBI}$ and Z^{SBI}. Use `detect_separation` to check for separation in Z and y_A.

3. If separation is detected in either (1) or (2), consider a bias-reduced estimator. Otherwise, the uncorrected estimators should be fine.

4. An additional post-estimation check can be conducted by using the lp-diagnostic to search for separation in $[Z^{SBI}, X_B]$ or $[Z^{FIML}, X_B]$ against each of the three outcomes (SQ, BD, SF), individually. If any of these three checks finds separation consider a bias-reduced estimator.

3 Performance

We now consider Monte Carlo experiments to evaluate the effectiveness of the PL estimators for strategic models. Within each experiment we compare the BR-SBI and BR-FIML estimators given by Eq. 2 and Eq. 4, respectively, to their unpenalized counterparts. The experimental setup is presented in Figure 3, where we consider four parameters.[6] The β parameters and the variable X_B characterize B's payoffs, while the α parameters and X_A are used to form A's payoffs. Both X_A and X_B are independently distributed Bernoulli with a mean of 0.5, while values of α and β vary to induce separation. In these two experiments we use Jeffreys prior to form the BR estimators; in the online materials we consider the Cauchy penalty.

[6]A more realistic setup with many covariates is considered in the online materials. Overall, our main conclusions hold under the more complex setting.

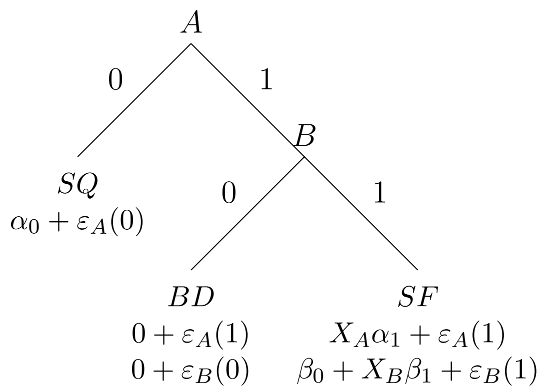

Figure 3: Monte Carlo version of the two-player deterrence game

3.1 Separation in B's choice

Our first set of Monte Carlos looks at the problem of separation in B's choice of 0 or 1. In this case B's choice is given by

$$\begin{aligned} y_B &= \mathbb{I}(-1 + 5X_B + \varepsilon_B(1) > 0 + \varepsilon_B(0)) \\ &= \mathbb{I}(-1 + 5X_B + \varepsilon_B(1) - \varepsilon_B(0) > 0). \end{aligned}$$

Each error term is independently and identically distributed standard normal, such that $p_B = \Phi\left(\frac{-1+5X_B}{\sqrt{2}}\right)$. Note that a large, but not unreasonable, coefficient on X_B will ensure that in most samples $y_B = 0$ only when $X_B = 0$.

The DGP for player A is

$$\begin{aligned} y_A &= \mathbb{I}(-2(X_A p_B) + \varepsilon_A(1) > 1.5 + \varepsilon_A(0)) \\ &= \mathbb{I}(-1.5 - 2(X_A p_B) + \varepsilon_A(1) - \varepsilon_A(0) > 0). \end{aligned}$$

In terms of Figure 3, the parameters of interest are $\alpha_0 = 1.5$, $\alpha_1 = -2$, $\beta_0 = -1$ and $\beta_1 = 5$. We repeat the Monte Carlo experiment 5,000 times with samples of size $D = 500$, and we only keep those results where the lp-diagnostic detects separation between x_B and ending the game at outcome BD.[7]

[7] As expected, in cases where the lp-diagnostic does not detect separation, the results from all four estimators are nearly identical.

The Monte Carlo results are reported in Table 1. The first thing to note is that the BR techniques makes a noticeable and positive impact on both the point estimates the standard errors. This translates into substantial decreases in the average root mean squared errors (RMSE) on the parameter estimates. For both the SBI and the FIML, the PL approach helps when separation is present. The improvements are slightly more pronounced with the BR-FIML, which has the smallest RMSE of all the estimators considered, while also having the least bias in estimating β_1.

Second, we see that both FIML estimators tend to outperform their SBI counterparts. One reason for this is that the FIML is a single-step estimator and will be more efficient by construction. However, it is also worth noting that the separation bias is worst in the unpenalized SBI and that while the FIML still exhibits bias, its RMSE is about 2/3 that of the SBI. This difference likely emerges because the first stage of the SBI has to deal with both separation and a smaller sample (only observations where $y_A = 1$). As mentioned above, smaller samples can exasperate separation problems, making this part of the SBI more vulnerable.

These differences among the unpenalized estimators lead to the another rule-of-thumb that analysts can use in addition to the lp-diagnostic. When separation is present, noticeable divergence can appear between the SBI and FIML estimates. The larger this divergence, the more likely a problem. Upon observing large differences, analysts should consider checking their data with the lp-diagnostic.

3.2 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

$$\begin{aligned} y_B &= \mathbb{I}(-0.5 + 5X_B + \varepsilon_B(1) > 0 + \varepsilon_B(0)) \\ &= \mathbb{I}(-0.5 + 5X_B + \varepsilon_B(1) - \varepsilon_B(0) > 0), \end{aligned}$$

Table 1: Monte Carlo results when separation is present in Player B's decision

Estimator	α_0	α_1	β_0	β_1	RMSE
Ordinary SBI	1.50	−2.34	−1.04	9.29	4.60
	(0.13)	(1.45)	(0.45)	(0.53)	
BR-SBI	1.50	−2.04	−1.00	3.83	1.55
	(0.13)	(0.85)	(0.38)	(0.40)	
Ordinary FIML	1.50	−2.43	−1.07	7.14	3.10
	(0.13)	(1.76)	(0.37)	(1.26)	
BR-FIML	1.50	−2.04	−0.98	3.95	1.34
	(0.13)	(0.68)	(0.33)	(0.33)	
True Values	1.5	−2	−1	5	

Note: Average coefficient estimates for each model with standard deviations in parentheses.

and $p_B = \Phi\left(\frac{-0.5+5X_B}{\sqrt{2}}\right)$. Likewise, player A's choice is given as

$$\begin{aligned} y_A &= \mathbb{I}(2.5(X_A p_B) + \varepsilon_A(1) > -3 + \varepsilon_A(0)) \\ &= \mathbb{I}(3 + 2.5(X_A p_B) + \varepsilon_A(1) - \varepsilon_A(0) > 0). \end{aligned}$$

In terms of Figure 3, we have $\alpha_0 = -3$, $\alpha_1 = 2.5$, $\beta_0 = -0.5$, $\beta_1 = 5$. Here, $y_A = 0$ is a rare event that is only likely when $X_A = 0$, while y_B is still only likely when $X_B = 0$. We keep results where the lp-diagnostic finds separation between either covariate and the outcomes SQ and BD.[8]

The Monte Carlo results are reported in Table 2. As before, there is striking improvement in both bias and RMSE when employing a BR estimator. Additionally, we continue to observe notable differences between the SBI and FIML; this divergence continues to serve as a warning that separation may be present.

3.3 Additional simulations

Several additional simulations are considered in the online appendix. The first two consider smaller sample sizes $D = 50$ in an effort to better match the Signorino and Tarar (2006) replication we consider below. With these smaller samples, separation becomes more likely

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

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

Estimator	α_0	α_1	β_0	β_1	RMSE
Ordinary SBI	-3.07	14.79	-0.50	8.87	13.37
	(0.35)	(3.53)	(0.12)	(0.13)	
BR-SBI	-3.07	1.43	-0.50	4.61	1.42
	(0.31)	(0.76)	(0.12)	(0.12)	
Ordinary FIML	-3.06	8.59	-0.50	8.31	7.31
	(0.31)	(2.22)	(0.12)	(0.57)	
BR-FIML	-3.07	1.55	-0.49	4.65	1.33
	(0.32)	(0.80)	(0.12)	(0.12)	
True Values	-3	2.5	-0.5	5	

Note: Average coefficient estimates for each model with standard deviations in parentheses.

to occur by chance and the bias in the unpenalized estimators is more pronounced. As in the above, we find that both BR estimators offer major improvements over their counterparts.

We also redo the large- and small-D simulations with the Cauchy penalty in place of the Jeffreys prior for the BR-FIML. We find that the choice of penalty makes little difference. The reason we consider the Cauchy at all is that FIML estimator evaluates a complicated likelihood, and we found situations where the evaluated Hessian is not negative definite. In these cases, the logged Jeffreys prior does not exist.[9] In these cases, the Cauchy approach provides an easy-to-use alternative that does not depend on either the data or likelihood, which means the penalty always exists. Overall, the similarity across penalties is important; we do not want the estimates to be overly sensitive to the penalty choice. Additionally, these simulations provide confidence in using the Cauchy penalty when the logged Jeffreys prior penalty fails to exist. We return to this point when we consider to the Signorino and Tarar (2006) replication.

The final simulation is based on the data from Signorino and Tarar (2006). Here, we want to capture the circumstances of real-world data to provide a more realistic Monte Carlo experiment where we have a smaller number of observations, many parameters, correlated

[9]This existence problem does not arise in the SBI as the binomial log-likelihood always has a strictly negative Hessian.

regressors, an alternative form of private information, and some numeric instability in the estimation. To conduct the experiment, we use the BR-FIML estimates from Table 5 as the "truth" and use the observed independent variables from Signorino and Tarar (2006) to simulate new values of the dependent variable. We then fit the model to the simulated dependent variable and the original regressors using the four estimators. Overall, this approach is similar to conducting a parametric bootstrap. Our main conclusions are unchanged in this more realistic exercise.

4 Applications

We now consider two replication studies. The first is from Leblang (2003) who uses a strategic model to examine currency crises. We find no evidence of separation in his data and the PL estimates closely match the original results. The second is from Signorino and Tarar (2006) who consider deterrence in interstate disputes. In this second example, we believe that separation bias is present in the original results. We start by applying the lp-diagnostic to both datasets. The diagnostic results are reported in Table 3.

Table 3: Checking for separation with the linear programming diagnostic

Regressors	Outcome	Leblang (2003)	Signorino and Tarar (2006)
X_B	$y_B \mid y_A = 1$	No	Yes
Z^{SBI}	y_A	No	No
$[Z^{FIML} \quad X_B]$	$\mathbb{I}(y_A = 0)$	No	Yes
$[Z^{FIML} \quad X_B]$	$\mathbb{I}(y_A = 1)\mathbb{I}(y_B = 0)$	No	Yes
$[Z^{FIML} \quad X_B]$	$\mathbb{I}(y_A = 1)\mathbb{I}(y_B = 1)$	No	Yes

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

4.1 Replication 1: 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 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.

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.

To specify B's utility for defending its pegged exchange rate, he 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 normalized.

Leblang (2003) uses standard normal, outcome-specific shocks rather than the action-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).

We fit the model with all four of the above estimators. The results are presented in

Table 4.[10] Note that there are slight differences between the results presented here and those printed Leblang (2003), however they are relatively minor and likely result from slight differences in software. Additionally, we checked the results against the BR-FIML with a Cauchy penalty 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.

4.2 Replication 2: Deterrence

We now turn our attention to Signorino and Tarar (2006) who study deterrence in interstate disputes using data on 58 crises between 1885 and 1983 from (Huth 1988). This game has two players: an aggressor and defender state. The aggressor (A) decides between attacking a protégé state of the defender (B) or preserving the status quo. If A chooses the latter, the game ends, but if A chooses the former, then the defender can either protect its protégé or back down. The dependent variable takes on three values: status quo, attack-war, attack-back down. Like Leblang (2003), they rationalize the data with outcome-specific private information.

In this example, 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). Additionally, A's utility when B backs down is given by only a constant term. When B stands firm and fights, A's utility

[10] 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 $1/\sqrt{1 + (1 - \hat{p}_B)^2 + \hat{p}_B^2}$.

Table 4: Leblang Replication

	FMLE	SBI	BR-FMLE	BR-SBI
U_A(SQ): Capital Controls	−0.45	−0.37	−0.41	−0.35
	(0.25)	(0.23)	(0.24)	(0.22)
U_A(SQ): Log(Reserves)	0.23	0.24	0.22	0.24
	(0.06)	(0.06)	(0.06)	(0.06)
U_A(SQ): Overvalued	−0.44	−0.41	−0.44	−0.39
	(0.09)	(0.16)	(0.08)	(0.13)
U_A(SQ): Credit Growth	−0.06	−0.06	−0.07	−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)	(0.05)	(0.05)
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	−0.12	−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	−3.66	−3.55
	(0.30)	(0.26)	(0.29)	(0.26)
U_A(Defend): Defense	−3.14	−3.01	−2.97	−2.94
	(0.29)	(0.30)	(0.32)	(0.30)
U_B(Defend): Const.	0.43	−0.20	0.42	−0.16
	(0.78)	(0.78)	(0.68)	(0.77)
U_B(Defend): Unified Gov't	−0.36	−0.06	−0.43	−0.04
	(0.36)	(0.45)	(0.32)	(0.44)
U_B(Defend): Log(Exports)	−0.20	−0.29	−0.17	−0.26
	(0.17)	(0.23)	(0.16)	(0.22)
U_B(Defend): Pre-election	1.66	2.23	1.22	1.90
	(0.75)	(0.93)	(0.62)	(0.86)
U_B(Defend): Post-election	1.06	1.10	0.91	0.96
	(0.59)	(0.72)	(0.54)	(0.70)
U_B(Defend): Right Gov't	−0.94	−1.54	−0.86	−1.34
	(0.45)	(0.66)	(0.42)	(0.62)
U_B(Defend): Real Interest	1.80	1.24	1.65	0.89
	(0.60)	(0.64)	(0.57)	(0.52)
U_B(Defend): Capital Control	0.07	0.66	0.05	0.57
	(0.76)	(0.80)	(0.66)	(0.79)
U_B(Defend): Log(Reserves)	0.31	0.59	0.21	0.51
	(0.17)	(0.20)	(0.15)	(0.19)
Observations	7240	7240	7240	7240

Notes: Standard errors in parenthesis

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 (Huth 1988, 61-2)), 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, B's 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.

There are several reasons to believe that separation bias is present in their reported estimates. First, as shown in Table 3, the lp-diagnostic detects separation in four of the five cases we consider in Table 3. Second, some of the coefficients they report are relatively large for discrete choice models (greater than 10). Third, when we attempted to fit model using the uncorrected SBI some estimates were over 100 and the two-step covariance matrix was not positive definite. All three of these warning signs suggest numerical issues. As a result, we expect that the BR models will produce different results.

Compounding the separation problem is the issue of fitting a complicated strategic model to a relatively small sample (58 aggressor-defender dyads). In replicating these results, we found that the determinant of the FIML information matrix is negative at many steps in the optimization process, making the logged Jeffreys prior penalty term undefined. As a result, we use the Cauchy penalty in this case as it does not rely on the curvature of the baseline log-likelihood. The BR-SBI continues to use the Jeffreys prior penalty here as the probit objective function does not have the same complexity as the FIML and the Jeffreys prior penalty always exists.

The results from the ordinary and the bias-reduced FIML and SBI estimators can be found in Table 5. Note that given the trouble we had with the ordinary SBI, the estimates and standard errors are the means and standard deviations from a non-parametric bootstrap where we threw away results that were greater than 50 in magnitude in order to keep the results on roughly the same scale across the estimators. The fact that we even had to consider this approach with the SBI is a warning against using an uncorrected model. As with the Leblang (2003) replication, there are slight differences between the replicated FIML and published results, which we attribute to slight differences in software implementation.

What is most striking about the results in Table 5 is that while many of the point estimates have the same sign across all four estimators, some results that were significant in their analysis are no longer significant at traditional levels. Additionally, we note that the estimates and standard errors on the uncorrected SBI are incredibly large despite the precautions we took to make the results appear more reasonable. Couple this with the fact that many of the point estimates vary in magnitude across the uncorrected estimators, and we have a good reason to suspect that a bias reduced estimator may be more appropriate. Indeed, the two BR estimators largely agree with each other in terms of magnitude and sign in 18 out of 21 estimates, although the BR-SBI finds fewer significant results. This difference may result from the relative inefficiency of the two-step estimator.

In examining player B's (the defender's) utility function, Signorino and Tarar (2006) find that the defender is more likely to support its protégé if B has nuclear weapons, if the protégé imports a lot of arms from B, and if there was a past, but unresolved, crisis between the defender and the aggressor (2006, 593). Our analysis concurs with these results in terms of sign, but only the effect of nuclear weapons remains significant at the 5% level. The overall decrease in coefficient magnitude and is consistent with our suspicion that separation bias is present in their original analysis.

Likewise, in player A's (the aggressor's) utility for the war we also find that while all the signs on our estimates match the original study, they are universally smaller in magnitude. In the original study, the aggressor's value for war is lower when the immediate and short-term balance of power favors the defender, but higher when the long-term balance of power

Table 5: Signorino and Tarar Replication

	FMLE	SBI	BR-FMLE	BR-SBI
U_A(SQ): Const.	−5.04	−5.76	−1.19	−3.09
	(2.39)	(6.01)	(1.69)	(1.85)
U_A(SQ): Tit-for-Tat	17.27	2.15	2.44	1.54
	(7.22)	(1.48)	(1.03)	(0.63)
U_A(SQ): Firm-Flex	6.59	1.05	1.23	0.68
	(3.26)	(1.76)	(0.88)	(0.60)
U_A(SQ): Democratic Attacker	15.75	−0.39	0.46	−0.51
	(8.60)	(2.66)	(1.56)	(0.95)
U_A(SQ): Year	−0.35	−0.03	−0.03	−0.02
	(0.18)	(0.05)	(0.02)	(0.01)
U_A(BD): Const.	13.51	−7.44	2.04	−2.88
	(12.76)	(7.05)	(2.90)	(2.25)
U_A(War): Nuclear	−9.13	−0.19	−0.91	−1.50
	(5.00)	(8.93)	(1.49)	(2.10)
U_A(War): Immediate Balance	−12.51	−2.93	−1.38	−1.19
	(5.26)	(7.28)	(1.09)	(0.67)
U_A(War): Short-term Balance	−6.22	−3.99	−1.81	−3.41
	(3.26)	(6.93)	(1.64)	(1.78)
U_A(War): Long-term Balance	3.35	0.86	0.42	0.74
	(1.57)	(2.92)	(0.50)	(0.54)
U_A(War): Military Alliance	12.62	1.92	2.56	1.53
	(5.23)	(3.95)	(1.39)	(1.18)
U_A(War): Arms Transfers	−0.86	−0.22	−0.16	−0.06
	(0.49)	(0.49)	(0.16)	(0.14)
U_B(War): Const.	−10.93	−20.21	−2.70	−1.07
	(5.88)	(15.98)	(1.31)	(1.43)
U_B(War): Nuclear	6.64	−3.66	2.41	0.13
	(2.62)	(19.77)	(1.10)	(1.74)
U_B(War): Immediate Balance	5.46	16.47	1.21	0.66
	(2.90)	(14.56)	(0.73)	(0.88)
U_B(War): Short-term Balance	4.16	3.72	1.23	0.25
	(2.37)	(15.11)	(0.79)	(1.09)
U_B(War): Military Alliance	13.39	12.54	1.54	2.39
	(7.61)	(19.33)	(1.55)	(1.86)
U_B(War): Arms Transfers	−1.75	−1.47	−0.27	−0.38
	(0.86)	(2.08)	(0.22)	(0.24)
U_B(War): Foreign Trade	4.85	5.96	0.87	0.71
	(2.55)	(2.81)	(0.53)	(0.41)
U_B(War): Stalemate	8.40	16.81	1.32	1.79
	(4.21)	(14.76)	(1.10)	(1.18)
U_B(War): Democratic Defender	5.93	1.60	1.03	−0.02
	(2.86)	(11.21)	(0.86)	(1.05)
Observations	58	58	58	58

Notes: Standard errors in parenthesis (Model 6 is bootstrapped) favors the defender. Additionally, B's nuclear capability reduced A's war payoff. While the signs on the BR estimates match these effects, none are statistically significant. Overall, this suggests that there is just not enough in the data to distinguish these effects absent separation bias.

Switching our attention to A's status quo payoff, we confirm the finding that tit-for-tat military preparations between the attacker and the defender/protégé make the status quo more desirable. This deterrence result against A is the only significant effect that the BR estimators find in the status quo payoff. Unlike Signorino and Tarar (2006), we do not find an effect on adopting firm-but-flexible bargaining postures.

Overall, the results from the BR estimators produce the coefficients with similar signs, but a smaller magnitude than the results obtained by Signorino and Tarar (2006). This difference is expected when separation bias is present. While some of their statistically significant results are reproduced by the BR estimators, many are not. These findings suggest that some of the original results may be numerical artifacts resulting from the separation problem. Many of these findings may, of course, still be true, but in many cases we cannot reject their various null hypotheses with these data once we correct for separation bias.

To better demonstrate these numeric issues and to illustrate how the BR corrections perform we consider the profiled log-likelihood of the FIML, the BR-FIML, and the BR-SBI for the coefficient on military alliance in B's utility function. We focus on this variable as the uncorrected coefficient estimate of about 13 suggests to us that it may be one cause of the separation bias. The profiling procedure works by fixing the value of a single parameter and refitting the model. By doing this at many fixed values for the parameter, we assess how sensitive model fit is to changes in this parameter. For a well-behaved problem, we would expect a classic upside-down U shape with a maximum at the estimated parameter value. The profiled results are shown in Figure 4. We focus on military alliance, as it appears to be a potential source of the separation problem. Specifically, when we profile the FIML (the top pane) with respect to this coefficient, we find that while there appears to be a local maxima at the estimate, model fit can be improved by increasing this estimate past positive 20. Put another way, while the estimate is a local maximum, it is not the global maximum; "better" fit can be found at estimates further toward ∞. This push towards $\pm\infty$ is the classic sign of the separation problem. Looking at the two BR profiles we see that, at least in the range considered, the estimates are at only maximum. Note that the BR-SBI has a flattish section at the right-hand end of the plot, however, this drops off quickly if we explore past this region, and we find no reason to suspect that there are better log-likelihood values beyond the range presented here.

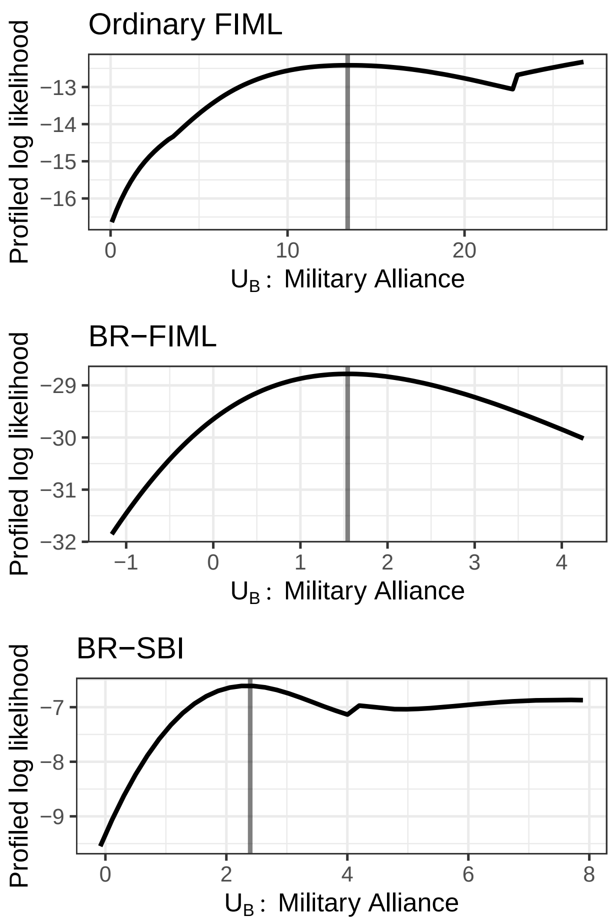

Figure 4: Profiled log-likelihood on the coefficient associated with how a military alliance affects B's decision to intervene.

These two applications illustrate two points. First, when separation is not detected in the data, the corrected estimators do not change the results. This result suggests that corrections are harmless in the sense that they can be used by the indiscriminate analyst without concern that it will lead to additional harm. Second, when separation is present, the corrections induce numerical stability into the log-likelihood, making it easier for the optimization software to find a solution.

5 Conclusions and Recommendations

Penalized likelihood methods provide an incredibly general blueprint for addressing separation bias in discrete choice modeling. In this paper, we adapt PL methods to estimate the parameters from extensive form games of incomplete information. Using Monte Carlo experiments and replication analysis we find that the BR estimators offer substantial gains in bias, RMSE, and numerical stability. We offer two estimators (BR-SBI and BR-FIML) to provide analysts with options for fitting games to data where separation appears to be a problem. The BR-SBI is easily implemented using the existing `brglm` package for R, while we offer our own software for fitting the BR-FIML.[11]

Additionally, we also discuss several ways to look for and diagnose separation in situations where software does not issue warnings. First, we repeat standard recommendations about visual inspection. When coefficients and standard errors seem unusually large, analysts should be concerned that separation bias may be present. Second, analysts can compare models, when the SBI and the FIML produce very dissimilar results there is likely a problem. Third, we introduce the linear programming diagnostic from Konis (2007) to political scientists, which is fully implemented in the `detectseparation` package. We detail five ways to use this tool in the case of fitting strategic probits that are very quick and easy to do.

[11] Specifically, we provide software built around the existing `games` package to provide both the Jeffreys prior penalty and the Cauchy penalty for three different forms of the two-player deterrence game here: action-specific normally distributed shocks (as in the Monte Carlos), outcome-specific normally distributed shocks (as in the applications and the Monte Carlo in Appendix D), and action-specific type-1 extreme value shocks (logit choice probabilities).

Two separate avenues of future work present themselves. First, researchers should further explore the trade-offs among different penalty terms. While we find that both Cauchy and Jeffreys prior penalties work very well, more work can be done to identify scope conditions for when one should be preferred (as in Rainey 2016). Second, researchers should consider extending the BR framework even further into the empirical analysis of discrete choice games. For example, signaling models based on Lewis and Schultz (2003) are also common in EITM studies of international relations (e.g., Crisman-Cox and Gibilisco 2021), extending the BR framework could be helpful for scholars interested in empirical models of strategic interactions.

References

Albert, A. and J.A. Anderson. 1984. "On the Existence of Maximum Likelihood Estimates in Logistic Regression Models." *Biometrika* 71(1):1–10.

Bas, Muhammet Ali, Curtis S. Signorino and Robert W. Walker. 2008. "Statistical backwards induction: A simple method for estimating recursive strategic models." *Political Analysis* 16(1):21–40.

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

Cook, Scott J., John Niehaus and Samantha Zuhlke. 2018. "A warning on separation in multinomial logistic models." *Research & Politics* 5(2):1–5.

Cook, Scott J., Jude C. Hays and Robert J. Franzese. 2020. "Fixed Effects in Rare Events Data: A Penalized Maximum Likelihood Solution." *Political Science Research and Methods.* 8(1):92–105.

Crisman-Cox, Casey and Michael Gibilisco. 2021. "Estimating Signaling Games in International Relations: Problems and Solutions." *Political Science Research and Methods* 9(3):565–582.

Firth, David. 1993. "Bias Reduction of Maximum Likelihood Estimates." *Biometrika* 80(1):27–38.

Gelman, Andrew, Aleks Jakulin, Maria Grazia Pittau and Yu-Sung Su. 2008. "A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models." *The Annals of Applied Statistics* 2(4):1360–1383.

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

König, Thomas and Lars Mäder. 2014. "The Strategic Nature of Compliance." *American Journal of Political Science* 58(1):246–263.

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

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

Lewis, Jeffrey B. and Kenneth A. Schultz. 2003. "Revealing Preferences: Empirical Estimation of a Crisis Bargaining Game with Incomplete Information." *Political Analysis* 11(4):345–367.

Murphy, Kevin M. and Robert H. Topel. 1985. "Estimation and Inference in Two-Step Econometric Models." *Journal of Business & Economic Statistics* 3(4):370–379.

Nieman, Mark David. 2016. "The Return on Social Bonds." *Journal of Peace Research* 53(5):665–679.

Rainey, Carlisle. 2016. "Dealing with Separation in Logistic Regression Models." *Political Analysis* 24:339–355.

Rust, John. 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica* 55(5):999–1033.

Signorino, Curtis S. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *American Political Science Review* 93(2):279–297.

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

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

Zorn, Christopher. 2005. "A Solution to Separation in Binary Response Models." *Political Analysis* 13(2):157–170.