Deterrence and Diplomacy

Anne E. Sartori
Princeton University

EITM Summer School, CBRSS, Harvard

July 17, 2002

1. China's Deterrence Failure in the Korean War; a Motivating Case Study

> The Chinese people absolutely will not tolerate foreign aggression nor will they supinely tolerate seeing their neighbors savagely invaded by the imperialists.
>
> –Chinese Foreign Minister Chou En-lai, one week before the US First Cavalry Division crossed the 38th Parallel into North Korea, and about two weeks before large numbers of Chinese communist "volunteers" crossed into North Korea, enlarging the Korean War and pitting Chinese against United States troops.

China's many threats were dismissed as bluffs. Why? Why do states ever believe each other's threats, despite incentives to bluff?

1.1. The late 1940s marked beginning of the Cold War, including in the Korean War

- Chinese communist forces defeated the Nationalists and proclaimed the PRC in 1949
- Nationalist forces retreated to offshore islands, primarily Taiwan
- US government preoccupied with Chinese threats against the Nationalist government
- One year after the last U.S. troops withdrew from Korea, on June 24 1950, North Korean leaders sent over 100,000 soldiers over the border into South Korea
- This attacked was countered by the UN, under US leadership
- As the tide of the war turned in favor of the UN side, US leaders began to contemplate crossing the 38th Parallel into North Korea

1.2. Chinese attempt to dissuade U.S./UN forces from crossing the 38th Parallel

1. Verbal warnings (examples)

   - Sept. 30: Chou En-lai: "the Chinese people... will not supinely tolerate seeing their neighbors savagely invaded by the imperialists" (publicized speech)

   - Oct 2 (day after South Korean forces cross the parallel), Chou summons Panikkar to a midnight meeting to tell him that "the South Koreans did not matter, but American intrusion into North Korea would encounter Chinese resistance." (Panikkar)

2. Low-Cost Military Signals

   - Troop movements towards North Korean border mid-May (one mo. before start of war) - mid-July.

   - Resume after UN Inchon landing: about 320,000 Chinese troops in Manchuria in mid-October

1.3. U.S. leaders don't want to fight China, but they dismiss the threats as bluffs

1. Truman's memoirs: The Panikkar warning (Oct .3), "might very well be no more than a relay of Communist propaganda"

2. John M. Allison (US Delegation to the UN) (Oct. 4): "In the Secretary's opinion [[Secretary of State Acheson]]...we should not be unduly frightened by what was probably a Chinese Communist bluff."

1.4. Why was China's diplomacy not credible?

- The balance of forces does not explain China's lack of credibility. U.S. military leaders drew distinction between capabilities and intentions. For example:
  - U.S. Army Chief of Staff General J. Lawton Collins, asked if MacArthur had information that the Chinese would intervene if we went to the Yalu, said "I wouldn't say necessarily that it was a likelihood... but it certainly would represent a capability."

Chinese communists engage in a series of threats over Taiwan, 1948-1950

1. Qualified declaration of war against the U.S. in ‘48

2. Chinese threats against Taiwan continue through the summer of 1950, including after the North Korean attack on South Korea

3. When Truman orders U.S. Seventh Fleet to Taiwan Straits (June 27 1950), Chou calls the move “armed aggression against the territory of China” and promises that China will “liberate Taiwan.”

4. Mao backs down from a self-imposed September 1, 1950 deadline for attacking Taiwan.

These threats were not carried out by the time of the Korean War.

China’s threats to enter the Korean War followed its series of threats over Taiwan.

1.5. The case suggests that:

1. The absence of credible diplomacy is harmful.

2. States can acquire reputations for bluffing.

3. When a state has a reputation for bluffing, its diplomacy is more likely to fail in the immediate future.

The question: When and why can states use diplomacy to convince an adversary of vital information in a crisis?

2. My Argument

Diplomacy is effective because of a "communications norm:"

A state that is caught bluffing is expected to be likely to bluff again.

When a state has a reputation for bluffing, its diplomacy is less likely to be believed. It is less likely to attain its goals.

A state wants to keep its reputation for honesty, so states usually use diplomacy honestly.

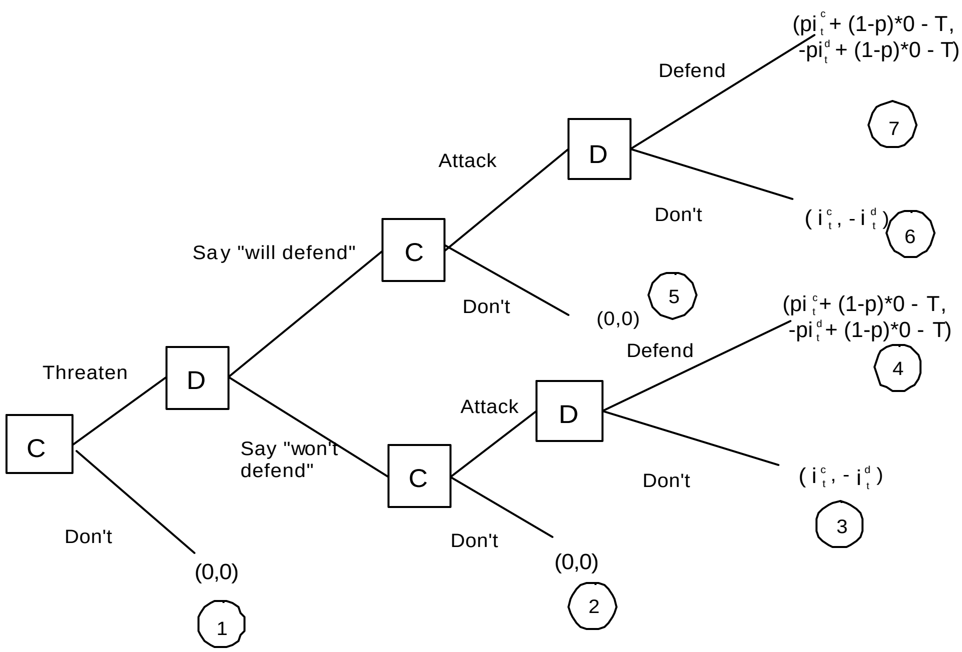
Diplomacy works (is believed) because it frequently is an honest indication of what an adversary intends to do.

3. Background: Deterrence Theory

- Rational/realist deterrence theory: Diplomacy is credible and likely to work insofar as the state possesses resolve.

$\Rightarrow$A state's ability to attain foreign-policy goals is largely based upon power and/or resolve. Diplomacy does not add to this ability

- Audience cost models: Signals must be costly to convey information (*What is a costly signal?*)



Payoffs: (Challenger's, Defender's)

C: Challenger's move
D: Defender's move

Stage-game Parameters

p: the probability with which the challenger is expected to win if war occurs.

T: the cost of fighting, which both states pay in case of war.

i^c_t, i^d_t: the challenger's and defender's values for the territory or other issue in this iteration, respectively. Note that the defender loses its value for the issue if it gives up through acquiescence, war, or backing down. In case of war, its expected value is p*(-i ^d_t) - T.

Figure 3.1: The Stage Game

4. The Formal Model

- Flow of a crisis: diplomatic stage, stage which corresponds to possible use of force/war

- 2 states wake up and are paired to interact. Some issue at stake.

- Next time, each state finds itself confronting a new potential adversary over a new issue.

- Challenger & defender. Defender has issues.

- What's unknown is whether or not adversary cares enough about the issue to fight. Defender trying to use diplomacy to convince challenger. (Obviously, how much a state cares about an issue varies from dispute to dispute.)

- Payoffs: happier when dispute resolved in your favor, happiest when that happens without war.

Game Tree

- Players
- Choices (CHEAP TALK)
- Payoffs
- Incomplete information
- Infinite repetition

Figure 2: Equilibrium Strategies

- Solution
- Choice of Equilibrium

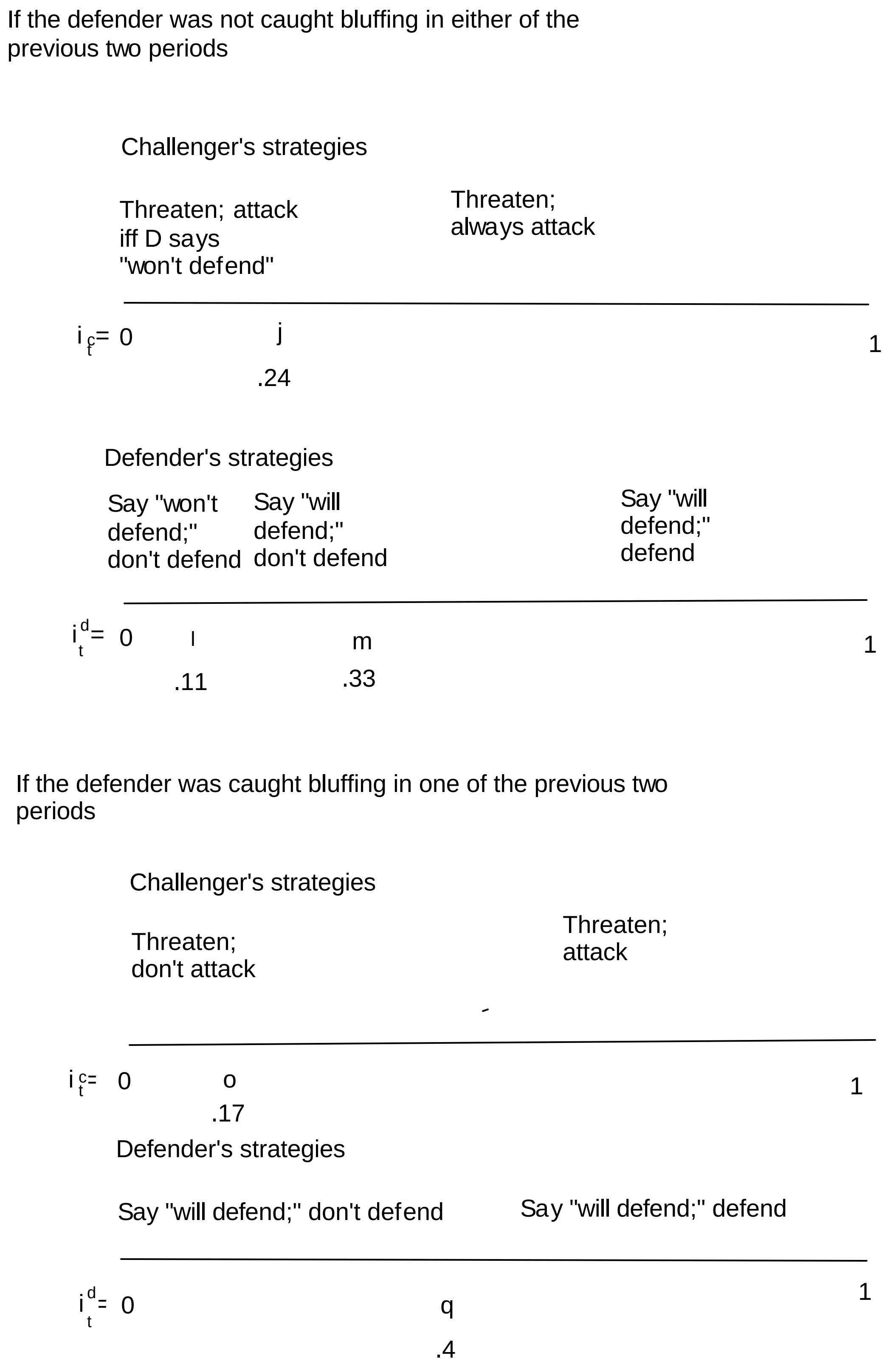


Figure 3.2: Equilibrium Strategies

5. Empirical Implications of the Formal Model

1. A defender with a reputation for honesty is more likely to defend following a deterrence failure than is a defender with a reputation for bluffing.

2. A challenger is more likely to attack following the defender's attempt at deterrence when the defender has a reputation for bluffing. (The defender's attempts at deterrence are more likely to fail when it has a reputation for bluffing.)

Some issues in choice of implications to test:

1. Alternative (partial) equilibrium

2. One eq. versus many?

1 Empirical Implications of the Formal Model

1. A challenger is more likely to attack following the defender's attempt at deterrence when the defender has a reputation for bluffing. (The defender's attempts at deterrence are more likely to fail when it has a reputation for bluffing.)

2. A defender with a reputation for honesty is more likely to defend following a deterrence failure than is a defender with a reputation for bluffing.

$$U_C(\text{ Attack} \mid \text{threatened, D tried deterrence}) = \alpha_1 + \beta_1 \text{ ln balance of forces} + \beta_2 \text{ def.'s reputation for bluffing} + u_2,$$

$$U_D(\text{Defend} \mid \text{ C threatened, tried deterrence, C attacked}) \; \alpha_1' + \beta_1' \text{ ln balance of forces} + \beta_2' \text{ def.'s reputation for bluffing} + u_2'.$$

An Estimator for Some Binary-Outcome Selection Models without Exclusion Restrictions

2 Introduction

2.1 The Data Generating Process

The selection eq. $U_{1i} = \boldsymbol{\gamma}'\mathbf{x}_i + u_{1i}$

The eq. of interest $U_{2i} = \boldsymbol{\beta}'\mathbf{x}_i + u_{2i}$

where x_i is a vector of explanatory variables for observation *i*, $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are vectors of parameters, the $u_i s$ are normally-distributed error terms, and the U_is are underlying utilities.

The reason for selection bias is that the u_is are correlated.

We observe:

$$Z_{1i} = \begin{cases} 1 \text{ if } U_{1i} \geq 0, \\ 0 \text{ otherwise;} \end{cases}$$

$$Z_{2i} = \begin{cases} 1 \text{ if } U_{2i} \geq 0, \\ 0 \text{ otherwise.} \end{cases}$$

The substantive/methodological problem:

- Nonrandom selection with correlated errors
  - OLS/probit inconsistent even for the observations in the selected sample
  - Heckman/Achen-type methods solve the problem

- Identical explanatory factors
  - Heckman/Achen methods inappropriate, identify from functional form alone

Current practice:

1. Dredge up an "extra" explanatory variable for the selection equation (possibly leading to specification error);

2. Identify from functional form alone;

3. Give up.

My addition to the menu of options, for problems with binary observed outcomes:

4. A new maximum likelihood estimator

   (a) Based on an additional identifying assumption: that the errors are perfectly correlated for an observation

   (b) Most likely to be reasonable when it's needed: when identical explanatory factors influence selection and the subsequent outcome of interest.

3 Overview of this hour

1. Briefly review selection bias

2. Introduce the estimator

3. Discuss properties of the estimator

4. Present simulation results showing that this is a better choice than the Heckman-type estimator for this problem

5. Next–data, variables, results

4 The Problem of Selection Bias

- We might estimate

Selection eq.	$\mathbf{U}_{1i} = \boldsymbol{\gamma}\mathbf{x}_i + \mathbf{u}_1$
Eq. of interest	$\mathbf{U}_{2i} = \boldsymbol{\beta}\mathbf{x}_i + \mathbf{u}_2$

where x_i is an independent variable and u_i are normally-distributed error terms. We observe:

$Z_{1i} = 0$ if $U_{1i} < 0$

$\quad 1$ if $U_{1i} \geq 0$;

$Z_{2i} = 0$ if $U_{2i} < 0$ and

$\quad 1$ if $U_{2i} \geq 0$.

- The source of selection bias: the two equations have correlated error terms. Here, the extent to which the state in question cares about the disputed issue is a key unmeasured variable in both equations.

- A observation enters the sample only if

$$\mathbf{U}_{1i} = \boldsymbol{\gamma}\mathbf{x}_i + \mathbf{u}_{1i} > \mathbf{0}$$

$\gamma \mathbf{x}_i$ "big enough"$\rightarrow$select in for wide variety of error terms

$\gamma \mathbf{x}_i$ small$\rightarrow$select in only if error is big

$\Longrightarrow$The independent variable is correlated with the error term in the equation of interest.

$\Longrightarrow$Probit estimation of the equation of interest leads to inconsistent estimates of the effect of the explanatory variable–*even for the observed sample*.

- Discuss: Identification problem

5 The Estimator

5.1 The Key Identifying Assumption

The error terms from the two equations are identical for an observation: $u_{1i}=u_{2i}$.

The assumption is initially plausible:

- The observed dependent variables are dichotomous;
- The underlying dependent variables are on the same scale;
- Both error terms are normally distributed.

The assumption is more likely to be reasonable when selection and the outcome :

- Are close in time and space;
- Have the same causes (identical explanatory factors);
- Represent similar decisions or goals.

5.2 The Likelihood and the Estimator

I define a random variable Y_{ij} such that:

$$\begin{cases} Y_{i0}=1 \text{ if } Z_1=0, \text{ 0 otherwise;} \\ Y_{i1}=1 \text{ if } Z_1=1 \text{ and } Z_2=0, \text{ 0 otherwise;} \\ Y_{i2}=1 \text{ if } Z_1=1 \text{ and } Z_2=1, \text{ 0 otherwise.} \end{cases}$$

Let $P_{ij} \equiv$ prob$[Y_{ij}=1]$.

- The likelihood function is proportional to the product of the probabilities of the observations, $\prod_{i=1}^{n} \prod_{j=0}^{2} P_{ij}^{Y_{ij}}$.

- Then,

$$L^* \equiv lnL \propto \sum_{i=1}^{n} \sum_{j=0}^{2} Y_{ij} lnP_{ij}$$

$Y_{ij}\ln P_{ij} \equiv 0$ if $P_{ij} \leq 0$. At the truth, this only happens if $Y_{ij} = 0$.

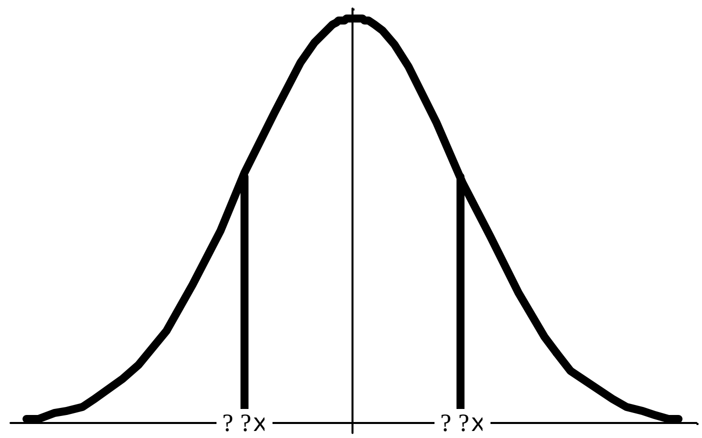


Figure 1: The Distribution of u_i When -$\boldsymbol{\gamma}'\mathbf{x}_i$<-$\boldsymbol{\beta}'\mathbf{x}_i$

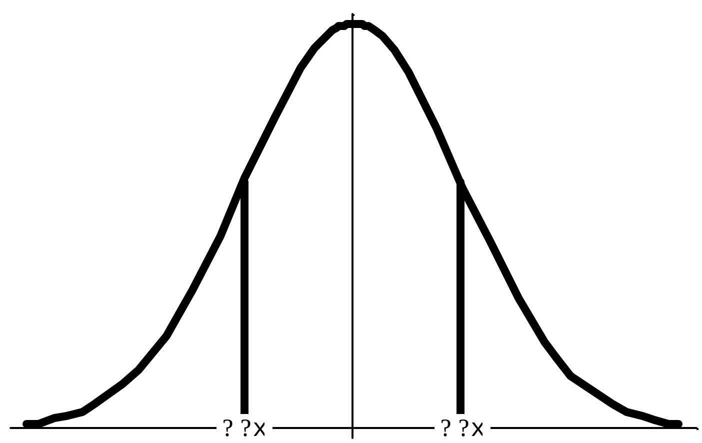


Figure 2: The Distribution of u_i When -$\boldsymbol{\beta}'\mathbf{x}_i$<-$\boldsymbol{\gamma}'\mathbf{x}_i$

To build the likelihood function, I substitute in the *ex ante* probability of seeing an observation, $P_{ij.}$ There are three cases:

1. j=0

$$P_{i0} = \Phi(-\boldsymbol{\gamma}'\mathbf{x}_i).$$

2. j=1

$$P_{i1} = \Phi(-\boldsymbol{\beta}'\mathbf{x}_i) - \Phi(-\boldsymbol{\gamma}'\mathbf{x}_i).$$

3. j=2

$$P_{i2} = min[\Phi(\boldsymbol{\gamma}'\mathbf{x}_i), \Phi(\boldsymbol{\beta}'\mathbf{x}_i)],$$

I now define the estimator:

$$\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}} := arg\ max_{\boldsymbol{\beta},\boldsymbol{\gamma} \in \Theta}\ L^* \tag{3}$$

where Θ is the parameter space.

6 Properties of the Estimator

Checking regularity conditions for MLE

- MLE's nice properties (consistency, asymptotic normality, asymptotic efficiency) automatically obtain *if the regularity conditions are met*.

- It is not standard to check them!

- The regularity conditions are sufficient but not necessary.

- My estimator does not meet the regularity conditions, so

- I prove consistency and normality. (Efficiency follows from normality for a MLE.)

MLE Regularity conditions (King 1990, 75)

1. The values of y for which f(y|θ)$>$0 (i.e., the sample space) do not depend on θ;

2. For each $\theta \in \Theta$, the probability distribution f(y|θ) is twice continuously differentiable for all y$\in$Y

3. The information matrix I is positive definite and finitely bounded, where

$$I_n(\theta) \equiv E\left[\left(\frac{\partial \ln f(y|\theta)}{\partial\theta}\right)\left(\frac{\partial \ln f(y|\theta)}{\partial\theta}\right)'\right].$$

4. For each $\theta \in \Theta$, $\left|\frac{\delta^i \ln f(y|\theta)}{\delta\theta^i}\right| \leq h_i(y)$, $i = 1, 2, 3$,

where

E[$h_i(y)$] $< \infty$, i=1,2 and [$h_3(y)$] does not depend on θ.

A couple of key steps in the proof of consistency:

1. I divide the likelihood up into observations i=1,..., m where $-\boldsymbol{\gamma}_{\mathbf{0}}'\mathbf{x}_i < -\boldsymbol{\beta}_{\mathbf{0}}'\mathbf{x}_i$ (Y_{1i}=1 has positive probability) and observations $i=m+1$,...,n where $-\boldsymbol{\beta}_{\mathbf{0}}'\mathbf{x}_i < -\boldsymbol{\gamma}_{\mathbf{0}}'\mathbf{x}_i$ (Y_{1i}=1 has zero probability). $\boldsymbol{\gamma}_{\mathbf{0}}$, $\boldsymbol{\beta}_{\mathbf{0}}$ represent the values at the truth.

$$\ln L \;\propto\; Q(\theta) = \sum_{i=1}^{m}\sum_{j=0}^{2} Y_{ij} ln P_{ij}\Big|_{-\boldsymbol{\gamma}_{\mathbf{0}}'\mathbf{x}_i<-\boldsymbol{\beta}_{\mathbf{0}}'\mathbf{x}_i} + \sum_{i=m+1}^{n}\sum_{j=0,2} Y_{ij} ln P_{ij}\Big|_{-\boldsymbol{\beta}_{\mathbf{0}}'\mathbf{x}_i<-\boldsymbol{\gamma}_{\mathbf{0}}'\mathbf{x}_i}$$

where I define $Y_{ij}\ln P_{ij}$ as 0 if $P_{ij} \leq 0$.

2. Now I can rule out a potential problem:

   (a) We saw above that $P_{i1} = \Phi(-\boldsymbol{\beta}'\mathbf{x}_i) - \Phi(-\boldsymbol{\gamma}'\mathbf{x}_i)$.

   (b) At the truth, this poses no problem, since $\Phi(-\boldsymbol{\beta}_0'\mathbf{x}_i) - \Phi(-\boldsymbol{\gamma}_0'\mathbf{x}_i) > 0$ when $Y_{i1}=1$.

   (c) At guesses of $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ not the truth, it could be that $P_{i1} \leq 0$ and the derivatives of the log likelihood function do not exist.

   (d) The proofs, however, require them to exist only in an open neighborhood of the true parameters. Since the truth is $-\boldsymbol{\gamma}_0'\mathbf{x}_i < -\boldsymbol{\beta}_0'\mathbf{x}_i$ for the observations where $Y_{i1}=1$, we can find a neighborhood around the true parameters such that $-\boldsymbol{\gamma}'\mathbf{x}_i < -\boldsymbol{\beta}'\mathbf{x}_i$. So the condition is satisfied.

7 Monte Carlo Results

7.1 The Simulations

The goal:

- To vary the "real" situation in order to evaluate the performance of probit, the Heckman-type estimator, and my estimator under different conditions, all of which include nonrandom selection and identical explanatory variables in the selection equation and the outcome equation;

- Especially to evaluate the estimates for the outcome equation

The experiments:

- I generate 1000 observations of an "independent variable," x, from a normal distribution with mean zero and variance 0.64.

- I consider situations in which the Heckman estimator's key assumption, normality, is met (using bivariate standard normal errors)

- I consider situations in which my estimator's key assumption, identical errors, is close to true (ρ=.9), then cases in which it is not:

  - The true ρ is .5, .1, or less

- I create the underlying utilities and the dichotomous dependent variables as follows:

$$U_1 = 1.25 * x + u_1$$

$$U_2 = -0.7 + 1.5 * x + u_2$$

  U_1 represents the actor's utility from "selecting in," U_2 represents its utility from "going on."

  One observes $\mathsf{Y}_1 = 1$ if $U_1 > 0$; $\mathsf{Y}_1 = 0$ otherwise, and Y_2=1 if $U_1 > 0$; $\mathsf{Y}_2 = 0$ otherwise.

  I refer to the intercept and slope parameters from the outcome equation as α_2 and β_2, respectively.

- I replicate 1000 times upon the same samples of 100 and 1000 using each correlation.

7.2 The Results

7.2.1 Coefficient Estimates

- The mean bias of my estimates is lower than that of Heckman's at high correlations and becomes higher than Heckman's as the correlation moves towards zero. The switch occurs somewhere between a true correlation of 0.3, where my estimates have a smaller mean bias, and one of 0.2, where Heckman's have a smaller mean bias.

- RMSE: <u>Remember:</u> The RMSE of an estimate of β_2 is $\sqrt{\frac{\sum_i(\widehat{\beta}_{i2}-\beta_2)^2}{n}}$, where β_2 is the true value, $\widehat{\beta}_{i2}$ is the estimate of β_2 in the ith replication, and n is the number of estimates.

  - My estimate has lower RMSE than Heckman's in each simulation down to ρ of -0.1. However, when the true correlation between the errors is small in magnitude (0.1 or -0.1 in these simulations), the ordinary probit estimate has lower RMSE than those of either of the selection estimators.

7.2.2 Coverage

- My estimator's coverage is more accurate at high correlations and becomes less-accurate at low correlations. The switch occurs somewhere between a true correlation of 0.4, where my estimator's coverage is more accurate, and 0.3, where Heckman's coverage is more accurate.

7.2.3 The Heckman estimator's estimates of ρ

- The estimates are quite biased and have large RMSEs

- The distribution is often quite flat with a couple of peaks

7.2.4 Summary

For the problem of identical explanatory variables, this is a better estimator. Even when the Heckman estimator's identifying assumption is met, and my estimator's key assumption is fairly far from true (the true ρ is 0.5 or lower, while I assume 1.0), my estimates have lower mean bias and RMSE and better coverage.

The decision about which estimator to use cannot be based on Heckman's estimate of ρ.

8 Conclusion

1. This is a consistent, maximum-likelihood estimator for use in selection models when identical explanatory factors affect the selection equation and a subsequent outcome of interest, and when both observed dependent variables are dichotomous.

2. The estimator additionally identifies from an assumption that the error terms in the selection equation and the outcome equation are identical for an observation.

3. This is a better estimator than Heckman or probit for this problem:

   (a) the Heckman-type models are best used only with an exclusion restriction. My identifying assumption is likely to be reasonable exactly when an exclusion restriction is inappropriate.

(b) The simulations show that this estimator is a better choice than the Heckman estimator for this substantive problem when the sample size is small.

4. Caveat: I'm still planning to do additional simulations with rare events.

9 The Escalation of International Disputes

9.1 The Data

- Information about each MID that occurs between the years of 1816 and 1993, a total of over 2000 disputes.

  What is a Militarized International Dispute (MID)? An event in which at least one state took overt militarized action against another, where a militarized action may be as minor as a threat to use force or as major as a full-scale war (Jones, Bremer, and Stuckey 1996)

- Information about each pair of states (dyads) that did not have a MID

9.1.1 Dependent Variables

1. The defender's willingness to follow through: If the defender tries deterrence, does the challenger nevertheless decide to attack (resulting in a deterrence failure), or does it back down?

2. Deterrence success/failure: if the challenger decides to attack, does the defender follow through on its deterrent threat and fight, or does it back down?

Also:

3. Did the "challenger" threaten the use of force?

4. Did the defender try deterrence?

- MID data: 5 hostility levels (in increasing order): no militarized action, threat to use force, display of force, use of force, and war.

9.1.2 Independent Variables

1. Reputations for honesty and for bluffing:

Dispute outcomes:

1) there is no dispute (neither state becomes a challenger by threatening the use of force); or a dispute exists (one state threatens) and:

2) the defender does not try deterrence,

3) the defender threatens and the challenger backs down,

4) the defender threatens, the challenger attacks, and the defender backs down

5) both states threaten and follow through on their threats to use force.

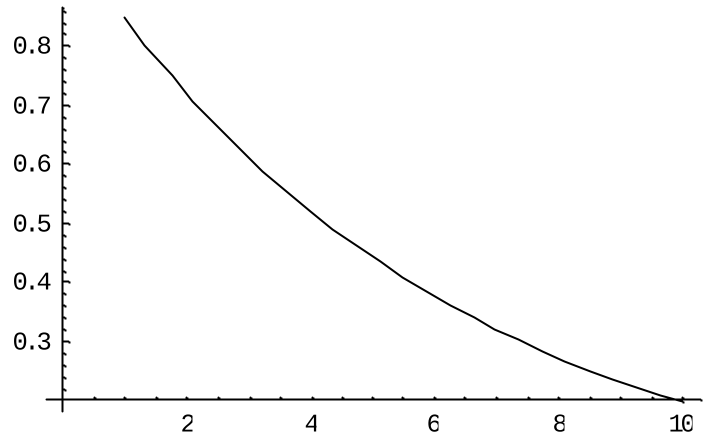


Figure 3: Coding of the Effect of Reputation with Years Passed

- A defender increases its reputation in a given year if the outcome is #4; either acquiescence or following through is honest behavior

- Reputations are temporary: A particular called bluff in the past ten years adds $.85^{\alpha}$ to a state's reputation for bluffing, where α is the time passed since the bluff was called.

2. The balance of forces: ln(challenger's capabilities/defender's capabilities)

Moving from the formal model to empirical definitions of reputations requires some judgement calls. I also explore the consequences of defining reputations differently. The first alternative definition is simpler: a defender's reputation is based on whether or not it was a defender in the previous year and was caught bluffing in that year. The second is more complicated: a defender's reputation is based upon past behavior when it was a challenger, in addition to its behavior when it was a defender.

9.2 Control Variables Not in the Model

(statistical model estimated with and without these)

It has become customary in statistical analyses of international relations to include a number of variables that are irrelevant to the theory being tested; these are "controls" suggested by competing or complementary theories. Including a long list of variables often is a substitute for careful thought about the factors implied by the theory. To convince readers who disagree with my simple specification, I also report results from analyses that include control variables suggested by the literature on international crises:

- whether or not each state is a major power, according to Singer and Small's definition [?, 44-45],

- whether or not the states are contiguous (either actually contiguous or across up to 12 miles of water),

- whether or not both states in the pair are democracies.

- with power parity instead of the balance of forces

10 Main Results

10.1 The Challenger's Decision about Whether or Not to Attack Following a Defender's Deterrent Threat

What is the effect of the defender's reputation on the challenger's decision about whether or not to attack if the defender tries deterrence, *holding constant the probability of selection* (the probability that the challenger threatens to use force to attain a change in the status quo)?

10.2 The Defender's Decision about Whether or Not to Defend if its Threat Fails to Deter an Attack

What is the effect of the defender's reputation on the defender's decision about whether or not to fight if deterrence fails, *holding constant the probability of selection* (the probability that the defender tries deterrence)?

11 Conclusion

The patterns in the data are consistent with the theory. Based on the results I've presented today:

- A typical defender with a reputation for honesty is more likely to deter an attack than a defender with a reputation for bluffing by roughly 21 percentage points.

- A typical defender with a reputation for honesty is more likely to follow through on its threats than a defender with a reputation for bluffing by five or six percentage points.

- The second fact explains the first. Defenders' deterrent threats are more likely to succeed (challengers are less likely to attack after hearing them) when those defenders have reputations for honesty precisely because defenders with reputations for honesty are more likely to mean what they say.

- States use diplomacy carefully because it is valuable: a state with a reputation for bluffing has more difficulty attaining its goals in its subsequent dispute.

	Prob(attack| threatened) defender rep. honesty second stage	Prob(attack| threatened) defender rep. bluffing second stage	Change in prob c deterren failure
capabilities 1-to-1	60.3%	82.2%	21.9%
capabilities 1.46-to-1	60.1%	81.7%	21.6%
capabilities 7-to-1	59.1%	80.3%	21.2%

Table 1: Relationship between defender's reputation and deterrence success or failure.

To perform these thought experiments, I hold the defender's reputation constant in the equation that represents the challenger's decision to threaten the use of force and vary its reputation in the equation that represents whether or not it attacks. I assume that the defender has a reputation for bluffing when the challenger decides whether or not to threaten the use of force. Here, a reputation for bluffing means that a defender's reputation has a value of .32, so its reputation is less-honest than those of 90% of defenders. A defender has a reputation for bluffing with value .85 if it was caught in a bluff in the previous year.

	Prob(defend| tried deterrence defender rep honesty second stage	Prob(defend| tried deterrence defender rep. bluffin, second stage
capabilities 1-to-1	84.0%	78.7%
capabilities 1.46-to-1	87.2%	81.8%
capabilities 7-to-1	86.3%	80.2%

Table 2: Relationship between defender's reputation and probability that the defender follows through on its threats.

ALTERNATIVE DEFINITIONS OF REPUTATIONS

I consider robustness to two alternative definitions, one more simple and one more complicated.

1. The first, and simpler, is: the defender has a reputation for bluffing if it was caught bluffing in the previous year (it tried deterrence, the challenger nevertheless attacked, and the defender backed down); otherwise, it has a reputation for honesty.

2. The second definition includes behavior as a challenger as part of the basis of the defender's reputation. The formal analysis argues that *defenders* obtain reputations for bluffing or for honesty, and defenders with reputations for honesty are more able to use threats effectively. The definition of the defender's reputation that I used earlier follows closely from these formal analyses. In practice, however, challengers clearly also do communicate. Though my formal work does not speak to this subject, it is possible that

states also acquire reputations for honesty or for bluffing from interactions in which they are challengers. It also is possible that challengers have reputations, and that these reputations affect the course of international disputes.

- In the second alternative definition, I still consider a state to increase its reputation for bluffing if it is caught bluffing when it is a defender.

- I now also consider a state to increase its reputation for bluffing in a given year if it is a challenger and it is revealed to be bluffing (it threatens the use of force; the defender indicates its willingness to fight, and the challenger backs down). I again code the reputations from a state's behavior over the previous ten years, with events farther in the past carrying less weight in the same way. However, I now code each state's reputation as based on all such behavior – regardless of whether the state is a challenger or a defender now and regardless of whether it was a challenger or a defender in a previous dispute.

This leads to a reputations variable for the challenger and one for the defender, each ranging between 0 and 4.55.

- In the statistical analyses that use this definition, I also include a variable that indicates the interaction between the two states' reputations. If the challenger's reputation does affect its ability to use diplomacy, one might expect each state's decisions to be different when both have severe reputations for bluffing than when only one has such a reputation.

$\Longrightarrow$My overall conclusions are the same with either alternative specification. In each case, the estimated effects of the defender's reputation are of the expected sign and quite precise.