

Supplementary Materials for
“Challenges to Inference in the Study of Crisis
Bargaining”

PHILIP ARENA

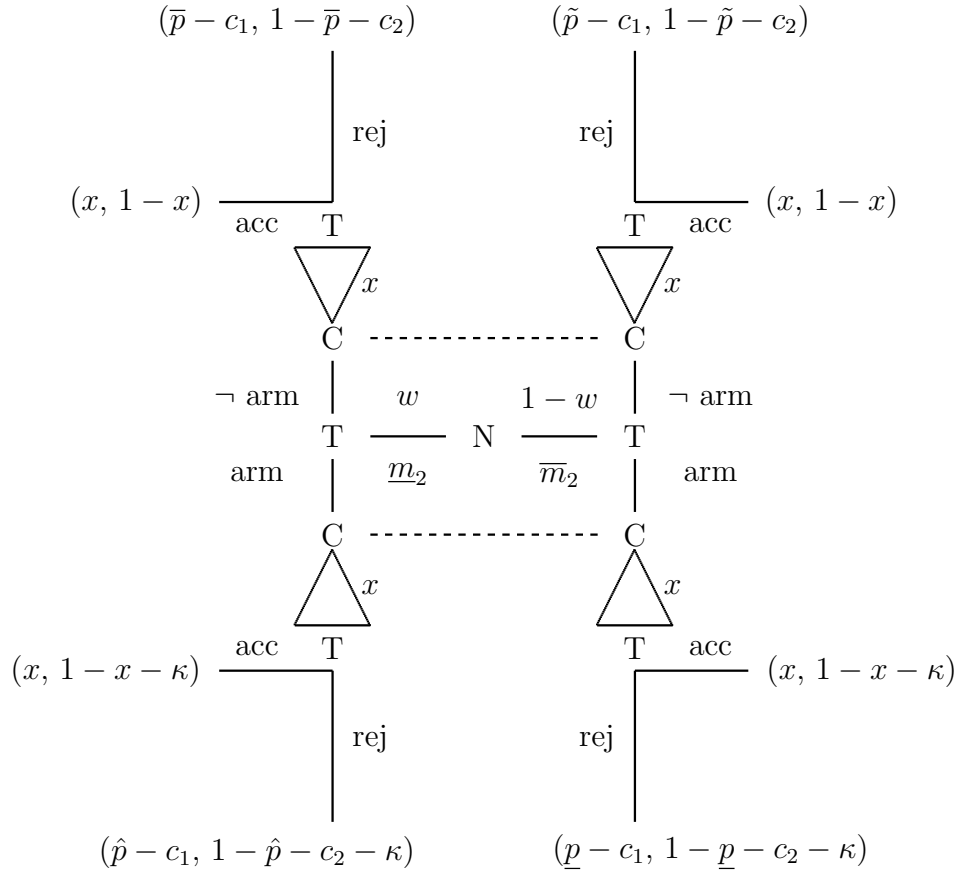
Department of Political Science, University at Buffalo (SUNY)

KYLE A. JOYCE

Department of Political Science, University of California, Davis

Extensive Form of the Crisis Bargaining Model

Below, we provide the extensive form of our crisis bargaining model.



Derivation of Equilibria in the Bargaining Model

Since we assume the Challenger is uninformed about the Target's type, we solve the model for perfect Bayesian equilibria, where players' strategies must be sequentially rational and their beliefs weakly consistent with Bayes' Rule.

First, consider the acceptance rule used by the Target (T). Generically, T accepts the demand the Challenger (C) issues after observing T choose not to arm, which we denote x_N to distinguish from demands issued after T arms, which we denote x_A , if and only if (iff) $u_T(\text{acc}) \geq EU_T(\text{war})$, which is equivalent to

$$1 - x_N \geq 1 - p - c_2, \tag{1}$$

which can be rewritten as

$$x_N \leq p + c_2. \tag{2}$$

Since p depends on m_2 , though, this actually gives us two acceptance rules. The relatively strong type accepts iff $x_N \leq \tilde{p} + c_2$ and the relatively weak type accepts iff $x_N \leq \bar{p} + c_2$. Let $\underline{x}_N \equiv \tilde{p} + c_2$ and $\bar{x}_N \equiv \bar{p} + c_2$. Thus, T accepts iff $x_N \leq \underline{x}_N$ when relatively strong and iff $x_N \leq \bar{x}_N$ when relatively weak. Note we assume all critical values of x are less than 1 to ensure interior solutions.

C can readily infer that, if T did not arm, then

$$Pr(\text{war}) = \begin{cases} 0 & \text{if } x \leq \underline{x}_N \\ 1 - w'' & \text{if } \underline{x}_N < x \leq \bar{x}_N \\ 1 & \text{if } x > \bar{x}_N. \end{cases}$$

We can immediately establish that C never sets $x < \underline{x}_N$, $x > \bar{x}_N$, or $x = x_1$ where $\underline{x}_N < x_1 < \bar{x}_N$. When C sets $x \leq \underline{x}_N$, T is certain to accept regardless of type, and so

$u_C(x \leq \underline{x}_N) = x$. Thus, it follows that $u_C(x < \underline{x}_N)$ is strictly dominated by $u_C(x = \underline{x}_N)$. Since $EU_C(x = x_1) \geq EU_C(x = \bar{x}_N)$ is equivalent to

$$w''x_1 + (1 - w'')(\tilde{p} - c_1) \geq w''\bar{x}_N + (1 - w'')(\tilde{p} - c_1), \quad (3)$$

or $x_1 \geq \bar{x}_N$, which cannot be true since setting $x = x_1$ is strictly dominated by $x = \bar{x}_N$.

Finally, because $EU_C(x > \bar{x}_N) \geq EU_C(x = \bar{x}_N)$ is equivalent to

$$w''(\bar{p} - c_1) + (1 - w'')(\tilde{p} - c_1) \geq w''\bar{x}_N + (1 - w'')(\tilde{p} - c_1), \quad (4)$$

or $c_1 + c_2 \leq 0$, which also cannot be true since setting $x > \bar{x}_N$ is strictly dominated by $x = \bar{x}_N$. Thus, the only values of x that C may select in equilibrium after T chooses not to arm are $x = \underline{x}_N$, which T accepts regardless of type, and $x = \bar{x}_N$, which T accepts iff $m_2 = \underline{m}_2$.

We now turn to evaluating $u_C(x = \underline{x}_N) \geq EU_C(x = \bar{x}_N)$, which is equivalent to

$$\underline{x}_N \geq w''\bar{x}_N + (1 - w'')(\tilde{p} - c_1), \quad (5)$$

or

$$w'' \leq \frac{c_1 + c_2}{\bar{p} - \tilde{p} + c_1 + c_2}. \quad (6)$$

Thus, we can say that C sets $x_N = \underline{x}_N$ if $w'' \leq \underline{w}$ and sets $x_N = \bar{x}_N$ if $w'' > \underline{w}$, where $\underline{w} \equiv \frac{c_1 + c_2}{\bar{p} - \tilde{p} + c_1 + c_2}$. Note that since $c_1 + c_2 < \bar{p} - \tilde{p} + c_1 + c_2 \Leftrightarrow \bar{p} > \tilde{p}$, it must be true that $\underline{w} \in (0, 1)$.

Following the same logic as above, it is straightforward to demonstrate that if T arms at the outset of the crisis, C sets $x_A = \underline{x}_A$ if $w' \leq \bar{w}$ and sets $x_A = \bar{x}_A$ if $w' > \bar{w}$, where $\underline{x}_A \equiv \underline{p} + c_2$, $\bar{x}_A \equiv \hat{p} + c_2$, and $\bar{w} \equiv \frac{c_1 + c_2}{\hat{p} - \underline{p} + c_1 + c_2}$.

With these preliminaries established, we now turn to the evaluation of candidate perfect Bayesian equilibria (PBE). We begin by ruling out separating equilibria.

First consider the following potential PBE: C sets $x = \underline{x}_A$ and believes $w' = 0$ if T arms and sets $x = \bar{x}_N$ and believes $w'' = 1$ if T does not arm; the weaker type of T accepts iff $x_N \leq \bar{x}_N$ when he does not arm, accepts $x_A \leq \bar{x}_A$ iff he does arm, but, in equilibrium, does not in fact arm; and the stronger type accepts $x_N \leq \underline{x}_N$ iff he does not arm, accepts $x_A \leq \underline{x}_A$ iff he does arm, and does in fact arm in equilibrium. In such an equilibrium, the decision to arm would signal that T is relatively strong. Holding T 's strategies constant, C 's beliefs follow immediately from Bayes' Rule and C 's proposals must be sequentially rational given those beliefs since $0 < \underline{w} < \bar{w} < 1$.

Incentive compatibility for the weak type of T requires that $u_T(\neg \text{arm} | x_N = \bar{x}_N) \geq u_T(\text{arm} | x_A = \underline{x}_A)$, or

$$1 - \bar{x}_N \geq 1 - \underline{x}_A - \kappa, \tag{7}$$

which simplifies to $\kappa \geq \bar{p} - \underline{p}$. For the strong type, $u_T(\text{arm} | x_A = \underline{x}_A) \geq u_T(\neg \text{arm} | x_N = \bar{x}_N)$, or

$$1 - \underline{x}_A - \kappa \geq 1 - \tilde{p} - c_2, \tag{8}$$

which simplifies to $\kappa \leq \tilde{p} - \underline{p}$. Note the strong type's incentive compatibility constraint is not the mirror image of the weak type's because the weak type accepts $x_N = \bar{x}_N$, whereas the strong type rejects.

Since $\bar{p} - \underline{p} > \tilde{p} - \underline{p} \Leftrightarrow \bar{p} > \tilde{p}$, for the equilibrium to hold, κ cannot simultaneously be large enough to dissuade the weak type from arming yet small enough to entice the strong type to arm. That is, κ cannot simultaneously be larger than one quantity while being smaller than the other, given the first quantity is itself larger than the second quantity. Thus, no separating equilibrium exists. If arming is to prevent war, it cannot, in this model,

be because it serves as a signal that T is strong.

Now consider the other potential separating PBE: C sets $x = \bar{x}_A$ and believes $w' = 1$ if T arms and sets $x = \underline{x}_N$ and believes $w'' = 0$ if T does not arm; the weak type T accepts $x_N \leq \bar{x}_N$ iff he does not arm, accepts $x_A \leq \bar{x}_A$ iff he does arm, and, in equilibrium, arms; and the strong type accepts $x_N \leq \underline{x}_N$ iff he does not arm, accepts $x_A \leq \underline{x}_A$ iff he does arm, and does not arm in equilibrium. In such an equilibrium, the decision to arm would signal that T is relatively weak. Again, holding T 's strategies constant, C 's beliefs follow immediately from Bayes' Rule and C 's proposals must be sequentially rational given those beliefs.

Incentive compatibility for the weak type requires $u_T(\text{arm}|x_A = \bar{x}_A) \geq u_T(\neg \text{arm}|x_N = \underline{x}_N)$, or

$$1 - \bar{x}_A - \kappa \geq 1 - \underline{x}_N, \quad (9)$$

which simplifies to $\kappa \leq \tilde{p} - \hat{p}$. For the strong type, $u_T(\neg \text{arm}|x_N = \underline{x}_N) \geq u_T(\text{arm}|x_A = \bar{x}_A)$, or

$$1 - \underline{x}_N \geq 1 - \underline{p} - c_2 - \kappa, \quad (10)$$

which simplifies to $\kappa \geq \tilde{p} - \underline{p}$. Note that the strong type's incentive compatibility constraint is not the mirror image of the weak type's because the weak type accepts $x_A = \bar{x}_A$, whereas the strong type rejects such a demand. Note that $\tilde{p} - \underline{p} > \tilde{p} - \hat{p} \Leftrightarrow \hat{p} > \underline{p}$. Thus, κ cannot simultaneously be large enough to dissuade the strong type from arming yet small enough to entice the weak type to arm. As above, κ cannot simultaneously be larger than one quantity while being smaller than the other quantity, given the first quantity is itself larger than the second quantity. Thus, no such separating equilibrium exists.

Now we consider pooling equilibria. Suppose $w' \leq \bar{w}$ and $w'' \leq \underline{w}$. In such an equilibria, C would set $x = \underline{x}_A$ if T arms and $x = \underline{x}_N$ if T does not arm. In a pooling PBE in which both types arm, incentive compatibility for the weak type would require $u_T(\text{arm}|x_A = \underline{x}_A) \geq$

$u_T(\neg \text{arm}|x_N = \underline{x}_N)$, or

$$1 - \underline{x}_A - \kappa \geq 1 - \underline{x}_N, \quad (11)$$

which simplifies to $\kappa \leq \tilde{p} - \underline{p}$. For the strong type, $u_T(\text{arm}|x_A = \underline{x}_A) \geq u_T(\neg \text{arm}|x_N = \underline{x}_N)$, or

$$1 - \underline{x}_A - \kappa \geq 1 - \underline{x}_N, \quad (12)$$

which, again, simplifies to $\kappa \leq \tilde{p} - \underline{p}$. Thus, as long as $\kappa \leq \tilde{p} - \underline{p}$, incentive compatibility is satisfied for both the weak and strong types and the equilibrium holds. It is straightforward to show, provided $\kappa > \tilde{p} - \underline{p}$, a pooling PBE in which both types choose not to arm holds.

Now suppose $w' > \bar{w}$ and $w'' \leq \underline{w}$. In this case, C would set $x = \bar{x}_A$ if T arms and $x = \underline{x}_N$ if T does not arm. In a pooling PBE in which both types arm, incentive compatibility for the weak type requires $u_T(\text{arm}|x_A = \bar{x}_A) \geq u_T(\neg \text{arm}|x_N = \underline{x}_N)$, or

$$1 - \bar{x}_A - \kappa \geq 1 - \underline{x}_N, \quad (13)$$

which simplifies to $\kappa \leq \tilde{p} - \hat{p}$. For the strong type, $u_T(\text{arm}|x_A = \bar{x}_A) \geq u_T(\neg \text{arm}|x_N = \underline{x}_N)$, or

$$1 - \underline{p} - c_2 - \kappa \geq 1 - \underline{x}_N, \quad (14)$$

which simplifies to $\kappa \leq \tilde{p} - \underline{p}$. Thus, as long as $\kappa \leq \tilde{p} - \underline{p}$, incentive compatibility is satisfied for both the weak and strong types and the equilibrium holds, and provided $\kappa > \tilde{p} - \hat{p}$, a pooling PBE in which both types choose not to arm holds.

Now suppose $w' \leq \bar{w}$ and $w'' > \underline{w}$. C would set $x = \underline{x}_A$ if T arms and $x = \bar{x}_N$ if T does not arm. In a pooling PBE in which both types arm, incentive compatibility for the weak

type requires $u_T(\text{arm}|x_A = \underline{x}_A) \geq u_T(\neg \text{arm}|x_N = \bar{x}_N)$, or

$$1 - \underline{x}_A - \kappa \geq 1 - \bar{x}_N, \quad (15)$$

which simplifies to $\kappa \leq \bar{p} - \underline{p}$. For the strong type, $u_T(\text{arm}|x_A = \underline{x}_A) \geq u_T(\neg \text{arm}|x_N = \bar{x}_N)$, or

$$1 - \underline{x}_A - \kappa \geq 1 - \tilde{p} - c_2, \quad (16)$$

which simplifies to $\kappa \leq \tilde{p} - \underline{p}$. Thus, as long as $\kappa \leq \tilde{p} - \underline{p}$, incentive compatibility is satisfied for both the weak and strong types and the equilibrium holds, and provided $\kappa > \bar{p} - \underline{p}$, a pooling PBE in which both types choose not to arm holds.

Finally, suppose $w' > \bar{w}$ and $w'' > \underline{w}$. C would thus set $x = \bar{x}_A$ if T arms and $x = \bar{x}_N$ if T does not arm. In a pooling PBE in which both types arm, incentive compatibility for the weak type requires $u_T(\text{arm}|x_A = \bar{x}_A) \geq u_T(\neg \text{arm}|x_N = \bar{x}_N)$, or

$$1 - \bar{x}_A - \kappa \geq 1 - \bar{x}_N, \quad (17)$$

which simplifies to $\kappa \leq \bar{p} - \hat{p}$. For the strong type, $u_T(\text{arm}|x_A = \underline{x}_A) \geq u_T(\neg \text{arm}|x_N = \bar{x}_N)$, or

$$1 - \underline{p} - c_2 - \kappa \geq 1 - \tilde{p} - c_2, \quad (18)$$

which simplifies to $\kappa \leq \tilde{p} - \underline{p}$. Thus, as long as $\kappa \leq \tilde{p} - \underline{p}$ and $\kappa \leq \bar{p} - \hat{p}$, incentive compatibility is satisfied for both the weak and strong types and the equilibrium holds, and provided $\kappa > \tilde{p} - \underline{p}$ and $\kappa > \bar{p} - \hat{p}$, a pooling PBE in which both types choose not to arm holds.

variable equal to 1 if $BOLD = 1$ and $WEAK = 0$.

Baseline Analysis

In the article we only reported the estimated effect for *ARM*. Below we present the results for $PROX_1$ and $PROX_2$. Also, in the article when we restricted the analysis to dyads where $BARG = 1$ we did not include the proxy variables since we already conditioned on the observable variable ($BARG$) whose influence they are meant to represent. Below we report the results for $PROX_1$ and $PROX_2$ when they are included in the models where $BARG = 1$.

Tables 1 and 2 present the results for the proxy variables referred to in footnote 15. Table 1 contains the results for $PROX_1$ for both Experiments 1 and 2 and Table 2 contains the results for $PROX_2$ for both Experiments 1 and 2. Unsurprisingly, $PROX_1$ and $PROX_2$ are positive and significant in the models where we analyze all dyads, that is, when $BARG = 0$. More often than not, these variables are not significant, and we would not expect them to be, when they are included in the models where $BARG = 1$.

In the article we presented the predicted probabilities of *ARM* when $PROX_1 = PROX_2 = 1$, which corresponds to the case where states are most likely to arm. Table 3 presents the results when $PROX_1 = PROX_2 = 0$ referred to in footnote 18. Table 3 contains the predicted probabilities and 95% confidence intervals of *ARM* when $PROX_1 = PROX_2 = 0$, as well as those when $PROX_1 = PROX_2 = 1$. The results when $PROX_1 = PROX_2 = 1$ correspond to the results presented in Table 3 in the article. The estimated effect for *ARM* is the same for both Experiments 1 and 2 because the proxy variables are not included when we only analyze those dyads that engaged in crisis bargaining.

The results when $PROX_1 = PROX_2 = 0$ are similar to the results presented in the article when $PROX_1 = PROX_2 = 1$. When we include all dyads, the change in the size of the effect in each case is larger when the dependent variable is *MID* compared to when it is *WAR*. However, the change in the size of the effect when we condition on $BARG$ is larger when the dependent variable is *WAR* compared to *MID*.

Table 1: Logit Models ($PROX_1$)

	$\hat{\beta}$	Standard deviation	Standard error	Rejection rate $\hat{\beta}$
Experiment 1				
All Observations: <i>MID</i>	1.45	0.02	0.002	1.0
All Observations: <i>WAR</i>	1.78	0.05	0.005	1.0
Only <i>BARG</i> : <i>MID</i>	0.002	0.05	0.005	0.05
Only <i>BARG</i> : <i>WAR</i>	-0.006	0.05	0.005	0.04
Experiment 2				
All Observations: <i>MID</i>	0.87	0.02	0.002	1.0
All Observations: <i>WAR</i>	1.15	0.05	0.005	1.0
Only <i>BARG</i> : <i>MID</i>	-0.004	0.05	0.005	0.07
Only <i>BARG</i> : <i>WAR</i>	0.000	0.05	0.005	0.03

Note: Results based on 100 independent data sets.

Table 2: Logit Models ($PROX_2$)

	$\hat{\beta}$	Standard deviation	Standard error	Rejection rate $\hat{\beta}$
Experiment 1				
All Observations: <i>MID</i>	0.84	0.03	0.003	1.0
All Observations: <i>WAR</i>	1.01	0.05	0.005	1.0
Only <i>BARG</i> : <i>MID</i>	0.003	0.06	0.006	0.05
Only <i>BARG</i> : <i>WAR</i>	-0.002	0.06	0.006	0.05
Experiment 2				
All Observations: <i>MID</i>	0.52	0.03	0.003	1.0
All Observations: <i>WAR</i>	0.68	0.05	0.005	1.0
Only <i>BARG</i> : <i>MID</i>	-0.001	0.06	0.006	0.03
Only <i>BARG</i> : <i>WAR</i>	-0.010	0.05	0.005	0.01

Note: Results based on 100 independent data sets.

Table 3: Logit Models (Predicted Probabilities)

	$ARM = 0$	$ARM = 1$	$ARM = 0$	$ARM = 1$
	$PROX_1 = 0$	$PROX_1 = 0$	$PROX_1 = 1$	$PROX_1 = 1$
	$PROX_2 = 0$	$PROX_2 = 0$	$PROX_2 = 1$	$PROX_2 = 1$
Experiment 1				
All Observations: <i>MID</i>	0.072 (0.070, 0.074)	0.132 (0.128, 0.137)	0.433 (0.419, 0.447)	0.599 (0.585, 0.615)
All Observations: <i>WAR</i>	0.009 (0.008, 0.010)	0.012 (0.011, 0.013)	0.129 (0.119, 0.141)	0.163 (0.147, 0.179)
Only <i>BARG: MID</i>	0.810 (0.800, 0.820)	0.784 (0.771, 0.797)	0.810 (0.800, 0.820)	0.784 (0.771, 0.797)
Only <i>BARG: WAR</i>	0.238 (0.227, 0.249)	0.134 (0.123, 0.144)	0.238 (0.227, 0.249)	0.134 (0.123, 0.144)
Experiment 2				
All Observations: <i>MID</i>	0.084 (0.082, 0.086)	0.160 (0.154, 0.165)	0.269 (0.257, 0.280)	0.432 (0.416, 0.448)
All Observations: <i>WAR</i>	0.012 (0.011, 0.013)	0.018 (0.016, 0.019)	0.069 (0.062, 0.077)	0.099 (0.088, 0.112)
Only <i>BARG: MID</i>	0.810 (0.800, 0.820)	0.784 (0.771, 0.797)	0.810 (0.800, 0.820)	0.784 (0.771, 0.797)
Only <i>BARG: WAR</i>	0.238 (0.227, 0.249)	0.134 (0.123, 0.144)	0.238 (0.227, 0.249)	0.134 (0.123, 0.144)

Note: Results based on 100 independent data sets.

Matching

In the article we focused our discussion on $\hat{\tau}_t$ after matching on $PROX_1 = PROX_2 = 1$, which corresponds to the case where states are most likely to arm. Below we present the results for $\hat{\tau}_c$ after matching on $PROX_1 = PROX_2 = 0$ referred to in footnote 19. When we match on $PROX_1 = PROX_2 = 0$, this corresponds to the case where states are least likely to arm.

In contrast to when we matched on $PROX_1 = PROX_2 = 1$, when we match on $PROX_1 = PROX_2 = 0$, we actually exacerbate the imbalance between observations in the treatment and control regimes, if only modestly. In our simulated data, $BARG$ is 2.8 times as likely to equal 1 in observations where $ARM = 1$ prior to matching, versus 3.03 after matching using the first proxy variable settings and 2.93 using the second proxy variable settings. Prior to matching, $BARG = 1$ in 21% of the observations when $ARM = 1$ and 7% when $ARM = 0$. After matching, the corresponding percentages are 10% and 3%, respectively, in Experiment 1 and 14% and 5%, respectively, in Experiment 2. Thus, matching on $PROX_1 = PROX_2 = 0$ provides a matched subsample with a relatively small percentage of observations where $BARG = 1$.

Table 4 summarizes the results and Table 5 shows the predicted probabilities and 95% confidence intervals when $ARM = 0$ and $ARM = 1$. These results show that $\hat{\tau}_c$ is *always* positive and significant, which should come as no surprise given that we have only exacerbated the problem caused by the presence of an unobservable confounder that correlates with treatment status. Thus, while we were able to identify some conditions under which we could recover the true negative $\hat{\tau}_t$ (see article), the $\hat{\tau}_c$ is always positive.

Additional Ways to do Exact Matching

In the article we focused on $\hat{\tau}_t$ and presented results after matching on $PROX_1 = PROX_2 = 1$. We presented the results after matching on $PROX_1 = PROX_2 = 0$ above. However, there

are four possible ways to do the exact matching: 1) $PROX_1 = PROX_2 = 0$; 2) $PROX_1 = 1, PROX_2 = 0$; 3) $PROX_1 = 0, PROX_2 = 1$; 4) $PROX_1 = PROX_2 = 0$. We chose to focus on $PROX_1 = PROX_2 = 1$ because these are the cases where ARM is most likely to equal 1. However, by doing so, we are not keeping all of the treated cases because some cases are treated (armed), when $PROX_1 = PROX_2 = 0$ and under the other two ways of matching as well and $BARG$ might equal 1 in these cases as well as the cases we matched on. In order to investigate whether our results change if we match on all possible combinations of $PROX_1$ and $PROX_2$, we used the Matching package in R (Sekhon 2011) to do exact matching (1-to-1) on $PROX_1$ and $PROX_2$.

Tables 6 and 7 show the balance statistics for $\hat{\tau}_t$ and $\hat{\tau}_c$, respectively, before and after performing exact matching on $PROX_1$ and $PROX_2$. These tables show the difference in means between the treatment and control groups, the t-test p-value for the difference in means, and the variance ratio.¹ If the variance ratio equals 1 there is perfect balance in the treatment and control groups when matching on $PROX_1$ and $PROX_2$.

When we focus on $\hat{\tau}_t$, exact matching isolates all of the systematic difference between the treatment and control groups. In our simulated data, $BARG$ is 2.8 times as likely to equal 1 in observations where $ARM = 1$ prior to matching, versus 1 after matching in Experiments 1 and 2, respectively, and regardless of whether MID or WAR is the dependent variable. Prior to matching, $BARG = 1$ in 21% of observations when $ARM = 1$ and 7% when $ARM = 0$. After matching, the corresponding percentages are 80% and 80%, respectively, in Experiment 1 and 58% and 58%, respectively, in Experiment 2. These percentages are the same regardless of whether MID or WAR is the dependent variable.² Thus, exact matching provides a matched subsample of observations where the percentage of observations in the treatment and control groups are the same when $BARG = 1$, that is, there is no disparity between the observations where the Target armed versus those where the Target did not.

¹For binary variables the Kolmogorov-Smirnov test is equivalent to the t-test.

²The reason why the balance statistics are the same regardless of whether MID or WAR is the dependent variable is because when we “code” a WAR as 1 we do not assume that a MID has previously occurred.

When we focus on the $\hat{\tau}_c$, exact matching isolates all of the systematic difference between the treatment and control groups. In our simulated data, $BARG$ is 2.8 times as likely to equal 1 in observations where $ARM = 1$ prior to matching, versus 1 after matching in Experiments 1 and 2, respectively, and regardless of whether MID or WAR is the dependent variable. Prior to matching, $BARG = 1$ in 21% of observations when $ARM = 1$ and 7% when $ARM = 0$. After matching, the corresponding percentages are 3% and 3%, respectively, in Experiment 1 and 5% and 5%, respectively, in Experiment 2. These percentages are the same regardless of whether MID or WAR is the dependent variable. Thus, exact matching provides a matched subsample of observations where the percentage of observations in the treatment and control groups are the same when $BARG = 1$, that is, there is no disparity between the observations where the Target armed versus those where the Target did not.

Table 8 summarizes the results and Table 9 shows the predicted probabilities and 95% confidence intervals when $ARM = 0$ and $ARM = 1$ for $\hat{\tau}_t$. These results show that $\hat{\tau}_t$ is *always* estimated to be positive and significant and we almost always reject the null hypothesis. Table 10 summarizes the results and Table 11 shows the predicted probabilities and 95% confidence intervals when $ARM = 0$ and $ARM = 1$ for $\hat{\tau}_c$. These results show that $\hat{\tau}_c$ is *always* estimated to be positive and significant and we *always* reject the null hypothesis.

Figure 1 presents the estimated $\hat{\tau}_t$ for our baseline analysis and each of the three empirical methods discussed in the article, where the matching results are those discussed above. The primary difference between the results shown in Figure 1 and those presented in the article is that the $\hat{\tau}_t$ is always positive when we match on all possible combinations of $PROX_1$ and $PROX_2$, whereas when we focus on matching on $PROX_1 = PROX_2 = 1$, the $\hat{\tau}_t$ is positive except when WAR is the dependent variable and the proxy variables in Experiment 1 are used, in which case it is negative. The $\hat{\tau}_t$ is also smaller than the results presented in the article when MID is the dependent variable.

Table 4: Logit Models after Matching

	$\hat{\beta}$	Standard deviation	Standard error	Rejection rate $\hat{\beta}$
Experiment 1				
<i>MID</i>	0.55	0.03	0.003	1.0
<i>WAR</i>	0.55	0.09	0.009	1.0
Experiment 2				
<i>MID</i>	0.65	0.03	0.003	1.0
<i>WAR</i>	0.52	0.08	0.008	1.0

Notes: Results based on 100 independent data sets. Results reported for *ARM* after matching on $PROX_1 = PROX_2 = 0$.

Table 5: Logit Models after Matching (Predicted Probabilities)

	<i>ARM</i> = 0	<i>ARM</i> = 1
Experiment 1		
<i>MID</i>	0.075 (0.073, 0.077)	0.124 (0.118, 0.129)
<i>WAR</i>	0.008 (0.007, 0.009)	0.014 (0.012, 0.016)
Experiment 2		
<i>MID</i>	0.086 (0.084, 0.089)	0.153 (0.147, 0.159)
<i>WAR</i>	0.011 (0.010, 0.012)	0.019 (0.017, 0.021)

Notes: Results based on 100 independent data sets. Results based on matching on $PROX_1 = PROX_2 = 0$.

Table 6: Balance Statistics: ATT ($\hat{\tau}_t$)

	Before Matching				After Matching			
	Mean Treatment	Mean Control	t-test p-value	Variance Ratio	Mean Treatment	Mean Control	t-test p-value	Variance Ratio
Experiment 1								
<i>MID, PROX₁</i>	0.18	0.13	0.0	1.32	0.18	0.18	1.0	1.0
<i>MID, PROX₂</i>	0.14	0.12	0.0	1.19	0.14	0.14	1.0	1.0
<i>WAR, PROX₁</i>	0.18	0.13	0.0	1.32	0.18	0.18	1.0	1.0
<i>WAR, PROX₂</i>	0.14	0.12	0.0	1.19	0.14	0.14	1.0	1.0
Experiment 2								
<i>MID, PROX₁</i>	0.20	0.17	0.0	1.15	0.20	0.20	1.0	1.0
<i>MID, PROX₂</i>	0.12	0.11	0.0	1.11	0.12	0.12	1.0	1.0
<i>WAR, PROX₁</i>	0.20	0.17	0.0	1.15	0.20	0.20	1.0	1.0
<i>WAR, PROX₂</i>	0.12	0.11	0.0	1.11	0.12	0.12	1.0	1.0

Note: Results based on 100 independent data sets.

Table 7: Balance Statistics: ATC ($\hat{\tau}_c$)

	Before Matching				After Matching			
	Mean Treatment	Mean Control	t-test p-value	Variance Ratio	Mean Treatment	Mean Control	t-test p-value	Variance Ratio
Experiment 1								
<i>MID, PROX₁</i>	0.18	0.13	0.0	1.32	0.13	0.13	1.0	1.0
<i>MID, PROX₂</i>	0.14	0.12	0.0	1.19	0.12	0.12	1.0	1.0
<i>WAR, PROX₁</i>	0.18	0.13	0.0	1.32	0.13	0.13	1.0	1.0
<i>WAR, PROX₂</i>	0.14	0.12	0.0	1.19	0.12	0.12	1.0	1.0
Experiment 2								
<i>MID, PROX₁</i>	0.20	0.17	0.0	1.15	0.17	0.17	1.0	1.0
<i>MID, PROX₂</i>	0.12	0.11	0.0	1.11	0.11	0.11	1.0	1.0
<i>WAR, PROX₁</i>	0.20	0.17	0.0	1.15	0.17	0.17	1.0	1.0
<i>WAR, PROX₂</i>	0.12	0.11	0.0	1.11	0.11	0.11	1.0	1.0

Note: Results based on 100 independent data sets.

Table 8: Logit Models after Matching: ATT ($\hat{\tau}_t$)

	$\hat{\beta}$	Standard deviation	Standard error	Rejection rate $\hat{\beta}$
Experiment 1				
<i>MID</i>	0.62	0.03	0.003	1.0
<i>WAR</i>	0.24	0.07	0.007	0.95
Experiment 2				
<i>MID</i>	0.71	0.03	0.003	1.0
<i>WAR</i>	0.39	0.07	0.007	1.0

Note: Results based on 100 independent data sets.

Table 9: Logit Models after Matching (Predicted Probabilities): ATT ($\hat{\tau}_t$)

	<i>ARM</i> = 0	<i>ARM</i> = 1
Experiment 1		
<i>MID</i>	0.119 (0.115, 0.124)	0.201 (0.195, 0.207)
<i>WAR</i>	0.022 (0.020, 0.024)	0.028 (0.025, 0.030)
Experiment 2		
<i>MID</i>	0.110 (0.106, 0.114)	0.201 (0.195, 0.207)
<i>WAR</i>	0.019 (0.017, 0.021)	0.028 (0.025, 0.030)

Note: Results based on 100 independent data sets.

Table 10: Logit Models after Matching: ATC ($\hat{\tau}_c$)

	$\hat{\beta}$	Standard deviation	Standard error	Rejection rate $\hat{\beta}$
Experiment 1				
<i>MID</i>	0.62	0.02	0.002	1.0
<i>WAR</i>	0.31	0.06	0.006	1.0
Experiment 2				
<i>MID</i>	0.70	0.02	0.002	1.0
<i>WAR</i>	0.41	0.06	0.006	1.0

Note: Results based on 100 independent data sets.

Table 11: Logit Models after Matching (Predicted Probabilities): ATC ($\hat{\tau}_c$)

	<i>ARM</i> = 0	<i>ARM</i> = 1
Experiment 1		
<i>MID</i>	0.106 (0.104, 0.108)	0.180 (0.178, 0.183)
<i>WAR</i>	0.018 (0.017, 0.019)	0.024 (0.023, 0.025)
Experiment 2		
<i>MID</i>	0.106 (0.104, 0.108)	0.194 (0.191, 0.196)
<i>WAR</i>	0.018 (0.017, 0.019)	0.026 (0.025, 0.027)

Note: Results based on 100 independent data sets.

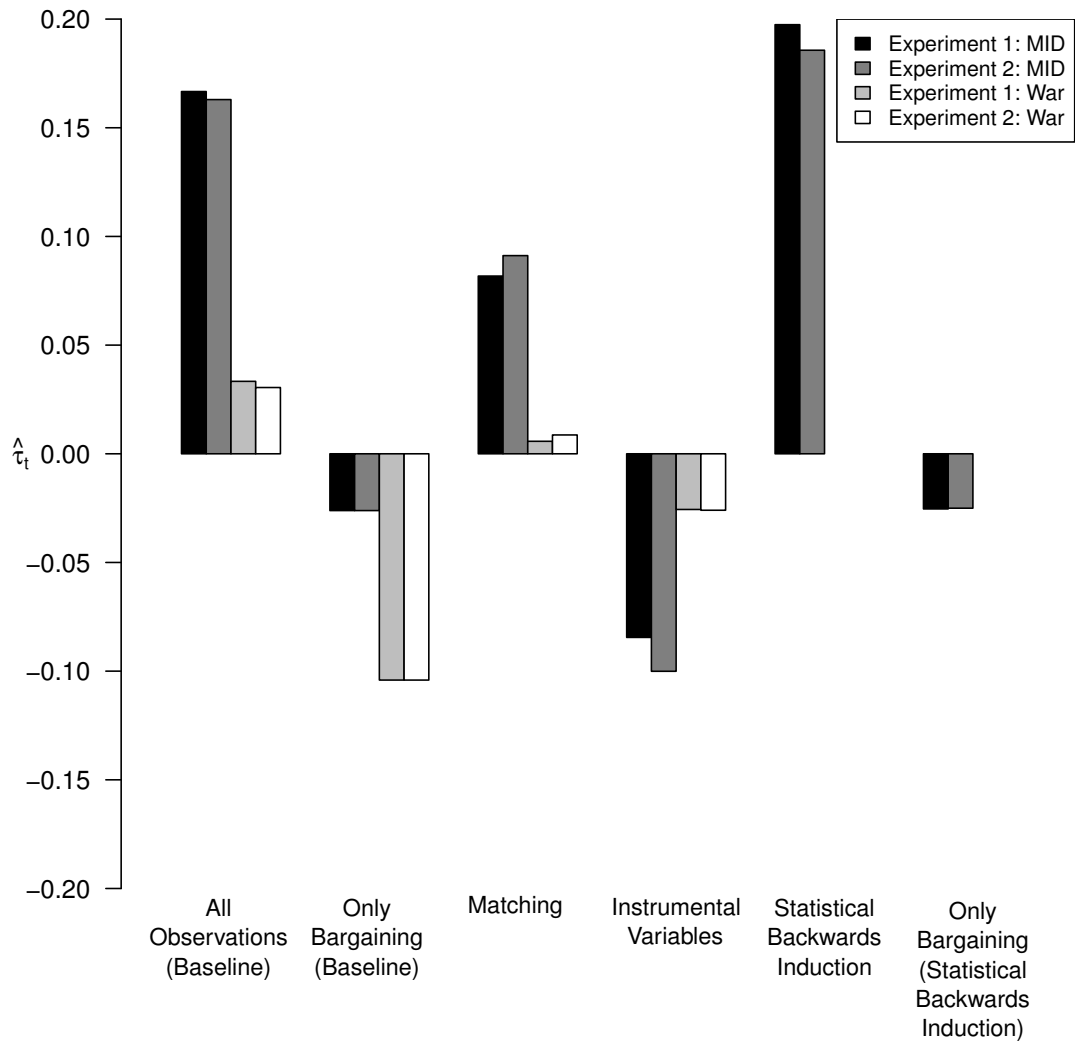


Figure 1: Comparison of Baseline Models and Three Empirical Methods

Instrumental Variables

In the article we presented the results for $\hat{\tau}_t$ holding IV_1 , $PROX_1$, and $PROX_2$ all constant at 1, which maximizes the probability of treatment. Table 12 shows the predicted probabilities and 95% confidence intervals when $IV_1 = 1$ and $PROX_1 = PROX_2 = 0$ (referred to in footnote 29) as well as the results when $IV_1 = 1$ and $PROX_1 = PROX_2 = 1$. The last two columns of Table 12 correspond to the results presented in the article.

In the article we presented the results for IV_1 since the results for IV_2 are very similar. Table 13 shows the predicted probabilities and 95% confidence intervals when $IV_2 = 1$ and $PROX_1 = PROX_2 = 0$ as well as the results when $IV_2 = 1$ and $PROX_1 = PROX_2 = 1$ referred to in footnote 31.

Table 12: Instrumental Variables Model (Predicted Probabilities, IV_1)

	$ARM = 0$	$ARM = 1$	$ARM = 0$	$ARM = 1$
	$PROX_1 = 0$	$PROX_1 = 0$	$PROX_1 = 1$	$PROX_1 = 1$
	$PROX_2 = 0$	$PROX_2 = 0$	$PROX_2 = 1$	$PROX_2 = 1$
Experiment 1				
All Observations: <i>MID</i>				
$Y_1=0, Y_2=0$	0.684 (0.680, 0.689)	0.676 (0.671, 0.681)	0.357 (0.347, 0.368)	0.338 (0.327, 0.349)
$Y_1=0, Y_2=1$	0.050 (0.048, 0.052)	0.059 (0.056, 0.061)	0.250 (0.240, 0.259)	0.269 (0.259, 0.280)
$Y_1=1, Y_2=0$	0.236 (0.231, 0.240)	0.231 (0.227, 0.236)	0.189 (0.181, 0.198)	0.177 (0.168, 0.185)
$Y_1=1, Y_2=1$	0.030 (0.029, 0.031)	0.034 (0.033, 0.036)	0.204 (0.195, 0.212)	0.216 (0.207, 0.226)
All Observations: <i>WAR</i>				
$Y_1=0, Y_2=0$	0.731 (0.726, 0.735)	0.735 (0.730, 0.740)	0.546 (0.534, 0.559)	0.597 (0.585, 0.610)
$Y_1=0, Y_2=1$	0.004 (0.004, 0.005)	0.000 (0.000, 0.000)	0.060 (0.055, 0.066)	0.009 (0.007, 0.011)
$Y_1=1, Y_2=0$	0.246 (0.246, 0.251)	0.262 (0.258, 0.267)	0.238 (0.227, 0.249)	0.344 (0.331, 0.356)
$Y_1=1, Y_2=1$	0.018 (0.018, 0.019)	0.002 (0.002, 0.003)	0.156 (0.147, 0.165)	0.050 (0.044, 0.056)
Experiment 2				
All Observations: <i>MID</i>				
$Y_1=0, Y_2=0$	0.669 (0.665, 0.674)	0.650 (0.645, 0.656)	0.489 (0.476, 0.499)	0.453 (0.440, 0.465)
$Y_1=0, Y_2=1$	0.059 (0.057, 0.060)	0.078 (0.075, 0.081)	0.170 (0.162, 0.178)	0.205 (0.196, 0.215)
$Y_1=1, Y_2=0$	0.239 (0.234, 0.243)	0.229 (0.225, 0.234)	0.226 (0.217, 0.236)	0.206 (0.197, 0.215)

$Y_1=1, Y_2=1$	0.033 (0.032, 0.034)	0.043 (0.041, 0.045)	0.116 (0.110, 0.122)	0.136 (0.129, 0.143)
All Observations: <i>WAR</i>				
$Y_1=0, Y_2=0$	0.723 (0.718, 0.727)	0.728 (0.724, 0.733)	0.622 (0.610, 0.635)	0.652 (0.640, 0.664)
$Y_1=0, Y_2=1$	0.006 (0.006, 0.007)	0.001 (0.000, 0.001)	0.035 (0.031, 0.039)	0.006 (0.004, 0.007)
$Y_1=1, Y_2=0$	0.250 (0.246, 0.255)	0.267 (0.263, 0.272)	0.255 (0.244, 0.266)	0.316 (0.304, 0.328)
$Y_1=1, Y_2=1$	0.021 (0.020, 0.022)	0.004 (0.003, 0.005)	0.087 (0.081, 0.094)	0.026 (0.023, 0.030)

Notes: Results based on 100 independent data sets. Results based on $IV_1 = 1$.

Table 13: Instrumental Variables Model (Predicted Probabilities, IV_2)

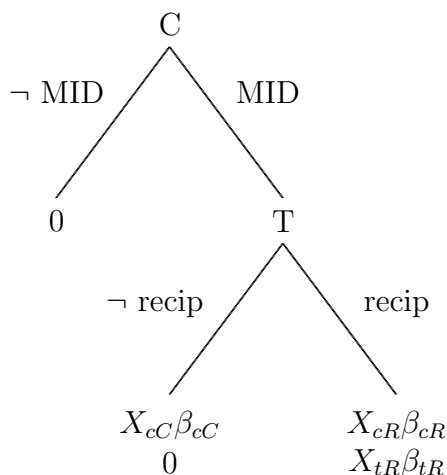
	$ARM = 0$	$ARM = 1$	$ARM = 0$	$ARM = 1$
	$PROX_1 = 0$	$PROX_1 = 0$	$PROX_1 = 1$	$PROX_1 = 1$
	$PROX_2 = 0$	$PROX_2 = 0$	$PROX_2 = 1$	$PROX_2 = 1$
Experiment 1				
All Observations: <i>MID</i>				
$Y_1=0, Y_2=0$	0.685 (0.680, 0.689)	0.679 (0.674, 0.684)	0.358 (0.347, 0.369)	0.344 (0.333, 0.356)
$Y_1=0, Y_2=1$	0.050 (0.048, 0.051)	0.055 (0.053, 0.058)	0.248 (0.239, 0.258)	0.262 (0.251, 0.272)
$Y_1=1, Y_2=0$	0.234 (0.230, 0.238)	0.231 (0.227, 0.235)	0.186 (0.178, 0.194)	0.177 (0.169, 0.186)
$Y_1=1, Y_2=1$	0.031 (0.030, 0.032)	0.034 (0.033, 0.036)	0.208 (0.200, 0.217)	0.217 (0.208, 0.226)
All Observations: <i>WAR</i>				
$Y_1=0, Y_2=0$	0.731 (0.727, 0.735)	0.735 (0.731, 0.739)	0.546 (0.534, 0.558)	0.596 (0.584, 0.609)
$Y_1=0, Y_2=1$	0.004 (0.004, 0.005)	0.000 (0.000, 0.000)	0.059 (0.054, 0.065)	0.009 (0.007, 0.011)
$Y_1=1, Y_2=0$	0.246 (0.242, 0.251)	0.262 (0.258, 0.266)	0.240 (0.229, 0.250)	0.344 (0.332, 0.357)
$Y_1=1, Y_2=1$	0.018 (0.017, 0.019)	0.002 (0.002, 0.003)	0.155 (0.146, 0.164)	0.050 (0.044, 0.057)
Experiment 2				
All Observations: <i>MID</i>				
$Y_1=0, Y_2=0$	0.670 (0.666, 0.674)	0.656 (0.651, 0.661)	0.488 (0.477, 0.500)	0.462 (0.449, 0.474)
$Y_1=0, Y_2=1$	0.058 (0.056, 0.060)	0.072 (0.069, 0.076)	0.169 (0.161, 0.177)	0.196 (0.186, 0.205)
$Y_1=1, Y_2=0$	0.237 (0.233, 0.241)	0.229 (0.225, 0.234)	0.222 (0.213, 0.232)	0.207 (0.198, 0.216)

$Y_1=1, Y_2=1$	0.035 (0.034, 0.036)	0.042 (0.040, 0.044)	0.120 (0.114, 0.127)	0.136 (0.129, 0.143)
All Observations: <i>WAR</i>				
$Y_1=0, Y_2=0$	0.723 (0.719, 0.727)	0.728 (0.724, 0.733)	0.623 (0.610, 0.635)	0.652 (0.640, 0.664)
$Y_1=0, Y_2=1$	0.006 (0.006, 0.007)	0.001 (0.000, 0.001)	0.035 (0.031, 0.039)	0.005 (0.004, 0.007)
$Y_1=1, Y_2=0$	0.250 (0.246, 0.254)	0.267 (0.263, 0.271)	0.255 (0.244, 0.266)	0.316 (0.304, 0.328)
$Y_1=1, Y_2=1$	0.021 (0.020, 0.022)	0.004 (0.003, 0.005)	0.088 (0.081, 0.094)	0.026 (0.023, 0.031)

Notes: Results based on 100 independent data sets. Results based on $IV_2 = 1$.

Statistical Backwards Induction

Below is the extensive form we used to estimate the Statistical Backwards Induction models.



In the article we focused on comparing the difference between the predicted probability the Challenger initiates a *MID* (\hat{p}_c) when $ARM = 0$ and $ARM = 1$ holding $PROX_1 = PROX_2 = 1$. Table 14 shows the predicted probabilities of *MID* and 95% confidence intervals for different settings of ARM , $PROX_1$, and $PROX_2$ for Experiments 1 and 2, respectively, when CAP_1 is held at its mean value (0.5), including the results where $PROX_1 = PROX_2 = 0$ referred to in footnote 43. The results when $PROX_1 = PROX_2 = 1$ correspond to the results presented in Table 3 in the article.

Tables 15 and 16 show the predicted probabilities of *MID* and 95% confidence intervals for different settings of ARM , $PROX_1$, and $PROX_2$ for Experiments 1 and 2, respectively, when CAP_1 is held at its minimum value (0.28) and when CAP_1 is held at its maximum value (0.74) referred to in footnote 44. In Tables 14, 15, and 16 the predicted probabilities are the same in the models were $BARG = 1$ because the proxy variables are not included in these models.

Table 14: Statistical Backwards Induction Models (Predicted Probabilities)

	$ARM = 0$	$ARM = 1$	$ARM = 0$	$ARM = 1$
	$PROX_1 = 0$	$PROX_1 = 0$	$PROX_1 = 1$	$PROX_1 = 1$
	$PROX_2 = 0$	$PROX_2 = 0$	$PROX_2 = 1$	$PROX_2 = 1$
Experiment 1				
All Observations: <i>MID</i>	0.074 (0.073, 0.074)	0.128 (0.127, 0.130)	0.423 (0.419, 0.427)	0.620 (0.616, 0.625)
Only <i>BARG</i> : <i>MID</i>	0.810 (0.809, 0.811)	0.784 (0.783, 0.786)	0.810 (0.809, 0.811)	0.784 (0.783, 0.786)
Experiment 2				
All Observations: <i>MID</i>	0.085 (0.085, 0.086)	0.157 (0.156, 0.159)	0.261 (0.257, 0.265)	0.447 (0.441, 0.452)
Only <i>BARG</i> : <i>MID</i>	0.810 (0.809, 0.811)	0.785 (0.783, 0.786)	0.810 (0.809, 0.811)	0.785 (0.783, 0.786)

Notes: Results based on 100 independent data sets. Results based on CAP_1 held at mean value of 0.5.

Table 15: Statistical Backwards Induction Models (Predicted Probabilities)

	$ARM = 0$	$ARM = 1$	$ARM = 0$	$ARM = 1$
	$PROX_1 = 0$	$PROX_1 = 0$	$PROX_1 = 1$	$PROX_1 = 1$
	$PROX_2 = 0$	$PROX_2 = 0$	$PROX_2 = 1$	$PROX_2 = 1$
Experiment 1				
All Observations: <i>MID</i>	0.077	0.134	0.444	0.642
	(0.076, 0.078)	(0.132, 0.136)	(0.439, 0.450)	(0.636, 0.648)
Only <i>BARG</i> : <i>MID</i>	0.815	0.789	0.815	0.789
	(0.812, 0.818)	(0.786, 0.793)	(0.812, 0.818)	(0.786, 0.793)
Experiment 2				
All Observations: <i>MID</i>	0.087	0.160	0.267	0.456
	(0.086, 0.088)	(0.158, 0.162)	(0.263, 0.272)	(0.449, 0.462)
Only <i>BARG</i> : <i>MID</i>	0.815	0.790	0.815	0.790
	(0.811, 0.818)	(0.786, 0.794)	(0.811, 0.818)	(0.786, 0.794)

Notes: Results based on 100 independent data sets. Results based on CAP_1 held at minimum value of 0.28.

Table 16: Statistical Backwards Induction Models (Predicted Probabilities)

	$ARM = 0$	$ARM = 1$	$ARM = 0$	$ARM = 1$
	$PROX_1 = 0$	$PROX_1 = 0$	$PROX_1 = 1$	$PROX_1 = 1$
	$PROX_2 = 0$	$PROX_2 = 0$	$PROX_2 = 1$	$PROX_2 = 1$
Experiment 1				
All Observations: <i>MID</i>	0.071 (0.070, 0.072)	0.123 (0.121, 0.125)	0.402 (0.397, 0.407)	0.598 (0.592, 0.603)
Only <i>BARG</i> : <i>MID</i>	0.803 (0.800, 0.807)	0.778 (0.775, 0.782)	0.803 (0.800, 0.807)	0.778 (0.775, 0.782)
Experiment 2				
All Observations: <i>MID</i>	0.084 (0.083, 0.085)	0.154 (0.152, 0.156)	0.255 (0.251, 0.259)	0.438 (0.431, 0.444)
Only <i>BARG</i> : <i>MID</i>	0.803 (0.799, 0.807)	0.779 (0.775, 0.782)	0.803 (0.799, 0.807)	0.779 (0.775, 0.782)

Notes: Results based on 100 independent data sets. Results based on CAP_1 held at maximum value of 0.74.

References

Sekhon, Jasjeet S. 2011. ‘Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching package for R’. *Journal of Statistical Software* 42(7):1–52.