

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

PHILIP ARENA

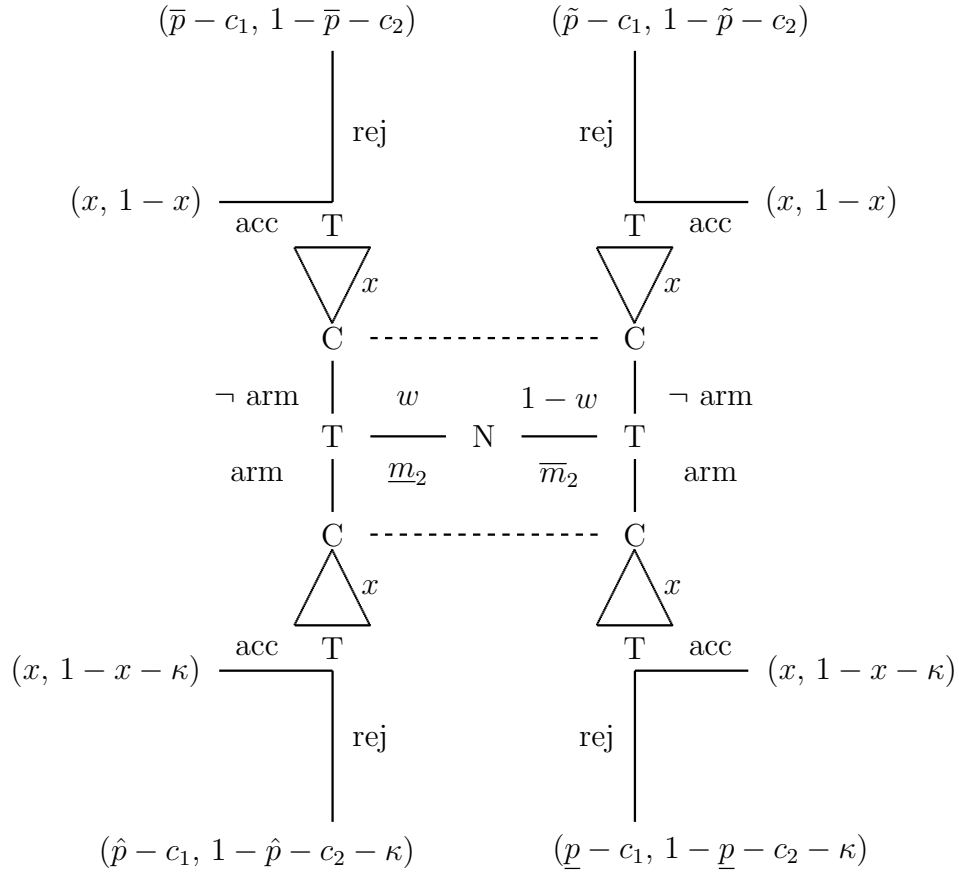
Department of Political Science, University at Buffalo (SUNY)

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.



Baseline Analysis

In the paper we only reported the estimated effect for *ARM*. Below we present the results for $PROX_1$ and $PROX_2$. Also, in the paper when we restricted the analysis to dyads where $BARG = 1$ we did not include the proxy variables since we had 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 25. Table 1 contains the results for $PROX_1$ for both settings (Experiment 1 and Experiment 2, respectively) and Table 2 contains the results for $PROX_2$ for both settings (Experiment 1 and Experiment 2, respectively). 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 paper 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 27. 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 2 in the paper. The estimated effect for *ARM* is the same for both Experiment 1 and Experiment 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 paper 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: Baseline Logit Models ($PROX_1$)

	$\hat{\beta}$	Standard deviation	Standard error	Rejection rate $\hat{\beta}$
Experiment 1				
All Observations, DV: <i>MID</i>	1.44	0.02	0.002	1.0
All Observations, DV: <i>WAR</i>	1.78	0.05	0.005	1.0
Only <i>BARG</i> , DV: <i>MID</i>	-0.004	0.05	0.005	0.07
Only <i>BARG</i> , DV: <i>WAR</i>	-0.007	0.05	0.005	0.04
Experiment 2				
All Observations, DV: <i>MID</i>	0.87	0.02	0.002	1.0
All Observations, DV: <i>WAR</i>	1.15	0.05	0.005	1.0
Only <i>BARG</i> , DV: <i>MID</i>	0.0004	0.05	0.005	0.06
Only <i>BARG</i> , DV: <i>WAR</i>	-0.001	0.05	0.005	0.05

Note: Results based on estimation of 100 independent data sets.

Table 2: Baseline Logit Models ($PROX_2$)

	$\hat{\beta}$	Standard deviation	Standard error	Rejection rate $\hat{\beta}$
Experiment 1				
All Observations, DV: <i>MID</i>	0.84	0.02	0.002	1.0
All Observations, DV: <i>WAR</i>	1.01	0.05	0.005	1.0
Only <i>BARG</i> , DV: <i>MID</i>	-0.003	0.06	0.006	0.05
Only <i>BARG</i> , DV: <i>WAR</i>	-0.002	0.05	0.005	0.04
Experiment 2				
All Observations, DV: <i>MID</i>	0.52	0.03	0.003	1.0
All Observations, DV: <i>WAR</i>	0.68	0.06	0.006	1.0
Only <i>BARG</i> , DV: <i>MID</i>	-0.002	0.06	0.006	0.03
Only <i>BARG</i> , DV: <i>WAR</i>	-0.001	0.06	0.006	0.03

Note: Results based on estimation of 100 independent data sets.

Table 3: Baseline Logit Models (Predicted Probabilities)

	<i>ARM</i> = 0	<i>ARM</i> = 1	<i>ARM</i> = 0	<i>ARM</i> = 1
	<i>PROX</i> ₁ = 0	<i>PROX</i> ₁ = 0	<i>PROX</i> ₁ = 1	<i>PROX</i> ₁ = 1
	<i>PROX</i> ₂ = 0	<i>PROX</i> ₂ = 0	<i>PROX</i> ₂ = 1	<i>PROX</i> ₂ = 1
Experiment 1				
All Observations, DV: <i>MID</i>	0.0721 (0.0703, 0.0740)	0.1328 (0.1282, 0.1374)	0.4309 (0.4174, 0.4445)	0.5985 (0.5837, 0.6133)
All Observations, DV: <i>WAR</i>	0.0090 (0.0083, 0.0096)	0.0118 (0.0105, 0.0130)	0.1286 (0.1172, 0.1400)	0.1623 (0.1465, 0.1781)
Only <i>BARG</i> , DV: <i>MID</i>	0.8087 (0.7987, 0.8187)	0.7830 (0.7704, 0.7957)	0.8087 (0.7787, 0.8187)	0.7830 (0.7704, 0.7957)
Only <i>BARG</i> , DV: <i>WAR</i>	0.2362 (0.2254, 0.2470)	0.1331 (0.1227, 0.1435)	0.2362 (0.2254, 0.2470)	0.1331 (0.1227, 0.1435)
Experiment 2				
All Observations, DV: <i>MID</i>	0.0840 (0.0819, 0.0860)	0.1599 (0.1548, 0.1650)	0.2674 (0.2558, 0.2790)	0.4310 (0.4151, 0.4469)
All Observations, DV: <i>WAR</i>	0.0117 (0.0109, 0.0125)	0.0175 (0.0158, 0.0191)	0.0688 (0.0611, 0.0765)	0.0997 (0.0876, 0.1117)
Only <i>BARG</i> , DV: <i>MID</i>	0.8087 (0.7987, 0.8187)	0.7830 (0.7704, 0.7957)	0.8087 (0.7787, 0.8187)	0.7830 (0.7704, 0.7957)
Only <i>BARG</i> , DV: <i>WAR</i>	0.2362 (0.2254, 0.2470)	0.1331 (0.1227, 0.1435)	0.2362 (0.2254, 0.2470)	0.1331 (0.1227, 0.1435)

Note: Results based on estimation of 100 independent data sets.

Matching

In the paper 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 28. 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 3.02 after matching using the first proxy variable settings and 2.95 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* estimated to be 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 paper), the $\hat{\tau}_c$ is always positive.

Additional Ways to do Exact Matching

In the paper 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 in the cases that 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 Experiment 1 and Experiment 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 Experiment 1 and Experiment 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 paper, where the matching results are those discussed above. The primary difference between the results shown in Figure 1 and those presented in the paper 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 paper when MID is the dependent variable.

Table 4: Logit Models after Matching

	$\hat{\beta}$	Standard deviation	Standard error	Rejection rate $\hat{\beta}$
Experiment 1				
DV: <i>MID</i>	0.55	0.03	0.003	1.0
DV: <i>WAR</i>	0.53	0.09	0.009	1.0
Experiment 2				
DV: <i>MID</i>	0.65	0.03	0.003	1.0
DV: <i>WAR</i>	0.52	0.07	0.007	1.0

Notes: Results based on estimation of 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		
DV: <i>MID</i>	0.0754 (0.0733, 0.0775)	0.1240 (0.1185, 0.1294)
DV: <i>WAR</i>	0.0079 (0.0072, 0.0086)	0.0134 (0.0115, 0.0153)
Experiment 2		
DV: <i>MID</i>	0.0862 (0.0840, 0.0885)	0.1530 (0.1471, 0.1590)
DV: <i>WAR</i>	0.0112 (0.0104, 0.0121)	0.0187 (0.0165, 0.0210)

Notes: Results based on estimation of 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								
DV: <i>MID, PROX</i> ₁	0.18	0.13	0.0	1.32	0.18	0.18	1.0	1.0
DV: <i>MID, PROX</i> ₂	0.14	0.12	0.0	1.19	0.14	0.14	1.0	1.0
DV: <i>WAR, PROX</i> ₁	0.18	0.13	0.0	1.32	0.18	0.18	1.0	1.0
DV: <i>WAR, PROX</i> ₂	0.14	0.12	0.0	1.19	0.14	0.14	1.0	1.0
Experiment 2								
DV: <i>MID, PROX</i> ₁	0.20	0.17	0.0	1.15	0.20	0.20	1.0	1.0
DV: <i>MID, PROX</i> ₂	0.12	0.11	0.0	1.10	0.12	0.12	1.0	1.0
DV: <i>WAR, PROX</i> ₁	0.20	0.17	0.0	1.15	0.20	0.20	1.0	1.0
DV: <i>WAR, PROX</i> ₂	0.12	0.11	0.0	1.10	0.12	0.12	1.0	1.0

Note: Results based on estimation of 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								
DV: <i>MID, PROX</i> ₁	0.18	0.13	0.0	1.32	0.13	0.13	1.0	1.0
DV: <i>MID, PROX</i> ₂	0.14	0.12	0.0	1.19	0.12	0.12	1.0	1.0
DV: <i>WAR, PROX</i> ₁	0.18	0.13	0.0	1.32	0.13	0.13	1.0	1.0
DV: <i>WAR, PROX</i> ₂	0.14	0.12	0.0	1.19	0.12	0.12	1.0	1.0
Experiment 2								
DV: <i>MID, PROX</i> ₁	0.20	0.17	0.0	1.15	0.17	0.17	1.0	1.0
DV: <i>MID, PROX</i> ₂	0.12	0.11	0.0	1.10	0.11	0.11	1.0	1.0
DV: <i>WAR, PROX</i> ₁	0.20	0.17	0.0	1.15	0.17	0.17	1.0	1.0
DV: <i>WAR, PROX</i> ₂	0.12	0.11	0.0	1.10	0.11	0.11	1.0	1.0

Note: Results based on estimation of 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				
DV: <i>MID</i>	0.62	0.03	0.003	1.0
DV: <i>WAR</i>	0.26	0.07	0.007	0.97
Experiment 2				
DV: <i>MID</i>	0.72	0.04	0.004	1.0
DV: <i>WAR</i>	0.39	0.07	0.007	1.0

Note: Results based on estimation of 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		
DV: <i>MID</i>	0.1193 (0.1148, 0.1239)	0.2013 (0.1957, 0.2069)
DV: <i>WAR</i>	0.0214 (0.0194, 0.0234)	0.0275 (0.0252, 0.0297)
Experiment 2		
DV: <i>MID</i>	0.1095 (0.1051, 0.1138)	0.2013 (0.1957, 0.2069)
DV: <i>WAR</i>	0.0187 (0.0168, 0.0206)	0.0275 (0.0252, 0.0297)

Note: Results based on estimation of 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				
DV: <i>MID</i>	0.62	0.03	0.003	1.0
DV: <i>WAR</i>	0.31	0.06	0.006	1.0
Experiment 2				
DV: <i>MID</i>	0.71	0.03	0.003	1.0
DV: <i>WAR</i>	0.41	0.05	0.005	1.0

Note: Results based on estimation of 100 independent data sets.

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

	$ARM = 0$	$ARM = 1$
Experiment 1		
DV: <i>MID</i>	0.1061 (0.1040, 0.1082)	0.1804 (0.1778, 0.1831)
DV: <i>WAR</i>	0.0174 (0.0165, 0.0183)	0.0237 (0.0226, 0.0247)
Experiment 2		
DV: <i>MID</i>	0.1061 (0.1040, 0.1082)	0.1939 (0.1912, 0.1966)
DV: <i>WAR</i>	0.0174 (0.0165, 0.0183)	0.0262 (0.0251, 0.0273)

Note: Results based on estimation of 100 independent data sets.

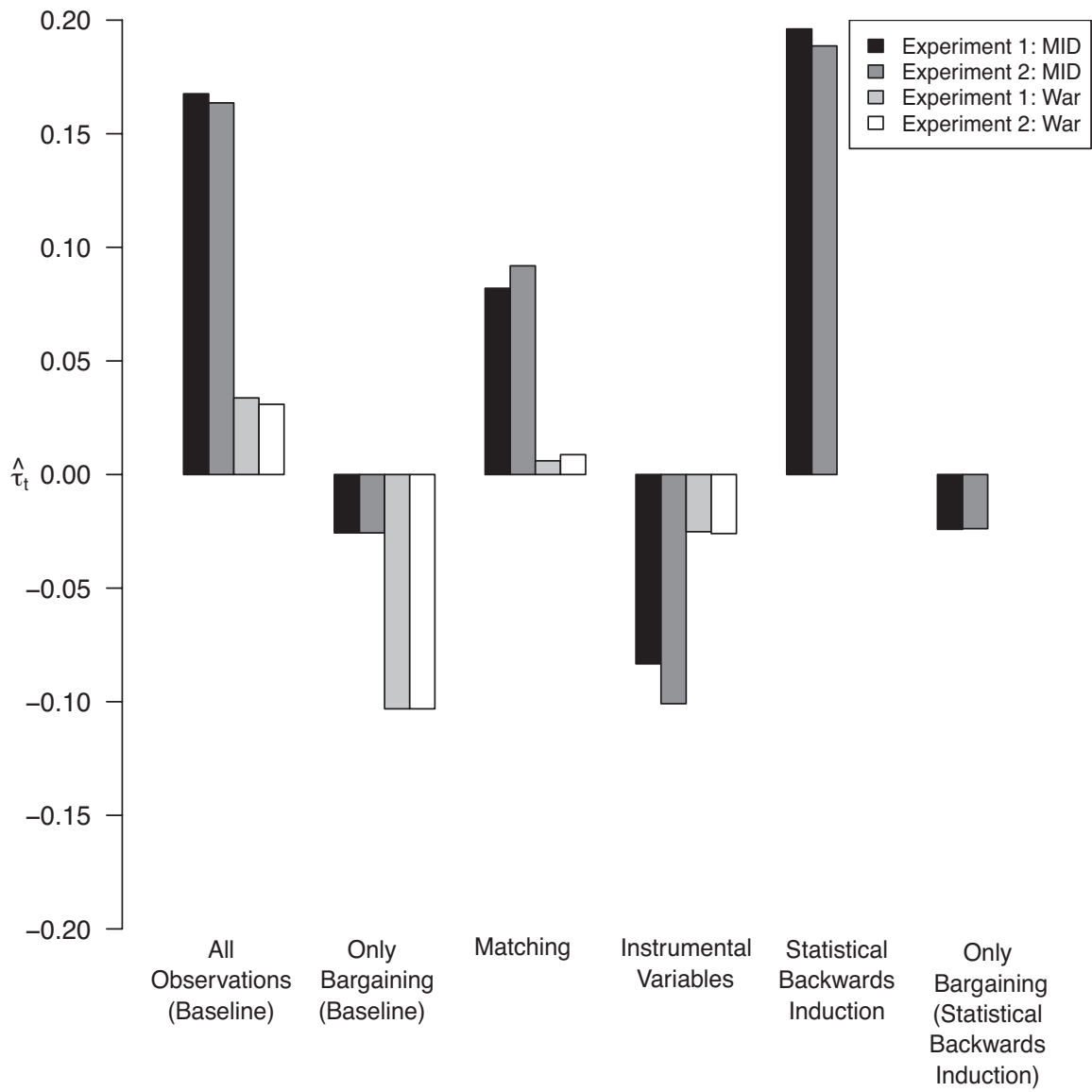


Figure 1: Comparison of Baseline Models and Three Empirical Methods

Instrumental Variables

In the paper 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 39) 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 paper.

In the paper 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 41.

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, DV: <i>MID</i>				
$Y_1=0, Y_2=0$	0.6841 (0.6795, 0.6886)	0.6756 (0.6706, 0.6807)	0.3591 (0.3482, 0.3700)	0.3396 (0.3282, 0.3511)
$Y_1=0, Y_2=1$	0.0500 (0.0485, 0.0516)	0.0585 (0.0558, 0.0613)	0.2484 (0.2389, 0.2580)	0.2678 (0.2573, 0.2785)
$Y_1=1, Y_2=0$	0.2358 (0.2315, 0.2402)	0.2314 (0.2270, 0.2358)	0.1897 (0.1814, 0.1981)	0.1769 (0.1685, 0.1855)
$Y_1=1, Y_2=1$	0.0301 (0.0289, 0.0313)	0.0345 (0.0328, 0.0362)	0.2029 (0.1945, 0.2114)	0.2157 (0.2065, 0.2249)
All Observations, DV: <i>WAR</i>				
$Y_1=0, Y_2=0$	0.7308 (0.7262, 0.7354)	0.7349 (0.7302, 0.7395)	0.5475 (0.5351, 0.5600)	0.5982 (0.5856, 0.6107)
$Y_1=0, Y_2=1$	0.0043 (0.0039, 0.0047)	0.0002 (0.0002, 0.0003)	0.0594 (0.0540, 0.0650)	0.0086 (0.0069, 0.0105)
$Y_1=1, Y_2=0$	0.2465 (0.2420, 0.2511)	0.2624 (0.2578, 0.2671)	0.2377 (0.2269, 0.2487)	0.3437 (0.3312, 0.3564)
$Y_1=1, Y_2=1$	0.0184 (0.0175, 0.0194)	0.0025 (0.0021, 0.0030)	0.1554 (0.1467, 0.1644)	0.0495 (0.0435, 0.0559)
Experiment 2				
All Observations, DV: <i>MID</i>				
$Y_1=0, Y_2=0$	0.6693 (0.6647, 0.6740)	0.6502 (0.6450, 0.6555)	0.4893 (0.4777, 0.5009)	0.4543 (0.4420, 0.4665)
$Y_1=0, Y_2=1$	0.0585 (0.0568, 0.0603)	0.0776 (0.0744, 0.0809)	0.1697 (0.1618, 0.1778)	0.2048 (0.1950, 0.2146)
$Y_1=1, Y_2=0$	0.2390 (0.2346, 0.2434)	0.2295 (0.2251, 0.2340)	0.2258 (0.2163, 0.2353)	0.2056 (0.1965, 0.2149)

$Y_1=1, Y_2=1$	0.0332 (0.0320, 0.0345)	0.0426 (0.0407, 0.0446)	0.1152 (0.1092, 0.1213)	0.1353 (0.1283, 0.1426)
All Observations, DV: WAR				
$Y_1=0, Y_2=0$	0.7225 (0.7178, 0.7272)	0.7280 (0.7233, 0.7327)	0.6238 (0.6116, 0.6360)	0.6533 (0.6410, 0.6655)
$Y_1=0, Y_2=1$	0.0061 (0.0056, 0.0066)	0.0006 (0.0004, 0.0007)	0.0350 (0.0311, 0.0391)	0.0055 (0.0044, 0.0069)
$Y_1=1, Y_2=0$	0.2507 (0.2461, 0.2553)	0.2676 (0.2629, 0.2723)	0.2544 (0.2433, 0.2655)	0.3150 (0.3030, 0.3271)
$Y_1=1, Y_2=1$	0.0207 (0.0196, 0.0217)	0.0038 (0.0032, 0.0045)	0.0868 (0.0804, 0.0933)	0.0262 (0.0224, 0.0302)

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

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

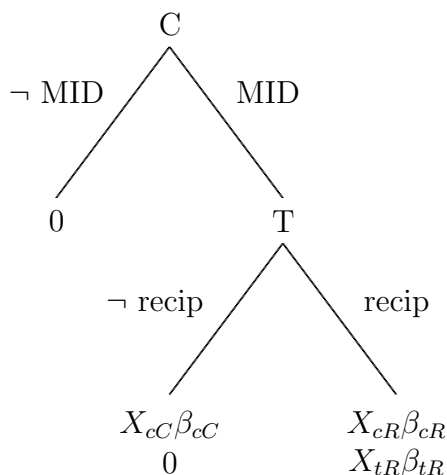
	$ARM = 0$	$ARM = 1$	$ARM = 0$	$ARM = 1$
	$PROX_1 = 0$	$PROX_1 = 0$	$PROX_1 = 1$	$PROX_1 = 1$
	$PROX_2 = 0$	$PROX_2 = 0$	$PROX_2 = 1$	$PROX_2 = 1$
Experiment 1				
All Observations, DV: <i>MID</i>				
$Y_1=0, Y_2=0$	0.6846 (0.6804, 0.6888)	0.6787 (0.6739, 0.6834)	0.3593 (0.3486, 0.3701)	0.3455 (0.3342, 0.3569)
$Y_1=0, Y_2=1$	0.0497 (0.0482, 0.0513)	0.0557 (0.0530, 0.0584)	0.2470 (0.2375, 0.2565)	0.2609 (0.2505, 0.2713)
$Y_1=1, Y_2=0$	0.2344 (0.2304, 0.2384)	0.2312 (0.2271, 0.2353)	0.1868 (0.1786, 0.1952)	0.1776 (0.1693, 0.1861)
$Y_1=1, Y_2=1$	0.0313 (0.0301, 0.0325)	0.0344 (0.0328, 0.0361)	0.2069 (0.1985, 0.2155)	0.2161 (0.2071, 0.2253)
All Observations, DV: <i>WAR</i>				
$Y_1=0, Y_2=0$	0.7311 (0.7269, 0.7354)	0.7351 (0.7308, 0.7394)	0.5470 (0.5346, 0.5592)	0.5969 (0.5893, 0.6094)
$Y_1=0, Y_2=1$	0.0042 (0.0038, 0.0046)	0.0002 (0.0002, 0.0003)	0.0587 (0.0533, 0.0643)	0.0088 (0.0071, 0.0107)
$Y_1=1, Y_2=0$	0.2466 (0.2424, 0.2508)	0.2621 (0.2579, 0.2665)	0.2403 (0.2294, 0.2513)	0.3447 (0.3321, 0.3573)
$Y_1=1, Y_2=1$	0.0180 (0.0171, 0.0189)	0.0025 (0.0021, 0.0030)	0.1540 (0.1453, 0.1629)	0.0497 (0.0436, 0.0560)
Experiment 2				
All Observations, DV: <i>MID</i>				
$Y_1=0, Y_2=0$	0.6699 (0.6656, 0.6743)	0.6557 (0.6507, 0.6607)	0.4899 (0.4783, 0.5016)	0.4635 (0.4513, 0.4758)
$Y_1=0, Y_2=1$	0.0581 (0.0564, 0.0598)	0.0723 (0.0692, 0.0754)	0.1687 (0.1607, 0.1767)	0.1951 (0.1857, 0.2047)
$Y_1=1, Y_2=0$	0.2369 (0.2328, 0.2409)	0.2296 (0.2254, 0.2338)	0.2218 (0.2126, 0.2311)	0.2063 (0.1972, 0.2157)

$Y_1=1, Y_2=1$	0.0351 (0.0339, 0.0364)	0.0424 (0.0405, 0.0444)	0.1196 (0.1135, 0.1259)	0.1351 (0.1281, 0.1423)
All Observations, DV: WAR				
$Y_1=0, Y_2=0$	0.7229 (0.7186, 0.7272)	0.7283 (0.7240, 0.7327)	0.6240 (0.6118, 0.6360)	0.6530 (0.6409, 0.6651)
$Y_1=0, Y_2=1$	0.0060 (0.0054, 0.0065)	0.0005 (0.0004, 0.0007)	0.0344 (0.0306, 0.0385)	0.0054 (0.0043, 0.0067)
$Y_1=1, Y_2=0$	0.2503 (0.2461, 0.2546)	0.2673 (0.2629, 0.2717)	0.2544 (0.2435, 0.2655)	0.3153 (0.3034, 0.3273)
$Y_1=1, Y_2=1$	0.0208 (0.0198, 0.0218)	0.0039 (0.0032, 0.0045)	0.0872 (0.0808, 0.0938)	0.0263 (0.0225, 0.0303)

Notes: Results based on estimation of 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 paper we focused on comparing the difference between the predicted probability that the Challenger initiates a *MID*, \hat{p}_c , when $ARM = 1$ and when $ARM = 0$ 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 Experiment 1 and Experiment 2 when CAP_1 is held at its mean value (0.5), including the results where $PROX_1 = PROX_2 = 0$ referred to in footnote 55. The results when $PROX_1 = PROX_2 = 1$ correspond to the results presented in Table 3 in the paper.

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 Experiment 1 and Experiment 2 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 56. 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, DV: <i>MID</i>	0.0740 (0.0734, 0.0747)	0.1283 (0.1269, 0.1297)	0.4218 (0.4175, 0.4262)	0.6180 (0.6133, 0.6225)
Only <i>BARG</i> , DV: <i>MID</i>	0.8082 (0.8070, 0.8094)	0.7841 (0.7827, 0.7855)	0.8082 (0.8070, 0.8094)	0.7841 (0.7827, 0.7855)
Experiment 2				
All Observations, DV: <i>MID</i>	0.0848 (0.0842, 0.0854)	0.1578 (0.1562, 0.1594)	0.2608 (0.2575, 0.2641)	0.4495 (0.4433, 0.4546)
Only <i>BARG</i> , DV: <i>MID</i>	0.8083 (0.8070, 0.8096)	0.7844 (0.7830, 0.7859)	0.8083 (0.8070, 0.8096)	0.7844 (0.7830, 0.7859)

Notes: Results based on estimation of 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, DV: <i>MID</i>	0.0766 (0.0758, 0.0774)	0.1332 (0.1315, 0.1349)	0.4401 (0.4348, 0.4454)	0.6366 (0.6309, 0.6423)
Only <i>BARG</i> , DV: <i>MID</i>	0.8134 (0.8102, 0.8165)	0.7890 (0.7850, 0.7929)	0.8134 (0.8102, 0.8165)	0.7890 (0.7850, 0.7929)
Experiment 2				
All Observations, DV: <i>MID</i>	0.0862 (0.0854, 0.0869)	0.1606 (0.1587, 0.1626)	0.2665 (0.2624, 0.2706)	0.4575 (0.4511, 0.4639)
Only <i>BARG</i> , DV: <i>MID</i>	0.8140 (0.8108, 0.8173)	0.7900 (0.7861, 0.7939)	0.8140 (0.8108, 0.8173)	0.7900 (0.7861, 0.7939)

Notes: Results based on estimation of 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, DV: <i>MID</i>	0.0716 (0.0708, 0.0723)	0.1236 (0.1219, 0.1253)	0.4037 (0.3980, 0.4095)	0.5982 (0.5923, 0.6042)
Only <i>BARG</i> , DV: <i>MID</i>	0.8014 (0.7974, 0.8055)	0.7779 (0.7747, 0.7811)	0.8014 (0.7974, 0.8055)	0.7779 (0.7747, 0.7811)
Experiment 2				
All Observations, DV: <i>MID</i>	0.0835 (0.0827, 0.0843)	0.1550 (0.1531, 0.1570)	0.2553 (0.2516, 0.2590)	0.4413 (0.4358, 0.4468)
Only <i>BARG</i> , DV: <i>MID</i>	0.8008 (0.7968, 0.8049)	0.7775 (0.7741, 0.7809)	0.8008 (0.7968, 0.8049)	0.7775 (0.7741, 0.7809)

Notes: Results based on estimation of 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.