Imamverdiyeva, U. and Shea, P. E. (2022) Re-examining women leaders and military spending. *Journal of Peace Research*, 59(5), pp. 679-693. (doi: 10.1177/00223433211055909).

This is the author's final accepted version.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

http://eprints.gla.ac.uk/299221/

Deposited on: 23 May 2023

Enlighten – Research publications by members of the University of Glasgow
http://eprints.gla.ac.uk

Re-examining women leaders and military spending

Ulkar Imamverdiyeva[1] and Patrick E. Shea[2]

[1]University of Houston
[2]University of Glasgow

Abstract

Do women leaders enact more hawkish foreign policies? Some research argues women leaders are more likely to adopt aggressive and masculine characteristics to obtain national office. As a result, women leaders should exhibit more hawkish behavior than men. In this study we re-examine the relationship between the women leaders and foreign policy by focusing on military spending behavior. We argue that conventional empirical methods, such as linear regression, are ill-suited to examine data on women leaders and military spending. These methods are sensitive to outliers and small sample sizes: two characteristics of women leadership. To address these issues, we use the synthetic control method to estimate the military spending behavior of women leaders. By creating unique synthetic counterfactuals for three prominent women leaders – Thatcher, Gandhi, and Meir – we analyze what would have happened if a particular state had a male leader. Generalizing beyond these cases, we also conduct a multiple-treatment test that examines the effect of women leadership jointly across multiple countries and time periods. We find that women leaders do not spend more on the military than men. We analyze plausible explanations of these null results and discuss their implications.

Keywords: gender; military spending; women leaders; foreign policy

*Corresponding author: patrick.shea@glasgow.ac.uk

Over the past two decades, the number of women leaders increased four-fold (Goemans, Gleditsch & Chiozza, 2009). Does this changing representative dynamic lead to different foreign policies and international relations outcomes? Some research on women leaders and foreign policy suggests dramatic changes, as women leaders have a positive effect on both conflict and military spending (Caprioli & Boyer, 2001; Koch & Fulton, 2011; Dube & Harish, 2017; Post & Sen, 2020; Schramm & Stark, 2020). There is no consensus in the literature, however, as some research shows no relationship between women leaders and conflict (for example, see Horowitz, Stam & Ellis, 2015).

This study re-examines the relationship between women leaders and foreign policy by focusing on military spending.[1] We argue that conventional methods, such as linear regression, are ill-suited to examine data on women leaders and foreign policy. These methods are sensitive to outliers, particularly the case of Israel. To avoid the shortcomings of regression analysis, we use synthetic control methods to estimate the military spending behavior of women leaders in general. In small samples, synthetic case methods allow for a quantitative comparison of cases of interest, systematically comparing units that are similar on a range of covariates.

This analysis avoids the pitfalls of previous research on women leaders and foreign policy that attempted to extrapolate inferences from too few observations. By creating unique synthetic counterfactuals for three prominent women leaders – Thatcher, Gandhi, and Meir – we analyze what would have happened if a particular state had a male leader. The analysis reveals that Thatcher, Gandhi, and Meir did not increase military spending more than their male counterparts. Generalizing beyond these cases, we also conduct a multiple-treatment test that examines the effect of women leadership jointly across multiple countries and time periods. These results show no relationship between gender and military spending. To explain these null results, we consider and test for several plausible explanations. For example, women and men may simply have similar preferences on military spending. Alternatively, if women leaders adopt more hawkish policies to overcome gender stereotypes, we should observe a convergence in foreign policy behavior between

[1] While others justifiably conceptualize gender as a continuum, we treat gender as a binary concept for abstraction and empirical purposes.

men and women leaders. We find evidence consistent with both these explanations, and we discuss potential extensions to our analysis in the conclusion.

Gender and international relations

There is a well-established literature that examines foreign policy and security issues in the context of gender.[2] Some of the literature focuses on women's general preferences towards conflict. For example, Eichenberg (2003) find that, on average, women are less supportive of the use of military force than men. This support varies given the stakes involved in conflict and whether the conflict has multilateral support (Brooks & Valentino, 2011).

Moving beyond individual attitudes, does gender affect foreign policy? At the societal level, states with higher female empowerment are generally less conflictual (Caprioli, 2000; Caprioli & Boyer, 2001; Goldstein, 2001; Regan & Paskeviciute, 2003). At the legislative level, more female representation is associated with lower military spending and less conflict (Caprioli, 2000; Caprioli & Boyer, 2001; Koch & Fulton, 2011).[3] Overall, a large swath of gender and foreign policy research suggests that more peaceful preferences and more peaceful state behavior is associated with female equality and female representation. Research focused on female executives, however, either find no relationship between the gender of leaders and hawkish behavior (Horowitz, Stam & Ellis, 2015) or a positive one (Caprioli & Boyer, 2001; Koch & Fulton, 2011; Dube & Harish, 2017; Post & Sen, 2020; Schramm & Stark, 2020).

For example, both Caprioli & Boyer (2001) and Koch & Fulton (2011) find that female leaders are associated with *more* conflict and higher military spending. While their respective research designs were more focused on other issues (i.e. gender equality and representation, respectively) the relationship between female executives and hawkish foreign policy stands in contrast to the rest

[2]For a review of recent research on gender and conflict see Reiter (2015) and Sjoberg, Kadera & Thies (2018). For a review of quantitative research in gender politics in general, see Stauffer & O'Brien (2019).

[3]There are exceptions to the negative relationship between female representation and conflict. Bjarnegård & Melander (2013) find no such relationship. Others suggest the relationship is contingent on the context or issues surrounding conflict (Conover & Sapiro, 1993; Shea & Christian, 2017).

of the gender and foreign policy literature. Caprioli & Boyer (2001) and Koch & Fulton (2011) argue that the positive relationship between female leaders and conflict is a result of the gender stereotypes that female leaders face. These gender stereotypes are most prominent at the executive level of national politics, as these leaders are expected to handle foreign policy issues. Gender stereotypes prompt voters to use gender as a heuristic to evaluate politicians' performance and infer policy positions (Herrnson, Lay & Stokes, 2003). These stereotypes portray men as better able to handle foreign policy issues because aggressiveness is deemed by the electorate as the appropriate strategy in foreign policy (Lawless, 2004; Swers, 2007, 2013).

Given the existence of gender stereotypes related to foreign policy, women face an obvious barrier in obtaining office. If the electorate generally perceives women as ill-suited to manage foreign policy issues, then selection effects condition the relationship between leaders' gender and leaders' foreign policy preferences. The women that we observe as state leaders are not exogenously selected into power. Instead, women leaders' selection is a function of the electorates' perceptions of their foreign policy preferences. This endogenous selection process makes it difficult to infer gender's effect on foreign policy outcomes.

Given this selection process, we expect two possibilities for women leaders. First, if women want to be elected as an executive, they need to combat gender stereotypes by adopting more hawkish attitudes, thus acting against their true preferences. As a result, executive female leaders are forced into taking positions consistent with their male counterparts (Eagly, 2007). Alternatively, the women selected to be national leaders already have the same preferences as men in regards to foreign policy, and thus their true preferences are already in line with men. Each explanation leads to the same conclusion: *we should expect men and women leaders to enact similar foreign policies.*

Whether women leaders actually have hawkish preferences or merely imitate hawkish preferences is difficult to empirically distinguish, but should lead to similar empirical implications. For example, both types of women face problems with the electorates. Since some women may only imitate or adopt hawkish preferences for election purposes, all women types face a credibility

problem of whether they can commit to campaign promises. Once in office women leaders merely emulating masculine characteristics on foreign policy still have incentives to maintain their façade because of future election considerations, for either themselves or their party.

Women leaders, in general, tend to be evaluated less favorably when particular leadership positions are associated with masculine styles (Eagly, Makhijani & Klonsky, 1992; Koenig et al., 2011). This is consistent with the 'double-bind' problem. Women adopt masculine styles to obtain leadership positions, but these same masculine styles contradict the electorate's perceptions of gender stereotypes, leading to lower evaluations (Eagly & Carli, 2007). The 'double-bind' obstructing women helps explain why we observe so few women as national leaders. Males, on the other hand, only need to act consistently to the gender stereotypes of strength to receive favorable evaluations (Koenig et al., 2011). As a result, we expect that male candidates act consistently with masculine styles to obtain leadership positions.

If women leaders' foreign policy strategies are consistent male leaders' foreign policy strategies, we should expect a convergence, not a divergence, of foreign policy behavior. If female candidates need to adopt stronger masculine strategies to signal toughness on foreign policy issues, men and women leaders should be articulating similar stances on foreign policy. Male leaders also face evaluation problems if their behavior is not consistent with the masculine stereotypes associated with leadership positions (Johnson et al., 2008). Consistent with this expectation, Horowitz, Stam & Ellis (2015) examine the likelihood that a leader will initiate a militarized dispute in a given year and find no discernible difference between male and female leaders. The authors provide a cautionary inference though, given the small number of female leaders in their data.

In sum, gender stereotypes regarding foreign policy issues make it difficult for females to become executives. To overcome these stereotypes, women adopt more aggressive, masculine foreign policy stances. Alternatively, women may already have foreign policy preferences similar to men. Consequently, the females we observe in office should exhibit similar, not different, foreign policy goals and strategies as their male counterparts. Foreign policies encapsulate varied policy portfolios, but we focus our expectations on military spending. Military spending is often a salient

issue in national elections (Philpot, 2018). Therefore, we expect gender stereotypes to be important on how voters evaluate candidates on military spending issues. Second, as we discuss in more detail below, alternative outcomes of interest such as conflict behavior are complicated by strategic considerations. Finally, the focus on military spending is consistent with previous research on female executives and foreign policy (Koch & Fulton, 2011).

Empirical analysis: re-examining women leaders and military spending

Following the discussion above, we expect a null relationship between leader's gender and various foreign policy dependent variables. Before describing our empirical strategy, we explain why we do not follow the regression strategies of previous research. In Koch & Fulton's (2011) analysis of twenty-two advanced democracies between 1970 and 2000, only twenty-two observation years have female leaders, and thus the results may be influenced by a few – or even just one – unusual leader.[4] For example, Golda Meir – the Israeli leader from 1969 to 1974 – presents a potential problem of statistical inference. Caprioli & Boyer (2001) note that out of the ten Interstate Crisis Behavior (ICB) crises that involved female leaders, seven involved Meir.

We replicate Koch & Fulton's (2011) analysis of military spending and conflict behavior. We describe the data, the models, and the replication results in more detail in the Online appendix, but we graph the residuals from Models 1 and 4 from Koch & Fulton (2011). Consistent with our concerns, the residuals in Figure 1 shows that Israel is not well explained and Meir's tenure has some of the largest residuals.[5]

[Figure 1 about here.]

[4]We thank the authors for sharing their data and their Stata code. We note that Koch & Fulton (2011) were not only focused on female leaders, but rather on the broader effects of women representation on foreign policy, in the context of party politics and political control. Our replication analysis does not speak to the overall implications of their article.

[5]The cloud of residuals on the right of Figure 1b belong to the United States.

Omitting Israel from this analysis or including Golda Meir as a binary control variable changes statistical inferences and improves model fit. These ad-hoc model specification choices, however, are not sufficiently justified.[6] If we are trying to make general inferences about the relationship of female leaders and foreign policy, Israel (and Golda Meir) are important observations that should not be discarded for empirical convenience. Instead of critiquing the robustness of previous statistical analysis, we follow an alternative empirical strategy. We rely on an empirical method that can (1) carefully compare cases with and without female leaders, (2) limit the influence of outliers, and (3) incorporate time dynamics directly into the estimation process. We use a *synthetic case design* to better illustrate the counterfactual effects of having a woman in office versus a man. Our results show no evidence that women leaders exhibit more hawkish behavior.

Cases, crises, and female leaders

There are a small number of female leader observations with which to make inferences about foreign policy preferences. This data feature can be problematic for some research designs, such as OLS, where estimated effects are greatly influenced by outliers. A small number of female year observations, however, is also an opportunity to explore female leaders' foreign policies. We take two approaches. First, following previous research we focus on three female leaders: Indira Gandhi, Golda Meir, and Margaret Thatcher. These leaders' tenures are often used as examples of female leadership in international relations (Steinberg, 2008). We believe that this attention is deserved. As noted above, Golda Meir's Israel accounts for the vast majority of International Crisis Behavior (ICB) crises events that involve female leaders, while Indira Gandhi's India and Margaret Thatcher's United Kingdom account for the other non-Israeli wars (Brecher et al., 2016).[7] Our second approach examines female leadership in general, consistent with previous cross-sectional research (Caprioli, 2000; Caprioli & Boyer, 2001; Koch & Fulton, 2011).

When examining the cases of female leaders, we want to determine whether a woman leader's

[6]The results from these alternative model specifications can be found in the Online appendix.
[7]See the list of cases in the Online appendix.

tenure deviates substantially from a counterfactual in which a male ruled during that tenure instead. For that purpose, we use the synthetic control methodology that is particularly useful with a small sample of female leaders (Abadie & Gardeazabal, 2003; Abadie, Diamond & Hainmueller, 2010, 2015). The synthetic control methodology has two additional advantages. First, synthetic control methodology circumvents the problem of heterogeneity by focusing on a particular country while facilitating reasonable comparisons. Second, the synthetic method provides an unbiased estimate with reasonable modeling assumptions, even when the potential effects of unobservable confounders vary with time (Abadie, Diamond & Hainmueller, 2010).

The method also offers a number of advantages over traditional comparative case studies. For one, instead of comparing a single country-unit to just one other country-unit with similar characteristics, the design constructs a 'synthetic' version of a country through a weighted combination, using several variables from several countries. The idea behind the synthetic control approach is that a combination of cases provides a better comparison to the case of interest rather than any single comparison alone. For example, instead of comparing Golda Meir's Israel to just one other country, we can construct a synthetic Israel using information from several countries. Given historical or structural differences, it may be difficult to find a reasonable comparison for Israel. Similarly, it may be difficult to determine which characteristics should form the basis of comparison and by how much. The advantage of using synthetic control methods is we can evaluate how well the data fit the cases before we make inferences, rather than assume the data fit the cases in order to make an inference.

The selection of cases into the synthetic control unit is not subjective. The synthetic control method systematically chooses comparison cases out of the 'donor pool,' producing a unique counterfactual that closely resembles the characteristics of the case of interest. In contrast to matching methodology, this approach compares countries that are similar not just in terms of covariates, but also in their pre-treatment outcome of interest. As such, Abadie, Diamond and Hainmuller (2010) show that synthetic control method is better in controlling for the effect of unobservable factors that have impact on the common time trend in the treated and donor units.

Another advantage of the synthetic control concerns the hypothesis test procedure. A typical hypothesis test assumes the null hypothesis is true and then examines whether an effect size is unusually large given the null. In this study, we expect a null relationship between female leaders and foreign policies, so we cannot simply assume that to be true. Instead, we derive a distribution of potential effects from permutating the analysis for all countries (led by either male or female leaders). We then compare how female leaders compare to male leaders using the distribution of effects as a basis to judge whether the results are unusual. This approach avoids the problem of looking for null results under an assumed null distribution (Gill, 1999).

We follow the following procedure to derive synthetic control cases. First, we identify important covariates that are consistently used in the military spending literature to accurately predict military spending. Next, we determine whether our covariates produce a good synthetic case that is not dependent on only one or two donor countries. After making inferences, we test the robustness of the donor pool using an alternative estimation techniques or alternative covariates.

Formally, we observe $\mathbf{J} + \mathbf{1}$ countries, one of which has a female leader. The $\mathbf{J}$ remaining countries serve as potential control countries. From those $\mathbf{J}$ control countries, we define $\mathbf{J} \times \mathbf{1}$ vector of weights, $\mathbf{W}$, to construct a synthetic case. $\mathbf{W}$ are nonnegative, sum to one, and represent the contribution each control country provides in constructing a synthetic case. The weights are selected so that a synthetic control unit best resembles the treated unit on pre-treatment characteristics. These characteristics are defined by control variables that are typically used in regression studies of military spending, which we discuss below.

For our study, the first year of a female executive's tenure marks the start of the treatment, so the years before she enters office are considered the pre-treatment period. Let $\mathbf{T_0}$ be the number of years before a female leader comes to power and $\mathbf{T}$ represent the whole time period analyzed, with $\mathbf{1 < T_0 < T}$. $\mathbf{Y^I_{it}}$ is the outcome of interest (e.g., military expenditures) for unit *i* at time *t*. Through optimization, we determine the weights, $\mathbf{W}$, that minimizes the distance between the treated and synthetic control for the pre-treatment period (Abadie, Diamond & Hainmueller, 2010).

Next, we use $\mathbf{W}$ to create the post-treatment outcome, $\mathbf{Y^N_{it}}$, for the synthetic case. The differ-

ence between the actual and the predicted synthetic outcome after the intervention is the estimated causal effect of treatment. For $\mathbf{t} > \mathbf{T_0}$, the effect of female leadership on the military spending of a country is given by[8]:

$$\alpha_{it} = Y_{it}^{I} - Y_{it}^{N} \tag{1}$$

The estimated effect depends on the fit between the actual and predicted synthetic unit before the treatment intervention. If the synthetic control provides a poor fit with the actual unit before a female takes office, the treatment effect is difficult to infer. The fit between the treated and synthetic control is typically evaluated in two ways. First, a graphical display of synthetic and actual outcomes before the treatment can provide evidence of gross misspecification. Second, we can analyze the fit through the root mean squared prediction error (RMSPE) for the pre-intervention period (Abadie, Diamond & Hainmueller, 2015). A small RMSPE in the pre-treatment period indicates a good model fit. The extent to which RMSPE is considered small or large depends on two comparisons. First, we compare it to the variance of the outcome variable. Second, we compare the prediction error from our main case studies to the overall distribution of RMSPE, which is estimated through placebo cases (i.e. male led countries).

Once the synthetic control is estimated, the trajectory of the counterfactual synthetic case is compared to our treated unit. We carry out a permutation test by applying the synthetic control method to every potential control unit in our sample. The intuition of this permutation test is an examination of placebo cases. For example, to determine whether observed changes in military spending during the tenure of Margaret Thatcher are gender specific, we apply the same synthetic method to a country which did not have female leadership under the same time period, such as Canada. We repeat this process for all the countries in our donor pool. This process generates a distribution of potential effects of leaders, which in turn helps evaluate how our treatment effect ranks in that distribution. More generally, we examine whether the estimated effect of female

[8]The equation above represent average treatment effect. A lead-specific causal effect is given by $\alpha_{i,T_0+1}, ..., \alpha_{i,T}$ for leads for $1, 2, ..., T - T_0$.

leadership is large relative to the distribution of the effects estimated for the countries without female leadership during the same sample period. From this distribution, we compute both a year-specific and average p-value of the treatment effect computed through the placebo case analysis. In line with previous studies, we approximate the significance test by using the RMSPE ratio of the treated unit and placebo cases(Abadie, Diamond & Hainmueller, 2010; Cavallo et al., 2013). The average p-value indicates the proportion of placebos that have a ratio of post-treatment RMSPE over pre-treatment RMSPE at least as large as the average ratio for the treated unit. This non-parametric approximation of significance level does not impose any distributional assumptions on the errors and effects calculated. More formally, we compute the average significance level as:

$$p\text{-value} = P(\frac{\overrightarrow{S}^{PL}}{\overleftarrow{S}^{PL}} > \frac{\overrightarrow{S}^{t}}{\overleftarrow{S}^{t}}) = \frac{\sum_{j=1}^{J} I(\frac{\overrightarrow{S}^{PL}}{\overleftarrow{S}^{PL}} > \frac{\overrightarrow{S}^{t}}{\overleftarrow{S}^{t}})}{J} \tag{2}$$

where $\overrightarrow{S}^{PL}$ represents the post-treatment RMSPE for placebo cases, while $\overleftarrow{S}^{PL}$ is the pre-treatment RMSPE. Thus,

$$\overrightarrow{S} = \sqrt{\frac{\alpha_{T_0}^2 + ...\alpha_{T_0+T_1}^2}{T_1+1}} \tag{3}$$

$$\overleftarrow{S} = \sqrt{\frac{\alpha_1^2 + ...\alpha_{T_0-1}^2}{T_0-1}}. \tag{4}$$

Data and sample

We leverage our small sample of countries lead by female leaders by considering 'what-if' scenarios had the country been led by men. To identify a gender of the head of state on a given year we rely on data from the Archigos dataset (Goemans, Gleditsch & Chiozza, 2009). In cases where several leaders assume office in a given year, we code gender according to a leader that has spent the highest fraction of days in office over the course of the respective year.[9]

To be consistent with earlier research, we focus on female leaders' association with military

[9]Following Abadie, Diamond & Hainmueller (2015) in constructing synthetic control for each case, we restrict the donor pool by dropping all other potential cases that have female leaders.

spending. *Military spending* data is obtained from the Stockholm International Peace Research Institute (SIPRI) and log-transformed to account for its skewed distribution (SIPRI, 2016). The dependent variable is measured as levels. However, we include a lagged dependent variable (LDV) as a control, which has two implications. First, the LDV addresses the problem that budgets are often a function of previous years' budgets. Second, the inclusion of LDV changes our focus on how female leaders *change* military spending. We note that electoral candidates often make promises about changes in military spending in elections. If a new leader is bringing new foreign policy preferences into office, changes in military spending would be a logical consequence.

As discussed above, the synthetic control method chooses a weighted combination of comparison units with respect to control characteristics to provide a counterfactual for a treated unit. The control characteristics used in constructing synthetic cases are standard for military spending predictors. We begin with economic indicators which should predict the ease at which leaders can change budgetary policy. We use *GDP per capita* (measured in constant US Dollars in 2011 prices) and *GDP growth* because those states with more wealth and better economic performance will have fewer budget constraints (data from Penn World Table, Feenstra et al. (2015)). We then include a measure of *trade openness* to proxy how engaged a state is with global affairs (data from Correlates of War International Trade dataset, Barbieri & Keshk (2016)). For similar reasons, we include *the number of allies* (data from Gibler (2009)), *number of rivals* (data from Diehl, Klein & Goertz (2006)), and *number of ongoing territorial disputes* (according to ICOW, Hensel et al. (2008)). These measures attempt to measure the demand states face for military spending, as alliances may ease that demand, while ongoing rivalries and territorial disputes increase that demand. Another measure that proxies the demand for military spending is *international threat*. This variable measure the likelihood that a country experiences a militarized dispute in a given year, a factor that directly predicts military expenditures (Nordhaus, Oneal & Russett, 2012; Zielinski, Fordham & Schilde, 2017).

We also control for regime type using *Polity* (Marshall, Jaggers & Gurr, 2002) and *military capacity* using COW National Material Capabilities (v5.0). Finally, to control for societal attitudes

that would make it more (less) likely to elect women as leaders and be more (less) likely to change military spending, we include *Women in legislatures*, which measures the percent of women in lower parliamentarian bodies. Caprioli & Boyer (2001) argue that female representations is a direct measure of gender equality in a state, which also indirectly proxies social and economic equality for women as well. In robustness tests, we also consider alternative measures for gender equality, as this concept serves as an obvious confounder to our model. To measure *Women in legislatures*, we rely on data from World Bank's World Development Indicators (1991, 1997 – 2015) and Paxton, Green & Hughes (2008) to fill in gaps in the data in all other years. See the Online appendix for the list of control units as well as weights each has on the synthetic counterfactuals.

United Kingdom - Margaret Thatcher (1979-1990)

Margaret Thatcher's tenure as the prime minister of the United Kingdom focused mostly on domestic policies, particularly austerity and deregulation. Thatcher, considered a radical Conservative because of her economic policies, was moderate in her handling of the Falklands War and mediation between the U.S.A. and Soviet Union (Steinberg, 2008: 223). She took a non-cooperative stance with the European Community, consistent with her domestic program's goals of reducing spending and regulations. In sum, while Thatcher's domestic policies were at the extreme of her party and country, her foreign policies were mostly consistent with her party's traditions.

To empirically analyze the effect of Thatcher on military spending of the UK, we construct the synthetic UK as a combination of the countries in the donor pool using the covariates discussed in previous section plus the lagged value of military spending. Table I reports how much the countries in the donor pool contributed to create a synthetic counterfactual of the UK. The contributions range from zero (Denmark, Finland) to 41 % (Canada), with all contributions summing to one. Again, instead of comparing the UK to just one case or to all the cases, we rely on the most relevant information from some cases to build a synthetic comparison. Table II reports the actual value of covariates for the UK, synthetic data values, and the sample mean of countries in the donor pool. The covariate values for the UK are closer to the synthetic values than just taking the

sample mean. Overall, this shows that the synthetic UK provides a better comparison for the UK than relying on the average of the countries in the donor pool.

[Table 1 about here.]

[Table 2 about here.]

To further evaluate the quality of the synthetic counterfactual by comparing military spending for the UK case and the synthetic UK case (dashed line) before Thatcher takes office in Figure 2a. Figure 2a shows that the synthetic UK closely follows the military spending for the UK during the pre-intervention period starting from 1960 until 1979. The fit of synthetic UK can also be seen from the small values of MSPE relative to the range of dependent variable which is displayed with rug plots on the y-axes.[10]

[Figure 2 about here.]

Additionally, we can analyze the synthetic UK's goodness of fit — pre-intervention RMSPE — by comparing it to the same outcome estimated for 33 placebo countries, as shown in Figure 2b. As described above, this placebo analysis allows us to derive a distribution of potential effects of male leaders, which in turn helps evaluate how our treatment effect ranks in that distribution. This comparison reveals that 87% of placebos have a pre-treatment RMSPE at least as large as that of UK, meaning that goodness of fit we observe for UK is centrally located in the distribution.[11]

Given the overall quality of the synthetic UK, we then compare the synthetic and actual military spending of Thatcher's UK in the post-treatment period. We observe minimal differences between the actual and synthetic military spending suggesting that she exerted little or no effect on the military spending of the UK. These differences are not statistically distinguishable from zero, as shown in Figure 2c. In short, the synthetic analysis provides support to our hypothesis of convergence of foreign policy between male and female leaders.

[10]The pre-intervention MSPE for UK is 0.004 and the military spending ranges from 0 to 13.3 with a standard deviation of 2.2.

[11]In an additional robustness test, we take a leave-one-out approach for each donor country that contributes any weight to the results. This approach produces a new set of donor countries, with a new set of weights, but produces similar inferences as our main results. See Online appendix for these results.

India - Indira Gandhi (1966-1977; 1980-1984)

The only child of India's first prime minister (Nehru), Indira Gandhi was appointed prime minister in 1966 after her predecessor died in office (Malhotra, 1989). The most formative foreign policy crisis that occurred during her tenure was the civil war between East and West Pakistan that ultimately led to the independence of Bangladesh. Her leadership during the Bangladesh crisis is cited as the pinnacle of her career (Steinberg, 2008: 36), while the rest of her tenure was marred by corruption charges, food shortages, inflation, and internal conflict. The remaining years of Gandhi's tenure focused more on the domestic cohesion of India rather than the rivalry between India and Pakistan. Unlike the tenures of Golda Meir and Margaret Thatcher, Indira Gandhi was less ideologically driven and more focused on keeping India's government under her control (Steinberg, 2008). We do not, however, observe Gandhi using hawkish policies to convince her electorate of her credibility of her office. Instead, we observed patience and calculated diplomacy during the Bangladesh crisis, where Gandhi chose not to push India's military advantages over Pakistan. Thus, we conclude that Gandhi was no more a hawk than other Indian leaders on international issues, nor did gender appear to affect her decision-making or policy positions. While there are some quotes from Pakistani leadership to suggest that Gandhi's gender escalated the Bangladesh conflict, it is difficult to envision another policy strategy that the Pakistani leadership could have taken without losing both Bangladesh and positions of power.

We analyze India's military spending during Gandhi's tenure and the counterfactual implied by the synthetic India constructed as a weighted combination of other countries from 1955 to 1985. We report the contributions from the donor pool countries and the corresponding covariate values in the Online appendix. Figure 3a illustrates the results of this analysis. The graph shows that India and its synthetic counterpart are closely aligned in the pre-treatment period. While there are small differences between actual and predicted military spending in the pretreatment period, in comparison to the distribution of military spending (displayed with rug plots on the y-axis) those differences are negligible. In addition, the pre-treatment MSPE for India is 0.058 and the military spending ranges from 0 to 13.2, with a standard deviation of 2.5, suggesting a good fit between

India and the synthetic control.

[Figure 3 about here.]

We do not observe a significant divergence between India and the synthetic case in the post-treatment periods. While we do observe a slight gap between synthetic and observed values of military spending, the placebo cases in Figure 3b reveal that those differences are not statistically significant. Figure 3c plots the *p-values* associated with each year of Gandhi's tenure derived from the placebo analysis. We fail to reject the null hypothesis as none of the *p-values* reach the conventional level of significance. This provides evidence that Gandhi did not exhibit unusual military spending.

Israel - Golda Meir (1969-1974)

The unexpected death of Prime Minister Eshkol brought Golda Meir into power in Israel one week after the War of Attrition started (Steinberg, 2008). Several other crises sprang up during her tenure, most notably the Yom Kippur War. In that war, Israel's slow mobilization and Egypt's improved anti-aircraft defense resulted in heavy initial losses for Israel. With U.S.A. support, Israel was able to reverse Egypt's early gains. A cease-fire was eventually orchestrated by the U.S., the Soviet Union, and the UN. While Israel recorded a military victory, its sense of invulnerability was crushed. Meir decided to resign from office and retired from politics for good in 1974 (Meir, 1975).

Golda Meir was a hawk, who refused to compromise on any foreign policy issue that she viewed was related to the existence to Israel. Steinberg (2008: 134) notes that Meir was viewed as a tough negotiator, who would not make concessions to Egypt. Thus it is reasonable to conclude that her foreign policy views helped her secure her leadership position. However, did gender help formulate these foreign policy preferences? Given that Meir's formative political years did not take place in a stable political environment, but rather a series of conflicts (first with the British, and then with the Arab states), it is reasonable to conclude that Meir's foreign policy preferences

were a function of these conflicts. In Steinberg's (2008) analysis of Meir's leadership style, the prime minister was mostly motivated by ideological concerns, rather than political survival. This is evident from how others viewed Meir. Henry Kissinger praised Meir's unrelenting defense of Israel's security and referred to her as a 'tiger' in negotiations. Even Sadat noted his admiration for Meir's tenacity in her defense of Israel's existence (Finklestone & Obe, 2013). While Meir's foreign policy preferences are hawkish, an important question is whether the foreign policy strategies of the Israeli government would have been different under male leadership.

To answer this question, we construct a synthetic Israel that mirrors the characteristics of Israel before Meir takes office.[12] Figure 4a graphs the actual and synthetic case predictions of military spending of Israel before and after the election of Golda Meir to office in 1969. We observe that Israel's military spending surpassed predictions of synthetic Israel, though the divergence starts years before Meir takes office. Despite this divergence between the synthetic and actual Israel in the pre-treatment period, we derive a good fit in the pre-treatment period. As in previous cases, the MSPE value shows the negligible amount of error relative to the range of military spending in our sample.[13]

[Figure 4 about here.]

In order to analyze the statistical significance of the gap, we carry out placebo studies in Figure 4b and compute lead specific *p-values* based on the distribution of effects in Figure 4c. While the p-values are closer to the conventional level of statistical significance (0.05) than the previous cases, they do not meet or exceed this threshold.[14]

The findings in Figure 4 inform our understanding of the positive association found in previous studies. First, it appears that during the tenure of Meir military spending was clearly above the average values Israel had before. Yet unlike regression tests, synthetic control allows us to observe

[12]The Online appendix compares the covariate values of actual Israel, synthetic counterfactual as well as averages of donor countries.

[13]Pre-intervention MSPE is 0.087. Analysis of placebo cases and comparison of prediction errors shows that 17% of countries in the donor pool have MSPE at least as large as the average of Israel's prediction error.

[14]The average p-value is 0.09, indicating that out of all the countries in the donor pool 9% have at least as large effect as the main treatment effect we observe for Israel.

the starting point of the increase in the military spending and determine it began to increase while Prime Minister Eshkol was still in power.[15]

The counterfactual for Israel during 1969-1974 is informative given that Meir's ascension into power was a consequence of (1) the unexpected death of Prime Minister Eshkol and (2) a political compromise to avoid party disunity given the close competition between Deputy Prime Minister Yigal Allon and Defense Minister Moshe Dayan (Steinberg, 2008: 134). If Eshkol had survived or Meir had stayed retired, it is easily imaginable that a man would have been in office. It is also easy to imagine that a non-Meir prime minister would result in a foreign policy consistent with the Meir government. Both Dayan and Allon were valuable advisors to Meir on foreign policy and military matters throughout her tenure. Given that Meir marginalized advisors who opposed her views, we can infer that Dayan and Allon's views were consistent with Meir (Steinberg, 2008: 184). Subsequently, results of synthetic control analysis yield similar findings. In sum, Figures 2-4 indicate that none of the female leaders under consideration had any substantive effect on the military spending of a country.

Multiple treatment synthetic control

While the previous three cases are of historical relevance, there are limitations of making a generalization from three cases on the relationship between gender of leaders and foreign policy. Thus, we extend our analysis to a larger sample of countries with female leaders and derive an average effect across a large **N** of countries with female leaders. So far, we have focused on the effect of the female leaders in specific countries and found no evidence that female leaders change military spending more than their male counterparts. We extend this analysis by estimating the average causal effect of female leadership on military spending of multiple treated units from 1960 to 2013.

We conduct a test that detects a treatment effect jointly across multiple countries and time

[15]We also construct the synthetic Israel by including time splines in addition to covariates discussed, to control for temporal dependence. Particularly, we include *t* which measures time since the last fatal conflict, t^2, and t^3 in creating synthetic Israel. The results of this estimation is presented in the Online appendix. The two models yield very similar results. The addition of time variables while slightly improves the RMSPE, yields p-value of 0.12 meaning we fail to see any effect of female leader on the military spending of Israel.

periods, while at the same time accounting for heterogeneity. Our sample is made up of countries that meet the following criteria: (a) no missing information in the outcome variables in the sample period of 53 years and (b) no missing values in any of the covariates in the same sample period. This yields 49 case studies, of which 19 have female leaders.[16] The rest of the cases consist of countries which did not experience female leadership, from which we construct a synthetic counterfactual for each treated unit. [17] Following the same procedure for the individual synthetic cases, we first calculate the average treatment effect, $\bar{\alpha}$.

Then we conduct placebo tests and obtain $\bar{\alpha}^{PL(j)}$. This design produces 30^{19} possible permutations of placebo assignments. To make computation feasible, we draw 10,000 such permutations to approximate the empirical null distribution. We then compare our treatment effect to the permutated distribution. As a result, the multiple treatment analysis allows us to identify the exact distribution of treatment effect under a null hypothesis, and then perform statistical inference by analyzing how large the $\bar{\alpha}$ is compared to the empirical null distribution (Cavallo et al., 2013).

Given the number of treated units, we only display the average effect of female leadership during the first ten years in office. We begin with reporting the significance of the average effect of female leadership on the military spending of a country. In Figure 5 we observe that none of the p-values reach statistical significance. On average, female leaders do not have any substantive impact on military spending.

[Figure 5 about here.]

Both single and multiple-treatment synthetic control analysis reveals that there are no differences in military spending of male and female leaders. Using a data-driven method appropriate for small samples, we find that gender does not necessarily result in any substantive changes in the military spending of a country.

[16]The list of the countries and treatment time is given in the Online appendix.

[17]For instance, United States serves as a control unit in the computation of the counterfactual for Chile (treatment period of 2006-2010) and New Zealand (treatment period of 2001-2008).

Additional tests and explanation of results

This section briefly summarizes a number of robustness tests we conducted to show that the main inferences are not sensitive to research design choices. These tests corroborate our expectations that female leaders enact similar foreign policies as male leaders. Because of space limitations, a more detailed description of these robustness tests are located in the Online appendix. Here is a brief summary of some of these tests: (1) We examine alternative dependent variables, including conflict and UN voting behavior; (2) We examine alternative independent variables; (3) We examine alternative leader characteristics to demonstrate that leaders matter to military spending; (4) We use multiple imputation to address missingness in the data. In sum, these robustness tests support our original inferences. We note that each variant to our original synthetic control analysis produce new synthetic comparisons and new donor pool compositions.

In addition, we conduct analysis to test the validity of the potential explanations for the null results. We put forward two plausible explanations: 1) if women want to be elected as an executive, they need to combat gender stereotypes by adopting more hawkish attitudes; and 2) women may simply already have the same preferences as men in regards to foreign policy, and thus their true preferences are already in line with men. It is empirically difficult to distinguish these two possibilities, but we attempt to test these mechanisms. First, we examine the effect women leadership in Norway and Denmark. These states are generally viewed as more gender egalitarian, and thus if women leaders have different preferences on military spending, we may be more likely to observe that effect here. We find no difference in military spending between men and women leaders. This lends support for the notion that men and women leaders have similar preferences on foreign policy. However, we cannot rule out the possibility that there is a sufficient amount of gender stereotypes in Norway and Denmark – as evidenced by the dearth of observations of female leaders in Nordic countries – to prompt female leaders to emulate males on foreign policy issues.

We also examine lame duck leaders' behavior, following Carter & Nordstrom (2017). It is possible that lame duck women leaders are more likely to behave consistent with their true policy preferences given that they do not have electoral concerns. Two countries –Philippines and Ar-

gentina – had lame duck women leaders at some point in their histories. We first examine Philippines' president Gloria Macapagal Arroyo. She offers a unique case to analyze because Philippines presidents usually only serve one term. Arroyo, however, served one term when she took power after her predecessor was impeached, and then was elected for a second term. Our analysis shows that Arroyo increased military spending in her first term, but when her lame duck term started, military spending was statistically lower than the synthetic case (see Online appendix for results). We replicate the analysis for Philippines president Cory Aquino and Argentina president Christina Kirchner, but we find no difference in military spending. We note that these lame duck presidents are not necessarily lame duck politicians, as these women ran for other offices after their presidency. We cannot, however, rule out that men and women leaders hold the same preferences on military spending.

Next, we briefly examine non-foreign policies, where gender stereotypes should not affect female leaders as much as military spending. In short, female leaders are more likely to act consistently with their preferences on these issues because they do not face the same level of gender stereotypes. We demonstrate that female leadership has a positive effect on health care spending. This result supports the claim that men and women leaders have different preferences on some policy dimensions. However, we cannot claim that the variance in preferences extend to foreign policies. In sum, our additional tests provide some suggestive evidence for the two plausible explanations, but cannot definitely validate or nullify one or the other. We suggest future research consider this question.

Conclusion

Our analysis demonstrated that typical regression techniques are limited in their ability to test hypotheses related to women leaders and foreign policy. The combination of the small sample size of female leaders and outliers lead to precarious inferences. Using synthetic control methods, we show that women leaders do not diverge significantly from their male counterparts on military

spending policy. Our results are consistent for individual cases – Thatcher, Gandhi, and Meir – and for the multiple-treatment test that examines the effect of women leadership across multiple countries and time periods.

The null results could be a result of gender stereotypes prompting women to adopt more masculine foreign policies. Alternatively, women and men leaders may simply have similar preferences on foreign policies to begin with. We ran additional tests related to each proposed mechanism, though we conclude that there is not enough evidence to dismiss one explanation over the other. We suggest that future research disentangle these mechanisms. For example, recent studies have demonstrated that elite level surveys are possible (Busby et al., 2020; Kertzer, 2020), which may shed light on foreign policy preferences of leaders.

While our research shows that women and men leaders enact similar military spending policies, our argument suggests two additional implications. First, male and female leaders do not necessarily converge on all policies. The more gender stereotypes surround a particular issue, the more we expect female leaders to adopt policy positions that are consistent with masculine characteristics. Military spending is usually a salient foreign policy issue during leader selection processes, where leaders can demonstrate how 'tough' they are on foreign policy. However, if other foreign policies to not trigger the same amount of gender stereotypes, then we may observe more policy differences between male and female leaders. We show in the appendix that male and female leaders differ on health care spending, but we suggest that future research consider the variance of foreign policies. Humanitarian interventions or foreign aid may not prompt the same gender stereotypes as military spending, allowing female leaders more political leeway to follow their own preferences.

Besides variance of gender stereotypes across issue, gender stereotypes may vary over time and space. Gender stereotypes may wane in some states as more and more females attain offices concerned with foreign policy (Barnes & O'Brien, 2018). We contend that we have not reached that point yet with female leaders and military spending given the lack of states that have appointed multiple female leaders.[18] However, female representation is changing dramatically,

[18]Only Lithuania, Switzerland, and Finland have had more than two female head of governments.

suggesting that gender stereotypes surrounding foreign policy may also change. As more female leaders emerge we expect gender stereotypes surrounding foreign policies to wane, which may lead observable differences in policy outcomes, including military spending. Current data limitations inhibited us from fully addressing the possibility of waning gender stereotypes, but these dynamics can and should be examined in the future, possibly using text-analysis of media content related to men and women candidates.

The existing comparative analyses of female executives in international relations is rather limited. The most obvious reason for this shortage was that women have been historically absent from executive offices. In this article, we show that with new methodologies, we can offer an example of how to analyze the effect of gender in international relations. Continued research on the relationship between female leaders and various foreign policy aspects is strongly warranted.

Replication data

The datasets, do-files, and R scripts for the empirical analysis in this article, along with the Online appendix, are available at https://www.prio.org/jpr/datasets/ and at Patrick Shea's website: https://patrickshea.weebly.com/research.html. All analyses were conducted using either R or Stata.

Acknowledgements

We thank the anonymous referees and the editor of *JPR* for their guidance. We also thank Scott Wolford, Tyson Chatagnier, and participants from the MPSA and Texas Triangle conferences for their helpful comments and feedback.

References

Abadie, Alberto; Alexis Diamond & Jens Hainmueller (2010) Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association* 105(490): 493–505.

Abadie, Alberto; Alexis Diamond & Jens Hainmueller (2015) Comparative politics and the synthetic control method. *American Journal of Political Science* 59(2): 495–510.

Abadie, Alberto & Javier Gardeazabal (2003) The economic costs of conflict: A case study of the Basque country. *The American Economic Review* 93(1): 113–132.

Barbieri, Katherine & Omar Keshk (2016) Correlates of war project trade data set codebook. (https://correlatesofwar.org/data-sets/bilateral-trade)

Barnes, Tiffany D & Diana Z O'Brien (2018) Defending the realm: The appointment of female defense ministers worldwide. *American Journal of Political Science* 62(2): 355–368.

Bjarnegård, Elin & Erik Melander (2013) Revisiting representation: Communism, women in politics, and the decline of armed conflict in East Asia. *International Interactions* 39 (4): 558–574.

Brecher, Michael; Jonathan Wilkenfeld, Kyle Beardsley, Patrick James & David Quinn (2016) International crisis behavior data codebook, version 11 (http://sites.duke.edu/icbdata/data-collections/).

Brooks, Deborah Jordan & Benjamin A Valentino (2011) A war of one's own: Understanding the gender gap in support for war. *Public Opinion Quarterly* 75(2): 270–286.

Busby, Joshua; Craig Kafura, Jonathan Monten & Jordan Tama (2020) Multilateralism and the use of force: Experimental evidence on the views of foreign policy elites. *Foreign Policy Analysis* 16(1): 118–129.

Caprioli, Mary (2000) Gendered conflict. *Journal of Peace Research* 37(1): 51–68.

Caprioli, Mary & Mark A. Boyer (2001) Gender, violence, and international crisis. *Journal of Conflict Resolution* 45 (4): 503 – 518.

Carter, Jeff & Timothy Nordstrom (2017) Term limits, leader preferences, and interstate conflict. *International Studies Quarterly* 61(3): 721–735.

Cavallo, Eduardo; Sebastian Galiani, Ilan Noy & Juan Pantano (2013) Catastrophic natural disasters and economic growth. *Review of Economics and Statistics* 95(5): 1549–1561.

Conover, Pamela Johnston & Virginia Sapiro (1993) Gender, feminist consciousness, and war. *American Journal of Political Science* 37(4): 1079–1099.

Diehl, Paul; James Klein & Gary Goertz (2006) The new rivalry data set: Procedures and patterns. *Journal of Peace Research* 43 (3): 331–348.

Dube, Oeindrila & SP Harish (2017) *Journal of Political Economy* 128(7): 2579–2652.

Eagly, Alice H (2007) Female leadership advantage and disadvantage: Resolving the contradictions. *Psychology of Women Quarterly* 31(1): 1–12.

Eagly, Alice H & Linda L Carli (2007) Women and the labyrinth of leadership. *Harvard Business Review* 85(9): 62-71.

Eagly, Alice H; Mona G Makhijani & Bruce G Klonsky (1992) Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin* 111(1): 3-22.

Eichenberg, Richard C (2003) Gender differences in public attitudes toward the use of force by the united states, 1990-2003. *International Security* 28(1): 110–141.

Feenstra RC, Inklaar R, & Timmer MP (2015) The next generation of the Penn World Table. http://www.ggdc.net/pwt/. *American Economic Review* 105(10): 3150–3182.

Finklestone, Joseph & Joseph Finklestone Obe (2013) *Anwar Sadat: Visionary Who Dared*. Abingdon:Routledge.

Gibler, Douglas M (2009) *International Military Alliances, 1648–200* . Washington, DC: CQ Press.

Gill, Jeff (1999) The insignificance of null hypothesis significance testing. *Political Research Quarterly* 52(3): 647–674.

Goemans, Henk E; Kristian Skrede Gleditsch & Giacomo Chiozza (2009) Introducing Archigos: A dataset of political leaders. *Journal of Peace Research* 46(2): 269–283.

Goldstein, Joshua S (2001) *War and Gender*. Cambridge: Cambridge University Press.

Hensel, Paul R; Sara McLaughlin Mitchell, Thomas E Sowers, & Clayton L. Thyne (2008) Bones of contention: Comparing territorial, maritime, and river issues. *Journal of Conflict Resolution* 52 (1): 117–143.

Herrnson, Paul S; J Celeste Lay & Atiya Kai Stokes (2003) Women running 'as women': candidate gender, campaign issues, and voter-targeting strategies. *Journal of Politics* 65(1): 244–255.

Horowitz, Michael C; Allan C Stam & Cali M Ellis (2015) *Why Leaders Fight*. Cambridge: Cambridge University Press.

Johnson, Stefanie K; Susan Elaine Murphy, Selamawit Zewdie & Rebecca J Reichard (2008) The strong, sensitive type: Effects of gender stereotypes and leadership prototypes on the evaluation of male and female leaders. *Organizational Behavior and Human Decision Processes* 106(1): 39–60.

Kertzer, Joshua D (2020) Re-assessing elite-public gaps in political behavior. *American Journal of Political Science*. (https://doi.org/10.1111/ajps.12583).

Koch, Michael T & Sarah A Fulton (2011) In the defense of women: Gender, office holding, and national security policy in established democracies. *The Journal of Politics 73(1)*: 1–16.

Koenig, Anne M; Alice H Eagly, Abigail A Mitchell & Tiina Ristikari (2011) Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin* 137(4): 616-642.

Lawless, Jennifer L (2004) Women, war, and winning elections: Gender stereotyping in the post-September 11th era. *Political Research Quarterly* 57(3): 479–490.

Malhotra, Inder (1989) *Indira Gandhi: A Personal and Political Biography*. Boston, MA: Northeastern Univiversity Press.

Marshall, Monty G; Keith Jaggers & Ted Robert Gurr (2002) Polity iv dataset. (http://www.systemicpeace.org/polity/polity4.htm).

Meir, Golda (1975) *My Life*. New York: GP Putnam's Sons.

Nordhaus, William; John R Oneal & Bruce Russett (2012) The effects of the international security environment on national military expenditures: A multicountry study. *International Organization* 66(3): 491–513.

Paxton, Pamela; Jennifer Green & Melanie Hughes (2008) Women in parliament, 1945-2003: Cross-national dataset. *Inter-university Consortium for Political and Social Research* 24340 (https://doi.org/10.3886/ICPSR24340.v1).

Philpot, Tasha S (2018) Race, gender, and the 2016 presidential election. *PS: Political Science & Politics* 51(4): 755–761.

Post, Abigail S & Paromita Sen (2020) Why can't a woman be more like a man? Female leaders in crisis bargaining. *International Interactions* 46(1): 1–27.

Regan, Patrick M & Aida Paskeviciute (2003) Women's access to politics and peaceful states. *Journal of Peace Research* 40(3): 287–302.

Reiter, Dan (2015) The positivist study of gender and international relations. *Journal of Conflict Resolution* 59(7): 1301–1326.

Schramm, Madison & Alexandra Stark (2020) Peacemakers or iron ladies? A cross-national study of gender and international conflict. *Security Studies*: 29(3): 515–548.

Shea, Patrick E & Charlotte Christian (2017) The impact of women legislators on humanitarian military interventions. *Journal of Conflict Resolution* 61(10): 2043–2073

SIPRI (2016) Stockholm International Peace Research Institute Military Expenditure Database (https://www.sipri.org/databases/milex).

Sjoberg, Laura; Kelly Kadera & Cameron G Thies (2018). Reevaluating gender and IR scholarship: Moving beyond Reiter's dichotomies toward effective synergies. *Journal of Conflict Resolution* 62(4): 848–870.

Stauffer, Katelyn E & Diana Z O'Brien (2019) Fast friends or strange bedfellows? Quantitative methods and gender and politics research. *European Journal of Politics and Gender* 2(2): 151–171.

Steinberg, Blema S (2008) *Women in Power: The Personalities and Leadership Styles of Indira Gandhi, Golda Meir, and Margaret Thatcher* Volume 4. Montreal: McGill-Queen's University Press.

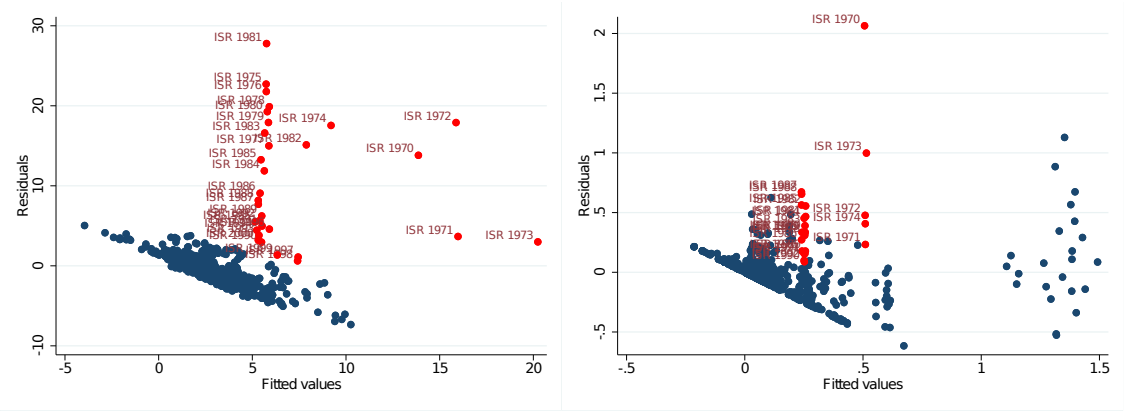
Swers, Michele L (2007) Building a reputation on national security: The impact of stereotypes related to gender and military experience. *Legislative Studies Quarterly 32(4)*: 559–595.

Swers, Michele L. (2013) *Women in the Club : Gender and Policy Making in the Senate*. Chicago, IL: University of Chicago Press.

Zielinski, Rosella Cappella; Benjamin O Fordham & Kaija E Schilde (2017) What goes up, must come down? The asymmetric effects of economic growth and international threat on military spending. *Journal of Peace Research* 54(6): 791–805.

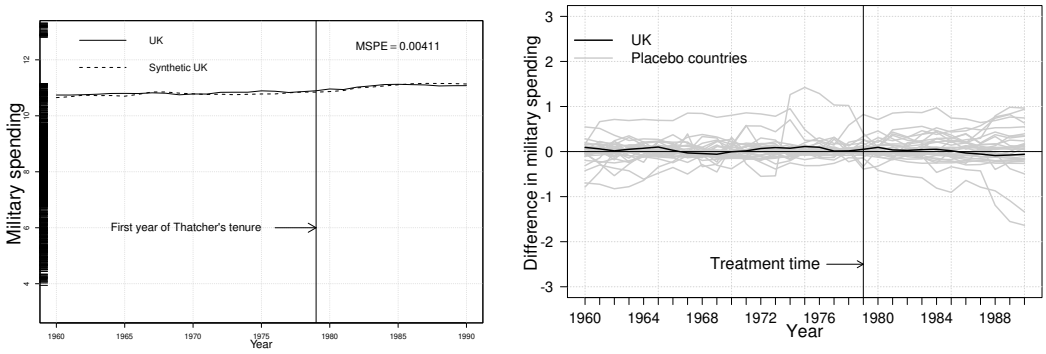
Biographical statements

ULKAR IMAMVERDIYEVA, PhD in Political Science (University of Houston, 2021); Senior Data Scientist at CarMax (2021-). PATRICK E SHEA, PhD in Political Science (Rutgers University, 2013); Associate Professor, University of Houston (2013-). Current main interest: domestic politics and international relations.

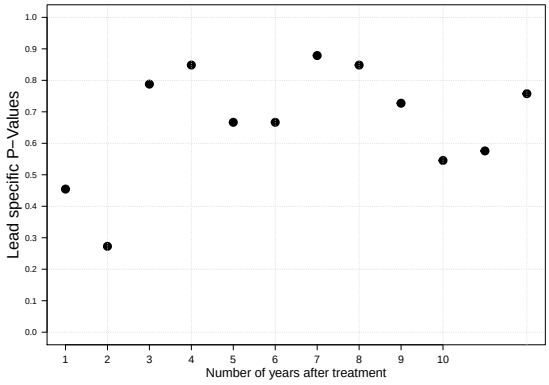


(a) Residuals of military spending OLS (b) Residuals of conflict OLS

Figure 1. Residuals from Koch & Fulton's (2011) Models 1 and 4

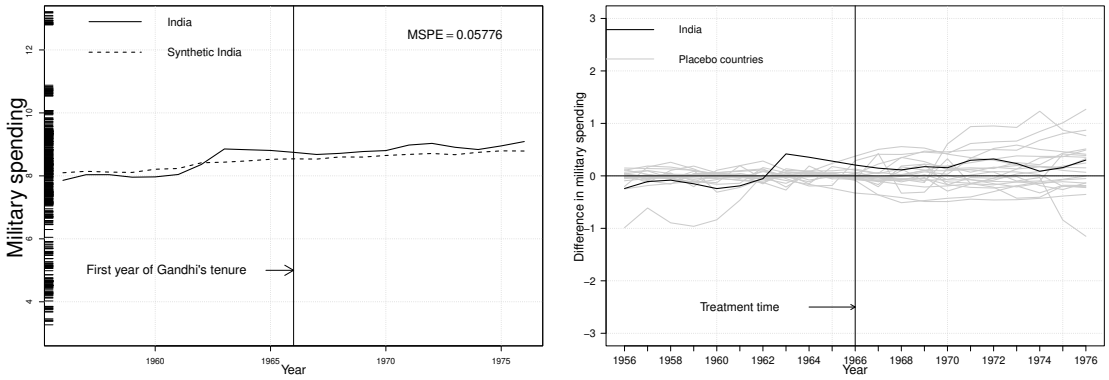


(a) Treatment effect (b) Placebo comparisons

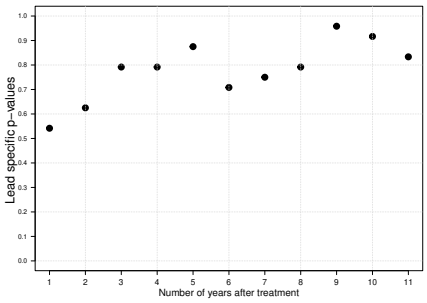


(c) Lead specific p-values

Figure 2. Actual UK and synthetic UK's military spending for Thatcher

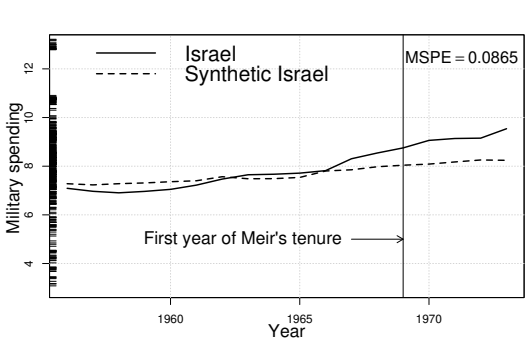


(a) Treatment effect (b) Placebo comparison

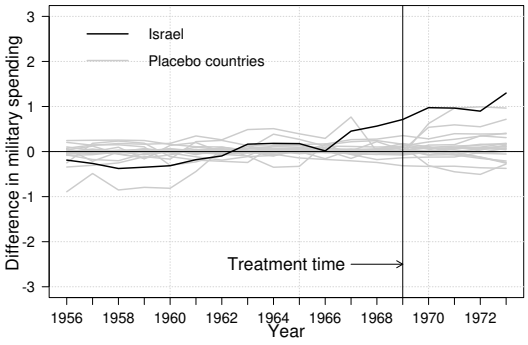


(c) Lead specific p-values

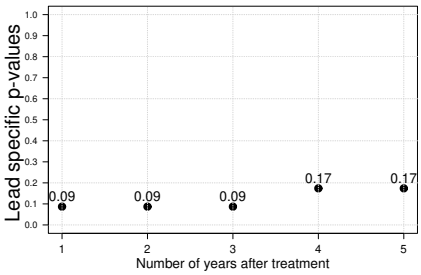
Figure 3. Actual and synthetic India military spending for Gandhi



(a) Treatment effect



(b) Placebo comparison



(c) Lead specific p-values

Figure 4. Actual Israel and synthetic Israel military spending for Meir

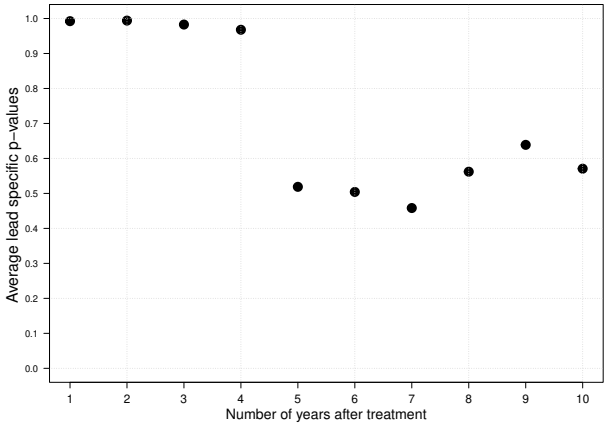


Figure 5. Significance levels of an average treatment effect

Table I. Synthetic weights for UK for military spending

Country names	**Weights**	**Country names**	**Weights**
United States	0.364	Spain	0
Canada	0.414	Portugal	0
Dominican Republic	0	Austria	0
Mexico	0	Italy/Sardinia	0
Guatemala	0	Greece	0
Costa Rica	0	Finland	0
Colombia	0	Sweden	0.092
Ecuador	0	Denmark	0
Peru	0	South Africa	0
Brazil	0	Morocco	0
Paraguay	0	Iran	0
Chile	0	Japan	0
Ireland	0	Thailand	0
Netherlands	0	Malaysia	0
Belgium	0.003	Australia	0.021
Luxembourg	0	New Zealand	0
France	0.107		

Table II. Military spending covariates for UK, synthetic UK and sample mean

	Actual UK	Synthetic UK	Sample mean
Military Spending.1978	10.86	10.85	7.98
Polity	20	20	14
GDPpc	15780	21102	11993
GDP growth	2.76	4.31	5.46
Women legislatures	4.11	3.78	4.41
National Capabilities	0.032	0.075	0.012
Trade	10.74	10.66	8.47
Threat	5.26	5.55	5.22
Number of allies	4.16	4.92	1.33
Number of issues	28	28.227	3.735
Number of rivals	8.32	5.94	1.14