Evaluating the Effects of Nonproliferation Agreements

Bradley C. Smith* William Spaniel†

January 30, 2019

Abstract

How do nuclear agreements keep parties from proliferating? One perspective is that such deals constrain these states' nuclear capacity, reducing their ability to produce weapons at an efficient cost. Current approaches cannot test this due to inappropriate measures of capacity and lack of appropriate comparison group. We address both of these points, gathering new data to estimate proficiency and adopting the generalized synthetic control method. We then evaluate the constraining effects of deals the United States implemented with Japan, South Korea, and Taiwan as well as the Declaration of Iguaçu between Brazil and Argentina. Our results indicate that the constraining effect is minimal.

*Department of Political Science, Vanderbilt University, Commons Center 329, Nashville TN 37203. (bradley.carl.smith@gmail.com)

†Department of Political Science, University of Pittsburgh, Posvar Hall 4446, Pittsburgh, PA 15260. (williamspaniel@gmail.com, http://williamspaniel.com).

1 Introduction

In 1969, President Richard Nixon announced the Guam Doctrine, indicating that U.S. allies were the primary party responsible for their own security. For the next few years, South Korea accelerated its nuclear weapons program, looking for civilian assistance from Belgium, Canada, and France and trying to poach American nuclear engineers of Korean descent (Oberdorfer, 2001, 69). Policymakers in Washington began worrying that another state would soon join the nuclear club just as they were gathering support for the Non-Proliferation Treaty. Concessions came forth, with the United States offering \$1.5 billion in military aid (Drezner, 1999, 255). Although the United States offered to assist South Korea in developing Kori-2, the latter's second nuclear power plant, the United States expected South Korea to stop making progress on nuclear activities it deemed objectionable.

The South Korean case suggests two distinct mechanisms for how agreements reduce proliferation temptations. First, the United States' concessions acted as a "carrot" to policymakers in Seoul. Violating the terms would risk the United States withholding aid, thereby raising the opportunity cost of proliferation. Second, South Korea also restricted its investment in native nuclear activities. With less research done in nuclear science, racing to a nuclear weapon at a later date would be harder than had South Korea continued its steady progress.

In this paper, we ask how nuclear agreements lead to fewer proliferation outcomes. More specifically, we evaluate the effectiveness of the second mechanism, the constraints on nuclear technologies short of atomic bombs. Our results indicate that constraint has minimal impact. At most, states may reduce their nuclear capacity by a small degree relative to where they would have been in the absence of a deal.

Although the literature identifies constraints as a possible nonproliferation mechanism (Reiss, 1995), researchers have yet to provide a rigorous investigation of the claim. Doing so requires both operationalizing nuclear capacity and identifying a proper comparison to the parties of the agreement. We contribute to the broader study of nuclear proliferation by providing solutions to each. First, we expand Smith and Spaniel's (2018) ν data. The new coding gives us both a more nuanced measure of nuclear proficiency and greater temporal coverage. Second, we bring the generalized synthetic control (GSC) method to the topic (Xu, 2017). This allows us to construct an estimate of the counterfactual version of the state that had not entered the agreement. Comparing the counterfactual state's nuclear capacity to the observed level provides an estimate of the agreement's effect.

In particular, we apply the method to four agreements. Three consist of the United States providing security guarantees to friendly East Asian countries during the Cold War: Japan, South Korea, and Taiwan. We find minimal support for the notion of a proficiency constraining effect of these agreements. The agreements between allied East Asian countries would seem to be a most likely case to observe a substantial effect. Unlike rivals, Japan, South Korea, and Taiwan had warm relations with the United States and therefore had less concern that policymakers in Washington would backtrack on a deal after receding their nuclear capacity. Nevertheless, we observe continued capacity increases following the agreements. They fall below the estimated counterfactual versions of these countries that did not broker a deal, but the overall effect is substantively minor and statistically indistinguishable from zero. Individually, only Taiwan shows some partial restraint.

The fourth is the 1985 Declaration of Iguaçu, a bilateral agreement between Argentina and Brazil to ease perceived hostilities and terminate a potential nuclear rivalry. Given the relative ineffectiveness of the East Asian agreements, one would expect to see little restraint between Argentina and Brazil. After all, these countries were rivals at the time of their agreement. The results confirm that speculation. Neither country increased their capacity following the agreement. However, we estimate the counterfactual countries would not have improved much either. Thus, the estimated effect is close to zero.

These subtle (non) effects demonstrate the utility of the GSC method. A naïve analysis of these agreements would suggest very different conclusions. For example, because Argentina's and Brazil's proficiency scores level off after 1985, a simple comparison before and after would suggest that the deal cooled both states' appetite for development. The GSC findings indicate no effect. Meanwhile, South Korea's score continue to increase after its deals. This would seem to imply that the agreement might have encouraged South Korea to continue. Yet the GSC method finds at worst no difference and at best marginal improvement over the counterfactual.

Our results have important implications for theories of proliferation and applications of agreements to policymakers. If deals do not constrict nuclear proficiency, concessions are the only nonproliferation motivation they provide to countries. Thus, terminating the conciliatory measures of nuclear agreements (e.g., friendlier relations and sanctions relief with the Iran Deal) would put the target countries back on a direct path to the bomb. The results also ought to temper expectations about how agreements can alter incentives for countries to develop nuclear know-how. For example, it may be unreasonable to expect the Joint Comprehensive Plan of Action to have an immediate impact on Iran's ability to produce nuclear technology.

To be clear, our findings speak to a state's general ability to produce nuclear technology, not any one specific technology in particular. Part of the goal of such agreements is to convince would-be proliferators to not develop enrichment and reprocessing technology, which is a critical step in the proliferation process (Fuhrmann and Tkach, 2015). Qualitative analysis of agreements indicate that this is often effective. But production of those technologies and the ability to produce them are analytically distinct concepts (Smith and Spaniel, 2018), and the Non-Proliferation Treaty explicitly gives states the right to pursue many forms of nuclear technology. We estimate that signing agreements has little effect on that overall nuclear capacity; the estimates are consistent with a state's nuclear industry not disappearing following a deal but rather shifting to uncontroversial activities. Still, in some ways, this gives reason for optimism: states are not pursuing those forbidden technologies despite the ability to do so. But it also further underscores the fact that technical incapacity is *not* what drives that reticience.

We proceed as follows. The next section outlines theories of nuclear agreement deals, comparing concessions to constraint mechanisms. Afterward, we give a full account of the measurement and comparison problems that serve as a barrier to inference. The fourth section details our method of creating counterfactual potential proliferators. We then give brief descriptions and motivations of the cases we investigate. Our final substantive section explores the results. A brief conclusion then ends the paper.

2 Concessions or Restraint? Two Perspectives on Counter-Proliferation

Nuclear agreements have two possible and distinct mechanisms: conciliatory measures from opposing states and nuclear capacity limitations from the potential proliferator.

As we detail later, the United States offered alliance benefits to each of their East Asian protégé states. In the immediate aftermath of the Cold War, the United States and Russia convinced Ukraine to move the leftover Soviet missiles on its soil in exchange for economic aid (Ellis, 2001, 9), energy subsidies (Reiss, 1995, 122), and security assurances (Cirincione, Wolfsthal and Rajkumar, 2011, 374-377). Argentina and Brazil exchanged favorable trade agreements in exchange for each other's non-nuclear compliance (Reiss, 1995, 55-58). And the Iran Deal offers Tehran a path toward integration into the world community if it gives up its program.

However, deals often come with the expectation that the potential proliferator will constrain or retract their nuclear growth. The Iran Deal provides the clearest example, requiring that Iran close multiple facilities, end ongoing construction projects, and downblend enriched uranium. In 2003, Libya handed over much of its nuclear infrastructure when struck a deal with the United States. As a lead-up to its agreement with Argentina, Brazil's 1988 constitution banned non-peaceful nuclear applications. Meanwhile, Argentina never opened a plutonium reprocessing facility it had started constructing in 1978. It also closed a uranium enrichment facility in 1994 that it had only operated for seven years.

The literature on nuclear weapons views both of these avenues as productive nonproliferation mechanisms. Many earlier works investigated concessions between allies as a cause of nuclear reticence, and this line of research has continued to the present (Reiss, 1988; Paul, 2000). More recent work expands the theory to rival states (Volpe, 2017), with Spaniel (2019) arguing that nonproliferation agreements exist under broad conditions. Opponents look at the potential proliferator's value for developing weapons and make offers commensurate with those costs and benefits. These deals remain credible as long as the potential proliferator maintains the ability to develop weapons at a later date. Any change to this alters the value of proliferating and therefore causes the opponent to adjust its concessions.

Meanwhile, the quantitative nuclear weapons literature finds a consistent link between capacity and proliferation. In the earliest work, Singh and Way (2004) use steel production and electricity generation as a crude proxy for technical sophistication and find a strong connection. Jo and Gartzke (2007) improved this measure by creating a seven-point index of nuclear-related infrastructure and show it predicts nuclear programs. As we detail further in a moment, Smith and Spaniel (2018) use a latent variable model that incorporates Jo and Gartzke's (2007) activities and others, recovering the same result.

A second group of capacity findings demonstrate that assistance also predicts proliferation behaviors. Kroenig (2009) shows this for states receiving sensitive technologies specifically geared toward the development of nuclear weapons. But the logic extends to civilian and "dual use" technologies that have civilian applications but could be of military interest with a small change in direction. This can be the result of states providing nuclear cooperation agreements to enhance infrastructure (Fuhrmann, 2009) or International Atomic Energy Agency technical cooperation (Brown and Kaplow, 2014), which fulfills a similar purpose.

The common logic behind these findings is that greater capacity makes developing nuclear weapons cheaper, and states are more likely to construct cheaper weapons. But there are two reasons to think that capacity constraint may not drive the long-term success of nuclear agreements. First, it is not clear whether a state can "forget" capacity. A state can shut down some of its facilities. However, most of the challenge of building nuclear weapons is not in making physical progress but rather in knowing what the conceptual challenges are and how to overcome them (Colaresi, 2014, 58-59). This knowledge would persist if scientists simply shifted to uncontroversial nuclear activities. Unless reaching an agreement causes those scientists to shift industries, we would expect proficiency to not change.

Second, the bargaining literature indicates that potential proliferators would demonstrate reluctance to follow through with such measures. To obtain compliance, opponents must offer deals commensurate with the state's ability to construct nuclear weapons. Moreover, these deals cannot be one-time measures. Instead, the flow of benefits must continue over time, lest the potential proliferator take the short-term influx and develop a weapon anyway. But reducing capacity means raising the costs of a nuclear program down the line. Opponents would therefore have incentive to cut back on agreements once signed. Anticipating this, the potential proliferator may refuse to take meaningful measures on that front.

This paper investigates whether that is the case. Do agreements slow a potential proliferator's nuclear capacity? To be clear, if the answer is "no," our research does not imply that such agreements are worthless. It does, however, inform the mechanism by which they achieve success. In particular, if the literature points to concessions as the alternative, then a "no" answer indicates that bargaining does the heavy lifting.

3 Empirical Challenges

Two barriers stand in the way of estimating whether nuclear agreements constrain a country's technical capacity. Fortunately, recent methodological developments solve the issues. We use this space to explain those challenges and demonstrate why our research design is effective for answering the question.

3.1 Measurement Problems

The first major challenge is in measurement of the dependent variable. In the quantitative nuclear weapons literature, binary proliferation behaviors—i.e., exploration, pursuit, or possession—tend to receive most of the focus. From a measurement perspective, binary variables lose much of the nuance in the proliferation process. From an inference perspective, binary measurements give us little leverage to explain the mechanisms through which states reach those outcomes. This is of particular importance for the cases we are interested in. Nonproliferation agreements the United States has reached with allies, for example, see consistent compliance. Explaining that success requires analysis of a state's nuclear program beyond just its proliferation status.

Smith and Spaniel (2018) provide provide some leverage here. Rather than conceptualizing proliferation as binary, they develop a continuous measure of nuclear proficiency. One of these measures, ν Infrastructure divorces nuclear science capability from nuclear weapons capability. It is a useful measure for a country's price tag for nuclear weapons; the higher a state's infrastructure, the fewer scientific hurdles the country needs to overcome to proliferate. It therefore appears to be an ideal variable for testing how nonproliferation agreements affect a state's nuclear proficiency. This, in turn, would tell us how much retarding proficiency enhancement causes the corresponding nuclear weapons reticence.

However, this is a bit hasty. Although the item response theory method Smith and Spaniel use is perfect for the construction of the variable we would like, their actual measurement is crude. In particular, the latent model only uses 12 activities to estimate their infrastructure measure. Two problems result. First, a country's score can fluctuate wildly upon beginning or quitting an activity. It is unclear whether a country's proficiency actually spikes or plummets in these years or if it is just an artifact of the small quantity of data. Second, and relatedly, the credibility intervals surrounding the ν measure are fairly large.

The only way to rectify this problem is to collect more data. As such, we have done just that. In supplemental appendix, we give a full accounting of the data expansion.[1] The data now cover previously ignored or underexploited activities. To begin, we went back through the original United Nations Statistical Yearbooks to disaggregate the industrial activities that Meyer (1986) originally coded and Jo and Gartzke (2007) condensed for their index, which Smith and Spaniel (2018) originally used.[2] The original data just had indicators for heavy water and non-heavy water power plants; we now include an indicator for reactors with each type of neutron moderator. The updated measure also captures whether a country uses the plant for desalination, if it meets a high production threshold, engagement with advanced power plant design, and membership in working groups to further develop the technology. As an entirely new feature, we coded information on countries' research reactors using the IAEA Research Reactor Database. Disaggregating Fuhrmann and Tkach (2015) uranium enrichment data, we separate whether states have advanced to centrifuge technology. And in our final major upgrade, we scanned the International Nuclear Information System database for countries publishing research related to nuclear science. In total, the dataset now contains 104 variables in total, including 98 that fall under the infrastructure category. We also expand the data through 2016, adding 15 years more coverage to the original measure.

Two improvements to the data make it particularly useful for our purposes. First, we code whether a state engages in key activities—e.g., power plants, uranium enrichment, and plutonium processing—on their own or with international partners. This better tracks whether a state's proficiency comes from their own knowledge or reliance on others. It is critical given that some nuclear agreements result in a transfer of nuclear assistance, which may otherwise appear to further develop the country's native

[1] This information appears in section 1 of the appendix, beginning on page 1.

[2] The earlier works had condensed the variables because an indexing approach would double the value of two tasks that serve similar purposes in the role of nuclear weapons construction. However, if two activities were functionally the same, the item response approach would identify that and not overvalue those activities.

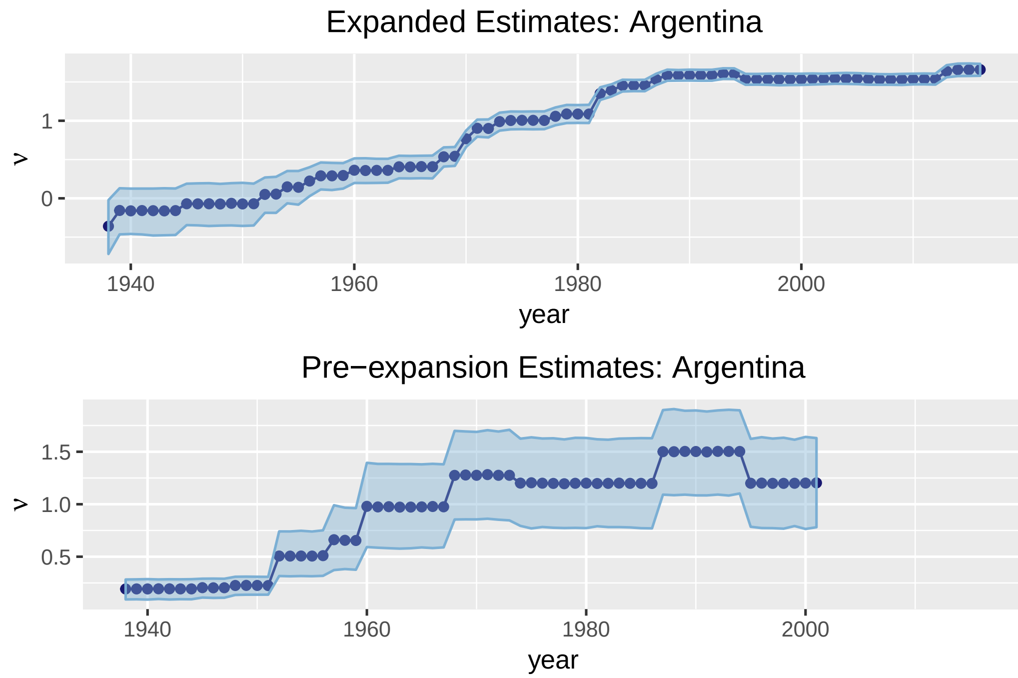

Figure 1: A comparison of our new estimates using the expanded data with the original estimates from Smith and Spaniel (2018). The plot indicates that the data expansion results in a smoother time trend as well as a substantial reduction in uncertainty.

knowledge than the actual change. Second, the data better plots the trajectory of the facilities. We add construction start dates of power plants from the IAEA's Reference Data Series and construction start dates of enrichment and reprocessing facilities from Fuhrmann and Tkach (2015). Again, digging deeper into Fuhrmann and Tkach's (2015) data, we include the scale of all enrichment and reprocessing facilities. Thus, the measure captures more subtle gains to nuclear capacity.

Figure 1 illustrates the improvement. It plots Argentina's estimated proficiency with the original and our new measure.[3] Two points stand out. First, the broad trends in the data are similar, with Argentina's proficiency generally increasing over time. In fact, the ν measures for all countries have a correlation of .905. Second, the new measure has fewer wild fluctuations and narrower credibility intervals. For example, the original formulation sees Argentina's capacity plateau from 1987 to 1994. This coincides with Argentina's opening of the Pilcaniyeu Enrichment Plant. In our new measure, these

[3]While it is tempting to compare the numbers directly, the IRT procedure used means that we cannot compare the magnitude of the estimates from the old to the new version of the estimates.

years marginally move the estimate. The original measure also treated all enrichment facilities identically. The updated data reduce the significance of Pilcaniyeu because it did not use superior centrifuge technology and it was only a pilot scale. Aside from that, our updated measure has more indicators to understand Argentina's underlying proficiency and therefore does not jump or enfeeble a score based on that single activity.

More broadly, the estimates are more precise than in the original analysis. The new data have more more activities are observed in each year, providing more information and reducing the corresponding uncertainty around the estimates. We can see this by comparing the credible intervals generated in each set of estimates. To do this, we calculate the ratio, for every country-year, of the absolute value of the ν estimate to the length of the corresponding 95% credible interval obtained from the posterior distribution. In the original data, this ratio is approximately 0.91, while in the new data this ratio is approximately 2.59. This indicates a substantial reduction in uncertainty resulting from the inclusion of additional activity indicators. We also redid the replications of Jo and Gartzke (2007) and Fuhrmann and Tkach (2015) and found that the new measure generates a better model fit. In sum, our data are the theoretically appropriate measure to answer our research question, and diagnostics provide confirmation of this.

3.2 Control Problems

Our second problem is the fundamental problem of causal inference. To understand how a deal changes nuclear proficiency trends, we need to compare an agreeing party to itself, except if it had not made the agreement. Of course, this is not possible, as we cannot observe the counterfactual, no-agreement outcome directly. Thus, the problem is essentially one of missing data. How do we impute this missing counterfactual value to obtain an estimate of the impact of such agreements? In one way or another, the solution relies on making comparisons between treated and untreated units. The key difficulty is that both treatment assignment and observed outcomes may be related to unobserved confounders, rendering direct comparisons unreliable estimates of the effect of interest. For example, a dip in proficiency following an agreement could be the result of either the agreement itself, or an unobservable factor that both made an agreement more likely *and* reduced proficiency. Because of the strategic and context-sensitive nature of these deals, it is unlikely that such confounding is not a concern. Without accounting for this, evaluating the effectiveness of nonproliferation agreements is difficult.

While this issue is ever-present when dealing with observational data, it is not insurmountable. Given certain assumptions it is possible to leverage the data to obtain an estimate of the influence of agreements. The nature of our data points toward a potential source of such assumptions. More precisely, our aforementioned capability estimates form a time-series cross-section (TSCS) of nuclear capability. The repeated observations in TSCS data allow us to leverage assumptions about how unobserved confounding factors might vary over time to eliminate the problem of unobserved confounding. However, the nature of nuclear proliferation means that the assumptions necessary to implement common approaches such as fixed effects estimation or differences-in-differences are unlikely to be satisfied.

A common approach to causal inference with time-series cross-section data is to assume that unobserved confounders, though they may be heterogeneous across units, are fixed across time within each unit. This approach is commonly referred to as fixed-effects estimation. When this assumption is plausible, concerns over confounding can be eliminated by simply including unit-specific dummy variables in a regression. For example, in our setting this would amount to running a regression including an indicator for each country. If unobservable factors within each country are invariant over time, these indicator variables account for all unobserved confounding.

However, the fixed effects assumption is implausible in our setting as it ignores a major trend in nuclear proliferation. Nuclear technology spreads over time. Thus, it would be unsurprising to observe an increase in proficiency from an older time period to a later time period. To demonstrate that time-varying, unit-specific factors likely drive some of the variation in our data, we estimate a simple regression of nuclear proficiency on time. If the relationship is not constant (i.e. the regression coefficient is distinguishable from zero) then the assumption necessary for fixed effects estimation is questionable. Table 1 plots the results. The estimated coefficient is positive and statistically distinguishable from zero, indicating an upward trend in proficiency over time. The simple fixed-effects approach is not suitable if even a small portion of this temporal trend is driven by unobserved factors.

When the assumption of time-invariant heterogeneity is implausible, as it is here,

Table 1: OLS regression to assess the relationship between time and the development of nuclear technology. Consistent with expectations, technology is trending upward over time.

	Dependent variable:
	Smith and Spaniel's (2018) ν Score
Time	0.012^{***}
	(0.0004)
Constant	-23.440^{***}
	(0.861)
Observations	10,986
R^2	0.063
F Statistic	740.411^{***} (df = 1; 10984)
Note:	*p<0.1; **p<0.05; ***p<0.01

a "differences-in-differences" (DID) strategy provides a possible alternative. This approach allows for time-varying confounders by making assumptions about the nature of the unobserved trend. In particular, the key identifying assumption in this setting is that of *parallel trends*. This amounts to assuming that the difference between pre-and post-treatment outcomes for the treated and control groups are the same in the absence of treatment. Put simply, parallel trends requires that the *trajectory* of unobserved confounders is the same.

The parallel trends assumption is problematic in our setting. While this assumption is untestable, it is more plausible when observed pre-treatment trends appear parallel in the data. This is not the case for nuclear proficiency—some states rush to develop nuclear proficiency, while others pursue a more leisurely path. The Soviet Union, for example, sacrificed efficiency for speed (Bundy, 1988, 177-178), worried that the United States may eventually take preventive measures (Spaniel, 2015) or lose access to the moles in the American nuclear program (Colaresi, 2014, 58-62). In contrast, India took a more leisurely approach, testing its first weapon in 1974 but not actually producing a deliverable weapon until 1987, even halting pursuit entirely between 1975 and 1980 (Bleek, 2017). Differences such as these suggest that the parallel trends assumption is implausible in the full data. While trends are not parallel in general, a possible solution is to select a plausible "control group" by hand, selecting states that appear to have parallel development trajectories in the pre-treatment period to our treated countries of interest. While this is possible in principle, it relies on ultimately arbitrary choices about which states constitute acceptable controls. Consequently, any results obtained from such an approach are at risk of being driven by these choices.

In sum, neither fixed effects nor differences-in-differences are appropriate in our setting. This presents a significant challenge. Fortunately, recent methodological developments provide a way forward. In particular, Xu (2017) provides a method to deal with potential time and unit-specific heterogeneity when the parallel trends assumption is untenable. In the following section, we provide an overview of this method and the data we use to implement it.

4 Creating Counterfactual Potential Proliferators

To overcome the difficulty of generating an appropriate counterfactual for states receiving nonproliferation agreements, we adopt the generalized synthetic control (GSC) method proposed by Xu (2017). This method overcomes the difficulties associated with both fixed effects and differences-in-differences approaches to deliver estimates of the causal impact of nonproliferation agreements on the development of nuclear proficiency. The GSC method obtains these estimates by generating an estimate of the counterfactual, no-agreement outcomes for states with agreements. Taking the difference between the observed outcome and the estimated counterfactual provides an estimate of the agreement's impact.

The GSC method achieves this in three steps. First, it estimates a linear, interactive fixed effects model in the spirit of Bai (2009) using only observations in the control group. The interactive fixed effects model produces estimates of both coefficients on observed variables and a set of unobserved factors and country-specific coefficients for these factors, typically referred to as factor loadings. Using these estimates, the second step incorporates the treated group, estimating factor loadings to minimize the predicted difference from the observed proficiency measure for the treated observations in the pre-treatment period. The final step uses the estimated coefficients and factor loadings to generate predicted, counterfactual values for the treated group in the post-treatment period, comparing these to the observed values to generate estimates.

This straightforward method allows us to overcome the difficulties outlined previously. Perhaps most importantly, the GSC approach allows us to significantly relax the "parallel trends" assumption by modeling unobserved factors in a flexible way. In particular, the method can capture unobserved factors that take the form of common shocks with potentially heterogeneous cross-sectional effects.[4] As Xu (2017) notes, this means that the GSC can capture difference-in-differences and fixed effects models as special cases. Additionally, while the spirit of the approach is similar to Abadie, Diamond and Hainmueller's (2015) synthetic control method, the GSC approach has a number of advantages that make it particularly well-suited for our application. The method allows us to deal with multiple treated groups in a single run and produces frequentist uncertainty estimates. Further, it removes the need to select the appropriate pre-treatment model specification by using a cross-validation procedure to select the appropriate model for unobserved factors.

Although this method is well-suited to deal with potential unobserved factors driving proliferation behavior, we must also take observed confounders into account. Indeed, the key identifying assumption of the GSC is a form of "strict exogeneity," which requires that the error term is independent of treatment assignment, observed covariates, and unobserved cross-sectional and time-variant factors. A failure to include all relevant observed confounding variables would constitute a failure of this assumption. Accordingly, building the synthetic control state requires us to condition on observed confounders. To do this, we gather a set of variables from key papers in the quantitative nuclear proliferation literature (Singh and Way, 2004; Jo and Gartzke, 2007; Fuhrmann, 2009; Brown and Kaplow, 2014) that are related to both agreements and nuclear outcomes:

- **Recent Disputes:** A count of the number of militarized interstates disputes that the state participated in over the last five years. Disputes give a measure of the country's security environment, which influences the attractiveness of nuclear weapons.

- **Number of Borders:** The number of other states directly contiguous with

[4]More precisely, allowing i to index country and t to index time, the GSC can account for an unobserved factor U_{it} as long as the unobserved factor can be decomposed into a multiplicative form $U_{it} = a_i * z_t$.

the state in question, as defined by Stinnett et al. (2002). Like disputes, this is a measure of the security environment, with more nearby states being more potential rivals.

- **U.S. Ally:** A binary variable whether the state had a defensive alliance with the United States in the given year. States receiving patronage from the U.S. may have less incentive to proliferate because the alliance could satisfy their security concerns (Bleek and Lorber, 2014). Alternatively, Debs and Monteiro (2016) argue that alliances can give protégés protection from rivals to build technology without the threat of preventive action. We code this variable based on ATOP (Leeds et al., 2002).

- **Soviet Union/Russia Ally:** Identical to U.S. ally but for a defensive alliance with the Soviet Union or Russia.

- **Major Power:** A binary variable whether Correlates of War recognizes the state as a major power.

- **Democracy:** The state's Polity score in the given year. Autocratic leaders may have greater incentive to develop weapons of mass destruction as a coup-proofing procedure (Brown, Fariss and McMahon, 2016).

- **NPT Ratification:** Whether the state ratified the Nuclear Non-Proliferation Treaty in the given year or any year prior. Although the grand bargain of the deal permits assistance, the nonproliferation regime aims to keep states below weapons thresholds.

- **Nuclear Cooperation Agreements:** The number of cumulative agreements that the state has signed up to and including the current year Fuhrmann (2009).

- **Logged GDP and Logged GDP Squared:** Self-explanatory. We include a squared term to account for potential nonmonotonic effects. Poor states may find investment in nuclear technology unaffordable. However, rich countries may avoid developing nuclear technology so as to avoid economic sanctions (Solingen, 2007).

- **Economic Openness:** A state's imports plus its exports, all divided by its GDP. Openness also captures exposure to the world economy and vulnerability to economic coercion. Data come from Singh and Way (2004).

- **Economic Liberalization:** Change in the state's economic openness over the past five years. States attempting to liberalize may wish to divert fewer economic resources to nuclear pursuits (Solingen, 1994). Data also come from Singh and Way (2004).

- **Logged Urban Population:** Also self-explanatory. Urban population captures some of the industrial capacity states have to build nuclear technology.

In addition, GSC model includes state fixed effects. Thus, the model also implicitly accounts for any other factor that stays constant for a state over time. Some that may seem relevant for our purposes include the state's geographical region and its land size.

Perhaps a more interesting discussion is about the variables we excluded. CINC scores appear to be an obvious choice. However, ν uses both electricity generation and iron and steel production, and so the same underlying variable generates both measurements. A broader problem is that the GSC method requires lead time before the treatment to train the expectation. Like nuclear cooperation agreements, IAEA technical cooperation would be an appropriate variable, but the data do not begin for that until 1972 (Brown and Kaplow, 2014).

5 Case Selection

Having developed the model to build counterfactual countries, we now explore which agreements to investigate. The method precludes a few cases. Negotiations between nuclear-equipped Soviet successor states and the United States and Russia seem like a good starting point. However, GSC requires having training data for a state prior to the treatment year. The United States passed the Cooperative Threat Reduction bill in 1992; the Soviet Union collapsed in 1991. We therefore cannot investigate nuclear proficiency for Belarus, Kazakhstan, or Ukraine.

On the other end of the spectrum, we need a sufficient time period after the treatment year to obtain a more complete picture of the agreement's effect. Many of the control variables we use end in 2000. This precludes the 1994 Agreed Framework between North Korea and the United States, as we would only have six post-treatment years.[5] It also eliminates Libya's 2003 agreement and the 2015 Iran Deal.[6]

Those logistical issues aside, we still have five prime cases to analyze. Each has received copious attention from the literature. We therefore briefly describe the intention of the agreements and how we can apply the treatment to the method we described earlier.

5.1 East Asian American Allies

First, consider three similar agreements that the United States reached with some of its East Asian allies in the 1970s. Negotiations with Japan came first. Despite being the victim of the only offensive nuclear attacks in world history, Japan flirted with the idea of building a stockpile after China's test in 1964 and observing the Vietnam War (Paul, 2000, 48). These security threats caused conservatives in the Japanese program to want to explore a Japanese nuclear deterrent.

Nevertheless, legal barriers stood in the way. Article 9 of the Japanese constitution outlaws war as a dispute resolution mechanism. Meanwhile, 1955's Atomic Energy Basic Law restricts Japan's nuclear activities to peaceful research. Yasuhiro Nakasone, then director of the Japanese defense agency and future Prime Minister, therefore commissioned a white paper to find loopholes to the restrictions. It concluded that tactical, defensive weapons would be legal (Campbell and Sunohara, 2004, 222).

For the United States, the Japanese nuclear push came at a bad time. The Nuclear Non-Proliferation Treaty opened for signature in 1968. Less than a decade earlier, at the third presidential debate, John F. Kennedy voiced his concern that the world could soon see up to twenty members of the nuclear club. Policymakers saw the creation of a worldwide nonproliferation regime that outlawed nuclear weapons as a solution. Japan obtaining a nuclear weapon just a few years into the movement would have struck a significant blow against it.

The United States therefore looked to convince Japan to adhere to the developing nonproliferation norm. To do so, it placed a key bargaining chip into the pot: Okinawa.

[5]Nevertheless, we add North Korea to the analysis in the appendix for the years available. Our main results do not change much. North Korea observes a small drop in its proficiency relative to the synthetic control, coinciding the temporary closing of its reprocessing facility.

[6]Libya also has a second disqualifying factor. 2003 also coincided with the United States' invasion of Iraq. Coding 2003 as a treatment year for Libya could not disentangle the treatment of the deal from its treatment of the broader geopolitical consequences of the war.

Washington maintained control over the island after World War II, and Japan wanted it returned. To begin that process, President Richard Nixon signed the "Agreed Minute." According to its terms, control of Okinawa would revert to Japan, with the condition that the U.S. could reacquire it in the event of a security crisis (Roehrig, 2017, 50). In 1972, the United States initiated the transfer. Since then, Japan mostly dropped its overt push for nuclear weapons. Still, some policymakers in Japan remind the United States that their reticence is conditional. According to former Prime Minister Morihiro Hosokawa, if the United States wants to halt the potential rise of a nuclear Japan, it must "maintain its alliance with Japan and continue to provide a nuclear umbrella" (Levite, 2003, 71).

South Korea had a similar experience. For South Korea, China's nuclear test was of less direct concern than its unending security concerns regarding North Korea. Following the Korean War, the United States maintained a large deployment of American soldiers in the region. But with popular support the Vietnam War a distant memory, Nixon began a new approach. With two events in 1969, the United States began playing a secondary role in the defense of its allies. The first, a press conference in July on Guam, lent its subject to the name of the strategy: the Guam Doctrine. A speech in November on Vietnamization solidified the new path.

Although the Guam Doctrine also affected Japan, policymakers in Seoul showed greater concern. Following through on his pledge, Nixon removed 20,000 soldiers from the Korean Peninsula (Hong, 2011, 487). President Park Chung Hee's searched for a nuclear solution, seeking help internationally and assistance from ethnic Koreans abroad. Put simply, if the United States would not provide military support, Park felt that a nuclear deterrent would compensate.

By 1974, American policymakers had become suspicious of South Korea's intents. A July 30 telegram from Ambassador Philip Habib warned that "based only on growing independence of Korean attitude toward defense matters and increasing doubts about [the] durability of U.S. commitments, that most senior ROK defense planners desire to obtain capability eventually to produce nuclear weapons" (Habib, 1974). The United States escalated pressure against Seoul the following year and began offering inducements. Washington deployed 600 American nuclear weapons to South Korea (Choi, 2014, 76) and $1.5 billion in military aid (Drezner, 1999, 255). It also threatened to withhold loans, which Seoul needed to keep its fast-rising economy on trajectory (Solingen, 2007). Standing firm also put the construction of its Kori-2 nuclear power plant in jeopardy (Drezner, 1999, 260-261). Meanwhile, Canada stipulated that South Korea had to ratify the NPT as a precondition for receiving a CANDU reactor (Choi, 2014, 75-76). Seoul did so in 1975.

The final case of United States intervention into East Asian allied nuclear pursuit is with Taiwan. Taiwan's major security concern was and remains China. Following Taiwan's loss in the Chinese Civil War, mainland China held a massive military advantage. Taiwan therefore pursued a military relationship with the United States, which the Sino-American Mutual Defense Treaty of 1954 formalized. According to its guidelines, the United States would provide aid and military support in the event of an attack on the island.

Taiwan found the terms satisfactory until China's 1964 weapons test. Leadership felt that their security environment had tightened, and the terms of the Mutual Defense Treaty were insufficient to counteract a potential mainland push for unification. A series of events that followed exacerbated the issue. The United Nations expelled Taiwan in 1971, precipitating its removal from the IAEA and the loss of the associated benefits. Nixon went to China the following year.

Concerns surrounding the Guam Doctrine and Nixon's warmer relations with China caused Taiwan to reconsider its posture. It began seeking foreign assistance, purchasing reactor technologies from Western nations and South Africa. The American embassy recognized Taiwan's progress in 1973 after noticing an incongruence between the quantity of technology transfered and Taiwan's civilian research programs (Hersman and Peters, 2006, 544). One year later, the CIA believed that Taiwan possessed a small scale weapons program (Albright and Gay, 1998, 57), and the U.S. removed its foreign-deployed nuclear weapons from the island.

By 1977, the United States clarified its nonproliferation preference and articulated its demands to Taiwan. By continuing to pursue its program, the Taiwanese leadership risked losing many of their current benefits (Hersman and Peters, 2006, 544). The U.S. already supplied the nuclear fuel necessary to run the island's nuclear power production. It also could weaken its security guarantee. And aside from any potential intervention in a conflict against mainland China, the U.S. could terminate its arms sales. Although Taiwan's formal nuclear program lingered years afterward, the threat marked the end of any significant consideration.

One key difference between Japan and South Korea is Taiwan's relatively weak bargaining position. Whereas Japan and South Korea looked for U.S. assistance, Taiwan's security hinged on maintaining a positive relationship with Washington. This perhaps put the United States in a better bargaining position, giving it more leverage to demand more stringent restrictions on future nuclear growth.

5.2 The Declaration of Iguaçu

Now consider the Declaration of Iguaçu. Both regional powers in South America, Argentina and Brazil faced a period of heightened tensions and increased hostilities from the 1970s. Neither was sure of the other's plans for the continent, and both had military governments during the period. Consequently, each began building its nuclear infrastructure to hedge against the other (Reiss, 1995, 45-52).

Nevertheless, toward the second half of that decade, Argentina and Brazil built a path toward rapprochement. The major breakthrough came in 1985 when Argentinian President Raúl Alfonsín met Brazilian President José Sarney at Foz do Iguaçu, a border city located near Iguaçu Falls. The summit aimed to reduce mistrust, work toward mutual economic development, and integrate each other's civilian nuclear programs to reduce fears of a military escalation. Before leaving, the presidents codified these goals in the the "Declaration of Iguaçu."

Whereas earlier attempts at an agreement had faltered, the Declaration of Iguaçu made steady, if slow, progress. The leaders made annual trips to each other's countries, and nuclear scientists started visiting previously restricted facilities (Reiss, 1995, 55). Trade policies improved and the states enjoyed warmer relations. A more finalized agreement came in 1991 with the Guadalajara Accords, which committed both parties' nuclear energy programs to peaceful applications and established a bilateral inspection regime. Brazil made positive steps by closing its Cachimbo test site in 1990. Meanwhile, Argentina ended its uranium enrichment in 1994.

5.3 Applying the Treatment

These sets of cases share similarities but also have distinct differences. For the East Asian countries, an ally provided the primary inducement to shift nuclear policy. With Argentina and Brazil, rival states built the agreement together, without much help from the outside.[7] As a result, we must make a choice on how to define the treatment variable. Given those differences, we adopt a conservative approach and split the analysis in two. One covers just the treatment of U.S. inducements in East Asia, and the other covers specifically the Declaration of Iguaçu. In each, we insert the other as a control for the estimated counterfactual.

We also must make a few choices about the year of the treatment. The above discussions explain our primary choices of 1972 for Japan, 1975 for South Korea, 1977 for Taiwan, and 1985 for Argentina and Brazil. However, a reasonable reading of some of these cases may yield a different treatment year. For example, Argentina and Brazil finalized their peaceful intentions with the creation of the Brazilian-Argentine Agency for Accounting and Control of Nuclear Materials in 1991. Alternatively, the Japanese Diet adopted Prime Minister Eisaku Satō's Three Non-Nuclear Principles in 1971, the same year Nixon signed the Agreed Minute. We defer these dates to the appendix as robustness checks.

As a more technical note, we shift the coding of the treatment by an additional year. For instance, we begin the effect of the Declaration of Iguaçu in 1986 even though Argentina and Brazil issued it in 1985. This is due to how Smith and Spaniel (2018) construct the ν measure. Under the coding rules—and as is common for any dataset broken down by year—if a country engages in an activity at any point during that period, it is coded as having done it for the observation. Thus, if signing the Declaration caused Argentina to scrap enrichment or reprocessing plans, we would not observe it in the data until the following year.

6 Synthetic Control Results

With the method and cases of interest described, we now turn to a presentation of the results. We begin with the East Asian agreements of the 1970s and then continue to the Declaration of Iguaçu. For each of the agreements, we first present individual effects, graphing comparisons of the observed and counterfactual development paths for each recipient state. Visualizing the individual effects provides an initial check on the

[7] The United States may have cast a shadow on the bargaining process between Argentina and Brazil, possibly tying economic programs from the International Monetary Fund and the World Bank to Argentina halting its weapons exploration.

validity of the results, allowing us to assess whether the observed and counterfactual levels of capability in the pre-treatment period are appropriately close. The state-level visualizations also allow us to detect possible heterogeneity in the effects that would be lost when we move on to consider estimated average effects.

After presenting the state-level results, we move on to consider average effects. Fortunately, the GSC method provides yearly estimates of the average effect of these agreements, along with associated confidence intervals. This means that we can assess not only whether these agreements slow or encourage development, but whether the effect exhibits temporal heterogeneity. In other words, the effect may not be constant, rather growing or diminishing as time passes from the date of agreement. Another possibility is that the effect is never distinguishable from zero, and that the agreements have no meaningful effect on nuclear development. Analyzing the estimated average yearly effects and associated confidence intervals allows us to address these competing possibilities.

6.1 East Asian Agreements

Our estimates for Japan, South Korea, and Taiwan appear in Figures 2, 3, and 4, respectively. In each of these graphs, the solid line plots the observed level of nuclear capacity, while the dashed line indicates the estimated counterfactual produced by the GSC estimates. The vertical line denotes the year of agreement. To avoid extrapolating too far, we focus on a ten-year window after the date of the agreement. For each of the recipients, it is reassuring that the observed and counterfactual levels of proficiency in the pre-treatment period are very close. This suggests that the estimated counterfactuals fit the data generating process.

Focusing on Japan's development as plotted in Figure 2 clarifies the utility of the GSC approach. Looking only at the observed path of Japan's nuclear development, the rate of growth clearly slows in the years after the agreement. Based solely on this information, one might conclude that the agreement caused this plateau in Japan's capacity. However, our estimates indicate that this is not the case. The estimated counterfactual also plateaus in the years immediately following 1972. The implication is clear: Japan would have slowed development even without American intervention. For South Korea, we see a similar plateauing effect in the years following the 1975

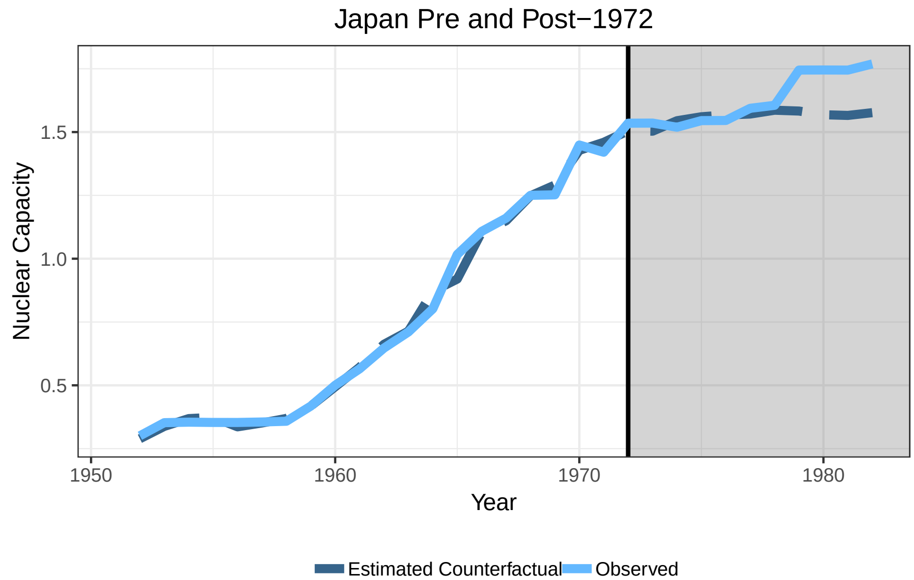

Figure 2: Observed and estimated nuclear capacity for Japan. The counterfactual line lies quite close to the observed level, indicating a minimal effect of the U.S. agreement.

agreement. However, our counterfactual estimate reveals an interesting wrinkle: this plateau only occurs in the aftermath of an *acceleration* of nuclear development. In the long-term, U.S. forward deployments and military aid appear to have reduced South Korea's capability by a modest amount. Taiwan followed a distinct pattern. In the immediate aftermath of the treatment, the observed and counterfactual paths are close. After five years, however, they diverge as Taiwan's development flattens.

While the individual graphs are informative with respect to each particular case, we have yet to address a crucial question: did these agreements limit the average development of nuclear capability among the recipients? Viewing the individual effects simultaneously suggests that these deals did little to constrain nuclear capability in the short term. In each case, the counterfactual estimate lies very close to the observed level of capability in the first five years after each agreement. In the case of South Korea, the observed level even rises slightly above the counterfactual, suggesting that the deal may have contributed to a modest *increase* in South Korea's capability in the short-term. The longer-term effects of these deals is murkier. In the case of Japan, the deal seems to have made little difference, as our counterfactual estimate deviates little from the observed capability, even in the long term. However, the point estimates

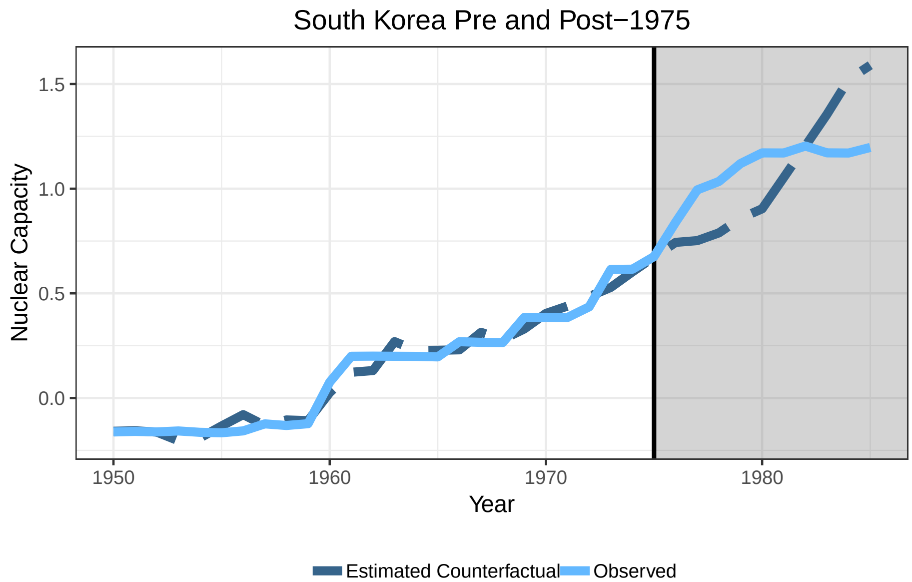

Figure 3: Observed and estimated nuclear capacity for South Korea. The estimates suggest that the agreement may have encouraged South Korea to develop nuclear technology in the immediate aftermath of the agreement, though growth slowed to be overtaken by the counterfactual estimate in subsequent years.

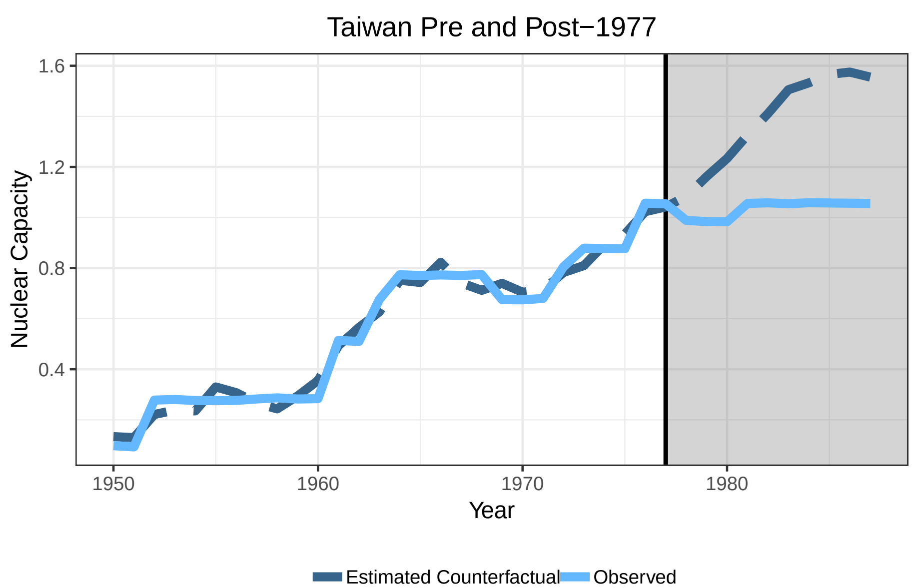

Figure 4: Observed and estimated nuclear capacity for Taiwan. The estimates suggest little change in the immediate aftermath of the clarification of U.S. expectations, with a modest level of constraint in the long-run.

are suggestive of moderate constraining effects in the long-term for South Korea and Taiwan.

Of course, we can only conclude so much by considering these point estimates of individual treatment effects. The individual point estimates tell us nothing about the average impact of these deals, nor do they reveal information about the statistical precision of our findings. Fortunately, the GSC method provides us with estimates of the average effect of these deals, as well as a relevant confidence interval of this average effect. Even more crucially for our purposes, the GSC provides *yearly* estimates of this effect. This allows us to draw conclusions about both the short and long-term impact of the agreements under consideration.

Figure 5 visualizes the average treatment on the treated (ATT) estimates for the East Asian deals of the 1970s. Estimated average effects for each year in the decade post-deal are presented, along with a 95% confidence interval. The point estimates of the ATT in each year align with expectations drawn from the individual plots above: the effects are small, but grow consistently through the post-agreement decade. More telling are the confidence intervals. Strikingly, in every year, we fail to reject the possibility that these agreements have no effect on the development of a recipient's nuclear capability.

While the estimates clearly indicate that we cannot rule out a null effect in each year of a post-agreement decade, this finding is only partially satisfactory. As Rainey (2014) notes, failure to reject the null hypothesis does not necessarily constitute evidence of a substantively negligible effect. Indeed, the simple fact that the confidence intervals include zero does not rule out the possibility that the data are also consistent with large positive or negative effects (Westlake, 1979). Put more directly: a confidence interval may include zero, but it may also include large positive or negative values as well.

Given this, how should we assess the range of effects that our estimates are consistent with? Fortunately, Rainey (2014) provides guidelines for researchers to rule out substantively large effects. The approach is straightforward: researchers first define what constitutes a "meaningful" effect, then check to see if such effects lie within the bounds of a 95% confidence interval. If the confidence interval does not include the effect size defined as substantively meaningful, then we can confidently conclude that the data are not consistent with large effects.

How do we define a substantively meaningful effect? We let the data guide us here.

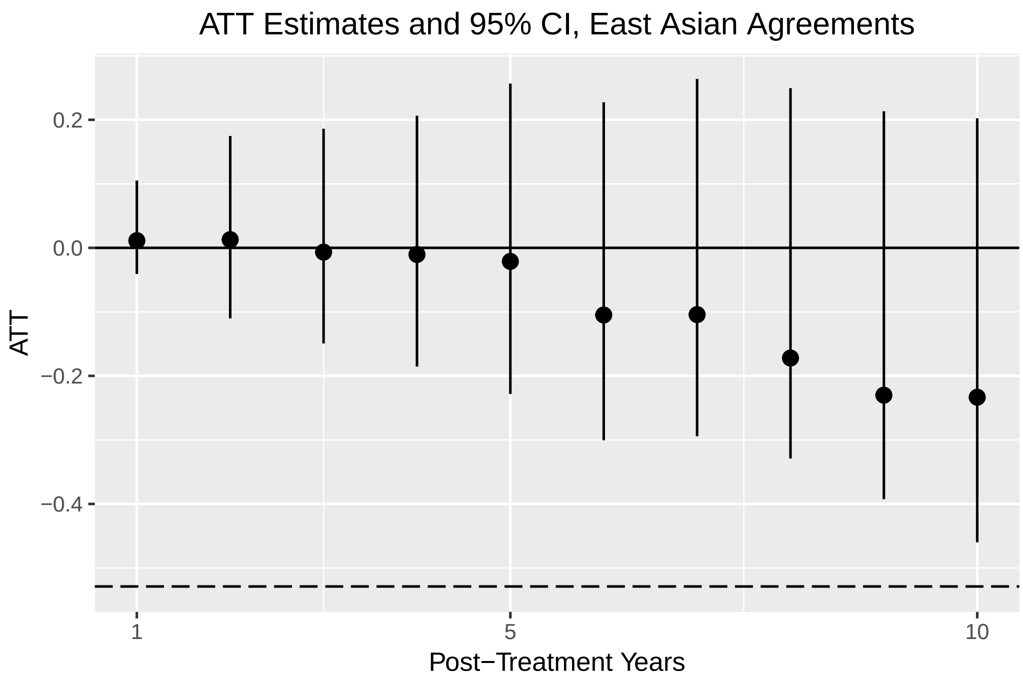

Figure 5: Yearly ATT estimates for East Asian agreements. The dotted line corresponds to the average difference in nuclear capacity in the first year of pursuit versus the year of successful test for all states possessing weapons as of 2016. Note that in each year, the confidence intervals include zero. Further, the confidence intervals fail to include the dotted line throughout the entire post-agreement decade.

Using the nuclear capability scores of Smith and Spaniel (2018), we gather data on the historical nuclear capability of all states that have successfully tested nuclear weapons as of 2016. Using this subset of capability scores, we calculate the average difference in each state's ν score at the time they began to actively pursue development of a nuclear weapon to their ν score at the time of their first successful test. We then take the average of these differences. The resulting number is approximately 0.53, which we can interpret as the average increase in a state's nuclear capability from the time it actively begins to pursue a weapon to the time of its first successful test. This amount gives us an idea of how much technological constraint is necessary, on average, to revert a state to its pre-pursuit level. Accordingly, we use this value as a benchmark for substantive significance.

Given that benchmark, we return to Figure 5. The dotted horizontal line marks the cutoff for a substantively meaningful effect. Following the guidelines above, for the first nine years after an agreement, we can rule out a large effect. In fact, for many years we can rule out substantially smaller effects. The interval does not include an effect half as large as the threshold until six years post-agreement. Only after a decade can we rule out a sizable effect, as the confidence interval barely overlaps the threshold at that point. This means that the largest possible effect consistent with the data after a decade barely meets the average level of development among states that have sought and obtained a weapon. We take this as evidence that the impact of these deals on the development of recipient nuclear capability in the first decade after an agreement is relatively small.

Summing up, our analysis of the East Asian agreements suggests two preliminary conclusions. First, the influence of these agreements is statistically indistinguishable from zero. Second, while our estimates grow less precise as the time after treatment grows, we can rule out substantively meaningful effects throughout the decade after agreement.

6.2 The Declaration of Iguaçu

What remains to be seen is whether these patterns hold for Argentina and Brazil. We plot the state-level observed and counterfactual paths of development in Figures 6 and 7. A first glance at the individual point estimates for Argentina and Brazil in the wake

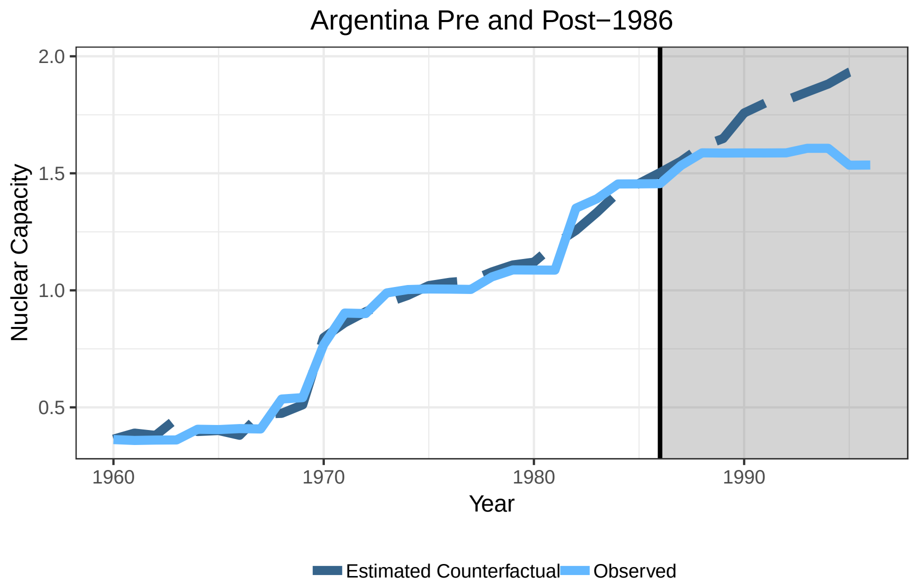

Figure 6: Observed and estimated nuclear capacity for Argentina.

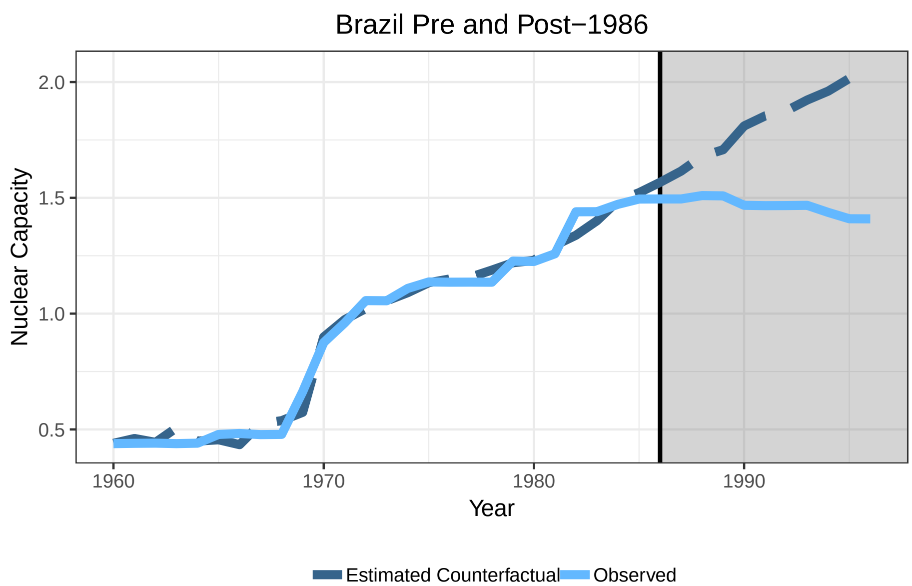

Figure 7: Observed and estimated nuclear capacity for Brazil.

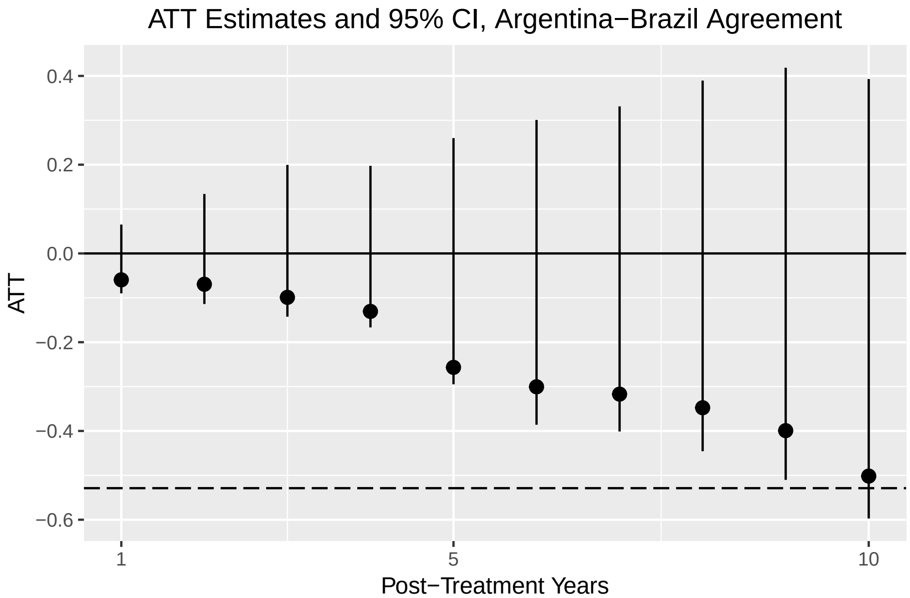

Figure 8: Yearly ATT estimates for Declaration of Iguaçu. The dotted line corresponds to the average difference in nuclear capacity in the first year of pursuit versus the year of successful test for all states possessing weapons as of 2016. Note that in each year, the confidence intervals include zero. Further, the confidence intervals fail to include the dotted line except for in the final year of the post-agreement decade.

of the agreement is suggestive of a consistent pattern. In the first five years after the agreement, Argentina's development lies slightly above its counterfactual, while Brazil's lies barely below. Much like the East Asian allied agreements, the bilateral agreement between Brazil and Argentina seems upon first blush to have made little difference in the short term. However, in the long run the point estimates in each case deviate noticeably from the counterfactual.

But is this a significant effect? Without addressing the uncertainty in our estimates, concluding that these effects are substantial is premature. As with the East Asian agreements, we present yearly point estimates and confidence intervals for the ATT. A similar pattern emerges. While the estimated ATT grows over time, we fail to reject the null hypothesis of no effect in every year throughout the post-agreement decade. Particularly striking are the near-zero point estimates in the first four years after the initial agreement. The point estimates only diverge from zero beginning in the fifth year, growing steadily for the remainder of the post-treatment decade. While the point estimates do grow over time, they are never distinguishable from zero, as the confidence intervals indicate.

In addition to this, we perform the same exercise as before, assessing the degree to which we can rule out substantively large effects as statistically implausible. Similar to with the East Asian agreements, the confidence intervals indicate that in nearly every year in the post-agreement decade, the Declaration of Iguaçu is inconsistent with anything but a negligible effect. Only in the ninth and tenth years do the confidence intervals include a meaningful effect. Even then, the interval barely covers the threshold.

With the results for both the East Asian agreements and the Declaration of Iguaçu on the table, some patterns begin to emerge. First, in the case of both agreements, the effects are indistinguishable from zero. However, the pattern is more nuanced. We can even more conclusively rule out large effects in the short-term. In particular, looking to Figures 5 and 8 indicates that our estimates are quite precise within the first five years post-agreement. Within this window, we can rule out all but extremely small effects. Moving out of this window, our estimates grow less precise, but are still only consistent with effects of relatively small substantive magnitude. Only towards the end of the post-treatment decade are we unable to convincingly reject the possibility of meaningful effects.

Finally, it is important to note that our inability to conclusively reject the possibility of meaningful effects in the long-term does not provide evidence that there is a substantively meaningful effect. While we cannot rule out a substantively meaningful effect ten years post-agreement, we also cannot rule out a null effect or a small positive effect. Our estimate of the long-term ATT is noisy. This is due to the GSC's accounting for the uncertainty about unobserved heterogeneity in the long-term. As the counterfactual estimates extrapolate further from the year of treatment, so too does uncertainty about these estimates.

6.3 Robustness

As always, it is important to assess whether our results hold under reasonable alternative specification. These key robustness checks address the possible ambiguities in the date of treatment. We recode the year of treatment for Japan to 1971 as it was in this year that the Japanese Diet adopted Three Non-Nuclear principles, as well as Nixon's signature of the Agreed Minute. We recode the year of treatment for Taiwan to 1979 to account for the clarification of U.S. policy that accompanied the passage of the Taiwan Relations Act. While the initial, informal agreement between Argentina and Brazil was formed in 1985, an official finalized agreement, the Guadalajara accords, occurred in 1991. To guard against the possibility that the formal agreement had an impact while the initial declaration did not, we recode the year of treatment from 1985 to 1991. In each case, recoding the date of treatment does not alter the substantive findings of our initial specification.

Taiwan's status provides another source of complication. The measure of capability that we adopt includes activities which Taiwan is precluded from participating in. We address this difficulty in two ways. First, we totally exclude Taiwan from the analysis, estimating the effect of U.S. agreements with Japan and South Korea. Second, we include Taiwan but reestimate our capability score, including all variables except for the IAEA research variables. In each case, our substantive results remain unchanged.

While there are a number of issues with traditional approaches such as fixed-effects regression in the context of this study, we also performed a simple fixed-effects regression for comparison. The results are qualitatively similar to those from our application of the GSC method. However, the fixed-effects results do not allow for us to analyze the year-by-year effects of these agreements, and they are still plagued by concerns over time-varying unit-specific confounding, so we relegate these results to the appendix.[8]

7 Conclusion

In this paper, we have evaluated whether nuclear agreements lead to fewer proliferation outcomes by constraining the development of nuclear technology. To address this question, we first sought out a suitable measure of technological development. We identified the scores of Smith and Spaniel (2018) as appropriate, and extended them through the collection of additional data on activities indicative of nuclear development. This data collection and generation exercise produced a measure suitable for evaluating the technological development processes we wish to study. With this measure, we still faced the difficulty of constructing the appropriate counterfactual. The GSC provides a desirable

[8]Results from all aforementioned robustness checks appear in section 2 of the supplementary information, beginning on page 17.

solution, and we therefore estimated the effect of nuclear agreements with it.

Our estimates indicate that agreements have little constraining effect. Any reduction in proficiency is statistically indistinguishable from zero. Put differently, we cannot rule out the possibility that these agreements do not constrain recipient development of nuclear technology. Going a step farther, we assessed the uncertainty in our estimates, analyzing what range of effects we *can* rule out. In the first decade after an agreement, our estimates are precise enough to eliminate all but substantively small effects. Comparing to documented cases of proliferation, our estimates suggest that the maximum level of constraint plausible given our estimates would not have stopped proliferation on average.

These results have important implications for policymakers. If nonproliferation agreements have an impact, our findings suggest that the mechanism is not one of technological constraint. Following the literature, this implies that concessions are the primary mechanism through which these deals prevent proliferation, especially in the short term. Continued technological progress does not indicate a failed agreement. Rather, states may continue to develop, while concessions keep them from taking the final steps towards a bomb. Our estimates suggest that this pattern is consistent with successful nonproliferation agreements.

Finally, while we can confidently rule out meaningful constraining effects in the short and medium term, our estimates of the long-term impact of such agreements are noisier. Thus, if these deals do have any efficacy at constraining the development of nuclear technology, such effects will only be borne out in the long term. This means that the success of efforts to constrain technology should not be judged based on short term changes. We suggest caution in declaring deals that do not result in immediate reversal a failure. Because of the importance of these long-run effects, future work should aim to develop tools to more precisely assess the long-term impact of these deals.

References

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

Albright, David and Corey Gay. 1998. "Taiwan: Nuclear nightmare averted." *Bulletin of the Atomic Scientists* 54(1):54–60.

Bai, Jushan. 2009. "Panel data models with interactive fixed effects." *Econometrica* 77(4):1229–1279.

Bleek, Philipp C. 2017. *When Did (and Didn't) States Proliferate?: Chronicling the Spread of Nuclear Weapons*. Belfer Center for Science and International Affairs.

Bleek, Philipp C and Eric B Lorber. 2014. "Security guarantees and allied nuclear proliferation." *Journal of Conflict Resolution* 58(3):429–454.

Brown, Cameron S, Christopher J Fariss and R Blake McMahon. 2016. "Recouping after coup-proofing: Compromised military effectiveness and strategic substitution." *International Interactions* 42(1):1–30.

Brown, Robert L and Jeffrey M Kaplow. 2014. "Talking peace, making weapons: IAEA technical cooperation and nuclear proliferation." *Journal of Conflict Resolution* 58(3):402–428.

Bundy, McGeorge. 1988. *Danger and survival: Choices about the bomb in the first fifty years*. New York: Random House.

Campbell, Kurt M and Tsuyoshi Sunohara. 2004. The Nuclear Tipping Point: Prospects for a World of Many Nuclear Weapons States. In *The Nuclear Tipping Point: Why States Reconsider Their Nuclear Choices*, ed. Kurt M Campbell, Robert J Einhorn and Mitchell B Reiss. Washington, DC: Brookings Institution Press pp. 218–253.

Choi, Lyong. 2014. "The first nuclear crisis in the Korean Peninsula, 1975–76." *Cold War History* 14(1):71–90.

Cirincione, Joseph, Jon B Wolfsthal and Miriam Rajkumar. 2011. *Deadly arsenals: Nuclear, biological, and chemical threats*. Washington DC: Carnegie Endowment.

Colaresi, Michael P. 2014. *Democracy Declassified: The Secrecy Dilemma in National Security*. New York: Oxford University Press.

Debs, Alexandre and Nuno P Monteiro. 2016. *Nuclear politics: The strategic causes of proliferation*. Vol. 142 New York: Cambridge University Press.

Drezner, Daniel W. 1999. *The sanctions paradox: Economic statecraft and international relations*. New York: Cambridge University Press.

Ellis, Jason D. 2001. *Defense by other means: The Politics of US-NIS threat reduction and Nuclear Security Cooperation*. Westport: Praeger.

Fuhrmann, Matthew. 2009. "Spreading temptation: proliferation and peaceful nuclear cooperation agreements." *International Security* 34(1):7–41.

Fuhrmann, Matthew and Benjamin Tkach. 2015. "Almost nuclear: introducing the nuclear latency dataset." *Conflict Management and Peace Science* 32(4):443–461.

Habib, Philip. 1974. "Record Group 59: Korean Accession to NPT." https://nsarchive2.gwu.edu//dc.html?doc=3513489-Document-01-U-S-Embassy-in-Republic-of-Korea.

Hersman, Rebecca KC and Robert Peters. 2006. "Nuclear U-turns: Learning from South Korean and Taiwanese rollback." *Nonproliferation Review* 13(3):539–553.

Hong, Sung Gul. 2011. The search for deterrence: Park's nuclear option. In *The Park Chung Hee era: The transformation of South Korea*, ed. Byung-Kook Kim and Ezra F Vogel. Cambridge: Harvard University Press pp. 483–512.

Jo, Dong-Joon and Erik Gartzke. 2007. "Determinants of nuclear weapons proliferation." *Journal of Conflict Resolution* 51(1):167–194.

Kroenig, Matthew. 2009. "Importing the bomb: Sensitive nuclear assistance and nuclear proliferation." *Journal of Conflict Resolution* 53(2):161–180.

Leeds, Brett, Jeffrey Ritter, Sara Mitchell and Andrew Long. 2002. "Alliance treaty obligations and provisions, 1815-1944." *International Interactions* 28(3):237–260.

Levite, Ariel E. 2003. “Never say never again: nuclear reversal revisited.” *International Security* 27(3):59–88.

Meyer, Stephen M. 1986. *The dynamics of nuclear proliferation*. Chicago: Chicago University Press.

Oberdorfer, Don. 2001. *The two Koreas: A contemporary history*. New York: Basic Books.

Paul, TV. 2000. *Power versus prudence: Why nations forgo nuclear weapons*. Montreal: McGill-Queen’s University Press.

Rainey, Carlisle. 2014. “Arguing for a negligible effect.” *American Journal of Political Science* 58(4):1083–1091.

Reiss, Mitchell. 1988. *Without the bomb: The politics of nuclear nonproliferation*. New York: Columbia University Press.

Reiss, Mitchell. 1995. *Bridled ambition: Why countries constrain their nuclear capabilities*. Baltimore: Johns Hopkins University Press.

Roehrig, Terence. 2017. *Japan, South Korea, and the United States Nuclear Umbrella: Deterrence After the Cold War*. New York: Columbia University Press.

Singh, Sonali and Christopher R Way. 2004. “The correlates of nuclear proliferation: A quantitative test.” *Journal of Conflict Resolution* 48(6):859–885.

Smith, Bradley C and William Spaniel. 2018. “Introducing ν-CLEAR: a latent variable approach to measuring nuclear proficiency.” *Forthcoming, Conflict Management and Peace Science* .

Solingen, Etel. 1994. “The political economy of nuclear restraint.” *International Security* 19(2):126–169.

Solingen, Etel. 2007. *Nuclear logics: contrasting paths in East Asia and the Middle East*. Princeton: Princeton University Press.

Spaniel, William. 2015. “Arms negotiation, war exhaustion, and the credibility of preventive war.” *International Interactions* 41(5):832–856.

Spaniel, William. 2019. *Bargaining over the bomb: The successes and failures of nuclear negotiations*. New York: Cambridge University Press.

Stinnett, Douglas M, Jaroslav Tir, Paul F Diehl, Philip Schafer and Charles Gochman. 2002. "The Correlates of War Project direct contiguity data, version 3.0." *Conflict Management and Peace Science* 19(2):59–67.

Volpe, Tristan A. 2017. "Atomic leverage: Compellence with nuclear latency." *Security Studies* 26(3):517–544.

Westlake, Wilfred J. 1979. "Statistical aspects of comparative bioavailability trials." *Biometrics* pp. 273–280.

Xu, Yiqing. 2017. "Generalized synthetic control method: Causal inference with interactive fixed effects models." *Political Analysis* 25(1):57–76.