

When are International Institutions Effective? The Impact of Domestic Veto Players on Compliance with WTO Rulings

Lauren Peritz*

January 1, 2015

DRAFT – Please do not circulate.

Abstract

When do international institutions promote economic cooperation among countries? The World Trade Organization (WTO) is central to the multilateral trade regime and a benchmark for international dispute resolution. Yet there are few attempts to measure its effectiveness in restoring trade cooperation. This paper uses WTO disputes to examine the impact of domestic politics in the defendant country on compliance with adverse legal rulings. I build a novel data set on compliance. Using the method of synthetic case control, I estimate the effect of adverse rulings on trade flows between disputant countries using product-level time-series trade data. I infer the defendant complied if trade flows increased after the dispute, relative to estimated levels that would have occurred in the absence of the ruling. The estimates show compliance problems are both widespread and systematically linked to domestic politics. Domestic political divisions—measured by veto players—hinder compliance.

*Department of Political Science, University of California, Davis, lperitz@ucdavis.edu.

1 Introduction

The World Trade Organization (WTO) Dispute Settlement Mechanism, created in 1995, promised a new institutional solution for trade disputes between countries and an advance in international economic cooperation. Compared to its predecessor, the General Agreement on Tariffs and Trade (GATT), the Dispute Settlement Mechanism provides a standardized process through which lawsuits are conducted, verdicts are delivered, and implementation is monitored. Nearly all WTO rulings identify infringements and require the respondent government modify its policies. Yet only some governments comply readily; others delay for years or defy the WTO altogether. What explains this variation?

For example, in 1995, the European Union (EU) sued Japan over an alcohol tax that favored domestic producers and inflated the price of comparable foreign products. Two years later, the EU sued Korea on nearly the same grounds. In both cases, the WTO ruled against the respondent governments, ordering them to modify their tax codes. The Korean government, with its recently consolidated political power, promptly complied. But Japan, with its domestic partisan discord, did not. Do features of these countries' domestic politics explain the divergent outcomes?

Many scholars argue that democracy facilitates international economic cooperation (e.g. Martin 2000; Mansfield, Milner, and Rosendorff 2002). Institutional divisions of power and partisan opposition in government—hallmarks of a functional democracy—create multiple veto points (Tsebelis, 1995). Multiple veto points narrow the set of international agreements that can be ratified, since more domestic actors must coordinate and consent to the agreement (Milner and Rosendorff, 1997).¹ This makes treaties enacted by democracies more credible commitments about governments' future behavior than those by non-democratic regimes. Veto players can lock in cooperative policies and make it difficult for governments

¹Some studies, e.g. Mansfield, Milner, and Pevehouse (2007), argue that veto players reduce the probability of forming a trade agreement but do not address whether the governments follow through on their commitments.

to renege on their international promises.² However, this lock-in mechanism has a second side. Contrary to a perspective prevalent in the literature, I argue veto players can sometimes hinder international cooperation.

Democratic leaders can and do break international commitments when the domestic pressure to do so is acute. For example, a government might grant farmers' request for import restrictions amid sharply declining wheat prices. Or, faced with a forthcoming election, a leader might introduce a subsidy that curries favor from key interest groups. Many scholars argue that institutions that allow leaders to respond to such pressure by temporarily violating their commitments are more stable (e.g. Staiger and Bagwell 1999; Rosendorff and Milner 2001; Bown 2002; Rosendorff 2005). By permitting some violations, flexible institutions allow members to manage temporary economic shocks or political circumstances without abandoning the institution altogether.

However, just as multiple veto players can lock in international commitments, they can also lock in violations of those commitments. Once an initial violation occurs, a government with more veto players is less likely to return to cooperation. For these cases, flexibility mechanisms may be less successful in generating long-term stability. In WTO disputes, a losing respondent government with many veto players is less likely to comply with the legal ruling and correct the initial breach. Defining features of democratic politics can actually obstruct international cooperation.

I evaluate government compliance with adverse WTO rulings by estimating an approximate causal effect of the rulings on product-level trade flows between the disputing countries. True causal effects are difficult to identify because they rely on specifying the correct counterfactual (Rubin, 2005). Using the method of synthetic case control (Abadie and Gardeazabal, 2003; Abadie, Diamond, and Hainmueller, 2010, 2014), I estimate a plausible and precise counterfactual that represents what product-level trade would have been without the WTO's legal verdict. It is built from the complainant government's exports of the disputed product

²And democratic leaders, beholden to voting publics, might be wary about breaking international agreements, lest they generate audience costs (Tomz, 2007; Dai, 2005).

to other countries not engaged in the dispute. The estimated counterfactual is matched to the actual trends in trade flows during the years leading up to the dispute. For every dispute, I identify the affected products and examine trade flows of just those products. Then by comparing actual trade to the estimated counterfactual after the dispute, I determine whether trade flows in the disputed products increased in relative terms, beyond ordinary fluctuations. Relative deviations in trade reveal the approximate causal effect of an adverse WTO ruling, with a positive deviation indicating the respondent government complied. I measure compliance in all 125 WTO disputes between 1995 and 2011 that received an adverse ruling on an import-restricting trade policy. This methodological approach offers an advance over matching—which does not account for temporal trends—and over traditional regression analysis which would entail many thousands of observations of product-level trade flows but merely 125 cases subjected to the intervention, an adverse legal ruling.

Using this novel metric, I show compliance patterns at the WTO reflect the cross-cutting impact of democratic politics. Some rulings from the WTO Dispute Settlement Mechanism (DSM) generate marked increases in trade indicative of compliance while others have no detectable effect. The variation is systematically linked to domestic politics. As the number of veto players in the respondent government increases, compliance with adverse rulings becomes less likely. Veto players appear to lock in initial violations and obstruct compliance, even when international pressure to capitulate is substantial. The result is large and robust, controlling for other potentially confounding factors such as size of the complainant's and respondents' economies and the extent of the adverse legal ruling. While prominent and influential, the DSM is not insulated from the domestic political controversies that lead to trade disputes in the first place.

Complementing the institutional design literature, this paper offers an empirically rigorous appraisal of institutional effects. Some scholars have puzzled over the seemingly insignificant impact of WTO disputes on trade (Chaudoin, Kucik, and Pelc, 2013). Measuring compliance through bilateral, product-level trade flows, I show there is substantial variation

in dispute outcomes that is obscured when one examines the compliance record on average. This heterogeneity suggests that international dispute settlement may not be uniformly effective or ineffective. Scholars continue to disagree about the impact of international institutions on cooperation (e.g. Mearsheimer 1994; Goldstein and Martin 2000; Hafner-Burton, von Stein, and Gartzke 2008; Voeten 2013). Inferential obstacles are partly to blame. When states appear to comply—for example by removing trade barriers—are they doing so in response to their international commitments? Or are they implementing policies they would have selected anyway? Because countries voluntarily enter international agreements they intend to follow, it is little surprise the compliance record *appears* to be good (Downs, Rocke, and Barsoom, 1996; von Stein, 2005). I weigh in on the debate by measuring *de facto* compliance in a precisely defined and particularly revealing domain: WTO disputes with adverse rulings. Construed this way, my evidence shows that WTO rulings can indeed have significant effects.

The following section situates my study in existing literature and describes dispute settlement at the WTO. The third section presents a theory of domestic politics and compliance with international institutions. Fourth, I discuss obstacles to assessing compliance and my analytic solution. In the final sections, I present statistical results and discuss the findings.

2 Cooperation, Compliance, and the WTO

International institutions aim to improve cooperation among countries by solving collective action problems (e.g. Keohane 1982; Stein 1990). In international trade, countries have a unilateral incentive to impose trade barriers but a mutual interest in exchanging goods and services freely. Trade agreements aim to move from a suboptimal outcome of high trade barriers to a more efficient one. Yet international institutions struggle with compliance. Because they represent voluntary agreements, they have difficulty ensuring governments abide by treaty terms. Sometimes noncompliance is the result of incomplete information.

For example, a violation may reflect ambiguous or imprecise legal obligations (Chayes and Chayes, 1993). Other times, governments are aware of obligations but behave opportunistically. Compliance problems have broad effects. For example, when governments violate trade agreements, they undermine the collaborative equilibrium: other states then prefer to impose trade protection. Noncompliance by a few leads to a worse outcome for many.

States often design international institutions with flexibility mechanisms in order to mitigate compliance problems (Abbott and Snidal, 1998; Koremenos, Lipson, and Snidal, 2001; Carrubba, 2005). By permitting short-term violations under certain circumstances, flexibility mechanisms can enhance long-term compliance if they prevent countries from leaving international agreements (e.g. Staiger and Bagwell 1999; Bown 2002). Dispute settlement mechanisms are institutional features that some scholars believe increase flexibility and thereby improve stability (e.g. Rosendorff 2005; Kucik and Reinhardt 2008; Pelc 2009).

Scholars have identified two forms of compliance in the context of international relations (Simmons, 1998). First-order compliance is when a state abides by its substantive legal obligations. In the WTO, a government complies in the first sense if its trade policy and practice reflect its commitment to low trade barriers. When a state fails to comply with treaty terms and a dispute arises, it is often required to correct the violation. Second-order compliance is when the state adjusts its policy and practice after an initial violation, often in response to international litigation. A government complies in the second sense if it loses a dispute and corrects the violation, e.g. removes an illegal trade barrier.

Governments that have been harmed by the actions of another state can use international dispute settlement mechanisms to address their grievances, especially when there are domestic political benefits from doing so (Goldstein and Steinberg 2008; Davis 2012; Allee and Huth 2006). Domestic politics factors into a government's decision to *initiate* litigation. It should also influence a government's *response* when it is convicted of violating international commitments. Many studies focus on dispute initiation; few consider compliance *ex*

post. The research that does largely ignores domestic politics, instead emphasizing important features of the international system (e.g. Bown 2004*a*).

The WTO is an ideal venue to examine compliance. First, it is the cornerstone of the multilateral trade regime. If any international institution is to have a demonstrable effect on its members' behavior, the WTO is it (Rose, 2004; Goldstein, Rivers, and Tomz, 2007; Subramanian and Wei, 2007).³ Second, its Dispute Settlement Mechanism is seen as the exemplar for international dispute resolution (Hudec, 1999; Busch and Reinhardt, 2000). With widespread interest among political scientists, lawyers, and economists alike, there is an implicit assumption that WTO legal verdicts matter. Yet scholars are only beginning to collect evidence of whether the rulings restore trade. Some find a large positive effect of WTO disputes⁴ while other recent work finds modest or no discernible trade gains (e.g. Bown and Ruta (2015); Chaudoin, Kucik, and Pelc (2013)).⁵ While these studies have made significant contributions, they do not account for the diversity in domestic political conditions that disputant governments face.

The primary objective of the WTO is to promote trade among countries by constraining the policies of member states. WTO disputes arise when one government alleges another has imposed trade barriers that violate treaty terms. Trade barriers differ in form. For example, subsidies lower production costs for domestic industries giving them a competitive advantage and reducing demand for imported goods. Tariffs inflate the price of imports, again improving domestic industry's competitive advantage. Quantitative restrictions place a limit on the volume of imported products. Regardless of form, these policies limit imports into the respondent country.

³Not all studies find a positive impact on trade and some find conditional effects, e.g. Gowa and Kim (2005).

⁴Bechtel and Sattler (2015) find a positive effect using more aggregated data and matching algorithms which, unlike the synthetic control method used here, do not explicitly account for temporal trends.

⁵Bown and Ruta 2015 emphasize the timing and types of violations that prompt WTO disputes Chaudoin, Kucik, and Pelc 2013 examine total imports into the respondent rather than dyadic import shares.

The DSM aims to restore trade cooperation. As the WTO itself states, “the priority is for the losing defendant to bring its policy into line with the ruling.”⁶ This distinguishes the DSM from other institutions which aim to punish violators, levy fines, etc. Adverse rulings require the respondent remove trade barriers, prompting an increase in imports. These trade flows can be observed and measured, making the WTO an excellent setting to study compliance.

WTO disputes lend special insight into institutional effectiveness. As Downs, Rocke, and Barsoom (1996) observed, “we do not know what a high compliance rate really implies. Does it mean that even in the absence of enforcement states will comply with any agreement...or does it mean that states only make agreements that do not require much enforcement?” (383). Ordinarily, it is problematic to infer that an institution is effective simply because states comply with its rules (Martin, 2013). But by focusing on situations where states do not want to cooperate and nevertheless do—second-order compliance—we *can* draw inferences about effectiveness. This makes WTO disputes revealing. By virtue of being sued and found guilty of violations, we know respondent governments prefer noncompliance. When the respondent corrects the violations in spite of this preference, we can infer the influence of the institution.

Besides the disputants, other states often have a stake in WTO disputes. The respondent’s trade policy might harm them in a similar way as it does the complainant. Or the case may establish standards akin to legal precedent (Pelc, 2014). The DSM allows states to enter the proceedings as “third parties”. Third party governments usually support the complainant. They may generate international pressure to comply by increasing the publicity of the dispute, monitoring to ensure the ruling is implemented faithfully, and making enforcement more likely.

WTO disputes have several stages. The first is formal negotiation. At least 60% of disputes have been settled early through bilateral agreements. When early settlement is

⁶“What is the WTO? Understanding the WTO,” available: <http://www.wto.org>.

impossible, the dispute moves into litigation. A panel evaluates the case on the merits and issues a ruling called the “panel report.” While rulings consist of multiple claims, they usually require the respondent change some aspect of its trade policy. The vast majority (94%) favor the complainant on at least one legal claim. The respondent is given a “reasonable period of time” to correct the violation. If a respondent fails to make required policy changes before the deadline, it has not complied.

The WTO has limited enforcement options. If the implementation deadline passes and the losing respondent fails to take adequate action, the complainant may initiate compliance proceedings. Complainants can only punish noncompliance after all other legal remedies are exhausted. When do governments implement adverse WTO rulings in a timely manner? When do they fail to comply?

3 A Theory of Domestic Veto Players and Compliance

3.1 Domestic Preferences, Power, and Institutions

Within a country, political actors have divergent preferences over international affairs. Some prefer trade protection; others trade liberalization. Many interest groups represent import-competing industries and prefer protectionist policies. Trade liberalization poses a threat to these groups by expanding imports, increasing competition, and driving down prices of the goods they produce. Voters sometimes have preferences over trade policy, but compared to interest groups, their positions are not well-formed or informed (Hiscox, 2006). As consumers, voters might prefer the availability of lower-cost goods that trade liberalization brings but they face well-known problems in mobilizing. Interest groups and voters impose political pressure on politicians.

Politicians vary in their own inherent preferences as well as in the political pressure they experience. Politicians in different branches of government experience domestic pressure in various ways, depending on the composition of their constituencies. Legislators often

prefer more protectionist trade policies than prime ministers or presidents do (Lohmann and O'Halloran, 1994). Legislators respond to local constituencies who benefit from protection for specific industries.⁷ By contrast, executives, with their broad constituencies, are more sensitive to the aggregate benefits of international trade cooperation.

Exogenous economic and political shocks can change the relative power of different interest groups and voters. An economic downturn that acutely harms an industry may lead groups to pressure their politicians for policies that benefit them. Typically, these groups are highly informed about whether current trade policy is helping or hurting their business. Subjected to increased pressure from mobilized industry groups, politicians often shift their preferences. Similarly, political shocks can shift different groups' relative power. Political responsiveness often changes for reasons unrelated to trade policy. For example, an upcoming election can make politicians sensitive to preferences in a key electoral district. Under these conditions, politicians are more likely to implement trade protection policies that provoke international disputes.⁸

Countries vary in the degree to which political authority is concentrated within government. In governments with substantial divisions in authority, it is harder to change existing policies because more political actors can block change. In part, the divisions arise from domestic political institutions. Checks and balances make policy change difficult. Divisions also arise from party politics. The partisan composition of the government can make it more or less difficult to change policies. When the legislature features a relatively strong opposition party or many opposition parties, there are significant obstacles to policy change. Taken together, domestic institutional and partisan divisions form *veto points* in government. For example, if one branch of government is held by a first party and another branch is dominated by a second party, the government has especially divisive domestic politics—many

⁷Export-oriented industries sometimes push policy in the opposite direction (Gilligan, 1997).

⁸For example, when the Canadian magazine industry suffered unprecedented decline, industry officials urged parliament to impose a controversial tax to stave off foreign competition. See: "Magazine Industry Urges Laws to Stem Foreign Competition" *The Toronto Star*, February 9, 1993.

veto points. As veto points increase, political authority becomes more fragmented and policy becomes more difficult to change (Tsebelis, 1995; Henisz, 2000; Tsebelis, 2003).

Veto points determine how responsive governments are to shifts in preferences and power. When a government has few veto points, only a few political actors must coordinate and it is relatively easy to change policies. So minor shifts in industry preferences or political power can have a potent influence on policy. Conversely, when government has many veto points, policy change is quite difficult because it requires coordination among many political actors. Only major shifts in preferences or power can alter policy.

3.2 International Cooperation and Compliance

Domestic veto players constrain governments when they join international agreements. Divisions make it more difficult for a government to join an international agreement because “[m]ultiple veto players...narrow the set of [treaty] proposals that can be domestically ratified” (Simmons, 2009, p. 69). When government becomes more divided, treaties are less likely to be ratified because there are more domestic actors willing to block the agreement. Domestic veto players can limit international cooperation in many domains. Because trade preferences of different domestic groups often vary widely, trade agreements are particularly susceptible to such divisions. Divided domestic politics tends to constrain trade liberalization (Milner and Rosendorff, 1996, 1997).

After a country has joined a trade agreement, veto players are thought to facilitate compliance with the agreement. Governments with substantial divisions in authority may be less responsive to political pressure and industry demands for new trade protection. In these governments, interest groups will have to persuade many political actors to support a new policy. Interest groups often pressure politicians for trade protection that benefits their targeted industry, sector, or region. Sometimes the protection they demand violates trade agreements. Veto players make it more difficult for a government to reverse existing trade policies. Following this logic, trade agreements formed by democratic governments

are thought to represent more credible commitments about future trade policy (Mansfield, Milner, and Rosendorff, 2002). By locking cooperative policies in place, veto players should improve a government's first-order compliance.

However, empirical obstacles make it difficult to test the impact of domestic politics on first-order compliance. First, it is difficult to observe when a leader is tempted to violate international commitments because domestic political pressure is idiosyncratic. Second, governments (and researchers) may be uncertain about what constitutes a violation of treaty terms. WTO disputes often focus on whether particular trade measures are legal or illegal; the answer is often nonobvious. Third, even if one could perfectly differentiate compliant trade policies from violations, not all trade violations are observed—only those which trigger a dispute.⁹ Governments impose trade barriers that go uncontested and these violations are hard to identify. As a result, an empirical analysis of first-order compliance may generate biased conclusions about the impact of domestic veto players.

Dispute settlement mechanisms make second-order compliance easier to examine than first-order compliance. International legal disputes reveal information about the initial violation and resolution. Although domestic political pressure is idiosyncratic, litigation publicizes these interests, revealing the domestic factors that lead to the violation. Litigation generates rulings that highlight infractions and explicitly differentiate legal from illegal trade measures. And the respondent government that loses a dispute is forced to respond. Some will adjust their policies and behavior to implement the legal ruling; others will not. Second-order compliance and defiance are both observable. Finally, by drawing attention to government behavior and specifying the requirements for implementation, DSMs provide criteria against which one can evaluate second-order compliance.¹⁰

⁹Some studies examine governments' choices to impose legal versus illegal protection that provokes a dispute (Bown, 2004*b*) but these measurable instances represent a small subset of cases.

¹⁰For example, a dispute between the US and Canada (DS31: Canada–Periodicals) highlighted political pressure from Canada's waning publishing industry. The WTO ruled that Canada's magazine tax violated treaty terms, establishing fault. And the ruling laid out requirements for compliance.

Just as veto players constrain a government's ability to form and violate trade agreements, they make it more difficult to restore cooperation, after a violation has occurred. When a government fails to comply with its substantive international commitments, domestic political divisions can lock in the violation. Sometimes, reversing the violation may require nullifying an old law or passing a new one. This is common for regulatory trade protection. Political actors who benefit from the trade barrier have incentive to block policy change. Veto players in domestic government create points where policy change can be obstructed. For example, some legislators may block legislation that implements a WTO ruling.¹¹ There will be some instances where policies can be changed by unilateral acts, like executive orders. But on average, second-order compliance becomes less likely as domestic veto points increase.

Faced with an adverse ruling from the WTO Dispute Settlement Mechanism, leaders decide whether and when to second-order comply. This decision is shaped by short-term economic conditions and long-term institutional constraints. Leaders who face acute exogenous shocks have little short-term *incentive* to comply with rulings. Leaders who face more domestic veto points experience greater institutional and partisan political *obstacles* to complying. Conditional on going to trial and losing the case, a leader who faces more veto points at home has an incentive to wait for the exogenous shock to pass. Rather than complying immediately, a leader who faces more veto points, and thus high obstacles to compliance, is more likely to violate the agreement until the political or economic conditions change. *Conditional on violating a treaty, domestic political divisions (veto points) should be associated with less compliance.*

Domestic veto players are not the only factor driving second-order (non)compliance. A country that loses a WTO dispute and does not change its policy before the implementation deadline may be subject to enforcement. To isolate the impact of domestic veto players,

¹¹In the magazine dispute, partisan fighting between Canada's Tory and Liberal parties and between the upper and lower houses of Parliament obstructed compliance with the WTO ruling. See: "Magazine Bill Angers Senate; Tories Say Liberals Trying to Shove Legislation Through," *The Hamilton Spectator* (Ontario, Canada), June 1, 1999.

one must control for international pressure which can improve the prospects for compliance. Litigation publicizes the initial violation, raising awareness among other countries. Other countries can reduce the respondent's benefits from noncompliance by denying future opportunities for cooperation or imposing retaliatory trade measures that increase the respondent's costs from noncompliance.

3.3 Illustrations

Two examples illustrate ways in which domestic political divisions—both institutional and partisan—can hinder compliance.¹² In the first case, the respondent government had few divisions and complied quickly. In the second, the respondent had many divisions and defied the WTO ruling for years before eventually capitulating.

In the first case, the European Union sued South Korea over its liquor tax system (DS 75: Korea - Taxes on Alcoholic Beverages, 1997). European whiskey sales to Korea had fallen sharply and EU officials argued this was because the Korean government violated its WTO obligations by levying a 130% tax on alcohol imports but not on the local product, soju.¹³ The WTO determined that the tax constituted a trade barrier and ordered Korea to revise its system.¹⁴ The Korean government was under political pressure to defy the WTO. As one newspaper reported, “at stake besides soju itself are votes....soju is the drink of choice for South Korea's poor, a group the government is wooing before next April's parliamentary elections.”¹⁵ But at that time, Korea had relatively few veto points. The ruling NCNP had created a parliamentary majority by incorporating as many opposition legislators from the GNP as possible. So by 1999, the ruling coalition held a majority in the National Assembly which facilitated the passage of economic reform bills (Kim 2000, p.176). With a unicameral legislature, only modest partisan divisions, and thus few veto points, Korea successfully

¹²The Japanese and Korean domestic political circumstances I describe are also mentioned in Mansfield and Milner (2012), who show that veto players impact a country's ability to form preferential trade agreements.

¹³“Commission Calls for WTO Talks on South Korean Alcohol Tax,” *European Report*, April 2, 1997.

¹⁴“Tax ruling boosts whiskey hopes,” *The Herald*, Glasgow Scotland, January 19, 1999.

¹⁵“S. Korea's most popular drink under fire: Government must raise excise tax on soju and cut import tariffs on whiskey” *The Vancouver Sun*, British Columbia, October 23, 1999.

revised its tax law.¹⁶ Korea had minor domestic obstacles and complied with the adverse ruling.

In the second case, the EU sued Japan on similar grounds (DS 8: Japan - Taxes on Alcoholic Beverages, 1995). Reviving an unresolved dispute under the GATT,¹⁷ the EU argued the Japanese tax system was discriminatory because it levied a substantially higher tax on foreign products than the local alcohol, shochu. The WTO ruled against Japan, ordering it to reform. At the time, Japan had many veto points. By the mid 1990's the 38-year period of dominance by the Liberal Democratic Party had ended, leading to "rampant party realignment" and a "massive overhaul" of the Japanese party system (Pempel 1997, p.333). In 1994, a new electoral system for the lower House of Representatives was introduced, further increasing uncertainty, fragmentation of political power, and sensitivity to local politics with its diverse preferences. With these substantial domestic obstacles, Japan did not comply. Industry leaders in the EU attacked the Japanese government for "dragging its feet on the implementation of [the] WTO ruling...over as long a period as they think they can get away with."¹⁸ Despite a compliance deadline of February 1998, Japan decided to "gradually increase the tax on shochu...[through] October 2001."¹⁹ With a bicameral legislature, divided domestic politics, and thus many veto points, Japan defied the WTO ruling for many years.²⁰

¹⁶"EU, S. Korea Becoming Closer, More Interdependent," *The Korea Herald*, October 16, 2000. The Ministry of Finance and Economy initially suggested an amended liquor law with 100% tax hikes on both domestic soju and imported spirits but due to political pressure, lowered that proposal before bringing it to the National Assembly for a vote. There is evidence of partisan controversy within the legislature, although the major parties agreed that some reform was necessary. The Grand National Party wanted to set the tax rate at 50-60% whereas the National Congress for New Politics members supported 70-80%. There was disagreement even within the NCNP as they faced legislative elections in April 2000. The legislators were under pressure from the constituents and the soju industry, which openly criticized the government for "attempting to raise the price of the drink of the common people" (Maeil Business Newspaper 1999). The ruling coalition reached a consensus and the Korean National Assembly passed two acts to apply the same rate of 72% to both whiskey and domestic soju (Liquor Tax Act, No. 6055 and partial amendment of Education Tax Act, No. 6050, both Dec. 28, 1999).

¹⁷"Japan - Customs Duties, Taxes, and Labeling Practices on Imported Wines and Alcoholic Beverages," GATT Report of the Panel, November 10, 1987 (L/6216).

¹⁸"Japan fails to remove whiskey tax," *The Scotsman*, Scotland, November 23, 1996.

¹⁹"WTO Ruling Pushes Shochu Makers to Reinvent Product" *The Nikkei Weekly*, Japan, March 31, 1997.

²⁰Ultimately, the disputants reached a "mutually acceptable solution" that included short-term compensation and long-term policy reform. "Japan - Taxes on Alcoholic Beverages: Arbitration under Article 21 (3)," WTO Document No. 97-0558, February 14, 1997. This agreement entailed a legislative amendment to the Liquor Tax Law which was eventually passed by a small margin. Within the House of Councilors, the

These examples illustrate the variation in outcomes. When the EU won the lawsuit against Korea (few veto points), the Korean government complied in a timely manner. Yet when the EU won the lawsuit against Japan (many veto points) there was a prolonged period of noncompliance. The difference between Korea and Japan's veto points is significant.²¹ As domestic government grows more divided through institutional and partisan divisions, the prospects for compliance wane. While the examples provide preliminary support for the theory, a more systematic test requires evaluating *de facto* compliance in all WTO disputes with adverse rulings. I develop a method for measuring compliance using the most objective metric available: product-level trade flows and demonstrate that conditional on an adverse ruling, domestic divisions obstruct a return to compliance on average.²²

4 Assessing Compliance with WTO Rulings

4.1 Measurement Strategy

An ideal test of the DSM's effectiveness would compare the outcome with litigation to the outcome of the same dispute where the DSM did not exist. Because nearly all countries are members of the WTO, it is impossible to construct such a counterfactual. Yet one can still examine effectiveness with respect to particular aspects of the institution. Adverse WTO rulings require a respondent government remove an impermissible trade barrier. If the trade barrier is removed, imports into the respondent should subsequently increase. The empirical challenge is to correctly identify such an increase.

LDP, Heisei-kai, Social Democrats, Democrats/New Green Wind parties favored the amendment while the Communist and New Socialists/Peace Union parties opposed the amendment.

²¹Veto points are measured on a scale of 0 to 1 with larger values denoting more divisions. At the time of the lawsuits, Korea had 0.45 veto points and Japan had 0.60, a difference (0.15) that exceeds one standard deviation in my sample (0.12).

²²There may be multiple ways that governments can alter their policies to comply. For example, in the Korean alcohol dispute, Korea could have complied by lowering the tax on foreign liquor, raising the tax on local soju, or a combination (actual outcome). Some governments may chose compliance strategies that activate fewer veto players. These cases should bias my results toward a null finding.

Absent a randomized experiment, the only way to estimate the causal effect of an adverse WTO ruling is by comparing trade between disputing countries to an estimated control that represents what trade would have been without the adverse ruling. This paper uses observational data to construct an approximate control in order to estimate the effect of the adverse ruling on trade and thereby assess compliance. Although adverse rulings are not randomly assigned and true treatment effects cannot be obtained, casting the analysis in a causal inference framework is informative and I refer to the adverse ruling as the “*treatment*” and post-ruling trade as the “*outcome*.”

One obstacle is that trade changes for reasons unrelated to the WTO dispute. Disputes may be prompted by economic trends that cannot be reversed, even when the respondent government complies. These factors may be confounded with the effects of the WTO dispute. For example, governments sometimes initiate disputes when their exports for a product are declining, even though the decline is partly driven by forces exogenous to the respondent’s trade barrier. My technique mitigates this problem by estimating a control from trade flows following a parallel trend.

Figure 1 illustrates two hypothetical scenarios for WTO disputes. In both, trade is decreasing over time and a WTO dispute yields an adverse ruling (treatment). Observed trade (outcome for the treated unit) continues to decrease after the ruling (post-treatment period). Did the respondent government comply?

[Figure 1 here.]

This depends on what trade *would have been* after the WTO dispute, if there were no adverse ruling. When trade for the treated unit exceeds trade for the control unit, as measured by a positive average yearly deviation after the treatment, I infer the respondent complied (Figure 1(a)). Conversely, when trade for the treated unit does not exceed trade for the control unit, I expect the respondent did not comply. Because trade data are noisy, I aim to avoid “false negatives” and infer noncompliance only when I detect a negative average yearly deviation, (Figure 1(b)).

Measuring *de facto* compliance with trade flows prioritizes economic outcomes. Like any metric, it is imperfect: governments sometimes settle WTO disputes through compensation in lieu of prompt implementation,²³ and these cases will not count as compliance per my methodology. But since the central goal of the WTO is to liberalize trade, my trade-based measure best reflects this goal.

4.2 Product-Level Trade Flows

I examine the 125 WTO disputes that (1) received an adverse ruling between 1995 and 2011 and (2) concerned import restrictions.²⁴ The sample consists of all disputes with a panel ruling that favored the complainant on at least one legal claim. I exclude disputes where the ruling was completely overturned on appeal. I include only disputes about import restrictions like tariffs, countervailing duties, anti-dumping measures, safeguards, quantitative restrictions, discriminatory tax schemes, or other barriers to trade.

WTO disputes cite specific products and services. I collect annual bilateral trade data for these, aggregating when multiple products are cited in a given dispute. Only the trade flows for disputed products and services enter the analysis.²⁵ Data are from the UN Commodity Trade Statistics Database, the UN Service Trade Statistics Database, and the European Commission’s Eurostat database on international trade.²⁶

The unit of analysis is the directed dyad-year. For each dispute, I use the complainant’s annual exports of that disputed product to the respondent. The counterfactual “control unit,” uses the complainant’s annual exports of the disputed product to other countries not engaged in the dispute. I use up to fifteen other countries to estimate a synthetic control unit.

²³Dispute Settlement Understanding Article 22.1

²⁴The data and technical appendix will be available upon publication on the author’s website.

²⁵Wherever possible, the six-digit level of Harmonized System codes are used. If disputes cite products at the four- or two-digit level or have insufficient coverage, I use the highest level of precision available.

²⁶For the EU, I use aggregate trade with membership updated by year.

Trade is measured as the “export share,” the complainant’s annual exports of disputed products to the respondent or other country, divided by its total annual exports of the products to the world. The value of exports varies widely from one dispute to the next: some concern products with a large value (e.g. gasoline) while others with a small value (e.g. preserved peaches). The export share improves comparability across disputes and controls for price fluctuations and variation over time in the complainant’s export volumes. A large export share means the respondent’s market was very important to the complainant.²⁷ For each dispute and in each year t , the complainant exports disputed products to the respondent and to other countries. Let $j = 1$ denote the respondent and let $j = 2, 3, \dots, J$ denote the other countries. Country j ’s export share of the products in year t is:

$$\text{Export Share}_{jt} = \frac{\text{Complainant's Exports of Product to Country } j_t}{\text{Complainant's Exports of Product to World}_t}.$$

Because export shares are compositional data, I adopt a common practice of converting to log ratios (Tomz, Tucker, and Wittenberg, 2002). In each dispute, I divide by the *ex ante* largest trade partner among control countries. The transformed unit of analysis is:

$$y_{jt} = \log \left(\frac{\text{Export Share}_{jt}}{\text{Export Share}_{2t}} \right), \quad \text{for } j \in \{1, 2, 3, \dots, J\}$$

where $j = 2$ denotes the control country (in $j = 2, 3, \dots, J$) with the largest export share.

The transformation factors out the proportional component and isolates the independent variation among the units. It mitigates bias from trade diversion, the increase (reduction) in trade some countries may experience when another country imposes (removes) trade barriers.

Countries other than the disputants often enter WTO disputes as third parties, but this is not a serious source of bias in my measurements. Like the complainant, third parties are typically concerned with their exports to the respondent (Bown, 2005). When the respondent complies, third parties should benefit in the same way as the complainant does: their exports

²⁷Standardizing relative to the respondent’s imports may generate biased results because in many WTO disputes, the respondent imposes trade a barrier against all imports, regardless of the country of origin.

to the respondent increase.²⁸ Because I estimate the counterfactual from the complainant’s exports to other countries; *not* other countries’ exports to the respondent, I do not use trade flows that directly concern third parties.²⁹

4.3 Synthetic Control Method to Measure Compliance

I use the synthetic control method to estimate the counterfactual (synthetic control unit) and infer the approximate causal effect of an adverse WTO ruling (Abadie and Gardeazabal, 2003; Abadie, Diamond, and Hainmueller, 2010).³⁰ When the units of analysis are a few aggregate entities like countries, a combination of comparison units often does a better job reproducing the characteristics of the unit of interest than any single comparison unit alone (Abadie, Diamond, and Hainmueller, 2014). The control unit in the synthetic control method (SCM) is constructed from a weighted average of all potential comparison units. I use the “`synth`” package in R (Abadie, Diamond, and Hainmueller, 2011).

For every dispute, I use a sample of countries observed over multiple years. The respondent country subject to an adverse ruling is the “treated unit.” Other countries form the “donor pool,” the potential comparison units used to approximate the counterfactual. The donor pool consists of countries (1) whose markets are similarly important to the complainant as measured by the “export share” (2) have adequate data and (3) are not engaged in similar WTO disputes. I ensure that several countries from the respondent’s geographical region are included. Each WTO ruling has an implementation deadline which I use to split the sample into a “pre-treatment period” and a “post-treatment period,” lasting five years after the deadline.

The synthetic control is created with a two-part optimization process. First, each country in the donor pool receives a country-weight that reflects the similarity to the respondent.

²⁸Complainant governments are most likely to file WTO complaints when their expected benefits exceed the significant legal costs. This implies that complainants tend to be the primary beneficiaries of WTO lawsuits; else a different country would have sued.

²⁹See Appendix A.

³⁰See Appendix A.

This optimizes the similarity between the respondent and the weighted average of the other countries on a number of covariates in the pre-treatment period (i.e. before the deadline). Second, each covariate receives a covariate-weight that minimizes the discrepancy in the pre-treatment period between the respondent’s trade and the synthetic control, using the country-weights from the first optimization step. Covariates that are important predictors of the respondent’s trade receive more weight. The optimal solution entails a set of country-weights and a set of covariate-weights.

The covariates are gross domestic product (GDP), GDP per capita, annual GDP growth, value added in agriculture, industry, manufacturing, and services, trade dependence, and the unemployment rate (WorldBank, 2013).³¹ In robustness checks, I include democracy measured by Polity IV (Marshall and Jaggers, 2012).

The goal is to create an estimated counterfactual—a synthetic control—from a weighted average of countries in the donor pool. The counterfactual is accurate when the export share for the synthetic control matches the respondent’s export share in the pre-treatment period. Trade for the synthetic control is then projected into the following five years. This projection approximates the counterfactual—trade the respondent would have had in the absence of the WTO ruling. Using a difference-in-difference approach, I then compare the respondent’s actual trade to the expected (synthetic control) trade.³²

I generate a compliance score S , the average yearly difference between the respondent’s actual and expected trade in the post-treatment minus the average yearly difference in the pre-treatment period. A positive score indicates the respondent’s actual trade after the implementation deadline was higher than expected and I infer the respondent complied. Otherwise, I infer noncompliance. I calculate a compliance score for each of the 125 WTO disputes and compute the standard deviation of these yearly measurements in the pre-treatment period d , to capture the stability of the estimator.

³¹Covariates for the EU are average across all member countries, with membership updated by year.

³²If the respondent and synthetic control trade follow parallel trends—are subjected to all the same systematic factors and shocks save the WTO ruling—this approach identifies the average treatment effect on the treated.

Compliance scores S are approximately normally distributed between -0.22 and 0.24 with a mean of -0.006 . On the upper bound, a compliance score of 0.24 indicates the ruling helped the complainant recover nearly one-quarter of its export market in the disputed product. For the average dispute, this export share translates into roughly \$80 million in recovered trade. Nevertheless, the sample mean indicates that when WTO disputes are considered on average, the effect of adverse rulings on trade may be negligible.

The examples in Section 3.3 also illustrate my coding method. Figure 2 displays trade patterns for these WTO disputes. In the first case, the EU sued South Korea over its alcohol tax and won the lawsuit. The trends in European alcohol exports indicate that Korea complied. Figure 2(a) shows European exports to Korea rose relative to the synthetic control after the implementation deadline. SCM yields a positive, significant compliance score.

[Figure 2 here.]

In the second case, the EU sued Japan over a similar alcohol tax and won the lawsuit. It took many years for Japan to reform its tax system, long after the WTO's deadline for implementation had passed. Figure 2(b) demonstrates Japan did not comply: European alcohol exports to Japan continued to fall, relative to the synthetic control. SCM yields a negative, significant compliance score.

4.4 Methodological Advantages

The synthetic control method has several advantages over standard matching or regression approaches. First, and unlike matching, SCM creates a counterfactual based on pre-treatment trends in the outcome of interest. This mitigates the problem of time-varying confounders. As Abadie, Diamond, and Hainmueller (2010) explain, SCM “extends the traditional linear panel data (difference-in-differences) framework, allowing that the effects of unobserved variables on the outcome vary with time” (494). Many exogenous factors affect

multiple countries in similar ways—for example, a worsening drought impacts an entire region over time. Where time-varying, unobserved confounding factors affect the donor pool in the same way as they do the treated unit, SCM controls for these factors without explicitly using them as the basis for a match. This reduces the risk that omitted variables, especially time-variant ones, introduce bias. Choosing a donor pool that is likely subject to the same systematic factors ensures that such confounders are unlikely to drive my inferences.

Second, SCM creates more similar controls than standard matching techniques can achieve with a small, heterogeneous set of units. When a sample consists of few aggregate units—like countries—matching can be ineffective because treated units cannot be paired to control units without substantial extrapolation. The matching criteria can heavily influence the conclusions drawn (Smith and Todd 2005; Imai and Ratkovic 2014). By contrast, SCM is appropriate when the units of analysis are a few aggregate and heterogeneous entities like countries.

Third, SCM is transparent and flexible.³³ It is transparent because it makes explicit the contribution of each comparison unit. It is flexible because it allows each WTO dispute to have a separate covariate weighting that reflects the products and industries involved. So SCM permits a direct analysis of similarities between the case of interest and the synthetic control. For example, in a dispute over computer chips, Korea’s exports to the EU looked much like Korea’s exports to Japan.³⁴ The covariates for GDP and industry value added received the most weight. In a dispute over cigarettes, Honduras’s exports to the Dominican Republic looked like a combination of Honduras’s exports to Canada and to Costa Rica.³⁵ Key covariates were agriculture value added and industry value added. These weightings comport with reasonable expectations about the sectors at stake in each dispute.

³³Abadie, Diamond, and Hainmueller (2010) explain, “because a synthetic control is a weighted average of available control units, [SCM] makes explicit: (1) the relative contribution of each control unit to the counterfactual of interest and (2) the similarities...between the unit affected by the event or intervention of interest and the synthetic control” (494).

³⁴DS299: Korea v. EU - DRAMs.

³⁵DS302: Honduras v. Dominican Republic - Cigarettes.

Finally, SCM provides a safeguard against extrapolation. Traditional regression analysis can lead to extrapolation outside the support of the data. By contrast, synthetic control units are computed as weighted averages (convex combinations) of the control units, here other countries' trade flows. SCM ensures all estimates are based on interpolation and well-supported by the data.

5 Analysis and Results

5.1 Dependent Variable - Compliance

Compliance is measured from the compliance score S , the difference between actual and expected trade in the five years after the implementation deadline minus the difference in years before. While the sign of this quantity is reliable, the magnitude is sensitive to data availability and therefore noisy. I transform the compliance score S , and associated standard deviation d , according to two alternative coding rules.³⁶

Under the first coding rule, COMPLIANCE_A is a binary variable. A positive compliance score $S > 0$ denotes compliance, which occurs in 46% of the cases (58 of 125).

Under the second coding rule, I account for the point estimate and associated uncertainty.³⁷ The standard deviation d reflects the variability in the match between the actual trade flows and the expected trade flows. Small standard deviations indicate the synthetic control unit precisely fits the observed data in the pre-treatment period and produces a more reliable compliance score. I code the case as compliance (2) if the score is positive and larger than the standard deviation, noncompliance (0) if the score is negative and larger in magnitude than the standard deviation, and inconclusive (1) otherwise. Let COMPLIANCE_B be an ordinal variable:

³⁶See Appendix C.

³⁷Methodologists are still developing a standardized approach for reporting uncertainty associated with synthetic controls (Xu, 2015).

$$\text{COMPLIANCE}_B = \begin{cases} 0 & \text{if } S \leq -d \\ 1 & \text{if } S \in (-d, d) , \\ 2 & \text{if } S \geq d \end{cases} \quad (1)$$

This indicates compliance in 35% of cases (44 of 125) and noncompliance in 45.6% of cases (57 of 125).

5.2 Explanatory Variables and Controls

The explanatory variable is domestic veto players in the respondent government that arise from institutional checks and partisan divisions. I measure this as VETO POINTS using the Political Constraints Index (Henisz, 2002). It accounts for the number of independent branches of government, federalism, the extent of partisan alignment across branches of government, and preference heterogeneity within each legislative body. Partisan alignment accounts for party composition and left/right preference which change over time. VETO POINTS range from zero (least constrained) to one (most constrained). This metric is ideal because it has comprehensive coverage, is widely-accepted among political scientists, and allows me to draw comparisons between otherwise dissimilar countries. Where the European Union is the respondent, I use the average with membership updated by year.³⁸

To control for international pressure I use the number of THIRD PARTY countries using Horn and Mavroidis (2008) data and WTO records. In my sample, 89 disputes have no third parties, 24 disputes have between one and three, and the remaining 18 have many third parties. I also use the COMPLAINANT GDP in the year the dispute was initiated because complainants with larger economies have a greater capacity to penalize respondents that defy a WTO ruling.

Additional controls include the RESPONDENT GDP in the year the dispute was initiated. The GDP data come from the World Bank and are normalized to improve comparability. I

³⁸Some WTO disputes address EU-wide policies so the average will tend to understate effective obstacles to compliance.

account for the extent of the % ADVERSE RULING, which measures the percentage of legal claims found in favor of the complainant. If the governments appeal the ruling, I count the claims that were sustained. I include a dummy variable for the 22 cases where the European Union is the respondent, EU RESPONDENT and lawsuits against FEDERAL RESPONDENT governments, since either could pose further obstacles to policy reform.

5.3 Results: Domestic Veto Points Hinder Compliance

5.3.1 Probit Model

Across all model specifications, veto points are associated with a lower probability of compliance.³⁹ Table 1 shows probit regression results with COMPLIANCE_A. Veto points has a negative, statistically significant coefficient.⁴⁰ The result holds when I include the level of democracy as a basis for a match in the synthetic control (Model 8). The significance of veto players diminishes due to the high correlation between veto players and democracy, suggesting some variation is attributable to checks and balances that democratic institutions provide. Federalism is associated with worse compliance rates, complementing the broader veto players argument, since federal governments have an additional layer of constraints on national policy-making. The number of third party countries is positively associated with compliance, suggesting international pressure matters. The extent of the adverse ruling is positively and significantly associated with compliance. It provides a useful proxy for the magnitude of the treatment. The more adverse, the more likely I am to detect trade patterns indicating compliance.

The complainant and respondent countries' GDPs are not strongly associated with compliance. Once governments engage in litigation, their relative economic power is not an

³⁹See Appendix B for robustness tests with flexibility measures, alternative coding methods for compliance, and groups of related disputes. I considered measures of “polarization” and “checks” from the Database of Political Institutions which yields ambiguous results, suggesting that neither partisan preferences nor formal checks and balances alone capture pertinent veto points in government.

⁴⁰Where measurement errors arise in coding compliance, Table 1 results will likely underestimate the magnitude of the effect. Random mis-assignment in the dependent variable will tend to attenuate the coefficient on the explanatory variable (Cox and Snell, 1989).

informative predictor of compliance. This reinforces WTO advocates' claim that the legal process has an equalizing impact on the relations among countries, reducing the importance of power politics.⁴¹

[Table 1 here.]

Figure 3 shows the predicted effect of veto points on compliance. As veto points increase, the predicted probability of compliance decreases. Estimates use the probit model with all controls, holding variables at their means. The rug at the bottom of the plot shows the distribution of observations. Most disputes involve respondents with a moderate number of veto points.

[Figure 3 here.]

Several examples are plotted. When Mexico was sued in 2003 over its imposition of anti-dumping duties, it had few veto points (0.284). Institutional constraints on the executive were modest. The Partido Acción Nacional (PAN) held the presidency and had pluralities in both houses of the legislature.⁴² Mexico's predicted probability of compliance was high (≈ 0.6) and in this instance it did comply. By contrast, when Brazil was sued in 1996 over its domestic aircraft program, it had many veto points (0.684). This reflects its federal system with many municipalities and the prevalence of coalition government among multiple political parties.⁴³ Brazil's predicted probability of compliance was low (≈ 0.2) and it did not comply. Most of the variation in veto points is across countries, which have many fixed institutions. There are modest variations across time for individual countries, due to partisan shifts. For example, in 2009 the United States has fewer veto points (0.397) than it did in 2011 (0.414) when the Republican party gained a majority of seats in the House of Representatives while the Democratic party maintained control of the Senate and Presidency.

⁴¹The EU does not have significantly different compliance rates.

⁴²In 2003 the PAN held 38.11% of the seats in the senate and 38.24% of the seats in the lower house compared to smaller shares held by the main opposition party.

⁴³The Brazilian Constitution treats its 5,570 municipalities as parts of the Federation, each with autonomous local government, and not simply dependent subdivisions of its 26 states and Federal District.

5.3.2 Ordered Probit Model

Domestic divisions are strongly associated with less compliance, even using the alternative coding rule that accounts for uncertainty. I fit an ordered multinomial probit model using the COMPLIANCE_B variable and display results in Table 2.

[Table 2 here.]

Table 2 reinforces the findings above. A respondent government with more veto points is less likely to comply with an adverse WTO ruling. The results also hold when I repeat the analysis using level of democracy as an additional covariate in the SCM calculations (column 7). The magnitude of the coefficient on VETO PLAYERS diminishes slightly, indicating that some—but certainly not all—of the variation in compliance is attributable to level of democracy. By accounting for democracy in the measurement of compliance and federalism in the regression analysis, these results show that the veto players explanation remains strong. Finally, using the compliance scores directly in the regression (column 8, OLS) confirms there is a tendency, if statistically insignificant, for countries with more veto points to comply less.

These predicted effects are large. If an average respondent government were to increase its domestic veto points from the least to the most, its predicted probability of compliance *decreases* by 0.44.⁴⁴

5.4 Robustness

5.4.1 Reverse Causality

Endogeneity bias might arise if international trade disputes affect domestic politics. Scholars have shown trade affects domestic political cleavages (Rogowski, 1989) and that international conditions can force a country to adapt its trade policy process (e.g. US fast track negotiating authority). This form of reverse causality is not a serious concern. Trade disputes occur over

⁴⁴This estimate uses Table 2(6) and conditions on a transition from the ambiguous outcome to compliance.

a short time-frame: usually less than five years. Changes in trade policy have a rapid impact on trade flows. By contrast, domestic veto points, change slowly over time. They reflect (1) the institutional separation of power which may change only a few times during a country's modern lifespan and (2) partisan divisions which are determined by unrelated and powerful macro-political issues. WTO disputes are unlikely to affect the respondent's veto points.

5.4.2 Selection Effects from Early Settlement

WTO disputes with adverse rulings are only a subset of the legal cases brought to the Dispute Settlement Mechanism. More than half of the disputes are resolved without any legal verdict because the disputants settle early. This reflects the WTO's aim to resolve trade disputes during consultations, before governments resort to costly litigation and compliance becomes an issue. Selection into litigation is not random, and potentially correlated with the outcome—compliance—raising the possibility of selection bias.⁴⁵

The direction of potential selection bias is *a priori* unclear. On the one hand, disputes that are especially intractable may be more likely to require a panel ruling and less likely to result in compliance.^{46,47} On the other hand, lawsuits against particularly obstinate respondents may be resolved prior to litigation if the complainant, realizing litigation is futile, capitulates. This type of selection effect is notoriously difficult to overcome using observational data.⁴⁸ The outcomes I observe may either understate or overstate the theoretical compliance rate, had all disputes gone through litigation.

I examine the selection process with a series of Heckman selection models (Heckman, 1979, 1990) in Appendix B. My results are robust in that I do not find any evidence that

⁴⁵Because nearly every WTO ruling is adverse, the important comparison group is the set of disputes that were settled early, without a legal verdict at all.

⁴⁶This expectation is probabilistic because potential complainants are not always well-informed and the respondent's policy preferences change with exogenous political and economic shocks. Even respondents with many veto players have a small chance of complying.

⁴⁷In addition, countries may choose whether to violate GATT/WTO rules based on expectations about retaliation (Bown, 2004a).

⁴⁸As strategic actors, governments in an international trade dispute have every incentive to anticipate their opponent's behavior and adjust their actions.

contradicts my theory: domestic veto points are still strongly associated with noncompliance. In these models, the selection stage estimates the effect of covariates on the probability the WTO issues an adverse ruling and the outcome stage estimates the probability of compliance, conditional on the adverse ruling. I use a number of different variables for identification. However, this robustness test is not decisive because the identification is imperfect.

5.4.3 Endogenous Dispute Timing

Governments prefer to initiate trade disputes when conditions are favorable. They may wait to file their complaints about treaty violations until there are exogenous shocks that lead to declines in trade. The WTO usually requires the complainant to demonstrate harm to its domestic industry and declining trade provides such evidence. If complainants prefer to file their disputes when there are exogenous declines in trade, dispute initiation may be endogenous to trade fluctuations.

Dispute timing is a potential source of bias. If complaints are filed when there are negative shocks, once the exogenous shock passes trade should increase, even absent a WTO ruling. Then I might mistakenly attribute to the WTO an increase in trade that is due to the unrelated passing shock. In other words, endogenous timing would bias my estimates toward compliance and thereby overestimate the efficacy of the WTO.

My methodology is robust to dispute timing because the synthetic control unit is constructed from many years of trade patterns for several countries and short-term fluctuations are moderated. Countries experience temporary import surges that pass. When the surge is viewed over a short time horizon, it might appear as though trade declined. However, viewed over a longer time span, it is clear the decline merely restores pre-surge trade levels.

[Figure 4 here.]

For example, Figure 4 shows trade flows in a dispute between Indonesia and Korea⁴⁹. Korea experienced an import surge around 2002 and imposed a trade barrier that broke

⁴⁹DS312: Indonesia v. Korea—Paper

WTO rules. Indonesian exports to Korea dropped precipitously and Indonesia filed a formal complaint at the WTO. Viewed over the two years leading up to the dispute, it would appear that Indonesian exports suffered a huge loss from Korea’s policy because the exogenous shock magnified the apparent impact of the trade barrier. Yet viewed over a longer time span, it is clear that Indonesian exports were unusually high before the dispute and Korea’s barrier had only a modest impact. SCM estimates are robust because I account for the longer time horizon and use data from many countries to construct the control. In Figure 4, the control unit trade (dashed line) does not increase during the import surge. I still detect Korea’s noncompliance. This comports with the legal record—compliance proceedings showed Korea failed to implement the ruling. Dispute timing does not appear to confound my estimation.

6 Conclusion

While the World Trade Organization’s Dispute Settlement Mechanism is central to the multilateral trade regime, relatively little is known about its effect on international cooperation. This paper evaluates the impact of adverse WTO rulings on trade flows between disputing countries by applying the method of synthetic case control. In every dispute, I estimate the causal effect of an adverse WTO ruling by constructing an estimated counterfactual against which to gauge trade fluctuations. The counterfactual incorporates specific product-level dyadic trade data for several years leading up to and following the ruling. Each is created with an optimal combination of economic control variables that reflects the products and issues in that particular case. Increases in actual trade relative to expected trade indicates the respondent government complied with the ruling. Using this methodology, I measure compliance for all 125 WTO disputes with adverse rulings between 1995 and 2011.

When its record is evaluated in the aggregate, it might appear that the WTO has had little success in restoring trade between disputing countries. However, my results demonstrate that many disputes actually prompt significant increases in trade, relative to expected levels.

Other disputes have no discernible effect. This study advances our collective knowledge on the impact of WTO rulings by accounting for the substantial heterogeneity in outcomes. My results show domestic politics explain a respondent government's behavior when it faces an adverse ruling. Conditional on violating a treaty, the more domestic veto points a government has, the less likely it is to return to compliance. This pattern is significant and persists even when the respondent faces international pressure to comply. It is robust across model specifications and coding schemes.

My findings have implications for research on the optimal design of international institutions. Many scholars have highlighted the advantages of flexibility mechanisms which provide a safety valve to countries facing acute domestic pressure to defy their international commitments. Governments are less likely to abandon an institution altogether when they are permitted temporary transgressions (Rosendorff and Milner, 2001; Carrubba, 2005; Rosendorff, 2005; Johns, 2014). The models that link flexibility to stability typically treat states as unitary actors that can violate rules and then easily return to their international commitments. However, opening the black box of domestic politics reveals that returning to compliance may not be so simple; multiple veto players can obstruct the process. This implies a potential hazard of institutional flexibility: it only promotes long-term cooperation when countries return to compliance after the temporary pressure to violate passes. When veto players lock in treaty violations, flexibility mechanisms may fail to restore cooperation in the long-term, undermining the institution's stability. Dispute settlement mechanisms may be particularly vulnerable due to the publicizing nature of litigation which forces the government into an often-fraught domestic political process. At this point, treaty violations become difficult to reverse.

Political scientists have long known that divisions in a government can obstruct policy change. The impact of institutional checks and balances and partisan conflict—veto players—also reaches beyond national borders. Veto players have been shown to impact the types of international commitments countries make and keep (Milner and Rosendorff,

1996, 1997; Rickard, 2010; Mansfield and Milner, 2012). Institutional checks and partisan opposition are integral to democracy. Many studies therefore conclude that democracies, with their multiple veto players and voter-based audience costs, comply more with international commitments (“first-order compliance”). In addition, democracies are thought to allow for the creation and growth of interest groups that support international cooperation (e.g. Finnemore and Sikkink 1998; Dai 2005; Simmons 2009).

That key features of democratic politics sometimes obstruct compliance sits uneasily with the prevailing view in the international relations literature. My findings point to a more nuanced relationship, suggesting that democratic politics can have cross-cutting effects on international cooperation. Veto players can actually decrease a government’s likelihood of complying with international legal rulings (“second-order compliance”). With domestic institutional divisions and partisan opposition blocking policy change, even trade policies that violate international obligations can be locked into place. Dispute settlement mechanisms are often unable to compel governments to reverse these violations. Ultimately, international dispute settlement is only as effective as domestic politics allows.

Table 1: Probit Models for Compliance with WTO Rulings, 1995–2011

	COMPLIANCE _A							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8) [†]
Veto Points	-2.320** (1.039)	-2.070* (1.097)	-2.804*** (1.083)	-2.245** (1.055)	-2.096** (1.058)	-2.449** (1.107)	-2.867** (1.352)	-2.361* (1.340)
Third Parties		0.138** (0.060)					0.135** (0.059)	0.500* (0.272)
% Adverse Ruling			0.702** (0.339)				0.713* (0.367)	0.696* (0.363)
Respondent GDP				0.069 (0.110)			0.193 (0.146)	0.225 (0.145)
Complainant GDP				0.016 (0.122)			0.001 (0.130)	0.036 (0.129)
EU Respondent					-0.295 (0.313)		0.046 (0.433)	0.326 (0.129)
Federal Respondent						-0.501* (0.276)	-0.638* (0.357)	-0.806** (0.358)
Constant	0.840* (0.443)	0.606 (0.476)	0.555 (0.471)	0.788* (0.455)	0.798* (0.442)	1.280** (0.540)	0.871 (0.663)	0.806 (0.670)
Observations	125	125	125	125	125	125	125	125
Log Likelihood	-83.457	-80.439	-81.274	-83.256	-83.007	-81.784	-75.752	-77.080

Notes: Compliance is coded using the synthetic control method with annual bilateral trade data for disputed products.

[†] Robustness check where compliance is coded using Polity IV as an additional covariate in the synthetic control calculation.

Probit regressions are calculated with the ‘sampleSelection’ package in R. * p<0.1; **p<0.05; ***p<0.01.

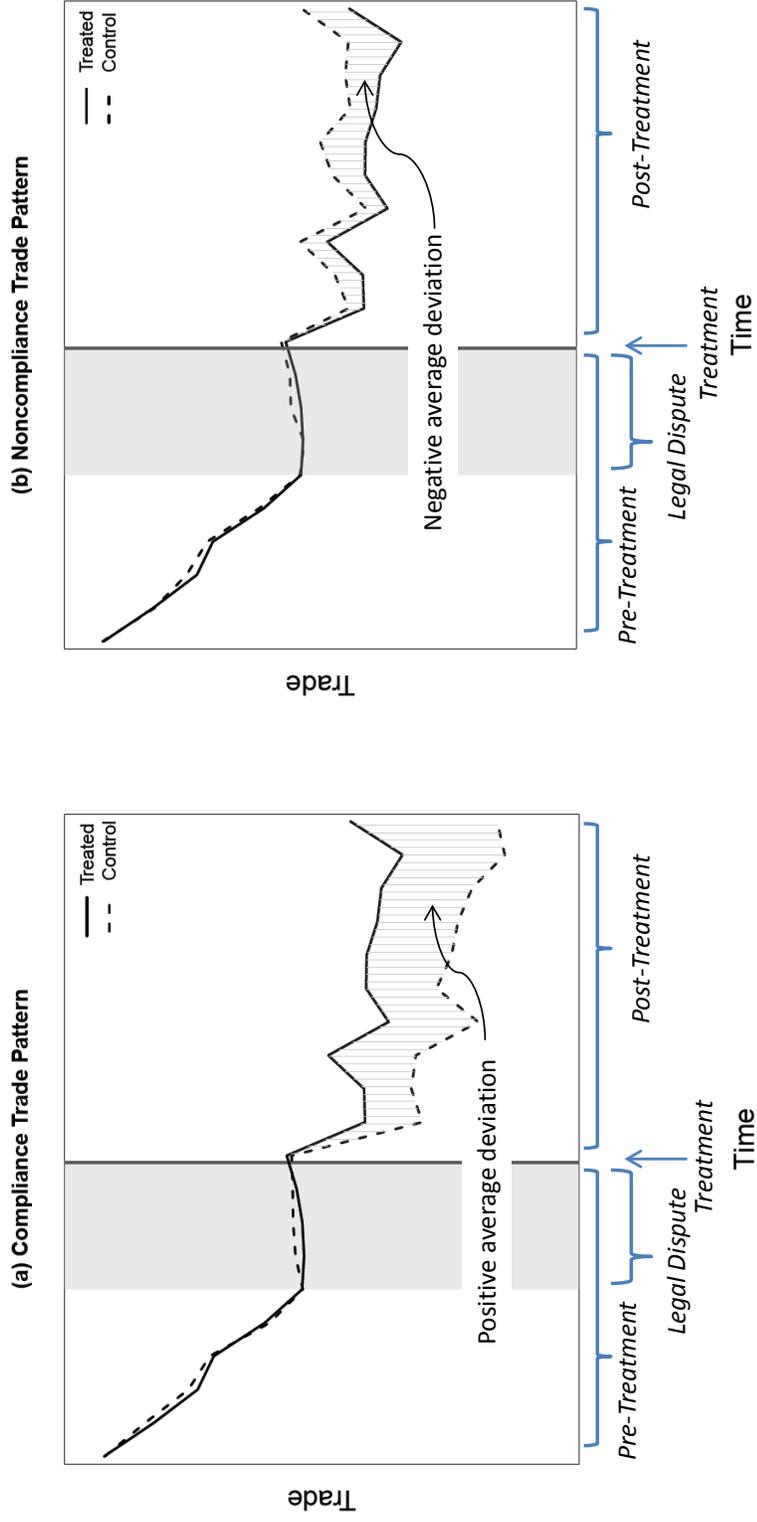
Table 2: Ordered Probit Models and Linear Model of Compliance with WTO Rulings, 1995–2011

	COMPLIANCE _B							COMPLIANCE
	(1)	(2)	(3)	(4)	(5)	(6)	(7) [†]	SCORE (S)
Veto Points	-1.938** (0.946)	-1.639* (0.988)	-2.280** (0.977)	-1.959** (0.970)	-2.012* (1.041)	-2.276* (1.206)	-1.968* (1.177)	-0.029 (0.057)
Third Parties		0.142** (0.056)				0.140** (0.056)	0.143*** (0.056)	0.005 (0.003)
% Adverse Ruling			0.507 (0.309)			0.433 (0.333)	0.264 (0.327)	0.011 (0.019)
Respondent GDP				0.020 (0.102)		0.139 (0.133)	0.059 (0.133)	-0.0002 (0.007)
Complainant GDP				-0.023 (0.113)		-0.039 (0.120)	0.003 (0.118)	-0.004 (0.006)
EU Respondent					-0.131 (0.304)	-0.028 (0.394)	0.025 (0.390)	-0.013 (0.021)
Federal Respondent					-0.643** (0.266)	-0.782** (0.325)	-0.760** (0.319)	-0.015 (0.017)
Constant (0:1)	-0.890 (0.407)	-0.641 (0.434)	-0.685 (0.429)	-0.892 (0.423)	-1.446 (0.510)	-1.185 (0.597)	-1.185 (0.584)	
Constant (1:2)	-0.431 (0.404)	-0.164 (0.432)	-0.219 (0.427)	-0.433 (0.419)	-0.969 (0.506)	-0.677 (0.594)	-0.589 (0.580)	
Constant								0.008 (0.027)
Compliance Measure	(0/1/2)	(0/1/2)	(0/1/2)	(0/1/2)	(0/1/2)	(0/1/2)	(0/1/2) [†]	Continuous
Model	OP	OP	OP	OP	OP	OP	OP	OLS
Observations	125	125	125	125	125	125	125	125
Log Likelihood	-126.534	-122.933	-125.176	-126.475	-122.926	-117.502	-121.012	

Notes: Compliance is coded using the synthetic control method with annual bilateral trade data for disputed products.

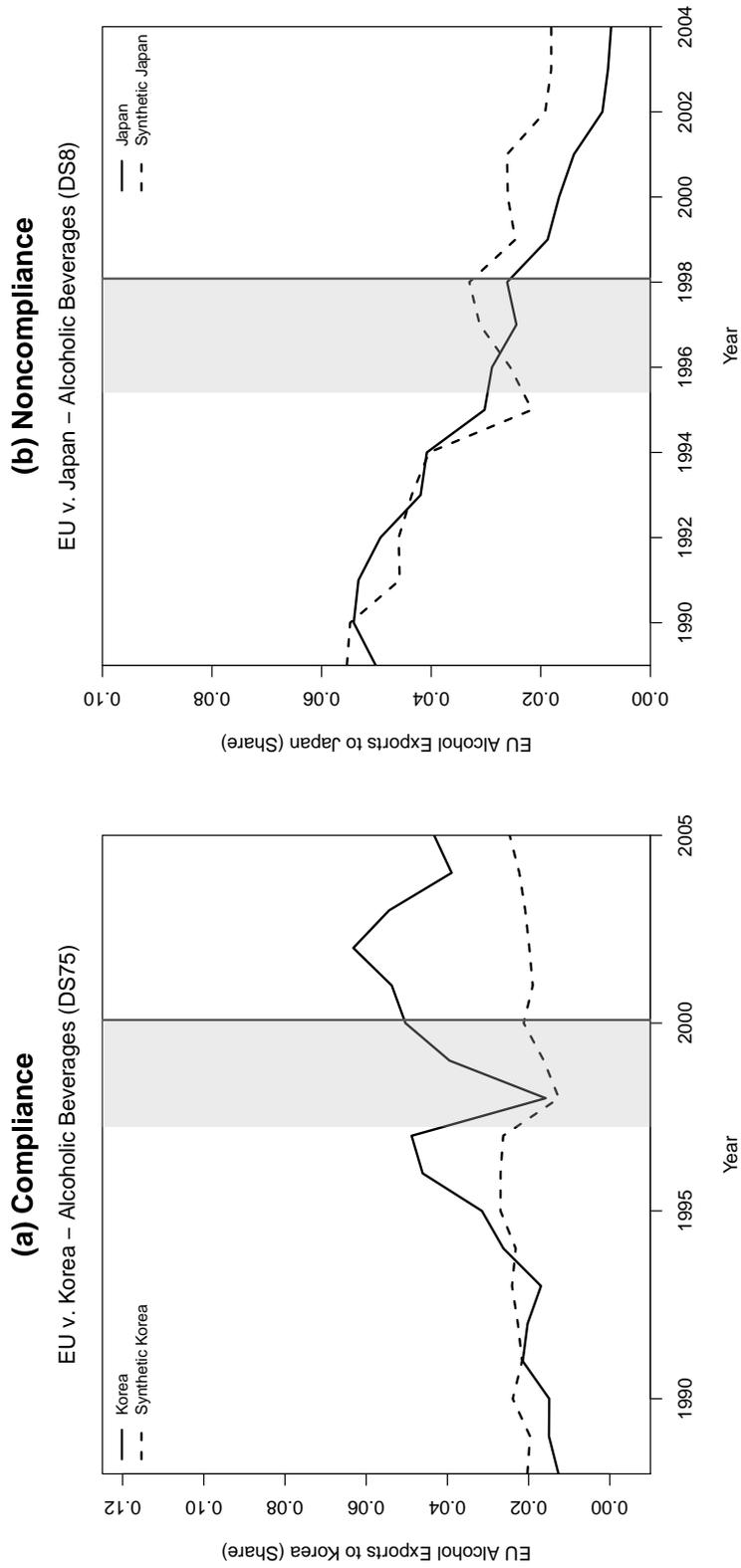
[†]Robustness check where compliance is coded using Policy IV as an additional covariate in the synthetic control calculation. Ordered probit regressions are calculated with the "MASS" package in R. *p<0.1; **p<0.05; ***p<0.01.

Figure 1: Hypothetical Compliance and Noncompliance Trade Patterns



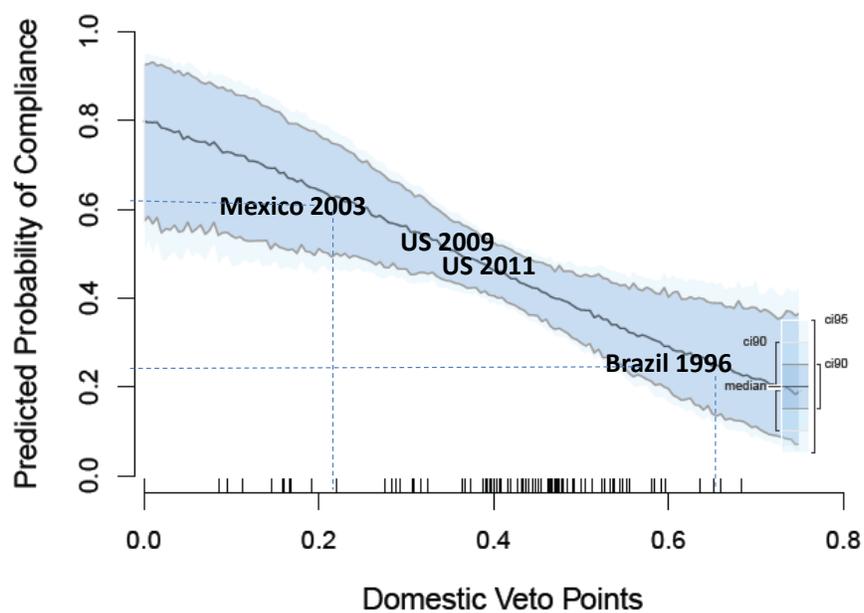
Note: Hypothetical scenarios for a dispute where trade is trending down over time. Panel (a) illustrates a compliance trade pattern, indicated by a positive average deviation. Panel (b) illustrates a noncompliance trade pattern. The shaded area denotes the duration of the legal dispute and the vertical line indicates the implementation deadline.

Figure 2: WTO Disputes with Compliance and Noncompliance Trade Patterns



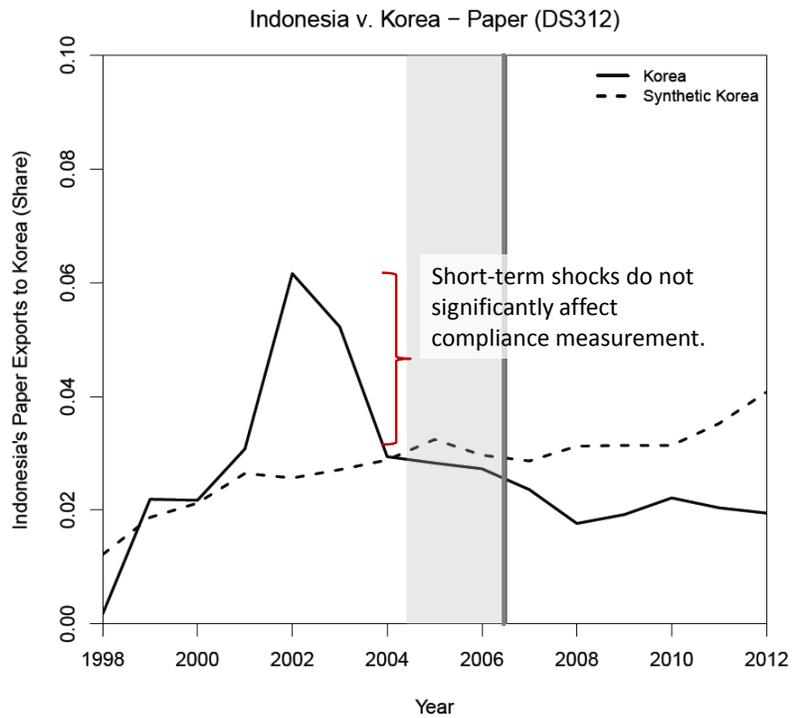
Note: WTO disputes with compliance (a) and noncompliance (b). The vertical shaded areas cover the dispute duration from the request for consultations to the implementation deadline (bold vertical line). Synthetic control estimates were created with “Synth” package in R.

Figure 3: Predicted Effect of Veto Points on Compliance



Note: Shading denotes 90% and 95% confidence intervals. Examples of veto points for selected country-years are plotted. The rug shows the distribution of observations. Predictions are based on probit model with all controls. Estimates were created with “Synth” and “Zelig” packages in R.

Figure 4: Robustness to Dispute Timing



Note: The vertical shaded area covers the dispute duration from the request for consultations to the implementation deadline (vertical line). Synthetic control estimates were created with “Synth” package in R.

References

- Abadie, Alberto, Alexis Diamond, and 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).
- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2011. "Synth: An R Package for Synthetic Control Methods in Comparative Case Studies." *American Statistical Association* .
- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2014. "Comparative Politics and the Synthetic Control Method." *American Journal of Political Science* Forthcoming.
- Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93(1): 113–132.
- Abbott, Kenneth, and Duncan Snidal. 1998. "Why States Act Through Formal International Organizations." *Journal of Conflict Resolution* 42(1): 3–32.
- Allee, Todd L, and Paul K Huth. 2006. "Legitimizing Dispute Settlement: International Legal Rulings as Domestic Political Cover." *American Political Science Review* 100(02): 219–234.
- Bechtel, Michael M, and Thomas Sattler. 2015. "What is litigation in the world trade organization worth?" *International Organization* 69(02): 375–403.
- Bown, Chad P. 2002. "The Economics of Trade Disputes, the GATT's Article XXIII, and the WTO's Dispute Settlement Understanding." *Economics & Politics* 14(3): 283–323.
- Bown, Chad P. 2004a. "On the Economic Success of GATT/WTO Dispute Settlement." *Review of Economics and Statistics* 86(3): 811–823.
- Bown, Chad P. 2004b. "Trade Disputes and the Implementation of Protection Under the GATT: An Empirical Assessment." *Journal of International Economics* 62(2): 263–294.
- Bown, Chad P. 2005. "Participation in WTO Dispute Settlement: Complainants, Interested Parties, and Free Riders." *The World Bank Economic Review* 19(2): 287–310.
- Bown, Chad P., and Michele Ruta. 2015. "Trade Flows and Trade Disputes." *World Bank Policy Research Working Paper* (6979).
- Brewster, Rachel. 2011. "The Remedy Gap: Institutional Design, Retaliation, and Trade Law Enforcement." *George Washington Law Review* 80: 102.
- Busch, Marc L, and Eric Reinhardt. 2000. "Bargaining in the Shadow of the Law: Early Settlement in GATT/WTO Disputes." *Fordham International Law Journal* 24: 158.
- Carrubba, Clifford J. 2005. "Courts and Compliance in International Regulatory Regimes." *Journal of Politics* 67(3): 669–689.
- Chaudoin, Stephen, Jeffrey Kucik, and Krzysztof Pelc. 2013. "Do WTO Disputes Actually Increase Trade?". In *American Political Science Association 2013 Annual Meeting*.
- Chayes, Abram, and Antonia Handler Chayes. 1993. "On Compliance." *International Organization* 47(02): 175–205.

- Cox, David Roxbee, and E Joyce Snell. 1989. *Analysis of binary data*. Vol. 32 CRC Press.
- Dai, Xinyuan. 2005. "Why Comply? The Domestic Constituency Mechanism." *International Organization* 59(02): 363–398.
- Davis, Christina L. 2012. *Why Adjudicate? Enforcing Trade Rules in the WTO*. Princeton: Princeton University Press.
- Downs, George W., David M. Rocke, and Peter N. Barsoom. 1996. "Is the Good News About Compliance Good News About Cooperation?" *International Organization* 50(03): 379–406.
- Finnemore, Martha, and Kathryn Sikkink. 1998. "International Norm Dynamics and Political Change." *International Organization* 52(04): 887–917.
- Fisher, Roger. 1981. *Improving Compliance with International Law*. Charlottesville: University Press of Virginia.
- Gilligan, Michael J. 1997. *Empowering Exporters: Reciprocity, Delegation, and Collective Action in American Trade Policy*. Ann Arbor: University of Michigan Press.
- Goldstein, Judith, and Lisa Martin. 2000. "Legalization, Trade Liberalization, and Domestic Politics: A Cautionary Note." *International Organization* 54(03): 603–632.
- Goldstein, Judith, and Richard Steinberg. 2008. "Negotiate or Litigate? Effects of WTO Judicial Delegation on US Trade Politics." *Law and Contemporary Problems* pp. 257–282.
- Goldstein, Judith, Douglas Rivers, and Michael Tomz. 2007. "Institutions in International Relations: Understanding the Effects of the GATT and the WTO on World Trade." *International Organization* 61(01): 37–67.
- Gowa, Joanne, and Soo Yeon Kim. 2005. "An exclusive country club: the effects of the GATT on trade, 1950–94." *World Politics* 57(04): 453–478.
- Hafner-Burton, Emilie M., Jana von Stein, and Erik Gartzke. 2008. "International Organizations Count: What Statistics Tell Us About IOs." *Journal of Conflict Resolution* 52(2).
- Heckman, James. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* pp. 153–161.
- Heckman, James. 1990. "Varieties of selection bias." *The American Economic Review* 80(2): 313–318.
- Henisz, Witold J. 2000. "The Institutional Environment for Economic Growth." *Economics & Politics* 12(1): 1–31.
- Henisz, Witold J. 2002. "The Political Constraint Index (POLCON) Dataset." Available: <http://www-management.wharton.upenn.edu/henisz/>.
- Hiscox, Michael J. 2006. "Through a Glass and Darkly: Attitudes Toward International Trade and the Curious Effects of Issue Framing." *International Organization* 60(7): 755–780.
- Horn, Henrik, and Petros C. Mavroidis. 2008. "The WTO Dispute Settlement Data Set 1995-2006." Available from: <http://go.worldbank.org/X5EZPHXJY0>.

- Hudec, Robert E. 1999. "The New WTO Dispute Settlement Procedure: An Overview of the First Three Years." *Minnesota Journal of Global Trade* 8: 1.
- Imai, Kosuke, and Marc Ratkovic. 2014. "Covariate Balancing Propensity Score." *Journal of the Royal Statistical Society* 76(1): 243–263.
- Johns, Leslie. 2014. "Depth versus Rigidity in the Design of International Trade Agreements." *Journal of Theoretical Politics* 26(3): 468–495.
- Keohane, Robert O. 1982. "The Demand for International Regimes." *International Organization* 36(02): 325–355.
- Kim, Sunhyuk. 2000. "The Politics of Reform in South Korea: The First Year of the Kim Dae Jung Government 1998-1999." *Asian Perspective* 24(1): 163–185.
- Kono, Daniel Y. 2006. "Optimal Obfuscation: Democracy and Trade Policy Transparency." *American Political Science Review* 100(3): 369–384.
- Koremenos, Barbara, Charles Lipson, and Duncan Snidal. 2001. "The Rational Design of International Institutions." *International Organization* 55(4): 761–799.
- Kucik, Jeffrey, and Eric Reinhardt. 2008. "Does Flexibility Promote Cooperation? An Application to the Global Trade Regime." *International Organization* 62(3): 477–505.
- Lohmann, Susanne, and Sharyn O'Halloran. 1994. "Divided Government and US Trade Policy: Theory and Evidence." *International Organization* 48(4): 595–632.
- Mansfield, Edward D, and Helen V Milner. 2012. *Votes, Vetoes, and the Political Economy of International Trade Agreements*. Princeton: Princeton University Press.
- Mansfield, Edward D., Helen V. Milner, and B. Peter Rosendorff. 2002. "Why Democracies Cooperate More: Electoral Control and International Trade Agreements." *International Organization* 56(3): 477–513.
- Mansfield, Edward D., Helen V. Milner, and Jon C. Pevehouse. 2007. "Vetoing Co-operation: The Impact of Veto Players on Preferential Trading Arrangements." *British Journal of Political Science* 37: 403–432.
- Marshall, Monty G., and Keith Jagers. 2012. "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2012."
- Martin, Lisa. 2013. "Against Compliance." In *Interdisciplinary Perspectives on International Law and International Relations: The State of the Art*, ed. Jeffrey L. Dunoff, and Mark A. Pollack. Cambridge University Press pp. 591–610.
- Martin, Lisa L. 2000. *Democratic Commitments: Legislatures and International Cooperation*. Princeton: Princeton University Press.
- Mearsheimer, John J. 1994. "The False Promise of International Institutions." *International Security* pp. 5–49.
- Milner, Helen V., and B. Peter Rosendorff. 1996. "Trade Negotiations, Information and Domestic Politics: The Role of Domestic Groups." *Economics & Politics* 8(2): 145–189.

- Milner, Helen V, and B Peter Rosendorff. 1997. "Democratic Politics and International Trade Negotiations Elections and Divided Government As Constraints on Trade Liberalization." *Journal of Conflict Resolution* 41(1): 117–146.
- Pelc, Krzysztof. 2009. "Seeking Escape: The Use of Escape Clauses in International Trade Agreements." *International Studies Quarterly* 53(2): 349–368.
- Pelc, Krzysztof J. 2014. "The Politics of Precedent in International Law: A Social Network Application." *American Political Science Review* 108(03): 547–564.
- Pempel, T. J. 1997. "Regime Shift: Japanese Politics in a Changing World Economy." *The Journal of Japanese Studies* 23(2): 333–361.
- Rickard, Stephanie J. 2010. "Democratic Differences: Electoral Institutions and Compliance with GATT/WTO Agreements." *European Journal of International Relations* 16(4): 711–729.
- Rogowski, Ronald. 1989. *Commerce and Coalitions: How Trade Affects Domestic Political Arrangements*. Princeton, NJ: Princeton University Press.
- Rose, Andrew. 2004. "Do We Really Know that the WTO Increases Trade?" *American Economic Review* 94(1): 98–114.
- Rosendorff, B. Peter. 2005. "Stability and Rigidity: Politics and Design of the WTO's Dispute Settlement Procedure." *American Political Science Review* 99(3): 389–400.
- Rosendorff, B. Peter, and Helen V. Milner. 2001. "The Optimal Design of International Trade Institutions: Uncertainty and Escape." *International Organization* 55(4): 829–857.
- Rubin, Donald B. 2005. "Causal Inference Using Potential Outcomes." *Journal of the American Statistical Association* 100(469).
- Simmons, Beth A. 1998. "Compliance with International Agreements." *Annual Review of Political Science* 1(1): 75–93.
- Simmons, Beth A. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics*. Cambridge: Cambridge University Press.
- Smith, Jeffrey, and Petra Todd. 2005. "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics* 125(1): 305–353.
- Staiger, Robert W., and Kyle Bagwell. 1999. "An Economic Theory of GATT." *American Economic Review* 89(1): 215–248.
- Stein, Arthur A. 1990. *Why Nations Cooperate: Circumstance and Choice in International Relations*. Ithaca: Cornell University Press.
- Subramanian, Arvind, and Shang-Jin Wei. 2007. "The WTO promotes trade, strongly but unevenly." *Journal of International Economics* 72(1): 151 – 175.
- Tomz, Michael. 2007. "Domestic Audience Costs in International Relations: An Experimental Approach." *International Organization* 61(04): 821–840.

- Tomz, Michael, Joshua A. Tucker, and Jason Wittenberg. 2002. "An Easy and Accurate Regression Model for Multiparty Electoral Data." *Political Analysis* 10(1): 66–83.
- Tsebelis, George. 1995. "Decision Making in Political Systems: Veto Players in Presidentialism, Parliamentarism, Multicameralism and Multipartyism." *British Journal of Political Science* 25(3): 289–325.
- Tsebelis, George. 2003. *Veto Players: How Political Institutions Work*. Princeton: Princeton University Press.
- Voeten, Erik. 2013. "Does Participation in International Organizations Increase Cooperation?" *The Review of International Organizations* pp. 1–24.
- von Stein, Jana. 2005. "Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance." *American Political Science Review* 99(4): 611–622.
- WorldBank. 2013. "World Development Indicators." Available from: <http://data.worldbank.org/indicator> .
- Xu, Yiquing. 2015. "Generalized Synthetic Control Method for Causal Inference with Time-Series Cross-Sectional Data." *MIT Working Paper* .