THE UNIVERSITY OF CHICAGO

EMPIRICAL ESSAYS ON THE INSITUTIONAL DESIGN OF

INVESTOR-STATE DISPUTE SETTLEMENT

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE LAW SCHOOL

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF JURISPRUDENCE

BY

WEIJIA RAO

CHICAGO, ILLINOIS

AUGUST 2019

EMPIRICAL ESSAYS ON THE INSTITUTIONAL DESIGN OF

INVESTOR-STATE DISPUTE SETTLEMENT

WEIJIA RAO

SUPERVISORS: TOM GINSBURG & WILLIAM H.J. HUBBARD

<div align="center">Table of Contents</div>

This J.S.D. dissertation provides empirical insights into existing concerns with the investor-state dispute settlement (ISDS) system. ISDS allows foreign investors to bring claims against sovereign states for alleged violations of international investment agreements before ad hoc arbitral tribunals. In recent years, ISDS has been criticized for, among other things, bias against developing countries, bias arising from ad hoc appointments, and lengthy and costly proceedings. This J.S.D. dissertation empirically examines each of these three critiques against ISDS. The empirical findings provide normative implications for the institutional design of the ISDS system.

The first chapter – Development Status and Decision-Making in Investment Treaty Arbitration – examines the criticism that ISDS is biased against developing countries. I first test the reproducibility and replicability of two representative empirical studies that examine this question. I then propose new methods to account for investors' case strength and selection bias caused by settlements. These are major confounding factors that previous studies fail to consider. I construct control variables that measure the level of investment protection in the underlying investment treaties and the governance quality of respondent states, which could be correlated with unobserved case strength. I also use the Heckman probit selection model to correct for potential selection bias caused by settlement. Taking into account these confounding factors, I find consistent evidence that more developed respondent states are significantly less likely to lose in investment treaty arbitration cases than less developed respondent states. The patterns indicate that there may exist biases against developing countries in investment treaty arbitration. The existence of these biases calls for institutional solutions like expanding the pool of arbitrators from developing countries and increasing the transparency of the dispute settlement proceeding.

1

The second chapter – Are Arbitrators Biased in ICSID Arbitration? A Dynamic Perspective – examines arbitrator bias from another angle: whether arbitrators are biased due to their reappointment interests or favoring of past appointers. Arbitrators may vote in biased ways to increase reappointments by shaping their reputations over time, or they may vote to the favor of appointing entities responsible for more of their past appointments. These two factors are closely tied to a defining feature of international arbitration: the appointment of adjudicators for individual cases by the disputing parties. Understanding how this feature might have negatively affected arbitrators' independence and impartiality is important for evaluating the reform proposal for establishing an investment court, because *ad hoc* appointment is an important feature that distinguishes arbitrators from judges. I find no evidence that arbitrators are biased due to either reappointment interests or appointment history. In addition, contrary to becoming more pro-investor or pro-state over time, I find arbitrator reputations tend to shift in a neutral direction towards the average of all ICSID arbitration cases. These findings lend credence to the view that an investment court system may not be more effective in achieving independence and impartiality among adjudicators of investor-state disputes.

The third chapter – Domestic Politics and Settlement in Investor-State Arbitration – explores potential causes of prolonged and costly ISDS proceedings by examining state settlement behavior. I posit that elected officials in respondent states suffer "domestic audience costs" for backing down in these high-stake investor-state disputes. Through elections and other mechanisms, domestic audiences can penalize elected officials for appearing weak internationally. The potential for such domestic backlash may constrain state leaders from pursuing settlement strategies that would otherwise be most efficient for the state. This in turn results in wasted legal expenditure and prolonged uncertainty for both the claimant and other potential investors. To empirically examine

the effects of domestic audience costs on case settlement, I use the time left until the next constitutionally mandated election in the respondent state (*Share of Term Left*) as a proxy for the size of anticipated domestic audience costs. As elections approach, domestic audience costs are magnified, because government decisions made closer to elections are more likely to impact voters at the ballot box and bring political consequences for the government. On the other hand, because a country's election term is usually stipulated in its constitution, *Share of Term Left* is unlikely to be correlated with other factors that may also affect settlement, such as case quality. Thus, the research design allows me to draw causal inferences about the effects of domestic audience costs on case settlement in ISDS. I find consistent evidence that case settlement probability decreases as elections approach in respondent states. This finding gives rise to the concern that due to domestic political influences, settlement may not function as effectively a filter mechanism as it should, which contributes to expensive and prolonged ISDS proceedings.

DEVELOPMENT STATUS AND DECISION-MAKING IN INVESTMENT TREATY ARBITRATION

I. INTRODUCTION

In recent years, developing countries like Bolivia, Ecuador and Venezuela have announced their withdrawal from the Convention on the Settlement of Investment Disputes between States and Nationals of Other States (the "ICSID Convention").[1] The ICSID Convention allows private investors to directly bring claims against host states for alleged violations of obligations under international investment agreements (IIAs) through arbitration. This investment treaty arbitration system has been characterized as a powerful sword in the hands of rich developed states against poor developing states.[2] In particular, foreign investors are allegedly more successful when litigating against developing states than developed states.[3] Due to this perception of bias, there is a backlash in the developing world against the investment treaty arbitration system. Not only have a few developing states withdrawn from the ICSID Convention, other developing states like India and Indonesia have terminated their Bilateral Investment Treaties (BITs) in reaction to continuous losses in investment treaty arbitration.[4]

---

[1] *See* ICSID News Release, Venezuela Submits a Notice under Article 71 of the ICSID Convention (January 26, 2012), *available at* https://icsid.worldbank.org/en/Pages/News.aspx?CID=47; ICSID News Release, Denunciation of the ICSID Convention by Ecuador (July 9, 2009), *available at* https://icsid.worldbank.org/en/Pages/News.aspx?CID=87; ICSID News Release, Denunciation of ICSID Convention (May 16, 2007), *available at* https://icsid.worldbank.org/en/Pages/News.aspx?CID=103.

[2] *See* Thomas Schultz and Cedric Dupont, *Investment Arbitration: Promoting the Rule of Law or Over-Empowering Investors? A Qualitative Empirical Study*, 25 THE EUROPEAN JOURNAL OF INTERNATIONAL LAW 1147, 1147 (2014); Leon Trakman, *ICSID Under Siege*, 45 CORNELL INTERNATIONAL LAW JOURNAL 603, 606 (2013).

[3] *See* Daniel Behn, Taraild Laudal Berge & Malcolm Langford, *Poor States or Poor Governance? Explaining Outcomes in Investment Treaty Arbitration* (2017), *available at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2978546.

[4] Alison Ross, India's Termination of BITs to Begin, Global Arbitration Review (March 22, 2017), https://globalarbitrationreview.com/article/1138510/indias-termination-of-bits-to-begin; Ben Bland and Shawn Donnan, Indonesia to Terminate More Than 60 Bilateral Investment Treaties, Financial Times (March 26, 2014), https://www.ft.com/content/3755c1b2-b4e2-11e3-af92-00144feabdc0.

Is the investment treaty arbitration system biased against developing states? A few studies have examined this issue from various perspectives but arrived at inconsistent conclusions.[5] These studies have been frequently relied on as evidence to vindicate each side's arguments in the debate about whether to reform the current system.[6] Yet, these studies have often been questioned for dataset and methodology limitations, which is especially concerning given the far-reaching policy impacts they have. Thus, it is important and necessary to replicate these studies to assess their credibility and evaluate the validity of corresponding policy proposals.

I select two widely-cited empirical studies whose results lead to contradictory implications: Susan Franck's *Development Status and Outcomes of Investor Treaty Arbitration* (2009) and Gus Van Harten's *Arbitrator Behavior in Asymmetrical Adjudication: An Empirical Study of Investment Treaty Arbitration* (2012).[7] These two studies provide valuable sources for replication studies because they are both influential in the field of international investment arbitration and

---

[5] *See* Susan Franck, *Development Status and Outcomes of Investment Treaty Arbitration*, 50.2 HARV. INT'L L. J., 435 (2009) [hereinafter *Development and Outcome*]; Gus Van Harten, *Arbitrator Behavior in Asymmetrical Adjudication: An Empirical Study of Investment Treaty Arbitration*, 50 OSGOODE HALL LAW JOURNAL, 211 (2012) [hereinafter *Asymmetrical Adjudication*]; Gus Van Harten, *Arbitrator Behavior in Asymmetrical Adjudication (Part Two): An Empirical Study of Investment Treaty Arbitration*, 53 OSGOODE HALL LAW JOURNAL, 540 (2016) [hereinafter *Asymmetrical Adjudication Part II*]; Susan Franck, *Conflating Politics and Development? Examining Investment Treaty Arbitration Outcomes*, 55 VA. J. INT'L L. 1,15 (2014) [hereinafter *Conflating Politics and Development*]; Schultz and Dupont, *supra* note 2; Anton Strezhnev, *Why Rich Countries Win Investment Disputes: Taking Selection Seriously* (2017), *available at* *https://static1.squarespace.com/static/5931baca440243906ef65ca3/t/ 59c55e2829f187ed71aba071/1506106921710/why_rich_countries_win_investment_disputes.pdf*; Behn, Berge & Langford, *supra* note 3.

[6] For studies citing Franck, see e.g. Stephen E. Blythe, *The Advantages of Investor-State Arbitration as a Dispute Resolution Mechanism in Bilateral Investment Treaties*, 47.2 THE INTERNATIONAL LAWYER 273, 278 (2013); Jason Webb Yackee, *Controlling the International Investment Law Agency*, 53.2 HARVARD INTERNATIONAL LAW JOURNAL 391, 430 (2012); Charles N. Brower & Sadie Blanchard, *What's in a Meme - The Truth about Investor-State Arbitration: Why It Need Not, and Must Not, Be Repossessed by States*, 52.3 COLUMBIA JOURNAL OF TRANSNATIONAL LAW 689, 711 (2014). For studies citing Van Harten, see e.g. Pia Eberhardt, Cecilia Olivet, Tyler Amos & Nick Buxton, Profiting from Injustice: How Law Firms, Arbitrators and Financiers Are Fueling An Investment Arbitration Boom, Corporate Europe Observatory and the Transnational Institute (November 2012), at 48, *available at* https://www.tni.org/files/download/profitingfrominjustice.pdf; Ciaran Cross & Christian Schliemann-Radbruch, *When Investment Arbitration Curbs Domestic Regulatory Space: Consistent Solutions through Amicus Curiae Submissions by Regional Organizations*, 6.2 LAW AND DEVELOPMENT REVIEW 67, 68 (2013).

[7] According to Google Scholar, by October 3, 2017, Franck (2009) has been cited 153 times, and Van Harten (2012) has been cited 102 times. Franck, *Development and Outcome*, *supra* note 5; Van Harten, *Asymmetrical Adjudication*, *supra* note 5.

controversial in terms of their methodologies and results. They have been considered as representative empirical works in investment treaty arbitration.[8] They have also been cited in policy reports by national government and international research institutes.[9] Meanwhile, these two studies have also been questioned because of the small size of datasets they used,[10] the lack of control for case merits,[11] and the failure to consider other measures of bias against developing countries.[12] Notably, the necessity of future replication studies has been acknowledged by authors of the original studies themselves.[13]

In this paper, I first attempt to reproduce the results in these two studies. To do so, I attempt to gather the original data used for the studies (the "reproduction datasets") and identify the original estimation specifications used in the studies.[14] I then focus on a common question examined by both studies: whether investment treaty arbitration is biased against developing respondent states. On this question, both studies have found null results. I replicate and extend the

---

[8] *See* Gregory Shaffer & Tom Ginsburg, *The Empirical Turn in International Legal Scholarship*, 106 AMERICAN JOURNAL OF INTERNATIONAL LAW 1, 36 (2012) (referring to Franck (2009) as an example of empirical studies in the field of investment arbitration); Suha Jubran Ballan, *Investment Treaty Arbitration and Institutional Backgrounds: An Empirical Study*, 34 WISCONSIN INTERNATIONAL LAW JOURNAL 31, 63 (2016) (citing both Franck (2009) and Van Harten (2012) when discussing empirical studies that evaluate challenges the investment treaty arbitration faces).

[9] *See* Report of the Subcommittee on Investment of the Advisory Committee on International Economic Policy Regarding the Model Bilateral Investment Treaty: Annexes, U.S. Department of State (September 30, 2009), *available at* https://2009-2017.state.gov/e/eb/rls/othr/2009/131118.htm#6n2; Eberhardt, Olivet, Amos & Buxton, *supra* note 6.

[10] *See* WON KIDANE, THE CULTURE OF INTERNATIONAL ARBITRATION 143 (2017).

[11] *See* Catherine A. Rogers, *The Politics of International Investment Arbitrators*, 12 SANTA CLARA JOURNAL OF INTERNATIONAL LAW 223, 234 (2013).

[12] *See* Kevin P. Gallagher & Elen Shrestha, *Investment Treaty Arbitration and Developing Countries: A Re-Appraisal*, 12 J. WORLD INVESTMENT & TRADE 919 (2011).

[13] *See* Franck, *Development and Outcome*, *supra* note 5, at 463, 475 ("Replicating the analysis with a larger sample could ascertain whether there is a detectable and reliable effect among the World Bank status of the respondent, the World Bank status of the presiding arbitrator, and the arbitration outcome…Replication with expanded data is necessary to avoid establishing a population parameter that may be due to chance alone."); Van Harten, *Asymmetrical Adjudication*, *supra* note 5, at 232, 251 ("For purposes of transparency and replication, the template is appended and case-by-case coded data are publicly available…Analysis of data on arbitrator decision-making provides one way in which to research arbitrator behavior, but should be applied in tandem with other methods").

[14] The reproduction dataset for *Franck (2009)* can be found at https://law2.wlu.edu/faculty/page.asp?pageid=1185. The dataset for *Van Harten (2012)* is available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2149256. This published dataset does not contain information on treaty types, arbitrator names, and total appointments per arbitrator, which are used in the original study. Through following correspondences with Professor Van Harten, I obtain the full dataset that contains information on all missing variables except for total appointments per arbitrator. I use the term "reproduction dataset" to refer to the full dataset.

original studies in the following way: (1) using the two reproduction datasets, I use alternative estimation strategies to examine the common question; (2) using a more up-to-date and comprehensive dataset, I reexamine the common question, taking into account different confounding factors.

A major confounding factor is potential differences in case strength between cases against developing states and cases against developed states. It is difficult to draw meaningful inferences about biases against developing states from case outcomes alone without considering case merits. On the legal side, it is possible that investors win more cases against developing states because developing states tend to commit to broader obligations in the IIAs they signed.[15] On the factual side, it is possible that developing states have poorer governance and thus are more likely to violate their IIA obligations.[16] It is also possible that developed states are better at filtering out stronger cases of investors at the settlement stage and thereby the cases that result in final awards against developed states are weaker and consequently have higher state win rates.[17]

In the extension, I take two steps to address this issue. First, I add controls that account for the investment protection level of the underlying IIA and the respondent state's governance quality, which could be correlated with unobserved case strength. Second, I use the Heckman probit selection model to correct for potential selection bias caused by settlement. I find that more developed respondent states are significantly less likely to lose in investment treaty arbitration cases than less developed respondent states. The results remain robust after adding different controls and using the Heckman probit selection model.

---

[15] Lauge Poulsen & Emma Aisbett, *When the Claim Hits: Bilateral Investment Treaties and bounded Rational Learning*, 65(2) WORLD POL. 273 (2013).

[16] *See* Behn, Berge & Langford, *supra* note 3.

[17] Strezhnev, *supra* note 5.

The contributions of this paper are threefold: First, by reproducing *Franck (2009)* and *Van Harten (2012)*, this paper validates some results and finds different results in the reproduction of the others. Second, by replicating *Franck (2009)* and *Van Harten (2012)* using the original datasets, this paper checks the robustness of the original findings to alternative estimation strategies. Finally, by extending *Franck (2009)* and *Van Harten (2012)* using a more up-to-date and comprehensive dataset, this paper takes into account potential confounds that could affect unobserved case strength. This paper identifies consistent negative correlations between respondent states' development status and their likelihood of losing, indicating potential biases against developing respondent states in investment treaty arbitration.

The remainder of this paper proceeds as follows: Part II briefly introduces the two original studies; Part III describes the three datasets used in this paper; Part IV reproduces major results in the two original studies; Part V replicates and extends the two original studies, using alternative estimation strategies and the expanded dataset; Part VI concludes.

II. DESCRIPTION OF THE ORIGINAL STUDIES

A. Franck (2009)

Franck (2009) investigates whether the outcome of an investment treaty arbitration case is affected by (1) the development status of the respondent state, (2) the development status of the presiding arbitrator's home state, and (3) the interaction of these two variables. The study applies two criteria to measure the development status of a state: whether the state is an OECD country ("OECD status") or which income level it falls under according to the World Bank's classification ("World Bank status"). [18] There are four income levels: High Income, Upper-Middle

---

[18] The World Bank's main criterion for classifying economies is GNI per capita. *See* The World Bank, Data and Statistics, Country Classification, http://www.worldbank.org/data/couneryclass/countryclass.html.

Income, Lower-Middle Income, and Low Income. For the World Bank status of the presiding arbitrator's home state, the study only distinguishes between High Income states and other states in its analysis, as there are few presiding arbitrators who are from developing countries. To denote case outcomes, the study uses both the total amounts awarded and the ultimate winners.[19]

The study uses a three-way cross-tabulation and chi-squared tests to see if there is a statistically significant pattern of relationship between the development status of respondent states, the development status of the presiding arbitrator's home country (for brevity, "the presiding arbitrator's development status"), and the case winner. The study finds no statistical significance in any of these relationships, though some estimated effect sizes are large. The study uses a two-way Analysis of Variance ("ANOVA") to examine the apparent effects of the respondent state and presiding arbitrator's development status on the damages awarded by tribunals. The reported F statistics and p-values do not reveal any statistically significant relationship between the damages awarded and the presiding arbitrator's development status, the respondent state's development status, or the interaction of these two factors. Using Tukey's HSD test,[20] the study also conducts pairwise comparisons of mean damages awarded against respondent states of different development status conditional on the presiding arbitrator's development status. The study finds that overall the means are not significantly different from each other.[21] The study concludes that neither the development status of the respondent nor the development status of the presiding

---

[19] According to *Franck (2009)*, the claimant is coded as the ultimate winner of a case if it is awarded more than $0. *See* Franck, *Development and Outcome*, *supra* note 5, at 456.

[20] Tukey's honestly significant difference ("HSD") test allows one to compute a single value that determines the minimum difference between treatment means that is necessary for significance. This value, called the honestly significant difference, or HSD, is then used to compare any two treatment conditions. If the mean difference exceeds Tukey's HSD, one can conclude that there is a significant difference between the treatments. *See* FREDERICK J. GRAVETTER & LARRY B. WALLNAU, ESSENTIALS OF STATISTICS FOR THE BEHAVIORAL SCIENCES 329 (2002).

[21] *Franck* (2009) notes that two exceptions exist when using the winsorized data: when the presiding arbitrator is from an Upper- or Lower-Middle Income state, significantly lower damages are awarded against High Income states than those awarded against Upper-Middle Income states or Low Income states. But this pattern is not observable from the raw data. *See* Franck, *Development and Outcome*, *supra* note 5, at 469.

arbitrator's home state affects the case outcome,[22] and that investment treaty arbitration is not per se biased in favor of either the developed or the developing world.[23]

B. Van Harten (2012)

Based on the expectations that arbitrators are incentivized to favor investor claimants to encourage more future cases, and that expansive interpretation of jurisdictional issues benefits investor claimants, *Van Harten (2012)* tests three hypotheses: (1) in investment treaty arbitration cases, arbitrators will adopt expansive approaches to jurisdictional issues more frequently than restrictive approaches; (2) the anticipated tendency toward expansive approaches will be accentuated in cases brought by claimants from France, Germany, the United Kingdom, and the United States (Western capital-exporting states); (3) the anticipated tendency toward expansive approaches will be reduced in cases brought against France, Germany, the United Kingdom, and the United States (Western capital-exporting states). The study identifies seven jurisdictional issues and expansive and restrictive approaches to each issue. The study then codes whether an arbitrator adopts an expansive or restrictive approach to each issue that appears in a case.

The study assumes that the percentage of expansive approaches should be 50%.[24] Using a one-sample binomial test, the study finds that arbitrators tend to adopt expansive approaches more frequently than restrictive approaches. Using a generalized linear mixed effects model,[25] the study

---

[22] *See* Franck, *Development and Outcome*, *supra* note 5, at 464.
[23] *See id.*, at 473.
[24] *See* Van Harten, *Asymmetrical Adjudication*, *supra* note 5, at 237.
[25] Generalized linear mixed effects models (or GLMMs) are an extension of linear mixed models to allow response variables from different distributions, such as binary responses. GLMMs can also be considered as an extension of generalized linear models to include both fixed and random effects. The general form of the model is: $y = X\beta + Zu + \varepsilon$, where y is a $N \times 1$ column vector; X is a $N \times p$ matrix of the p predictor variables; β is a $p \times 1$ column vector of the fixed-effects regression coefficients; Z is the $N \times q$ design matrix for the q random effects; u is a $q \times 1$ vector of the random effects; and ε is a $N \times 1$ column vector of the residuals, the part of y that is not explained by the model, $X\beta + Zu$. For a more detailed introduction of the generalized linear mixed effects model, see CHARLES MCCULLOCH, GENERALIZED, LINEAR AND MIXED MODELS, 220 – 246 (2001).

finds "evidence of a strong tendency in favor of an accentuated expansive approach where the claimant was a national of a Western capital-exporting state," but finds no evidence of a reduced tendency toward expansive approaches where the respondent is a Western capital-exporting state. The study also uses OECD vis-à-vis non-OECD states as an alternative grouping and finds similar results. It concludes that on issues of jurisdiction and admissibility, there is tentative evidence for systematic bias towards Western capital-exporting states.[26]

III. DATASETS

A. Franck (2009)

Franck (2009) uses a dataset of 102 investment treaty awards from 82 cases that were publicly available before June 1, 2006. According to *Franck (2009)*, since 50 out of the 102 investment treaty awards did not resolve investment treaty claims and another 3 cases did not publish nationalities of presiding arbitrators, only 49 awards were analyzed in the study.[27] The reproduction dataset covers the OECD status and World Bank income level of respondent states and presiding arbitrators' home states. It also includes information of the IIA invoked, the total amount awarded to claimants in USD and the ultimate winners of the cases. Table 1.1 presents summary statistics as reported in the original study and of the reproduction dataset. The reproduction dataset summary statistics match the data reported in the original study.

---

[26] *See* Van Harten, *Asymmetrical Adjudication*, *supra* note 5, at 249, 251. A follow-up study conducted by the same author tested the hypotheses on both jurisdictional and substantive issues in the same awards and found similar results. *See* Van Harten, *Asymmetrical Adjudication Part II*, *supra* note 5.

[27] *See* Franck, *Development and Outcome*, *supra* note 5, at 455.

Table 1.1: Summary Statistics in Original Study and of Reproduction Dataset (*Franck (2009)*)

Original Study

Panel A: Breakdown of Presiding Arbitrator's and Respondent State's OECD Status

	Respondent State		
Presiding Arbitrator	*Non-OECD*	*OECD*	*Total*
Non-OECD	10	3	13
OECD	21	15	36
Total	31	18	49

Panel B: Breakdown of Presiding Arbitrator's and Respondent State's World Bank Status

	Respondent State				
Presiding Arbitrator	*High*	*Upper-Middle*	*Lower-Middle*	*Low*	*Total*
High Income	9	13	12	2	36
Upper-Middle Income	1	4	3	0	8
Lower-Middle Income	1	0	2	2	5
Low Income	0	0	0	0	0
Total	11	17	17	4	49

Reproduction Dataset

Panel A: Breakdown of Presiding Arbitrator's and Respondent State's OECD Status

	Respondent State		
Presiding Arbitrator	*Non-OECD*	*OECD*	*Total*
Non-OECD	10	3	13
OECD	21	15	36
Total	31	18	49

Panel B: Breakdown of Presiding Arbitrator's and Respondent State's World Bank Status

	Respondent State				
Presiding Arbitrator	*High*	*Upper-Middle*	*Lower-Middle*	*Low*	*Total*
High Income	9	13	12	2	36
Upper-Middle Income	1	4	3	0	8
Lower-Middle Income	1	0	2	2	5
Low Income	0	0	0	0	0
Total	11	17	17	4	49

B. Van Harten (2012)

Van Harten (2012) identifies 100 investment treaty arbitration cases that result in publicly available English-language awards with at least one jurisdictional issue resolved expansively or restrictively by June 1, 2010. According to *Van Harten (2012)*, there are seven jurisdictional issues that are common in investment treaty arbitration cases. For each issue, the original study publishes

a list of related approaches that are categorized as expansive or restrictive, based on whether the approach enhances or reduces "the compensatory promise of the system for claimants and the risk of liability for states."[28] The reproduction dataset is organized at the arbitrator-issue-vote level. For each arbitrator on the tribunal of each case, a dummy variable is created to denote whether the arbitrator voted expansively or restrictively on an issue that appears in the case. The reproduction dataset also contains information on the nationality of claimants, the identity of respondent states, treaty types, issue types, count of issues per case, case names, arbitrator names, and jurisdictional outcomes.[29]

Table 1.2 presents summary statistics as reported in the original study and of the reproduction dataset. As Table 1.2 shows, the reproduction dataset contains one more case than the original study. Probably due to this difference, the statistics on issue resolutions, claimant nationalities and respondent state identities in the reproduction dataset are slightly different from those reported in the original study. Another variable that generates different statistics is the total appointments per arbitrator. This variable is not included in the reproduction dataset. I calculate each arbitrator's total appointments based on the frequency of their appearances in the dataset.

Table 1.2: Summary Statistics in Original Study and of Reproduction Dataset (*Van Harten (2012)*)[30]

	Original Study		Reproduction Dataset	
Issues	*Percentage*	*N*	*Percentage*	*N*
Corporate Person Investor	13.4	69	13.8	72
Natural Person Investor	1.2	6	1.2	6
Investment	22.5	116	22.8	119
Minority Shareholder Interests	14.0	72	14.4	75
Permissibility of Investment	5.2	27	5.2	27
Parallel Claims	32.0	165	31.1	162
Scope of MFN Treatment	11.7	60	11.5	60
Total	100.0	515	100.0	521

---

[28] *See* Van Harten, *Asymmetrical Adjudication*, *supra* note 5, at 266.
[29] *Jurisdictional Outcome* is a dummy variable that equals 1 if the arbitral tribunal has found jurisdiction in a case.
[30] Percentages may not add up to 100% due to rounding.

Table 1.2 (Continued)

Issue Resolutions	Expansive(%)	Restrictive(%)	Expansive(%)	Restrictive(%)
Corporate Person Investor	81.9	18.1	84.7	15.3
Natural Person Investor	0.0	100.0	0.0	100.0
Investment	72.3	27.7	69.8	30.2
Minority Shareholder Interests	92.0	8.0	92.0	8.0
Permissibility of Investment	66.7	33.3	66.7	33.3
Parallel Claims	82.7	17.3	84.0	16.0
Scope of MFN Treatment	50.0	50.0	50.0	50.0
Total	76.1	23.9	76.2	23.8
Nationality of Claimant	**Percentage**	**N**	**Percentage**	**N**
France	8.7	45	8.6	45
Germany	4.7	24	4.6	24
United Kingdom	9.5	49	9.4	49
United States	27.2	140	26.3	137
All Other States	49.9	257	51.1	266
Total	100.0	515	100.0	521
Identity of Respondent State	**Percentage**	**N**	**Percentage**	**N**
France	0.0	0	0.0	0
Germany	0.0	0	0.0	0
United Kingdom	0.0	0	0.0	0
United States	2.9	15	2.9	15
All Other States	97.1	500	97.1	506
Total	100.0	515	100.0	521
Treaty Type	**Percentage**	**N**	**Percentage**	**N**
BITs	Unclear	Unclear	79.3	413
NAFTA	Unclear	Unclear	7.5	39
Energy Charter Treaty	Unclear	Unclear	11.5	60
ASEAN Agreement for Promotion and Protection of Investments	Unclear	Unclear	1.2	6
BIT and Energy Charter Treaty[31]	NA	NA	0.6	3
Total	Unclear	Unclear	100.0	521
Count of Issues per Case				
Min	1		1	
Max	4		4	
Mean	2		2	
Median	2		2	
Total Appointments per Arbitrator				
Min	1		1	
Max	14		12	

---

[31] This category is not reported in the original study but is included in the reproduction dataset.

Table 1.2 (Continued)		
Mean	5	4
Median	3	2
Cases		
Total	100	101
Arbitrator		
Total	172	172

C. The Expanded Dataset

In the extension, I use a more up-to-date dataset that covers 739 investment treaty arbitration cases registered before August 12, 2016. The dataset contains detailed information of the case outcomes, the types of IIA obligation violations alleged and found, the amount of damages claimed and awarded, the economic sectors and subsectors the disputes arise in, the IIAs invoked, the respondent states and home states of investors, the GDP per capita of the respondent state in 2005 USD, and the polity score of the respondent state.[32] Based on which IIAs are invoked in a case, I merge this dataset with United Nations Conference on Trade and Development (UNCTAD) IIA Database, which contains detailed information of the types of obligations covered in an IIA and the limitations added to each obligation in the IIA.[33]

Most investment treaty claims are based on certain obligations that the respondent states have committed to in the IIAs they signed. These IIAs usually contain similar sets of obligations and use standardized language to describe them. A major variance in IIA content is the number and types of limitations added to an obligation, which are used by states to restrict their obligations. For example, for the fair and equitable treatment obligation, contracting states could limit their

---

[32] The investment treaty arbitration case information is mainly collected from the UCTAD IIA Arbitration Database. http://investmentpolicyhub.unctad.org/ISDS. The polity score and GDP data are imported from the Polity IV project and the World Development Indicators datasets, respectively. *See* Polity IV project, http://www.systemicpeace.org/polity/polity4.htm; World Development Indicators, https://data.worldbank.org/products/wdi.
[33] Of the 739 cases, I am able to find IIA obligation information for 687 cases. The UNCTAD IIA Database is available at http://investmentpolicyhub.unctad.org/IIA.

obligations by specifying that they only commit to the treatment standard required by the customary international law,[34] or by enumerating which state actions constitute a violation of this obligation.[35] Investors whose claims rely on IIAs with more restrictive obligations could have weaker cases on the law and a lower chance to win, because there is less room for broad interpretation of state obligations.

To measure the restrictiveness of the investment protection obligations in the underlying IIA, I identify seven major substantive obligations that are frequently invoked in investment treaty arbitration cases.[36] For each obligation, the most restrictive scenario is when the IIA does not include the obligation, and the least restrictive scenario is when the IIA includes the obligation and does not include any limitation. For the intermediary scenarios, the restrictiveness of the obligation depends on the number of limitations added. Based on this spectrum of restrictiveness, for each obligation in each IIA, I assign a score of 0 if the IIA includes the obligation and a score of 1 if not. For each limitation added to that obligation, I increase the score by 0.1. An obligation with two limitations added will have a score of 0.2, whereas an obligation with five limitations added will have a score of 0.5. Adding up the score for each obligation and dividing it by seven, I generate a new variable *Restrictiveness*, which measures the overall restrictiveness of the obligations in the underlying IIA.[37] The larger the value of *Restrictiveness* is, the more restrictive the obligations in the underlying IIA are, the less room there exists for broad interpretation of IIA obligations and

---

[34] *See e.g.* Peru-Belgium Luxembourg BIT, art. 3 ("All investments made by investors of one Contracting Party shall enjoy a fair and equitable treatment in the territory of the other Contracting Party, in accordance with customary international law.")

[35] *See e.g.* Canada-Republic of Korea FTA, art. 8.5.

[36] These seven obligations are: most favored nations, national treatment, fair and equitable treatment, direct expropriation, indirect expropriation, full protection and security and performance requirement.

[37] To take into account diminishing returns in restrictiveness for additional limitations added, I have also used an alternative way to measure the restrictiveness of the underlying IIA. For each obligation in each IIA, I assign a score of –1 if the IIA includes the obligation, and a score of 0 if not. For each limitation added to that obligation, I multiply the score by 0.9. The regression results using this alternative control is quantitatively and qualitatively comparable to those in Table 1.8.

16

for findings of violation of such obligations. Thus, investors whose claims rely on more restrictive IIAs are likely to have weaker cases on the law.

Panel A of Table 1.3 presents summary statistics on *Restrictiveness*. Panel B of Table 1.3 presents respondent win rate and mean *Restrictiveness* for respondent states of different development statuses. Within decided cases, the mean *Restrictiveness* for OECD respondent states is lower than non-OECD respondent states, which indicates that the underlying IIAs of cases involving OECD respondent states are less restrictive than the underlying IIAs of cases involving non-OECD respondent states. Nevertheless, OECD respondent states have a much higher win rate than non-OECD respondent states. In terms of World Bank status, the mean *Restrictiveness* for High Income and Upper-Middle Income respondent states is lower than Lower-Middle Income respondent states, which indicates that the underlying IIAs of cases involving High Income and Upper-Middle Income respondent states are less restrictive than those of cases involving Lower-Middle Income respondent states. However, High Income and Upper-Middle Income respondent states have a higher win rate than Lower-Middle Income respondent states.

Supplementary summary statistics of other variables are presented in the Appendix (Table A1.1).

Table 1.3: Summary Statistics of The New Dataset

Panel A: Summary Statistics, Restrictiveness						
Variable	Mean	Median	Max.	Min.	Std. Dev.	N
Restrictiveness	0.209	0.200	0.714	0.029	0.109	686
Panel B: Summary Statistics, by Respondent Development Status (Decided Cases)						
Respondent OECD Status	OECD		Non-OECD		Total	
Respondent Win Rate	73.3% (N = 90)		53.9% (N = 219)		59.5% (N = 309)	
Restrictiveness	0.173 (N = 83)		0.215 (N = 199)		0.203 (N = 282)	

Table 1.3 (Continued)

Respondent World Bank Status	High Income	Upper-Middle Income	Lower-Middle Income	Low Income	Total
Respondent Win Rate	76.5% (N = 81)	54.6% (N = 141)	53.3% (N = 75)	41.7% (N = 12)	59.5% (N = 309)
Restrictiveness	0.206 (N = 72)	0.196 (N = 134)	0.220 (N = 66)	0.151 (N = 10)	0.203 (N = 282)

IV. REPRODUCING RESULTS OF THE ORIGINAL STUDIES

A. Franck (2009)

Franck (2009) conducts multiple tests to examine the relationship between the development status of respondent states, the development status of presiding arbitrators' home states, and the case outcome. I have reproduced all those tests.[38] I present the reproduction results of the three-way cross tabulation here.[39] Like *Franck (2009)*, for each category of presiding arbitrator development status, I report the chi-squared test statistics that indicate whether the observed difference in winning frequencies is statistically significant for different respondent state development statuses.[40] I report Cramér's V as a measure of effect sizes.[41]

Table 1.4 presents the original estimates and reproduction estimates. As Table 1.4 shows, the reproduction estimates are mostly the same as the original estimates. Panel A of Table 1.4 focuses on OECD status. For cases with presiding arbitrators from either OECD or non-OECD

---

[38] The reproduction results for the other tests are mostly the same as the results reported in the original study. These reproduction results are on file with the author.

[39] In Stata, I use command "table Winner Respondent_OECD President_OECD, c(freq)" to examine the relationship between the OECD status of respondent states, the OECD status of the presiding arbitrator, and the case outcome; I use command "table Winner Respondent_WB President_WB, c(freq)" to examine the relationship between the World Bank status of respondent states, the World Bank status of the presiding arbitrator, and the case outcome.

[40] In Stata, I use command "tab Winner Respondent_OECD if President_OECD == 1, all label" and "tab Winner Respondent_OECD if President_OECD == 0, all label" to calculate the Pearson χ^2 test for the independence of OECD and non-OECD respondent states' wins; I use command "tab Winner Respondent_WB if President_WB == 1, all label" and "tab Winner Respondent_WB if President_WB == 2, all label" to calculate the Pearson χ^2 test for the independence of OECD and non-OECD respondent states' wins.

[41] *Franck (2009)* reports effect sizes, but does not specify which measure of effect sizes she uses.

states, the original and reproduction estimates of chi-squared test statistics and effect sizes are exactly the same, reporting no statistically significant relationship between the OECD status of respondent states and case outcomes.[42] Panel B of Table 1.4 focuses on World Bank status. For cases with presiding arbitrators from either High Income or Upper/Lower-Middle Income countries, the original and reproduction estimates of chi-squared test statistics are exactly the same, reporting no statistically significant relationship between the World Bank status of respondent states and case outcomes. The reproduction estimates of effect size are slightly different from those reported in the original study, though both estimates are pretty large.[43]

Table 1.4: Development Status of Respondent State, Development Status of Presiding Arbitrator, and Frequency Breakdown of the Case Winner

Presiding Arbitrator	Respondent State	Claimant Wins	Respondent Wins	Total	State Win Rate (%)
Panel A: OECD Status					
Original Estimates					
Non-OECD	$(\chi^2(1) = .258; p = .61; r = .14)$				
	Non-OECD	5	5	10	50.00
	OECD	2	1	3	33.33
OECD	$(\chi^2(1) = .045; p = .83; r = .04)$				
	Non-OECD	7	12	19	63.16
	OECD	5	10	15	66.67
Total		19	28	47[44]	59.57
Reproduction Estimates					
Presiding Arbitrator	Respondent State	Claimant Wins	Respondent Wins	Total	State Win Rate (%)
Non-OECD	$(\chi^2(1) = .258; p = .61; r = .14)$				
	Non-OECD	5	5	10	50.00
	OECD	2	1	3	33.33
OECD	$(\chi^2(1) = .045; p = .83; r = .04)$				
	Non-OECD	7	12	19	63.16
	OECD	5	10	15	66.67
Total		19	28	47	59.57

---

[42] As *Franck (2009)* notes, the effect size for cases with presiding arbitrators from non-OECD states is relatively large here, suggesting potential statistical power problems that require a larger sample. *See* Franck, *Development and Outcome*, *supra* note 5, at 457 – 458.

[43] As *Franck (2009)* acknowledges, the large effect size suggests that the results may have weak statistical power. *See* Franck, *Development and Outcome*, *supra* note 5, at 463.

[44] Two cases are dropped from the sample because the awards contain settlement agreements. At 455.

Table 1.4 (Continued)

Panel B: World Bank Status

Original Estimates

Presiding Arbitrator	Respondent State	Claimant Wins	Respondent Wins	Total	State Win Rate (%)
High Income	(χ^2(df>1) = 2.216; p = .53; r = .25)				
	High Income	3	6	9	66.67
	Upper-Middle Income	5	8	13	61.54
	Lower-Middle Income	3	8	11	72.73
	Low Income	1	0	1	0.00
Upper-Middle and Lower-Middle Income	(χ^2(df>1) = 1.130; p = .77; r = .28)				
	High Income	1	1	2	50.00
	Upper-Middle Income	3	1	4	25.00
	Lower-Middle Income	2	3	5	60.00
	Low Income	1	1	2	50.00
Total		19	28	47	59.57

Reproduction Estimates

Presiding Arbitrator	Respondent State	Claimant Wins	Respondent Wins	Total	State Win Rate (%)
High Income	(χ^2(3) = 2.216; p = .53; r = .255)				
	High Income	3	6	9	66.67
	Upper-Middle Income	5	8	13	61.54
	Lower-Middle Income	3	8	11	72.73
	Low Income	1	0	1	0.00
Upper-Middle and Lower-Middle Income	(χ^2(3) = 1.130; p = .77; r = .295)				
	High Income	1	1	2	50.00
	Upper-Middle Income	3	1	4	25.00
	Lower-Middle Income	2	3	5	60.00
	Low Income	1	1	2	50.00
Total		19	28	47	59.57

B. Van Harten (2012)

I have reproduced results for all three hypotheses that *Van Harten (2012)* examines.[45] I present the reproduction results for Hypothesis (2) here because it is the focus in *Van Harten (2012)*. I use the same generalized linear mixed effects model as *Van Harten (2012)* does.[46] Like

---

[45] The reproduction results for the other two hypotheses are qualitatively similar to the original results, but for some results the significance level is different. These reproduction results are on file with the author.

[46] In the paper, *Van Harten (2012)* does not mention which distribution has been specified for the generalized linear mixed effects model. The R codes that Professor Van Harten provides reveal that a binomial distribution has been

Van Harten (2012), I control for the following fixed effects predictors: claimant nationality, issue type, count of issues per case, treaty type, and total appointments per arbitrator. Like *Van Harten (2012)*, I code the claimant nationality as a categorical variable that represents five groups of states: France, Germany, United States, United Kingdom and all other states. Issue type and treaty type are also specified as categorical variables. Like *Van Harten (2012)*, I control for random effects at the case-level and arbitrator-level.

Table 1.5 presents the original and reproduction estimates of expansive issue resolution probability for each claimant nationality category.[47] Other States is set as the baseline category.[48] Statistical significance of the coefficients for the other categories is measured relative to the baseline. As Table 1.5 shows, the magnitudes of expansive issue resolution probabilities for most claimant nationality categories are qualitatively similar to the original estimates. However, in my reproduction results, the probability of expansive issue resolutions for claimants from France, United Kingdom, or United States are not significantly higher than claimants from other states. The probability of expansive issue resolutions for German claimants is significantly lower than claimants from other states.

Table 1.5: Effect of Claimant Nationality on the Likelihood of Expansive Resolution

	Original Estimates	*Reproduction Estimates*
	Expansive	*Expansive*
France	0.86	0.90
	(p = 0.005)	(p = 0.539)
Germany	0.47	0.03
	(p = 0.38)	(p = 0.008)

specified. However, when I use the generalized linear mixed effects model with a binomial distribution in Stata, no results were generated because the interactions do not converge. Thus, for Hypotheses (2) and (3), I use R instead.

[47] The R codes that Professor Van Harten provides show that claimant nationality, issue type and treaty type have been specified as categorical variables. Following his estimation strategy, in R, I use command "fitFactor <- glmer(Resolution~as.factor(ClaimantNationality)+as.factor(Issue)+as.factor(Treaty)+Total_issues+Total_appointme nts+(1|Arbitrator/CaseName),dataM,family="binomial",na.action=na.omit)".

[48] The R codes that Professor Van Harten provides show that Other States is set as the baseline category in the original study.

Table 1.5 (Continued)

United Kingdom	0.95	0.95
	(p < 0.001)	(p = 0.074)
United States	0.98	0.86
	(p < 0.001)	(p = 0.528)
Other States	0.69	0.77
	(p < 0.001)	
Issue Type	Yes	Yes
Treaty Type	Yes	Yes
Total Issues per Case	Yes	Yes
Total Appointments per Arbitrator	Yes	Yes
Observations	515	521

Note: Like *Van Harten (2012)*, I report the predicted probability of expansive approaches here. It appears that the p-values reported in *Van Harten (2012)* reflect statistical significance different from zero. In the reproduction test, Other States is set as the baseline category. Statistical significance of the coefficients for the other categories is measured relative to the baseline.

V. REEXAMINING BIASES AGAINST DEVELOPING RESPONDENT STATES

A. Replication

Using the reproduction datasets, I run linear and logistic regressions of case outcomes on the development status of respondent states, controlling for covariates used in the original studies. Linear and logistic regressions are familiar statistical models in empirical legal studies (unlike generalized linear mixed effects models), and using these models should make the results more comparable with the wider literature on the determinants of arbitration and litigation results. This empirical strategy allows us to at least understand if there is a statistically significant correlation between case outcomes and the development status of respondent states. I will address causal identification issues in the following section using the expanded dataset.

Using the *Franck (2009)* dataset, I examine the relationship between investor wins and the OECD status and World Bank status of respondent states. All variables I use are from the *Franck (2009)* dataset. The outcome variable is a dummy variable that equals 1 if the claimant is coded as the ultimate winner of a case. The primary explanatory variable of interest is the respondent state's

OECD status or World Bank status. I correct a few mistakes of respondent states' World Bank status in the reproduction dataset, pursuant to the 2008 World Bank classification.[49] I run both linear and logit regressions (reporting marginal effects). A coefficient of 0.01 would indicate a 1 percentage point increased chance of investor winning. All standard errors are clustered at the IIA level.[50]

Table 1.6 presents the results. Columns (2) and (4) add the presiding arbitrator's development status as a control variable. Panel A focuses on OECD status. All results are null. None of the coefficient magnitudes is particularly large, especially after controlling for the presiding arbitrator's development status. Panel B focuses on World Bank status. Lower-Middle Income states are specified as the baseline category.[51] None of the results are statistically significant. The point estimates of the likelihood of investor wins for Low Income states are fairly large in magnitude relative to the likelihood of investor wins for Lower-Middle Income states. Overall, there is no evidence suggesting a statistically significant correlation between the respondent state's OECD status or World Bank status and investor wins. It is worth noting that the standard errors in most specifications are pretty large, suggesting that the results are not precise enough to rule out the possibility of a correlation.

---

[49] Based on the 2008 World Bank classification, I find three respondent states whose status is different from the status in the reproduction dataset: Algeria (upper-middle income according to the 2008 World Bank classification but coded as lower-middle income in the original dataset), Peru (upper-middle income according to the 2008 World Bank classification but coded as lower-middle income in the original dataset), Slovak Republic (high income according to the 2008 World Bank classification but upper-middle income in the original dataset). The 2008 World Bank classification of country incomes is available at http://web.worldbank.org/WBSITE/EXTERNAL/ DATASTATISTIC/0,,contentMDK:20487070~menuPK:64909257~pagePK:64909151~piPK:64909148~theSi tePK:6950074~isCURL:Y,00.html.

[50] Each IIA that has been invoked in cases is a cluster.

[51] I did not specify Low Income states as the baseline category because there are very few Low Income states in the dataset.

Table 1.6: Respondent's Development Status and Case Outcome, *Franck (2009)* Dataset

Panel A: OECD Status

	Logit		Linear	
	(1)	(2)	(3)	(4)
Respondent_OECD	-0.0511	0.0085	-0.0509	0.00847
	(0.122)	(0.125)	(0.122)	(0.127)
President_OECD		-0.180		-0.187
		(0.154)		(0.171)
Observations	50	47	50	47

Panel B: World Bank Status

	Logit		Linear	
	(1)	(2)	(3)	(4)
High Income	-0.0495	0.00724	-0.0495	0.00779
	(0.153)	(0.153)	(0.158)	(0.159)
Upper-Middle Income	0.0929	0.110	0.0929	0.110
	(0.163)	(0.159)	(0.168)	(0.168)
Low Income	0.310	0.261	0.310	0.258
	(0.301)	(0.343)	(0.311)	(0.340)
President_High Income		-0.159		-0.166
		(0.152)		(0.172)
Observations	50	47	50	47

Note: The outcome variable is a dummy variable *Investor Win*, which equals 1 if the claimant is awarded more than \$0. Logit coefficients are reported as marginal effects. "Lower-Middle Income" is specified as the baseline category. Standard errors, in parentheses, are clustered at the IIA level; $^{*} p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$.

Using the *Van Harten (2012)* dataset, I examine the relationship between jurisdictional outcomes and the OECD status and World Bank status of respondent states. The outcome variable is a dummy variable that equals 1 if the arbitral tribunal has found jurisdiction in a case. The primary explanatory variable of interest is the respondent state's OECD status or World Bank status. Since there are only three cases against United States and no cases against other "western capital-exporting states" in the dataset, I do not include this grouping in the analysis. I adjust the dataset to the case-level. I also create seven dummy variables to denote whether each of the seven

jurisdictional issues identified in *Van Harten (2012)* appears in a case. These issue type dummies are controlled for in alternative specifications. I run both linear and logit regressions (reporting marginal effects). All standard errors are clustered at the IIA level.

Table 1.7 presents the results. Treaty type dummies, total issues per case, and issue type dummies are added as control variables in Columns (2) – (4) and (6) – (8). Panel A focuses on OECD status. Although all coefficients for respondent's OECD status are negative and their magnitudes are fairly consistent across regressions, the significance disappears after more controls have been added. Thus, there is no consistent evidence of a negative correlation between the respondent's OECD status and the arbitral tribunal finding jurisdiction in a case. Panel B focuses on World Bank status. Lower-Middle Income states are specified as the baseline category. None of the results are statistically significant. The magnitude of coefficients for High Income states become very small after controlling for issue type dummies. The coefficients for Upper-Middle Income states are all positive. Overall, there is no consistent evidence suggesting that less developed respondent states are more likely to lose at the jurisdiction phase than more developed respondent states. Again, the standard errors in most specifications are pretty large, suggesting that the results are not precise enough to rule out the possibility of a correlation.

Table 1.7: Respondent's Development Status and Jurisdiction Outcome, *Van Harten (2012)*
Dataset

Panel A: OECD Status

	Logit				Linear			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Respondent_OECD	-0.198***	-0.190	-0.193	-0.157	-0.253**	-0.235	-0.236	-0.177
	(0.056)	(0.102)	(0.101)	(0.119)	(0.092)	(0.164)	(0.163)	(0.162)
Treaty Dummies	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Total Issues per Case	No	No	Yes	Yes	No	No	Yes	Yes
Issue Type Dummies	No	No	No	Yes	No	No	No	Yes
Observations[52]	99	91	91	66	99	99	99	99

Panel B: World Bank Status

	Logit				Linear			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High Income	-0.178	-0.145	-0.148	-0.008	-0.178	-0.152	-0.152	-0.061
	(0.133)	(0.149)	(0.148)	(0.153)	(0.135)	(0.139)	(0.140)	(0.135)
Upper-Middle Income	0.126	0.137	0.131	0.123	0.126	0.124	0.125	0.092
	(0.090)	(0.102)	(0.103)	(0.117)	(0.091)	(0.093)	(0.095)	(0.099)
Low Income								
Treaty Dummies	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Total Issues per Case	No	No	Yes	Yes	No	No	Yes	Yes
Issue Type Dummies	No	No	No	Yes	No	No	No	Yes
Observations[53]	99	91	91	66	100	100	100	100

Note: The outcome variable is a dummy variable *Jurisdictional Outcome*, which equals 1 if the tribunal has found jurisdiction in a case. Logit coefficients are reported as marginal effects. "Lower-Middle Income" is specified as the baseline category. The coefficients for "Low Income" are missing because there is only one observation in this category. Standard errors, in parentheses, are clustered at the IIA level; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

B. Extension

---

[52] One case is omitted in all specifications because their jurisdictional outcomes are unavailable in the dataset.
[53] One case is omitted in all specifications because their jurisdictional outcomes are unavailable in the dataset.

The absence of evidence for biases against developing respondent states in the replication can result from the small size of the two reproduction datasets. In the extension, I use the expanded dataset, which contains more recent cases and allows one to control for a number of confounding factors that could be correlated with case strength in terms of the law and facts. I run linear and logit regressions similar to the ones in the replication to examine whether investment treaty arbitration is biased against developing respondent states. The outcome variable is a dummy variable that equals 1 if the claimant is awarded more than $0. The primary explanatory variable of interest is the respondent state's OECD status or World Bank status. In alternative specifications, in addition to treaty type dummies, total issues per case, and issue type dummies, I also account for the restrictiveness of the underlying IIAs (*Restrictiveness*) and the respondent's governance quality (*Polity Score*)[54]. All standard errors are clustered at the IIA level.

Table 1.8 presents regression results. Columns (1) – (4) represent the same sets of regressions as those used in the replication. Columns (5) – (6) add *Restrictiveness* and *Polity Score* as control variables. Panels A and B focus on OECD status. The coefficients for respondent's OECD status are significantly negative in all specifications, suggesting that OECD respondent states are significantly less likely to lose than non-OECD respondent states, even after controlling for the restrictiveness of the underlying IIAs and the respondent's governance quality. The magnitudes of the estimated coefficients are fairly consistent across specifications. It appears that the respondent state's OECD status is associated with approximately an 18 percentage point decrease in the probability for respondents to lose a case. Panels C and D focus on World Bank Status. Lower-Middle Income states are specified as the baseline category. The coefficients for

---

[54] The *Polity Score* variable from the Polity IV index is often used in political science literature as a proxy for general levels of governance quality. It is based on the competitiveness of political participation, the openness of executive recruitment and constraints on the chief executive. *See* Monthy Marshall, Ted Gurr & Keith Jaggers, Polity IV Project: Dataset Users' Manual, www.systemicpeace.org/inscrdata.html.

High Income states are significantly negative in all specifications, suggesting that High Income respondent states are significantly less likely to lose than Lower-Middle Income respondent states by approximately 22%. None of the coefficients for Upper-Middle Income respondent states are statistically significant, and the point estimates are small in magnitude. Most coefficients for Low Income respondent states are not statistically significant. But the point estimates are not small. Overall, there is no consistent evidence suggesting statistically significant difference in the likelihoods of losing a case between Lower-Middle Income respondent states and Upper-Middle or Low Income respondent states.

Table 1.8: Respondent's Development Status and Case Outcome, Expanded Dataset

Panel A: OECD Status (Logit)						
	(1)	(2)	(3)	(4)	(5)	(6)
Respondent_OECD	-0.200***	-0.195*	-0.200*	-0.179*	-0.172*	-0.177*
	(0.058)	(0.074)	(0.079)	(0.069)	(0.073)	(0.084)
Treaty Dummies	No	Yes	Yes	Yes	Yes	Yes
Total Issues per Case	No	No	Yes	Yes	Yes	Yes
Issue Type Dummies	No	No	No	Yes	Yes	Yes
Restrictiveness	No	No	No	No	Yes	Yes
Polity Score	No	No	No	No	No	Yes
Observations	309	309	309	309	282	275
Panel B: OECD Status (Linear)						
	(1)	(2)	(3)	(4)	(5)	(6)
Respondent_OECD	-0.195***	-0.192*	-0.197**	-0.182**	-0.179*	-0.189*
	(0.056)	(0.074)	(0.075)	(0.070)	(0.074)	(0.088)
Treaty Dummies	No	Yes	Yes	Yes	Yes	Yes
Total Issues per Case	No	No	Yes	Yes	Yes	Yes
Issue Type Dummies	No	No	No	Yes	Yes	Yes
Restrictiveness	No	No	No	No	Yes	Yes
Polity Score	No	No	No	No	No	Yes
Observations	309	309	309	309	282	275
Panel C: World Bank Status (Logit)						
	(1)	(2)	(3)	(4)	(5)	(6)
High Income	-0.232***	-0.220**	-0.225**	-0.230**	-0.232**	-0.262***
	(0.070)	(0.072)	(0.073)	(0.071)	(0.071)	(0.072)
Upper-Middle Income	-0.013	0.007	-0.012	-0.021	0.013	-0.029

	(0.076)	(0.078)	(0.077)	(0.073)	(0.074)	(0.072)
Low Income	0.117	0.107	0.110	0.056	0.207	
	(0.163)	(0.164)	(0.165)	(0.163)	(0.122)	
Treaty Dummies	No	Yes	Yes	Yes	Yes	Yes
Total Issues per Case	No	No	Yes	Yes	Yes	Yes
Issue Type Dummies	No	No	No	Yes	Yes	Yes
Restrictiveness	No	No	No	No	Yes	Yes
Polity Score	No	No	No	No	No	Yes
Observations	309	309	309	309	282	275

Panel D: World Bank Status (Linear)

	(1)	(2)	(3)	(4)	(5)	(6)
High Income	-0.232**	-0.222**	-0.227**	-0.225**	-0.222**	-0.254**
	(0.070)	(0.074)	(0.074)	(0.072)	(0.072)	(0.076)
Upper-Middle Income	-0.013	-0.009	-0.013	-0.022	-0.010	-0.023
	(0.076)	(0.079)	(0.078)	(0.078)	(0.080)	(0.079)
Low Income	0.117	0.108	0.110	0.064	0.211	0.300**
	(0.163)	(0.166)	(0.168)	(0.171)	(0.129)	(0.114)
Treaty Dummies	No	Yes	Yes	Yes	Yes	Yes
Total Issues per Case	No	No	Yes	Yes	Yes	Yes
Issue Type Dummies	No	No	No	Yes	Yes	Yes
Restrictiveness	No	No	No	No	Yes	Yes
Polity Score	No	No	No	No	No	Yes
Observations	309	309	309	309	282	275

Note: The outcome variable is a dummy variable *Investor Win*, which equals 1 if the claimant is awarded more than \$0. Logit coefficients are reported as marginal effects. "Lower-Middle Income" is specified as the baseline category. The coefficients for "Low Income" are missing in Column (6) because there is there is no variation in the outcome variable. Standard errors, in parentheses, are clustered at the IIA level; $^{*} p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$.

To address the concern that more developed respondent states are better at filtering out stronger cases of investors at the settlement stage, I also use the Heckman probit selection model to correct for potential selection bias caused by settlement. To estimate the probability of a case reaching the award, I include in the first stage regression an exclusion restriction *Alleged Amount Public*. *Alleged Amount Public* denotes whether the amount of damages sought by investors is

made public.[55] The publicity of the amount of damages alleged makes settlement less likely,[56] but is unlikely to have an independent impact on the case outcome at the award stage.[57] As Table A1.2 shows, in the first stage, the coefficients for *Alleged Amount Public* are significantly positive across specifications, which is consistent with the expectation that settlement becomes less likely when the amount alleged is made public. However, the first stage regression results reveal no evidence that less developed respondent states are more likely to settle cases than more developed respondent states. In the second stage, the coefficients for OECD states and High Income states are significantly negative in all specifications. Their magnitudes are similar to those reported in Table 1.8. The finding of decreased likelihood to lose for more developed respondent states still holds under the Heckman probit selection model.

I conduct two more robustness checks. First, I use respondent states' logged GDP per capita as a proxy for their development status and run the same linear and logit regressions. I find similar results: in all specifications, the coefficients for respondent states' logged GDP per capita are significantly negative (Table A1.3). Second, I substitute polity score with six alternative proxies for respondent states' governance quality and run the same linear and logit regressions.[58] I find similar results: in all specifications, the coefficients for respondent states' logged GDP per capita

---

[55] This variable has been used in Pelc (2017) to estimate the probability of a case reaching the award stage in Heckman probit selection model. See Krzysztof J. Pelc, *What Explains the Low Success Rate of Investor-S-State Disputes?* 71 INTERNATIONAL ORGANIZATION 559 (2017).

[56] A number of studies show that privacy increases settlement likelihood. *See id. See also* Chad Brown, *Participation in WTO Dispute Settlement: Complainants, Interested Parties, and Free Riders*, 19 (2) WORLD BANK ECONOMIC REVIEW 287 (2005); Marc Busch and Eric Reinhardt, *Bargaining in the Shadow of the Law: Early Settlement in GATT/WTO Disputes*, 24 (1) FORDHAM INT'L L.J. 158 (2000).

[57] The publicity of the amount of damages alleged is unlikely to affect the *direction* of the ruling. The tribunal may take this factor into account when determining the amount of damages being awarded, but is unlikely to change the direction of its ruling (i.e., rule in favor of the other side) simply because the amount of damages alleged is public. After all, the existence of the dispute is often public.

[58] These six proxies are the Liberal democracy index, the judicial constraints on the executive index, the legislative constraints on the executive index, the high court independence indicator, the lower court independence indicator, and the political corruption index. All of them are imported from the Varieties of Democracy (V-Dem) Project. *See* V-Dem Project, https://www.v-dem.net/en/.

are significantly negative (Table A1.4). The extension results suggest that after taking into account different confounds, more developed states are still significantly less likely to lose in investment treaty arbitration cases than less developed states. Therefore, I find tentative evidence suggesting that investment treaty arbitration is biased against developing respondent states.

VI. CONCLUSION

In this paper, I reproduce, replicate and extend two representative empirical studies that use different strategies to examine biases against developing states in investment treaty arbitration and arrive at contradictory conclusions. Using the reproduction datasets, I find no evidence of biases against developing states. I find similar results when using alternative estimation strategies in the replication. However, when I change to the expanded dataset, the estimated coefficients for respondent states' development status become significantly negative, suggesting potential biases against developing respondent states. The results remain robust after I control for confounding factors that could be correlated with unobserved case strength, use the Heckman probit selection model to correct for potential selection bias caused by settlement, and use alternative proxies for respondent states' development status. As a statistical matter, the extension results are consistent with the replication results. The extension estimates are all within the 95% confidence intervals of the replication results. Nevertheless, the extension tests have stronger statistical power due to the use of a larger dataset. They allow me to infer with better precision and confidence that the effect of respondent states' development status on case outcomes is not zero. The results affirm that the authors of the two original studies are correct about the necessity of replication.

Perceived biases against developing states have generated widely-held criticisms and backlashes against investment treaty arbitration in the developing world. Finding credible empirical evidence of such biases is important for informing policy debates and potential reforms

of the investment treaty arbitration system. This paper contributes to this mission by testing the reproducibility and replicability of representative empirical studies that examine the issue of biases against developing states. It also provides a significant step forward by proposing new methods to better identify any causal effect that respondent states' development status has on case outcomes. The paper applies the new methods to a larger and more up-to-date dataset. The new results reveal consistent negative correlations between respondent states' development status and their likelihood of losing across specifications. The patterns indicate that there may exist biases against developing states in investment treaty arbitration, an issue that deserves closer examination in future research.

Table A1.1: Additional Summary Statistics of The Expanded Dataset

Respondent State's OECD Status	
OECD	211
Non-OECD	528
Total	739
Respondent State's World Bank Status	
High Income	210
Upper-Middle Income	318
Lower-Middle Income	189
Low Income	22
Total	739
Treaty Type	
BITs	580
Energy Charter Treaty	95
NAFTA	58
Other Treaties	28
Total	761[59]
Issue Type	
Direct Expropriation	75
Fair and Equitable Treatment	363
Full Protection and Security	194
Indirect Expropriation	321
Most Favored Nations	84
National Treatment	104
Performance Requirement	12
Total	1153[60]
Count of Issues Per Case	
Min	0
Max	5
Mean	2
Median	1
Respondent State's Polity Score	
Min	-10
Max	10
Mean	5
Median	8

---

[59] In some cases, more than one treaty was invoked.
[60] In some cases, more than one issue was raised.

Table A1.2: Respondent's Development Status and Case Outcome and Case Outcome (Heckman Probit Selection Model)

Panel A: OECD Status

	(1)	(2)	(3)	(4)	(5)	(6)
Respondent_OECD	-0.205**	-0.190*	-0.199*	-0.182*	-0.171*	-0.164*
	(0.063)	(0.191)	(0.077)	(0.077)	(0.082)	(0.099)
Treaty Dummies	No	Yes	Yes	Yes	Yes	Yes
Total Issues per Case	No	No	Yes	Yes	Yes	Yes
Issue Type Dummies	No	No	No	Yes	Yes	Yes
Restrictiveness	No	No	No	No	Yes	Yes
Polity Score	No	No	No	No	No	Yes

Award Rendered

	(1)	(2)	(3)	(4)	(5)	(6)
Alleged Amount Public	0.431***	0.423***	0.290***	0.256***	0.283***	0.246***
	(0.043)	(0.044)	(0.049)	(0.050)	(0.051)	(0.059)
Respondent_OECD	0.009	0.031	0.013	0.034	0.031	0.084
	(0.417)	(0.051)	(0.054)	(0.057)	(0.059)	(0.071)
Treaty Dummies	No	Yes	Yes	Yes	Yes	Yes
Total Issues per Case	No	No	Yes	Yes	Yes	Yes
Issue Type Dummies	No	No	No	Yes	Yes	Yes
Restrictiveness	No	No	No	No	Yes	Yes
Polity Score	No	No	No	No	No	Yes
Observations	739	739	739	739	686	572

Panel B: World Bank Status

	(1)	(2)	(3)	(4)	(5)	(6)
High Income	-0.226**	-0.209**	-0.219**	-0.245**	-0.225**	-0.260**
	(0.073)	(0.078)	(0.080)	(0.081)	(0.079)	(0.090)
Upper-Middle Income	-0.017	-0.010	-0.021	-0.043	-0.024	-0.039
	(0.071)	(0.074)	(0.077)	(0.083)	(0.082)	(0.088)
Low Income	0.158	0.155	0.163	0.106	0.272	0.369
	(0.154)	(0.158)	(0.159)	(0.171)	(0.187)	(0.176)
Treaty Dummies	No	Yes	Yes	Yes	Yes	Yes
Total Issues per Case	No	No	Yes	Yes	Yes	Yes
Issue Type Dummies	No	No	No	Yes	Yes	Yes
Restrictiveness	No	No	No	No	Yes	Yes
Polity Score	No	No	No	No	No	Yes

Award Rendered

	(1)	(2)	(3)	(4)	(5)	(6)
Alleged Amount Public	0.444***	0.435***	0.307***	0.269***	0.303***	0.274***
	(0.044)	(0.045)	(0.049)	(0.051)	(0.052)	(0.059)
High Income	0.008	0.017	-0.015	-0.016	0.002	0.052
	(0.051)	(0.055)	(0.059)	(0.062)	(0.064)	(0.077)

Table A1.2 (Continued)						
Upper-Middle Income	0.036	0.040	-0.034	-0.054	-0.028	-0.004
	(0.047)	(0.047)	(0.050)	(0.052)	(0.053)	(0.060)
Low Income	0.311**	0.311**	0.255*	0.272*	0.292*	0.236
	(0.111)	(0.112)	(0.118)	(0.122)	(0.128)	(0.133)
Treaty Dummies	No	Yes	Yes	Yes	Yes	Yes
Total Issues per Case	No	No	Yes	Yes	Yes	Yes
Issue Type Dummies	No	No	No	Yes	Yes	Yes
Restrictiveness	No	No	No	No	Yes	Yes
Polity Score	No	No	No	No	No	Yes
Observations	739	739	739	739	686	572

Notes: Probit coefficients are reported as marginal effects. The first stage estimates likelihood of an award being rendered. The second stage estimates likelihood of investor wins. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A1.3: Respondent's Logged GDP Per Capita in 2005 USD and Case Outcome, Expanded Dataset

Panel A: Logit

	(1)	(2)	(3)	(4)	(5)	(6)
Logged GDP Per Capita	-0.094***	-0.092***	-0.096***	-0.092***	-0.098***	-0.119***
	(0.022)	(0.026)	(0.026)	(0.025)	(0.025)	(0.025)
Treaty Dummies	No	Yes	Yes	Yes	Yes	Yes
Total Issues per Case	No	No	Yes	Yes	Yes	Yes
Issue Type Dummies	No	No	No	Yes	Yes	Yes
Restrictiveness	No	No	No	No	Yes	Yes
Polity Score	No	No	No	No	No	Yes
Observations	299	299	299	299	277	274

Panel B: Linear

	(1)	(2)	(3)	(4)	(5)	(6)
Logged GDP Per Capita	-0.095***	-0.095***	-0.098***	-0.094***	-0.098***	-0.113***
	(0.021)	(0.026)	(0.026)	(0.025)	(0.024)	(0.023)
Treaty Dummies	No	Yes	Yes	Yes	Yes	Yes
Total Issues per Case	No	No	Yes	Yes	Yes	Yes
Issue Type Dummies	No	No	No	Yes	Yes	Yes
Restrictiveness	No	No	No	No	Yes	Yes
Polity Score	No	No	No	No	No	Yes
Observations	299	299	299	299	277	274

Note: The outcome variable is a dummy variable *Investor Win*, which equals 1 if the claimant is awarded more than \$0. Logit coefficients are reported as marginal effects. Standard errors, in parentheses, are clustered at the IIA level; $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table A1.4: Respondent's Logged GDP Per Capita in 2005 USD and Case Outcome (Alternative Proxies), Expanded Dataset

Panel A: Logit

	(1)	(2)	(3)	(4)	(5)	(6)
Logged GDP Per Capita	-0.103***	-0.085***	-0.091***	-0.087**	-0.072**	-0.097**
	(0.028)	(0.026)	(0.024)	(0.025)	(0.027)	(0.029)
Treaty Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Total Issues per Case	Yes	Yes	Yes	Yes	Yes	Yes
Issue Type Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Restrictiveness	Yes	Yes	Yes	Yes	Yes	Yes
Liberal Democracy	Yes	No	No	No	No	Yes
Judicial Constraint	No	Yes	No	No	No	No
Legislative Constraint	No	No	Yes	No	No	No
High Court Independence	No	No	No	Yes	No	No
Low Court Independence	No	No	No	No	Yes	No
Political Corruption	No	No	No	No	No	Yes
Observations	273	273	273	273	273	273

Panel B: Linear

	(1)	(2)	(3)	(4)	(5)	(6)
Logged GDP Per Capita	-0.104***	-0.086***	-0.092***	-0.088***	-0.074**	-0.101**
	(0.028)	(0.026)	(0.024)	(0.025)	(0.028)	(0.030)
Treaty Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Total Issues per Case	Yes	Yes	Yes	Yes	Yes	Yes
Issue Type Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Restrictiveness	Yes	Yes	Yes	Yes	Yes	Yes
Liberal Democracy	Yes	No	No	No	No	No
Judicial Constraint	No	Yes	No	No	No	No
Legislative Constraint	No	No	Yes	No	No	No
High Court Independence	No	No	No	Yes	No	No
Low Court Independence	No	No	No	No	Yes	No
Political Corruption	No	No	No	No	No	Yes
Observations	273	273	273	273	273	273

Note: The outcome variable is a dummy variable *Investor Win*, which equals 1 if the claimant is awarded more than \$0. Logit coefficients are reported as marginal effects. Standard errors, in parentheses, are clustered at the IIA level; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

CHAPTER TWO

ARE ARBITRATORS BIASED IN ICSID ARBITRATION? A DYNAMIC PERSPECTIVE

I. INTRODUCTION

The rise of the International Centre for Settlement of Investment Disputes (ICSID) as a major forum to resolve investment-related disputes has generated scrutiny over its neutrality and legitimacy.[1] Arbitrators, as *ad hoc* appointees, may be incentivized to vote in biased ways that enhance career and financial prospects.[2] For example, since investor-state arbitration may be initiated only by investors, arbitrators may exhibit bias against host states to enhance the attractiveness of investor-state arbitration to investors and increase reappointments as the total number of cases grows.[3] Arbitrators' independence and impartiality could also be compromised due to multiple appointments by the same parties or counsel.[4] In particular, party-appointed arbitrators may choose to favor their appointers out of gratitude.[5]

---

[1] The major concerns raised so far include the incoherence and inconsistence of the substantive jurisprudence that ICSID tribunals have generated, the integrity of decision makers, and the curtailing of state public interests. *See e.g.* Sandra L. Caruba, *Resolving International Investment Disputes in a Globalized World*, 13 N.Z. BUS. L.Q. 128, 150; George K. Foster, *Collecting from Sovereigns: The Current Legal Framework for Arbitral Awards and Court Judgments Against States and Their Instrumentalities, and Some Proposals for its Reform*, 25 ARIZ. J. INT'L & COMP. L. 665, 705 (2008); William W. Park, *Arbitrator Integrity: The Transitory and the Permanent*, 46 SAN DIEGO L. REV. 629, 658 (2009); Gus Van Harten et al., *Public Statement on the International Investment Regime* (Aug. 31, 2010), *available at* http://www.osgoode.yorku.ca/public-statement-international-investment-regime-31-august-2010/; Susan D. Franck, *The Legitimacy Crisis in Investment Treaty Arbitration: Privatizing Public International Law Through Inconsistent Decisions*, 73 FORDHAM L. REV. 15, 21 (2005).

[2] *See e.g.* Pia Eberhardt, Cecilia Olivet, Tyler Amos & Nick Buxton, *Profiting from Injustice: How Law Firms, Arbitrators and Financiers Are Fueling An Investment Arbitration Boom*, CORPORATE EUROPE OBSERVATORY AND THE TRANSNATIONAL INSTITUTE (November 2012), at 35, 67, https://www.tni.org/files/download/profitingfrominjustice.pdf; M. Sornarajah, *Power and Justice in Foreign Investment Arbitration*, 14 J. INT'L ARB. 103, 118 (1997).

[3] *See* EU Commission, *Concept Paper: Investment in TTIP and Beyond – The Path for Reform*, 6–7, *available at* http://trade.ec.europa.eu/doclib/docs/2015/may/tradoc_153408.PDF; GUS VAN HARTEN, INVESTMENT TREATY ARBITRATION AND PUBLIC LAW (2007) 152-153; Filip De Ly et al, *Who Wins and Who Loses in Investment Arbitration? Are Investors and Host States on a Level Playing Field?* 6 J WORLD INVESTMENT & TRADE 59, 69 (2005).

[4] *See* Vanessa Giraud, *The Problem of Repeat Arbitrators in Investment Arbitration*, Kluwer Arbitration Blog (September 1, 2014), http://kluwerarbitrationblog.com/2014/09/01/the-problem-of-repeat-arbitrators-in-investment-arbitration/; Fatima-Zahra Slaoui, *The Rising Issue of "Repeat Arbitrators": A Call for Clarification*, 25 ARB. INT'L 103 (2009); Jan Paulsson, *Ethics, Elitism, Eligibility* 14(4) J INT'L ARB 14 (1997).

[5] *See e.g.* Yves Derains, *Fifth Annual Lecture on International Commercial Arbitration: The Arbitrator's Deliberation*, 27:4 AM. U. INT'L L. REV. 911, 915; David Branson, *Sympathetic Party-Appointed Arbitrators: Sophisticated*

In part due to such concerns, voices favoring investor-state dispute settlement reform have grown more pronounced. On October 25, 2017, in a letter to the White House, over 200 professors of economics or law called on President Trump to strip the investor-state arbitration system from NAFTA and other trade agreements.[6] Recent EU trade agreements have included provisions to establish an investment court system consisting of first instance and appellate tribunals, members of which will be appointed exclusively by states and paid a monthly retainer fee.[7] International investment court proponents have argued that tenured judges, free from incentives related to possible reappointments, would be more impartial and independent compared to arbitrators appointed *ad hoc*.[8]

Strangers and Governments Demand Them, 25 ICSID REV. FOREIGN INVESTMENT LAW JOURNAL. 367, 368 (2010); Jan Paulsson, *Moral Hazard in International Dispute Resolution*, 25 ICSID REV. FOREIGN INVESTMENT LAW JOURNAL 339 (2010); Hans Smit, *The Pernicious Institution of the Party-Appointed Arbitrators*, 33 COLUM. FDI PERSP. (2010), *available at* https://academiccommons.columbia.edu/catalog/ac:134503; Carlos G. Garcia, *All the Other Dirty Little Secrets: Investment Treaties, Latin America, and the Necessary Evil of Investor-State Arbitration*, 16 FLA. J. INT'L L. 301, 353 (2004).

[6] *See* Alex Lawson, *Academics Line up to Sink NAFTA's Investment Arbitration*, LAW360 (October 26, 2017), https://www.law360.com/internationalarbitration/articles/978639/academics-line- up-to- sink-nafta-s-investment-arbitration; The complete letter is available at https://www.citizen.org/system/files/ case_documents/isds-law-economics-professors-letter-oct-2017_2.pdf.

[7] According to CETA (the Comprehensive Economic and Trade Agreement) between Canada and the European Union, the most recent agreed drafts of TTIP (Trans-Atlantic Trade and Investment Partnership) between the United States and the European Union and the European Union-Vietnam Free Trade Agreement, the first instance tribunal is to be composed of fifteen judges: five nationals of EU Member States, five nationals of the other Contracting Party and five nationals from third countries, who are appointed for a six-year term, renewable once. The tribunal will hear cases in divisions consisting of three judges: one national from each Contracting Party, chaired by a third country national. The appellate tribunal is to be composed of six members: two nationals of EU Member States, two nationals of the other Contracting Party and two nationals of third countries. The appeal tribunal members are also appointed for a six-year term, renewable once. *See* arts. 8.27, 8.28 of CETA, *available at* http://ec.europa.eu/trade/policy/in-focus/ceta/ceta-chapter-by-chapter; arts. 9, 10 of Draft TTIP Investment Chapter (November 12, 2015), *available at* http://trade.ec.europa.eu/doclib/docs/2015/ november/tradoc_153955.pdf; arts. 12, 13 of EU-Vietnam FTA Draft Investment Chapter, *available at* http://trade.ec.europa.eu/doclib/docs/2016/february/ tradoc_154210.pdf.

[8] *See* Gus Van Harten, *Perceived Bias in Investment Treaty Arbitration*, in CLAIRE BALCHIN, LIZ KYO-HWA CHUNG, ASHA KAUSHAL & MICHAEL WAIBEL EDS., THE BACKLASH AGAINST INVESTMENT ARBITRATION: PERCEPTIONS AND REALITY 433, 445 – 446 (2010); Gus van Harten, *A Case for an International Investment Court*, at 30 (2008), *available at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1153424; VAN HARTEN, *supra* note 3, at 175; M. Sornarajah, *The Neo-Liberal Agenda in Investment Arbitration*, in WENHUA SHAN ET AL EDS., REDEFINING SOVEREIGNTY IN INTERNATIONAL ECONOMIC LAW 218 (2008); UNCTAD, *Reform of Investor-State Dispute Settlement: In Search of a Roadmap, International Investment Agreement Issues Note No. 2* (June 2013), at 9, *available at* http://unctad.org/en/PublicationsLibrary/webdiaepcb2013d4_en.pdf; Gabrielle Kaufmann-Kohler and Michele Potesta, *Can the Mauritius Convention Serve as A Model for the Reform of Investor-State Arbitration in Connection with the Introduction of a Permanent Investment Tribunal or An Appeal Mechanism? Analysis and Roadmap*, at 18, *available at* http://www.uncitral.org/pdf/english/CIDS_Research_Paper_Mauritius.pdf.

In this paper, I empirically examine ICSID arbitrator bias due to two factors: reappointment interests and favoring of frequent appointers. I focus on these two factors not only because they are primary concerns behind proposals to replace arbitration with international investment courts, but also because they are closely tied to a defining characteristic of international arbitration: the appointment of adjudicators for individual cases by the disputing parties.[9] International investment court proponents attempt to improve adjudicator impartiality and independence by completely removing *ad hoc* appointment, which has long been considered a main attraction of international arbitration vis-à-vis national courts or other forms of dispute resolution.[10] I empirically examine the extent to which bias arising from *ad hoc* appointment exists and discuss the implications of such bias, if any, on the institutional design of investor-state dispute settlement.

In this paper, I distinguish three sources of bias: pre-existing bias, prospective bias and retrospective bias. Judges and arbitrators may both exhibit predispositions formed through experiences prior to their appointments by way of academic standing, scholarly publication record, practical experiences, business or political connections, etc.[11] I use the term "pre-existing bias" to refer to this type of bias. For the purpose of this study, pre-existing bias is considered time-invariant. By contrast, certain biases are specific to arbitrators due to the *ad hoc* nature of their appointments. For instance, investors might be more likely to appoint arbitrators with pro-investor reputations, and arbitrators hoping to capitalize on this might thereby strategically vote in an attempt to shape a pro-investor reputation.[12] I use the term "prospective bias" to refer to this type

---

[9] *See* Michael Waibel, *Arbitrator Selection: Towards Greater State Control*, in ANDREAS KULICK ED., REASSERTION OF CONTROL OVER THE INVESTMENT TREATY REGIME 333 (2017).

[10] *See id.*, at 334.

[11] Prior literature has also referred to these characteristics as an arbitrator's symbolic capital. *See* YVES DEZALAY & BRYANT G. GARTH, DEALING IN VIRTUE 19 (1996); JOSHUA KARTON, THE CULTURE OF INTERNATIONAL ARBITRATION AND THE EVOLUTION OF CONTRACT LAW 27 (2013).

[12] Serving as an arbitrator is a lucrative and prestigious job. According to ICSID's Memorandum on the Fee and Expenses, arbitrators on arbitral tribunals and ad hoc committees receive $3,000 per meeting day or 8-hour day of other work. *See* International Center for Settlement of Investment Disputes, *Memorandum on the Fee and Expenses*

of bias. To measure "reputation", I take each arbitrator's investor-state vote ratio from previous decisions and compare to the arbitrator population average (i.e. the average investor-state win ratio) at a point in time.[13] I assume that the average investor-state win ratio in the dataset at a point in time is a reference point for arbitrators and appointing entities when assessing the degree to which an arbitrator's reputation is pro-investor, pro-state or neutral.[14] Considering another angle, arbitrators can also exhibit bias towards whoever has appointed them more often in the past, due to reasons such as loyalty or reciprocity. I use the term "retrospective bias" to refer to this type of bias. Prospective and retrospective biases are specific to arbitration (in so far as judges are not selected by parties) and thus particularly important to considerations of reforming investor-state arbitration to more resemble a judiciary.

In this study, I use a hand-coded dataset of arbitrator votes and appointments from all publicly available ICSID arbitration cases registered before January 1, 2016.[15] I test the prospective bias prediction by examining the relationship between an arbitrator's reputation at case decision time and their voting behavior. I test the retrospective bias prediction by examining the relationship between an arbitrator's appointment rate by an entity at case decision time and their voting behavior. I find no evidence in my study for the existence of either bias. Therefore, the data

---

(July 2005), *available at* https://icsid.worldbank.org/en/Pages/icsiddocs/Memorandum-on-the-Fees-and-Expenses-FullText1.aspx.

[13] While there are other alternative proxies for an arbitrator's reputation, such as their publications and presentations at conferences, an arbitrator's votes in previous decisions are the most easily obtained information and are unlikely to deviate much from the information the arbitrator has produced through other means. Cohen, Klement, and Neeman (2015) has adopted a similar approach by comparing the average sentences that a judge imposed the preceding year to the average sentences imposed in the previous year in the judge's district. *See* Alma Cohen, Alon Klement, and Zvika Neeman, *Judicial Decision Making, A Dynamic Reputation Approach*, 44 J. LEGAL STUD. 133, 147.

[14] Figures A2.1 shows changes in the average state win ratio after 2000 as more cases have been decided over time. Cases decided before 2000 are not shown in the graph because there are very few of them which causes large fluctuation in average state win ratio. As Figure A2.1 shows, after 2000, the average state win ratio remains relatively stable.

[15] ICSID arbitration cases constitute a majority but not all investment arbitration cases. I only analyzed ICSID arbitration cases because the award or appointment information in most investment arbitration cases administered by other institutions are not publicly available.

does not suggest that replacing the current arbitration system with an investment court system would achieve greater adjudicator independence and impartiality due to removing biases associated with *ad hoc* appointment.

Distinguished from previous studies that have focused on static (i.e. time-invariant) arbitrator bias,[16] this paper considers a dynamic approach towards arbitrator behavior. Examining how arbitrators behave in relation to different incentives in ICSID arbitration helps us to evaluate concerns about arbitrator bias associated with *ad hoc* appointment and related reform proposals. This paper also adds to earlier literature which considered incentives from investor or state appointments only, by also considering incentives arising from other entities with authority to make appointments. Methodologically, this paper develops new empirical strategies to measure arbitrator reputation on a time-variant basis. The paper also introduces a placebo test that reveals (and overcomes) the difficulty of measuring dynamic voting behavior in fixed effect models. Note that any causal relationship identified using this method does not rely on exogenous variation in the data. Instead, it relies on structural identification, by comparing arbitrator voting behavior in the real data with that in a placebo data which is manipulated in a way that makes it impossible for any bias to exist. Normatively, this paper proposes that dynamic biases deserve greater attention when evaluating the institutional design of investor-state dispute settlement. It provides empirical evidence alleviating certain concerns of arbitrator bias and calls into question the superiority of courts over arbitration in improving the independence and impartiality of investor-state adjudication.

---

[16] *See e.g.* Daphna Kapeliuk, *The Repeat Appointment Factor: Exploring Decision Patterns of Elite Investment Arbitrators*, 96 CORNELL L. REV. 47 (2012) [hereinafter *The Repeat Appointment Factor*]; Daphna Kapeliuk, *Collegial Games: Analyzing the Effect of Panel Composition on Outcome in Investment Arbitration*, 31:2 THE REVIEW OF LITIGATION, 267 (2012) [hereinafter *Collegial Games*]; Susan Franck, *Development Status and Outcomes of Investor Treaty Arbitration*, 50:2 HARV. INT'L L. J., 435 (2009).

The remainder of this paper proceeds as follows: Part II provides a brief overview of ICSID arbitration, literature related to arbitrator bias and investor-state dispute settlement reform proposals, and the research predictions in this paper; Part III describes the data; Part IV introduces the empirical results; and Part V provides concluding remarks.

II. BACKGROUND

A. ICSID Arbitration

ICSID was established in October 1966 under the ICSID Convention, the drafting and negotiation of which was initiated by the World Bank, "to provide facilities for conciliation and arbitration of investment disputes between Contracting States and nationals of other Contracting States".[17] There are currently 153 Contracting States (161 signatories) of the ICSID Convention.[18] For investors from or in non-contracting states, ICSID provides the possibility of requesting arbitration or conciliation under the Additional Facility framework.[19] ICSID also provides annulment proceedings, where parties can request an *ad hoc* committee to set aside an award if certain grounds are met[20] and resubmission proceedings, where after an award is annulled

---

[17] ICSID Convention, art. 1(2).
[18] *See* ICSID List of Contracting States and Other Signatories of the Convention, https://icsid.worldbank.org/en/Pages/about/Database-of-Member-States.aspx.
[19] The procedural rules for arbitrations under the additional facility are similar to those of the ICSID in large part. One major difference is that in addition to the provisions of the Additional Facility Rules, the laws of the place of arbitration apply to the proceeding. If a provision of the Additional Facility Rules conflicts with a mandatory provision of the applicable law, that law prevails. *See* Rules Governing the Additional Facility for the Administration of Proceedings by the Secretariat of the International Centre for Settlement of Investment Disputes (Additional Facility Rules) art. 1, 2, 20, *available at* http://icsidfiles.worldbank.org/ICSID/ICSID/StaticFiles/facility/partA-article.htm#a02.
[20] According to Article 52(1) of the ICSID Convention, there are five grounds where an award of a tribunal may be annulled: (a) that the Tribunal was not properly constituted; (b) that the Tribunal has manifestly exceeded its powers; (c) that there was corruption on the part of a member of the Tribunal; (d) that there has been a serious departure from a fundamental rule of procedure; or (e) that the award has failed to state the reasons on which it is based.

investors can resubmit their claims and have a new tribunal rehear the case.[21] As of 2015, 525

cases have been registered.[22]

Dispute proceedings usually arise when foreign investors file a request for arbitration with

the Secretary-General against host states for alleged violation of obligations under investment

treaties or national investment laws. Once the Secretary-General registers the case, the parties have

90 days to constitute the tribunal.[23] While tribunals can consist of a sole arbitrator or any odd

number of arbitrators,[24] most are composed of three: one appointed by each party and a third

presiding arbitrator appointed by mutual agreement between the parties or between the two party-

appointed arbitrators.[25] If the tribunal is not constituted within the required period, the Chairman

of ICSID's Administrative Council[26] ("ICSID") will appoint any remaining arbitrators from a

Panel of Arbitrators, the members of which are designated by Contracting States and ICSID.[27]

Thus, five entities can potentially have authority to appoint arbitrators: investors, states, both

parties in mutual agreement, co-arbitrators in mutual agreement, and ICSID.[28] The investor and

state when in mutual agreement (the "parties") are considered as a separate appointing entity to

---

[21] *See* ICSID Convention, art. 52.6.

[22] 516 of these cases are arbitration cases and 9 of them are conciliation cases. *See* International Center for Settlement of Investment Disputes, *ICSID 2015 Annual Report* 5, *available at* https://icsid.worldbank.org/en/Documents/resources/ICSID_AR15_ENG_CRA-highres.pdf.

[23] Unless the Secretary-General finds the claim to be manifestly outside the jurisdiction of ICSID, the Secretary-General will register the case. Parties may agree to any other period. *See* ICSID Convention, art. 36(3), 38.

[24] *See* ICSID Convention, art. 37(2)(a).

[25] *See* ICSID Arbitration Rules, Rule 3.

[26] The Chairmanship of the Administrative Council is filled by the President of the World Bank pursuant to Article 5 of the ICSID Convention. In practice, the mandate of appointing arbitrators is fulfilled by the Secretary-General of ICSID. For the sake of succinctness, in following paragraphs, they will be referred to as ICSID.

[27] *See* ICSID Arbitration Rules, Rule 4. ICSID Convention, Art. 38, 40. Under Article 13 of the ICSID Convention, each Contracting State may designate four individuals, who need not be its nationals, to the Panel of Arbitrators. In addition, the Chairman of the Administrative Council may designate ten individuals to the Panel of Arbitrators, each of a different nationality. Under Article 15 of the ICSID Convention, panel members shall serve for renewable periods of six years. As of March 24, 2017, there are 285 designees in total on the Panel of Arbitrators. *See* Database of ICSID Panels, https://icsid.worldbank.org/en/Pages/about/Database-of-Panel-Members.aspx.

[28] ICSID retains the authority to appoint arbitrators when the tribunal is not successfully constituted, when there is a vacancy caused by the resignation, disqualification, death, or incapacity of an arbitrator during the arbitral proceeding, and during the annulment proceeding where parties could request an ad hoc committee to set aside an award if certain grounds are met.

account for the fact that mutual agreement will exhibit very different preferences compared to independent selection of an arbitrator.

B. Literature Review

Various concerns regarding arbitrator bias have been raised by the literature over the past decade. A commonly cited rationale is bias arising from the *ad hoc* nature of arbitrator appointment. Van Harten (2007) points out the possibility of systematic bias in favor of claimant-investors among arbitrators seeking to promote growth of investor-state proceedings and thereby encourage reappointments.[29] Paulson (1997) raises the concern that arbitrators tend to favor parties that have already appointed them previously.[30] Branson (2010) suggests that party-appointed arbitrators are likely to tilt the arbitral process and deliberations in favor of the appointing party.[31] Dammann and Hansmann (2008) believe that arbitrators could be incentivized to render compromise awards that increase both parties' satisfaction in the arbitration process.[32]

In response to such concerns, practitioners have raised various arguments in defense of investor-state arbitration. Park (2009) offers that desire for reappointment may actually incentivize arbitrators to be more accurate and neutral instead, since both host state and investor have a role in the appointment process.[33] Brower and Rosenberg (2013) also note that arbitrators achieve "success" by retaining a reputation for honesty, independence and impartiality, as a reputation for bias will undercut their credibility (and hence influence) within a tribunal.[34]

---

[29] *See* VAN HARTEN, *supra* note 3.
[30] *See* Paulsson, *supra* note 4.
[31] *See* Branson, *supra* note 5.
[32] *See* Jens Dammann & Henry Hansmann, *Globalizing Commercial Litigation*, 94 CORNELL L. REV. 1, 34 (2008).
[33] *See* Park, *supra* note 1.
[34] Charles N. Brower & Charles B. Rosenberg, *Death of the Two-Headed Nightingale: Why the Paulsson—Van den Berg Presumption that Party-Appointed Arbitrators are Untrustworthy is Wrongheaded*, 29 ARB. INT'L, 7, 15 – 16 (2013).

A small number of empirical studies have examined arbitrator bias in investor-state arbitration from various perspectives. Kapeliuk (2012a) finds that tribunals involving repeatedly appointed arbitrators do not exhibit a tendency to render compromise awards, nor do they exhibit a tendency to rule in favor of investors.[35] Yet among repeatedly appointed arbitrators, Kapeliuk (2012b) finds that presiding arbitrators are less averse to extreme outcomes than party-appointed arbitrators.[36] Kapeliuk (2012b) also divides party-appointed arbitrators into those with arbitration experience and newcomers, then examines the impact of experience on case outcomes and issuance of dissenting opinions, but finds no significant correlations.[37] Van Harten (2012) finds that frequently appointed arbitrators are more likely to resolve jurisdictional issues expansively in terms of interpretation, benefitting investors.[38] Van Harten (2012, 2016) also finds that arbitrators are more likely to adopt an expansive approach to legal interpretation of both jurisdictional and substantive issues if the investor is a national of a major western capital exporting country.[39] Franck (2009) finds that presiding arbitrators from the developing world make smaller awards against developed states in certain circumstances, but finds no significant relationship between the development status of the respondent state and amount awarded.[40] Franck (2015) also finds no significant relationship between case outcomes and tribunal gender composition, tribunal development status composition, or repeat player status of the president.[41] Waibel and Wu (2017)

---

[35] *See* Kapeliuk, *The Repeat Appointment Factor*, *supra* note 16.

[36] *See* Kapeliuk, *Collegial Games*, *supra* note 16.

[37] *See* Kapeliuk, *Collegial Games*, *supra* note 16.

[38] *See* Gus Van Harten, *Arbitrator Behavior in Asymmetrical Adjudication: An Empirical Study of Investment Treaty Arbitration*, 50 OSGOODE HALL LAW JOURNAL, 211, 248 (2012).

[39] *See id.* at 240 – 245; Gus Van Harten, *Arbitrator Behavior in Asymmetrical Adjudication (Part Two): An Empirical Study of Investment Treaty Arbitration*, 53 OSGOODE HALL LAW JOURNAL, 540, 557–559 (2016).

[40] *See* Franck, *supra* note 16.

[41] *See* Susan D. Franck & Lindsey E. Wylie, *Predicting Outcomes in Investment Treaty Arbitration*, 65 DUKE L. J., 459 (2015).

find that arbitrator decisions appear influenced by policy preferences, among which home country development status matters greatly.[42]

While evidence of arbitrator bias has thus far been mixed, practitioners and scholars have nevertheless put forward a number of reform proposals. Proposed reforms include establishment of an investment court of tenured judges (Van Harten (2007)),[43] appointment of all three tribunal members by arbitral institution (Smit (2010), Paulsson (2010)),[44] confidentiality of appointment sources towards nominees (Puig (2016)),[45] and amendment of arbitrator challenge rules to strengthen selection procedure and enlarge arbitrator pools (Giorgetti (2014)).[46] These proposals share a common thread of amending aspects of the *ad hoc* nature of arbitrator appointment. From among this collection of proposals, the idea of an investment court has attracted significant attention and has already been implemented in recent free trade agreements.

C. Predictions

1. Prospective Bias

I first consider whether arbitrators vote in biased ways to increase reappointments by shaping reputations over time ("prospective bias"). I assume that reputations are likely to be shaped in one of three directions if arbitrators are actually shaping their reputations to encourage reappointment: a pro-investor, pro-state or neutral direction. Arbitrators may try to shape a pro-investor reputation as has been suggested,[47] or they may try to shape a pro-state reputation to attract

---

[42] *See* Michael Waibel & Yanhui Wu, *Are Arbitrators Political? Evidence from International Investment Arbitration*, *available at* http://www-bcf.usc.edu/~yanhuiwu/arbitrator.pdf.
[43] *See* VAN HARTEN, *supra* note 3.
[44] *See* Smit, *supra* note 5; Paulsson, *supra* note 5, at 352.
[45] *See* Sergio Puig, *Blinding International Justice*, 56 VA. J. INT'L L. 647 (2016).
[46] *See* Chiara Giorgetti, *Who Decides Who Decides in International Investment Arbitration?* 35 U. PA. J. INT'L L. 431, 474–485 (2014).
[47] *See* VAN HARTEN, *supra* note 3.

more state appointments. If either case is true, we should observe arbitrator vote casting become increasingly polarized over time. Alternatively, arbitrators may try to shape a neutral reputation to encourage presiding arbitrator appointments if the appointing entities are expected to value neutrality.[48] The literature has not accounted for the scenario whereby arbitrators may purposefully decide certain cases against their merits for the sake of maintaining a neutral reputation over the arbitrator's entire body of decisions. If that is the case, we should observe that arbitrators avoid voting for either side too frequently in order to not appear overly biased towards one side. Hence, if arbitrators exhibit prospective bias, dependent on the target of the reputation shaping, existing reputations should be predictive of subsequent votes. This yields the following alternative predictions:

Prediction 1(a): arbitrators with stronger pro-investor or pro-state reputation at case decision time are more likely to vote in favor of investors or states, respectively.

Prediction 1(b): arbitrators with stronger pro-investor or pro-state reputation at case decision time are less likely to vote in favor of investors or states, respectively.

---

[48] Presiding arbitrators are attractive appointments as they play a more dominant and prestigious role in decision-making. According to rule 14 of the ICSID Arbitration Rules, the president of the tribunal is in charge of conducting hearings and presiding at deliberations, and fixing the date and hour of sittings. A number of scholars and practitioners have pointed out that presiding arbitrators take the lead in drafting opinions, influence the style of an international arbitration, and make critical procedural decisions. *See* Urs Martin Laeuchli, *Civil and Common Law: Contrast and Synthesis in International Arbitration*, 62 DISP. RESOL. J. 81, 82 (2007); Michael Black, Wendy Kennedy Venoit & George J. Pierson, *Arbitration of Cross Border Disputes*, 27 CONSTRUCTION LAWYER 5, 17 (2007) (observing how the chair of a tribunal permitted the introduction of documents that were arguably privileged); Winston Stromberg, *Avoiding the Full Court Press: International Commercial Arbitration and Other Global Alternative Dispute Resolution Processes*, 40 LOY. L.A. L. REV. 1337, 1367 (2007) (observing how tribunals may rely on "testimonial summaries prepared by the presiding arbitrator and presented to the witness for approval and signature"). Lawrence W. Newman, *A Practical Assessment of Arbitral Dispute Resolution*, *in* THOMAS E. CARBONNEAU ED. LEX MERCATORIA AND ARBITRATION: A DISCUSSION OF THE NEW LAW MERCHANT 5–6 (1998) (suggesting that when parties appoint arbitrators who "blatantly will favor one side" this can polarize the tribunal and "leave the chair to decide"); DEZALAY & GARTH, *supra* note 11, AT 8–9 (1996) (discussing how the selection of the chair is a "key decision in winning or losing"); Hans Smit, *Comments on Public Policy in International Arbitration*, 13 AM. REV. INT'L ARB. 65, 67 (2002) (suggesting that "if no majority decision seems possible, the chair's vote is decisive").

In making these predictions, I expect appointing entities to have certain preferences. Investors or states often appoint arbitrators due to perceived predispositions to the appointing party, [49] which are largely obtained from arbitrators' prior decisions. [50] ICSID is likely to select arbitrators known for independence and impartiality to support its institutional reputation and compete for investment cases against other dispute settlement institutions. [51] This goal is specifically set out in Article 14 of the ICSID Convention, which requires arbitrators to be persons "who may be relied upon to exercise independent judgment."[52] Parties and co-arbitrators are also more likely to agree on presiding arbitrators with reputations for neutrality than those with reputations favoring one side. Hence, I also investigate a corollary:

Corollary 1: Arbitrators with a stronger reputation at case registration time for investor, state, or neutrality are more likely to be appointed, respectively, by investor, by state, or by ICSID/parties/co-arbitrators/as president.

I also investigate whether there is a dynamic relationship between reputation and appointment likelihood. Arbitrator reputation reflects both pre-existing bias and prospective bias. It seems probable that arbitrators will only shape their reputations (exhibit prospective bias) if they perceive benefit due to this behavior in the form of increased reappointment likelihood. Hence, I

---

[49] *See* R. Doak Bishop & Lucy Reed, *Practical Guidelines for Interviewing, Selecting, and Challenging Party-Appointed Arbitrators in International Commercial Arbitration*, 14 ARBITRATION INT'L 395, 395; Kapeliuk, *The Repeat Appointment Factor*, *supra* note 16, at 67; Seth H. Lieberman, *Something's Rotten in the State of Party-Appointed Arbitration: Healing ADR's Black Eye That Is "Nonneutral Neutrals"*, 5 CARDOZO J. CONFLICT RESOL. 215, 222 (2004); Andreas F. Lowenfeld, *The Party-Appointed Arbitrator in International Controversies: Some Reflections*, 30 TEX. INT'L L. J. 59, 65 (1995); William W. Park, *A Fair Fight: Professional Guidelines in International Arbitration*, 30 ARBITRATION INTERNATIONAL 409, 413 (2014).

[50] A survey of users of international arbitration has found that prior arbitration experience is a key factor in the selection of party-appointed arbitrators. *See* Queen Mary University of London, *2010 International Arbitration Survey: Choices In International Arbitration* 25 (2010), at 26, *available at* http://www.arbitration.qmul.ac.uk/docs/123290.pdf (reporting that 58% of survey states considered prior experience with arbitration a key factor in arbitrator selection).

[51] *See* Park, *supra* note 1, at 658.

[52] The ICSID Convention, art. 14.

examine whether prospective bias will increase an arbitrator's reappointment likelihood by each appointing entity:

Corollary 2: An appointing entity will be more likely at case appointment time (T_n) compared to prior case appointment time (T_{n-1}) to appoint an arbitrator whose reputation has shifted in the direction favored by that entity at T_n compared to T_{n-1}.

2. Retrospective Bias

Asides from the desire to shape reappointment encouraging reputations, Arbitrators may also exhibit bias to cater to those entities which have generated more past appointments out of, for example, loyalty or reciprocity. Such loyalty effect has been shown to exist among Supreme Court Justices.[53] In ICSID arbitration, there is also a view that ICSID arbitrators appear to be earning a reputation of lacking independence and impartiality due to multiple appointments by the same parties or counsel.[54] I examine whether arbitrator voting behavior is affected by appointment history as follows:

Prediction 2: at case decision time an arbitrator appointed by an appointing entity a higher percentage of the time will more likely vote in the way favored by that entity.

III. DATA

A. Dataset Description

The dataset consists of panel data at arbitrator-vote level[55] of all publicly available concluded and pending ICSID cases registered before January 1, 2016, mainly collected from the

---

[53] See Lee Epstein and Eric A. Posner, *Supreme Court Justices' Loyalty to the President*, 45 J. LEGAL STUD. 401 (2016).

[54] See Giraud, *supra* note 4.

[55] Each arbitrator, not each case, constitutes an observation in the dataset. Specifically, for each case, there are usually three units of arbitrator-level variables relating to the appointment, vote, and other information corresponding to each of the three arbitrators in that case. The dataset also includes appointment information related to arbitrators who were

ICSID website, with supplemental information from ITALaw website and UNCTAD's Investment

Dispute Settlement Navigator.[56] Summary statistics of the data appear in Table 2.1. The data

covers 1873 appointments concerning 449 unique individual arbitrators and 500 cases.[57] Figure

A2.2 reports the number of registered cases in the dataset by year.[58] The number of appointments

per arbitrator varies widely. 236 of 449 arbitrators are one-shot players, while certain arbitrators

have been appointed over 30 times. Figure A2.3 shows the number of repeat players at each

frequency of appointment.

Table 2.1: Summary Statistics

Appointments		
	Investor	508
	State	507
	ICSID	539
	Parties	217
	Co-arbitrators	87
	Source unclear	15
	Total	1873
Individual arbitrators		
	One-time players	236
	Repeat players	213
	Total	449
Cases		
	Pending	131
	Concluded[59]	369
	Total	500
Votes[60]		

replaced, removed due to incapacity or resignation, or disqualified during the arbitration in accordance with Rules 7, 8 or 9 of ICSID Arbitration Rules.

[56] The dataset records the status of the case as of March 24, 2017. The dataset also includes all publicly available information on annulment and resubmission of cases registered before January 1, 2016. The dataset does not include correction, interpretation, rectification, revision or supplementary decision proceedings as they usually do not involve new arbitrator appointments or substantive rulings on the merits of the case, thereby provide no valuable information to the appointing entities. *See* International Centre for Settlement of Investment Disputes, https://icsid.worldbank.org/en/Pages/cases/AdvancedSearch.aspx; ITALaw, https://www.italaw.com/ search/site; UNCTAD, http://investmentpolicyhub.unctad.org/ISDS/.

[57] Since this paper discusses arbitrators' biases due to *ad hoc* appointment, cases settled or discontinued before the appointment of any arbitrator are excluded.

[58] In some years the number differs from the statistics published by ICSID because those cases that were settled or discontinued before the appointment of any arbitrator are excluded from the dataset.

[59] Of the 369 concluded cases, 266 cases reach the final award stage and 103 cases were discontinued.

[60] For a detailed description of how the votes are coded, see Part 3 of Section III.

Investor	378
State	443
Total	821

The data also allows one to identify the following information related to the merits or procedure of each case: the type of instrument invoked, category of case facts, economic sector of dispute, and proceeding type (Tables A2.1, A2.2, A2.3, A2.4). I control for these variables because they may be correlated with arbitrator reputation and may also affect the arbitrator's vote and appointment in each case.[61] In the Appendix, I rerun all regressions with different combinations of these controls and find all results robust to alternative controls.

B. Independent Variables

I generate three reputation-measuring variables which describe the extent of arbitrators' pro-investor, pro-state, or neutral leanings, which I call *Reputation for Investor, Reputation for State,* and *Reputation for Neutrality*. I measure these variables at both case registration and decision time to reflect when appointing entities make appointment decisions and when arbitrators vote. *Reputation for Investor* and *Reputation for State* are the percentage difference between an arbitrator's votes for investor or state and the overall proportion of investor or state wins in the dataset. The larger *Reputation for Investor* or *Reputation for State* the more pro-investor or pro-state the arbitrator as they have voted proportionally more often than average for investor or state, respectively. *Reputation for Neutrality* is the negative absolute value of *Reputation for Investor*, so as to ensure the coefficient sign applies consistently with respect to *Reputation for Neutrality*,

---

[61] The regressions do not control for the claims involved in the case, such as fair and equitable treatment, national treatment, direct and indirect expropriation, while an arbitrator's vote or appointment might be claim-specific. An arbitrator might possess a reputation with respect to his or her views on a particular claim instead of being pro-investor or pro-state generally. In addition, the collegial politics among the three members of a tribunal may also affect an arbitrator's vote. While arbitrators always have the option to dissent if they do not agree with the majority, appointing entities may not consider the final outcome of a case attributable to a single arbitrator, but a product of compromise among the three. These claim-specific dynamics and collegial politics among tribunal members are interesting and should be left for a future project to address.

Reputation for Investor and *Reputation for State*. Arbitrators with a larger (i.e. closer to zero) *Reputation for Neutrality* are more neutral because they deviate less from the overall winning rate of either side in the dataset.

Figure A2.4 shows the distribution of average *Reputation for Investor* based on all votes of individual arbitrators who have been appointed more than once in the dataset. There are more arbitrators with a strong reputation for neutrality than those with a strong reputation in favor of either side. There are also slightly more arbitrators with a pro-state reputation than those with a pro-investor reputation.

Appointment rates by the five appointing entities at case decision time are represented by the variables: *Investor Appointment Rate*, *State Appointment Rate*, *ICSID Appointment Rate*, *Parties Appointment Rate*, and *Co-Arbitrators Appointment Rate*. I also generate a variable, *Appointment Rate as President*, which is the percentage that an arbitrator has been appointed as president at case decision time.

C. Dependent Variables

Dummy variables *Vote Investor*, *Vote State*, and *Vote Neutral* denote an arbitrator's vote in a case. In original proceedings, additional facility proceedings, and resubmission proceedings, *Vote Investor* is set to one when the investor is awarded more than $0. *Vote State* is set to one when the investor is not awarded damages. *Vote Investor* in annulment proceedings is set to one when an award is annulled in a proceeding initiated by the investor, or when an award is not annulled in a proceeding initiated by the state. *Vote State* in annulment proceedings is set to one when an award is annulled in a proceeding initiated by the state, or when an award is not annulled in a proceeding initiated by the investor. If the arbitrator dissents on liability, the dummy variable will reflect the arbitrator's dissenting opinion, not the majority ruling. The dataset records 43 publicly available

dissenting opinions and 20 publicly available separate opinions issued by individual arbitrators. While it is more common for arbitrators to render awards in consensus, the consensus award should reflect the arbitrators' votes in the case, given that one did not opt for writing a dissenting or separate opinion when there is such an option. In all proceedings, *Vote Neutral* is set to one for each case when *Reputation for Neutrality* becomes larger (i.e. closer to zero) after voting. This indicates whether the arbitrator is becoming more neutral in the sense of aligning his or her votes with the overall average results of all ICSID arbitration cases at that point.

Figure A2.5 shows average voting rates of individual arbitrators appointed more than once for investors, states, and in a neutral direction.[62] Some arbitrators have exclusively voted for investor or state.[63] On average, arbitrators having voted more for states exceed those having voted more for investors. 105 out of the 157 arbitrators with publicly available decisions have voted in a way that increases *Reputation for Neutrality* more than half the time. Hence, a larger proportion of arbitrators have aligned votes with the average results of all ICSID arbitration cases at the time they vote.

I generate five dummy variables denoting arbitrator appointment, which equal 1 if the arbitrator is appointed in the current case by a particular appointing entity, and 0 if not: *Investor Appointment*, *State Appointment*, *ICSID Appointment*, *Parties Appointment*, and *Co-Arbitrators Appointment*. I also generate another dummy variable *Appointment as President* to denote whether the arbitrator is appointed as the presiding arbitrator.

Figure A2.6 shows the average appointment rate of individual arbitrators appointed more than once by appointing entity, as well as average appointment rate as president. Some arbitrators

---

[62] Some repeat players who have been appointed multiple times have not rendered a publicly available final decision in a case yet, thus there is no record of their vote in the dataset.

[63] Among the repeat players, 35 arbitrators have exclusively voted for investor, 41 arbitrators have exclusively voted for state.

have been exclusively appointed by certain types of entities.[64] No arbitrator has been exclusively appointed by co-arbitrators or appointed over half the time by co-arbitrators. 53 out of the 213 repeat players have been appointed as president over half the time.

IV. METHODOLOGY AND RESULTS

A. Prediction 1: Prospective Bias

1. Empirical Design

To test Prediction 1, I employ a linear probability model with the following specification:[65]

$$Vote_{i,c} = \beta_0 + \beta_1 Reputation_{i,c} + X_{i,c} + \Phi_i + \varepsilon_{i,c}, \tag{1}$$

where i is an indicator for the arbitrator, c for the case. $Vote_{i,c}$ represents *Vote Investor*, *Vote State*, or *Vote Neutral. Reputation*$_{i,c}$ represents the arbitrator's *Reputation for Investor*, *Reputation for State*, or *Reputation for Neutrality* at case decision time. $X_{i,c}$ is a vector of control variables *Instrument Dummies*, *Case Fact Dummies*, *Economic Sector Dummies* and *Proceeding Dummies*. All standard errors are clustered for individual arbitrators. Φ_i is an arbitrator-specific fixed effect, which controls for unobserved arbitrator heterogeneity in pre-existing bias. Pre-existing bias can affect an arbitrator's vote and needs to be teased out of a regression that measures effects of prospective bias alone.

Due to how the reputation variables are constructed, even absent causal relationship between reputation and voting behavior, $\hat{\beta}_1$ can nevertheless appear to be negative after demeaning

---

[64] Among the repeat players, 14 arbitrators have been exclusively appointed by investor, 25 arbitrators have been exclusively appointed by state, 25 arbitrators have been exclusively appointed by ICSID, 2 arbitrators have been exclusively appointed by parties.

[65] I use linear probability models rather than logit models for ease of interpretation. The results are statistically and substantively comparable when using logit models.

variables in fixed effect models.[66] The true value of β_1 should be equivalent to the difference between $\hat{\beta}_1$ and the estimated coefficient in a placebo regression where by construction there is no causal relationship between reputation and voting behavior. Hence, I generate 1,000 placebo datasets where the assignment of votes for each arbitrator is randomized across cases in 1,000 different ways. In each placebo dataset, I generate a *Reputation for Investor* and *Reputation for Neutrality* for each arbitrator based on the new sequence of their votes. In this way, the placebo datasets have the same number of observations for each arbitrator, and each arbitrator has the same rate of deciding in favor of investors as the real data. However, the randomization ensures that there is no causal relationship between reputation and voting behavior. I rerun regressions using these 1,000 placebo datasets and generate a distribution of the estimated coefficients for *Reputation for Investor* and *Reputation for Neutrality*. I then compare $\hat{\beta}_1$ with the distribution of placebo coefficients to determine the true value of β_1. If $\hat{\beta}_1$ is significantly larger than the distribution of placebo coefficients, then there is evidence that arbitrators have prospective bias for polarity. If $\hat{\beta}_1$ is significantly smaller than the distribution of placebo coefficients, then there is evidence that arbitrators have prospective bias for neutrality. If $\hat{\beta}_1$ is not significantly different from the distribution of placebo coefficients, then there is no evidence of prospective bias, as the arbitrator behavior in the real data is not different from that in the placebo data which is manipulated to ensure there is no prospective bias in either direction.

---

[66] An arbitrator's reputation at the decision time of a case (*Reputation$_{i,c}$*) derives from a cumulative average of her votes before the case. The mean reputation of an arbitrator ($\overline{Reputation}_i$) derives from a cumulative average of all her votes in the dataset. In fixed effects model, the covariant is equivalent to *Reputation$_{i,c}$* $-$ $\overline{Reputation}_i$. When *Reputation$_{i,c}$* $-$ $\overline{Reputation}_i$ is positive, it is more likely that in case c the arbitrator's vote is 0, because there must be more 0s in following cases for $\overline{Reputation}_i$ to be smaller than *Reputation$_{i,c}$*. When *Reputation$_{i,c}$* $-$ $\overline{Reputation}_i$ is negative, it is more likely that in case c the arbitrator's vote is 1, because there must be more 1s in following cases for $\overline{Reputation}_i$ to be larger than *Reputation$_{i,c}$*. Thus, $\hat{\beta}_1$ can be significantly negative even when there is no actual causal relationship between reputation and voting behavior.

2. Results

Table 2.2 presents regression results for Prediction 1. The estimated coefficients $(\hat{\beta}_1)$ in all three regressions are negative and significant at the 1% level, which seem to indicate that arbitrators exhibit prospective bias for neutrality. However, in placebo regressions the estimated coefficients are also negative and statistically significant.[67] To compare between $\hat{\beta}_1$ and the distribution of placebo coefficients, I report a pseudo p-value equal to the proportion of placebo estimates whose absolute value is greater than the absolute value of $\hat{\beta}_1$. As Figures 2.1 and 2.2 show, $\hat{\beta}_1$ for *Reputation for Investor* (-0.386) and *Reputation for Neutrality* (-1.505) fall at the 92nd and 78th percentiles of the placebo distribution, respectively. The pseudo p-value is 0.16 and 0.44 respectively, which suggests that $\hat{\beta}_1$ are not significantly different from the distribution of placebo coefficients. Thus, I cannot reject the null hypothesis that the true value of β_1 is likely 0. To ensure that the results are not driven by infrequent players who have not had an opportunity to establish a reputation yet, in Appendix B, I repeat the analysis after excluding arbitrators who have been appointed for fewer than three times and five times. I find similar results. Therefore, Predictions 1(a) and 1(b) are unsupported by the data and Prediction 1 returns "null" result, indicating no evidence of prospective bias.

---

[67] The placebo estimates for *Reputation for Claimant* appear to be significant at the 1% level 98.5 percent of the time, the 5% level 1.3 percent of the time, and the 10% level 0.2 percent of the time. The placebo estimates for *Reputation for Neutrality* appear to be significant at the 1% level 100 percent of the time.

Table 2.2: Effects of Arbitrator's Reputation on Vote

	(1) Vote Claimant	(2) Vote Respondent	(3) Vote Neutral
Reputation for Claimant	-0.386**		
	(0.107)		
Reputation for Respondent		-0.386**	
		(0.107)	
Reputation for Neutrality			-1.505**
			(0.124)
Instrument Dummies	Yes	Yes	Yes
Case Fact Dummies	Yes	Yes	Yes
Economic Sector Dummies	Yes	Yes	Yes
Proceeding Dummies	Yes	Yes	Yes
Arbitrator Fixed Effect	Yes	Yes	Yes
Observations	539	539	539

Note: standard errors, in parentheses, are clustered at arbitrator level.
[†] Significant at the 10% level; [*] Significant at the 5% level; [**] Significant at the 1% level.

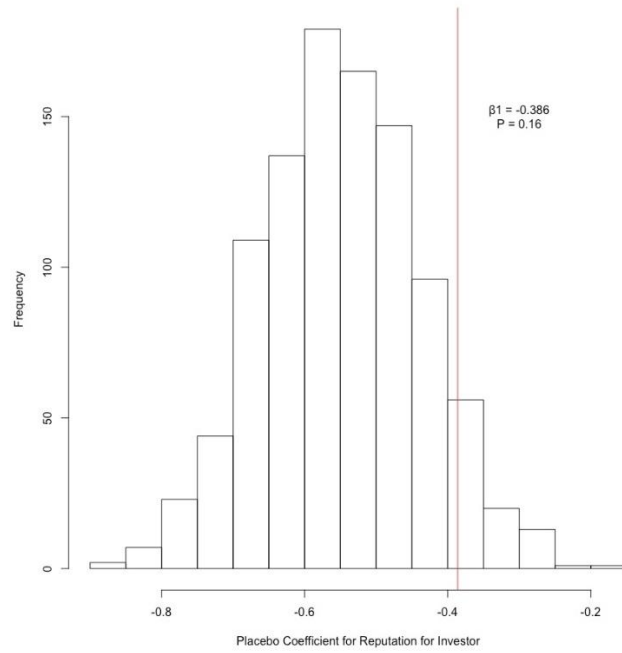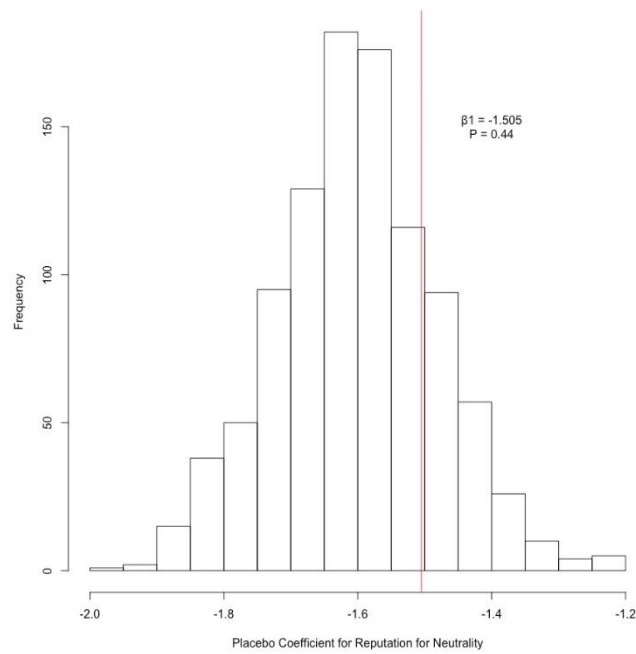Figure 2.1: Distribution of placebo coefficients for Reputation for Investor

Figure 2.2: Distribution of placebo coefficients for Reputation for Neutrality

The absence of evidence for prospective bias does not necessarily mean arbitrators possess no bias. A peripheral note which we can identify from the regression results is that a significant number of ICSID arbitrators are pro-investor or pro-state, but this is more determined by their predispositions than strategic reputation-shaping. Figure A2.7 shows the distribution of Φ_i, which indicates an arbitrator's likelihood of voting for investor because of pre-existing bias, and thereby reflects the level and direction of individual arbitrators' pre-existing bias (or lack thereof). There are more arbitrators with pre-existing bias in favor of states than those with pre-existing bias in favor of investors. Around 20 percent of arbitrators do not exhibit pre-existing bias in favor of either side.

Although not directly related to Prediction 1, I also empirically examine the time trend of an arbitrator's reputation development. The time trend does not lend support to the idea that arbitrators attempt to shape reputations to promote reappointment by voting increasingly pro-investor or pro-state. Figure A2.8 shows changes in individual *Reputation for Investor* and *Reputation for State* over time as votes are cast for the twenty most frequently appointed arbitrators in the dataset, of which nine are average pro-investor and eleven are average pro-state. Most of these repeat players tend to shift from a polarized reputation towards a more neutral one as they cast more votes. Regression of the time trend of arbitrator reputation development further shows that reputation tends to shift towards a neutral direction (Table A2.5).

3. Corollaries

In order to further understand the incentives behind arbitrators' behavior, I investigate the two corollaries.

Turning to Corollary 1, Figure 2.3 presents variations in *Reputation for Investor* at appointment time sorted by appointing entities and president appointment. Investor-appointed

arbitrator reputations tend to skew pro-investor and state-appointed arbitrator reputations tend to skew pro-state. Reputations for parties-appointed and co-arbitrator-appointed arbitrators as well as presiding arbitrators tend to skew neutral. Reputations of ICSID-appointed arbitrators seem to skew pro-state. [68] These observations are largely consistent with predicted preferences of appointing entities.

Figure 2.3: Reputation for Investor by the time of appointment from each appointing entity
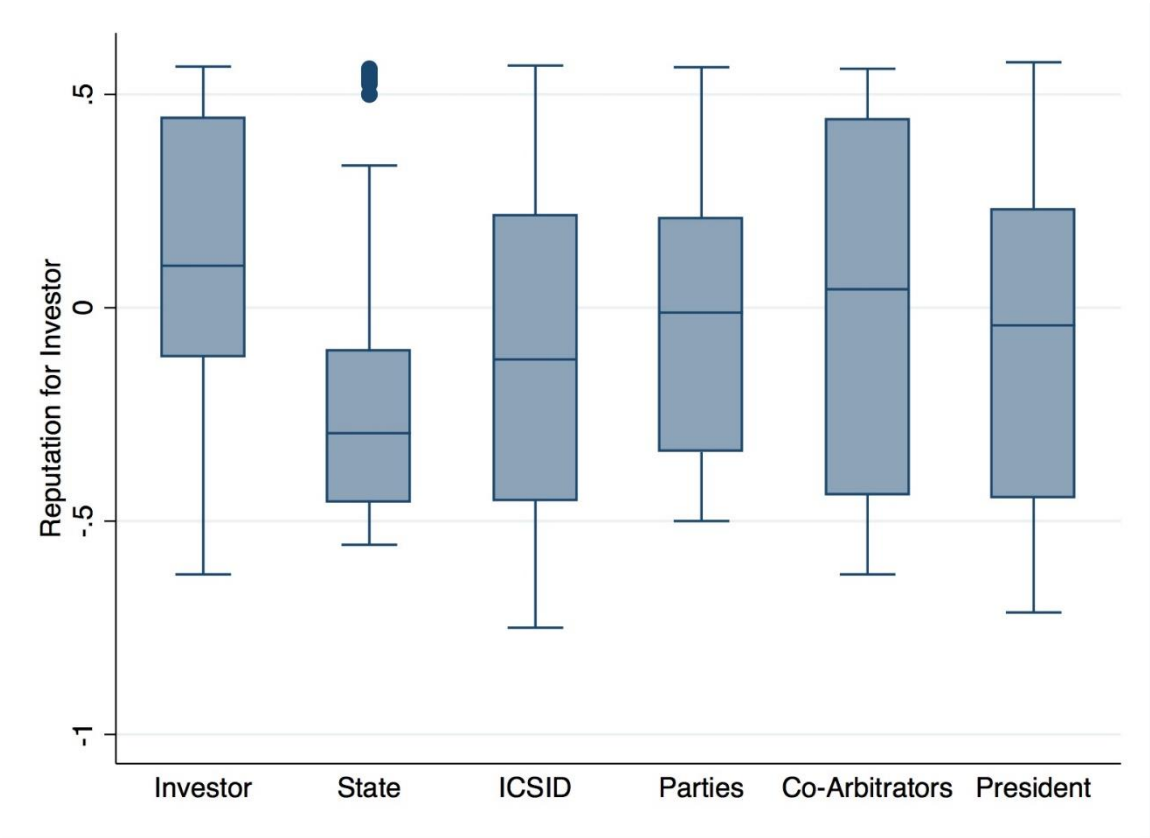
To test Corollary 1, I employ a linear probability model with the following specification:

$$Appointment_{i,a} = \beta_0 + \beta_1 Reputation_{i,a} + X_{i,a} + \varepsilon_{i,a}, \qquad (2)$$

---

[68] According to ICSID Arbitration Rule 4(4), the Chairman of the Administrative Council will have to appoint arbitrators from a Panel of Arbitrators. Since a vast majority of the members of the Panel of Arbitrators were designated by Contracting States, it is possible that the arbitrators appointed by the Chairman of the Administrative Council are more state-friendly. However, as shown in Table A2.6, the regression results do not suggest that arbitrators with a stronger reputation for state are significantly more likely to be appointed by ICSID.

where i is an indicator for the arbitrator, a for the appointment. *Appointment$_{i,a}$* represents *Investor Appointment*, *State Appointment*, *ICSID Appointment*, *Parties Appointment*, *Co-Arbitrators Appointment*, or *Appointment as President*. *Reputation$_{i,a}$* represents an arbitrator's *Reputation for Investor*, *Reputation for State*, or *Reputation for Neutrality* at case registration time. *X$_{i,a}$* represents the same set of control variables as equation (1). All standard errors are clustered for individual arbitrators.

I do not include arbitrator-specific fixed effects because appointing entities usually compare between different candidates when making appointment decisions. Thus, apart from prospective bias, the heterogeneity in pre-existing bias between individual arbitrators also goes into appointing entities' preference calculations. Without controlling for arbitrator-specific fixed effects, $\hat{\beta}_1$ measures effects of reputation (pre-existing bias and prospective bias) on appointment.

There would be a selection bias if I tested Corollary 1 by considering arbitrators in the dataset appointed to a case without considering those not appointed. To resolve this issue, for each case, I assign every non-appointed arbitrator who has appeared on ICSID tribunals previously a *Reputation for Investor*, *Reputation for State* and *Reputation for Neutrality* by case registration time.[69] I then assign an appointment variable equal to 0 for each of the non-appointed arbitrators. However, if in a case an appointing entity (e.g. ICSID) appoints no arbitrator, then the corresponding appointment variable is recorded as "missing" instead of 0. I then use the newly constituted dataset to run the regressions that examine the relationship between the arbitrator's reputation and whether they are appointed by the appointing entity expected to favor that reputation. In this way, all arbitrators who have appeared on ICSID tribunals previously but are not appointed to a specific case are accounted for in the empirical analysis.

---

[69] Non-appointed arbitrators who had not been previously appointed to any ICSID tribunals are not included here as they do not have a reputation based on prior decisions.

The regression results in Table 2.3 largely support Corollary 1. Arbitrators with a stronger reputation for investor at case registration time are significantly more likely to be appointed by investor in that case, though only at the 10% level.[70] Arbitrators with a stronger reputation for state at case registration time are significantly more likely to be appointed by state in that case. Arbitrators with a stronger reputation for neutrality at case registration time are significantly more likely to be appointed by parties and as president of the tribunal in that case. The two appointing entities whose appointments are not significantly affected by the reputations that they are expected to favor are ICSID and co-arbitrators, though the results have the predicted sign.[71]

The coefficient magnitudes in Table 2.3 may seem small, but these are large relative to the baseline probability of any given arbitrator being appointed by a specific appointing entity in a given case.[72] The increase in likelihood of being appointed by states relative to the baseline for arbitrators who are more pro-state is 34%, the increase in likelihood of being appointed as president relative to the baseline for arbitrators who are more neutral is 49%, and the increase in likelihood of being appointed by parties relative to the baseline for arbitrators who are more neutral is 88%. Thus, Table 2.3 provides evidence that appointment patterns respond to differences in arbitrator reputation, although it does not distinguish between the effect of pre-existing bias and prospective bias. In Appendix B, I exclude infrequent players and repeat the analysis. The results remain largely similar except for the coefficient for *Appointment as President*, which loses significance after excluding infrequent players.

---

[70] As shown in Table A2.8, after dropping arbitrators who have been appointed for fewer than three times, the result is significant at the 5% level.

[71] Since ICSID can only make appointments from the Panel of Arbitrators, the lack of significance might result from the fact that the dataset does not have information about whether an arbitrator is on the Panel of Arbitrators at the time of appointment. As for co-arbitrators, since they have only made 87 appointments in total, the result might be insignificant due to the small size of data.

[72] The baseline average probability of an arbitrator being appointed by investor, state, ICSID, parties, co-arbitrators and as president in the newly-constituted dataset are respectively 0.0176, 0.0134, 0.0233, 0.0164, 0.00596, and 0.0165.

Table 2.3: Effects of Arbitrator's Reputation on Appointment without Arbitrator-Specific Fixed Effects

	(1) Investor Appointment	(2) State Appointment	(3) ICSID Appointment	(4) Parties Appointment	(5) Co-arbitrators Appointment	(6) Appointment as President
Reputation for Investor	0.00263[†]					
	(0.00145)					
Reputation for State		0.00452[**]				
		(0.00168)				
Reputation for Neutrality			0.00615	0.0145[**]	0.00682	0.00813[**]
			(0.00399)	(0.00490)	(0.00527)	(0.00289)
Instrument Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Case Fact Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Economic Sector Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Proceeding Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Arbitrator Fixed Effect	No	No	No	No	No	No
Observations	54823	53502	38563	25847	8169	74696

Note: standard errors, in parentheses, are clustered at arbitrator level.
[†] Significant at the 10% level; [*] Significant at the 5% level; [**] Significant at the 1% level.

I then investigate Corollary 2. I employ a linear probability model identical to equation (2) except for including arbitrator-specific fixed effects:

$$Appointment_{i,a} = \beta_0 + \beta_1 Reputation_{i,a} + X_{i,a} + \Phi_i + \varepsilon_{i,a}, \tag{3}$$

I include arbitrator-specific fixed effects here because after holding constant the level of pre-existing bias, $\hat{\beta}_1$ measures effects of prospective bias alone on appointment. The results reveal whether prospective bias plays a role in determining an appointing entity's appointment decision.

The regression results in Table 2.4 reject most of Corollary 2. As Table 2.4 shows, after adding arbitrator-specific fixed effects, when an arbitrator has a stronger reputation for state than before, the arbitrator will be significantly more likely to be appointed by states. However, when an arbitrator has a stronger reputation for neutrality than before, the arbitrator will be significantly less likely to be appointed by ICSID or as president.[73] In other words, the estimated coefficient has the "wrong" sign. The other three regressions related to investors, parties and co-arbitrators return "null" results. Table 2.4 indicates that, after controlling for pre-existing bias, there is little evidence that appointment patterns respond to changes in reputation. The results remain similar after excluding infrequent players in Appendix B.

Taken together, Tables 2.3 and 2.4 suggest that while most appointing entities are more likely to appoint arbitrators with the types of reputation that they are expected to favor when selecting from a pool of candidates with different reputations, such preference should be largely attributed to the heterogeneity in pre-existing bias. Prospective bias alone appears ineffective in generating reappointments except in the case of state appointments. This provides a potential explanation for why arbitrators seem to not display any prospective bias by consciously shaping their reputations.

---

[73] The reason why arbitrators who become more neutral turn out to be significantly less likely to get reappointments from ICSID or as president may be that candidates for ICSID appointees and presiding arbitrators are subject to more scrutiny with respect to relational conflicts, while arbitrators who become more neutral after having sit in more cases (as Table A2.5 shows) are likely to be more vulnerable under such scrutiny. In addition, ICSID has been endeavoring to promote diversity in the pool of its appointees. Thus, ICSID may avoid appointing arbitrators who have appeared for too many times, even though they are also developing a reputation for neutrality. *See* Eloise Obadia and Frauke Nitscheke, *The Role of The Secretariat*, *in* Chiara Giorgetti ed. Litigating International Investment Dispute: A Practitioner's Guide 115; Sergio Puig, *Emergence & Dynamism in International Organizations: ICSID, Investor-State Arbitration & International Investment Law*, 44 Geo. J. Int'l L. 531, 552—553 (2013).

Table 2.4: Effects of Arbitrator's Reputation on Appointment with Arbitrator-Specific Fixed Effects

	(1) Investor Appointment	(2) State Appointment	(3) ICSID Appointment	(4) Parties Appointment	(5) Co-arbitrators Appointment	(6) Appointment as President
Reputation for Investor	-0.00643					
	(0.00484)					
Reputation for State		0.00703*				
		(0.00335)				
Reputation for Neutrality			-0.0226*	-0.000282	-0.0157	-0.00888*
			(0.00883)	(0.00732)	(0.0155)	(0.00430)
Instrument Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Case Fact Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Economic Sector Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Proceeding Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Arbitrator Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	54823	53502	38563	25847	8169	74696

Note: standard errors, in parentheses, are clustered at arbitrator level.
[†] Significant at the 10% level; [*] Significant at the 5% level; [**] Significant at the 1% level.

B. Prediction 2: Retrospective Bias

To test Prediction 2, I employ a linear probability model with the following specification:

$$Vote_{i,c} = \beta_0 + \beta_1 Appointment\ Rate_{i,c} + X_{i,c} + \Phi_i + \varepsilon_{i,c}, \qquad (4)$$

where i is an indicator for the arbitrator, c for the case. $Vote_{i,c}$ represents *Vote Investor*, *Vote State*, or *Vote Neutral. Appointment Rate$_{i,c}$* represents an arbitrator's *Investor Appointment Rate*, *State Appointment Rate*, *ICSID Appointment Rate*, *Parties Appointment Rate*, *Co-Arbitrators*

Appointment Rate, or *Appointment as President Rate* at case decision time. $X_{i,c}$ represents the same set of control variables as equation (1). All standard errors are clustered for individual arbitrators.

The regression results in Table 2.5 suggest that none of the appointing entities appears to exert an influence on the voting behavior of their frequent appointees. The point estimates for an arbitrator's likelihood of voting in a neutral direction are fairly large. Thus, the results do not allow one to completely reject a relationship between an arbitrator's previous appointment rate by ICSID, by parties, by co-arbitrators, or as president, and their subsequent vote that shifts reputations in a neutral direction. The results are interpreted as "null" results: they offer no evidence that in ICSID arbitration, on average an arbitrator is voting to cater to the appointing entity that has given appointments a higher percentage of times in the past.

Table 2.5: Effects of Appointment Rate by an Appointing entity on Arbitrators' Votes

	(1) Vote Investor	(2) Vote State	(3) Vote Neutral	(4) Vote Neutral	(5) Vote Neutral	(6) Vote Neutral
Investor Appointment Rate	-0.0388					
	(0.196)					
State Appointment Rate		0.0657				
		(0.228)				
ICSID Appointment Rate			0.238			
			(0.330)			
Parties Appointment Rate				0.438		
				(0.325)		
Co-Arbitrators Appointment Rate					-0.885†	
					(0.518)	
Appointment Rate as President						0.241
						(0.274)
Instrument Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Case Fact Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Economic Sector Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Proceeding Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Arbitrator Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	780	780	537	537	537	537

Note: standard errors, in parentheses, are clustered at arbitrator level.

† Significant at the 10% level; * Significant at the 5% level; ** Significant at the 1% level.

V. CONCLUSION

In this paper, I look for empirical evidence of prospective and retrospective bias due to the *ad hoc* nature of arbitrator appointment in ICSID arbitration and find none. No evidence suggests that arbitrators vote in biased ways to increase reappointment by shaping reputations over time or to cater to entities responsible for a greater number of past appointments. In addition, contrary to becoming more pro-investor or pro-state over time, I find arbitrator reputations tend to shift in a neutral direction towards the average of all ICSID arbitration cases.

A number of explanations may account for why the presence of prospective or retrospective biases arising from *ad hoc* appointment is unsupported by the data. Firstly, appointment appears to be influenced by pre-existing bias instead of how arbitrator reputations develop over time. Arbitrators may not be incentivized to adapt their reputations or voting patterns to encourage reappointment, as there may not be a belief that such behavior will be rewarded. Secondly, another consideration may be deterrence built into the ICSID system. Parties may challenge an arbitrator for disqualification on account of evidence indicating "manifest lack of the qualities" required to serve as an arbitrator, [74] which would include previous decisions or appointments indicative of bias.[75] Although the "manifest lack of [the] qualities" standard has been considered an extremely high bar for challenging an arbitrator,[76] the existence of this challenge procedure may nevertheless

---

[74] According to Article 57 of the ICSID Convention, a party may propose to a Commission or Tribunal the disqualification of any of its members on account of any fact indicating a manifest lack of the qualities required by paragraph (1) of Article 14 ("Persons designated to serve on the Panels shall be persons of high moral character and recognized competence in the fields of law, commerce, industry or finance, who may be relied upon to exercise independent judgment. Competence in the field of law shall be of particular importance in the case of persons on the Panel of Arbitrators").

[75] *See e.g. Blue Bank International & Trust (Barbados) Ltd. v. Bolivarian Republic of Venezuela*, Decision on the Parties' Proposal to Disqualify a Majority of the Tribunal, ICSID Case No. ARB/12/20, 12 November 2013, at 10 – 13; *Caratube International Oil Company LLP & Mr. Devincci Salah Hourani v. Republic of Kazakhstan*, Decision on the Proposal for Disqualification of Mr. Bruno Boesch, ICSID Case No. ARB/13/13, 20 March 2014, at 15 – 35.

[76] *See* LUCY REED, JAN PAULSSON & NIGEL BLACKABY, GUIDE TO ICSID ARBITRATION 80 (2004).

discourage appearance of bias. Finally, anecdotal evidence also suggests that reputation for bias might undercut an arbitrator's credibility within a tribunal and erode their career prospects long term.[77]

The results of this study provide tentative evidence that an investment court system may not be more effective in achieving independence and impartiality among adjudicators of investor-state disputes if the judges are to be drawn from the same pool as the arbitrators. It is true that judges could have more *perceived* impartiality because of the institutional safeguards built into the court system. But if we believe that actual impartiality matters, then only evidence of biases directly arising from *ad hoc* appointment would justify a wholesale change in the *ad hoc* nature of appointment inherent to arbitration. Pre-existing bias will exist regardless of how adjudicators are appointed. Investment court reforms could also cause problems not currently present in the arbitration system, such as the formation of pro-state biases among adjudicators.[78] Meanwhile, other criticisms of investor-state arbitration typically have potential solutions which fall short of abandoning *ad hoc* appointment completely.[79] For instance, legitimate concerns like double-hatting of arbitrators can be addressed by less radical means such as adopting Codes of Conduct for Arbitrators.

The ability of disputing parties to appoint adjudicators of their own choice has long been one of the most attractive aspects of investor-state arbitration.[80] When disputing parties come from different legal and cultural background, their input in adjudicator selection is essential for ensuring

---

[77] *See* Brower & Rosenberg, *supra* note 34, 13 – 14; KARTON, *supra* note 11, at 55, 117.

[78] *See* Nikos Lavranos, *How the European Commission and the EU Member States Are Reasserting Their Control over Their Investment Treaties and ISDS Rules*, *in* ANDREAS KULICK ED., REASSERTION OF CONTROL OVER THE INVESTMENT TREATY REGIME 323 (2017).

[79] For a summary of existing criticism about the investor-state dispute settlement system, see Kaufmann-Kohler and Potestà, *supra* note 8, at 12-15.

[80] *See* Loukas Mistelis & Crina Mihaela Baltag, *Trends and Challenges in International Arbitration: Two Surveys of In-House Counsel of Major Corporations*, 2 WORLD ARB. & MED. REV. 83, 94 (2008).

their confidence in the proceedings and respect for the awards.[81] Absent concrete empirical evidence of bias arising from *ad hoc* appointment, reform efforts should focus instead on improvements to existing investor-state arbitration systems rather than embarking on a potentially unnecessary search for a replacement.

---

[81] *See* Giorgio Sacerdoti, *Is the Party-Appointed Arbitrator a "Pernicious Institution"? A Reply to Professor Hans Smit*, 35 COLUMBIA FDI PERSPECTIVE (2011), *available at* http://ccsi.columbia.edu/files/2014/01/FDI _35.pdf.

A. Gambler's Fallacy

While there is no evidence that arbitrators are strategically shaping a reputation for neutrality, arbitrators' reputations might be becoming more neutral because of the tendency of arbitrators to avoid voting in streaks as they tend to underestimate the likelihood of streaks occurring by chance and attempt to appear fair by engaging in negatively auto-correlated decision-making.[82] I examine whether the arbitrator's current vote is significantly negatively correlated with the lagged vote, which suggests whether the arbitrator may be suffering from the gambler's fallacy. The empirical specification is presented below. *Vote Investor$_{i,c-1}$* is the lagged variable. $X_{i,c}$ represents the same set of control variables as equation (1).

$$\textit{Vote Investor}_{i,c} = \beta_0 + \beta_1 \textit{Vote Investor}_{i,c-1} + X_{i,c} + \Phi_i + \varepsilon_{i,c} \qquad (5)$$

Table A2.7 does not report any significant result with respect to the relationship between the current vote and the lagged vote. Therefore, it lacks empirical support to claim that arbitrators' reputations are becoming more neutral because of the effect of the gambler's fallacy.

B. Infrequent Players

Arbitrators who have only been appointed for a few times may not have had an opportunity to establish a reputation. To reduce the noise created by these infrequent players, I reexamine the relationship between arbitrators' reputations and their subsequent vote and appointment as well as the time trend of *Reputation for Neutrality* after dropping all arbitrators who have been appointed for fewer than three times from the dataset. As Figure A2.9.1 and A2.9.2 show, after dropping

---

[82] For an empirical study of decision-making under the gambler's fallacy in the case of asylum judges, see Daniel L. Chen, Tobias J. Moskowitz & Kelly Shue, *Decision Making under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, And Baseball Umpires*, 131 QUARTERLY JOURNAL OF ECONOMICS 1181 (2016).

arbitrators who have been appointed for fewer than three times, $\hat{\beta}_1$ for *Reputation for Investor* (-0.386) and *Reputation for Neutrality* (-1.505) fall at the 92nd and 78th percentiles of the placebo distribution, respectively. The pseudo p-value is 0.16 and 0.72 respectively, which suggests that $\hat{\beta}_1$ are not significantly different from the distribution of placebo coefficients. As Figure A2.10.1 and A2.10.2 show, after dropping arbitrators who have been appointed for fewer than five times, $\hat{\beta}_1$ for *Reputation for Investor* (-0.366) and *Reputation for Neutrality* (-1.460) fall at the 95th and 65th percentiles of the placebo distribution, respectively. The pseudo p-value is 0.09 and 0.70 respectively. $\hat{\beta}_1$ for *Reputation for Investor* is significantly different from the distribution of placebo coefficients only at the 10% level. $\hat{\beta}_1$ for *Reputation for Neutrality* is not significantly different from the distribution of placebo coefficients. Overall, after dropping infrequent players, I still cannot reject the null hypothesis that the true value of β_1 is likely 0.

As Table A2.8 shows, without arbitrator-specific fixed effect, arbitrators with a stronger reputation for investor, for state and for neutrality are still significantly more likely to be appointed by investor, by state, and by parties. Yet the coefficient related to appointment by investor is now significant at the 5% level, while the coefficient related to appointment by parties is now only significant at the 10% level. The coefficient related to appointment as president has lost significance. After controlling for arbitrator-specific fixed effect, as Table A2.9 shows, the coefficient related to appointment by ICSID is now significant at 1% level. The coefficient related to appointment by parties becomes positive, though it remains insignificant. This suggests that the significant results related to appointment by parties and as president can be driven by those infrequent players who have a weak reputation for neutrality and have been appointed for few times.

Meanwhile, infrequent players do not create much noise with respect to the time trend of *Reputation for Neutrality*. As suggested by Table A2.10.1, the significance level and magnitude of the coefficient that denotes the changes of an arbitrator's *Reputation for Neutrality* over time are the same as those in Table A2.5. In Table A2.10.2, where arbitrators who have been appointed for fewer than five times are also dropped, the results are still similar to those in Table A2.5.

C. Different Control Variables

In order to make sure that the findings do not result from the particular combination of control variables used, I rerun the regressions with various controls excluded and included. As shown in Tables A2.11, A2.12, A2.13, A2.14, alternate specifications generate results with respect to the key findings that are largely similar to those exhibited in Tables 2.2, 2.3, 2.4, 2.5.

Figure A2.1: Changes in average state win rate 2000 – 2017
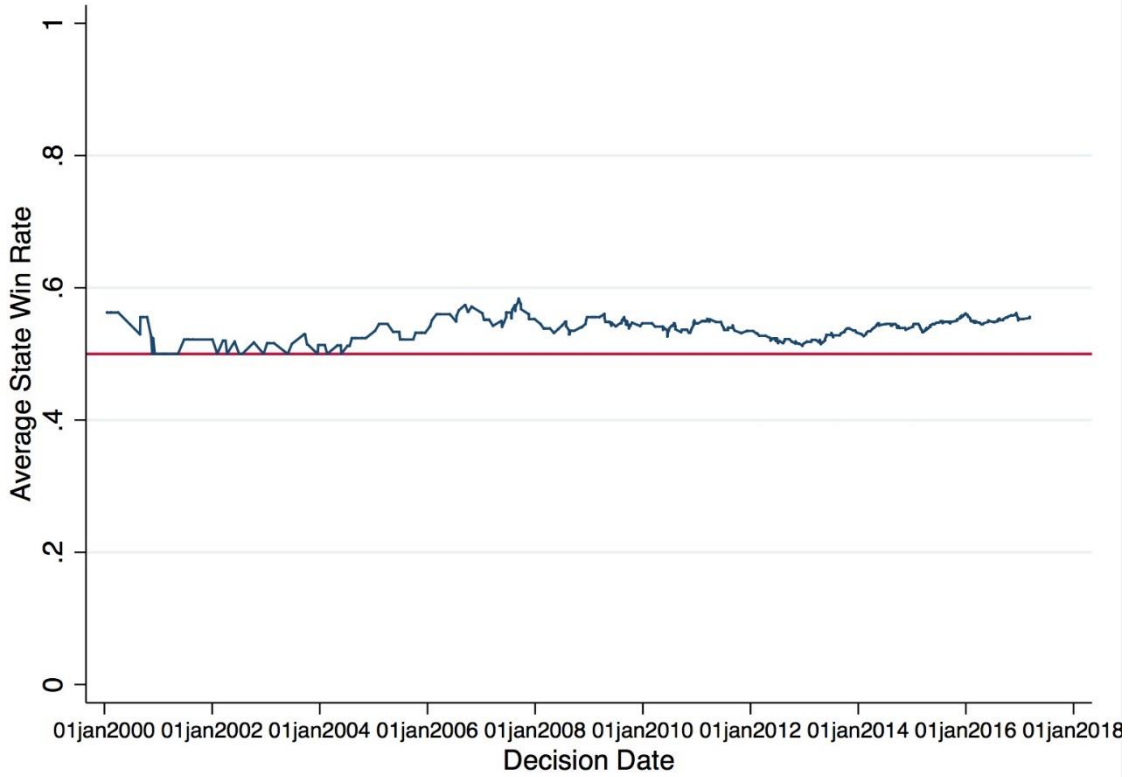
Figure A2.2: Registered cases 1972 – 2015
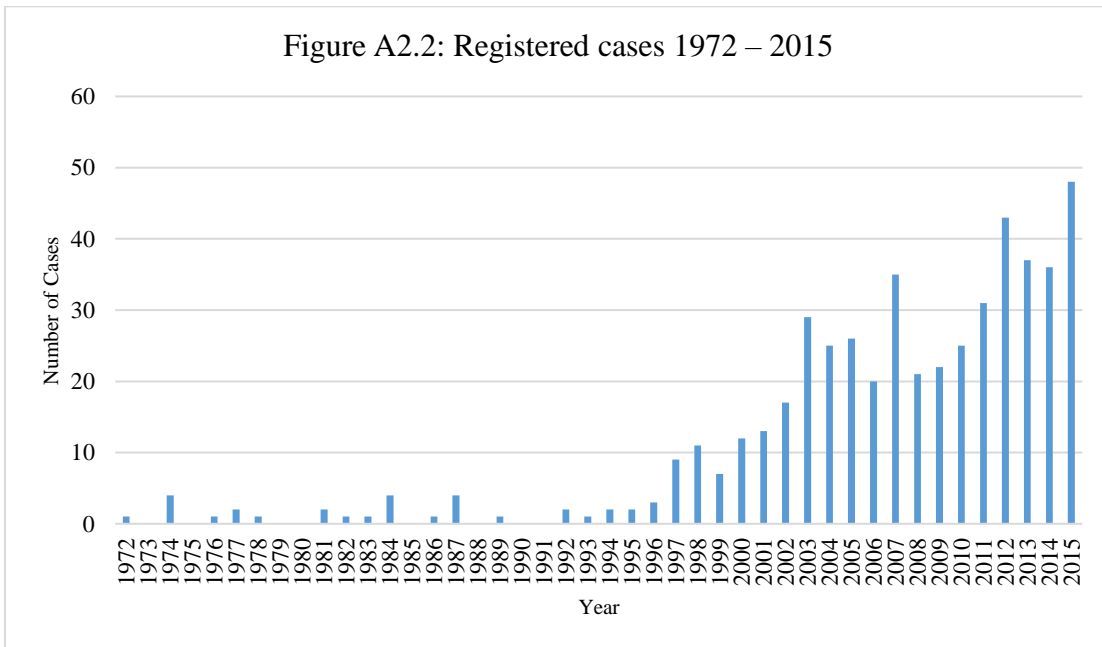
Figure A2.3: Number of repeat players at each frequency of appointment
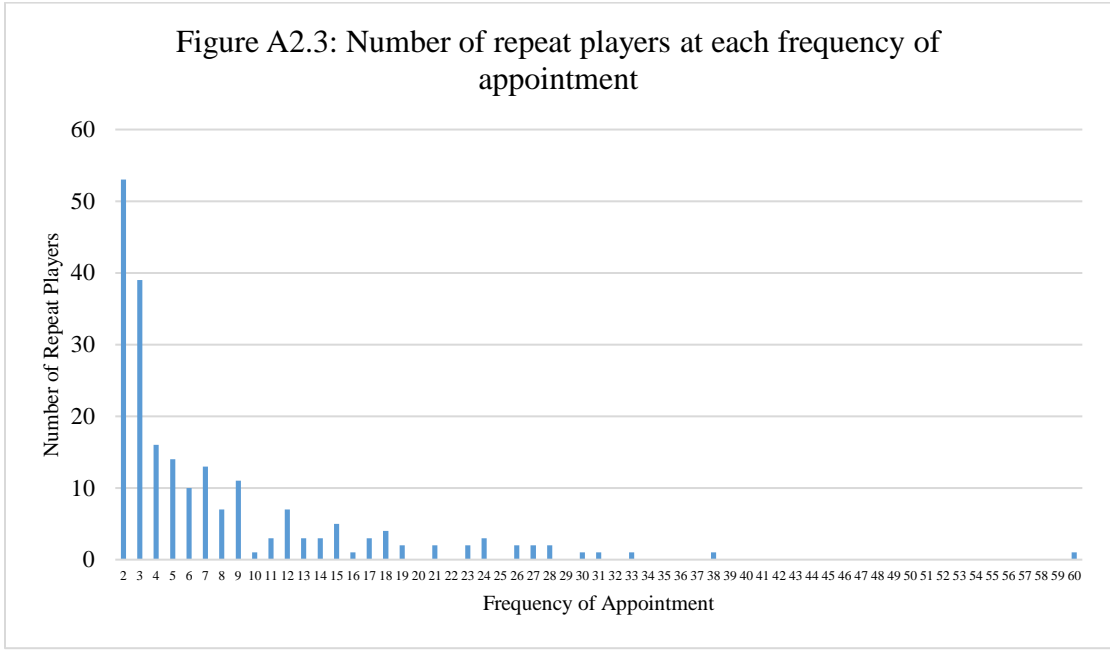
Figure A2.4: Distribution of individual arbitrator's average reputation
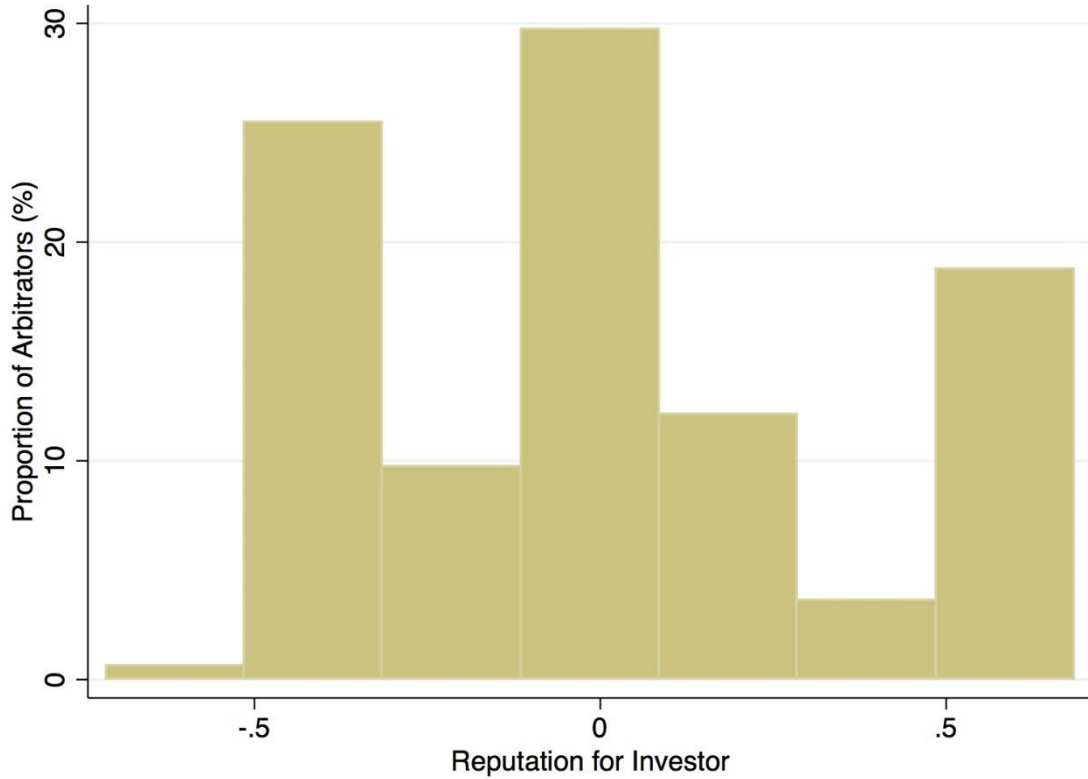
Note: The x-axis is the arbitrator's average *Reputation for Investor* based on all of her votes in the dataset, which runs from extremely pro-state on the far left to extremely pro-investor on the far right. The y-axis represents the proportion of arbitrators that have a particular *Reputation for Investor*.

Figure A2.5: Individual arbitrator's average vote rate
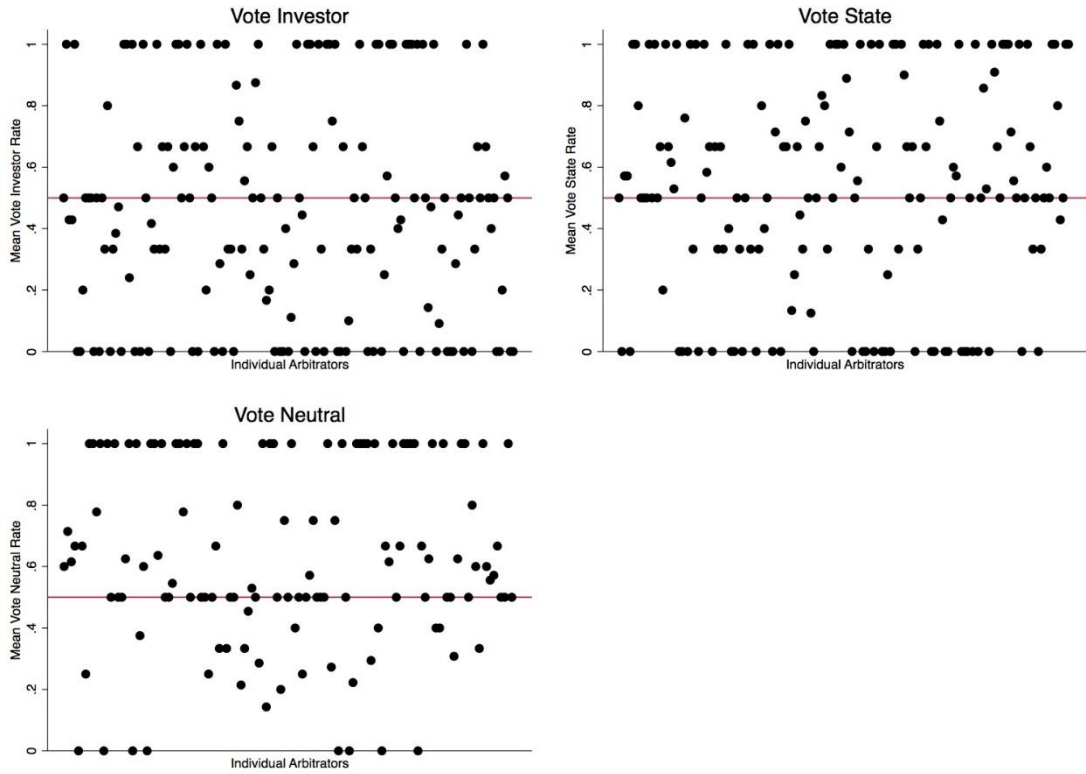
Figure A2.6: Individual arbitrator's average appointment rate
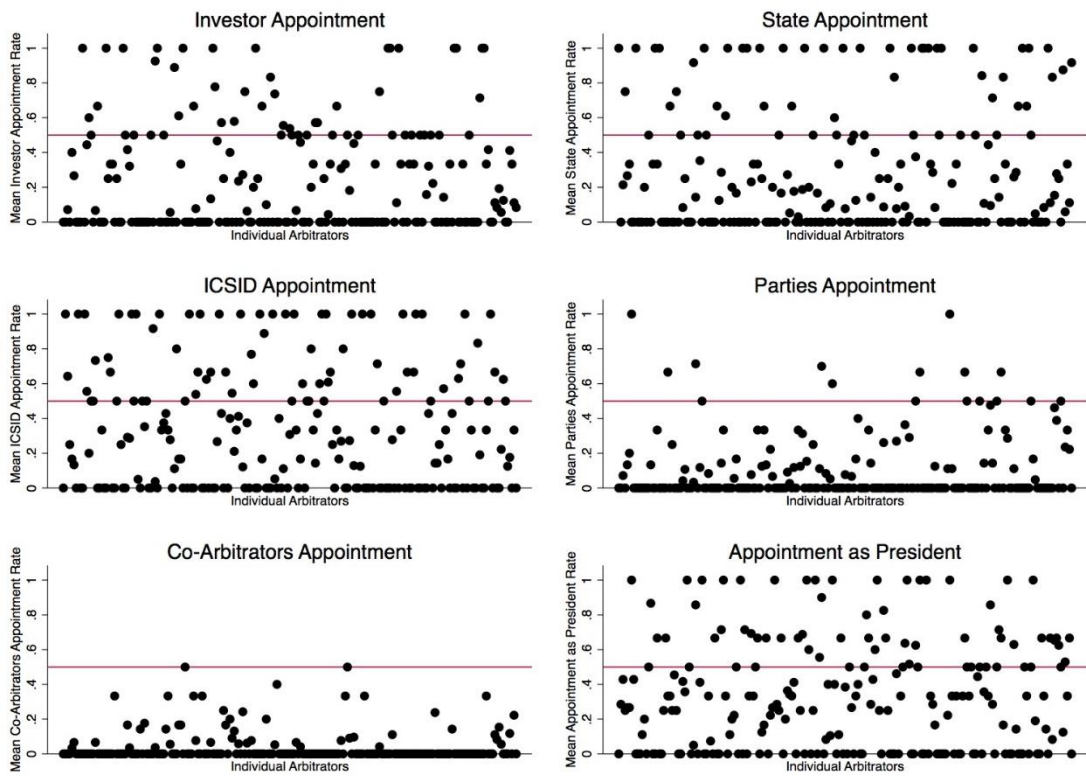
Figure A2.7: Distribution of fixed effects for Vote Investor
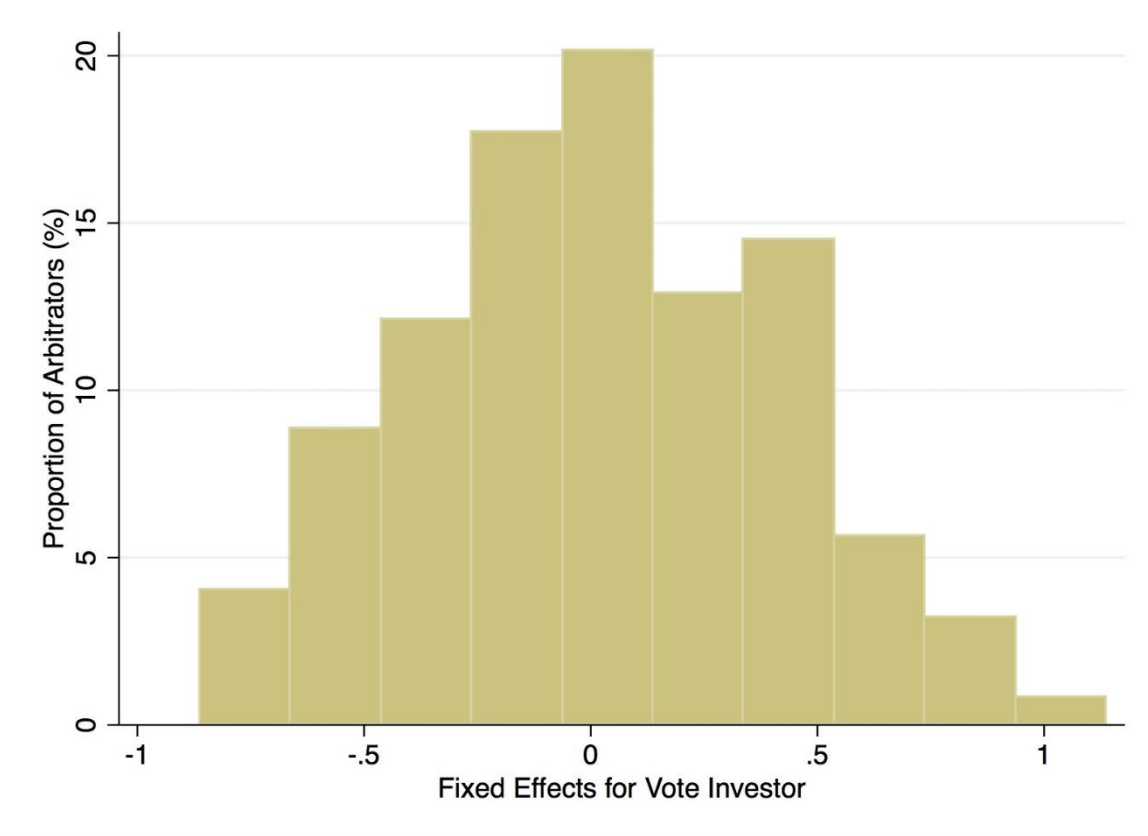
Note: The x-axis is the fixed effects for *Vote Investor*, which indicates an arbitrator's likelihood to vote for investors because of their pre-existing bias (or lack thereof). It runs from extremely unlikely to vote for investors on the far left to extremely likely to vote for investors on the far right. The y-axis represents the proportion of arbitrators with a particular likelihood.

Figure A2.8: Changes in reputation of twenty most-frequently appointed arbitrators
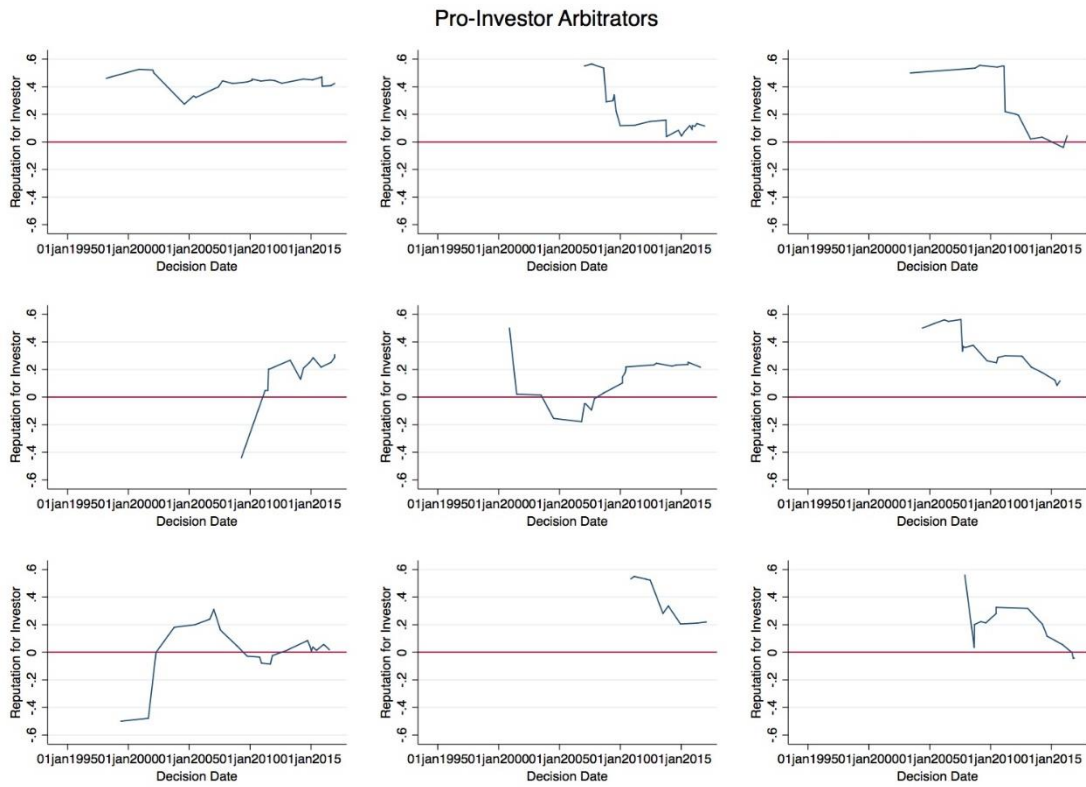
Pro-State Arbitrators
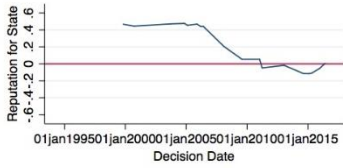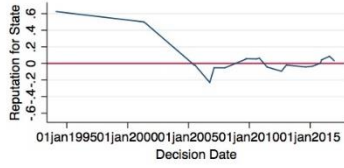
Figure A2.9.1: Distribution of placebo coefficients for Reputation for Investor after Dropping

Fewer than Three Times Players
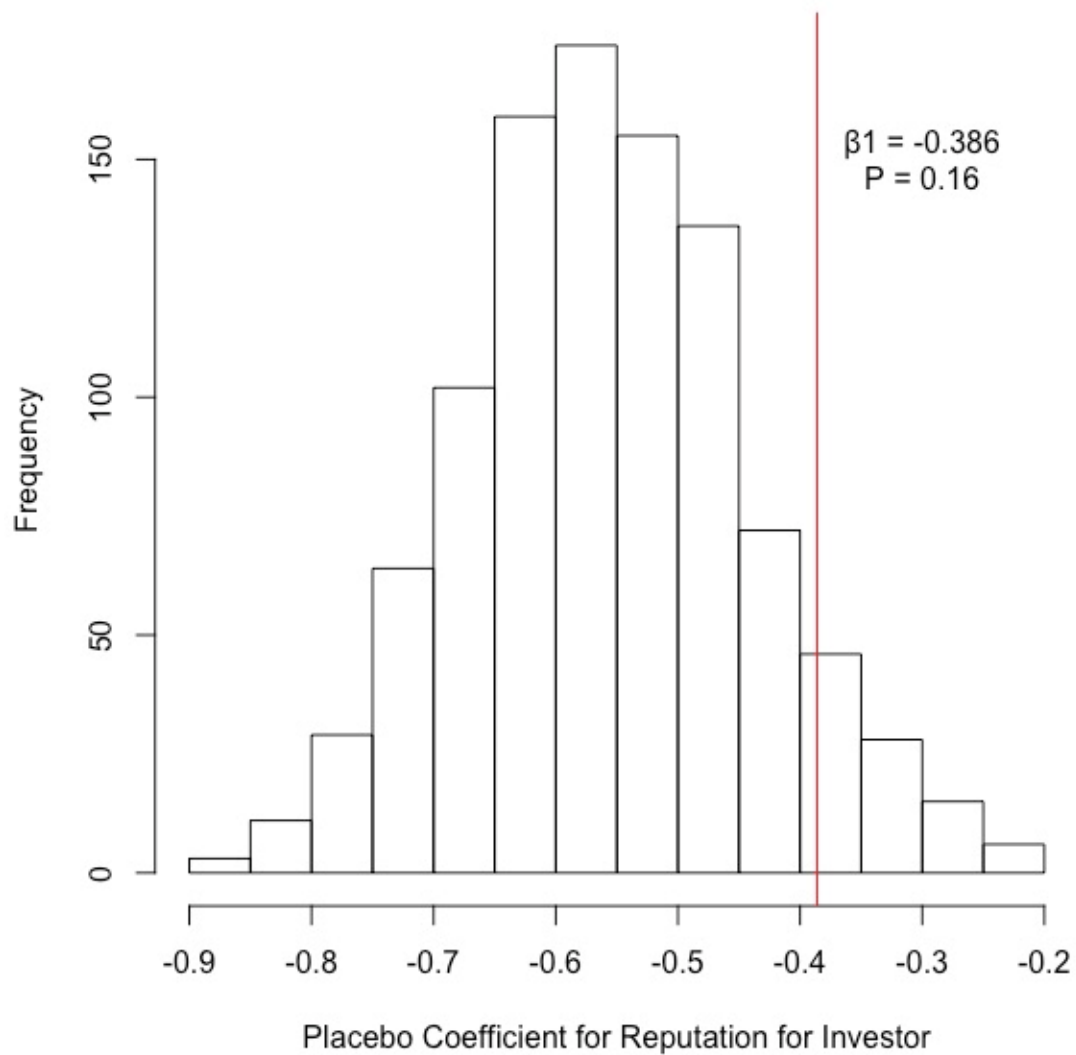
Figure A2.9.2: Distribution of placebo coefficients for Reputation for Neutrality after Dropping

Fewer than Three Times Players

$\beta1 = -1.505$
$P = 0.72$

Placebo Coefficient for Reputation for Neutrality
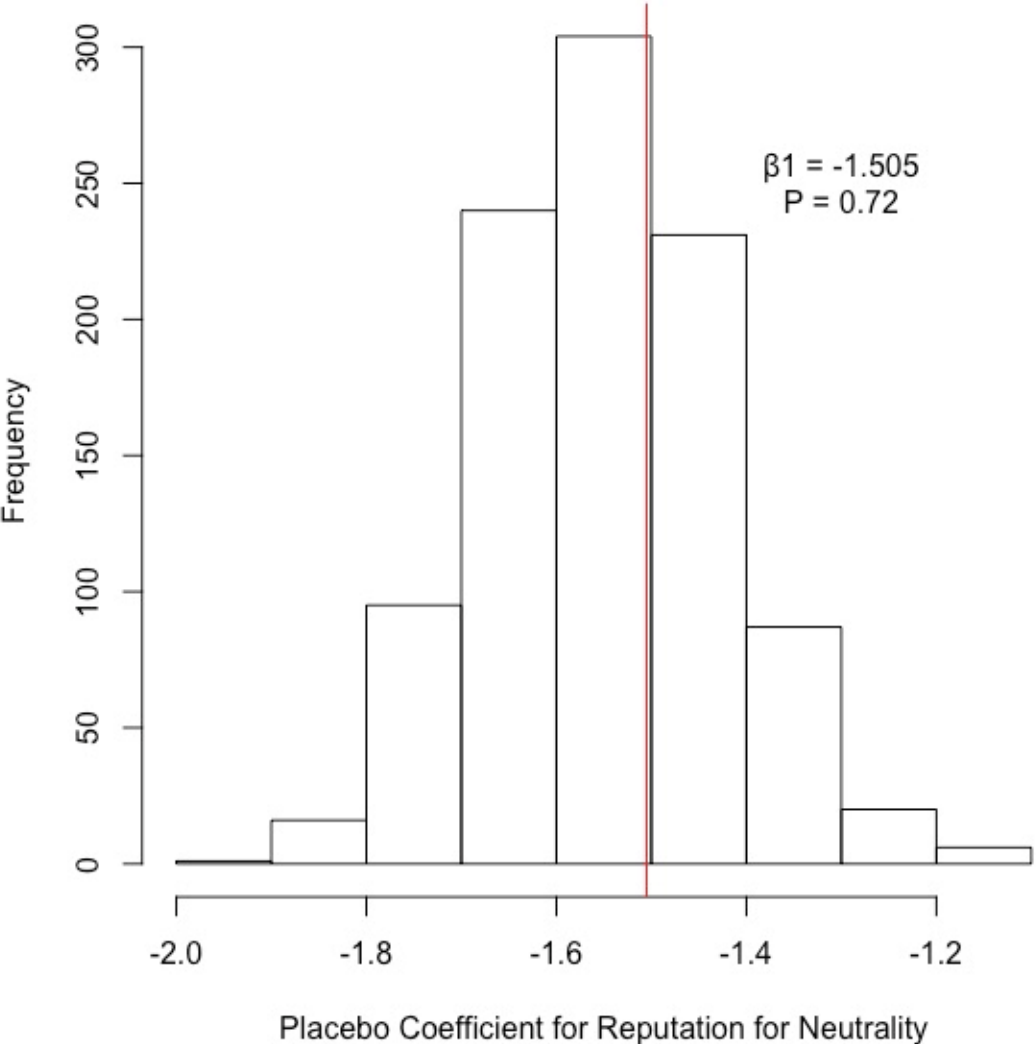
Figure A2.10.1: Distribution of placebo coefficients for Reputation for Investor after Dropping

Fewer than Five Times Players

Placebo Coefficient for Reputation for Investor
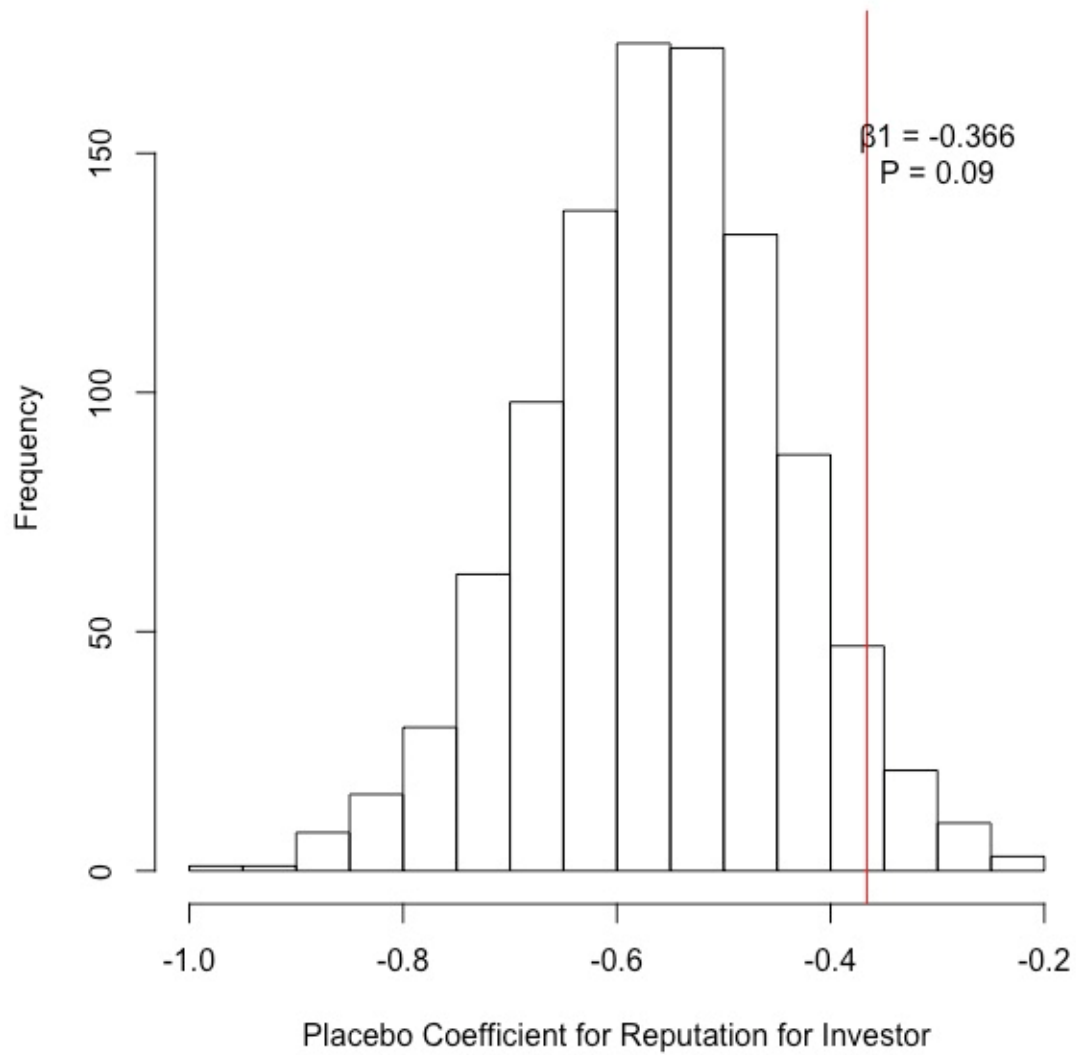
Figure A2.10.2: Distribution of placebo coefficients for Reputation for Neutrality after Dropping

Fewer than Five Times Players

Placebo Coefficient for Reputation for Neutrality
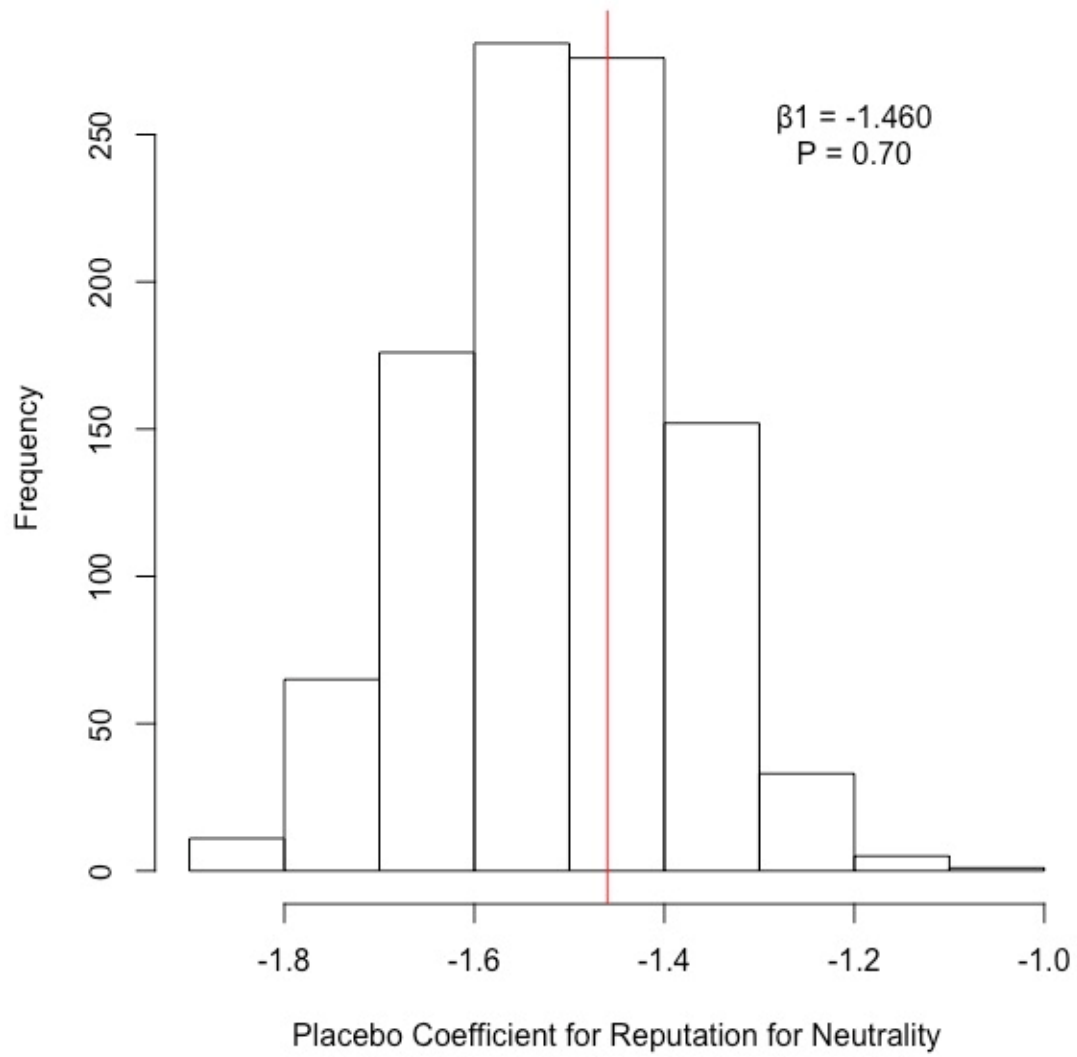
Table A2.1: Instrument Invoked[83]

Instrument Invoked	Frequency	%
BIT	343	68.88
Contract	78	15.66
Other Treaties	75	15.06
National Investment Law	44	8.84
Total	498	108.44[84]

Table A2.2: Case Fact[85]

Case Fact	Frequency	%
Termination or voiding of agreement or concession	58	14.01
Contract dispute	52	12.56
Occupation or requisition of property	45	10.87
Tax-related measures	45	10.87
Government inaction	39	9.42
Other administrative or legislative measures	29	7.00
Expropriate shares or nationalization	25	6.04
Tariff or other import or export measures	22	5.31
Deprival of license	19	4.59
Government interference	17	4.11
Debt default or restructuring	16	3.86
Court ruling	13	3.14
Financial measures	13	3.14
New restrictions or requirements related to existing agreement or concession	9	2.17
Government investigations or criminal proceedings	7	1.69
Price regulation	5	1.21
Total	414	100

---

[83] The information and categorization derive from ICSID's website. See International Centre for Settlement of Investment Disputes, https://icsid.worldbank.org/en/Pages/cases/AdvancedSearch.aspx. Two cases in the dataset do not have information about the instrument invoked in the case available online.

[84] Since in some cases more than one instrument are invoked, the total cumulative percentage is higher than 100%.

[85] I created these categories of case facts based on the description of facts in the publicly available awards and news reports.

Table A2.3: Economic Sector[86]

Economic Sector	Frequency	%
Oil, gas and mining	114	22.94
Electric power and other energy	85	17.10
Other	66	13.28
Transportation	48	9.66
Finance	37	7.44
Construction	36	7.24
Information and communication	31	6.24
Water, sanitation and flood protection	27	5.43
Agricultural, fishing and forestry	20	4.02
Tourism	18	3.62
Services and trade	15	3.02
Total	497	100.00

Table A2.4: Type of Proceeding[87]

Proceeding	Frequency	%
Ordinary Proceeding	455	76.21
Annulment Proceeding	91	15.24
Additional Facility	45	7.54
Resubmission Proceeding	6	1.01
Total	597	100.00

---

[86] The information and categorization derive from ICSID's website. *See* International Centre for Settlement of Investment Disputes, https://icsid.worldbank.org/en/Pages/cases/AdvancedSearch.aspx. Three cases in the dataset do not have information about the economic sector involved in the case available online.

[87] Some of the 500 cases in the dataset involve multiple proceedings. The information and categorization derive from ICSID's website. *See* International Centre for Settlement of Investment Disputes, https://icsid.worldbank.org/en/Pages/cases/AdvancedSearch.aspx. There are three other types of proceedings: the revision proceeding, the interpretation proceeding, the rectification or supplement proceeding in ICSID arbitration. Since none of these proceedings involve new arbitrator appointments or substantive rulings on the merits of the case, thereby provide no valuable information to the appointing entities, they are not included in the dataset.

Table A2.5: Time Trend of Reputation for Neutrality[88]

	(1) Reputation for Neutrality
Number of Cases	0.00889**
	(0.00196)
Instrument Dummies	Yes
Case Fact Dummies	Yes
Economic Sector Dummies	Yes
Proceeding Dummies	Yes
Arbitrator Fixed Effect	Yes
Observations	1343

Note: standard errors, in parentheses, are clustered at arbitrator level.
[†] Significant at the 10% level; [*] Significant at the 5% level; [**] Significant at the 1% level.

Table A2.6: Effects of Arbitrator's Reputation for State on Appointment by ICSID without Arbitrator-Specific Fixed Effect

	(1) ICSID Appointment
Reputation for State	0.00185
	(0.00159)
Instrument Dummies	Yes
Case Fact Dummies	Yes
Economic Sector Dummies	Yes
Proceeding Dummies	Yes
Arbitrator Fixed Effect	No
Observations	38563

Note: standard errors, in parentheses, are clustered at arbitrator level.
[†] Significant at the 10% level; [*] Significant at the 5% level; [**] Significant at the 1% level.

---

[88] Since the variable *Reputation for Neutrality* is a percentage between 0 and 1, its change relative to the change in the number of cases the arbitrator has decided ("1") would be small. Also, for cases that were discontinued before reaching the final decision stage or that are still pending, the arbitrator's *Reputation for Neutrality* will remain the same as she does not vote in that case. For these reasons, the coefficient for *Reputation for Neutrality* is expected to be small.

Table A2.7: The Relationship between Current Vote and Lagged Vote

	(1) Vote Investor
Vote Investor in the Last Case	-0.0284
	(0.0655)
Instrument Dummies	Yes
Case Fact Dummies	Yes
Economic Sector Dummies	Yes
Proceeding Dummies	Yes
Arbitrator Fixed Effect	Yes
Observations	377

Note: standard errors, in parentheses, are clustered at arbitrator level.
[†] Significant at the 10% level; [*] Significant at the 5% level; [**] Significant at the 1% level.

Table A2.8: Effects of Arbitrator's Reputation on Appointment without Arbitrator-Specific
Fixed Effects after Dropping Infrequent Players

	(1) Investor Appoint-ment	(2) State Appoint-ment	(3) ICSID Appoint-ment	(4) Parties Appoint-ment	(5) Co-arbitrators Appoint-ment	(6) Appoint-ment as President
Reputation for Investor	0.00766*					
	(0.00304)					
Reputation for State		0.00881**				
		(0.00336)				
Reputation for Neutrality			-0.00711	0.0101†	-0.00198	0.00168
			(0.00550)	(0.00609)	(0.00771)	(0.00364)
Instrument Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Case Fact Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Economic Sector Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Proceeding Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Arbitrator Fixed Effect	No	No	No	No	No	No
Observations	30077	29348	21139	14211	4479	40986

Note: standard errors, in parentheses, are clustered at arbitrator level.
† Significant at the 10% level; * Significant at the 5% level; ** Significant at the 1% level.

Table A2.9: Effects of Arbitrator's Reputation on Appointment with Arbitrator-Specific Fixed Effects after Dropping Infrequent Players

	(1) Investor Appointment	(2) State Appointment	(3) ICSID Appointment	(4) Parties Appointment	(5) Co-arbitrators Appointment	(6) Appointment as President
Reputation for Investor	-0.00669					
	(0.00484)					
Reputation for State		0.00685*				
		(0.00338)				
Reputation for Neutrality			-0.0249**	0.000101	-0.0163	-0.00900*
			(0.00916)	(0.00773)	(0.0162)	(0.00443)
Instrument Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Case Fact Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Economic Sector Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Proceeding Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Arbitrator Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30077	29348	21139	14211	4479	40986

Note: standard errors, in parentheses, are clustered at arbitrator level.
[†] Significant at the 10% level; [*] Significant at the 5% level; [**] Significant at the 1% level.

Table A2.10.1: Time Trend of Reputation for Neutrality after Dropping Fewer than Three Times Players

	(1) Reputation for Neutrality
Number of Cases	0.00889[**]
	(0.00196)
Instrument Dummies	Yes
Case Fact Dummies	Yes
Economic Sector Dummies	Yes
Proceeding Dummies	Yes
Arbitrator Fixed Effect	Yes
Observations	1212

Note: standard errors, in parentheses, are clustered at arbitrator level.
[†] Significant at the 10% level; [*] Significant at the 5% level; [**] Significant at the 1% level.

Table A2.10.2: Time Trend of Reputation for Neutrality after Dropping Fewer than Five Times Players

	(1) Reputation for Neutrality
Number of Cases	0.00879[**]
	(0.00193)
Instrument Dummies	Yes
Case Fact Dummies	Yes
Economic Sector Dummies	Yes
Proceeding Dummies	Yes
Arbitrator Fixed Effect	Yes
Observations	1122

Note: standard errors, in parentheses, are clustered at arbitrator level.
[†] Significant at the 10% level; [*] Significant at the 5% level; [**] Significant at the 1% level.

Table A2.11: Effects of Arbitrator's Reputation on Vote (Different Controls)

	(1)	(2)	(3)	(4)
Reputation for Investor (Vote Investor)	- 0.456**	- 0.436**	- 0.411**	- 0.409**
	(0.113)	(0.111)	(0.113)	(0.111)
Reputation for State (Vote State)	- 0.456**	- 0.436**	- 0.411**	- 0.409**
	(0.113)	(0.111)	(0.113)	(0.111)
Reputation for Neutrality (Vote Neutral)	-1.422**	- 1.426**	-1.457**	- 1.478**
	(0.124)	(0.126)	(0.127)	(0.119)
Instrument Dummies	No	Yes	Yes	Yes
Case Fact Dummies	No	No	Yes	Yes
Economic Sector Dummies	No	No	No	Yes
Proceeding Dummies	No	No	No	No
Arbitrator Fixed Effect	Yes	Yes	Yes	Yes
Observations	559	557	539	539

Note: The left-hand-side variable for each regression is listed in parentheses under the right-hand-side variable. Standard errors, in parentheses, are clustered at arbitrator level.
[†] Significant at the 10% level; [*] Significant at the 5% level; [**] Significant at the 1% level.

Table A2.12: Effects of Arbitrator's Reputation on Appointment without Arbitrator-Specific Fixed Effects (Different Controls)

	(1)	(2)	(3)	(4)
Reputation for Investor (Investor Appointment)	$0.00246^{\dagger}$	$0.00245^{\dagger}$	$0.00265^{\dagger}$	$0.00263^{\dagger}$
	(0.00136)	(0.00135)	(0.00144)	(0.00145)
Reputation for State (State Appointment)	0.00425^{**}	0.00425^{**}	0.00452^{**}	0.00452^{**}
	(0.00163)	(0.00164)	(0.00168)	(0.00168)
Reputation for Neutrality (ICSID Appointment)	0.00652	0.00654	0.00652	0.00648
	(0.00412)	(0.00412)	(0.00404)	(0.00405)
Reputation for Neutrality (Parties Appointment)	0.0127^{**}	0.0128^{**}	0.0137^{**}	0.0145^{**}
	(0.00430)	(0.00432)	(0.00478)	(0.00490)
Reputation for Neutrality (Co-arbitrators Appointment)	0.00848	0.00853	0.00687	0.00683
	(0.00516)	(0.00515)	(0.00524)	(0.00527)
Reputation for Neutrality (Appointment as President)	0.00742^{**}	0.00748^{**}	0.00794^{**}	0.00815^{**}
	(0.00265)	(0.00265)	(0.00287)	(0.00289)
Instrument Dummies	No	Yes	Yes	Yes
Case Fact Dummies	No	No	Yes	Yes
Economic Sector Dummies	No	No	No	Yes
Proceeding Dummies	No	No	No	No
Arbitrator Fixed Effect	No	No	No	No
Observations[89]	NA	NA	NA	NA

Note: The left-hand-side variable for each regression is listed in parentheses under the right-hand-side variable. Standard errors, in parentheses, are clustered at arbitrator level.
† Significant at the 10% level; * Significant at the 5% level; ** Significant at the 1% level.

---

[89] The number of observations is not listed here because it varies with respect to each left-side variable.

Table A2.13: Effects of Arbitrator's Reputation on Appointment with Arbitrator-Specific Fixed Effects (Different Controls)

	(1)	(2)	(3)	(4)
Reputation for Investor (Investor Appointment)	- 0.00313	- 0.00315	- 0.00588	- 0.00643
	(0.00448)	(0.00450)	(0.00481)	(0.00484)
Reputation for State (State Appointment)	0.00532	0.00536	0.00702*	0.00704*
	(0.00340)	(0.00341)	(0.00337)	(0.00335)
Reputation for Neutrality (ICSID Appointment)	- 0.0216*	- 0.0214*	- 0.0203*	- 0.0199*
	(0.00931)	(0.00934)	(0.00886)	(0.00883)
Reputation for Neutrality (Parties Appointment)	- 0.000723	- 0.000367	- 0.00199	- 0.000334
	(0.00647)	(0.00652)	(0.00726)	(0.00731)
Reputation for Neutrality (Co-arbitrators Appointment)	- 0.00848	- 0.00847	- 0.0155	- 0.0156
	(0.0116)	(0.0116)	(0.0155)	(0.0156)
Reputation for Neutrality (Appointment as President)	- 0.00920*	- 0.00903*	- 0.00930*	- 0.00874*
	(0.00408)	(0.00407)	(0.00433)	(0.00431)
Instrument Dummies	No	Yes	Yes	Yes
Case Fact Dummies	No	No	Yes	Yes
Economic Sector Dummies	No	No	No	Yes
Proceeding Dummies	No	No	No	No
Arbitrator Fixed Effect	Yes	Yes	Yes	Yes
Observations[90]	NA	NA	NA	NA

Note: The left-hand-side variable for each regression is listed in parentheses under the right-hand-side variable. Standard errors, in parentheses, are clustered at arbitrator level.
[†] Significant at the 10% level; [*] Significant at the 5% level; [**] Significant at the 1% level.

---

[90] The number of observations is not listed here because it varies with respect to each left-side variable.

Table A2.14: Effects of Appointment Rate by an Appointing entity on Arbitrators' Votes
(Different Controls)

	(1)	(2)	(3)	(4)
Investor Appointment Rate (Vote Investor)	- 0.107	- 0.0868	- 0.132	- 0.0583
	(0.216)	(0.197)	(0.224)	(0.203)
State Appointment Rate (Vote State)	0.177	0.137	0.0785	0.149
	(0.202)	(0.179)	(0.213)	(0.221)
ICSID Appointment Rate (Vote Neutral)	0.255	0.304	0.197	0.124
	(0.281)	(0.282)	(0.295)	(0.297)
Parties Appointment Rate (Vote Neutral)	0.394	0.369	0.435	0.495
	(0.319)	(0.325)	(0.359)	(0.327)
Co-arbitrators Appointment Rate (Vote Neutral)	- 0.980*	- 0.909*	- 0.860†	- 0.855†
	(0.415)	(0.419)	(0.469)	(0.513)
Appointment Rate as President (Vote Neutral)	0.224	0.249	0.296	0.295
	(0.274)	(0.270)	(0.270)	(0.261)
Instrument Dummies	No	Yes	Yes	Yes
Case Fact Dummies	No	No	Yes	Yes
Economic Sector Dummies	No	No	No	Yes
Proceeding Dummies	No	No	No	No
Arbitrator Fixed Effect	Yes	Yes	Yes	Yes
Observations[91]	NA	NA	NA	NA

Note: The left-hand-side variable for each regression is listed in parentheses under the right-hand-side variable. Standard errors, in parentheses, are clustered at arbitrator level.
† Significant at the 10% level; * Significant at the 5% level; ** Significant at the 1% level.

---

[91] The number of observations is not listed here because it varies with respect to each left-side variable.

DOMESTIC POLITICS AND SETTLEMENT IN INVESTOR-STATE ARBITRATION

I. INTRODUCTION

States are the participants of international dispute settlement. In areas ranging from international trade to international investment disputes, states' litigation strategies are often formulated and approved by elected officials and their subordinates. Due to the high profile and significant attention these international disputes usually receive, elected officials are likely to be affected by views from domestic audiences when making strategic decisions in these high-stakes disputes. State leaders may suffer "domestic audience costs" for backing down from such international disputes, as domestic audiences can penalize them for appearing weak internationally by voting in elections or through public opinion. The potential for such domestic backlash may constrain state leaders from pursuing the litigation strategies that would otherwise be most efficient for the state. Such impacts of domestic politics may lead to suboptimal case outcomes for state actors in international dispute settlement. This paper identifies causal effects of domestic politics on state litigation strategies in the context of settlement behavior in investor-state arbitration. The findings reveal significant inefficiencies arising from suboptimal decision-making due to domestic political influences.

Although no studies have examined the relationship between domestic politics and state behavior in investor-state dispute settlement (ISDS), the relationship in WTO dispute settlement has been examined. Busch (2000) found that highly democratic dyads are more likely than other dyads to settle WTO disputes through concessions at the consultation stage, but they are not more

likely to settle at the panel stage.[1] Busch posited that escalation from consultation to panel stage is more likely to generate large domestic "audience costs" due to backing down in democracies compared to non-democracies.[2] Chaudoin (2014) examined the relationship between WTO dispute initiation by plaintiff countries and domestic audience features in defendant countries. Chaudoin found that U.S. trading partners are more likely to initiate trade disputes against the United States during election years with lower unemployment, when U.S. domestic audiences support free trade and the government is most sensitive to those preferences.[3]

However, these existing studies do not attempt to identify causal effects of domestic politics on state litigation behavior. International relations scholars have long posited that domestic politics shape foreign policy decision making via audience costs: domestic constituents can penalize leaders who appear weak internationally by voting at elections or via public opinion.[4] Leaders' awareness of such audience costs in turn shapes their foreign policy incentives. In the context of international dispute settlement, domestic audience costs may arise when domestic constituents punish incumbent governments electorally or through public opinion. Settlement can make states look weak on the world stage, which may give rise to negative or disapproving reactions from domestic audiences. Anticipating such costs, governments may end up not settling international disputes.

---

[1] The consultation stage is a required first step in the GATT dispute settlement process. Article 4.7 of the Dispute Settlement Understanding provides that "[i]f the consultations fail to settle a dispute within 60 days after the date of receipt of the request for consultations, the complaining party may request the establishment of a panel. The complaining party may request a panel during the 60-day period if the consulting parties jointly consider that consultations have failed to settle the dispute."

[2] *See* Marc Busch, *Democracy, Consultation, and the Paneling of Disputes under GATT*, 44 J. CONFLICT. RESOL. 425 (2000).

[3] *See* Stephen Chaudoin, *Audience Features and the Strategic Timing of Trade Disputes*, 68 INT'L ORG 877 (2014).

[4] *See* James Fearon, *Domestic Political Audiences and the Escalation of International Disputes*, 88 AM. POL. SCI REV. 577 (1994).

Investor-state arbitration cases are particularly suitable for studying the interplay between domestic politics and international dispute settlement, due to the significant attention they have generated in recent years.[5] Under investor-state arbitration, only foreign investors may bring claims against host states, not vice versa. Thus, states are always respondents defending against claims from foreign investors, which typically amount to billions of dollars and can have far-reaching policy impacts for the host state in areas ranging from environmental policies to human rights protections.[6] Indeed, investor-state arbitration is increasingly being used to challenge domestic regulations.[7] As a result, host state constituents can be strongly antagonistic towards these claims, which may be perceived as a way for well-financed foreign investors, usually multinational corporations, to extort public money and erode state regulatory space.[8]

Settlement by paying out large sums or rescinding domestic regulations may further inflame public sentiment by causing controversial headlines portraying capitulation to large corporations, secret deals with foreign investors, misuse of public funds, and discrimination against domestic investors.[9] Thus, settlement in investor-state arbitration is likely to generate

---

[5] *See e.g.*, Meg Kinnear and Aissatou Diop, *Use of the Media by Counsel in Investor-State Arbitration*, in ALBERT JAN VAN DEN BERG ED., ARBITRATION ADVOCACY IN CHANGING TIMES 40 – 51 (2011); Ryan Brutger and Anton Strezhnev, *International Disputes, Media Coverage, and Backlash against International Law* (2018), *available at* https://www.internationalpoliticaleconomysociety.org/sites/default/files/conference_files/IPES_Proposal_2018_Brutger_Strezhnev_0.pdf.

[6] *See* HALEY EDWARDS, SHADOW COURTS: THE TRIBUNALS THAT RULE GLOBAL TRADE (2016).

[7] *See* Krzysztof Pelc, *What Explains the Low Success Rate of Investor-State Disputes?*, 71 INT'L ORG. 559 (2017).

[8] *See* Gus Van Harten and Pavel Malysheuski, *Who Has Benefited Financially from Investment Treaty Arbitration? An Evaluation of the Size and Wealth of Claimants* (2016), *available at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2713876.

[9] For negative media reports about settlement agreements in investor-state arbitration, see e.g. Chris Hamby, *The Secret Threat That Makes Corporations More Powerful Than Countries*, BUZZFEED NEWS (August 30, 2016), https://www.buzzfeednews.com/article/chrishamby/the-billion-dollar-ultimatum\#.sbbnblBGYP (criticizing governments for "giving in to the demands [of] corporations that have turned the threat of ISDS legal action into a fearsome weapon" and revealing that "the threatened governments — afraid of appearing weak or sparking a public backlash — are loath to admit they capitulated"); Mark Allix, *Economic sense 'must override BEE' in foreign investing*, BUSINESS DAY (November 8, 2013), https://www.businesslive.co.za/bd/companies/2013-11-08-economic-sense-must-override-bee-in-foreign-investing/ (commenting on a settlement agreement South Africa signed with Italian investors "No other mining company has been treated so generously"); Bette Hileman, *Canada Capitulates on MMT, Settles with Ethyl*, CHEMICAL AND ENGINEERING NEWS (July 27, 1998),

domestic audience costs for respondent state governments, which may in turn lead to a failure to settle in the first place.

To empirically test whether domestic audience costs affect case settlement in investor-state arbitration, the problem of endogeneity must be addressed. Domestic politics may be correlated with other state or case characteristics that could also affect settlement. To address this challenge, I employ a variable that is correlated with domestic audience costs but otherwise unrelated to settlement decisions: the share of term left until the next mandated election for the leader of the respondent state ("*Share of Term Left*"). When elections approach, domestic audience costs are magnified. Through voting in elections, domestic constituents can hold governments accountable for unfavorable settlement decisions.[10] As a result, as elections draw nearer, host state leaders face larger audience costs in the sense that politically they would suffer more from settling with foreign investors.[11] Previous studies on election timing showed that the public control over the leader is highest when an election is imminent, and government decisions made closer to elections are more likely to impact voters at the ballot box.[12] The closer the settlement timing is to the next election, the more political consequences the government is likely to suffer for settling with foreign investors. Thus, I use the time left until the next mandated election as a proxy for the size of anticipated domestic audience costs. Note that in parliamentary systems, although elections can be called early, it is still true that the closer it is to the next constitutionally mandated election, the

---

https://pubs.acs.org/doi/pdf/10.1021/cen-v076n030.p013a; Uchenna Awom And Patience Akpuru, *Nigeria: Malabu Deal Latest – Shell's Dirty Lies*, Leadership (May 24, 2012), https://allafrica.com/stories/201205240952.html.

[10] Although it is usually government employees in the legal department that are engaged in the negotiation of settlement agreements, they receive instructions from the department head who is usually a member of the government leader's cabinet. If the government leader has a general policy against settlement or activities that would generate bad press before elections, it is unlikely that their cabinet members will act against that policy.

[11] *See* Laron Williams, *Flexible Election Timing and International Conflicts*, 57 Int'l Stud. Q. 449 (2013).

[12] *See e.g.*, Kurt Gaubatz, *Election Cycles and War*, 35 J. Conflict Resol. 212 (1991); David Leblang and Steve Chan, *Explaining Wars Fought by Established Democracies: Do Institutional Constraints Matter?*, 56 Pol. Res. Q. 385 (2003); Dan Reiter and Tillman Erik, *Public, Legislative, and Executive Constraints on the Democratic Initiation of Conflict*, 64 J. Pol. 810 (2002).

more likely an election will be called, as both removal costs and expected benefits of staying in office decrease over time.[13] Thus, even for parliamentary democracies with flexible election timing, *Share of Term Left* can still predict the anticipated timing of an election, and thus influences the expected opportunity to hold the government accountable and thereby the size of anticipated domestic audience costs. On the other hand, since a country's election term is usually stipulated in its constitution, *Share of Term Left* should be exogenous to case quality and related to case settlement only through its impact on anticipated domestic audience costs.

Using a dataset of all publicly available investor-state arbitration cases registered before 2016, I find that, consistent with my prediction, the probability of settlement decreases when there is less time left until the next constitutionally mandated election for the respondent state leaders. The finding suggests that domestic politics of respondent states affect case settlement in investor-state arbitration. As elections approach, the increase in anticipated domestic audience costs in respondent states decreases settlement probability in investor-state disputes.

I also investigate whether domestic audience costs affect settlement decisions of democracies and non-democracies similarly. I find that while less democratic respondent states tend to settle more often, the higher settlement rate in less democratic respondent states is largely attributed to those non-democratic respondent states that do not hold regular elections. The regression results further suggest that overall, non-democracies that hold elections are not significantly less affected by anticipated audience costs than democracies when it comes to case settlement in investor-state arbitration.

---

[13] *See* Daniel Diermeier and Randolph Stevenson, *Cabinet Terminations and Critical Events*, 94 AM. P. SCI. REV. 627 (2000); Arthur Lupia and Kaare Strom, *Coalition Termination and the Strategic Timing of Parliamentary Elections*, 89 AM. P. SCI. REV. 648 (1995); Michael Koch, *Governments, Partisanship, and Foreign Policy: The Case of Dispute Duration*, 46 J. PEACE RES. 799 (2009).

The findings in this paper lead to some broader implications. If states base their settlement offers solely on case merits, we would expect to observe no relationship between proximity to the next constitutionally mandated election and settlement probability. The patterns identified in this study indicate that due to the influence of domestic political considerations, elected officials in respondent states are not making settlement decisions in ways that would be financially optimal for the state. The decrease in settlement probability caused by proximity to the next constitutionally mandated election may be explained by two mechanisms: due to the influence of domestic audience costs, first, some cases that should have been settled earlier may be settled only after elections have been held; and, second, cases that should have been settled based on their merits may not be settled at all.

Either delay or failure to settle could result in excessive litigation expenses and considerable uncertainty for both the claimant investor and other potential investors contemplating investing in the respondent state. According to the United Nations Commission on International Trade Law (UNCITRAL), the average litigation cost per case for claimant investors was $6,019,000.[14] The average litigation cost per case for respondent states was $4,855,000.[15] Such costs could be especially burdensome for developing states with limited financial resources, who are usually the target of investment claims. Without a settlement delay or failure, the amount of public funding spent in litigating investor-state disputes would be significantly reduced. In addition, ongoing arbitration proceedings, which last on average for three to four years, create prolonged uncertainty for claimant investors with respect to the value of their investments. Other potential investors may also be deterred because of the lack of clarity over the political and economic risks associated with

---

[14] *See* UNCITRAL Working Group III, *Possible Reform of Investor-State Dispute Settlement (ISDS) – Cost And Duration* (August 31, 2018), *available at* https://undocs.org/en/A/CN.9/WG.III/WP.153.
[15] *See id.*

investing in the respondent state, which may also deter investment in the respondent state long term.[16]

The findings also indicate that due to domestic political influences, settlement may not be completely effective in filtering out cases with clear outcomes. hence, clear cut cases which would otherwise typically be settled may be decided by arbitrators, leading to the creation of additional precedents that will be relied upon by future tribunals.[17] Precedents where the claimant investor will clearly prevail can negatively affect respondent states, because they are usually the repeat players in these disputes. Further, assuming that arbitrators are fallible (or biased), wrong decisions made in cases with clear outcomes (which should have been settled) are likely to generate more dissatisfaction and distrust in the dispute settlement system. Finally, most investor-state arbitration cases are decided by a small number of stellar arbitrators, who are repeatedly appointed by disputing parties.[18] The surplusage of arbitrated cases resulting from the ineffective filter of settlement may add to their caseload, and further prolong the lengths of proceedings. Partly for this reason, investor-state arbitration, which was designed to be a cheap and efficient dispute settlement mechanism to the benefit of both investors and states, has become more costly and lengthy a system than intended that is subject to intense scrutiny and criticism.

The remainder of this paper proceeds as follows: Section 2 elaborates on the relationship between domestic audience costs and settlement in investor-state arbitration. Section 3 describes

---

[16] *See* Todd Allee and Clint Peinhardt, *Contingent Credibility: The Impact of Investment Treaty Violations on Foreign Direct Investment*, 65 INT'L ORG. 401 (2011).

[17] While there is no doctrine of *stare decisis* or binding precedents in investor-state arbitration, the precedential value of arbitration decisions have been widely acknowledged, see e.g. Susan Franck, *The Legitimacy Crisis in Investment Treaty Arbitration: Privatizing Public International Law through Inconsistent Decisions*, 73 FORDHAM L. REV. 1521 (2004); Jeffrey Commission, *Precedent in Investment Treaty Arbitration – A Citation Analysis of a Developing Jurisprudence*, 24 J. INT'L ARB. 129 (2007).

[18] *See* Malcolm Langford, Daniel Behn, and Runar Hilleren Lie, *The Revolving Door in International Investment Arbitration*, 20 J. INT'L ECON. L. 301 (2017).

the data; Section 4 introduces the research design and presents the empirical results; and Section 5 provides concluding remarks.

II. DOMESTIC AUDIENCE COSTS AND SETTLEMENT IN INVESTOR-STATE ARBITRATION

In this section, I discuss the relationship between domestic audience costs and settlement in investor- state arbitration. I first elaborate on why settlement in investor-state arbitration cases tends to generate domestic audience costs for respondent state leaders. I then elaborate upon why such costs to the respondent state alone can lead to a failure to settle, when these costs are not expected to affect the decision of investors. Finally, I explain why proximity to the next constitutionally mandated election can be used as a proxy for the size of domestic audience costs to respondent states.

ISDS cases have made headlines in major news outlets, such as the New York Times, the Economist, and the Guardian.[19] These news periodicals criticized ISDS for allowing profit-driven corporations to bring excessive claims against sovereign states, often for policy decisions that fall within state regulatory space. Due to the high stakes and strong domestic hostility to these claims, settlement in these investor-state arbitration cases is susceptible to public criticism. Settlement can make governments look weak and corrupt, and thus generate domestic audience costs for respondent state leaders. Indeed, as a leading practitioners in the field has commented, "a lot of what goes into the respondent state government's calculus is political – how is it going to be read to the public that we are settling, which generally means we are paying a significant amount, and

---

[19] *See* Manuel Pérez-Rocha, *When Corporations Sue Government*, NEW YORK TIMES (December 3, 2014), https://www.nytimes.com/2014/12/04/opinion/when-corporations-sue-governments.html; *Investor-State Dispute Settlement, The Arbitration Game*, THE ECONOMIST (October 11, 2014), https://www.economist.com/finance-and-economics/2014/10/11/the-arbitration-game; Claire Provost and Matt Kennard, *The Obscure Legal System That Lets Corporations Sue Countries*, THE GUARDIAN (June 10, 2015), https://www.theguardian.com/business/2015/jun/10/obscure-legal-system-lets-corportations-sue-states-ttip-icsid.

how can that be defensible. So I would say in most instances, you rarely settle cases with states."[20] Because state leaders make decisions in the shadow of potential punishment from domestic audiences, the presence of such audience costs can lead to a failure to settle in investor-state arbitration cases.

One theoretical issue that arises is why domestic audience costs for respondent states alone should affect settlement outcome, when the investor's acquiescence is also a prerequisite to settlement. One might expect that only the settlement amount would change as a result of domestic audience costs, not the settlement rate itself. However, the settlement costs relative to litigation costs have changed as a result of domestic audience costs. When one of the parties is penalized for settling cases, this penalty can drive a wedge between the parties, and lead to a failure to settle.[21]

Empirical analysis on the effects of domestic audience costs on case settlement presents a challenge, because domestic political factors can be correlated with other, unobserved state and case characteristics that may also affect settlements. To address this challenge, I employ a variable *Share of Term Left*, which denotes proximity to the next constitutionally mandated election in the respondent state. *Share of Term Left* is a proxy for the size of domestic audience costs because as elections approach, electoral disapproval is more likely to translate into loss of political power. The use of election timing to measure public constraints over the executive is well established in

---

[20] Carolyn Lamm, *A State's Many Roles in International Adjudication*, The 2019 American Society of International Law Annual Meeting. The audio track of the panel discussion is available at https://www.asil.org/events/2019-asil-annual-meeting.

[21] The necessary condition for settlement is that plaintiff's minimum settlement demand is smaller than defendant's maximum settlement offer. In a simple litigation and settlement model (divergent expectations or asymmetric information model), plaintiff's minimum settlement demand is $P_p J - C_p + S_p$, defendant's maximum settlement offer equals $P_d J + C_d - S_d$, where J is the damages that defendant pays plaintiff if plaintiff prevails, C_p and C_d are litigation costs to plaintiff and defendant, respectively, and S_p and S_d are the respective settlement costs. P_p and P_d are plaintiff's and defendant's estimates of the probability that plaintiff will prevail, respectively. The settlement condition can be rewritten as $P_p - P_d < \frac{C-S}{J}$, where $C = C_p + C_d$ and $S = S_p + S_d$. The domestic audience costs constitute a separate settlement costs to the respondent state alone. That is, domestic audience costs add to S_d, which increases S and thus can make $\frac{C-S}{J}$ smaller than $P_p - P_d$. As a result, when domestic audience costs are large enough, the settlement condition, $P_P - P_d < \frac{C-S}{J}$, will no longer hold, and settlement will fail.

international relations literature.[22] Governments are generally considered more sensitive to public opinion and more constrained by domestic constituents as elections approach. On the other hand, *Share of Term Left* should be exogenous to case quality because the term length is largely predetermined by the Constitution. In Section 4.3, I rule out potential confounds and show that *Share of Term Left* affects settlement probability only through its impacts on domestic audience costs. Due to the exogeneity of *Share of Term Left*, the research design is similar to a natural experiment, which allows me to draw causal inferences about the effects of domestic audience costs on state settlement behavior.

III. DATA

I compiled a dataset that covers all publicly available investor-state arbitration cases registered before January 1, 2016, collected from the ICSID website, UNCTAD's Investment Dispute Settlement Navigator and the ITALaw website. The data covers case procedural history and outcome information of 801 investor-state arbitration cases involving 129 respondent states. The dataset is adjusted so that the unit of analysis is case-year: each case includes an annual observation from the year of registration until the year of conclusion. If a case is still pending by 2017, the annual observation ends in 2017.

The dependent variable, *Settle*, is a dummy variable that equals one if the case was settled in year t, and zero otherwise. A case is coded as settled if the arbitral proceeding is discontinued due to settlement.[23] Settlement can occur at any stage throughout the arbitration proceeding before

---

[22] *See* Gaubatz, *supra* note 12; Leblang and Chan, *supra* note 12; Williams, *supra* note 11.

[23] For ICSID arbitration cases, a case is coded as settled if the arbitral proceeding is discontinued according to Rule 43 or Rule 44 of the ICSID Arbitration Rules. For non-ICSID arbitration cases, a case is coded as settled if its outcome is "settled" on UNCTAD's Investment Dispute Settlement Navigator and there exists document in support of the settlement status.

a final award is rendered.[24] Of the 801 cases in the dataset, 147 cases have been settled (Table A3.1). The settlement rate is 18.35%. Figure 3.1 plots the number of settled cases relative to the number of cases that are concluded without settlement each year.[25] Before 2000, very few cases were decided each year. As the number of cases decided each year started to grow dramatically after 2000, the portion of decided cases that ended in settlement remains at a relatively low level.

Figure 3.1: Case Outcome by Decision Year
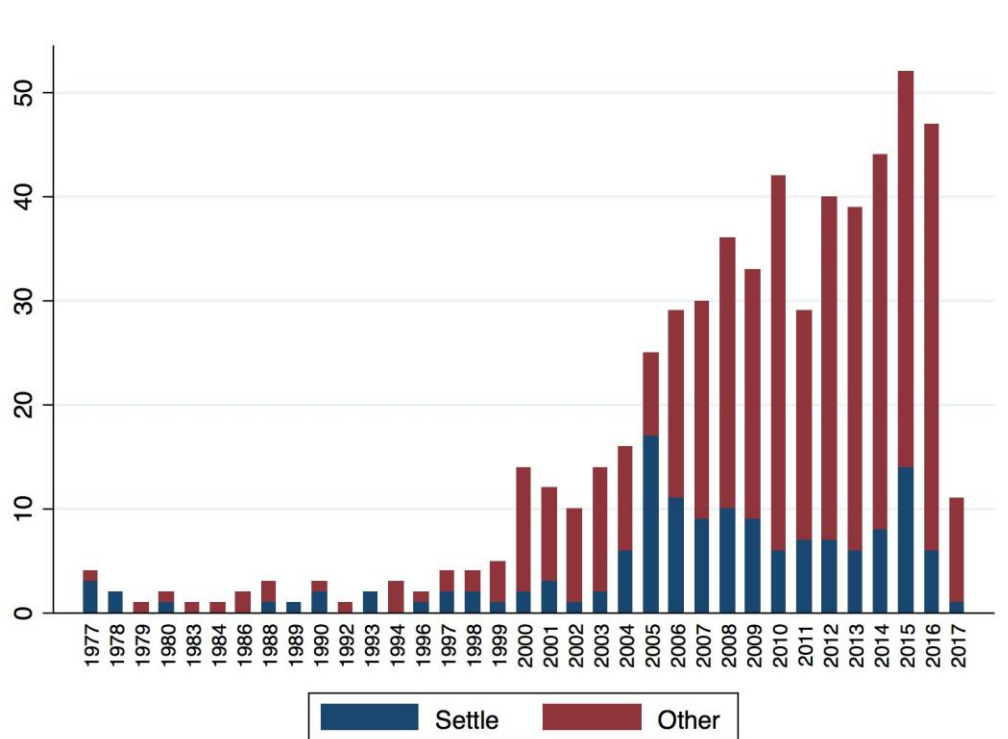
I then merged this dataset with the Database of Political Institutions 2017,[26] which has recorded the years left in the constitutional term length for the leaders of 180 countries for each

---

[24] Note that here we only consider settlements occurred after the registration of a case, which our dataset has information on. Settlements that occurred before case registration are less likely to be affected by electoral accountability concerns as most of them are confidential and not subject to public scrutiny.

[25] There were fewer cases in 2017 because I recorded the status of case outcomes by March 2017.

[26] The Database of Political Institutions 2017 is available at https://publications.iadb.org/handle/11319/8806.

year between 1975 and 2017 (*Years Left in Current Term*).[27] Note that I made one change to this variable. In the original dataset, the variable is coded "0" in an election year. This coding method is consistent with the actual number of years left in the constitutional term length in countries with fixed election timing (i.e. presidential systems). But in countries with flexible election timing (i.e. parliamentary systems), when an election is called early, this coding method would not represent the actual number of years left in the constitutional term length (the maximum number of years the leader can serve before new elections must be called). This might in turn cause endogeneity problems. To address this issue, I adjusted this variable so that in an election year when the election has been called early, *Years Left in Current Term* still equals the number of years left in the constitutional term length for the leader. For instance, if in the respondent state an election has been called in year t, which is two years prior to the next constitutionally mandated election, then for that case in year t, *Years Left in Current Term* is coded "2" instead of "0". In this way, *Years Left in Current Term* is *de jure* exogenous to any government influence that could affect settlement decisions.

The independent variable of interest, *Share of Term Left* is the share of constitutional term left until the next mandated election for the leader of the respondent state in year t. Since the Database of Political Institutions 2017 does not include constitutional term length information, I collected this information from the Comparative Constitution Project Data,[28] the Political Handbook of the World Online Edition,[29] and the Parline Database.[30] I confirmed *Years Left in Current Term* against each respondent state's constitutional term length in year t and generated

---

[27] According to the codebook of the Database of Political Institutions 2017, *Years Left in Current Term* is coded as "0" in an election year, and n-1 in the year after an election, where n is the length of the term. If a country is in the midst of civil war or political crisis, *Years Left in Current Term* is coded as "not applicable", in which case the observation is dropped from the regression analysis.
[28] The Comparative Constitution Project Data is available at http://comparativeconstitutionsproject.org.
[29] The Political Handbook of the World Online Edition is available at http://library.cqpress.com.
[30] The Parline Database is available at http://archive.ipu.org/parline-e/parlinesearch.asp.

Share of Term Left based on these two variables. For example, for a case where the respondent state's leader has three years left in a four-year term, *Share of Term Left* equals .75, which indicates that three quarters of the constitutional term remain for the leader. Note that I do not use *Years Left in Current Term* as the main independent variable of interest because *Years Left in Current Term* may not be directly comparable across countries which have different constitutional term lengths. However, the results do not change when using *Years Left in Current Term* (Table A3.2).

In the main analysis, I also controlled for respondent states' democracy level, GDP per capita (Logged), and foreign direct investment as percentage of GDP, which may affect respondent states' settlement likelihood. Data for these control variables is provided by the Polity IV index,[31] World Bank Open Data on GDP per capita,[32] and World Bank Open Data on FDI net inflows,[33] respectively. I also controlled for the number of claims, i.e. investment protection standard violations investor claimants have brought against respondent states in a case. I expect that the more claims the two parties have to agree on, the less likely a settlement agreement will be reached.

Table 3.1 presents summary statistics regarding case duration, constitutional term length for the leader of the respondent state, the respondent state's Polity Score, GDP per capita, FDI relative to GDP, and the number of claims brought by the investor. On average, arbitration proceedings last for 3.87 years, which is shorter than the average constitutional term length in respondent states (4.84 years). The median Polity Score of respondent states is 8, which indicates that a majority of respondent states are democracies. The average GDP per capita of respondent states is $10,905 in 2005 USD. This is lower than the threshold for a country with a developed

---

[31] The Polity IV index dataset is available at http://www.systemicpeace.org/ polity/polity4.htm.
[32] World Bank Open Data on GDP per capita is available at
https://data.worldbank.org/indicator/NY.GDP.PCAP.KD.
[33] World Bank Open Data on FDI net inflows is available at
https://data.worldbank.org/indicator/BX.KLT.DINV.WD.GD.ZS.

economy. The median GDP per capita is even lower, showing that developing countries constitute

a majority of respondent states. On average, FDI net inflows (new investment inflows less

disinvestment) from foreign investors constitute 3.87% of the respondent state's GDP. The number

of claims brought against respondent states per case ranges from 1 to 7, with an average of 2.5

claims per case.

Table 3.1: Summary Statistics, Continuous Variable

Variable	Mean	Median	Max.	Min.	N
Case Duration (Year)	3.87	4	18	0	801
Constitutional Term Length (Year)	4.84	5	8	2	768
Polity Score	4.72	8	10	-10	788
GDP per capita (in 2005 USD)	10,905	7,352	60,435	218	784
FDI (% of GDP)	3.87	2.65	55.49	-10.66	774
Number of Claims	2.49	2	7	1	697

Of the 129 countries that have a respondent history in the dataset, 35 have only been named

a respondent once. 21 countries have appeared as respondent state more than ten times. The three

most frequent players are Argentina, Venezuela and the Czech Republic, which have had 60, 44

and 33 cases before investment arbitration tribunals, respectively (Figure A3.1). In terms of

settlement, 62 of the 129 countries have never settled a case. Five countries have settled cases

more than five times: Argentina (14 times), Bolivia (9 times), India (9 times), Egypt (7 times) and

Venezuela (6 times) (Figure A3.2).

The constraints an upcoming election imposes on leaders may vary across different regimes.

To examine whether the effects of domestic audience costs on settlement decisions differ between

democracies and non-democracies, I also take into account the respondent state's regime type,

which is based on the categorization from Cheibub, Gandhi and Vreeland (2010).[34] They divide

---

[34] The information on respondent state's regime type is collected from the Democracy and Dictatorship Database and the Database on Democratic Electoral System. The Database on Democratic Electoral System has updated the

countries into six regime types based on their form of government and power structure: parliamentary democracy, semi-presidential democracy, presidential democracy, civilian dictatorship, military dictatorship, and royal dictatorship. Table A3.3 provides summary statistics on respondent state regime types. 72.5 % respondent states in the 801 cases are democracies. The remaining 27.5% respondent states are dictatorships. Royal dictatorship, the only regime type in the dataset where no election has been held, constitutes only 2% of respondent states.

IV. RESULTS

A. Domestic Audience Costs and Settlement

Because respondent states will suffer higher domestic audience costs when elections are closer, I expect to observe lower settlement probability when elections approach in respondent states.

I begin with a graphical presentation before turning to regression results. The first graph in Figure 3.2 presents the distribution of case settlement based on the number of months relative to the nearest mandated election. The second graph in Figure 3.2 presents the distribution of case settlement based on the quartiles of the constitutional term length relative to the nearest mandated election. The line represent the time of the nearest mandated election. The bars on the left of the line represent cases settled prior to the nearest mandated election. The bars on the right represent cases settled after the nearest mandated election. In both graphs, we observe a sharp increase in settlements after the nearest mandated election. More than 60% of settlements occur in the first half of the election cycle. We also observe a moderate decrease in settlements in the period immediately prior to the nearest mandated election. The two graphs in Figure 3.2 provide visual

---

regime variable through 2016. For detailed description about the categorization of country regime type, see Jose Cheibub, Jennifer Gandhi and James Vreeland, *Democracy and Dictatorship Revisited*, 143 PUB. CHOICE 67 (2010).

evidence that settlements are more likely to occur immediately following an election rather than before an election.

Figure 3.2: Settlement Timing and The Nearest Mandated Election
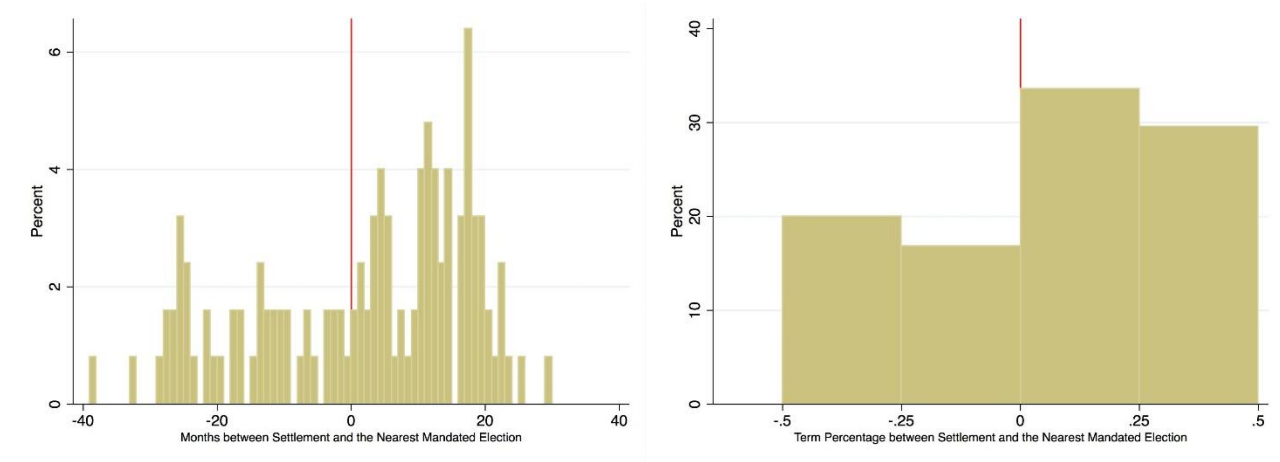
I now turn to regression analysis. Given the nature of my hypothesis which is about duration before settlement, I employ a Cox proportional hazards model to evaluate the relationship between Share of Term Left and the duration of cases before they are settled. This type of survival analysis estimates the "risk" that an event, in our case settlement, will take place as time elapses. This approach has the advantage of being able to estimate the effects of the explanatory variables on the risk of settlement, while leaving the underlying, or baseline risk, of settlement during year t, h(t), unspecified. The empirical specification is as follows:

$$h_c(t) = h_0(t)e^{\beta_1 Share\ of\ Term\ Left_c + \delta X_c + \theta_i + \varepsilon} \tag{1}$$

In this model, $h_c(t)$ represents the probability of case c being settled conditional on having continued until year t. $h_0(t)$ models the baseline hazard of settlement. X_c is a vector of control variables that include the Polity Score of the respondent state in case c in year t (Polity Score), the log of GDP per capita of the respondent state in case c in 2005 USD in year t (GDP per capita

113

(Logged)), the amount of FDI net inflows relative to the GDP of the respondent state in case c in year t (FDI (% of GDP)), and the number of claims the claimant investor brought against the respondent state in case c (Number of Claims). θ_i is a country-specific frailty parameter, which accounts for country-specific heterogeneity that might affect the hazard rate. This is equivalent to estimating a model with country-specific random effects in a more standard time-series cross-sectional framework.[35] A positive coefficient β_1 would imply a reduction in the hazard $h(t)$ of settlement as less time is left in the constitutional term length for the leader of the respondent state.

Table 3.2 reports the regression results using all the data. I add one more control variable in each column. To account for possible calendar year trends, in column (6), I include a counter variable that begins at 1 for the first calendar year in the dataset. In all specifications, I control for the country-specific frailty parameter. The results support the prediction that settlement risks are lower when elections draw near. In all specifications, increased proximity to the next constitutionally mandated election in the respondent state lowers the risk of settlement. The coefficients for *Share of Term Left* are positive and statistically significant across specifications. The coefficient magnitudes are quite large. The decrease in settlement likelihood for respondent states that are one-quarter-term closer to the next constitutionally mandated election is approximately 32.5%. The results hold when accounting for the respondent state's Polity Score, GDP per capita, amount of FDI relative to GDP, claims number and the calendar year trend.

The results provide support for the theoretical prediction that settlement behavior in investor-state arbitration is affected by domestic political calculations of the respondent state. As domestic constituents have more opportunities to impose audience costs on the respondent state

---

[35] Recent work in the investment treaties literature has used such shared-frailty parameter for each country to account for country heterogeneity in cox proportional hazard models, see e.g. Soumyajit Mazumder, *Can I Stay a BIT Longer? The Effect of Bilateral Investment Treaties on Political Survival*, 11 REV. INT'L ORG. 477 (2016). In alternative specifications, I also used state fixed effects and find similar results (Table A3.5).

government in the face of an upcoming election, settlement becomes more costly for respondent states. When such domestic audience costs become large enough at a certain point in the respondent state's election cycle, the settlement range could cease to exist, and settlement would fail. Thus, we observe lower settlement risks when there is less time left until the next constitutionally mandated election.

Table 3.2: Effects of Share of Term Left on Settlement

	(1)	(2)	(3)	(4)	(5)	(6)
Share of Term Left	0.874**	0.881**	0.841**	0.850**	0.825*	0.822*
	(0.321)	(0.322)	(0.322)	(0.322)	(0.364)	(0.359)
Polity Score		-0.0286	-0.00444	-0.00617	0.0156	0.0162
		(0.0178)	(0.0186)	(0.0188)	(0.0223)	(0.0221)
GDP per capita (Logged)			-0.279**	-0.281**	-0.356**	-0.290**
			(0.0930)	(0.0937)	(0.104)	(0.105)
FDI (% of GDP)				-0.0174	-0.00961	-0.0115
				(0.0223)	(0.0237)	(0.0262)
Claims Number					-0.342**	-0.405**
					(0.0748)	(0.0789)
Year						-0.0878**
						(0.0185)
Observations	3221	3172	3107	3087	2758	2758

Note: Coefficient estimates from Cox proportional hazards model. All specifications include a country-specific frailty parameter. Standard errors in parentheses. + $p < .10$, * $p < .05$, ** $p < .01$

B. Effects of Domestic Audience Costs in Democracies and Non-Democracies

The main specification does not differentiate the impacts of Share of Term Left on settlement decisions in democracies and non-democracies. Democracies are often perceived to be more sensitive to domestic audience costs than non-democracies.[36] Indeed, a projection of the relationship between settlement rate and respondent state democracy level shows that more

---

[36] *See* Busch, *supra* note 2; Fearon, *supra* note 4.

democratic respondent states tend to have lower settlement rates in investor-state arbitration cases.[37] The fitted line in Figure 3.3 shows that an increase in the respondent state's Polity Score corresponds to a moderate decrease in settlement rate.[38]

Figure 3.3: Scatterplot of Settlement Rate and Polity Score with Prediction Line
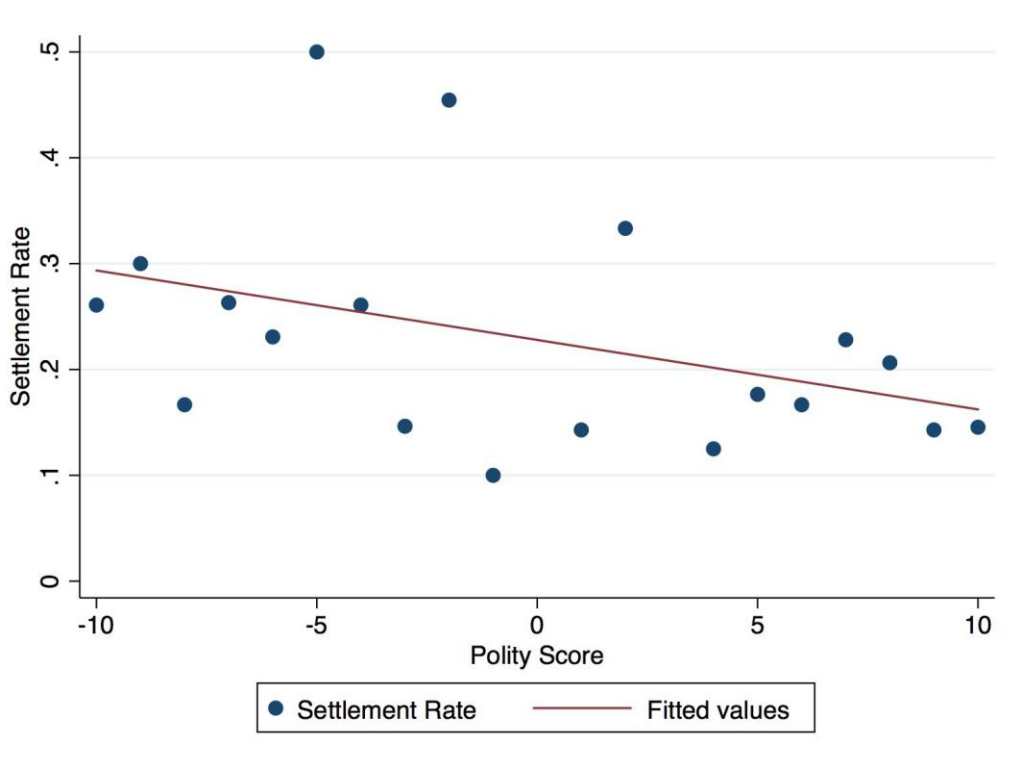
However, not all non-democracies are the same in that some non-democracies also hold elections. Elections in these non-democracies may also impose constraints on their settlement decisions. If that is true, it is possible that the higher settlement rate in non-democratic respondent states shown in Figure 3.3 is driven by those that do not regularly hold elections. Thus, based on

---

[37] In Table A3.4, I examine the effects of respondent states' Polity Score on their settlement likelihood using case-level data. The coefficient for *Polity Score* is negative, small in magnitude and not statistically significant in most specifications. This is likely due to the lack of variation in *Polity Score*. As Figure A2.3 shows, more than 50% of the respondent states have a polity score of 9 or 10, less than 10% of the respondent states have a polity score smaller than 5. This may explain why the coefficient for *Polity Score* is not significant.

[38] The Polity Score from the Polity IV index is often used to measure a country's democracy level, see Monty Marshall, Keith Jaggers and Ted Gurr, *Polity IV Project: Characteristics and Transitions, 1900 – 2009. Dataset Users' Manual* (2010), available at www.systemicpeace.org/inscr/pvmanualv2010.pdf.

the categorization from Cheibub, Gandhi and Vreeland (2010), I break respondent states into

different regime types and plot each type's settlement rate. As Figure 3.4 shows, while overall

democracies have a lower settlement rate than non- democracies,[39] this difference is largely

attributed to the high settlement rate of military dictatorships and royal dictatorships, which seldom

hold elections.[40] Civilian dictatorships, many of which regularly hold elections, have a settlement

rate that is similar to those of parliamentary and presidential democracies.[41] This suggests that

election appears to be the institution that matters in influencing settlement decisions.

Figure 3.4: Settlement Rate by Regime Type

---
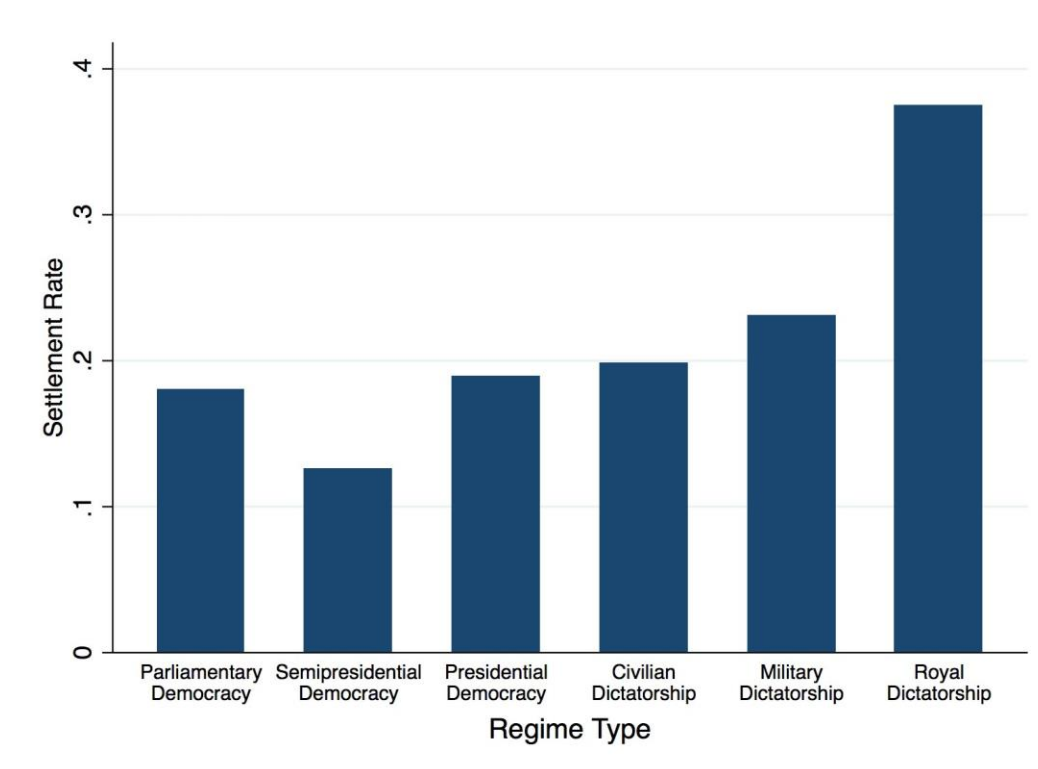
[39] The settlement rate of democracies is 17.07%. The settlement rate of dictatorships is 22.27%.
[40] 40% of military dictatorship respondent states have held elections during their cases in the dataset. None of the royal dictatorship respondent states have held elections during their cases in the dataset.
[41] Except for two where the election data is not available, all civilian dictatorship respondent states regularly held elections during their cases in the dataset.

To further examine whether the impacts of *Share of Term Left* on settlement probability is similar between democracies and non-democracies, I rerun the regressions in Table 2 after adding an interaction term *Share of Term Left × Democracy*.[42] Democracy equals one if the respondent state falls under one of the three democracy regime types according to the categorization by Cheibub, Gandhi and Vreeland (2010), and zero otherwise. If the effects of *Share of Term Left* on settlement risks is stronger in democracies than in non-democracies, then we expect the interaction to be positive. However, as Table 3 shows, *Share of Term Left × Democracy* is negative and statistically insignificant in all specifications. This indicates that democracies are not more sensitive to election pressures than non-democracies during settlement negotiations. The coefficient for *Share of Term Left* is positive and statistically significant in most specifications. This suggests that in cases with non-democratic respondent states, there is evidence that settlement risks decrease when it is closer to the next constitutionally mandated election. Thus, both democracies and non-democracies appear to be affected by upcoming elections which make settlement more costly due to increasing domestic audience costs.

---

[42] To address the concern that *Democracy* and *Polity Score* may be collinear, I have also run the same regressions without controlling for *Polity Score*, and find similar results (Table A3.6).

Table 3.3: Effects of Share of Term Left on Settlement (with Interaction Term)

	(1)	(2)	(3)	(4)	(5)	(6)
Share of Term Left	0.948[+]	0.960[+]	0.919	0.976[+]	1.503[+]	1.464[+]
	(0.575)	(0.578)	(0.577)	(0.578)	(0.796)	(0.777)
Democracy	-0.316	-0.0913	0.0394	0.0240	0.184	0.348
	(0.421)	(0.544)	(0.512)	(0.512)	(0.658)	(0.615)
Democracy × Share of Term Left	-0.116	-0.118	-0.114	-0.183	-0.869	-0.831
	(0.692)	(0.696)	(0.696)	(0.696)	(0.896)	(0.878)
Polity Score		-0.0199	-0.00351	-0.00243	0.0301	0.0206
		(0.0300)	(0.0283)	(0.0286)	(0.0343)	(0.0321)
GDP per capita (Logged)			-0.279**	-0.280**	-0.355**	-0.290**
			(0.0937)	(0.0944)	(0.105)	(0.106)
FDI (% of GDP)				-0.0182	-0.0125	-0.0132
				(0.0224)	(0.0240)	(0.0264)
Claims Number					-0.346**	-0.408**
					(0.0751)	(0.0790)
Year						-0.0870**
						(0.0185)
Observations	3221	3172	3107	3087	2758	2758

Note: Coefficient estimates from Cox proportional hazards model. All specifications include a country-specific frailty parameter. Standard errors in parentheses. + p < .10, * p < .05, ** p < .01.

C. Extension

1. Governance Quality

One possible confound is that *Share of Term Left* may be correlated with the governance quality of respondent states, which may affect case merits and settlement probability. It is possible that when elections approach, governments also become better-governed and less likely to engage in behavior that would violate their obligations under international investment treaties. Table 4 reports the correlation between *Share of Term Left* in year t and the respondent state's governance quality in that year, which is measured by six governance quality indicators from the World Bank's

Worldwide Governance Indicator Dataset.[43] These six indicators measure a country's governance quality from the perspectives of voice and accountability, political stability and absence of violence, governance effectiveness, regulatory quality, rule of law, and control of corruption. As Table 4 shows, none of the coefficients is statistically or substantively significant except for Political Stability, which is statistically significant at the 0.1 level. Overall, there is little evidence that *Share of Term Left* is endogenous through its correlation with respondent states' governance quality.

Table 4: Effects of Share of Term Left on Governance Indicators

	(1)	(2)	(3)	(4)	(5)	(6)
Voice and Accountability	-0.00679 (0.00613)					
Political Stability		-0.0125$^+$ (0.00646)				
Governance Effectiveness			-0.00603 (0.00654)			
Regulatory Quality				0.00000933 (0.00564)		
Rule of Law					-0.00333 (0.00579)	
Control of Corruption						-0.00317 (0.00646)
Constant	0.417** (0.00527)	0.414** (0.00549)	0.416** (0.00529)	0.417** (0.00533)	0.416** (0.00554)	0.416** (0.00558)
Observations	2924	2924	2924	2924	2924	2924

Standard in parentheses, $+$ $p < .10$, * $p < .05$, ** $p < .01$.

2. Selective Case Filing

[43] The Worldwide Governance Indicator Dataset is available at https://datacatalog.worldbank.org/dataset/worldwide-governance-indicators. Note that the Worldwide Governance Indicator Dataset only covers the time span of 1996--2016. This is unlikely to be an issue because only a very small number of cases are registered before 1996.

Another possible confound is that case filing timing may be deliberately selected based on the election timing, which may in turn affect the distribution of unobserved case merits across time. One might expect the dip in settlement prior to elections, as shown in Figure 3.2, is caused by a dip in case filings before elections.

However, as Figure 3.5 shows, this does not seem to be a concern. The first graph in Figure 3.5 presents the distribution of case registration based on the number of months relative to the nearest mandated election. The second graph in Figure 3.5 presents the distribution of case registration based on the quartiles of the constitutional term length relative to the nearest mandated election. The line represents the time of the nearest mandated election. The bars on the left of the line represent cases registered prior to the nearest mandated election. The bars on the right of the line represent cases registered after the nearest mandated election. In contrast to Figure 3.2, we do not observe a difference in case registration frequency in the periods before and after the mandated election. This suggests that case filing rate is not correlated to election timing, which largely rules out the possibility of selective case filing.

Figure 3.5: Case Registration Timing and The Nearest Mandated Election
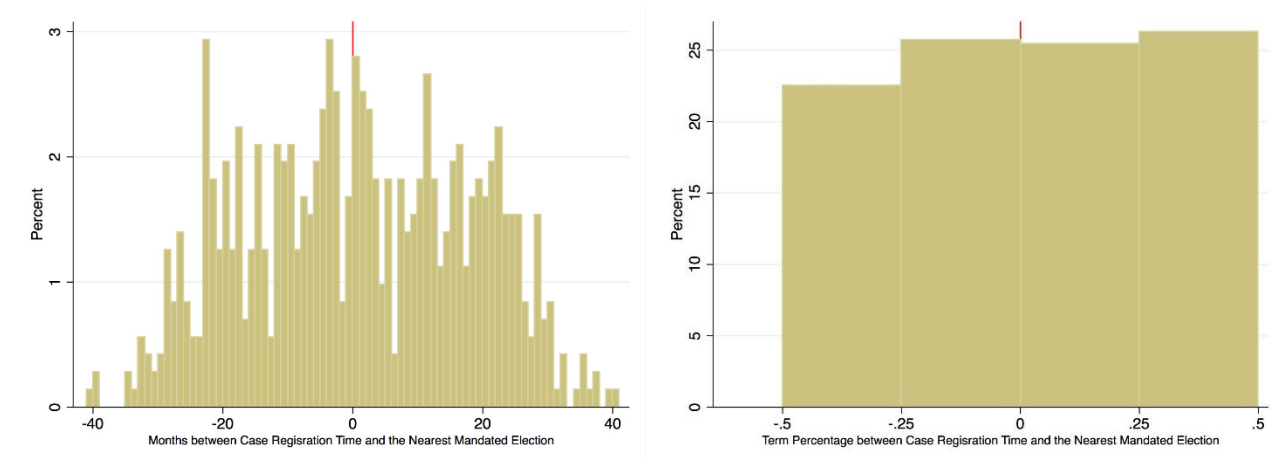
Nevertheless, there is still a residual concern. That is, even though the case filing rate does not change relative to election timing, the composition of cases being filed might have changed relative to the proximity to the next election due to strategic behavior of claimant investors. To address this concern, in Figure 3.6, I plot the distribution of the number of claims in the cases being filed based on the number of months or the quartiles of the constitutional term length relative to the nearest mandated election. The number of claims is a proxy for the underlying merits of cases being filed. Again, we do not observe a difference in the number of claims in the cases being filed in the periods before an after the mandated election. This further indicates that there is no evidence of selective case filing based on election timing.

Figure 3.6: Average Number of Claims of Cases Filed Relative to The Nearest Mandated
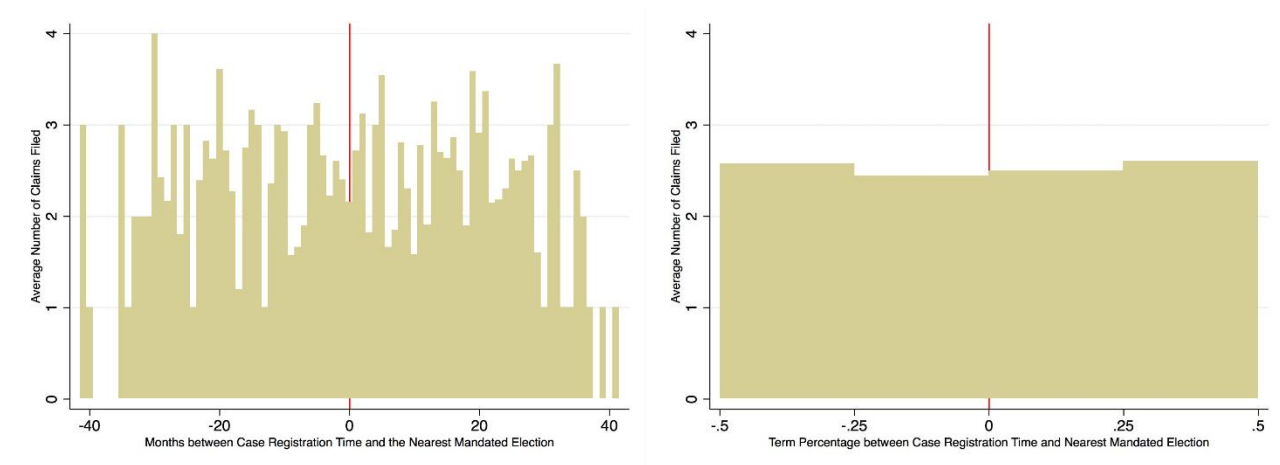
Election

3. Other Respondent State Activities

The pattern observed might reflect the overall reduction of government activities prior to an election, instead of a trend specific to settlement behavior. To address this concern, I examine the effect of proximity to the next constitutionally mandated election on another respondent state activity in investor-state arbitration proceedings: the appointment of respondent-side arbitrator. If we observe similar patterns with respect to arbitrator appointments as elections come near, then the decreased probability of settlement we observe might be explained by overall reduction of government activities prior to elections. However, the absence of such pattern would not rule out the influence of overall government activeness, as arbitrator appointment is only one example of government activities, and is subject to the time limitation of 90 days from case registration time.[44] Due to the short time limitation required for arbitrator appointment, I use monthly-level data here to allow for more variation in the data. Although imperfect, by examining whether these two types

[44] ICSID Convention, Art. 38.

of respondent state activities in the same proceeding exhibit similar patterns during the election cycle, the proposed test still provides implications as to whether the effects of proximity to the next constitutionally mandated election is more specific to settlement behavior.

The investor-state arbitration case dataset records arbitrator appointment information in 508 of the 801 cases. Similar to the data processing method for examining settlement, I adjusted the dataset to case-month level and merged it with the Database on Political Institution 2017. Each case includes a monthly observation from the month in which it was registered until the month in which the respondent state appoints its arbitrator. I use the same Cox proportional hazard model, except for changing the dependent variable to Appoint, an indicator for whether the respondent state in a case appoints its arbitrator in month t. Table A7 presents the regression results. None of the coefficients for Share of Term Left is statistically significant, although the coefficients remain positive.[45] Overall, I do not find consistent evidence that respondent states become less likely to appoint arbitrators when there is less time left until the next constitutionally mandated election as in the case of settlement. In Figure A3.3, I plot the distribution of arbitrator appointments by respondent states based on the number of months and quartiles of the constitutional term length relative to the nearest mandated election. Although more appointments have been made in the first quarter of the election cycle than in the last quarter of the election cycle, the difference is not as large as that shown in Figure 3.2. 16 appointments were made during the election month. Overall, the findings do not support the concern that the decreased settlement likelihood is entirely driven by overall reduction of government activities before elections.

D. Robustness Check

---

[45] When restricting the sample of cases to those involving democracy respondent states only, the coefficient is negative in all specifications.

1. Linear Probability Model

As an alternative estimation strategy, I also employ a linear probability model with the following specification:[46]

$$Settle_{c,t} = \beta_0 + \beta_1 Share\ of\ Term\ Left_{c,t} + \delta X_{c,t} + \gamma_i + \delta_y + E \qquad (2)$$

where $Settle_{c,t}$ is an indicator for whether case c is settled in year t. $Share\ of\ Term\ Left_{c,t}$ denotes the share of the constitutional term length left for the leader of the respondent state in case c in year t. $X_{c,t}$ is the vector of control variables. γ_i is a state-specific fixed effect. δ_t is a year-specific fixed effect. All standard errors are clustered at the case level.

The results are similar to those in the main specification. As Table A9 shows, In all specifications, I find that respondent states are significantly less likely to settle cases when there is less time left in the constitutional term length of their leaders. The coefficients for Share of Term Left are positive and statistically significant across specifications. The coefficient magnitudes are quite large, considering the baseline probability of settlement in a given year, which is 4.30%. The decrease in settlement likelihood relative to the baseline for respondent states that are one-quarter-term closer to the next constitutionally mandated election is around 23%.

2. Month-Level Data

To ensure that proximity to the next constitutionally mandated election is accurately measured, I also use month-level data in an alternative specification. Here the unit of analysis is case-month: each case includes a monthly observation from the month in which it was registered until the month in which it was concluded. If a case is still pending by March 2017, the monthly observation ends in March 2017. Similar to the main specification, I use the Cox Proportional

---

[46] I have also run logit regressions and the results are similar (Table A3.8).

Hazard Model. The dependent variable *Settle* is a dummy variable that equals one if the case is settled in month t, and zero otherwise. I generate a new variable, *Months Left in Current Term*, based on the timing of the previous election, which also comes from the Database of Political Institutions 2017, and the respondent state leader's constitutional term length. For instance, if the constitutional term length for a respondent state leader is four years, then the variable *Months Left in Current Term* is recorded as "47" for the month immediately following an election. I then use this variable to generate the independent variable of interest Share of Term Left.

Similar to the main specification, in all specifications, I control for the country-specific frailty parameter to account for country-specific heterogeneity that might affect hazard rate. In column (6), I include a counter variable Month that begins at 1 for the first calendar month in the dataset, to control for possible time trend.

Table A10 reports the regression results using month-level data. Again, the coefficients for Share of Term Left are positive and statistically significant in all specifications. The decrease in settlement risks for respondent states that are one-quarter-term closer to the next constitutionally mandated election is approximately 24%. The results hold when accounting for the respondent state's Polity Score, GDP per capita, amount of FDI relative to GDP, claims number and the time trend.

V. CONCLUSION

This paper provides robust evidence that proximity to the next constitutionally mandated election lowers settlement likelihood in investor-state arbitration. The paper also shows that non-democracy respondent states are no less responsive to election pressure than democracies in terms of case settlement. The findings suggest that case settlement in investor-state arbitration is affected by domestic political calculations. The findings in this paper are consistent with those in the

literature on state behavior in GATT/WTO proceedings: domestic politics appear to affect state decision-making in international dispute settlement. The influence of domestic audience costs may generate significant inefficiencies in ISDS. Due to such influence, elected officials in respondent states do not always make settlement decisions in ways that would be financially optimal for the state. As a result, settlement may not function as effectively a filter mechanism as it should. This could result in excessive litigation expenses and considerable uncertainty for disputing parties and third parties, which could have been avoided had states based their settlement calculations solely on case merits. Failure to settle may also contribute to some of the institutional problems surrounding ISDS, such as perceived biases and lengthy proceedings.

The cost arising from the settlement behavior itself has not been formally studied in previous literature on settlement and litigation. The findings in this paper indicate that the presence of such costs may prevent a case that would otherwise result in settlement from being settled or delay settlement. These costs may also exist not just in the context of investor-state arbitration, but in a broader range of circumstances where settlement is practiced.[47] Future research could further explore the effects of such settlement costs on settlement and litigation behavior in both domestic and international contexts.

---

[47] For example, some scholars have suggested that repeat defendants may strategically choose to not settle so as to develop a "reputation for toughness" and deter future nuisance suits, see e.g. Marc Galanter, *Why the Haves Come Out Ahead: Speculations on the Limits of Legal Change*, 9 L. SOC. REV. 95 (1974); Ivan P'ng, *Strategic Behavior in Suit, Settlement, and Trial*, BELL J. ECON 539 (1983); Thomas Miceli, *Optimal Deterrence of Nuisance Suits by Repeat Defendants*, 13 INT'L REV. L. ECON. 135 (1993).

Figure A3.1: Appearing as Respondent in Investor-State Arbitration by Country
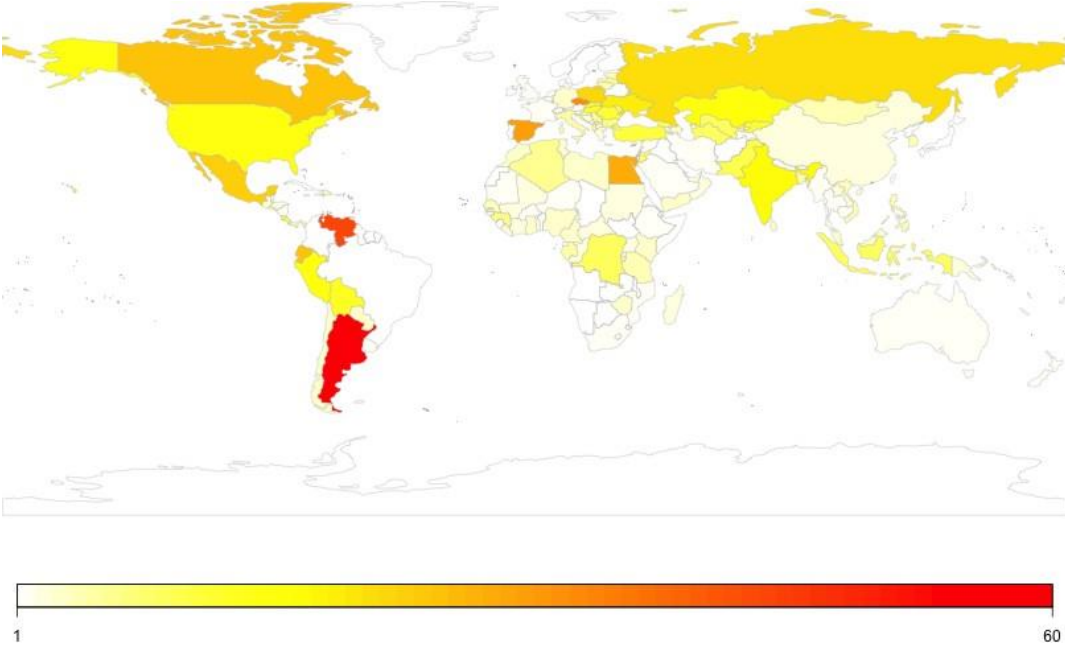
Figure A3.2: Settlement in Investor-State Arbitration by Country
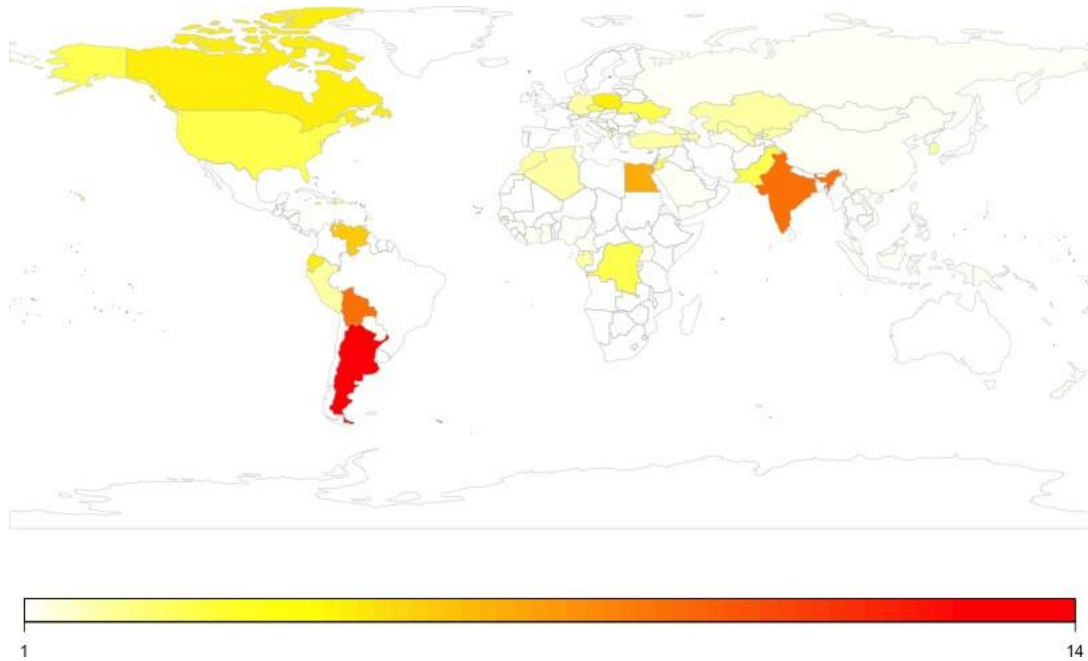
Figure A3.3: Appointment Timing and The Nearest Mandated Election
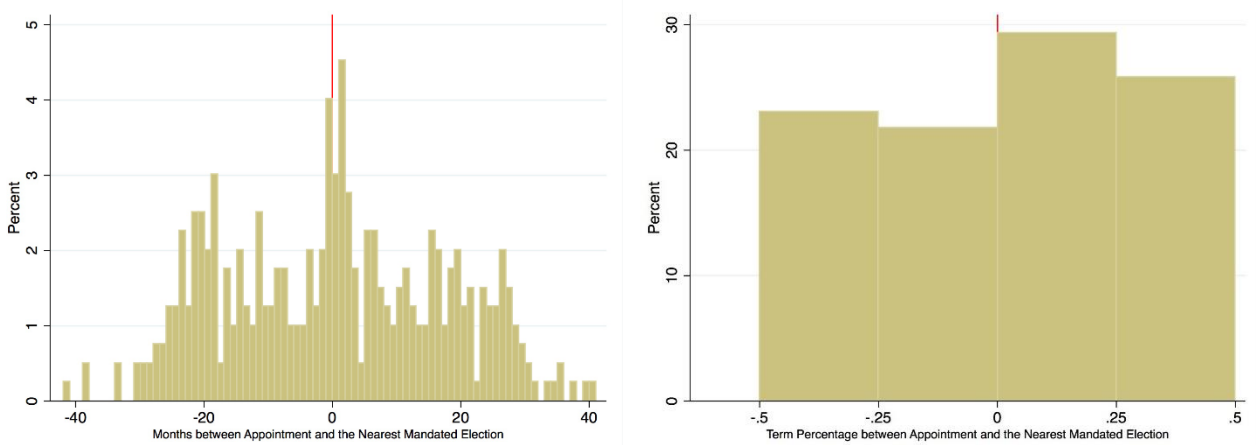
Table A3.1: Case Outcome

Case Outcome	Frequency	Percentage
Data Not Available	11	1.37
Decided	363	45.32
Discontinued	51	6.37
Pending	229	28.59
Settled	147	18.35
Total	801	100.00

Table A3.2: Effects of Years Left on Settlement

	(1)	(2)	(3)	(4)	(5)	(6)
Years Left	0.169**	0.163*	0.147*	0.148*	0.139+	0.132+
	(0.0636)	(0.0637)	(0.0632)	(0.0631)	(0.0722)	(0.0705)
Polity Score		-0.0247	-0.00164	-0.00328	0.0177	0.0183
		(0.0179)	(0.0187)	(0.0189)	(0.0224)	(0.0222)
GDP per capita (Logged)			-0.271**	-0.272**	-0.346**	-0.281**
			(0.0936)	(0.0942)	(0.105)	(0.106)
FDI (% of GDP)				-0.0172	-0.00968	-0.0116
				(0.0223)	(0.0238)	(0.0264)
Claims Number					-0.344**	-0.407**
					(0.0749)	(0.0791)
Year						-0.0873**
						(0.0185)
Observations	3221	3172	3107	3087	2758	2758

Note: Coefficient estimates from Cox proportional hazards model. All specifications include a country-specific frailty parameter. Standard errors in parentheses. + $p < .10$, * $p < .05$, ** $p < .01$.

Table A3.3: Respondent State Regime

Respondent State Regime	Frequency	Percentage
Parliamentary Democracy	183	22.85
Semi-Presidential Democracy	143	17.85
Presidential Democracy	254	31.71
Civilian Dictatorship	126	15.73
Military Dictatorship	78	9.74
Royal Dictatorship	16	2.00
Unclear	1	0.12
Total	801	100.00

Table A3.4: Effects of Polity Score on Settlement (Case Level)

	(1)	(2)	(3)	(4)	(5)	(6)
Polity Score	-0.00229$^+$	-0.00126	-0.00185	0.000595	-0.00259	-0.00177
	(0.00132)	(0.00186)	(0.00197)	(0.00232)	(0.00414)	(0.00453)
GDP per capita (Logged)		-0.0418*	-0.0399*	-0.0301	-0.175	-0.277
		(0.0186)	(0.0186)	(0.0198)	(0.137)	(0.264)
FDI (% of GDP)			-0.00916*	-0.00546	-0.00151	0.00122
			(0.00387)	(0.00371)	(0.00517)	(0.00561)
Number of Claims				-0.0902**	-0.0864**	-0.0908**
				(0.0118)	(0.0128)	(0.0138)
State Fixed Effect					x	x
Year Fixed Effect						x
Observations	571	478	470	407	407	407

Note: Standard errors in parentheses. $+$ p < .10, * p < .05, ** p < .01.

Table A3.5: Effects of Share of Term Left on Settlement (State Fixed Effect)

	(1)	(2)	(3)	(4)	(5)	(6)
Share of Term Left	0.924**	0.924**	0.833*	0.834*	0.824*	0.860*
	(0.327)	(0.329)	(0.329)	(0.328)	(0.376)	(0.375)
Polity Score		-0.0249	0.0379	0.0372	0.0271	0.0220
		(0.0412)	(0.0468)	(0.0468)	(0.0741)	(0.0713)
GDP per capita			-1.979**	-1.913**	-2.784**	-0.786
(Logged)			(0.532)	(0.534)	(0.661)	(1.216)
FDI (% of GDP)				-0.00904	-0.00150	-0.0114
				(0.0302)	(0.0340)	(0.0362)
Claims Number					-0.381**	-0.406**
					(0.0863)	(0.0888)
Year						-0.0875+
						(0.0452)
State Fixed Effect	x	x	x	x	x	x
Observations	3221	3172	3107	3087	2758	2758

Note: Coefficient estimates from Cox proportional hazards model. Standard errors in parentheses. $+ \, p < .10$, $^* \, p < .05$, $^{**} \, p < .01$.

Table A3.6: Effects of Share of Term Left on Settlement (with Interaction Term, without Polity Score)

	(1)	(2)	(3)	(4)	(5)	(6)
Share of Term Left	0.948[+]	0.948[+]	0.922	0.980[+]	1.539[+]	1.486[+]
	(0.575)	(0.575)	(0.576)	(0.577)	(0.803)	(0.782)
Democracy	-0.316	-0.316	0.00418	0.00236	0.468	0.526
	(0.421)	(0.421)	(0.423)	(0.424)	(0.562)	(0.545)
Democracy × Share of Term Left	-0.116	-0.116	-0.129	-0.199	-0.907	-0.849
	(0.692)	(0.692)	(0.693)	(0.694)	(0.902)	(0.883)
GDP per capita (Logged)			-0.283**	-0.283**	-0.330**	-0.265**
			(0.0921)	(0.0926)	(0.102)	(0.0989)
FDI (% of GDP)				-0.0181	-0.0143	-0.0158
				(0.0221)	(0.0249)	(0.0271)
Claims Number					-0.344**	-0.405**
					(0.0749)	(0.0788)
Year						-0.0892**
						(0.0186)
Observations	3221	3221	3156	3136	2795	2795

Note: Coefficient estimates from Cox proportional hazards model. All specifications include a country-specific frailty parameter. Standard errors in parentheses. + $p < .10$, * $p < .05$, ** $p < .01$.

Table A3.7: Effects of Share of Term Left on Appointment

	(1)	(2)	(3)	(4)	(5)	(6)
Share of Term Left	0.153	0.182	0.181	0.193	0.112	0.108
	(0.168)	(0.169)	(0.169)	(0.169)	(0.183)	(0.182)
Polity Score		-0.00106	0.00130	0.00178	-0.00320	0.00264
		(0.00965)	(0.0109)	(0.0110)	(0.0118)	(0.0112)
GDP per capita (Logged)			-0.0315	-0.0335	-0.00340	-0.103
			(0.0584)	(0.0595)	(0.0699)	(0.0677)
FDI (% of GDP)				-0.00643	0.00411	0.00487
				(0.00963)	(0.0123)	(0.0119)
Claims Number					-0.0545	-0.0165
					(0.0341)	(0.0339)
Month						0.0732**
						(0.0141)
Observations	2192	2162	2156	2154	1838	1838

Note: Coefficient estimates from Cox proportional hazards model. All specifications include a country-specific frailty parameter. Standard errors in parentheses. + p < .1, * p < .05, ** p < .01.

Table A3.8: Effects of Share of Term Left on Settlement (Logit)

	(1)	(2)	(3)	(4)	(5)	(6)
Share of Term Left	0.0590**	0.0588**	0.0611**	0.0605**	0.0576**	0.0389*
	(0.0182)	(0.0182)	(0.0185)	(0.0184)	(0.0179)	(0.0179)
Polity Score		-0.002024	0.000336	0.000344	0.000746	-0.003203
		(0.002173)	(0.002587)	(0.002599)	(0.00344)	(0.00302)
GDP per capita (Logged)			-0.0678**	-0.0649**	-0.0950**	
			(0.0241)	(0.0025)	(0.0291)	
FDI (% of GDP)				-0.001084	-0.000700	
				(0.00111)	(0.00161)	
Number of Claims					-0.0170**	-0.0210**
					(0.00406)	(0.00419)
State Fixed Effect	x	x	x	x	x	x
Year Fixed Effect						x
Observations	2348	2340	2279	2264	1996	2033

Note: Standard errors in parentheses. Logit coefficients are reported as marginal effects. In column (7), I do not control for GDP per capita (Logged) and FDI (% of GDP) because the interactions do not converge. + p < .10, * p < .05, ** p < .01.

Table A3.9: Effects of Share of Term Left on Settlement (Linear)

	(1)	(2)	(3)	(4)	(5)	(6)
Share of Term Left	0.0433**	0.0442**	0.0455**	0.0453**	0.0421**	0.0395**
	(0.0124)	(0.0125)	(0.0126)	(0.0126)	(0.0126)	(0.0128)
Polity Score		-0.00196	-0.000492	-0.000470	0.000599	-0.000904
		(0.00175)	(0.00186)	(0.00189)	(0.00215)	(0.00216)
GDP per capita (Logged)			-0.0693**	-0.0669**	-0.111**	-0.133**
			(0.0230)	(0.0232)	(0.0264)	(0.0473)
FDI (% of GDP)				-0.000620	-0.000585	-0.00108
				(0.000886)	(0.000868)	(0.000884)
Claims Number					-0.0111**	-0.0130**
					(0.00232)	(0.00239)
State Fixed Effect	x	x	x	x	x	x
Year Fixed Effect						x
Constant	0.0229**	0.0330**	0.633**	0.615**	1.035**	1.115**
	(0.00620)	(0.0110)	(0.200)	(0.202)	(0.235)	(0.403)
Observations	3221	3172	3107	3087	2758	2758

Note: Standard errors in parentheses. $+ p < .10$, $^* p < .05$, $^{**} p < .01$.

Table A3.10: Effects of Share of Term Left on Settlement (Month-Level Data)

	(1)	(2)	(3)	(4)	(5)	(6)
Share of Term Left	0.634*	0.655*	0.659*	0.665*	0.682^{+}	0.634^{+}
	(0.313)	(0.314)	(0.314)	(0.313)	(0.353)	(0.348)
Polity Score		-0.0216	0.00660	0.00624	0.0292	0.0289
		(0.0185)	(0.0192)	(0.0193)	(0.0222)	(0.0224)
GDP per capita (Logged)			-0.314**	-0.316**	-0.364**	-0.324**
			(0.0948)	(0.0952)	(0.103)	(0.105)
FDI (% of GDP)				-0.0175	-0.00885	-0.0106
				(0.0224)	(0.0232)	(0.0257)
Claims Number					-0.378**	-0.435**
					(0.0755)	(0.0793)
Month						-0.0815**
						(0.0188)
Observations	29580	29144	28586	28538	25631	25631

Note: Coefficient estimates from Cox proportional hazards model. All specifications include a country-specific frailty parameter. Standard errors in parentheses. + p < .10, * p < .05, ** p < .01.