Title
Empirical Essays on the Enforcement of Domestic and International Contracts

Author
Nyarko, Julian

Publication Date
2018

Peer reviewed|Thesis/dissertation

Empirical Essays on the Enforcement of Domestic and International Contracts

By

Julian Nyarko

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Jurisprudence and Social Policy

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Katerina Linos, Co-Chair
Professor Kevin M. Quinn, Co-Chair
Professor Robert P. Bartlett, III
Professor Robert D. Cooter
Professor Andrew T. Guzman
Professor Robert L. Powell

Summer 2018

Abstract

Empirical Essays on the Enforcement of
Domestic and International Contracts

by

Julian Nyarko

Doctor of Philosophy in Jurisprudence and Social Policy

University of California, Berkeley

Professor Katerina Linos, Co-Chair

Professor Kevin Quinn, Co-Chair

Each essay in this dissertation analyzes an original data set to empirically investigate widely held beliefs about the enforcement of contracts in a domestic and in an international context.

Chapter 1 tests the assumption that sophisticated parties prefer arbitration over litigation in international agreements. Relying on over half a million contracts filed with the SEC between 2000 and 2016, the results suggest that arbitration clauses are less frequently adopted than clauses referring parties to the domestic court system. If they are included, arbitration clauses serve the specific purpose of strategically reducing the discretion granted to the courts enforcing the decision. Absent serious threats to enforcement, parties prefer courts over arbitration, making arbitration a second-best alternative to a well-functioning domestic judiciary.

A striking result arising out of the investigation in Chapter 1 is that most commercial contracts lack a choice-of-forum provision altogether, even though the stakes in these contracts are typically very high. Chapter 2 investigates the reasons for this omission and shows that it is the consequence of a "sticky" drafting process used by the representing law firms. This phenomenon leads to the widespread use of incomplete and suboptimal agreements between even the most sophisticated commercial actors. The deficiency is the result of a lack of firm-level policies on how contracts should be drafted, paired with a strong tendency for law firms to stick to existing but often deficient templates. The findings imply that default rules matter not only for the distribution, but also for the final allocation of goods, even if the parties are highly sophisticated and the transaction costs are negligible.

Chapter 3 shifts the focus from contracts between parties to contracts between states. Scholars have argued that there is little use in the treaty instrument

as a modern policy tool and that the executive agreement is a more reliable commitment device that comes at a reduced cost. In contrast, Chapter 3 uses survival time analysis to demonstrate that agreements concluded in the form of a treaty are more durable than those concluded as executive agreements. The analysis suggests that this is the result of increased political costs imposed by the treaties' Advice and Consent procedure. Together, the findings imply that treaty usage signals a higher level of commitment than executive agreements. Abolishing the treaty would lock negotiators out of the possibility to indicate their intended level of compliance, potentially leading to fewer agreements with less favorable terms.

Chapter 4 addresses a long-standing debate about the relative merits of lawyers and non-lawyers as adjudicators in international dispute settlement. Some argue that lawyers would encourage predictability and coherence in jurisprudence. Others believe that non-lawyers would better protect state interests. Both sides of the debate assume that lawyers are more formalist, while non-lawyers are more instrumentalist. However, combining multiple-imputation, matching, and post-matching regression analysis, the analysis shows that panel chairs without a law degree and without much experience make a greater effort to signal adherence to formalist rules and competence in WTO jurisprudence than lawyers. The Appellate Body deems the signal credible, in turn rewarding inexperienced non-lawyers with a decrease in reversal rates. Overall, the results suggest that non-lawyers display similar (if not greater) levels of formalism as lawyers, calling into question a central reservation against non-lawyers in adjudicatory positions.

To those who make coding more accessible.

Acknowledgement

I would like to acknowledge those who have played a central role in the process of writing and finishing this thesis. First and foremost, I thank my academic advisors. Professor Kevin Quinn built the foundation for my empirical training, provided invaluable advice on all things empirical and was always available to answer one of my many, many (, many, many) questions. Professor Katerina Linos stepped in and co-chaired my committee, read and provided detailed and honest feedback on my drafts and guided me in the process of better addressing a legal audience in my writings. In addition, I would like to thank my former chair, Dean and Professor Andrew Guzman, for introducing me to the economic analysis of international law and for making time to provide insightful and prompt feedback on both my research and my academic career, even after becoming a dean. I would also like to thank Professor Robert Cooter for his advice, both academic and personal, and for his efforts in securing summer funding to allow me to focus on completing my thesis. Professor Robert Bartlett provided detailed guidance on the second chapter of this thesis and invaluable contributions stemming from his practical experience, to which I am most grateful.

Finally, I would like to acknowledge my family and friends who supported me while completing this thesis. First and foremost, I thank my wife Nurhan, who encouraged and supported my plans of completing a Ph.D. in the United States, even though it meant we would not be able to see each other for the majority of the next five years. In addition, I would like to thank my brother Benjamin, who made it financially viable for me to travel to the United States in the first place. And last but not least, I am grateful to my friends in Hamburg, who I have known since high school and who made sure I always had something to look forward to when I traveled back to Germany.

Table of Contents

Introduction **1**

1 We'll See You in... Court! The Lack of Arbitration Clauses in International Commercial Contracts **4**
1.1 Introduction . . . 5
1.2 Theory and Literature Review . . . 6
1.3 Data & Methodology . . . 14
1.4 Analysis . . . 20
1.4.1 Greater Efficiency of the Arbitration Process . . . 20
1.4.2 Concerns for Court Biases . . . 22
1.4.3 Enforcement Concerns . . . 23
1.5 Discussion . . . 26
1.6 Conclusion . . . 30
1.7 Figures and Tables . . . 32

2 Stickiness of Contractual Gaps: Explaining the Lack of Forum Selection Clauses in Commercial Agreements **52**
2.1 Introduction . . . 53
2.2 Theory . . . 54
2.1 Omission of Forum Selection Clauses: A Rational Gap? 59
2.2 Hypotheses . . . 66
2.3 Data Description & Methodology . . . 70
3.1 Contracts Data . . . 70
3.2 General Counsel Data . . . 75
3.3 Litigation Data . . . 76
3.4 Analytical Approach . . . 76
2.4 Results . . . 85
2.5 Discussion . . . 89
2.6 Conclusion . . . 91
2.7 Figures and Tables . . . 92

3 Giving the Treaty a Purpose: Comparing the Durability of Treaties and Executive Agreements **109**
3.1 Introduction . . . 110
3.2 Theory . . . 111
2.1 Legal Substitutability . . . 112
2.2 Differences in Reliability . . . 114
2.3 Prior Empirical Work . . . 117
3.3 Data Description and Methodology . . . 119
3.1 The Data . . . 119
3.2 Methodology . . . 121
3.3 Summary Statistics . . . 124
3.4 Results . . . 126
3.5 Discussion . . . 129
3.6 Conclusion . . . 132
3.7 Figures and Tables . . . 133

4 Conforming Against Expectations: The Formalism of Non-Lawyers at the WTO **144**
4.1 Introduction . . . 145
4.2 Theory . . . 147
2.1 Formalism in Legal Opinions . . . 147
2.2 Lawyers versus Non-Lawyers . . . 150
2.3 How WTO Reports are Drafted . . . 153
4.3 Data Description and Methodology . . . 154
3.1 Population . . . 154
3.2 Data . . . 155
3.3 Methodology . . . 157
4.4 Results . . . 164
4.1 Primary Results . . . 164
4.2 Robustness Checks . . . 167
4.3 Precedent and Reversal . . . 168
4.4 Summary of Results . . . 170
4.5 Implications and Conclusion . . . 170
4.6 Figures and Tables . . . 172

Conclusion **193**

References **195**

Appendix **213**
A.1.1 Detailed Description of Text Analysis Procedures . . . 214
A.1.2 Arbitration Clauses Illustrating Cosine Similarity . . . 220
A.1.3 Published ICC Award No. 10947 223
A.1.4 Names and Abbreviations of Arbitration Organizations 225
A.1.5 Detailed Regression Results for Arbitration and Court Clauses 226
A.2.1 Formalizing Incentive Costs 234
A.2.2 Demonstrating Heterogeneity in Predictability 235
A.3.1 Groupings of Treaties by Subject Area 238
A.3.2 Complete List of Agreement Type by Subject Area . . 241
A.3.3 Complete List of Agreements by Country 242
A.3.4 Graphical Representaiton of Schoenfeld Residuals . . . 246
A.4.1 Reduction in Bias Obtained by Matching with Different Calipers 250
A.4.2 Robustness of Results for the Analysis of Reversal Rates 251
A.4.3 Sensitivity of Results for the Analysis of Reversal Rates 254

List of Figures

1.1 Forum Selection Clause Usage over Time 35
1.2 Arbitration Clause Usage over Time 36
1.3 Court Clause Usage over Time 37
1.4 Governing Law Clause Usage over Time 38
1.5 Density Plot of Similarity Scores 41
1.6 Forum Selection Clause Usage over Rule of Law Scores 51

2.1 Use of Forum Selection Clauses Over Time 93
2.2 Density Plot of Within-Company Means 98
2.3 Density Plot of Within-Company Variance 99
2.4 ARIMA-Model of Forum Selection Clause Usage 104
2.5 Synthetic Controls Method 105

3.1 Agreement Types over Time 140
3.2 Governing Law Clause Usage over Time 141
3.3 Governing Law Clause Usage over Time 142
3.4 Omitting Sole Executive Agreements 143

4.1 Citation Frequency by Panelist Type 172
4.2 Listwise Deletion vs. Multiple-Imputation 174
4.3 Balance Improvement Through Matching 176
4.4 The Effect of Lawyer Chair Across Calipers 186
A.3.1 Schoenfeld Residual Plots 246

List of Tables

1.1 Summary Statistics 32
1.2 Agreement Types 33
1.3 Industries 34
1.4 Most Popular Arbitration Institutions 39
1.5 Most Popular Court Forums 40
1.6 Significance Tests for Differences between Forum Selection Clauses 42
1.7 Logit-Regression on Arbitration Clause Usage 43
1.8 Logit-Regression on Court Clause Usage 45
1.9 Average Marginal Difference in Forum Selection Clause Usage . 47
1.10 Selected Contracting Countries 48
1.11 Logit-Regression of Forum Selection Usage on Judicial Institutions 49

2.1 Summary Statistics 92
2.2 Industries 94
2.3 Agreement Types 95
2.4 Most Popular Court Forums 96
2.5 Law Firms 97
2.6 Law Firm / General Counsel Influence on Forum Selection Clauses 100
2.7 Clause Stickiness by Industry / Contract Type 101
2.8 Transition in Forum Selection Clauses 102
2.9 Textual Similarity of Clauses 103
2.10 Learning Effects 106
2.11 Learning Effects, Robustness 107
2.12 Learning Effects, Substantive Significance 108

3.1 Summary Statistics 134
3.2 Agreement Use by Subject Area 135
3.3 Agreement Use by Partner Country 136
3.4 Cox Proportional Hazard Model 137
3.5 Complementary Log-Log Model 138
3.6 Cox Model by Senatorial Support 139

4.1 Summary Statistics . . . 173
4.2 Combining Multiple Imputation and Matching . . . 175
4.3 Post-Matching Regression Analysis (Lawyer Chair) . . . 177
4.4 Regression Analysis Using Alternative Outcome Measures . . . 180
4.5 Alternative Model Specifications . . . 183
4.6 Sensitivity Analysis . . . 187
4.7 Regression of Reversal Rate on Citation Frequency . . . 189
4.8 Summary of Results . . . 192
A.1.1 Agreement Types and their Terms . . . 216
A.1.2 Arbitration Organizations . . . 225
A.1.3 Logit-Regression on Arbitration Clause Usage . . . 226
A.1.4 Logit-Regression on Court Clause Usage . . . 230
A.2.1 Relative Use of Courts . . . 237
A.3.1 Agreement Type by Subject Area . . . 241
A.3.2 Agreement Use by Partner Country . . . 242
A.4.1 Bias Reduction by Caliper . . . 250
A.4.2 Robustness Test for Reversal Analysis . . . 251
A.4.3 Sensitivity Analysis for Reversal Rates . . . 254

Introduction

Contracts are the central pillar of any modern economy. At the domestic level, they allow for the future exchange of goods and services between strangers; they reallocate the potential harm of contingencies, e.g. from risk-averse to risk-neutral parties; they ensure liquidity by facilitating borrowing and lending activity; and they promote the production of information by allowing those with diverging views to bet on the future price of a real or financial asset. At the international level, contracts between states serve many additional purposes. For instance, free trade agreements reduce tariffs and promote the flow of goods and services across borders; environmental treaties increase the supply of global public goods, such as clean air; and weapons treaties can help to reduce the use, research and stocking of weapons of mass destruction.

But not every contract is born equal. A promise can induce reliance and promote cooperation only if the promisee has the means to enforce it, making contract enforcement a vital aspect of any practical considerations in contract law. Whereas much scholarship has focused on theoretical models of contract enforcement, a lack of comprehensive data historically made it difficult to ascertain if and how these contractual promises are enforced in practice. This dissertation seeks to address this shortcoming. By making use of recent advancements in computational methods, it creates three novel, comprehensive data. This data is then used to systematically test theoretically motived, popular beliefs about different aspects of contract enforcement, both at the national and at the international level.

Chapter 1 combines web scraping, text analysis and statistical analyses to examine choice-of-forum provisions in over half a million material business-to-business contracts in order to evaluate the prevalence of and motivation for choosing arbitration over courts in international commercial agreements. Whereas arbitration has often been portrayed as a broadly applicable solution to idiosyncratic problems arising out of complex international business agreements, the results of this chapter imply that arbitration serves a much more limited and specific function. In particular, parties treat international

arbitration as a second-best alternative to a well-functioning domestic court system that is used not in order to avoid foreign courts, but in an attempt to avoid supposedly dysfunctional court systems.

A curious finding arising out of the data set under investigation in Chapter 1 is that most commercial contracts, both domestic and international, lack a choice-of-forum provision altogether, even though the stakes in these contracts are typically very high. Chapter 2 investigates the reasons for this omission. After finding that it seems implausible to characterize the lack of forum selection clauses as a rational gap, it is shown that the prevalence of choice-of-forum provision can be explained almost exclusively by the law firm that has drafted the contract. In particular, it is found that law firms use templates when writing the agreements and that these templates tend to be sticky, such that gaps that existed in the past are likely to persist in the future. Overall, the results cast doubt on the common assumption by courts and in the empirical literature on contract law that sophisticated parties in high-stakes interactions routinely contract around inefficient rules in an effort to create an optimal legal frame work. Instead, the findings imply that default rules matter not only for the distribution, but also for the final allocation of goods, even if the parties are highly sophisticated and the transaction costs are negligible.

Chapter 3 shifts the focus from contracts between private parties to contracts between states. It compiles a novel data set on the durability of international agreements between the United States and other countries concluded between 1982 and 2012. With the use of survival time analysis, the chapter demonstrates that agreements concluded in the form of a treaty are more durable than those concluded as executive agreements. Recent studies have called for the abolishment of the treaty in favor of the executive agreement, but the findings suggest that the implementation of such proposals may lock negotiators out of the possibility to indicate their intended level of compliance, potentially leading to fewer agreements with less favorable terms in the future.

The fourth chapter considers the role of an often neglected actor in the process of contract enforcement, the adjudicator. Using an original data set on the citations in all reports issued by the World Trade Organization's Dispute Settlement Body, the chapter tests the popular assumption that non-lawyers would be less beholden to formalistic ideals and thus would use a more instrumentalist adjudicatory decision making process than lawyers when assessing compliance with trade agreements. Relying on a combination of multiple-imputation, matching, and post-matching regression analysis, the findings suggest that panel chairs without a law degree and without much experience make a greater effort to signal adherence to formalist rules and competence in WTO jurisprudence than lawyers. The Appellate Body deems the signal credible,

in turn rewarding inexperienced non-lawyers with a decrease in reversal rates. Overall, the findings imply that non-lawyers display similar (if not greater) levels of formalism as lawyers, calling into question a central reservation against non-lawyers in adjudicatory positions.

Chapter 1

We'll See You in... Court! The Lack of Arbitration Clauses in International Commercial Contracts

1.1 Introduction

International arbitration is an intriguing phenomenon. Some view it as the hallmark for the settlement of cross-border disputes, arguing that it promotes efficiency of the dispute settlement process by providing a reliable enforcement mechanism (Fisher and Haydock, 1995). Arbitration is also viewed as an important instrument to overcome "hostage-taking", which originates from exchanges that require specialized investments (Williamson, 1983). Others, both in- and outside of the international commercial setting, are skeptical, pointing to the fact that arbitration creates a private and largely secret alternative court system. Some argue that, since decisions tend to not be available to the public and thus, are inadequate to establish precedent, arbitration does not provide a public good through its rendering of a decision (Landes and Posner, 1979). Others criticize the intransparency associated with the arbitral process that would diminish its legitimacy (Lew, 1982; Buys, 2003; Gruner, 2003). However, for all debate on the normative desirability of international arbitration, it is striking how little we know about its actual use in practice.

Empirical uncertainty surrounds the question of how prevalent international arbitration is to begin with. Estimates for the share of international agreements that include arbitration clauses range from 15% to 90% (Casella, 1996; Weidemaier, 2015; Eisenberg and Miller, 2007), leading to considerable disagreement on whether arbitration is an essential pillar in the landscape of international dispute settlement or a mechanism that is used in only a narrow, more or less coherent subset of contracts.

In addition to the question *if* parties are relying on arbitration to a significant degree, it is also unclear *why* parties would prefer arbitration over courts in a commercial context. The standard narrative suggests three sets of reasons. First, arbitral proceedings, being privately organized, are believed to be more susceptible to the specific needs of the business community, in turn allowing for a more efficient resolution of the dispute. Second, it is assumed that parties have an incentive to avoid foreign courts due to concerns for home biases, and that arbitration allows them to remove the dispute from the domestic court system of either party. And third, arbitral awards are viewed as more easily enforceable than foreign court decisions. However, while all these motivations might seem theoretically appealing, so far none of them have been validated empirically.

This chapter then provides the most comprehensive look at the practice and relevance of international commercial arbitration to date. It uses the population of over half a million material contracts of publicly held companies registered with the SEC between 2000 and 2016 to examine the role of arbitra-

tion in international contracts. The analysis yields two main findings. First, U.S. parties and those with close economic ties to the U.S. only rarely rely on arbitration. Whereas 25% of international agreements include arbitration clauses, 34% include clauses referring parties to domestic courts. Under the assumption that parties routinely opt for the optimal procedure, this implies that parties do not view arbitration procedures as more efficient than litigation. Second, there is little evidence to suggest that litigating in another countries' court is a general concern for parties. Instead, companies strategically use arbitral tribunals only if the contractual partner comes from a country with judicial institutions that pose a risk to the enforcement of a U.S. court decision. If the quality of the foreign judicial institutions is not in doubt, parties are much more likely to refer disputes to the U.S. judiciary than to arbitration.

Together, the findings shine a new light on the relevance of international arbitration in a commercial context. Whereas arbitration has often been portrayed as a broadly applicable solution to idiosyncratic problems arising out of complex international business agreements, the results of this study imply it serves a much more limited and specific function. In particular, parties treat international arbitration as a second-best alternative to a well-functioning domestic court system that is used not in order to avoid foreign courts, but in an attempt to avoid supposedly dysfunctional court systems.

The rest of the chapter proceeds as follows: The next section provides the theoretical underpinnings and an overview of the empirical literature on arbitration clause usage. Section 1.3 describes the data used in this study. Section 1.4 presents the analytical results, Section 1.5 discusses them and a last section concludes.

1.2 Theory and Literature Review

The perfect contract necessarily leads to a pareto improvement. That is, it makes no party worse off while making at least one party better off than it would have been without the contract. However, in reality, such a perfect contract does not exist. Language is, by definition, imprecise and requires interpretation. Parties lack precise information about the future and thus, are unable to specify the desired outcome for every possible contingency that might be realized. And even if they could write down a precise contract and foresee all possible contingencies, monitoring is necessarily imperfect, giving rise to disagreements over whether performance in accordance to the terms of the contract occurred or not. In short, wherever there is a contract, there is the possibility for a contractual dispute.

The Market for Dispute Settlement

Parties to a contract have the possibility to define the rules by which to solve potential disputes *ex ante*. In doing so, they act similar to consumers on a market for contractual instruments (Landes and Posner, 1979; Miller and Eisenberg, 2009; Ribstein and O'Hara, 2009). On this "market for contracts",[1] business entities are the consumers, shopping for dispute settlement forums (among others). The two main goods offered on the market that parties can use to solve their cross-border disputes are international arbitration and domestic court litigation. The most significant suppliers of arbitration are large, often private organizations such as the American Arbitration Association (AAA) and the International Chamber of Commerce's arbitration division. The supplier of court forums are the states.

The theory of optimal contract design (see e.g. Schwartz and Scott, 2003), which has been extended to the negotiation of procedural rules between sophisticated parties (Shavell, 1995; Scott and Triantis, 2006; Dodge, 2011) assumes that parties will agree on the dispute settlement mechanism that maximizes their joint utility. The market framework then assumes that arbitration organizations and states compete for a market share generated by the demand for international dispute settlement. Competition takes place on two levels. First, in the wake of *inter-industry* competition, states compete with arbitration organizations. Second, there is *intra-industry* competition, with states competing with one another and similarly arbitration organizations competing with each other. Both inter- and intra-industry competition exerts pressure on the suppliers of dispute settlement procedures to improve the efficiency of their respective settlement processes. However, while the competitive forces are comparable on a number of dimensions, the specific incentives differ between the different types of suppliers.

Arbitration organizations are incentivized by the monetary benefits they receive when being selected as the forum of choice. These benefits can be quite substantial. For example, at the International Chamber of Commerce (ICC), a dispute over $20,000,000 creates average liabilities of about $450,000 for administrative expenses and arbitrators' fees.[2] For states, the benefits can be both monetary and non-monetary. A study conducted by Cornerstone Research for the State of New York projects that international dispute settlement alone creates 2 billion dollars in annual revenue for law firms headquartered

[1] As coined by Miller and Eisenberg (2009).

[2] These numbers originate from Jones and Lloyd (2011) and have been consolidated in an ICC Arbitration Cost Calculator available at https://www.international-arbitration-attorney.com/icc-arbitration-costs-calculator/.

in the state, which amounts to about 10% of their total revenue. This number does not yet take into account additional revenue created by hotels, the gastronomy, the transportation industry etc. In addition, being the primary state forum for the settlement of international disputes allows states to spearhead the development of international business law, in turn solidifying their position as a commercial hub with substantial economic and political power (Ribstein and O'Hara, 2009).

It should be noted, however, that these benefits are reserved to a small subset of states, namely those which are most frequented. The reason lies in the existence of strong positive externalities. For instance, lawyers have a strong incentive to become well-versed in the laws and competent litigators before the courts of the states that currently attract the most litigation, because the expected return on their educational investment is highest in said states. A small court with a negligible amount of commercial disputes will have difficulties attracting the legal profession, even if it increases the efficiency of its dispute settlement procedure significantly. Against the backdrop of these positive externalities, it becomes clear that only a limited number of countries and states such as the U.K., New York, Delaware, California or Texas have an incentive to compete on the market for dispute settlement provisions, whereas smaller courts do not have a substantial incentive to attract more litigation and might in fact be better off reducing their docket as much as possible.

Comparing Arbitration to Courts

In the literature, it is a widely held assumption that the recent increase in the transnational movement of goods and services has lead to an increase in the popularity of arbitration as the primary dispute settlement mechanism in international contracts (Knull III and Rubins, 2000). Many even believe that arbitration retains the majority of the international dispute settlement market (Knull III and Rubins, 2000; Stipanowich, 2009; Craig, 2010; Menon, 2014; Wagner, 2014). There are several theoretical arguments for this conjecture which can broadly be collected under two distinct categories.

The first category of arguments pertains to supposed efficiency of arbitration and applies both in an international and in a domestic context. Arbitration is believed to be more flexible than courts, allowing disputes to be settled faster and cheaper. The reason is that arbitration is not bound to the same procedural rules as courts. Many arbitration institutions offer their users a lot of room to customize the dispute settlement process, e.g. by limiting or avoiding discovery, preventing the use of motion practice or by setting fixed time limits for each stage of the process (Stipanowich, 2009). It is believed

that this discretion is used by the parties to streamline the procedure, leading to a fast and efficient resolution of disputes (Fisher and Haydock, 1995). In addition, arbitration as a private dispute settlement process is confidential, which is especially relevant to commercial disputes in which parties often have a significant interest not to reveal certain information pertaining to their businesses to the public. Also, arbitration commonly allows parties to choose their own arbitrators and many large arbitration organizations provide their users with a subject specific list of experts from which the parties choose their arbitrators (Franck, 2006). This is believed to result in greater expertise of the adjudicator who is then better suited to resolve a dispute in the interest of the parties (Knull III and Rubins, 2000).

The second set of arguments for why arbitration is viewed as superior to litigation pertains specifically to dispute settlement in an international context, where arbitration is perceived as especially relevant for two reasons. First, parties are assumed to have a fundamental distrust in each others' court systems due to the possibility for home bias (Drahozal, 1999). Arbitration in a neutral, third country is viewed as a way for parties to circumvent these potential biases. Second, the enforcement of arbitral awards is considered easier than the enforcement of foreign judicial decisions. Most commentators see the reason in the existence of an international legal regime that governs the enforcement of arbitral awards, most importantly the New York Convention on the Recognition and Enforcement of Foreign Arbitral Awards, which is often referred to as a "cornerstone" in the international transactional environment (Van den Berg, 1981). The treaty greatly reduces the discretion domestic courts have in reviewing the legality of arbitral awards rendered in another member state. In this way, arbitration is viewed as helping parties incorporated in states with a weak judiciary to overcome the commitment problem they face. Foreign judgments lack a comparable international institutional framework and thus, their enforcement faces a greater threat of non-compliance (Bühring-Uhle et al., 2006; Wagner, 2014).

However, while all these considerations seem intuitively appealing, the underlying assumptions leave room for doubt. Consider first the claim that arbitral proceedings are more efficient and better able to cater to the preferences of commercial parties. The underlying scholarly debate often focuses on efforts by arbitration organizations to improve the efficiency of their procedures while neglecting that states have powerful incentives to retain a large share of the market as well, as discussed above. This then leads to the incorrect assumption that arbitration organizations have successfully catered to the preferences of commercial parties while traditional court systems are inflexible and too cumbersome to be frequently relied upon in a business environment. There is

substantial evidence to suggest otherwise. For instance, since 1992, twenty-eight states in the U.S. have created specialized business courts designed to handle complex commercial disputes more efficiently and with greater expertise. Dammann (2016) shows that the creation of these courts is associated with a subsequent increase in firm performance. State courts, especially those characterized as "textualist" such as New York (Dammann and Hansmann, 2009; Gilson et al., 2014), often pay great attention to develop a jurisprudence that maximizes predictability with the goal to minimize uncertainty in business dealings.[3] Contractual waivers of provisions viewed with skepticism by the business community, such as jury trials and punitive damages, find increasing acceptance by courts. And even under the assumption that arbitration provides parties with more flexibility in designing their preferred dispute settlement process, some scholars are doubtful that parties adequately take advantage of this flexibility. For instance, Stipanowich (2009) cautions of the increased tendency to turn arbitration into just another form of litigation, with discovery processes and submissions of evidence comparable to U.S. litigation. This, in turn, would annul many of the benefits in speed and costs arbitration is supposed to provide. Yet other commentators criticize that arbitration has the potential to lead to unexpected outcomes, since arbitrators are paid by the parties and thus have an incentive to "split the baby" (Farber and Bazerman, 1984; Dammann and Hansmann, 2009).[4] Together, these rationales

[3]See e.g. the position of courts in New York in regards to the "four corners" rule: *[The four corners] rule imparts "stability to commercial transactions by safeguarding against fraudulent claims, perjury, death of witnesses * * * infirmity of memory * * * [and] the fear that the jury will improperly evaluate the extrinsic evidence." (Fisch, New York Evidence § 42, at 22 [2d ed].) Such considerations are all the more compelling in the context of real property transactions, where commercial certainty is a paramount concern.* (W.W.W. Assocs., Inc. v. Giancontieri, 77 N.Y.2d 157, 162, 566 N.E.2d 639, 642 (1990)); Regarding refiling financial statements after a name change as an expression of a general duty to act in good faith: *While UCC 1-203 provides that every contract or duty within the UCC imposes an obligation of good faith in its performance or enforcement, to impose a generalized duty to refile, not fairly precisely fixed in a particular section of the UCC, would upset the preference for definiteness, regularity and predictability in commercial dealings.* (Fleet Factors Corp. by Ambassador Factors Div. v. Bandolene Indus. Corp., 86 N.Y.2d 519, 519–20, 658 N.E.2d 202 (1995)); On the applicability of a choice of law clause: *New York has an overriding and paramount interest in the outcome of this litigation. It is a financial capital of the world, serving as an international clearinghouse and market place for a plethora of international transactions, such as to be so recognized by our decisional law (Intercontinental Planning v Daystrom, Inc., supra, at pp 383-384). (...) In order to maintain its pre-eminent financial position, it is important that the justified expectations of the parties to the contract be protected.* (J. Zeevi & Sons, Ltd. v. Grindlays Bank (Uganda) Ltd., 37 N.Y.2d 220, 227, 333 N.E.2d 168 (1975))

[4]However, the validity of this argument remains questionable in the light of a wave of

raise serious doubts as to the supposed advantage in efficiency of the arbitral process.

Consider now the claim that arbitration is especially popular in international contracts due to concerns for home biases. While home bias might be of concern in some commercial contracts, many countries are perceived to have a well developed judicial system that is largely immune to influences from political or private parties. Given that most cross-border business is conducted between highly developed countries with an independent judiciary, it seems questionable why the potential for disparate treatment should be a major concern in a majority of commercial contracts. Instead, it seems reasonable to assume that arbitration is especially popular in contracts between companies from jurisdictions with vastly different judicial quality, whereas it is of less relevance if judicial quality is not a concern.

Regarding the consideration that the New York Convention would ease the enforcement of a decision rendered outside of the country in which it is enforced, it should be noted that even for arbitration awards, the domestic judiciary does not necessarily become a passive bystander. While the New York Convention intends to prevent domestic courts from reevaluating the merits of an arbitral award, survey evidence suggests that the perception of the ease with which an award can actually be enforced differs significantly by country. Indeed, respondents in a 2008 survey on the enforcability of arbitration awards described many developing countries as hostile towards the enforcement of foreign arbitral awards, echoing concerns that the domestic judiciary will not assist foreign parties in their efforts to enforce.[5] Due to the impossibility to eliminate the role of the judiciary as the final decision maker in the process of enforcing a claim arising out of an international contract, it is thus possible that assertions focusing on the easier enforcability under the New York Convention are overstated. To be sure, this does not necessarily imply that enforcement considerations are irrelevant. Even absent the New York Convention and similar treaties, arbitral awards might be easier to enforce simply because the enforcement of a privately issued opinion does not infringe on a nation's sovereignty in the same way that the enforcement of a foreign decision does. Indeed, ratification of the New York Convention could merely be a reflection of states' greater willingness to enforce arbitral awards in the first place, with the treaty not changing state preferences over enforcement in an observable way (Downs et al., 1996).

Existing Empirical Evidence

most recent studies on the amount granted in arbitration awards, see Weber et al. (2014).

[5]See Lagerberg and Mistelis (2008).

While a theoretical assessment does not provide a clear answer as to the relevance of arbitration in an international commercial framework, the empirical landscape on arbitration usage is similarly inconclusive. Scholars often point to a reported increase in the rates of arbitration filings at the large arbitral organizations as an indication for its increasing popularity and widespread acceptance. For example, in 2016 alone, the International Chamber of Commerce (ICC) reports an increase in their caseload by 20%, compared to the previous year.[6] However, caseload alone is a poor indicator for the popularity of international arbitration, as the difference could be solely driven by an increase in international commercial activity and does not take into account the number of cases that are resolved through informal means such as negotiations (Wagner, 2014).

A second piece of empirical evidence often employed are surveys. In surveys on arbitration usage in practice, respondents commonly report quite high usage rates. For instance, the periodic International Arbitration Survey consistently reports that about 90% of their respondents prefer arbitration over other forms of dispute settlement.[7] The usage rates in the Litigation Trends Annual Survey are somewhat more modest, even though arbitration is more popular than litigation here as well, with 48% of surveyed companies in the latest survey reporting a preference for arbitration over litigation in international contracts.[8] However, both surveys struggle with a significant share of non-respondents. They ask detailed questions about a companies' arbitration practice that imposes a considerable research cost on its respondents. It is thus likely that those most interested in international arbitration are the most likely to respond, potentially subjecting the studies to severe response bias. In addition, since the surveys prime the respondents to trade off arbitration against courts, reported rates could differ significantly from actual usage rates.

For a long time, conducting quantitative studies analyzing the *de facto* prevalence of international arbitration clauses was difficult due to a lack of comprehensive and accessible data on international commercial contracts. Scholars were thus limited to anecdotal evidence[9] or samples provided by third parties, which tend to employ an intransparent procedure to preselect the contracts they make available.[10] It is only more recently that scholars have begun to

[6]See https://iccwbo.org/media-wall/news-speeches/icc-reveals-record-number-new-arbitration-cases-filed-2016/

[7]For the latest survey, see Friedland and Mistelis (2015).

[8]See Pecht (2015).

[9]For example, Casella (1996) relies on anecdotal evidence by the Netherlands Arbitration Institution, which states that 80% of international contracts include arbitration clauses.

[10]For example, Drahozal and Naimark (2005) study 17 international joint venture agree-

sample contracts either directly from the SEC or through the use of LexisNexis, which is not subject to any known selection biases.[11]

Among the quantitative studies on the usage rate of arbitration clauses, two stand out in particular for their extensive and rigorous approach to the analysis. The first is a study by Eisenberg and Miller (2007). They analyze 272 international contracts filed with the SEC in 2002 and find that arbitration clauses are used in only 20% of international agreements. The other is a recent study by Weidemaier (2015) relying on a hand-coded sample of 136 international contracts in the SEC database filed between 2000 and 2012. Weidemaier finds that 61% of those contracts include arbitration clauses. Both studies employ a rigorous process to analyze the contracts at hand, but the results vary widely. There are several potential reasons for this discrepancy. The overall sample size for international contracts is modest and because neither study controls for contract characteristics, it is possible that observed differences in arbitration clause usage are caused by differences in the contracts studied.[12] For instance, joint venture agreements are particularly likely to include an arbitration clause and are disproportionately concluded between parties from different countries. Further, the study by Eisenberg and Miller (2007) is a cross-section of 2002, whereas Weidemaier (2015) tracks usage rates over time. Lastly, Eisenberg and Miller (2007) sample their contract directly from the SEC, whereas Weidemaier (2015) accesses the agreements through Bloomberg Law using search terms and selection algorithms capable of introducing biases.[13]

Overall, it appears that prior empirical studies do not provide clear guidance on the importance of international arbitration, a state lamented by scholars calling for more comprehensive empirical evidence as a precondition to understanding the role of arbitration in today's business environment Born (2014); Drahozal (2016).

---

ments between 1993 and 1996 that have been preselected by the University of Missouri-Columbia's Contracting and Organizations Research Institute. They find that 88% include arbitration clauses, but because it is unclear how the Institute preselects its agreements, extrapolation to any broader population of contracts other than those studied is problematic.

[11] See e.g. Drahozal and Ware (2010), who procure a sample of 31 joint venture agreements submitted in 2008 through LexisNexis and find that 71% of them include arbitration clauses.

[12] While Eisenberg and Miller (2007) run regressions controlling for contract type, those are limited to the full set of contracts which predominantly include domestic contracts.

[13] In particular, instead of sampling from all agreements, a search query subsets agreements to those of a specific type, resulting in only 700 agreements per year from which a sample of 40 is drawn.

1.3 Data & Methodology

The data set studied here is based on all filings of 'material contracts' with the SEC through its electronic filing system EDGAR between 2000 and 2016. The SEC requires registered companies to report every "material contract", which encompasses "[e]very contract not made in the ordinary course of business which is material to the registrant."[14]. Companies registered with the SEC are those that made a public offering or have "total assets exceeding $10,000,000 and a class of equity security (...) held (...) by five hundred or more persons".[15]. The lack of a precise definition of the word "material" provides these companies with some discretion in deciding which agreements to disclose. However, this discretion is limited by general principles established in judicial decisions or administrative guidelines taken into account by the companies.[16] For instance, since the purpose behind this and similar disclosure rules is to remove information asymmetries and allow investors to make informed investment decisions, the SEC staff typically applies the standard established by the Supreme Court in *Basic v. Levinson*[17] when determining whether information falling under a disclosure requirement is "material". Accordingly, materiality implies that "'there is a substantial likelihood that a reasonable shareholder would consider [the contract] important' in making an investment decision."[18] In practice, contracts that meet this definition are often asset and stock purchasing agreements, loan contracts as well as agreements governing the employment and compensation of key employees such as CEOs. SEC staff actively monitors the compliance of companies with the contract disclosure requirement and notifies them if the financial statements indicate an omission.[19]

Companies attach the agreements to their annual reports (Form 10-K), quarterly reports (Form 10-Q) and to reports filed due to important events and changes between quarterly reports (Form 8-K). Similar provisions exist for foreign companies, who have the option to report using Forms 20-F and 6-K. In addition, during Mergers & Acquisitions, the relevant contracts are reported as exhibits to Form S-4. The electronic forms and exhibits are avail-

[14] 17 C.F.R. § 229.601(b)(10)(i)

[15] See Securities Exchange Act § 12(g)

[16] See Correspondence between Marketo, Inc. and the SEC staff about Marketo's procedure on how to determine disclose requirements, available at https://www.sec.gov/Archives/edgar/data/1490660/000110465914004115/filename1.htm (dated January 24st 2014).

[17] Basic Inc. v. Levinson, 485 U.S. 224, 108 S. Ct. 978, 99 L. Ed. 2d 194 (1988).

[18] SEC Release No. 33-7881 (Aug. 75, 2000).

[19] Id.

able for all registered companies through the SEC Electronic Data Gathering, Analysis, and Retrieval system (EDGAR) from its establishment in 1996 to 2016. Because the SEC has continuously changed and extended the filing requirements through EDGAR pursuant to the system's phase-in in 1996, this study is limited to contracts filed between 2000 and 2016, when filing requirements were largely uniform for the forms examined here.[20].

Overall, the data set includes 780,689 agreements. Of those, 272,837 filings are dropped because they are duplicates or mere amendments to already existing contracts, leaving a total of 507,852 unique contracts submitted by a total of 18,641 companies.

Identifying International Agreements

In order to identify which agreements are international and which are domestic, it is necessary to obtain information on the parties of the agreement. EDGAR includes data on the party that made the filing and its industry. The filing party is assumed to be the first party to the contract and its industry is assumed to be the industry pertaining to the contract. A search algorithm based on regular expressions then identifies the paragraph in the contract that includes the parties to the dispute. The algorithm is described in detail in the Appendix. For purposes of illustration, below is one of those paragraphs:

This Note Exchange Agreement (this "Agreement") is made and entered into as of April 2, 2009, by and among (i) Sculptor Finance (MD) Ireland Limited, Sculptor Finance (AS) Ireland Limited and Sculptor Finance (SI) Ireland Limited (the "Existing Noteholder"), (ii) OZ Master Fund, Ltd., OZ Asia Master Fund, Ltd. and OZ Global Special Investments Master Fund, L.P. (collectively, the "Existing Warrant Holders," and together with the Existing Noteholders, the "Holders"), and (iii) Network CN Inc., a Delaware corporation (the "Company").

The program scans this and similar paragraphs for the mentioning of any of the 630,106 companies and individuals that have ever disclosed information through filings with the SEC, as well as the mentioning of foreign companies that are not registered with the SEC by searching for country names in their noun and adjective form. For example, in the paragraph above, the parties under (i) are not registered with the SEC but are found by and associated

[20]For example, in 1999, the SEC allowed submission of filings in HTML format (and the attachment of PDFs), which made filing much easier and is by far the most frequent form of submission today.

with the country of Ireland due the mention of the country in its noun form. The parties' place of incorporation then determines whether the contract is a contract only between U.S. parties (*U.S.–U.S.*), an international contract (*U.S.–Foreign*) or an entirely foreign contract (*Foreign–Foreign*).[21]

Identifying Contract Characteristics

Next, it is necessary to identify whether a given agreement includes a forum selection clause and if so, what type of dispute settlement provision the parties agreed on. Due to the large number of contracts, a machine learning algorithm is required that is able to identify forum selection provisions. To achieve this goal, 5,226 paragraphs are coded by hand for their inclusion of dispute settlement clauses. The paragraphs are then randomly divided into a training set and a test set. Using the training set, a Naive-Bayes classifier[22] is trained to identify words and word-combinations that are most indicative of each type of dispute settlement clause. The classifier is then used to predict the types of paragraphs in the test set, which in turn allows for an assessment of the classifier's performance.

The approach correctly classifies 99.88 percent of the paragraphs. However, the correct classification rate alone can be misleading, since it does not take into account the number of relevant items. For instance, for a test set consisting of 99 irrelevant and 1 relevant paragraphs, a simple algorithm that always considers all paragraphs irrelevant would achieve a correct classification rate of 99 percent. This is why –in addition to the correct classification rate– studies in information retrieval and machine learning use precision, recall, F_1 scores and Matthew's Correlation Coefficients (MCC) to assess the quality of

[21] When the place of incorporation is not available in the SEC database, the location on file is used.

[22] For a thorough examination of the performance of the Naive-Bayes classifier, see Rish (2001). While there are other popular options available, the Naive-Bayes classifier yields the best results.

automated classification procedures.[23] Together, these can be thought of as relative measures of performance that take into account the total number of relevant items. The classifier trained here achieves a precision of 0.89, a recall of 0.94, and an F1-Score as well as a Matthew's Correlation Coefficient of 0.91. It can thus be considered as very accurate, with no strong tendency for false positives or negatives.

A similar process is used to identify whether a contract includes a clause specifying the substantive law governing the contract and if so, which law governs. In a last step, a combination of search terms / phrases and regular expressions is used to identify the type of the document (e.g. employment contract, credit / loan agreement etc.) and the form of the document (e.g. agreement, plan, policy). The entire procedure is described in greater detail in the Appendix.

Summary Statistics

Tables 4.1, 1.2 and 1.3 contain summary statistics describing the data. As can be seen, 10% of contracts are international in nature and only 1% of contracts does not include a U.S. party at all. Overall, 44% of contracts specify some sort of dispute resolution mechanism, where 30% specify that dispute resolution should take place before national courts and 19% opt for arbitration.[24] At the same time, clauses that specify the governing law are very common,

[23]Let TP be the number of true positives, i.e. the number of correctly classified forum selection clauses; FP the number of false positives, i.e. the number of paragraphs that have incorrectly been classified as forum selection clauses when they are not; TN the number of true negatives, i.e. the number of correctly classified paragraphs that are not forum selection clauses; and FN the number of false negatives, i.e. the number of paragraphs that have been classified as not containing a forum selection clause when in fact they do. Then

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision+Recall}$$

$$MCC = \frac{TP \cdot TN - FT \cdot FN}{\sqrt{(FT+FN)(TP+FN)(TN+FP)(TN+FN)}}$$

[24]Note that the dispute resolution mechanisms are non-exclusive. For example, a contract might refer only a subset of issues to arbitration.

appearing in 74% of the contracts. Most contracts in the data set are employment contracts (21%) but international employment contracts are relatively rare (5%). Lease, consulting, employment, licensing and joint venture agreements have a higher propensity to include an arbitration clause than a court selection clause, with joint venture agreements being the contract type most likely to include an arbitration clause (44%) and also with the highest share of international contracts (41%). The agricultural industry is the industry most likely to rely on arbitration, though with only 13 observations, those results are of questionable reliability. Other than agriculture, all industries are more likely to rely on courts than on arbitration.

Figures 1.1, 1.2, 1.3 and 1.4 depict the use of forum selection and governing law clauses over time. First, it should be noted that the use of forum selection clauses overall increases over time, which could indicate an increase in the awareness of the parties for the dangers of leaving the forum in which disputes should be settled unspecified. Second, international (U.S.–Foreign) contracts are more likely to include a forum selection clause than domestic contracts. However, both of these differences are largely driven by trends in dispute resolution through a national court system, where there is both a sizable difference in usage rates between domestic and international agreements as well as an increasing trend over time. For arbitration, the rates between domestic and international contracts are very similar and remained stagnant over the period of examination. If no U.S. party is involved, arbitration is most likely to occur. Lastly, it is somewhat more likely to find clauses specifying the substantive law in international contracts than it is to find them in domestic contracts, though the rate of such clauses remained high for both kinds of contracts over the entire period of observation.

Next, Tables 1.4 and 1.5 indicate which dispute settlement forums are the most prominent in domestic, international and foreign contracts. Note that these numbers represent dispute settlement forums conditional on the parties opting for the respective dispute settlement device (e.g. the propensity to choose New York, given that courts are the forum of choice). New York courts are by far the most popular, with 34% of domestic contracts designating New York as their court forum of choice. Interestingly, with 45% this share is even greater in international contracts. For both domestic and international contracts, Delaware courts are the second most opted for, with California courts third. The other jurisdictions are rarely used. As for arbitration organizations, first note that for international contracts, if parties opt for arbitration, it is usually outside of one of the established arbitration organizations. Surveying a random sample of 100 of those contracts suggests that parties commonly tend to either (1) specify the arbitral proceeding in great detail, including the num-

ber of the arbitrators, the venue and the applicable procedure, thus making the specification of an international arbitration organization partially obsolete; or (2) make a vague remark indicating that their disputes should be solved through arbitration, without specifying anything about the arbitral procedure. Second, while naturally, domestic contracts rely on the American Arbitration Association (AAA) more frequently than international contracts, the AAA is still the most popular established arbitration organization in international and in foreign contracts. The International Chamber of Commerce (ICC), which is often characterized as the most important arbitration organization for international commercial contracts, is relied upon only rarely.

Limitations

Before moving forward, it is important to highlight two limitations of the data set. First, the contracts studied here are part of the filing requirement of the SEC. As such, all contracts in the data set include at least one party with substantial economic ties to the U.S. Similarly, contracts with some sort of relationship to the U.S. are overrepresented in the sample. Even though the data set includes a subset of contracts with only non-U.S. parties, even these contracts should not be understood as being representative of commerce conducted outside of the U.S. Instead, the data is best understood as a representation of how U.S. companies and those with strong economic ties to the U.S. act in a domestic and international commercial context.

A second limitation lies in the fact that all contracts reported here are "material" and outside the "ordinary course of business". As such, they do not represent contracts concluded in the day-to-day business dealings of a company, but only those with significant potential interest to shareholders. Some have argued that this subsets the population of contracts to those contracts that are least likely to include arbitration clauses, because arbitration is predominantly used for transactions of small value (Drahozal and Ware, 2010). However, while it is difficult to ascertain the merits of this claim without data on small stakes contracts, it should be noted that the underlying theory is at least inconsistent with the descriptive statistics presented above. For example, the same theory assumes –by extension– that arbitration is rarely found in M&A contracts due to their character as bet-the-company contracts, which regularly involves very significant economic stakes. Table 1.2 does not support this conjecture, indicating that, with 27%, M&A agreements regularly include clauses referring parties to arbitration. In addition, the contracts studied here are the ones most likely to receive significant care and attention in their design (Eisenberg and Miller, 2007). If the economic stakes are small,

transaction costs incurred by elaborate negotiations can quickly exceed the marginal gain attained by specifying an efficient forum, making it more likely that the issue of specifying a dispute settlement forum is solved through standard clauses or not at all. The observed choice of forum is then less reflective of the parties' preferences towards dispute settlement and more reflective of other factors such as norms, standards and convenience. Nonetheless, it is acknowledged that inferences drawn based on this data set should be cabined to material contracts and more research is necessary to examine whether the findings extend to non-material contracts as well.

1.4 Analysis

The descriptive statistics alone are sufficient to refute the claim that arbitration is the predominant dispute settlement mechanism in either domestic or international commercial contracts.[25] However, they do not directly speak to the motivation for parties to choose between both available instruments. This section supplements the descriptive statistics with additional analyses aimed at understanding when and why arbitration is used in an international commercial context.

1.4.1 Greater Efficiency of the Arbitration Process

As discussed above, one often suggested motivation for parties to rely on international arbitration is its flexibility, which would allow parties to tailor procedures to their individual preferences. Under the assumption that companies choose the forum that is optimal for their needs, the descriptive statistics seem sufficient to refute this conjecture. Given that court clauses are more prevalent than arbitration clauses in both domestic and international agreements, it appears that parties, on average, view courts as the more efficient instrument to settle their disputes. However, one observation that might give pause is the fact that the dispute settlement process is strongly centered around the U.S., with New York being the most popular court forum and the AAA being the most popular arbitration organization. This could indicate a disparity in bargaining power, with U.S. companies successfully imposing their preferred dispute settlement process onto their foreign counterparts. If true, the specific forum choice would then be a mere reflection of U.S. company preferences to

[25] With the mentioned caveat that these are all material contracts with some relationship to the U.S.

litigate most of their disputes in the U.S. However, the observed choice of dispute settlement device would then be a poor proxy for the overall efficiency of the instrument.

In order to investigate the possibility that U.S. companies impose their preferred dispute settlement provisions, I consider the influence that U.S. and foreign companies have on the wording of the dispute settlement clause. In particular, I calculate similarity scores for each company in the data set that represent how similar the forum selection clauses a company uses are. If U.S. companies get to dictate the terms of the forum selection clauses to non-U.S. companies, it should be the case that clauses used by U.S. companies look more similar to one another than clauses used by non-U.S. companies, since U.S. companies get to reuse their preferred clause repeatedly when contracting with non-U.S. companies, whereas non-U.S. companies use forum selection clauses that change with each U.S. counter party. U.S. companies should thus have higher similarity scores than foreign parties.

The similarity of two documents can be measured as the cosine similarity of their respective frequency-inverse document frequency (tf-idf) vectors (Baeza-Yates and Ribeiro-Neto, 1999; Manning et al., 2008). tf-idf vectors are a representation of how important a word is in a document, given its prevalence in all other documents. They can be thought of as a numeric representation of the characteristic terms in a given document using vectors in a high-dimensional space. The cosine between two vectors is a representation of the angle between them, with a small angle between two very similar vectors having a cosine close to 1 and a wide angle between two dissimilar vectors having a cosine close to -1. tf-idf vectors are restricted to the positive occurrence of words, such that the cosine similarity is bounded between 0 and 1, with a value close to 1 indicating a high degree of similarity and a value close to 0 indicating a lot of dissimilarity. To illustrate how the cosine similarity translates into differences in the actual wording of a clause, the Appendix includes two clauses that have a cosine similarity that is close to 1 and two clauses with a similarity of 0.6. The similarity score for each company in the data set is computed by collecting all its forum selection clauses, computing their cosine similarity pairwise and then taking the average over all similarities. If companies do not negotiate forum selection clauses, there should be a substantial amount of companies with similarity scores close to 1. If, on the other hand, dispute resolution clauses are negotiated every time, then the clauses a company uses in its contracts should look largely dissimilar, with most companies receiving a low similarity score.

Table 1.6 compares the difference in the means and distributions of similarity scores for U.S. companies and for foreign companies using a T-test and

a Kolmogorov-Smirnov Test. If U.S. companies impose their clauses, then they should receive higher similarity scores than foreign companies, as U.S. companies get to repeatedly use the same (or a similar) clause in their international contracts, whereas foreign companies have to use the clause offered by their changing U.S. counterpart. As can be seen, neither the means nor the distributions of similarity scores are substantively or statistically significantly different between U.S. companies and those incorporated outside of the U.S.. Hence there is no evidence to support the conjecture that U.S. companies impose their preferred dispute settlement mechanism on foreign parties. Under the assumption that parties are choosing the dispute settlement process that is most beneficial to them, the descriptive statistics thus suggest that, on average, litigation is indeed viewed as more efficient than arbitration.

1.4.2 Concerns for Court Biases

As discussed, many commentators believe that international contracts are more likely to include arbitration clauses than domestic contracts because parties are generally skeptical of another nation's courts. If true, we should be able to observe that parties substitute court provisions for dispute settlement clauses if the the agreement is international. At the same time, they should substitute arbitration clauses for court clauses in domestic agreements. In order to examine whether such a dynamic is at play, I run two separate logit regressions. The first regresses arbitration clause usage on an indicator variable for whether a contract is international. The second regress court clause usage on the same international indicator variable.[26] I then compare the results of both regressions.

Table 1.7 depicts the outcome for arbitration clause usage rates.[27] Models (1) to (3) include time-fixed effects. The first Model does not include any further control variables. Model (2) includes fixed effects for the 11 industries in which the contracts can be concluded, the 15 types of contracts and 8 different formats. Model (3) includes interaction effects between the indicator for international contracts and the industry, as well as the type of the contract. This is in order to allow the rate at which international contracts use arbitration clauses differently from domestic contracts to vary with industry and type. Model (4) includes time not as a fixed effect, but as a numeric variable in order to analyze time trends in arbitration clause usage. It also interacts this linear time trend with the dummy variable on international contracts to

[26]Recall that a contract can include both arbitration and court selection clauses and thus, a single regression will not obtain the relevant quantity of interest.

[27]More detailed tables specifying the fixed effects are included in the Appendix.

analyze whether the gap in arbitration clause usage between domestic and international contracts increased or decreased over time. Foreign contracts with no U.S. party are omitted from the analysis. As can be seen, regardless of the model specification, international contracts do include more arbitration clauses than domestic contracts. However, contrary to popular belief, this is a trend that has not been increasing over time. Indeed, the coefficient on the time trend is negative, suggesting that the use of arbitration has slightly declined over the past years. There also is no statistically significant difference in time trends for domestic and international contracts, suggesting that that both types of contracts used arbitration at a decreasing rate for time.

Table 1.8 depicts the outcome for court clause usage and paints a similar picture. International contracts are significantly more likely to include a clause that refers dispute settlement to the courts than domestic contracts. Contrary to the slight decrease in usage rates over time for arbitration, the use of forum selection clauses referring parties to courts has increased for both for domestic and international contracts, though for international contracts at a slower rate.

In order to make the coefficients in both regressions comparable, Table 1.9 translates the findings into average marginal changes in dispute settlement clause usage rates across all contracts. What can be seen is that the average change in usage rates for arbitration clauses between domestic and international contracts is 4-5%. This difference is more pronounced in the case of court clauses, where the average difference between usage rates is 5-17%, depending on model specification. This finding suggests that parties' primary response to the internationality of an agreement is not the inclusion of an arbitration clause, but that the inclusion of a court clause is at least as common, if not more common. Concerns for home biases thus seem an unlikely motivation for parties to rely on international arbitration.

1.4.3 Enforcement Concerns

A third supposed reason to prefer arbitration over litigation in international agreements is a suggested ease in enforcability of arbitration clauses. If true, then arbitration should be especially relevant in a context that appears particularly challenging to the enforcement of foreign court decisions. At the same time, parties should be more likely to rely on courts in environments that pose no serious risk to the enforceability of a foreign judgments.

To assess the presence of this dynamic, I first subset the data to all international contracts, where each observation is a contract with a country dyad. Each dyad consists of at least one U.S. party and one foreign party (i.e.

U.S.–Canada, *U.S.–France* etc.).[28] The country dyads allow for combining the contractual data with country-specific covariates.

Table 1.10 displays the number of dyads between U.S. companies and selected countries in the data set, together with the rate at which arbitration clauses and court selection clauses occur and how often disputes are settled in U.S. courts. What can be noticed immediately is that arbitration rates vary widely by country, with the Virgin Islands having the lowest rate with 7% of contracts and Ghana having the highest rate with 60%. A second noticeable feature is that litigation occurs almost exclusively in U.S. courts, with the only striking exception being Argentina, where U.S. courts are referred to in only half the contracts. Unsurprisingly, most cross-border relationships to which a U.S. company is a party are concluded with Canadian companies, followed by those incorporated in the United Kingdom and China. Another important aspect of the table is that arbitration is particularly rare in countries that are considered tax havens.

I now complement the data set with two measures of legal institutional quality. The first measure is the World Bank's World Governance Indicator on the Rule of Law. The indicator "captures perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence." (Kaufmann et al., 2009). It is one of the standard indicators for measuring the rule of law in the literature on institutional economics and development (see e.g. Rigobon and Rodrik, 2004; Ginsburg, 2005; Licht et al., 2007; Berggren and Bjørnskov, 2013). Note that the indicator is not only influenced by judicial performance, but also by perceptions of other organs such as the police. To ameliorate potential concerns with how broadly this measure is defined, the analysis is supplemented with a popular indicator for *de facto* judicial independence introduced by Linzer and Staton (2015), measuring the underlying latent quantity of judicial independence that is the subject of the investigation of several other studies up to the year 2012. In doing so, the indicator combines information derived from state departments, expert surveys and objective measurements.

Compiling the data in this way, I regress arbitration- and court-clause usage on the different measures of judicial institutional quality. The underlying rationale is that enforcement concerns should make it likely for companies to rely on U.S. courts if and only if the counter-party is incorporated in a country with a reliable judiciary that does not pose a serious threat to the enforcement

[28] Contracts with more than one foreign party are excluded, since it is not possible to accurately determine country dyads.

of foreign court decisions. If, however, judicial quality poses a serious threat to the enforcement, companies should opt for arbitration instead.

To be sure, both measures for judicial quality used here are not a direct proxy for enforcement probability. In particular, one might contend that a low rule of law score is correlated not only with difficulties at the enforcement stage, but also at the initial trial stage. Thus if companies opt for arbitration in the face of low judicial quality, this could simply mean that these companies do not wish to litigate initial disputes before a dysfunctional court system. However, the previous analysis makes this interpretation unlikely. In particular, what was shown is that parties generally prefer litigation over arbitration. If their only concern was to avoid initial litigation before a dysfunctional court system, parties could simply opt for U.S. courts instead of the dysfunctional court system. However, if they respond to low judicial quality not by choosing courts but by opting for arbitration as a second-best alternative, it would suggest that parties are not worried about the initial stage of litigation, but about consecutive enforcement of the decision.

As mentioned, one important feature of the data is that arbitration is particularly rare in contracts between U.S. companies and those inorporated in tax heavens. These contracts could potentially look very different from the rest of the contracts, given that the contractual partner could closely resemble a U.S. company by substance and be incorporated outside of the U.S. only for tax purposes. All regressions thus control for whether the country of incorporation is considered a tax heaven, where the categorization is adopted from Dyreng and Lindsey (2009). In addition, it is possible that forum choice is a function of the marginal costs of litigation, with companies being more likely to litigate if they share a common legal system. The regressions thus control for the legal system in the country of incorporation. Also included are economic covariates on the country level that are potentially relevant for the amount of cross-border commerce. This information is obtained from the World Bank's World Development Indicators.[29] In particular, the economic indicators include GDP, inflow of foreign direct investments and the current account balance (CBA). Because data on economic indicators is sometimes incomplete, missing values are imputed by multiple imputation based on the EMB algorithm (Honaker and King, 2010), using the Amelia II package for R (Honaker et al., 2011).

All models also control for contract type, industry and format, as well as an indicator for whether a country is a party to the New York Convention. Though desirable, a model specification including country-fixed effects

[29] Available at http://databank.worldbank.org/.

is not informative due to very low within-country variation of the outcome measures that is often caused by changes in the individuals surveyed.[30] The within-country variance can thus not be explained meaningfully. To nonetheless address concerns about unobserved heterogeneity, all model specifications include region-fixed effects.

The results are presented in Table 1.11. As can be seen, arbitration clauses are more prevalent if the foreign company scores low on the rule of law or judicial quality index and less prevalent if it scores high. On the flipside, court selection clauses become more prevalent if the quality of legal institutions is high. This suggests that companies strategically use arbitration as an instrument if problematic legal institutions could endanger the enforcement of a court decision.

To further understand the relationship between institutional quality and forum selection clause usage, Figure 1.6 plots the mean difference between court and arbitration clauses over institutional quality for a typical contract in the data set.[31] Since institutional quality is measured on different scales and is ordinal in nature, scores have been standardized using percentiles, such that 50 on the x-Axis indicates the median rule of law score and judicial quality rating. The graph indicates that for low institutional quality, contracts consistently include more arbitration clauses than court selection clauses. For institutional quality above the 30th percentile, court clauses are the primary forum of choice.

1.5 Discussion

The findings in this paper cast a new light on the role of arbitration, its relevance in an international contractual business environment and its relation to domestic court systems. A widely held belief among scholars of international arbitration is that arbitration dominates the international dispute culture because domestic court systems are ill-equipped to handle disputes between entities of different nations to the parties' satisfaction. As far as U.S. companies and those with economic ties to the U.S. are concerned, this view does not withstand empirical scrutiny, as arbitration clauses are absent in a majority of international contracts between two business entities. In addition, it was

[30] Whereas the between-country variance is 0.94 for rule of law scores and 0.09 for judicial independence ratings, the average within-country variance is 0.02 for rule of law scores and 0.001 for judicial independence ratings.

[31] Numerical variables are centered at their mean, for categorical variables, the most frequent category is used.

shown that parties of international contracts can and do use arbitration clauses strategically when an unbiased court trial or the enforcement of a foreign judgment is called into question due to weak judicial institutions in a parties' home state. However, if the quality of judicial institutions is not in doubt, companies registered with the SEC substitute arbitral proceedings for court proceedings. Together, the evidence provided here can be interpreted as parties treating arbitration not as a one-size-fits-all approach to cross-border challenges, but as a second-best solution to a well-function court system that is primarily of relevance to solve commitment problems related to weak judicial institutions. This result is particularly important given the strong scholarly focus in the literature on international arbitration, while the domestic judiciaries' role in the settlement of international disputes is often neglected.

What might explain the striking and increasing popularity of court selection clauses in material contracts in contrast to arbitration? To understand this phenomenon, it is useful to recall the supposed advantages of arbitration, which is often described to be more flexible, cheaper, faster and staffed with more experts than courts. What is notable is that all these supposed advantages can and have successfully been copied by courts in the wake of increased inter-industry competition. As mentioned initially, many states made a considerable effort to tailor their procedures to the preferences of the commercial world by establishing business courts, allowing for customization of procedural rules and in turn streamlining the dispute settlement process. One particularly striking illustration of the competitive pressure exerted on states by the growing number of arbitration organizations is the 2011 Task Force of New York Law in International Matters Report. The report was conducted to assess how New York can continue to attract international dispute settlement and explicitly warns:

> *It is significant that jurisdictions around the world, many with government support, are taking steps to increase their arbitration case load. New arbitration laws were enacted in 2010 and 2011 in France, Ireland, Hong Kong, Scotland, Ghana and other nations to enhance their attractiveness as seats of arbitration. Maintaining New York's position, which already generates hundreds of millions of dollars in revenues for law firms and related businesses and millions of dollars of tax revenues, and which complements and reinforces New York's position as a center of commerce and finance, requires that attention be directed to the measures discussed in this Report.*[32]

[32]See Hurlock et al. (2011, at 4).

These recommended measures included, among others:[33]

- The specialization of judges in the Commercial Division on international matters
- A "rocket docket" to fast-track disputes for parties that do not wish to make use of the entire array of procedures commonly available
- The possibility for New York courts to make "judicial referee" decisions on matters submitted to them by other courts that involve the interpretation of New York law

In addition to competitive pressures exerted on adjudicative systems, what is often overlooked is that courts, too, have idiosyncratic advantages over arbitration. Some of these advantages can and have been emulated by arbitration organizations. For instance, while Drahozal and Ware (2010) argue that a court system's advantage over arbitration is the possibility for review by a court of higher instance, today, numerous arbitration organizations have established an appellate level that allows parties have arbitration decisions reviewed for legal errors. Among them are the International Institute for Conflict Prevention & Resolution, the AAA's International Centre for Dispute Resolution and JAMS, all of which created optional appeal mechanisms that parties can use if they so desire. This development is still quite recent and it can be expected that other arbitral organizations will follow if the measure proves to be successful. However, other key advantages of courts cannot or have not been emulated by arbitration organizations and it can be assumed that these advantages are a significant contributor to the continuous popularity of courts. In particular, three defining characteristics stand out.

First, many court opinions are published. Courts strive to act consistently in order to appear legitimate. The principle of *stare decisis* even explicitly invokes consistency as grounds for legitimacy. Thus, previous opinions on comparable issues provide parties with a credible signal on how courts would decide a similar legal question in the future. The publication of previous judicial decisions allows parties to update their priors about the outcome of a potential legal dispute, in turn reducing uncertainty associated with a contract (Dammann and Hansmann, 2009). Risk-averse parties are naturally drawn to courts as the instrument with a higher degree of certainty. But even for risk-neutral parties, uncertainty increases the possibility of a legal dispute (Priest

[33]See Hurlock et al. (2011, at 27–28).

and Klein, 1984), making the contract more costly. Arbitration organizations could, in theory, publish opinions in a way similar to general court practice. However, the organizations who have done so are highly selective in the choice of decisions and awards they publish, making it difficult for observers to acquire a coherent set of rulings that would be representative of legal doctrine that helps conditioning one's expectations. And even those published decisions are often so heavily redacted that they include hardly any useful information at all.[34] The arbitral practice is understandable, given that arbitration prioritizes confidentiality of the parties. But even if arbitration organizations were more liberal in their publication practice, it is questionable whether extending the accessibility of previous decisions would help parties update their priors in a meaningful way. Given that arbitrators are not subject to the same legitimacy concerns that courts face and are drawn from a more heterogeneous population of individuals, consistency is not of paramount interest in arbitral proceedings, thus making it more likely that two similar cases come out differently despite the existence of prior case law.

A second important market advantage that many domestic court systems have over arbitral institutions is the cross-subsidy that parties to a dispute receive from the general public (Drahozal and Ware, 2010). Studies indicate that about 20% of the total expenditures by the parties to an arbitration are paid to the arbitration organization and arbitrators.[35] In court proceedings, these costs are born almost entirely by the general public. Indeed, Kakalik and Ross (1983) find that the court fees that parties pay are roughly sufficient to cover court expenditures only if a case is immediately dismissed after it has been filed. Any additional work by the court creates costs that are paid through public subsidies. Given that legal expenditures are of paramount concern of publicly registered companies (Simkin, 2005), these subsidies provide an important advantage when settling disputes in domestic courts over arbitration institutions. While in theory, it is imaginable that arbitration is similarly subsidized, in practice, to most this is normatively undesirable and politically unfeasible. As pointed out above, a court decision creates important positive externalities as the development of an accessible and coherent body of legal decisions allows future parties to condition their behavior (Landes and Posner, 1979). This positive externality legitimizes public expenditures in exchange for a public good. Without the provision of a public good, it appears difficult to justify sizable subsidies for arbitral organizations which exclusively provide

[34]For an example obtained by Westlaw's collection on arbitral awards, see ICC Award No. 10947, attached in the Appendix.

[35]See Jones and Lloyd (2011); Wolrich (2011).

private goods by settling a private dispute between the parties (Ware, 2013).

A third advantage of court systems over arbitral tribunals has traditionally been the possibility for interim relief that arbitration tribunals did not grant in the same way. This is due to the fact that arbitrators have to be appointed prior to making any decision, and this process alone can take a substantial amount of time if the parties cannot initially agree on the arbitrators (Bennett, 2002). It should be noted that, over the past decade, many arbitration organizations have created different forms of emergency arbitration that address these concerns. However, as of today, these emergency instruments have only rarely been used (Savola, 2016) and can thus not be considered well-established.

The benefits discussed above give court systems an important advantage over arbitral tribunals.[36] Because the virtues of arbitration are non-exclusive while it is difficult to emulate the advantages of a domestic court system, it is reasonable that parties view arbitration as a second-best alternative to a well-functioning and efficient court system. However, these findings should not lead one to believe that there is no profound role for arbitration in the international commercial business environment. The fact that arbitration is the predominant instrument to solve disputes where courts are less likely to make and enforce decisions impartially suggest two distinct functions of arbitration organizations. First, they exert competitive pressure on courts to increase their own judicial proceedings in order to retain a high share on the market for dispute settlement provisions, as exemplified by the case of New York. Second, arbitration enables companies from states with weak judicial institutions to nonetheless engage in international commercial relationships they would otherwise be locked out of due to a fundamental distrust in their local courts.

1.6 Conclusion

Arbitration as an alternative dispute resolution mechanism is associated with many hopes and many concerns. To supporters, arbitration offers a cheap, quick possibility to have disputes decided by qualified experts, with a process that is tailored to parties' individual needs. To critics, arbitration creates a quasi-legal, parallel settlement system that foregoes all the positive externalities that the public court systems are associated with. Most importantly,

[36] Scholars have recognized that these are especially pronounced in the context of intellectual property, see O'Connor and Drahozal (2014).

since arbitration decisions are generally private, they do not establish useful precedent that helps parties to condition their behavior.

As this chapter shows, the reality is that the impact of arbitration on the commercial environment is often overstated. Neither for domestic nor for international contracts, the rate at which arbitration clauses are included is particularly high, alleviating both hopes and concerns that arbitration will replace the domestic judiciary in the near future. In those instances where parties rely on arbitration, their choice is motivated neither by efficiency concerns nor by a general desire to avoid litigating before another's domestic courts. Instead, the evidence presented suggests that arbitration has a narrower purpose as a tool that addresses concerns arising out of dysfunctional courts at the enforcement stage.

The overall attractiveness of courts is best explained through states' deliberate effort over the past decades to make their court proceedings amicable to the resolution of international business disputes by offering an efficient, predictable and sophisticated framework. In this way, state courts emulate many of the benefits that are often said to be exclusive to arbitration while retaining the benefits of a heavily subsidized judiciary, high predictability and interim relief. Nonetheless, one should not conclude from these findings that the international business community has outgrown the need for arbitral organizations. That is because, first, arbitration organizations on the market for contracts exert strong incentives on governments to improve the efficiency of their domestic court proceedings in order to retain a significant market share. And second, arbitration allows parties from countries with weak judicial institutions to participate in a market they would otherwise be locked out of.

1.7 Figures and Tables

Table 1.1: Summary Statistics

	Mean	SD	Min	Max	Med	IQR
Year	2008	4.35	2000	2016	2008	7
DRM	0.44	0.50	0	1	0	1
Courts Selection Clause	0.30	0.46	0	1	0	1
Court Clause Length	220	154	29	809	181	196
Arbitration Clause	0.19	0.39	0	1	0	0
Arb. Clause Length	324	245	27	1,128	255	313
Governing Law Clause	0.75	0.43	0	1	1	1
GLC Length	79	77	16	401	47	66
International Contract	0.10	0.30	0	1	0	0
Foreign Contract	0.01	0.09	0	1	0	0
Domestic Contracts	0.89	0.31	0	1	1	0

Summary Statistics for non-categorical variables used in the analysis.

Table 1.2: Agreement Types

Type	Obs	Freq	Int'l	Foreign	Arb.	Courts
Joint Venture	1,399	0.00	0.41	0.01	0.44	0.26
					(0.58)	(0.20)
Licensing	9,431	0.02	0.25	0.01	0.39	0.32
					(0.49)	(0.31)
Employment	108,313	0.21	0.05	0.01	0.33	0.23
					(0.31)	(0.30)
Consulting	7,860	0.02	0.13	0.01	0.28	0.25
					(0.47)	(0.18)
M&A	29,013	0.06	0.09	0.01	0.28	0.43
					(0.20)	(0.61)
Lease	16,076	0.03	0.07	0.01	0.27	0.21
					(0.32)	(0.25)
Transportation	1,313	0.00	0.16	0.02	0.24	0.30
					(0.43)	(0.42)
Sales	89,257	0.18	0.16	0.01	0.22	0.44
					(0.32)	(0.47)
Legal	10,002	0.02	0.16	0.01	0.11	0.30
					(0.13)	(0.32)
Loan	57,086	0.11	0.19	0.01	0.10	0.53
					(0.10)	(0.62)
Security	21,084	0.04	0.16	0.01	0.09	0.44
					(0.11)	(0.49)
Incentives	130,236	0.26	0.03	0.01	0.09	0.12
					(0.19)	(0.26)
Neg. Instrument	14,024	0.03	0.09	0.00	0.05	0.36
					(0.05)	(0.42)
Other	12,758	0.03	0.10	0.01	0.09	0.40
					(0.11)	(0.41)

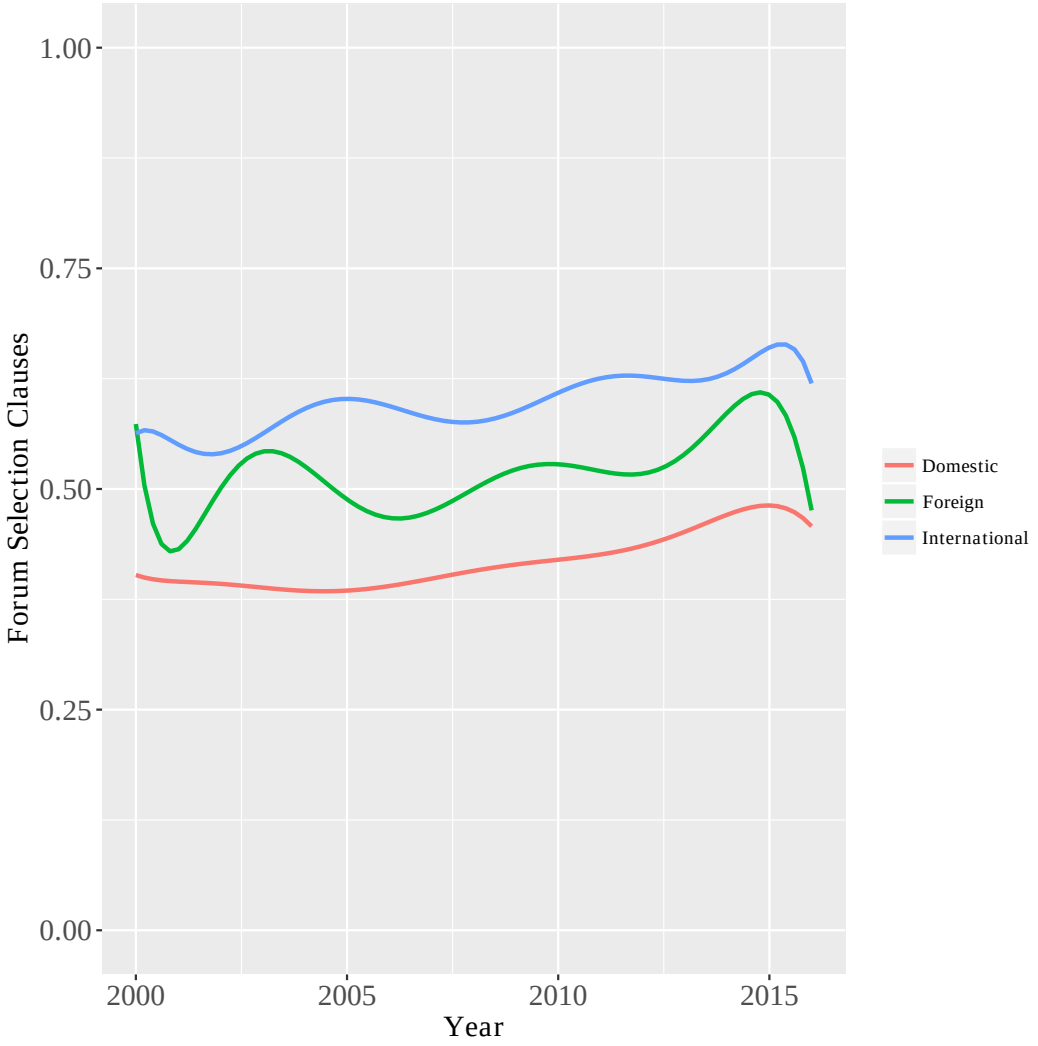
Agreement types in the data set, as well as the frequency of their occurrence, their share of international contracts, of arbitration and of court selection clauses. Statistics in parentheses are based on the subset of international contracts. Sorted by arbitration clause frequency.

Table 1.3: Industries

Industry	Obs	Freq	Int'l	Foreign	Arb.	Courts
Agriculture	13	0.00	0.15	0.00	0.38 (0.50)	0.23 (1.00)
Services	99,596	0.20	0.10	0.00	0.21 (0.24)	0.32 (0.44)
Manufacturing	175,413	0.35	0.11	0.01	0.19 (0.26)	0.30 (0.44)
Finance	100,608	0.20	0.07	0.01	0.19 (0.21)	0.28 (0.44)
Trade	40,671	0.08	0.09	0.00	0.18 (0.21)	0.30 (0.45)
Mining	31,451	0.06	0.15	0.02	0.18 (0.21)	0.32 (0.42)
Transportation	45,472	0.09	0.11	0.01	0.18 (0.20)	0.31 (0.51)
Construction	5,268	0.01	0.07	0.01	0.15 (0.23)	0.30 (0.41)
Other	9,360	0.02	0.12	0.00	0.18 (0.21)	0.33 (0.47)

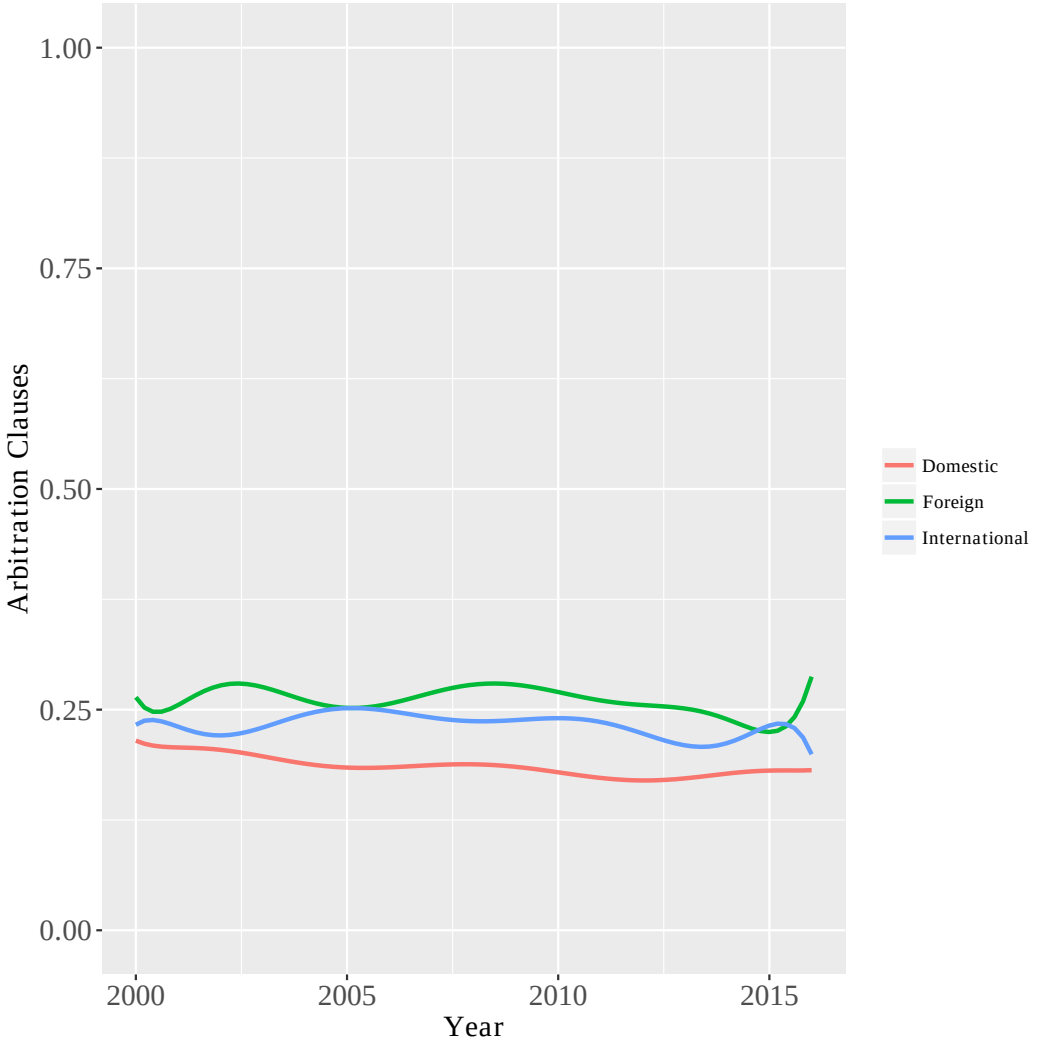
Industries of the contracts in the data set, as well as the frequency of their occurrence, their share of international contracts, of arbitration and of court selection clauses. Statistics in parentheses are based on the subset of international contracts. Sorted by arbitration clause frequency.

Figure 1.1: Forum Selection Clause Usage over Time



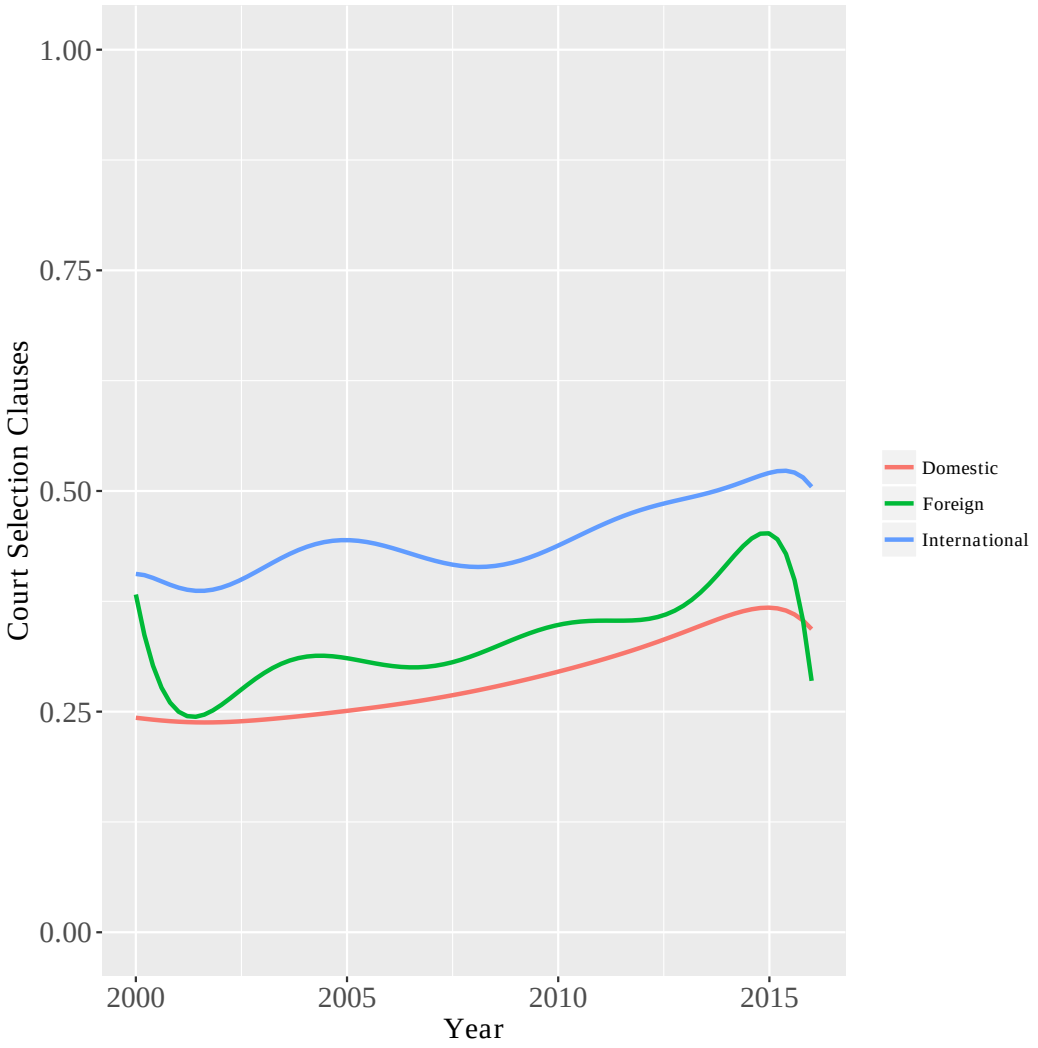
This graph depicts the proportion of contracts that include a forum selection clause over time. The minimum number of contracts per year is 11,489 for U.S.–U.S. contracts, 1,225 for U.S.–Foreign contracts and 42 for Foreign–Foreign contracts. The lines are smoothed using an 8th degree polynomial.

Figure 1.2: Arbitration Clause Usage over Time



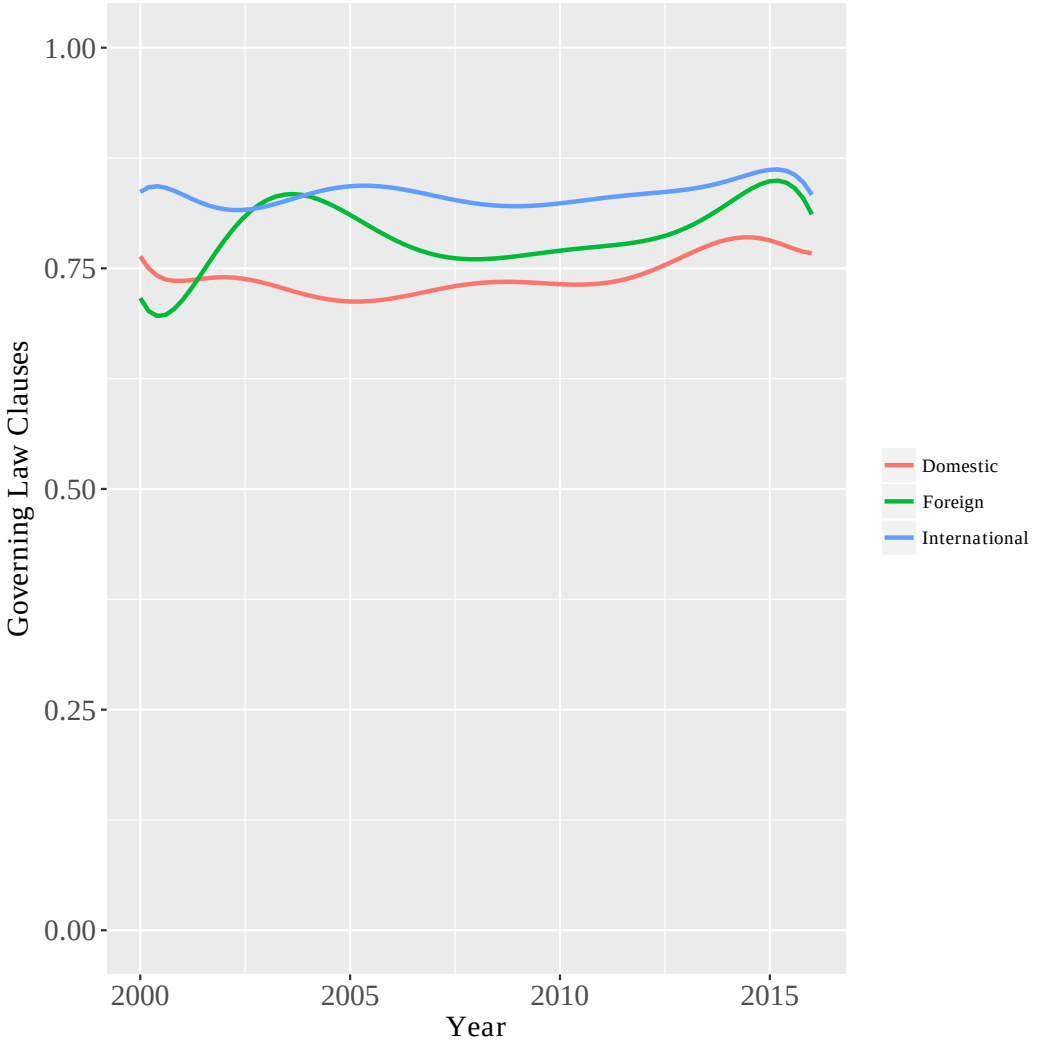
This graph depicts the proportion of contracts that include an arbitration clause over time. The minimum number of contracts per year is 11,489 for U.S.–U.S. contracts, 1,225 for U.S.–Foreign contracts and 42 for Foreign–Foreign contracts. The lines are smoothed using an 8th degree polynomial.

Figure 1.3: Court Clause Usage over Time



This graph depicts the proportion of contracts that include a court clause over time. The minimum number of contracts per year is 11,489 for U.S.–U.S. contracts, 1,225 for U.S.–Foreign contracts and 42 for Foreign–Foreign contracts. The lines are smoothed using an 8th degree polynomial.

Figure 1.4: Governing Law Clause Usage over Time



This graph depicts the proportion of contracts that include a governing law clause over time. The minimum number of contracts per year is 11,489 for U.S.–U.S. contracts, 1,225 for U.S.–Foreign contracts and 42 for Foreign–Foreign contracts. The lines are smoothed using an 8th degree polynomial.

Table 1.4: Most Popular Arbitration Institutions

	Overall	U.S.–U.S.	U.S.–Foreign	Foreign–Foreign
AAA	0.51	0.54	0.28	0.30
JAMS	0.06	0.06	0.04	0.02
ICC	0.02	0.01	0.08	0.06
CIETAC	0.01	0.00	0.05	0.06
LCIA	0.00	0.00	0.01	0.04
HKIAC	0.00	0.00	0.02	0.02
SIAC	0.00	0.00	0.00	0.01
SCC	0.00	0.00	0.00	0.01
Other	0.40	0.38	0.51	0.49

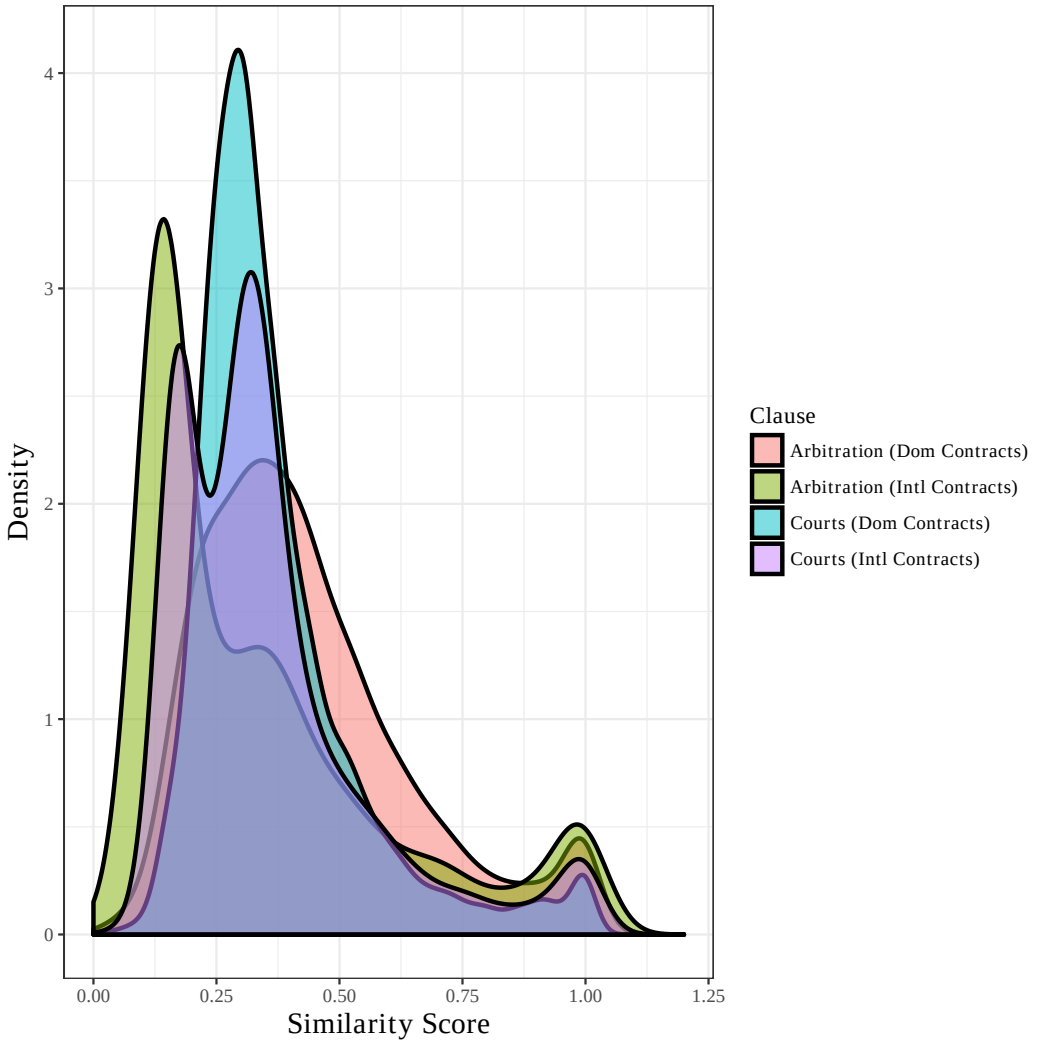
The table depicts the most popular arbitration organizations, conditional on parties opting for arbitration, across all contracts. Full names of arbitration organizations in the Appendix.

Table 1.5: Most Popular Court Forums

	Overall	U.S.–U.S.	U.S.–Foreign	Foreign–Foreign
New York	0.37	0.34	0.51	0.37
Delaware	0.11	0.12	0.09	0.04
California	0.08	0.08	0.05	0.02
Texas	0.05	0.05	0.03	0.03
Florida	0.03	0.04	0.02	0.02
Illinois	0.03	0.03	0.02	0.01
Nevada	0.02	0.02	0.02	0.02
New Jersey	0.02	0.02	0.01	0.01
Massachusetts	0.02	0.02	0.01	0.02
Pennsylvania	0.02	0.02	0.01	0.00
Ohio	0.01	0.02	0.01	0.00
Colorado	0.01	0.01	0.01	0.02
Minnesota	0.01	0.01	0.01	0.00
Georgia	0.01	0.01	0.01	0.01
England*	0.01	0.01	0.02	0.07
Canada*	0.01	0.00	0.03	0.11
Hong Kong*	0.00	0.00	0.01	0.00

The table depicts the most popular courts, conditional on the parties opting for court litigation, across all contracts. Jurisdictions that are used less than 0.5% of the time in international contracts have been omitted.

Figure 1.5: Density Plot of Similarity Scores



This graph depicts a weighted density plot of the company similarity scores for four different forum selection clauses: (1) arbitration clauses in international contracts; (2) arbitration clauses in domestic contracts; (3) court clauses in international contracts; (4) court clauses in domestic contracts. The unit of observation is a company. The density is weighted by the number of observations for each company.

Table 1.6: Significance Tests for Differences between Forum Selection Clauses

	Mean Dom	Mean Intl	T-Test	KS-Test
Arbitration	0.50	0.53	0.14	0.55
Courts	0.51	0.49	0.14	0.37

The p-values in this table relate to the Null-hypothesis that the forum selection clauses domestic and foreign companies use are of equal similarity. More technically, comparing domestic to foreign companies, the Null-hypothesis is that there is no difference with regard to the mean cosine-difference in tf-idf vectors of the companies' respective forum selection clauses.

Table 1.7: Logit-Regression on Arbitration Clause Usage

	Dependent variable:			
	Arbitration Clause			
	(1)	(2)	(3)	(4)
International	0.297^{***}	0.317^{***}	0.337^{***}	0.345^{***}
	(0.011)	(0.012)	(0.046)	(0.052)
Year				-0.005^{***}
				(0.001)
Year*International				-0.001
				(0.003)
Constant	-1.318^{***}	-1.767^{***}	-1.786^{***}	5.281^{**}
	(0.022)	(0.026)	(0.026)	(1.752)
Type-Fixed Effects		✓	✓	✓
Industry-Fixed Effects		✓	✓	✓
Format-Fixed Effects		✓	✓	✓
Time-Fixed Effects	✓	✓	✓	
Interactions			✓	✓
Observations	504,119	504,119	504,119	504,119
Log Likelihood	$-244{,}467$	$-220{,}155$	$-219{,}517$	$-219{,}550$
Akaike Inf. Crit.	488,970	440,401	439,181	439,219
Note:				*p<0.05; **p<0.01; ***p<0.001

The table depicts the estimates for a logit regression of a dummy indicating whether a contract includes an arbitration clause on a dummy indicating whether a contract is an international contract. Standard errors in parentheses. Model (1) includes year-fixed effects. Model (2) additionally controls for type, industry and form of

the agreement. Model (3) includes interaction effects between the dummy for international contracts and the type of agreement, as well as the industry. Model (4) imposes a linear time trend and interacts it with the dummy for international contracts. Other interaction effects are omitted to increase readability. The reference categories for categorical variables are the most prevalent categories. For type, that is Incentives; for industry, it is Manufacturing; for format, it is agreement.

Table 1.8: Logit-Regression on Court Clause Usage

	Dependent variable:			
	Court Clause			
	(1)	(2)	(3)	(4)
International	0.696^{***}	0.220^{***}	0.592^{***}	0.818^{***}
	(0.010)	(0.010)	(0.042)	(0.047)
Year				0.060^{***}
				(0.001)
Year*International				-0.026^{***}
				(0.002)
Constant	-1.126^{***}	-1.901^{***}	-1.916^{***}	-1.939^{***}
	(0.021)	(0.024)	(0.024)	(0.013)
Type-Fixed Effects		✓	✓	✓
Industry-Fixed Effects		✓	✓	✓
Format-Fixed Effects		✓	✓	✓
Time-Fixed Effects	✓	✓	✓	
Interactions			✓	✓
Observations	504,119	504,119	504,119	504,119
Log Likelihood	−304,258	−270,089	−269,666	−219,550
Akaike Inf. Crit.	608,552	540,270	539,479	439,219
Note:				*p<0.05; **p<0.01; ***p<0.001

The table depicts the estimates for a logit regression of a dummy indicating whether a contract includes an arbitration clause on a dummy indicating whether a contract is an international contract. Standard errors in parentheses. Model (1) includes year-fixed effects. Model (2) additionally controls for type, industry and form of

the agreement. Model (3) includes interaction effects between the dummy for international contracts and the type of agreement, as well as the industry. Model (4) imposes a linear time trend and interacts it with the dummy for international contracts. Other interaction effects are omitted to increase readability. The reference categories for categorical variables are the most prevalent categories. For type, that is Incentives; for industry, it is Manufacturing; for format, it is agreement.

Table 1.9: Average Marginal Difference in Forum Selection Clause Usage

	Arbitration	Courts
Model (1)	0.05	0.16
Model (2)	0.05	0.05
Model (3)	0.04	0.07
Model (4)	0.04	0.07

The table depicts the average marginal difference between domestic and international contracts with respect to their usage of forum selection clauses.

Table 1.10: Selected Contracting Countries

Country	Arb Freq	Ct Freq.	Ct In U.S.	# Dyads
Anguilla	0.11	0.63	1.00	27
Argentina	0.25	0.36	0.60	27
Australia	0.24	0.47	0.87	338
Brazil	0.31	0.37	0.89	181
Canada	0.18	0.44	0.98	4395
Chile	0.33	0.33	0.93	61
China	0.42	0.25	0.96	2404
Colombia	0.32	0.33	1.00	48
Cuba	0.27	0.42	1.00	16
Egypt	0.51	0.40	1.00	30
France	0.25	0.41	0.98	577
Germany	0.26	0.42	0.96	734
Ghana	0.61	0.22	0.80	31
Hong Kong	0.33	0.39	0.75	584
India	0.42	0.41	0.95	510
Ireland	0.21	0.51	0.90	495
Israel	0.19	0.39	0.77	420
Italy	0.31	0.36	0.94	198
Japan	0.32	0.39	0.93	661
Mexico	0.27	0.42	0.92	415
Netherlands	0.20	0.50	0.98	665
Russian Federation	0.51	0.29	0.95	72
Singapore	0.26	0.39	0.90	170
Spain	0.28	0.42	0.95	175
Sweden	0.47	0.39	0.93	125
Switzerland	0.20	0.40	0.96	612
United Arab Emirates	0.36	0.18	1.00	14
United Kingdom	0.20	0.53	0.88	2551
Virgin Islands, U.S.	0.07	0.42	1.00	21

The table depicts a selection of contract-dyads between U.S. and foreign companies, their arbitration clause frequency, court selection clause frequency, how often courts within the U.S. are opted for and the number of contract-dyads.

Table 1.11: Logit-Regression of Forum Selection Usage on Judicial Institutions

	Dependent variable:			
	Arbitration Clause		Court Clause	
	(1)	(2)	(3)	(4)
Rule of Law	-0.136^{***}		0.118^{***}	
	(0.036)		(0.033)	
Judicial Independence		-0.423^{***}		0.257^{*}
		(0.110)		(0.103)
NYConvention	-0.263	-0.373	0.201	0.364^{*}
	(0.226)	(0.201)	(0.208)	(0.184)
Common Law	-0.221^{***}	-0.261^{***}	0.322^{***}	0.336^{***}
	(0.050)	(0.053)	(0.044)	(0.046)
Tax Haven	-0.165^{**}	-0.152^{*}	-0.121^{*}	-0.130^{*}
	(0.057)	(0.061)	(0.049)	(0.054)
GDP	-0.843	-1.968	3.208^{*}	6.146^{***}
	(1.555)	(1.641)	(1.368)	(1.464)
FDI Inflow	0.004^{***}	0.004^{***}	0.002^{*}	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Current Account Balance	0.276	0.079	-0.487^{**}	-0.397
	(0.198)	(0.213)	(0.189)	(0.206)
Constant	1.184	1.370	8.767	8.577
	(1.442)	(1.447)	(84.454)	(84.422)
Type Controls	✓	✓	✓	✓
Industry Controls	✓	✓	✓	✓
Format Controls	✓	✓	✓	✓

Continued on next page

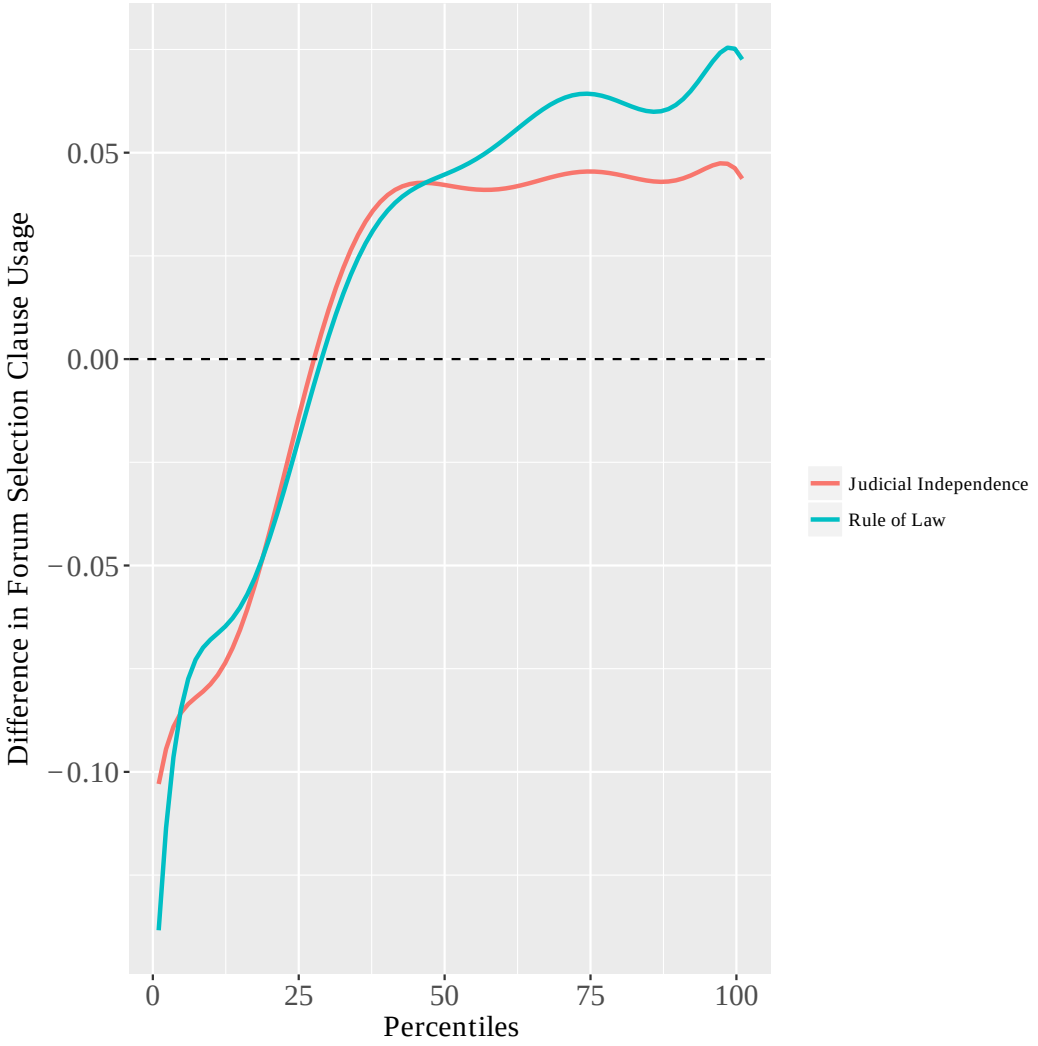
Table 1.11 – continued from previous page

	(1)	(2)	(3)	(4)
Region-Fixed Effects	✓	✓	✓	✓
Time-Fixed Effects	✓	✓	✓	✓
Observations	28,170	25,523	28,170	25,523
Log Likelihood	−14,433	−13,119	−17,555	−15,813
Akaike Inf. Crit.	28,978	26,348	35,222	31,736

Note: *p<0.05; **p<0.01; ***p<0.001

The table depicts the estimates for a logit regression of a dummy indicating whether a contract includes an arbitration or court clause on rule of law scores and judicial independence ratings for the companies' country of origin. Standard errors in parentheses. GDP in mio $, Current Account Balance in trio $, FDI Inflow as percentage of GDP.

Figure 1.6: Forum Selection Clause Usage over Rule of Law Scores



This graph depicts the difference between the probability to include a court selection clause and the probability to include an arbitration clause over measures of judicial quality. Negative values indicate a higher probability to include arbitration, whereas positive values indicate a higher probability to include a court selection clause. The lines are smoothed using an 8th degree polynomial.

Chapter 2

Stickiness of Contractual Gaps: Explaining the Lack of Forum Selection Clauses in Commercial Agreements

2.1 Introduction

When publicly traded companies negotiate contracts, much is at stake. A purchasing agreement can govern the transaction of assets worth multiple billion dollars; a loan agreement can engulf the parties in a creditor-debtor relationship for decades, allowing the lender to exert significant influence on the borrower; and joint venture agreements specify the terms for cooperative endeavors often yielding high profits to the participants. The predominant view is that these contracts of sophisticated commercial actors can best be understood through the lens of utilitarian contract theory (Hart and Holmstrom, 1986; Posner, 2004, 2014), with parties writing optimal agreements that maximize their joint surplus. Legal scholars commonly apply this framework to the analysis of contract negotiation and design between sophisticated parties (see Schwartz and Scott, 2003) and in particular have applied it to the negotiation of forum selection clauses between sophisticated parties (see e.g. Shavell, 1995; Slottje, 2006; Dodge, 2011). The Supreme Court goes even one step further. In *Bremen*,[1] it contends that "(...) it would be unrealistic to think that the parties did not conduct their negotiations, including fixing the monetary terms, with the consequences of the forum clause figuring prominently in their calculations."[2] This expresses the courts belief that parties would carefully consider the implications of individual clauses and even trade them off against seemingly unrelated provisions such as the price in a purchasing contract.

Without a doubt, utilitarian contract theory has made great and significant contributions to our understanding of how sophisticated companies should design contracts in order to maximize the joint surplus of their interactions. However, whether and when commercial actors factually do design agreements in such a manner remains unclear.

This chapter joins the studies of a group of corporate legal scholars in arguing that many commercial contracts are deficient. The deficiency is not the inevitable result of a careful balance of costs and benefits, but rather stems from agency problems between companies and the law firms that represent them. In particular, when law firms draft contracts, they rely heavily on the provision of templates and making changes to these templates for the good of the client is an exception rather than the norm. Using the example of forum selection clauses which are absent in more than half of all material commercial agreements reported to the SEC, this study shows that a gap that existed in one contract is highly likely to persist in future iterations. Though

[1] M/S Bremen v. Zapata Off-Shore Co., 407 U.S. 1, 92 S. Ct. 1907, 32 L. Ed. 2d 513 (1972)

[2] Id. at 14.

through monitoring, internal counsel could guarantee that each agreement included a forum selection clause at relatively low cost, it seems that company employees are hesitant to get involved in the drafting process unless a firm makes significantly negative experiences with their current practice.

The results contribute to the literature on the economics of contract design and the role of the legal profession in several aspects. First, the stickiness of contractual gaps implies that the initial distribution of rights through default rules can be important for the final allocation of the surplus even when the transaction costs are negligible, a finding that stands in contrast to implications drawn from the popular Coase theorem. Second, the results contradict the claim that a clauses' prevalence easily translates into its efficiency, an assumption often encountered at least implicitly in empirical studies on contract design. And third, this study adds to the growing body of literature that emphasizes the importance of the law firm's role in the allocation of contractual rights.

The chapter proceeds in six parts. After this introduction, the next section provides the institutional and theoretical background underlying this study, considers whether leaving the forum unspecified is a rational gap and develops several hypotheses. Section 2.3 introduces the data set and the methodology used. Section 2.4 presents the results and Section 2.5 discusses their implications. A last section concludes.

2.2 Theory

In both the legal and the economic discipline, it is a widely held view that sophisticated actors draft contracts in ways that maximize the joint surplus of the parties. But of course, negotiating each term individually in every contract is not always feasible due to possibly high transaction costs, which leads to the creation of standardized agreements. It is often assumed that these standardized, or "boilerplate", agreements evolve over time, with unfavorable clauses being removed and favorable clauses added, such that any time-tested standard contract is efficient from the perspective of the party that proposes it (Smith and Warner, 1979).

A small group of corporate legal scholars challenges this view. Through a series of empirical studies focusing mostly on covenants in corporate and sovereign bonds, they show that many high-value contracts are not written individually and independently from one another in a way that would reflect a careful consideration of individual terms by the parties. Instead, drafters use templates that they are hesitant to deviate from. This effectively results in

a development process that is highly path-dependent and resistant to change. Rather than trying to write the optimal agreement that can be realized under given resource constraints, most contracts follow a "good enough" approach where contracts that seem to have worked in the past are recycled repeatedly (Choi et al., 2013; Richman, 2011; Anderson and Manns, 2016). Scholars refer to this as the "stickiness" of contract provisions.

There is significant evidence to show that this dynamic can ultimately lead to the adoption of suboptimal agreements. For instance, Klein et al. (1993) analyze call provisions for corporate bonds, which allow the issuer to repurchase the bond at a prespecified date. The authors find that a complex provision capable of optimizing incentives and bond prices is foregone in favor of a simpler rule that tends to overprice the embedded call option. Kahan (1995) analyzes anti-dilution provisions, which are intended to protect holders of convertible securities who have the right to change their investment into a common stock. The value of the right to convert naturally depends on the value of the common stock. Companies can use different measures to devalue common stock without devaluing the common stock holders' investments, e.g. by paying a dividend. Anti-dilution provisions are intended to protect against such redistributions of wealth, but Kahan finds that the clauses are often boilerplate provisions that do not provide adequate protection. Kahan and Klausner (1997) examine 101 investment-grade bond issues with event risk covenants, which are clauses designed to provide remedies to bondholders for declines in the value of their bonds caused by predefined events, usually related to takeovers. They find that standardizations of the convenants lead to a suboptimal compensation scheme that only 17% of contracts improved upon.[3]

Scholars in the late 1990s believed that the stickiness of suboptimal covenants can ultimately be rationalized through the economics of networks. That is, contract standardization, even if suboptimal, would provide positive network externalities that create "lock-in" effects, making deviations from the norm prohibitively costly (Kahan and Klausner, 1997). In this sense, deficient indentures were characterized as an almost naturally evolving suboptimal equilibrium that no single issuer has an incentive to deviate from, though efficiency gains could be realized through collective action. However, a recent book by Gulati and Scott (2012) challenges this rational choice explanation. Through a comprehensive study including over 200 interviews and 1,500 sovereign debt contracts, the authors analyze the evolution of *pari passu*, a clause included in

[3]In particular, the covenants gave investors a "put at par" option, allowing them to sell their bonds at the market rate without factoring in changes in the bond value due to changes in market interest or credit risk rates.

virtually all sovereign debt contracts. The authors show that law firms who included the clause in their contracts generally did not think it had any particular meaning in the debt context. However, in fact, the clause poses a significant and unintended litigation risk.[4] Traditional theory would have suggested that, after its harmful potential was revealed, the *pari passu* clause would be rewritten to prevent adverse effects in the future. Instead, the authors find that 90% of contracts continued to include the clause without making any changes to it.[5]

Gulati and Scott (2012) test different hypotheses about stable, suboptimal equilibriums such as network externalities and learning effects that might explain their results, to no avail. Instead, the interviews reveal great frustration of some lawyers over an increased commoditization of contract drafting that does not leave much room for alterations and does not familiarize young attorneys with the intricacies of drafting clauses from the ground up. In addition, the authors provide convincing evidence to support the claim that lawyers are extremely risk averse and susceptible to herd behavior, favoring conservatism with respect to the drafting process. These characteristics are particularly common among junior associates, who work under the belief that a standard, time-tested contract must be efficient and that any apparent deficiencies must be rooted in their lack of experience and understanding of the matter.

The combined evidence presented in the literature to this date has demonstrated that the existence of time-tested standard clauses creates a stickiness that is able to cement a suboptimal clause even in agreements drafted by the most sophisticated commercial actors. However, it remains the subject of debate whether this stickiness is the product of a rational process characterized by learning and network effects or whether risk aversion and herd behavior is responsible for the rigidity often found in the provisions. To illuminate this issue and extend upon previous research, this chapter examines whether contractual gaps are as resistant to change as time-tested standard clauses.

[4]More specifically, when restructuring their debt, countries can propose a change to the lending terms. Even though these changes benefit the borrowing country, creditors often accept in hopes to recoup some of their investment. *pari passu* can be invoked by a small group of holdout investors who do not accept the new terms in order to prevent the sovereign from repaying the larger group of creditors that have accepted the new terms. In effect, holdout investors are able to force countries into either paying large returns on their investments or to default. The strategy was successfully employed by the hedge fund *Elliott Management Corporation* in Peru and Argentina.

[5]To be sure, lawyers do make numerous edits to the templates they rely upon. However, Anderson and Manns (2016) describe a majority of these edits as "editorial churnings" that, while increasing billable hours, do not change the substance of the agreement in a meaningful way.

Specifically, it analyzes the stickiness of using or not using forum selection clauses in commercial contracts. Studying gaps in forum selection clause usage is particularly informative for several reasons.

First, a key advantage of this study is that it overcomes the lack of generalizability characterizing previous inquiries. Specifically, note that past inquiries analyze publicly issued corporate or sovereign bonds. Though both bond issuers and holders can be large and sophisticated financial actors, the bond indentures for publicly issued bonds are rarely the result of a traditional bargaining process. Instead, bond issuers and underwriters cement the indentures while bondholders do not participate directly.[6] While underwriters have an incentive in creating marketable bonds, they are also interested in preserving their relationship with the issuer, who wants to minimize constraints on the companies' or governments' future conduct. As such, bond indentures typically start with terms strongly favoring the issuer and amendments are made in favor of bondholders only to the degree necessary to ensure marketability (Riger, 1991). But since bondholders tend to neither read nor price in bond indentures,[7] there is little discipline imposed by the market to ensure terms that balance the interest of the stakeholders.[8] The absence of a traditional bargaining process makes the study of bonds a special case of contracts that allows issuers and underwriters to simply reuse old indentures with minimal constraints or market discipline, making path-dependence and rigidity in contract terms especially likely. In contrast, studying forum selection clauses in a broad range of commercial contracts between economic actors provides more generalizable insights that inform our understanding in the broader set of contracts that are the outcome of a traditional bargaining process between the contracts' stakeholders.

Another aspect that makes bond indentures, specifically for corporate bonds, especially sticky and conclusions drawn from their analysis difficult to generalize is the existence of several model indentures which are widely used across the industry. The American Bar Association has published the ABF Model

[6] "[T]he holders of public bond issues (...) often enter the market after the indentures have been negotiated and memorialized. Thus, those indentures are often not the product of face-to-face negotiations between the ultimate holders and the issuing company. (...) [U]nderwriters ordinarily negotiate the terms of the indentures with the issuers.", Metro. Life Ins. Co. v. RJR Nabisco, Inc., 716 F. Supp. 1504, 1509 (S.D.N.Y. 1989)

[7] A former seller of bonds tells a colorful story about the rarity with which investors complained about covenants, typically to no avail. See https://www.bloomberg.com/view/articles/2017-01-12/bond-covenants-and-skeptic-skepticism.

[8] To be sure, the market begins to react to this deficiency. For example, in 2012, Moody's began to score the covenant quality of high-yield bonds.

Debenture Indentures (1965), the ABA Model Simplified Indenture (1983, revised in 2000) and the Model Negotiated Covenants and Related Definitions (2006). It is believed that the model indentures provide a widely used template across the industry,[9] again increasing the probability for sticky covenants to evolve. In contrast, the vast majority of contracts does not evolve out of an industry-wide model agreement, making results of the contracts under study here more representative and generalizable.

It should also be mentioned that studying bond indentures means studying one of 'the most involved financial document[s] that has been devised' (Kennedy, 1961, 1). The covenants which are the subject of previous studies typically deal with complex issues that do not only require knowledge of the relevant legal rules, but also a significant level of expertise in the relevant financial market dynamics and incentive effects (Klein et al., 1993). The impenetrability of the underlying legal issues makes it especially likely for suboptimal rules favoring the issuer to emerge, given that most investors neither fully process, nor have an incentive to invest in identifying how each covenant may affect their return or the default risk.

The lack of a traditional bargaining process, the existence of widely used templates and the high degree of complexity make the conclusions drawn from the previous literature difficult to generalize, thus leaving open whether contracts in other areas that are the product of negotiations between equals over simpler terms are subject to a comparably path-dependent drafting process. The study of forum selection clauses addresses these limitations. By analyzing a broad range of corporate agreements across multiple issue areas, it provides a picture of how contracts are written outside of the area of bond issuances, allowing to test whether rigidity is a characteristic of contract provisions more generally or whether it is specific to certain issue areas. In addition, choice-of-forum clauses lie at the core of legal expertise and touch upon an issue that is comparatively simple to comprehend and taught in every first year law school curriculum. Hence finding deficiencies in forum selection clause usage would make for an especially compelling case of suboptimal contract drafting.

Another advantage of this study is that the analysis of contractual gaps significantly reduces the number of potential explanations for supposedly deficient design. Previous studies focused on the wording of a covenant and how it relates to the presumed goal of the contract, concluding that commercial actors are incapable of optimizing the wording of a clause. But choosing the

[9] "The 1983 MSI and the 1983 Notes were promulgated with the hope that having a common form for the most standard provisions of indentures would reduce the need for significant negotiation of such provisions, and, in large part, the 1983 MSI accomplished that objective.", Revised Model Simplified Indenture, 55 Bus. Law. 1115 (2000).

optimum wording of contractual language is a choice from a space with virtually infinite alternatives. Trying to find the optimum choice among a great number of alternatives in such a setting quickly becomes economically infeasible, incentivizing actors to settle for contract terms that are good enough to achieve their goal without the need to optimize the text, a decision-making process also known as "satisficing" (Bolton and Faure-Grimaud, 2010). In contrast, this study focuses not on the efficient wording of the a clause, but on its inclusion.

Focusing on the inclusion of simple contractual terms rather than the wording of a particular clause drastically reduces the choice space. Indeed, contractual gaps can generally be explained in only two ways. Either they are rational, such that closing the gaps through negotiating and drafting a choice-of-forum clause is more costly than going with the default rule; or they reflect an omission that is not in the best interest of the client and thus should be filled. The concept of satisficing is an unsuitable explanation for the existence of gaps, as parties should have clearly defined preferences over the inclusion or non-inclusion of a clause.

The rational explanations provided by scholars in the 1990s are similarly unsuitable explanations for the existence of contractual gaps. In particular, parties who do not include a choice-of-forum clause can neither gradually improve upon the clause, nor can they feasibly be described as any coherent network. That is because–as will be discussed momentarily–the default rule is imprecise and gives parties little control or predictability in a way that would allow them to benefit from one another. Thus, network or learning effects are unlikely explanations for the occurrence of the gaps studied here. If it can be shown that law firms consistently fail to use forum selection clauses and that this omission is not the result of a careful balance of costs and benefits, this is can be seen as compelling evidence for the existence of agency problems allowing those writing a contract to not act in the best interest of those who are bound by its terms.

2.1 Omission of Forum Selection Clauses: A Rational Gap?

A study by Eisenberg and Miller (2007) has revealed that there is a distinct lack of choice-of-forum clauses in material commercial contracts reported to the SEC in 2002, with only 39% of agreements including such a clause. The previous chapter confirmed the high frequency of this gap in an analysis of all contracts filed between 2000 and 2016. The lack of forum selection provisions

is especially striking when comparing their usage rate (44%) to that of clauses specifying the substantive law governing the contract (75%), given that both address similar concerns arising out of uncertainty over the regulatory system that will ultimately govern the contract. As pointed out, these findings motivate the question of whether leaving the forum unspecified constitutes a rational gap. This can be determined by comparing the costs of negotiating and drafting a choice-of-forum provision to the costs of leaving the forum unspecified.

Leaving the dispute settlement forum unspecified can produce different types of costs, which is caused by the fact that default rules in the absence of forum selection clauses provide considerable uncertainty. If the parties do not include a forum selection clause in their contract, then the plaintiff can sue the defendant in any court that has personal jurisdiction over the defendant. Rules by which courts can exert personal jurisdiction in any given dispute are not conclusive and overall lack clarity, especially in the period under study here. Nonetheless, one can try to formulate a few broad principles that apply to company contracts of the type analyzed here.

Courts exert jurisdiction over a defendant either based on the principles of *general* or of *specific* jurisdiction. A court that has general jurisdiction over a defendant can hear any case against that defendant, whether it arises out of a contract, product liability, property or others. Courts all over the country have long differed in the level of intensity of the relationship between a company and the state that is sufficient to establish general jurisdiction. The most expansive view is expressed in the widely used "doing business" test. Under that test, it is sufficient for a company to do business "with a fair measure of permanence and continuity" in a state in order for the courts in that state to exert general jurisdiction.[10]

The "doing business" test was the most prevalent view held by the courts until 2011. The period from 2011 until 2017 is characterized by much turmoil, when a series of Supreme Court decisions[11] began to cabin the expansive "doing business test", instead establishing a narrower "essentially at home" test. This test limits general jurisdiction to the state of incorporation and the state of principal business and allows courts to extend upon those only under

[10]See e.g. Tauza v. Susquehanna Coal Co., 220 N.Y. 259, 115 N.E. 915 (1917).

[11]Goodyear Dunlop Tires Operations, S.A. v. Brown, 564 U.S. 915, 131 S. Ct. 2846, 180 L. Ed. 2d 796 (U.S. 2011); Daimler AG v. Bauman, 134 S. Ct. 746, 187 L. Ed. 2d 624 (2014); BNSF Ry. Co. v. Tyrrell, 137 S. Ct. 1549, 198 L. Ed. 2d 36 (2017); Bristol-Myers Squibb Co. v. Superior Court of California, San Francisco Cty., 137 S. Ct. 1773, 198 L. Ed. 2d 395 (2017)

"exceptional" circumstances.[12] Initially, the Supreme Court's attempt to limit general jurisdiction was met with hostility by some states. For instance, New York courts began to more frequently invoke an old doctrine[13] by which companies registering in New York are assumed to implicitly subject themselves to the general jurisdiction of New York courts. Their efforts were supported by the legislature, which proposed a bill codifying the doctrine by making consent by registration explicit in the law.[14] Similarly, California courts tried to circumvent the "essentially at home" test by expanding specific jurisdiction to a degree that resembles the broad general jurisdiction doctrine.[15] However, the Supreme Court firmly rejected all of these attempts, such that since 2017, it appears settled that the "essentially at home" test is the only test by which general personal jurisdiction can be established.[16]

Specific jurisdiction can be exerted if a defendant's actions in a state give rise to the specific claim at hand and is limited to that claim. For contract disputes, the Supreme Court established that the contract itself as well as "the prior negotiations and contemplated future consequences, along with the terms of the contract and the parties' actual course of dealing, must be evaluated to determine whether a defendant purposefully established minimum contacts"[17] with the state in which the plaintiff sues. This test has led to considerable confusion among courts, with some emphasizing the parties' bargaining power, some considering whether the out-of-state party is the buyer or seller and again others focusing on the "passive" or "aggressive" nature of the litigants in order to determine specific jurisdiction (see Stephens 1986). However, in cases where it is unchallenged that a contract between the parties exists and the only question is how to interpret the contract's terms, some courts show a

[12]Daimler AG v. Bauman, 134 S. Ct. at 761, 187 L. Ed. 2d 624 (2014).

[13]Bagdon v. Philadelphia & Reading Coal & Iron Co., 217 N.Y. 432, 111 N.E. 1075 (1916).

[14]At the time of this writing, the bill was ordered to its third and last reading and could become law in the near future, see A06714. Assemb. Reg. Sess. 2015-2016 (N.Y. 2016).

[15]Bristol-Myers Squibb Co. v. Superior Court, 1 Cal. 5th 783, 377 P.3d 874 (2016).

[16]There are a number of post-*BMS* decisions to show that courts all over the country reject the concept of consent by registration, *see* e.g. W. Express, Inc. v. Villanueva, No. 3:17-CV-01006, 2017 WL 4785831 (M.D. Tenn. Oct. 24, 2017); Gulf Coast Bank & Tr. Co. v. Designed Conveyor Sys., L.L.C., No. 17-30062, 2017 WL 6553374 (5th Cir. Dec. 22, 2017); Fundamental Innovation Systems International LLC v. LG Electronics, Inc. et al, No. 2:2016cv01425, 2018 WL 279091 (E.D. Tex. Jan. 3, 2018); Travelers Prop. Cas. Co. of Am. v. Hume Lake Christian Camps, Inc., No. 17-CV-1600 JLS (KSC), 2018 WL 280025 (S.D. Cal. Jan. 3, 2018); Harter v. Ascension Health, No. CV-15-00343-TUC-RM, 2018 WL 496911 (D. Ariz. Jan. 22, 2018).

[17]Burger King Corp. v. Rudzewicz, 471 U.S. 462, 463, 105 S. Ct. 2174, 2177, 85 L. Ed. 2d 528 (1985)

tendency to emphasize the place of performance over all other factors (Rhodes 2005).

What can then be taken away from this description of the default rule is that it induces much uncertainty. In contracts between large public companies in which the place of incorporation, the principle place of business and the location of performance often diverge, parties that do not specify the dispute settlement forum potentially open themselves up to a multitude of court forums, making it difficult to foresee which court(s) will ultimately hear the dispute.[18] In the worst case, parties litigate the same contract in different forums at the same time. The most popular example of this type of litigation is multi-forum shareholder litigation, especially after mergers, a practice often criticized by academics and practitioners alike that affects companies that do not have an exclusive forum selection clause in their charters (see Armour et al., 2012; Micheletti and Parker, 2012; Romano and Sanga, 2017). But even outside of the realm of shareholder litigation, uncertainty about the applicable forum can produce a number of additional costs and in turn diminish the welfare gains.

Costs for Settling on a Forum

Perhaps the most obvious costs resulting from omitting a forum selection clause are the costs caused by additional litigation over the jurisdiction. Consider the case of *Abbott Laboratories v. Takeda Pharmaceutical Company LTD*,[19] a case in which the parties –both of which are publicly traded, large pharmaceutical companies– negotiated a contract including two forum selection clauses, one in Japan and one in Illinois. Because the choice of which clause applies would grant a significant home turf advantage to one of the parties, the issue of forum selection was seen as an important predictor of the outcome of the trial. It was fought over fiercely by the parties, initiating a separate lawsuit over the question of forum selection alone that would take almost two years to resolve, consuming a substantial amount of legal resources on the way.

As the example illustrates, challenging the forum can be quite costly. A rational party will weigh the costs of challenging a forum against its expected gains. Challenges thus are most likely where adjudicators in the challenged forum and in the targeted forum are expected to come to widely different

[18]While it is possible for a defendant to bring a motion to transfer in an attempt to change court forums, the judge applies a balancing test to determine whether the motion is granted or not. See e.g. 28 U.S. Code § 1404(a) for the federal judiciary.

[19]Abbott Labs. v. Takeda Pharm. Co., 476 F.3d 421 (7th Cir. 2007).

conclusions.[20] Note, however, that any challenge to a court forum produces avoidable costs. Indeed, any post-challenge outcome can be achieved cheaper by settling on the forum that will finally hear the case *ex ante* through the inclusion of a forum selection clause.

Incentive Costs

The choice of forum has important effects on the parties' incentives to perform with the contractual terms. Companies that routinely fail to specify their preferred forum miss an important opportunity to incentivize an optimal level of performance (Shavell, 1995; Hylton, 2000; Drahozal and Hylton, 2003; Dodge, 2011). To develop an intuition for this result, note that the parties' incentive to breach a contract is a function of the costs they face if they breach. These costs generally come in the form of dispute settlement expenses and damages awarded by the court. Both expected dispute settlement expenses and damages vary from one jurisdiction to the other. This is certainly true for the difference in expenses between litigation and arbitration, provided that only in arbitration, parties bear the full costs of their disputes. Indeed, studies indicate that about 20% of the total costs of complex arbitral proceedings are paid to the arbitration institution and the arbitrators,[21] an amount largely subsidized by the public in the domestic court system. But even within forums of a particular type, costs can vary substantially. For example, most corporate legal firms have a significant presence in and familiarity with the courts of New York, lowering the costs for disputes litigated in the state, compared to litigation in a state corporate lawyers are much less familiar with. Further, different states have different procedural laws which in turn alter their costs. For example, it is well known that civil jury trials on average take twice as long as bench trials (Posner, 2014; Kakalik and Ross, 1983; Sipes and Oram, 1988), but that not all states enforce jury waiver clauses, potentially exposing parties to longer and more costly litigation.[22] In addition to dispute settlement

[20] e.g. due to different views on the type of damages that the defendant ought to pay for a contractual breach or due to differences in standards of proof which can make it either hard or easy to substantiate a claim.

[21] For survey data, see Jones and Lloyd (2011) and Wolrich (2011).

[22] In North Carolina, jury waiver clauses are unenforcable by statute, see N.C. Gen. Stat. §22B-10. California as well as Georgia courts often hold them unconscionable as a matter of common law, see Grafton Partners, L.P. v. Superior Court, 116 P. 3d 479 (2005); 264 Ga. 339, 444 S.E.2d 799 (1994); In re County of Orange, No. 14-72343 (9th Cir. April 16, 2015). Even those states that enforce jury waivers often invoke a presumption against the enforcability of a waiver, limiting enforcement to those clauses that are narrowly construed (Posner, 2004).

costs, damage awards can also vary with the forum of choice. Again, the most significant difference exists between courts and arbitration, where some evidence suggests that arbitrators might be susceptible to granting awards that 'split the baby' in order to maximize their chances of reappointment.[23] Even within the domestic judiciary, awarded damages can differ, e.g. due to a difference in the interpretation of vague contractual terms relevant for the determination of whether a breach occurred (Drahozal and Hylton, 2003).

Parties that choose their forum have the possibility to optimize the incentives provided in order to guarantee that a contract is only breached if it is efficient to do so. Parties that do not agree on a forum forego this possibility, allowing plaintiffs to unilaterally choose court forums that are particularly favorable to her claim. Whether the expected dispute settlement expenses and damages awarded by the court chosen by the plaintiff unilaterally exceed those awarded by the court that is chosen *ex ante* by mutual agreement cannot be determined generally. On one hand, it is evident that the plaintiff will have an interest to choose a forum that is particularly favorable to her claims. On the other hand, not choosing the forum *ex ante* significantly diminishes the set of jurisdictions that the plaintiff can sue in absent consent by the defendant, such that the plaintiff's options are severely limited. However, what should be noted is that only in exceptional circumstances will the plaintiff's choice of jurisdiction provide efficient incentives to the defendant. In all other cases, the defendant may be over- or underdeterred, leading to an expected welfare loss for the contractual parties. The Appendix formalizes this dynamic, presenting an extension to the standard model of *ex ante* choice of forum provided by Hylton (2000) and Drahozal and Hylton (2003).

Predictability of the Courts

It is well known that two parties with identical information and the same expectation about the outcome of a legal proceeding have no reason litigate, as they could achieve a similar outcome through negotiations without incurring substantial litigation costs (see e.g. Cooter and Rubinfeld, 1994). Hence, in forums that are highly predictable, parties can be expected to initiate fever lawsuits than in forums for which the outcome of a lawsuit is difficult to predict. Parties who forgo the opportunity to specify the forum in advance subject themselves to the threat of having to litigate in a less predictable forum. For instance, parties who intend to enforce an interpretation of a clause in the contract that lies outside of its four corners may be inclined to nonetheless

[23]See Farber and Bazerman (1984) and Dammann and Hansmann (2009). However, for contradicting evidence see Weber et al. (2014).

pursue their claim in contextualist jurisdictions which take a liberal approach to the introduction of extrinsic evidence, increasing the risk of litigation.

In the Appendix, I demonstrate empirically that jurisdictions vary widely in their propensity to deter litigation. Consistent with much of the theory (see e.g. Romano, 1987; Fisch, 1999; Slottje, 2006),[24] New York and Delaware are among the most predictable jurisdictions. In addition, the previous chapter showed that these are also the jurisdictions most often opted for in forum selection clauses. Overall, these findings lend credence to the notion that forum selection clauses are primarily used to move disputes into predictable jurisdictions, effectively lowering the risk of litigation.

Loss of Competitive Advantage

Lastly, and related to the predictability of the courts, there are costs associated with the risk of having a court decide a matter by a law that it is not familiar with. As McClendon (2012) notes, courts have a competitive advantage in deciding their own state law. Indeed, contracts that specify both a governing law and a court forum hardly ever create a dispute resolution process in which courts apply a law from another state. Interviews have shown that the divergence of the substantive law and forum are the primary concern for lawyers considering in which forum to settle their disputes (Cain & Solomon 2010). However, if parties do specify a law but do not specify a forum, the chance for such a constellation to arise increases significantly, making the outcome less predictable and potentially longer due to the unfamiliarity of the judges.[25]

The Benefits

The costs described above appear to be substantial. Indeed, the fact that an unspecified forum alters the incentive for parties to perform in accordance with the contractual terms implies that most contracts without forum selection clauses are negatively affected, even if litigation never occurs. Compare these costs to the benefits of leaving the forum unspecified, i.e. the costs of negotiating and drafting a forum selection clause. While comprehensive data on the drafting costs is not available, we can rely on evidence for a subset of

[24]See also Fisch, New York Evidence § 42, at 22 [2d ed].

[25]To be sure, the governing law is one factor taken into consideration as part of considerations of *forum non conveniens*, which allow courts to dismiss a case in favor of a more suitable forum. However, taken by itself, the governing law is generally insufficient to justify dismissal, see Piper Aircraft Co. v. Reyno, 454 U.S. 235, 102 S. Ct. 252, 70 L. Ed. 2d 419 (1981).

contracts to see that high negotiation costs are an unlikely cause for the lack of forum selection clauses. In particular, a 2013 Survey of general counsels in the Public Utility, Communications and Transportations Industries (PUCAT) conducted by the American Bar Association inquired about the time devoted to the negotiation of dispute resolution clauses in "significant commercial contracts".[26] 59% of respondents said that their company allots less than one hour on the negotiation of dispute resolution clauses and 82% spends less than four hours. Even if one were to assume that the most senior partners at law firms are responsible for drafting forum selection clauses, the costs of their inclusion in accordance with current practice would not exceed $5,000 per company.[27] This finding is in line with evidence provided by Gulati and Scott (2012), who find that, in the cases in which the *pari passu* clause has been modified in corporate agreements, this was usually done without incurring elaborate costs going through painstaking negotiations.

2.2 Hypotheses

Overall, it seems implausible to characterize the striking lack of choice-of-forum clauses as a rational gap. Instead it is proposed here that law firms have considerable discretion with regard to the drafting of commercial contracts. Law firms use this discretion to write clauses that reflect their own preferences, not the preferences of the company bound by the agreement. Building on Gulati and Scott (2012), it is suggested that the drafting process within a law firm may be characterized by risk aversion and herd behavior, which leads to an aversion to innovation and a preference for conservatism. This preference persists even when the contract includes obvious gaps. To be sure, companies employ sophisticated inhouse counsel who could closely monitor forum selection provisions or limit the law firm's action space through firm-wide policies. However, anecdotal evidence suggests a striking lack of interest of inhouse counsel in these clauses. To illustrate, consider the 2010 version of the International Arbitration Survey on Choices in International Arbitration. The authors were interested in the experience of inhouse counsel with respect to the inclusion of arbitration clauses into corporate contracts and even though it remains a side note of the study, the authors find that many of the interviewees referred to arbitration clauses as "2am clauses" that are expected to

[26] ABA ADR Survey of Companies in PUCAT Industries, Fall 2014 Report of the Alternative Dispute Resolution Committee.

[27] Assuming an hourly rate of $1,500 per hour, see `http://www.abajournal.com/news/article/top_partner_billing_rates_at_biglaw_firms_nudge_1500_per_hour` (February 8th, 2016).

be drafted and included quickly *after* the commercial terms are set.[28] Though only a first step, the survey indicates that inhouse counsel often does not attach much value to choice-of-forum provisions and thus, lacks motivation to closely monitor their inclusion. In effect, forum selection clause usage then lies in the law firms' discretion.

Hypothesis I: Forum selection clause usage does not reflect company preferences, but law firm preference for using templates

If it can be shown that the presence (or lack) of forum selection clauses is a function of the law firm writing the contract, a next question is how sticky these clauses or gaps are. That is, will parties change their practice pursuant to the experience of shocks? In general, there are two types of shocks that could cause parties to change their practice.

System Wide Shocks

System wide shocks are unforeseen changes that affect all parties. As pointed out above, the law governing forum selection clauses underwent an important change following a series of Supreme Court decisions. The first of these decisions was *Goodyear v Brown*, which was decided in June 2011. In that case, plaintiffs were estates of two American boys killed in a bus accident in France. They alleged faulty tires and proceeded to sue the manufacturers, Goodyear's affiliate in Luxembourg and its branches in Turkey and France, in the courts of North Carolina. The plaintiffs argued that North Carolina courts had personal jurisdiction because the defendants' parent company and distributor, Goodyear U.S., is a United States company. Goodyear U.S. operates plants and is commercially active in North Carolina, but the subsidiaries argued that this was enough to their parents' activity was sufficient to establish jurisdiction over them.

In a unanimous decision written by Justice Ginsburg, the Supreme Court sided with the defendants, holding that a companies' connections to a state must be so "'continuous and systematic' as to render them essentially at home" in the state exerting general jurisdiction.[29]

While the specific circumstances in *Goodyear* left some doubt as to the holdings' generalizability, the subsequent decision of *Daimler v. Bauman* made it abundantly clear that the "essentially at home" test would be the new test

[28]Friedland and Mistelis (2010, 10)

[29]Goodyear Dunlop Tires Operations, S.A. v. Brown, 564 U.S. 915, 919, 131 S. Ct. 2846, 2851, 180 L. Ed. 2d 796 (U.S. 2011)

courts were required to apply when determining general jurisdiction. *BNSF v. Tyrrell* clarified that the almost universally adopted state laws which establish jurisdiction by way of (registering for) doing business in a state violate the Fourteenth Amendment. And *Bristol Myers v. Superior Court of California* stopped attempts by Californian courts to apply a "sliding scale" approach to the interpretation of specific jurisdiction. Under this approach, California courts sought to argue that doing business in a state was a relevant and sufficient factor to establish specific jurisdiction, effectively mimicking the "doing business" test under general jurisdiction.

Even though none of these Supreme Court decisions were contracts cases, they had profound implications for forum selection clauses in contracts. Indeed, not only were they followed by a great number of court opinions which applied the "essentially at home" test to determine jurisdiction in contract disputes.[30] Many practitioners also took the decisions as an indication that the inclusion of a forum selection clause would be more important than ever in order to ensure that a dispute can be litigated in a competent forum. For instance, the Association of Corporate Counsel urged its members to pay especially close attention to forum selection clauses in their contracts following *Bauman*.[31] As such, the series of Supreme Court decisions can be viewed as a drastic decrease in the number and variety of jurisdictions a plaintiff could sue in absent a forum selection clause.

If companies or the law firms paid close attention to the legal framework surrounding forum selection clauses, then the four decisions should have had two effects which lead to testable hypotheses. First, we might expect that the Supreme Court decisions increased the salience of the importance of forum selection clauses, incentivizing parties to review their drafts in order to make sure that a forum selection is present. This should then lead to an overall increase in the number of forum selection clauses in contracts. Second, since the Supreme Court decisions drastically decreased the number of forums parties could litigate in absent forum selection clauses, we would expect the rate of these clauses to increase simply because parties want to preserve the option to litigate in their preferred forum. However, note that this second effect should not apply to all contracts. In particular, we would expect the narrowing of general jurisdiction to have an effect only where parties seek to litigate outside of their state of incorporation and their principal place of business. If it is the parties' intention to litigate in a state that has jurisdiction both under the "essentially at home" test and under the "doing business" test, then there

[30] CITE a bunch of contracts cases

[31] See `https://www.acc.com/chapters/ncr/upload/Slides-ForumSelection2.pdf`

should not be any effect absent that of higher overall salience.

Hypothesis II.1.1: The introduction of the "essentially at home" test lead to an increase in the number of forum selection clauses.

Hypothesis II.1.2: The introduction of the "essentially at home" test lead to a further increase in the number of forum selection clauses for parties seeking to litigate outside of their state of incorporation or principal place of business.

Company Specific Shocks

Aside from shocks affecting all companies, it may also be instructive to investigate whether parties can be incentivized to change their practice with respect to forum selection clauses when they experience shocks at the individual level. The shock to a company that is considered here is that of being sued. If a company is sued and this suit is particularly costly, we would expect the company to try to avoid similar suits in the future using forum selection provisions. One way in which companies can opt out of litigation is by including an arbitration clause. It thus follows that companies who make negative experiences with litigation should be more likely to rely on arbitration in the future.[32] To be sure, specialization causes transactional lawyers to only rarely be involved in the contracts they drafted. In fact, the litigators often do not even come from the same law firm as the drafters. It is thus unlikely that transactional lawyers will ever directly experience the negative consequences of the contracts they have drafted. However, an insufficiently drafted contract can cause companies to change their preferences over forum selection clauses and mandate the inclusion of a specific type of provision if their previous practice has not worked well.

Hypothesis II.2: Companies increase their use of arbitration clauses after frequent and costly trials.

[32]In general, the same applies in the reverse, i.e. companies that make negative experiences with arbitration should be more likely to rely on the courts. However, since a companies' arbitration experience is not publicly documented, the focus will lie on arbitration clause usage after negative trial experiences.

2.3 Data Description & Methodology

3.1 Contracts Data

The Data

The analysis uses a combination of three large data sets. The first is a the data set of all filings of 'material contracts' with the SEC through its electronic filing system EDGAR between 2000 and 2016.[33] The SEC requires registered companies to report every "material contract", which encompasses "[e]very contract not made in the ordinary course of business which is material to the registrant."[34]. Companies registered with the SEC are those that made a public offering or have "total assets exceeding $10,000,000 and a class of equity security (...) held (...) by five hundred or more persons".[35]. The lack of a precise definition of the word "material" provides these companies with some discretion in deciding which agreements to disclose. However, this discretion is limited by general principles established in judicial decisions or administrative guidelines taken into account by the companies.[36] For instance, since the purpose behind this and similar disclosure rules is to remove information asymmetries and allow investors to make informed investment decisions, the SEC staff typically applies the standard established by the Supreme Court in *Basic v. Levinson*[37] when determining whether information falling under a disclosure requirement is "material". Accordingly, materiality implies that "'there is a substantial likelihood that a reasonable shareholder would consider [the contract] important' in making an investment decision."[38] In practice, contracts that meet this definition are often asset and stock purchasing agreements, loan contracts as well as agreements governing the employment and compensation of key employees such as CEOs. SEC staff actively monitors the compliance of companies with the contract disclosure requirement and notifies them if the financial statements indicate an omission.[39]

[33]A more detailed description is given in the previous chapter and is only recited in abbreviated form here to allow the reader to read individual essays in isolation.

[34]17 C.F.R. § 229.601(b)(10)(i)

[35]See Securities Exchange Act § 12(g)

[36]See Correspondence between Marketo, Inc. and the SEC staff about Marketo's procedure on how to determine disclose requirements, available at https://www.sec.gov/Archives/edgar/data/1490660/000110465914004115/filename1.htm (dated January 24st 2014).

[37]Basic Inc. v. Levinson, 485 U.S. 224, 108 S. Ct. 978, 99 L. Ed. 2d 194 (1988).

[38]SEC Release No. 33-7881 (Aug. 75, 2000).

[39]Id.

Companies attach the agreements to their annual reports (Form 10-K), quarterly reports (Form 10-Q) and to reports filed due to important events and changes between quarterly reports (Form 8-K). Similar provisions exist for foreign companies, who have the option to report using Forms 20-F and 6-K. In addition, during Mergers & Acquisitions, the relevant contracts are reported as exhibits to Form S-4. The electronic forms and exhibits are available for all registered companies through the SEC Electronic Data Gathering, Analysis, and Retrieval system (EDGAR) from its establishment in 1996 to 2016. Because the SEC has continuously changed and extended the filing requirements through EDGAR pursuant to the system's phase-in in 1996, this study is limited to contracts filed between 2000 and 2016, when filing requirements were largely uniform for the forms examined here.[40].

Overall, the data set includes 780,689 agreements. Of those, 272,837 filings are dropped because they are duplicates or mere amendments to already existing contracts, leaving a total of 507,852 unique contracts submitted by a total of 18,641 companies.

EDGAR includes data on the party that filed a contract and its industry. The filing party is assumed to be the first party to the contract and its industry is assumed to be the industry pertaining to the contract. A search algorithm based on regular expressions then identifies the paragraph in the contract that includes the parties to the dispute. The algorithm is described in detail in the Appendix. This paragraph is then scanned for the mentioning of any of the 630,106 companies and individuals that have ever disclosed information through filings with the SEC in order to supplement the information on the parties to the contract.

Next, it is necessary to identify whether a given agreement includes a forum selection clause and if so, what type of dispute settlement provision the parties agreed on. Due to the large number of contracts, a machine learning algorithm is required that is able to identify forum selection provisions. To achieve this goal, 5,226 paragraphs are coded by hand for their inclusion of dispute settlement clauses. The paragraphs are then randomly divided into a training set and a test set. Using the training set, a Naive-Bayes classifier[41] is trained to identify words and word-combinations that are most indicative of each type of dispute settlement clause. The classifier is then used to predict the types

[40]For example, in 1999, the SEC allowed submission of filings in HTML format (and the attachment of PDFs), which made filing much easier and is by far the most frequent form of submission today.

[41]For a thorough examination of the performance of the Naive-Bayes classifier, see Rish (2001). While there are other popular options available, the Naive-Bayes classifier yields the best results.

of paragraphs in the test set, which in turn allows for an assessment of the classifier's performance.

The approach correctly classifies 99.88 percent of the paragraphs. However, the correct classification rate alone can be misleading, since it does not take into account the number of relevant items. For instance, for a test set consisting of 99 irrelevant and 1 relevant paragraphs, a simple algorithm that always considers all paragraphs irrelevant would achieve a correct classification rate of 99 percent. This is why –in addition to the correct classification rate– studies in information retrieval and machine learning use precision, recall, F_1 scores and Matthew's Correlation Coefficients (MCC) to assess the quality of automated classification procedures.[42] Together, these can be thought of as relative measures of performance that take into account the total number of relevant items. The classifier trained here achieves a precision of 0.89, a recall of 0.94, and an F1-Score as well as a Matthew's Correlation Coefficient of 0.91. It can thus be considered as very accurate, with no strong tendency for false positives or negatives.

A similar process is used to identify whether a contract includes a clause specifying the substantive law governing the contract and if so, which law governs. In a last step, a combination of search terms / phrases and regular expressions is used to identify the type of the document (e.g. employment contract, credit / loan agreement etc.) and the form of the document (e.g. agreement, plan, policy). The entire procedure is described in greater detail in the Appendix.

[42]Let TP be the number of true positives, i.e. the number of correctly classified forum selection clauses; FP the number of false positives, i.e. the number of paragraphs that have incorrectly been classified as forum selection clauses when they are not; TN the number of true negatives, i.e. the number of correctly classified paragraphs that are not forum selection clauses; and FN the number of false negatives, i.e. the number of paragraphs that have been classified as not containing a forum selection clause when in fact they do. Then

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

$$MCC = \frac{TP \cdot TN - FT \cdot FN}{\sqrt{(FT+FN)(TP+FN)(TN+FP)(TN+FN)}}$$

Summary Statistics

Table 2.1 contains summary statistics describing the contracts in the data set. The vast majority of contracts is concluded exclusively between U.S. parties (89%). Only 44% of agreements in the period of observation include forum selection clauses, even though 75% include a clause specifying the substantive law of the contract. As pointed out above, this seems puzzling, given that both types of clauses seek address issues arising out of uncertainties of the relevant and applicable legal framework. Among forum selection clauses, those that refer parties to courts are more prevalent than arbitration clauses (30% vs. 19%).

Figure 2.1 plots the use of different types of forum selection clauses over time. What can be seen is that contracts became more likely to include forum selection clauses as time passed. However, there is a difference between the propensity to include a forum selection clause referring parties to courts and one that refers to arbitration. In particular, the higher propensity to include forum selection clauses is exclusively driven by the increased presence of clauses referring parties to courts. In contrast, arbitration clauses became less common with the years. This finding contradicts some of the claims found in the literature contending that arbitration is becoming increasingly popular.

Table 2.2 breaks the prevalence of forum selection clauses down by industry. Most of the contracts in the sample come from the manufacturing industry, followed by the financial industry and the service industry. What can be seen is that the agricultural industry is the only industry where forum selection clauses are more likely to be included than not included. However, with only 13 observations, these numbers are not particularly reliable. In all other industries, it is more likely *not* to find a forum selection clause in the contract (between 41% and 46%), even though it is very likely to find a governing law clause in contracts across all industries (between 73% and 85%). Throughout all industries, arbitration clauses are relatively rare, with clauses referring parties to courts dominating the landscape of dispute settlement provisions.

Breaking contracts down by agreement type as in Table 2.3 paints a somewhat different picture. Joint venture, M&A, licensing, loan and sales and purchasing agreements are more likely than not to include a forum selection clause. At the same time, contracts providing incentives to key employees, such as employee stock option plans, pension plans and 'golden parachute' agreements are the least likely to include a forum selection clause. While one should proceed with caution when interpreting descriptive statistics, these findings are at least consistent with the idea that contracts of great economic importance are more like to be carefully drafted and parties make a greater

effort to anticipate contingencies. The descriptives are also consistent with isolated findings in the literature on the relevance of dispute settlement clauses in specific settings. For instance, it has previously been argued that contracts over innovative goods, among them joint venture and licensing agreements, are particularly sensitive to the issue of legal enforcement due to a high level of dependence on injunctive and emergency relief (O'Connor and Drahozal, 2014). For M&A, it has been argued that the close entanglement of contract law with corporate, securities and antitrust law provides incentives for parties to pay especially close attention to harmonize the legal framework surrounding their deal. In effect, this often means that forum selection clauses refer disputes to Delaware (Eisenberg and Miller, 2007; Cain and Davidoff, 2012; Coates, 2012).

Next, Table 2.4 depicts how frequently different court forums are chosen. Consistent with previous findings in the literature (Eisenberg and Miller, 2007), New York is by far the most popular forum, with 37% of forum selection clauses referring parties to New York courts. It is commonly assumed that the reason for this dominance is the high level of predictability that New York courts have, paired with their expertise in adjudicating complex commercial disputes. In addition, most big law firms are headquartered in New York and economies of scale incentivize attorneys interested in practicing business law to seek admission to the New York bar, making it an unsurprising primary choice for dispute settlement. Other popular forums include Delaware (11%), California (8%) and Texas (5%).

If a contract includes both a choice-of-forum and a choice-of-law clause, parties consistently match the substantive law to the forum. This confirms interviews conducted by Cain and Davidoff (2012) in which lawyers stated that one of their primary concerns in drafting these clauses is to avoid an incoherence between the law governing the contract and the forum that interprets it.

Consider now the question of which law firm assisted in drafting a contract. While contracts often do not disclose the law firm responsible for drafting the contract, there are many instances in which they do. Typically, the drafting law firm is disclosed in the notice clause, which requires a copy of any written communication regarding the contract to be submitted to the counsel that assisted in drafting the contract. Other instances include fee shifting clauses in which one party agrees to pay for the administrative costs of the other's counsel or clauses stating where the contract will be signed, which is often in one of the advising law firm's offices. I exploit this fact using a list of 7,708 law firms with at least 50 employees collected through LexisNexis Academic to identify the external counsel involved in the drafting of the agreement. This

approach successfully identifies participating law firms for 102,475 contracts. It should be noted that this is not a random sample of all contracts. Indeed, contracts identifiably drafted by law firms tend to be longer and more likely to include coice-of-forum and choice-of-law clauses than the average contract.[43] This makes sense, given that external counsel is more likely to be required in especially complex matters.

Table 2.5 depicts the 30 most frequently relied upon law firms. By far the most contracts are drafted by Latham & Watkins and Skadden, Arps, Slate, Meagher & Flom LLP, with 4,509 and 4,438 contracts, respectively. Choice-of-law clauses are almost universally adopted, with most law firms including them in 97% of their contracts. Choice-of-forum clauses are less common, with most law firms including them in 70-80% of contracts. The majority of contracts utilize a domestic court system, with arbitration being relied upon typically in only 20-30% of agreements.

3.2 General Counsel Data

In order to obtain data on a companies' general counsel, I rely on FactSet. Though it is one of the most comprehensive data sets on general counsels, it has two important limitations. First, the data set contains information only on individuals who are currently active as a general counsel. Hence I do have information about the current general counsel and how long she worked for a company, but have no information on who was the general counsel for a company prior to its current counsel.[44] Second, the general counsel information on FactSet is limited to companies publicly traded on large U.S. stock exchanges such as the NYSE and NASDAQ. In total, the data set includes information on 4,201 general counsels for 4,670 companies drafting a total of 138,617 agreements. Because the SEC uses a company central index key (CIK) to identify companies, whereas FactSet uses the security identifiers CUSIP and ISIN, I rely on Compustat to translate CIKs to ISINs and merge the two data sets.

[43]The average length is 9,207 overall and 23,512 if a law firm drafted the contract. 44% of contracts include a choice-of-forum clause, 73% if a law firm is involved. 75% of contracts include a choice-of-forum clause, 94% if a law firm is involved.

[44]For example, if a company employed GC1 from 1999-2004, GC2 from 2004-2008 and GC3 from 2008- today, my data shows that GC3 worked for the company since 2008, but I lack information on the general counsels prior to 2008, i.e. on GC1 and GC2.

3.3 Litigation Data

Lastly, I collect data from AuditAnalytics on material federal litigation disclosed to the SEC by registered issuers. Registrants are required to disclose "material pending legal proceedings, other than ordinary routine litigation incidental to the business".[45] AuditAnalytics collects all those disclosures and additional information relevant to the lawsuit, such as the court in which the suit is filed, the date at which the lawsuit concluded and the settlement amount. The full data set includes 63,261 cases filed between 2000 and 2016. When filing a case, plaintiff's attorney is required to specify one (and only one) Nature of Suit (NoS) code that best describes the nature of filed lawsuit. This NoS code is used to identify the 4,4778 cases that are contract disputes.

3.4 Analytical Approach

Testing of Hypothesis I

Consider now the first hypothesis, that the contracts reflect law firm as opposed to company preferences. We can start analyzing this question using simple descriptive statistics. As was suggested, inhouse counsel could limit law firm practice and guarantee that forum selection clause usage reflects company preferences by enacting firm-wide policies. If such policies exist and are enforced, we should be able to observe consistency in forum selection clause usage on the company level. In other words, if a company has policies on the usage of forum selection clauses, it should include the same type of clause in most of its contracts. For instance, a company that has a policy favoring arbitration should have an arbitration clause usage rate close to 1 and a low variance in its use of arbitration clauses. Figures 2.2 and 2.3 depict the distribution of within-company means and variances for the frequency with which different types of clauses occur. What can be seen is that virtually no company consistently uses the same type of clause repeatedly, which is indicated by the fact that almost no company has a mean usage rate that is anywhere close to 1 and by the substantial variance of clause usage within a company. The only exception to this rule is arbitration. Indeed, the results leave open the possibility that some companies have a "no arbitration" policy, given that there are many companies that do not use arbitration and are also consistent in this practice. However, note that the consistent omission of arbitration clauses is not met with a consistent use of court clauses. A company policy that is solely

[45] 17 CFR 229.103 - (Item 103) Legal proceedings.

"negative" in that it asks drafters not to include arbitration clauses without mandating the inclusion of court clauses would be puzzling in many ways. It thus seems reasonable to assume that the consistent lack of arbitration clauses is not the result of a widely enforced company policy, but could rather be an expression of the drafters' aversion to arbitration. Overall, the data suggests that the vast majority of companies lacks a coherent and widely enforced policy on forum selection clauses.

The descriptive statistics lend preliminary support to hypothesis I insofar as the contracts do not reflect a consistent company preference. Instead, most firms sometimes use and sometimes do not use forum selection clauses. However, descriptives alone are insufficient and are thus complemented with formal statistical tests. Intuitively, hypothesis I can be examined by comparing the variation of forum selection clause usage of contracts written by different law firms ("between variation") to the variation of contracts written by the same law firm ("within variation"). If choice-of-forum clauses have a high degree of between variation but a low degree of within variation, this indicates that a) law firm preferences over forum selection clauses is expressed in the contracts; and b) law firms are internally consistent in their use or non-use of forum selection clauses.

Any comparison of two contracts has the potential to introduce omitted variable bias. In particular, the companies studied here regularly conclude contracts in a very broad range of commercial contexts. For example, a single company can conclude several licensing, sales and loan agreements, all with different contractual partners and in different years. Comparing two contracts that differ on these dimensions makes it impossible to tell whether changes in forum selection clauses are due to changes in the law firm drafting the contract or due to changes of other observable characteristics. To address this concern, I employ a double robust exact matching procedure.

Matching is a popular method in the social sciences and causal inference that seeks to pair two units that look similar on a number of dimensions, with the only exception being the covariate of interest. The estimates derived from these pairs decrease potential omitted variable bias and guarantee common support (Rubin, 1973). Among the different matching algorithms, exact matching is the most restrictive, as it requires each pair of observations to be exactly the same across all covariates. This has advantages and disadvantages. The main disadvantage is that an exact matching algorithm omits a lot of data. However, in very large data sets such as this one, omitting data is not a primary concern as long as reliable standard errors can be obtained. The main advantage of exact matching is that it is able to achieve perfect balance across all observable covariates, making treated and control units perfectly

comparable on those observables.

I match with replacement each contract to another contract that is concluded between the same parties, is of the same type and written at the same time.[46] This leads to a data structure in which each observation is a dyad of highly comparable contracts, with the only possible observable difference between contracts being the law firm who drafted them. The dyadic structure has the significant advantage that it is now possible to define a single treatment indicator. Treated dyads are those in which both contracts were written by the same law firm. In contrast, control dyads are those in which both contracts were written by different law firms.

After creating contract dyads in this way, I employ exact matching again, this time on the dyadic level. In particular, I match each dyad in which both contracts have been written by the same law firm to another dyad in which both contracts have been written by different law firms. To maximize comparability, I impose the additional restriction that treated and control dyads have to share one contract. This guarantees that all contracts in both treated and control dyads are contracts between the same companies, are of the same type and written at the same time, with the only observable difference being the law firms who drafted the contracts. If multiple control dyads match a treated dyad, I average across controls.

Intuitively, the procedure can be thought of in the following way: "Take a contract i and compare it to a similar contract j written by the same law firm and to a similar contract k written by a different law firm. Compare the variation between i and j to the variation between i and k.". The following example illustrates how the procedure is implemented:

Let $Y_{i,L}$ denote the outcomes of contracts of the same type, written between the same companies in the same year. i is an index identifying one of four contracts, i.e. $i \in \{1, 2, 3, 4\}$. L identifies whether the contract was written by law firm l, hence $L \in \{0, 1\}$. Assume that of the given contracts, two were written by law firm l, so the contracts are $Y_{1,1}$, $Y_{2,1}$, $Y_{3,0}$ and $Y_{4,0}$. I first employ the matching procedure, leading to the following matches:

[46] Contracts are considered to be written at the same time if they have been reported within a one-year period from one another.

$$Y_{1,1} - Y_{2,1}$$
$$Y_{1,1} - Y_{3,0}$$
$$Y_{1,1} - Y_{4,0}$$
$$Y_{2,1} - Y_{3,0}$$
$$Y_{2,1} - Y_{4,0}$$
$$Y_{3,0} - Y_{4,0}$$

In order to estimate the law firm influence on contract 1, I consider only dyads that include contract 1. Those are

$$Y_{1,1} - Y_{2,1}$$
$$Y_{1,1} - Y_{3,0}$$
$$Y_{1,1} - Y_{4,0}$$

I now compare the difference in choice-of-forum clauses in dyad $Y_{1,1} - Y_{2,1}$ to the average difference in dyads $Y_{1,1} - Y_{3,0}$ and $Y_{1,1} - Y_{4,0}$.

The described method is applied in a similar way to the general counsel in a company. That is, I compare a pair of contracts written under the same general counsel to a pair of contracts written under different general counsels. Focusing on the general counsel follows the rationale that she is the person primarily responsible for supervising and monitoring contract policy within a company. As such, the general counsel is the individual best suited to enforce company preferences on forum selection clause usage.

Matching has its roots in the literature on causal inference and can rule out the existence of omitted variable bias caused by observable characteristics. However, methodology alone can never guarantee causal identification in observational studies, as it is still possible for observations to differ in important but unobservable ways. In particular, it cannot be ruled out that both general counsels and law firms are appointed for reasons that correlate with their forum selection clause usage and that the method employed does not capture. Consider first the appointment of a general counsel. The relevant literature suggests that a general counsel has four main tasks (DeMott, 2005). In addition to being the chief legal adviser, she takes a senior managerial role, forming corporate strategy at the highest hierarchical level (Chayes and Chayes, 1985). She has an administrative function, managing a companies' budget of the legal

department, hiring new lawyers and setting internal policies (Basri and Kagan, 2004). And she has a representative function, frequently dealing and negotiating with third parties. The use of dispute resolution clauses is just one small aspect of the general counsel's role as legal adviser and could thus only play a nominal role in her hiring decision. Now in addition, note that the hiring decision of a general counsel is typically made by the CEO.[47] Choosing the forum for a dispute is an entirely legal problem that a CEO typically neither is nor has an incentive to become well versed in. It thus is plausible to assume that general counsel practice on forum selection clauses plays no relevant part in a hiring decision that is designed to introduce any substantial biases.

For law firms, omitted variable bias seems somewhat more plausible. To be sure, if companies directly observed or cared deeply about the inclusion of choice-of-forum clauses into contracts, one would expect firm-level policies and a near universal inclusion of such clauses, both of which we can't find. However, it is possible that secondary considerations play a role. A company might notice that contracts drafted by a certain law firm lead to more litigation or break down more frequently than those drafted by others, without being aware of the fact that this is due to the omission of choice-of-forum clauses. If that was the case, we would expect forum selection clause usage to correlate positively with the frequency with which a law firm is chosen. However, this correlation is merely 0.029 and is also statistically insignificant. It thus seems likely that even law firms are chosen independently of their use of forum selection clauses. Overall, while it is not possible to entirely rule out the presence of omitted variable bias, theoretical considerations make the presence of substantial biases caused by unobservable differences unlikely.

The primary outcome measure on which dyads are compared is the inclusion of a forum selection clause, with an indicator taking the value 1 if a (specific type of) forum selection clause is included in the contract and 0 otherwise. However, to also assess differences not in *whether* a contract includes a forum selection clause, but also in *how* forum selection clauses are drafted, I define two alternative outcome measures which depict the similarity between two clauses. The similarity of two texts can be measured in several ways.

The most popular way to measure similarity with respect to the content of the text is by way of the cosine similarity of the tf-idf vectors for the preprocessed text (Baeza-Yates and Ribeiro-Neto, 1999; Manning et al., 2008). In a first step, the text is preprocessed by converting all letters to lower case, removing punctuation and stopwords and by stemming the words in the text

[47] While technically appointed by the board of directors, it is factually the CEO who has a 'substantial if not exclusive role' in the hiring decision (DeMott, 2005).

in order to remove morphological affixes. The process is described in greater detail and illustrated by way of an example in the Appendix, where I describe information retrieval from the contracts. A second step translates the stemmed words into high-dimensional vectors, with each occurrence of a word receiving a tf-idf weight.[48] Having thus transformed two texts into two vectors, one can measure the similarity of the texts by way of the angle between the vectors. The cosine similarity is a representation of that angle that is bounded between 0 and 1, with 1 indicating a high level of similarity and 0 indicating a low level of similarity. An example for similar and dissimilar clauses is included in the Appendix.

An alternative popular way to measure the similarity of two texts is by way of the Levenshtein distance (Levenshtein, 1966). Intuitively, the Levenshtein distance is simply a measure of the single-character edits that are required to make two strings of text look identical. Different from the cosine similarity, this similarity measure is not focused on the content of the text, but sensitive to how a text is written. As such, it is well suited to determine whether a text has been copy-pasted with minimal alterations (in which case the Levensthein distance is small) or whether it has been drafted anew.

Both of these distance measures are used as alternative ways in measuring the influence of law firms on how forum selection clauses are drafted.

Testing of Hypothesis II.1

Consider now hypothesis II.1.1 and II.1.2, that system wide shocks influence forum selection clause practice. As these are shocks that should affect the entire population of companies, they can be assessed using time series analysis, where the unit of observation is the use of forum selection clauses of all companies over time. Since the reporting of contracts to the SEC is periodic, I group observations into quarterly periods, starting with Q1 in 2000 and ending with Q2 in 2016.

The general analytical frame work to assess the time series data is the same for both hypotheses. In a first step, it is required to fit a model to the observed frequency of forum selection clauses in contracts using appropriate control data. In a second step, it is then necessary to predict how the use of forum selection clauses would have developed if there had been no system

[48] tf-idf stands for 'term frequensy - inverse document frequency'. In contrast to simply counting the words in a text, it reweighs the frequency based on their occurrence in the entire sample. For instance, a word that is especially prevalent in one document but never mentioned in the other receives more weight than a word that is very prevalent in both documents. This is intended to represent the words in the document based on their importance for said document.

wide shocks. In a third step, we can assess the difference between the actual observed rate and the predicted rate and determine how likely it is that the court decisions affected the use of forum selection clauses.

Hypothesis II.1.1 predicts more contracts to include a forum selection clause after the recent wave of Supreme Court decisions, which began with *Goodyear* in June 2011. We can thus use time series intervention analysis and fit the popular ARIMA model (Box et al., 1970) using pre-*Goodyear* data.[49] This model can then be used to forecast the post-*Goodyear* prevalence of forum selection clauses and compare the difference between actual and forecasted rates.

Hypothesis II.1.2 predicts more forum selection clauses under the "essentially at home" test for parties that intend to litigate outside of their state of incorporation or principal place of business. The parties' intent is not directly observable. However, an implication of hypothesis II.1.2 is that post-*Goodyear*, the number of contracts including forum selection clauses referring parties to courts outside of the state of incorporation and principal place of business ("*outside* court selection clause") should increase at a higher rate than forum selection clauses referring parties to courts inside the state of incorporation or principal place of business ("*inside* court selection clauses"). Since the prevalence of inside court selection clauses should be unaffected by the change in the legal frame work, it can serve as a control group, whereas the prevalence of outside court selection clauses serves as a treated group. Having two time series, one treated and one control, allows for the application of the synthetic controls method (Brodersen et al., 2015). In the synthetic controls method, the control data is used to create a synthetic time series that looks like the treated time series. But since the synthetic time series was never subject to the intervention, it allows the researcher to interpret differences in the post-intervention period as causal effects of the intervention. The advantage of the synthetic controls method over forecasting using an ARIMA model is that the accuracy of the model can be validated. In particular, in a successful implementation, the time series of the treated and control units should be largely identical prior to the intervention. If there is a causal effect, then we should be able to observe a divergence post-intervention.

Testing of Hypothesis II.2

Consider now hypothesis II.2, that a companies' arbitration clause usage

[49]The recommended order of first-differencing, as well as the model orders p and q in the ARIMA model can be determined using the `forecast` package in `R`. The order of first-differencing is chosen based on a KPSS test, the order of p and q are chosen based on the Akaike information criterion. For details, see Hyndman & Khandakar (2007)

changes after experiencing many costly trials. As pointed out above, the focus lies only on federal litigation. Though data on the entire litigation experience–both in state and federal courts–is preferable, examining only federal litigation experience is the result of a preference for reliable results obtained from a smaller subset over less reliable results for a larger subset. For the defendant, litigation is typically an unforeseeable event.[50] It is thus common to characterize the filing of a suit as an external shock for the defendant that provides a convincing identification strategy in research on public companies (see e.g. Fich and Shivdasani, 2007). In contrast, the plaintiff can choose when and where to sue, raising concerns rooted in possible endogeneity. I thus limit the analysis to negative litigation experience as a defendant.

In addition, note that companies typically report litigation as a loss contingency. Loss contingencies are among the most sensitive information that companies provide as part of their disclosure requirements and many companies fail to report litigation altogether, a fact the SEC has criticized repeatedly in the past and is increasingly scrutinizing. It is impossible to derive any useful information from firms that categorically fail to report any lawsuits as it is unclear whether their failure to report is caused by the fact that they never experienced any material litigation–which is unlikely for most of the firms analyzed here–or whether they are categorical "never reporters". The analysis thus only considers companies that have reported litigation at least once, as non-reporting of other firms could be caused by reasons other than the non-occurrence of material litigation. Results for all companies are included only as a robustness check.

Having cabined the observations in this way, I convert the data set to panel data, where each observation is a company-year with information on the number of contracts reported, the share of forum selection clauses in the reported contracts, the number of federal lawsuits filed against a company and the settlement amount of those lawsuits as a proxy for their economic importance.[51] The panel includes 105,142 observations of 6,186 companies. Intuitively, hypothesis II.2 can be tested similar to hypotheses II.1, but only on the individual company level. However, the data structure presents two additional complexities. First, trial experience is not binary. In particular, each company can experience more than one trial, and in more than one period. For instance, if a company experiences one trial in the year 2009 and two in 2010, it does not make sense to compare arbitration clause usage rates in

[50] At least with respect to the specific point in which the lawsuit is initiated

[51] Using quarterly data as above would render the panel too sparse to conduct meaningful analyses.

2008 to 2010 in order to estimate the learning effect. Similarly, the economic stakes of each trial can vary dramatically. Second, learning effects can be immediate, but can also take some time to fully come into effect. A simple before-after comparison might incorrectly capture longer-term learning effects. For these two reasons, I impose a structural model on the data that preserves the intuitive idea of comparing usage rates before and after an intervention as presented above, but that takes into account longer-term effects and the possibility of non-binary treatment.

The outcome of interest is a rate, namely the frequency with which a company uses a clause, relative to its contracts. Changes in rates can be estimated using models for integer outcomes where the denominator is included as an offset (i.e. with a coefficient fixed at 1). I fit a negative binomial fixed effects panel regression (Hausman et al., 1984) where I regress the rate of choice-of-forum clause usage on the number of disputes that occurred and on their settlement amount. The logged number of contracts concluded in a year is included as an offset. The inclusion of time and company fixed effects control for all time-invariant company effects and overall time trends. The estimate can be thought of as the average of all company-level changes in response to negative litigation experience.[52]

Motivated by the fact that the average trial in the data set lasts for about 2.5 years, I allow learning to take effect over a period of 3 years, such that the model includes the numbers / economic volume of trials with a lag of 0, 1 and 2. Standard errors are derived from the observed information matrix. In order to test the hypothesis that litigation experience matters, it is then of interest whether the variables indicating the (lagged) litigation experience are jointly significant. This hypothesis is assessed using a Wald chi-square test for joint significance. As a robustness check, the full model (3) is estimated without omitting firms that never report lawsuits and as a random effects model.

[52]Formally, following Hausman et al. (1984), the conditional likelihood function maximized is

$$lnL = \sum_{i=1}^{n} \omega_i [ln\Gamma(\sum_{t=1}^{n_i} \lambda_{it}) + ln\Gamma(\sum_{t=1}^{n_i} y_{it} + 1) - ln\Gamma(\sum_{t=1}^{n_i} \lambda_{it} + \sum_{t=1}^{n_i} y_{it}) + \sum_{t=1}^{n_i} \{ln\Gamma(\lambda_{it} + y_{it}) - ln\Gamma(y_{it} + 1)\}]$$

with ω_i being the weight for the ith company and $\lambda_{it} = exp(\boldsymbol{\beta x}_{it} + \text{offset}_{it})$.

2.4 Results

Table 2.6 depicts the results for the test of hypothesis I. As can be seen, the probability that two contracts differ in their forum selection clause usage even though they are drafted by the same law firm is very small. Only 4% of contracts add or remove an arbitration clause and 9% add or remove a court selection clause. Overall, 8% of contracts either add or remove the forum selection clause when switching from one contract to the next. In contrast, the probability that two contracts differ with respect to the forum selection clause is comparably very high if the contracts have not been drafted by the same law firm. Indeed, 12% of contracts add or remove an arbitration clause, 48% add or remove a court selection clause and 41% of contracts add or remove the forum selection clause altogether. The difference between the variation in clause usage within a law firm and the between law firm variation is statistically significant and, with factor of 3 to 5, substantively very large, implying that law firm preferences have much influence on whether a contract includes a forum selection clause or not. Further, the small variation in clause usage within a law firm implies that contracts used within a law firm are sticky, with the drafters closely following a given standard. At the same time, the large variation between law firms implies that different law firms have very different standards, with some including forum selection clauses and others not including them.

Compare these results to the findings for the general counsel. Variation both under different general counsels and under the same general counsel is modest, with 7-27%. While the variation of two contracts written by the same general counsel is smaller than if the contracts were written under different general counsels, this difference is small, with a factor of about 0.2. This implies that the general counsel, while imposing some consistency on the use of forum selection clauses, exerts much less influence than law firms.

Table 2.7 breaks down the stickiness of choice-of-forum clauses by industry and contract type. Contract types for which there was no sufficiently large number of similar contracts are omitted. Employment and incentive contracts are omitted because the employee, and with it one party to the contract, can change without the change being reflected in the data.[53] What can be seen is that in all industries and for all contract types, law firms are internally very consistent in the use of choice-of-forum clauses. This indicates that clause stickiness is not an idiosyncratic feature of the bond indentures that have

[53]Recall that parties are determined by scanning the contracts for entities registered with the SEC.

been studied in the past, but is a widespread phenomenon that characterizes a variety of commercial contracts.[54]

It was shown that contracts drafted by law firms are sticky with regards to choice-of-forum clauses. But how sticky is the *omission* of a clause when compared to the *inclusion* of a clause? This is what Table 2.8 examines. It defines different transitory processes from contractual gaps to forum selection clauses, e.g. from a contract with no clause to a contract with an arbitration clause, and considers how often they occur. It then contrasts this process to the opposite transition, in this case from an arbitration clause to no clause, in order to estimate whether a transition from no clause to an arbitration clause is more likely to occur than a transition from an arbitration clause to no clause. If gaps are as sticky as written clauses, there should not be a significant difference between the probability for each transition to occur. In contrast, if clauses are stickier than gaps, we should find a significant negative difference.[55] As can be seen, overall clauses are a little more sticky than gaps. In particular, there is an 12% probability that a law firm will fill a gap in its contracts, whereas the probability to create new gaps when there were previously none is 6%. However, this difference is largely driven by a stronger tendency for law firms to include court selection clauses. When we consider arbitration clauses, it is in fact more likely for law firms to omit a previously existing arbitration clause and leave a gap than it is that a previous gap will be filled by using an arbitration clause. Overall, the results suggest that court clauses are stickier than gaps, whereas arbitration clauses are less sticky.

So far, it has been shown that the inclusion or non-inclusion of forum selection clauses is sticky and largely a function of law firm preferences. Consider now the way in which law firms draft their clauses. It was suggested that the drafting of these clauses is heavily based on templates, with law firms reusing previous contracts repeatedly while making minimal alterations to the language. The presence of such a mechanism is observable by considering the

[54]It is worth mentioning one category of outliers. Security agreements, which relate to mortgages and other collateral and are predominantly used in the financial industry, almost always change their forum when the law firm changes. The reason is that virtually all matched security agreements are drafted either by the law firm Orrick, Herrington & Sutcliffe, LLP, or by K&L Gates LLP. Both law firms are internally very consistent, but agreements drafted by Orrick always include court selection clauses, where those drafted by K&L never include forum selection clauses. Hence if companies switch between law firms, this is almost always accompanied by a change in forum selection clauses. If we were to omit these agreements, the probability for forum selection clauses to change with the law firm in the financial industry would drop from 0.82 to 0.55.

[55]For some pairs, it is unclear which contract was signed first. In this case, both possibilities are considered, though each observation only receives half the weight.

way in which forum selection clauses are drafted. Table 2.9 considers only dyads in which both contracts include a forum selection clause and assesses whether clauses drafted by the same law firm are more similar to one another than those drafted by different law firms. As described above, the similarity of two clauses is measured by their cosine similarity and their Levenshtein distance. If law firms use templates, then the results should indicate a significant difference in Levenshtein distance when a clause is drafted by different law firms and a small difference when it is drafted by the same law firm. At the same time, the difference in cosine similarity between contracts drafted by the same and drafted by different law firms should be substantively small, as both types clauses should be highly comparable in their content.

As can be seen, the results are consistent with a model in which law firms use templates. In particular, the cosine similarity of forum selection clauses is always high, regardless of whether these clauses are drafted by the same law firm or not. In contrast, the Levenshtein distance when a clause is drafted by the same law firm is much smaller if two contracts are drafted by the same law firm, compared to a scenario in which they are drafted by different law firms. This suggests that law firms tend to copy-paste their clauses from previous contracts with minimal alterations, leading to a small Levenshtein distance. When the law firm changes, the clause remains substantively largely similar, as indicated by the high cosine similarity. However, since the clause is based on a different template, the Levenshtein distance changes drastically.

Figures 2.4 and 2.5 display results for hypotheses II.1.1 and II.1.2, respectively. As can be seen, the overall use of forum selection clauses does not demonstrate any anomalies. Neither has the overall prevalence of these clauses increased after *Goodyear*, nor has there been a significant change in the use of outside court selection clauses in reaction to the narrowing of the doing business test. As such, there is no evidence to support the conjecture that the practice of forum selection clauses changed in reaction to system wide shocks.

Table 2.10 depicts the results for the test of hypothesis II.2. Of importance for the interpretation of the results is the Wald test indicating whether litigation experiences in their different lagged forms are jointly significant. The model specifications show that both the (lagged) number of trials as well as the (lagged) settlement amount are jointly significant at conventional levels and positively associated with the number of arbitration clauses. This holds true when treating disputes and settlement amounts separately (Models (1) and (2)), but also when adding both to the same Model (3). The result implies that companies are more likely to include arbitration clauses in their contracts after having made experience with costly trials at the federal level.

Table 2.11 presents the results of the robustness tests. As can be seen, both

a model including all companies that never have reported a lawsuit (4) and a random effects model (5) yield similar results with respect to the number of lawsuits. That is, experiencing more federal litigation leads to a higher probability to rely on arbitration. The settlement amounts are not jointly significant in either specification, providing no support for the hypothesis that the outcome of the trial matters, as opposed to the mere fact that a lawsuit was filed. However, there are two reasons to be at least cautious with these results. Note that a random effects model makes the underlying assumption that all companies originate from a common, multinomial distribution. It thus pools information from different companies, whereas the fixed-effects model considers each company in isolation. The pooling assumption as well as the underlying distribution from which the company-specific effects originate are often difficult to justify and the random effects model is thus typically used in order to address concerns arising out of a lack of data availability. Data availability is no concern here and thus, there is little reason to prefer results obtained from the random effects model over the fixed effects estimation. With regards to the model considering all firms, note that, as pointed out above, it is likely that even those companies that have not reported material litigation have made experience with such litigation nonetheless. Their underreporting could cause changes due to costly trials to be attributed to the error term, in turn biasing the results downwards. Given this possibility, it is especially reassuring that the coefficient on the number of trials remains robust.

To illustrate the substantive significance of the findings, Table 2.12 reports the estimated marginal effects of litigation experience at the means. They reflect the increase in the probability to include an arbitration clause into a contract if the number of disputes increases from 0 to 1. Similarly, they depict the increase in arbitration clause usage if settlement amounts paid increase from 0 to 10 million USD. Having experienced a single federal trial is estimated to increase the probability to include an arbitration clause by 17-19 percent over a three-year period. Similarly, an increase in the settlement amount by 10 million USD increases arbitration clause usage by 9-10 percent. For reference, the average settlement amount conditional on a settlement being paid is 11.5 million USD. It is important to highlight, however, that these estimates, like any substantive estimates derived from fixed-effects models, need to be treated with caution. That is because the fixed effects are not estimated directly and thus, are assumed to be 0 when computing average effects at the mean.[56]

[56]There is no direct meaningful substantive interpretation of this constraint.

2.5 Discussion

Overall, the results provide strong evidence in support of hypothesis I. They suggest a model in which important company contracts are drafted by external counsel. Companies themselves are not entirely ignorant to the drafting procedure and do occasionally change their patterns in response to significant negative shocks. However, for the most part, law firms are able to draft these contracts under no constraint of firm-wide policies and with only modest participation of internal counsel. At the same time, law firms heavily rely on templates and show little impetus to make changes to these templates. This holds true regardless of whether the template happens to include a forum selection clause or whether it contains a gap. Indeed, a lack of support for hypotheses II.1.1 and II.1.2 suggests that subtle changes to the law pertaining to forum selection clauses are insufficient for companies and law firms to change their practice. Instead, it seems that companies are only willing to revise their practice and pay particular attention to the issue of forum selection after being subjected to direct and immediate costs, as suggested by the strong support for hypothesis II.2. The findings have important policy implications for legal scholarship, legal education and the provision of legal services.

Legal and economic scholarship applies the theory of revealed preferences to contracting, making the fundamental assumption that observable agreements concluded between commercial actors must be the result of a sophisticated and complex cost-benefit analysis that leads to an optimal result. The findings in this sutdy suggest that this assumption is incorrect and that even the most sophisticated actors frequently fail to have optimal agreements drafted. The popular "Coase Theorem" teaches students that default rules do not matter for the final allocation of goods in important commercial agreements because transaction costs are comparatively small and parties will simply contract around unfavorable defaults. However, the fact that not only drafted clauses but also contractual gaps are sticky suggests that default rules can have a significant economic impact even in agreements over multiple millions of dollars. Their significance implies that default rules deserve more scrutiny than they currently receive, as designing and improving on defaults can greatly increase overall welfare gains not only for consumers, but for commercial actors as well.

In addition, the findings suggest that empirical scholarship needs to be more mindful in its tendency to deduct the efficiency of a clause from its prevalence and longevity. While the widespread use of contractual terms is certainly indicative of their ability to allocate a right to the party that values it the most, the inflexibility of the drafting process suggests that contracts will react only slowly to changes in the external framework that governs an

agreement. It is thus essential that we as researchers rely on other factors beyond historical and current prevalence to determine a clauses' efficacy.

Economies of scale suggest that the commoditization of legal services is inevitable and greatly increases the productivity of law firms. However, the results of this study imply that the commoditization needs to be implemented with reflection. Templates should be designed and periodically reviewed with great care in order to examine whether, on average, they serve the clients' interest. Substantive knowledge and understanding is indispensable to make adjustments when needed and this expertise has to be fostered within the law firm by ensuring that more experienced lawyers transfer their knowledge to young attorneys. At the same time, inhouse counsel has the possibility to ensure a minimum quality of the written agreements by enacting firm-wide standards. The data suggests that most companies lack such a policy, even though their implementation is of relatively low cost and can provide for a standard that is far superior to the opaque default rules in most jurisdictions. In addition, enacting a firm-wide policy on the use of forum selection clauses is also likely to increase the importance that parties attach to these clauses because drafters will become accountable for their choices. For instance, if it is firm policy to litigate disputes in New York but the other party wants to include a choice-of-forum clause to litigate in California, those who negotiate the contract will need convincing reasons to derogate from firm policy, making it more likely that choice-of-forum clauses become a bargaining chip that adequately reflects their economic importance than is implied under current practice.

Law schools have often been confronted with claims that they would not prepare students well enough for practice, raising doubts as to whether a legal education should not best be considered a mere signaling device while the real conveyance of practically relevant skills happens on the job (Thorner, 1987; Garth and Martin, 1993). This criticism has intensified with the 2007 Carnegie Report on the Legal Profession (see also Thies, 2010). The identified heterogeneity and suboptimality in drafting practices demonstrate an opportunity to create value through legal services and the important role law schools have in it. Interviews conducted by Gulati and Scott (2012) suggest that young attorneys sometimes spot supposed deficiencies in contracts but conclude that the fault most likely falls with them, given that the template is time-tested and must have been written by lawyers much more experienced than they are. This is especially concerning given that young lawyers are more likely to be free of a status quo bias, putting them in an ideal position to reassess contract clauses independently and thus correct insufficient but cemented terms. Law schools are best situated to break the resulting cycle of the perpetuation of

inefficient contract terms by providing students with the necessary skills and confidence to reevaluate the efficiency of contractual terms in commonly used drafts and propose adjustments where necessary, in turn increasing the value creation of their graduates to their client (Gilson, 1984).

2.6 Conclusion

Relying on a novel data set including information on material contracts, federal litigation and corporate counsel, this chapter finds strong evidence to support the hypothesis that contracts even of the most sophisticated parties are drafted suboptimally with regards to forum selection clauses and that this deficiency is the consequence of a reliance on flawed templates by external counsel. The inefficiencies are economically substantial but avoidable by making a conscious effort to enact firm-wide policies on forum selection clauses and by increasing awareness for the deficiencies associated with an inflexible and overly path-dependent drafting procedure.

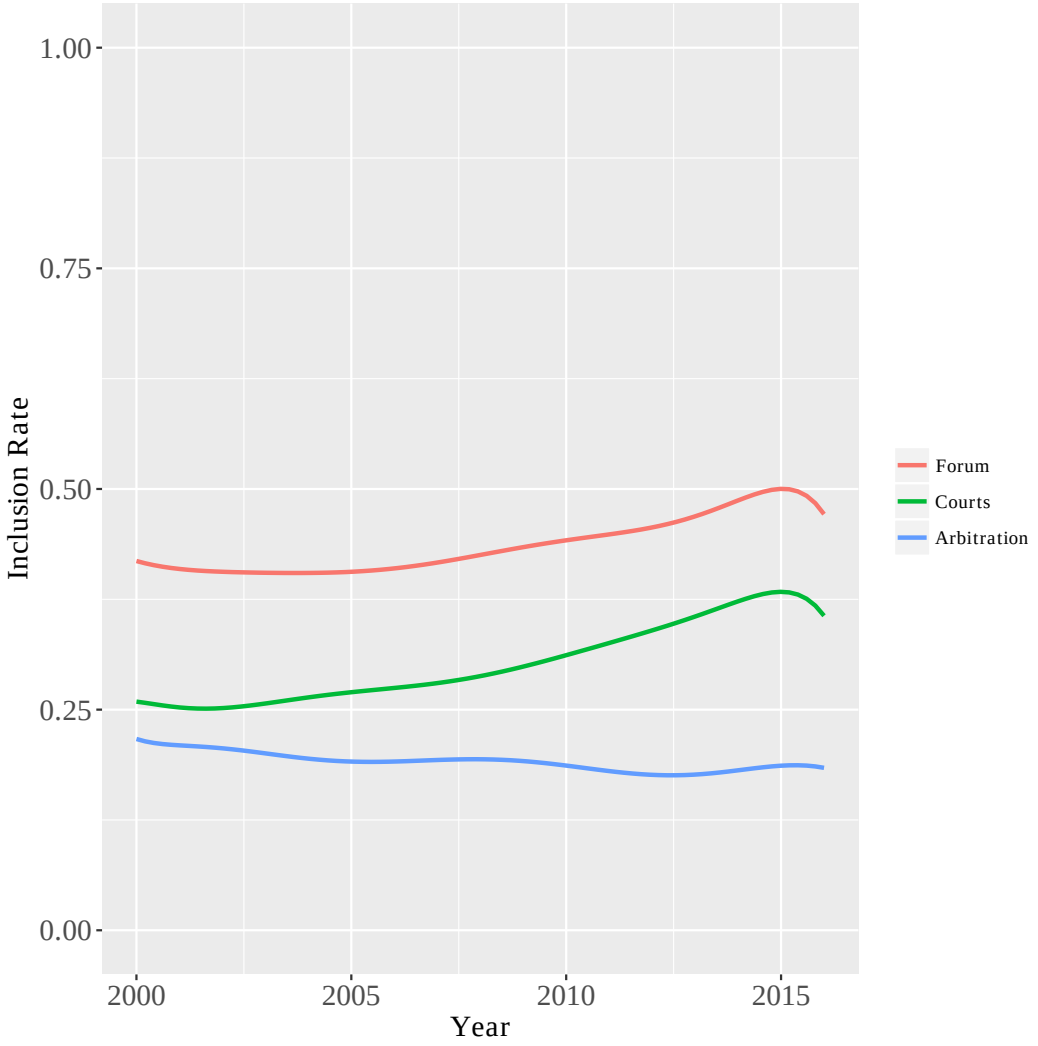
2.7 Figures and Tables

Table 2.1: Summary Statistics

	Min	Max	Mean	SD	Med	IQR
Year	2000	2016	2008	4.35	2008	7
Forum Selection Clause	0	1	0.44	0.50	0	1
Courts Selection Clause	0	1	0.30	0.46	0	1
Court Clause Length	29	809	220	154	181	196
Arbitration Clause	0	1	0.19	0.39	0	0
Arb. Clause Length	27	1,128	324	245	255	313
Governing Law Clause	0	1	0.75	0.43	1	1
GLC Length	16	401	79.45	76.76	47	66
U.S.–U.S. Contract	0	1	0.89	0.31	1	0
U.S.–Foreign Contract	0	1	0.10	0.30	0	0
Foreign–Foreign Contract	0	1	0.01	0.09	0	0

Summary Statistics for non-categorical variables used in the analysis.

Figure 2.1: Use of Forum Selection Clauses Over Time



This graph depicts the prevalence of forum selection clauses from 2000 to 2016. Lines have been smoothed using an 8th degree polynomial.

Table 2.2: Industries

Industry	Obs.	Freq.	Courts	Arb.	Forum	Law
Agriculture	13	0.00	0.23	0.38	0.54	0.85
Services	99,596	0.20	0.32	0.21	0.46	0.75
Other	9,360	0.02	0.33	0.18	0.46	0.83
Mining	31,451	0.06	0.32	0.18	0.44	0.77
Transportation	45,472	0.09	0.31	0.18	0.43	0.74
Manufacturing	175,413	0.35	0.30	0.19	0.43	0.73
Trade	40,671	0.08	0.30	0.18	0.43	0.75
Finance	100,608	0.20	0.28	0.19	0.42	0.77
Construction	5,268	0.01	0.30	0.15	0.41	0.74

Industries in the data set, as well as the frequency of their occurrence, their choice-of-forum clause frequency and their choice-of-law clause frequency. Sorted by overall choice-of-forum clause frequency.

Table 2.3: Agreement Types

Type	Obs.	Freq.	Courts	Arb.	Forum	Law
Joint Venture	1,399	0.00	0.26	0.44	0.61	0.74
M&A	29,013	0.06	0.43	0.28	0.61	0.87
Licensing	9,431	0.02	0.32	0.39	0.61	0.80
Loan	57,086	0.11	0.53	0.10	0.58	0.80
Sales	89,257	0.18	0.44	0.22	0.56	0.83
Security	21,084	0.04	0.44	0.09	0.49	0.82
Employment	108,313	0.21	0.23	0.33	0.49	0.75
Consulting	7,860	0.02	0.25	0.28	0.48	0.78
Transportation	1,313	0.00	0.30	0.24	0.47	0.70
Other	12,758	0.03	0.40	0.09	0.46	0.91
Lease	16,076	0.03	0.21	0.27	0.40	0.66
Negotiable Instrument	14,024	0.03	0.36	0.05	0.39	0.81
Legal	10,002	0.02	0.30	0.11	0.38	0.71
Incentives	130,236	0.26	0.12	0.09	0.19	0.62

Agreement types in the data set, as well as the frequency of their occurrence, their choice-of-forum clause frequency and their choice-of-law clause frequency. Sorted by overall choice-of-forum clause frequency.

Table 2.4: Most Popular Court Forums

Forum	Mean FSC	Mean GLC	Overlap
New York	0.37	0.26	0.91
Delaware	0.11	0.15	0.89
California	0.08	0.09	0.87
Texas	0.05	0.05	0.89
Florida	0.03	0.03	0.91
Illinois	0.03	0.02	0.89
Nevada	0.02	0.02	0.92
New Jersey	0.02	0.02	0.94
Massachusetts	0.02	0.02	0.92
Pennsylvania	0.02	0.02	0.86
Ohio	0.01	0.02	0.89
Colorado	0.01	0.01	0.88
Minnesota	0.01	0.01	0.84
Georgia	0.01	0.02	0.91
Virginia	0.01	0.01	0.80

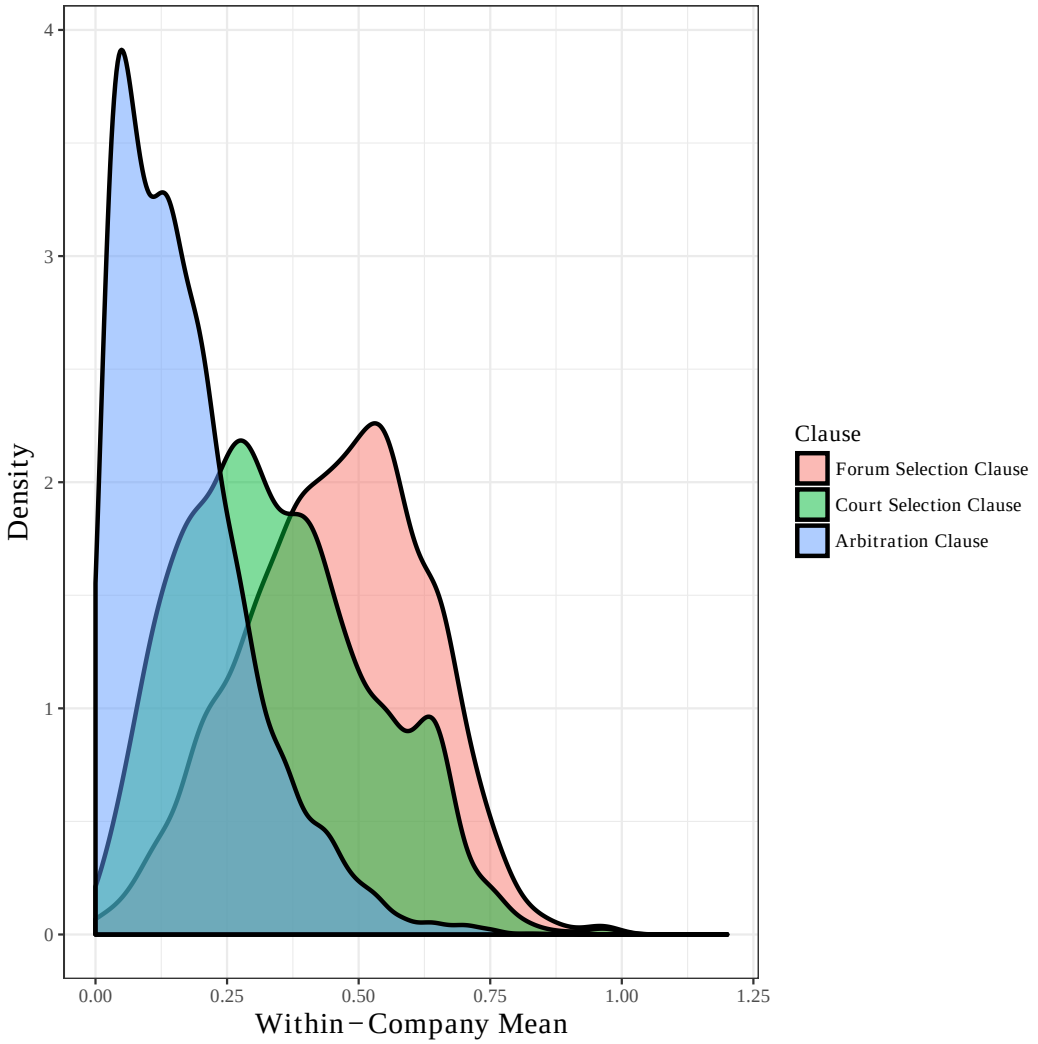
The table depicts the 15 most popular court forums, conditional on the parties opting for court litigation (Mean FSC), and the most popular substantive laws, conditional on contracts including a choice-of-law provision (Mean GLC). 'Overlap' refers to the probability that the substantive law matches the forum, given that a contract includes both a choice-of-law and a choice-of-forum clause.

Table 2.5: Law Firms

Law Firm	# Contracts	Forum	Courts	Arbitration	Law Clause
Latham & Watkins	4,509	0.76	0.65	0.25	0.97
Skadden, Arps, Slate, Meagher & Flom LLP	4,438	0.82	0.69	0.27	0.97
Kirkland & Ellis LLP	3,067	0.73	0.64	0.22	0.97
Simpson Thacher & Bartlett	2,856	0.81	0.71	0.21	0.97
Greenberg Traurig, LLP	2,470	0.77	0.68	0.20	0.95
Weil, Gotshal & Manges LLP	2,085	0.81	0.73	0.21	0.96
Wilson Sonsini Goodrich & Rosati	1,931	0.79	0.68	0.31	0.96
Jones Day	1,859	0.82	0.74	0.25	0.94
Vinson & Elkins	1,858	0.76	0.66	0.23	0.97
Shearman & Sterling LLP	1,834	0.72	0.65	0.15	0.98
Wachtell, Lipton, Rosen & Katz	1,730	0.83	0.74	0.21	0.96
Dla Piper Ltd	1,681	0.82	0.70	0.27	0.95
Davis Polk & Wardwell	1,667	0.81	0.79	0.14	0.97
Gibson Dunn & Crutcher	1,627	0.78	0.67	0.26	0.97
Sidley Austin	1,622	0.79	0.73	0.23	0.97
Morgan, Lewis & Bockius LLP	1,596	0.74	0.64	0.28	0.96
Sichenzia Ross Friedman Ference LLP	1,553	0.82	0.74	0.16	0.96
Ropes & Gray LLP	1,513	0.72	0.64	0.19	0.96
Mayer Brown	1,510	0.74	0.68	0.17	0.95
Akin Gump Strauss Hauer & Feld LLP	1,481	0.77	0.65	0.26	0.96
Paul Hastings LLP	1,395	0.82	0.70	0.28	0.94
Morrison & Foerster L.L.P.	1,381	0.79	0.68	0.26	0.96
O'Melveny & Myers LLP	1,346	0.77	0.60	0.30	0.96
Sullivan & Cromwell LLP	1,345	0.82	0.73	0.23	0.97
Cravath, Swaine And Moore LLP	1,296	0.85	0.82	0.14	0.97
Goodwin Procter LLP	1,291	0.80	0.69	0.29	0.96
Cahill Gordon & Reindel LLP	1,188	0.74	0.72	0.07	0.98
Paul, Weiss, Rifkind, Wharton & Garrison LLP	1,170	0.82	0.74	0.21	0.97
Willkie Farr & Gallagher LLP	1,167	0.79	0.71	0.22	0.96
Proskauer Rose LLP	1,146	0.75	0.66	0.20	0.96

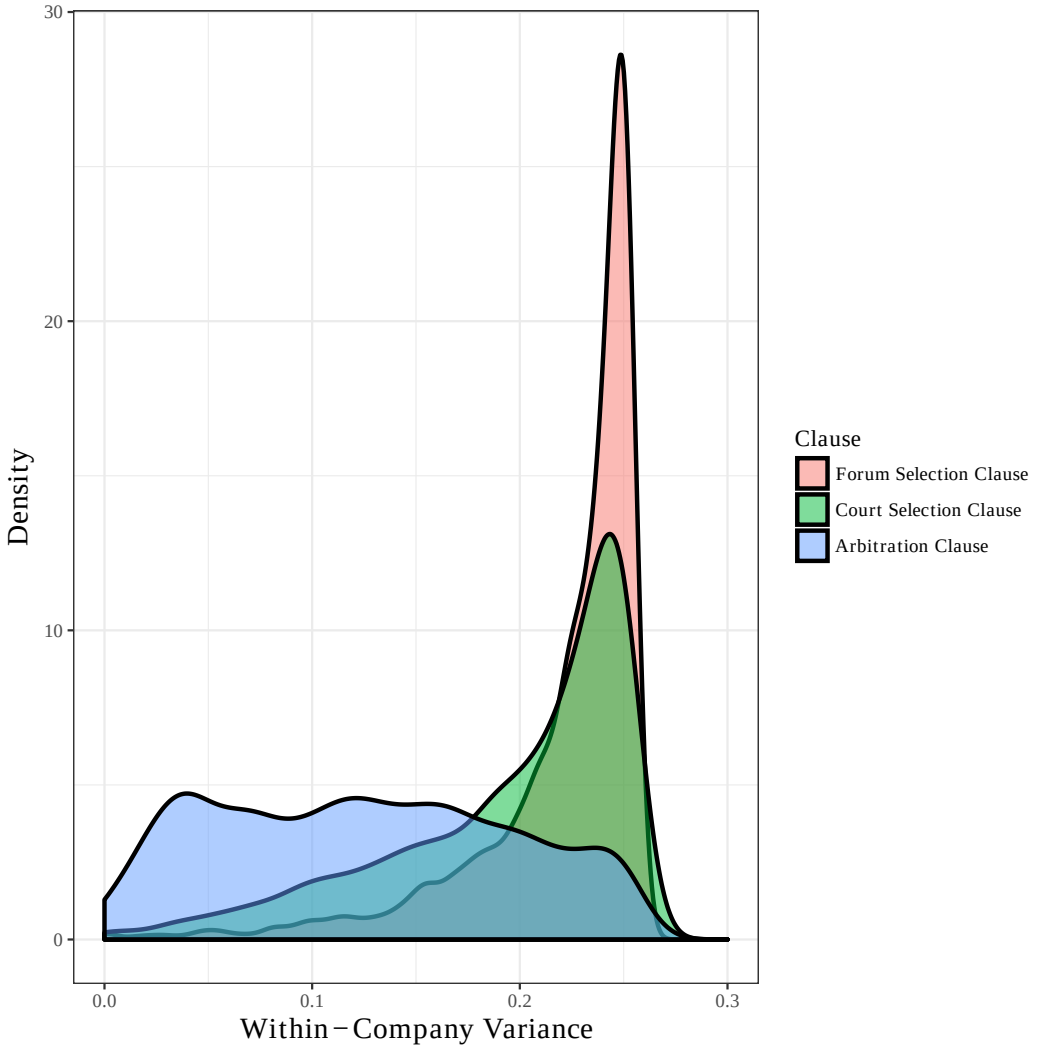
The table details law firms identified in the contracts.

Figure 2.2: Density Plot of Within-Company Means



This graph depicts a weighted density plot of the within-company usage rates for different forum selection clauses. The density is weighted by the number of contracts for each company.

Figure 2.3: Density Plot of Within-Company Variance



This graph depicts a weighted density plot of the within-company variance in usage of different forum selection clauses. The density is weighted by the number of contracts for each company.

Table 2.6: Law Firm / General Counsel Influence on Forum Selection Clauses

	Law Firm (N =1,184)			General Counsel (N = 348)		
	¬ Same	Same	Diff	¬ Same	Same	Diff
Any Forum	0.41	0.08	0.33***	0.26	0.18	0.08***
			(0.01)			(0.007)
Courts	0.48	0.09	0.39***	0.27	0.17	0.10***
			(0.01)			(0.007)
Arbitration	0.12	0.04	0.08***	0.09	0.07	0.02***
			(0.007)			(0.005)

Note: *p<0.05; **p<0.01; ***p<0.001

The table analyzes matched pairs of contract dyads. The matching algorithm is exact one-to-one matching, weighted if multiple perfect matches exist. Estimates are average probabilities that two contracts differ in their use of forum selection clauses if they are written by the same law firm / general counsel ("Same") or by different law firms / general counsels ("¬ Same"). Significance levels are obtained via a weighted Student's t-test.

Table 2.7: Clause Stickiness by Industry / Contract Type

	¬ Same LF	Same LF	Difference
		Industry	
Finance	0.82	0.04	-0.79***
			(0.01)
Construction	0.50	0.00	-0.50
			(0.25)
Other	0.56	0.16	-0.40***
			(0.06)
Trade	0.59	0.25	-0.34***
			(0.07)
Mining	0.33	0.06	-0.27***
			(0.05)
Agriculture	0.21	0.05	-0.17***
			(0.02)
Manufacturing	0.23	0.12	-0.10***
			(0.02)
Services	0.15	0.10	-0.05*
			(0.02)
Transportation	0.16	0.15	-0.01
			(0.04)
		Contract Type	
Security	0.97	0.00	-0.97***
			(0.01)
Legal	0.31	0.00	-0.31***
			(0.07)
Loan	0.44	0.20	-0.24***
			(0.02)
Sales	0.28	0.04	-0.24
			(0.01)
Other	0.00	0.00	
Lease	0.00	0.00	
M&A	0.06	0.07	0.02
			(0.01)

Note: *p<0.05; **p<0.01; ***p<0.001

The table analyzes matched pairs of contract dyads. The matching algorithm is exact one-to-one matching, weighted if multiple perfect matches exist. Estimates are average probabilities that two contracts differ in their use of forum selection clauses if they are written by the same law firm ("Same LF") or by different law firms ("¬ Same LF"). Significance levels are obtained via a weighted Student's t-test.

Table 2.8: Transition in Forum Selection Clauses

Contract 1	Contract 2	$\Longrightarrow$	$\Longleftarrow$	Difference
No Clause	Any Forum	12%	6%	$6\%^{***}$
				(0.014)
No Clause	Courts	12%	7%	$5\%^{**}$
				(0.014)
No Clause	Arbitration	2%	4%	$-2\%^{*}$
				(0.013)

Note: *p<0.05; **p<0.01; ***p<0.001

The table depicts the frequency with which different transitions in forum selection clause usage between two contracts occur. The right arrow indicates a transition from the clause named under "Contract 1" to the clause named under "Contract 2". The left arrow indicates a transition from "Contract 2" to "Contract 1". For observations that were observed at the same point in time receive, both transitions ("left" and "right") are considered and each is given half weight. Significance levels are computed using a weighted Student's t-test.

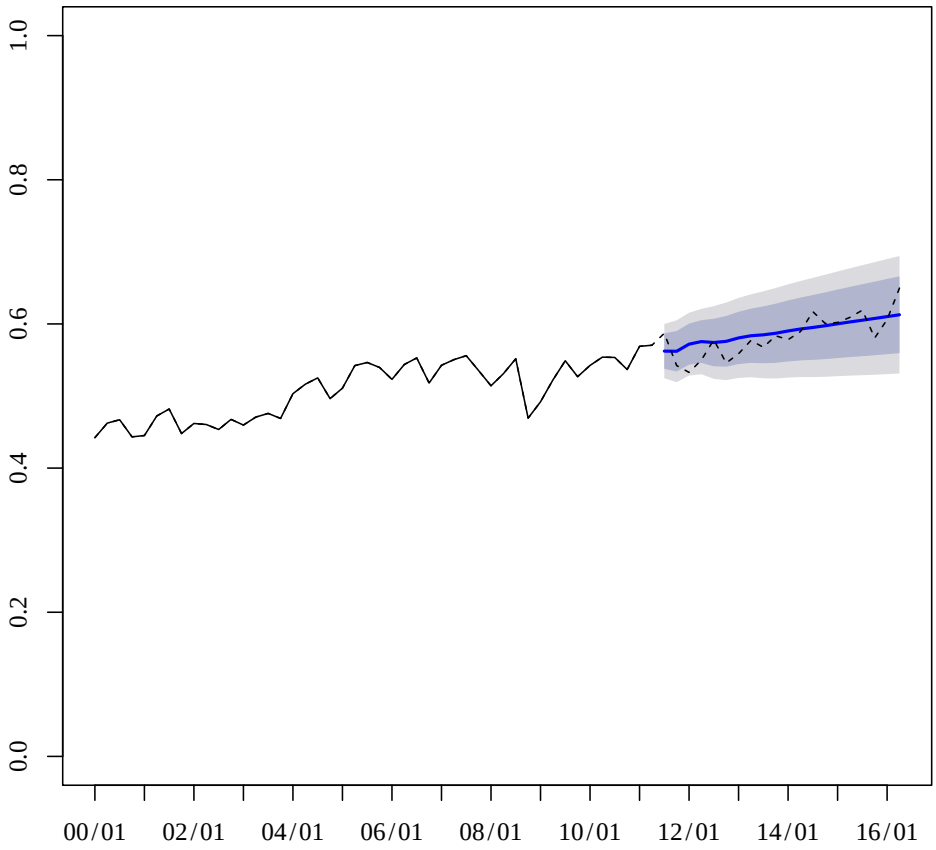
Table 2.9: Textual Similarity of Clauses

Similarity Measure	¬ Same LF	Same LF	Difference
Cosine Similarity	0.814	0.900	-0.086^{***} (0.023)
Levensthein Distance	155.929	87.935	67.994^{**} (25.265)

Note: *p<0.05; **p<0.01; ***p<0.001

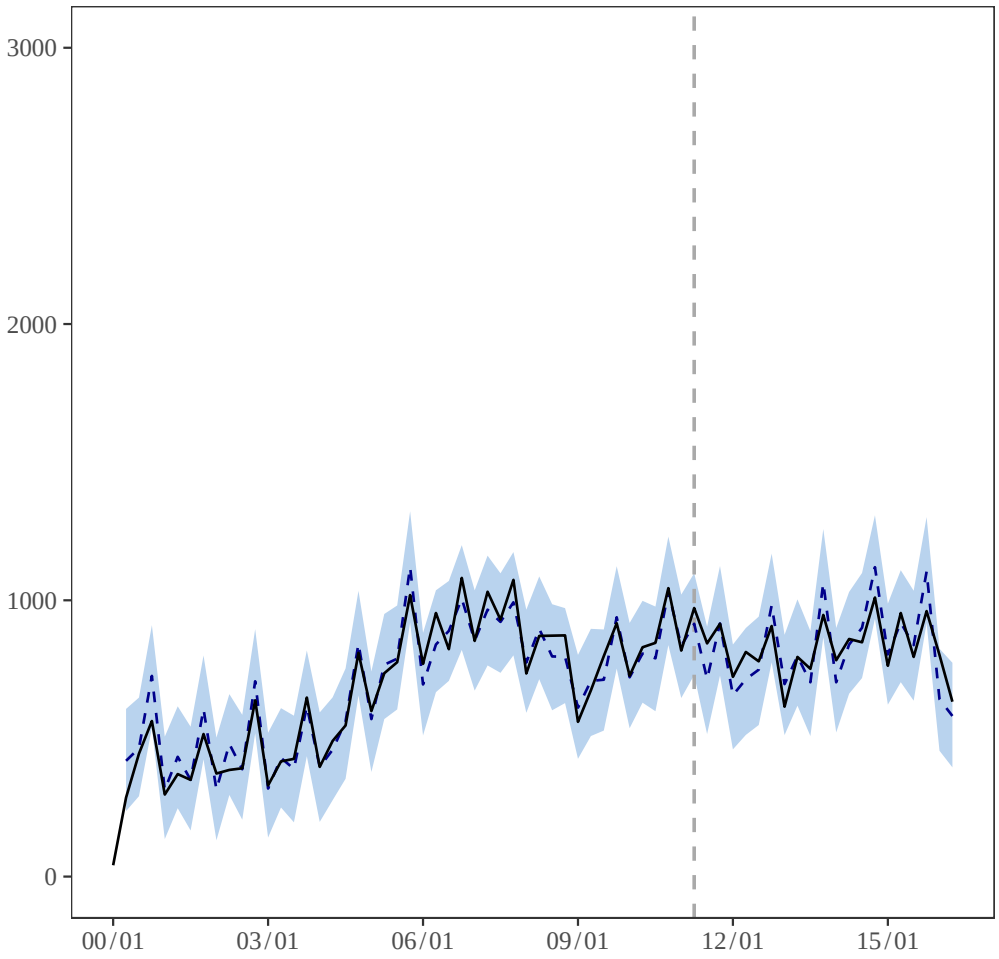
The table depicts the similarity between two forum selection clauses drafted by different law firms (¬ Same LF) and by the same law firm (Same LF). Significance levels for the differences are computed using a Student's t-test.

Figure 2.4: ARIMA-Model of Forum Selection Clause Usage



This graph depicts the prevalence of choice-of-forum clauses prior to and after Goodyear. Predictions in blue are based on pre-Goodyear data.

Figure 2.5: Synthetic Controls Method



This graph depicts the prevalence of outside court selection clauses prior to and after Goodyear. It also uses inside court selection clauses to create a synthetic unit which has never been treated.

Table 2.10: Learning Effects

	Dependent Variable:		
	# Arbitration Clauses		
	(1)	(2)	(3)
# Disputes	0.019		0.010
	(0.033)		(0.034)
# Disputes$_{-1}$	0.037		0.034
	(0.034)		(0.034)
# Disputes$_{-2}$	0.100**		0.092**
	(0.033)		(0.034)
Settlement Amount		3.380*	3.365*
		(1.319)	(1.335)
Settlement Amount$_{-1}$		1.990	1.823
		(1.468)	(1.484)
Settlement Amount$_{-2}$		2.650*	2.110
		(1.132)	(1.151)
Constant	0.266***	0.265***	0.266***
	(0.067)	(0.067)	(0.067)
Offset	✓	✓	✓
Company Fixed Effects	✓	✓	✓
Year Fixed Effects	✓	✓	✓
Observations	105,142	105,142	105,142
Companies	6,186	6,186	6,186
Log Likelihood	−39,032	−39,032	−39,028
Akaike Inf. Crit.	78,099	78,099	78,097
Note:			*p<0.05; **p<0.01; ***p<0.001

9.83* 8.06* 18.68** 10.62* 8.76*

The table depicts the results of a two-way fixed effects panel regression of arbitration clause usage on federal litigation and settlement amounts in those litigations. Joint significance indicated by vertical brackets spanning multiple lines. The χ^2 statistic for the joint significance test as well as the significance level is reported vertically. The logged number of contracts is included as an offset. Standard errors derived from the observed information matrix. Settlement amounts in bio USD.

Table 2.11: Learning Effects, Robustness

	Dependent Variable:				
	# Arbitration Clauses				
	(4)			(5)	
# Disputes	−0.008			−0.013	
	(0.028)			(0.028)	
# Disputes$_{-1}$	0.028			0.027	
	(0.028)	9.18**		(0.028)	8.34*
# Disputes$_{-2}$	0.078**			0.073**	
	(0.027)			(0.028)	
Settlement Amount	1.571			1.128*	
	(1.113)			(1.106)	
Settlement Amount$_{-1}$	1.976			1.156	
	(0.994)	5.83*	17.42**	(0.978)	3.74
Settlement Amount$_{-2}$	1.370			1.023	
	(1.025)			(1.016)	
Constant	1.125***			0.527 ***	
	(0.086)			(0.074)	
Offset	✓			✓	
Company Effects	Fixed			Random	
Year Fixed Effects	✓			✓	
Observations	699,754			105,142	
Companies	41,162			6,186	
Log Likelihood	−87,085			−58,726	
Akaike Inf. Crit.	174,213			117,499	
Note:				*p<0.05; **p<0.01; ***p<0.001	

The table depicts the results of two robustness checks for the full model (3). Model (4) includes all companies, including those who have never reported any litigation. Model (5) uses assumes random instead of fixed company effects. Joint significance indicated by vertical brackets spanning multiple lines. The χ^2 statistic for the joint significance test as well as the significance level is reported vertically. The logged number of contracts is included as an offset. Standard errors derived from the observed information matrix. Settlement amounts in bio USD.

Table 2.12: Learning Effects, Substantive Significance

	(1)	(2)	(3)
# Disputes	19%*		17%*
Settlement Amount		10%*	9%*

Note: *p<0.05; **p<0.01; ***p<0.001

The estimated influence of litigation experience on the probability that a contract includes an arbitration clause. The estimates depict derived from comparing the average company with no prior disputes to an average company that experienced a single material dispute. Similarly, they compare the average company having paid no settlement amount to the average company having paid 10 million USD in settlement amount. All covariates are fixed at their mean. The company fixed effect is assumed to be 0.

Chapter 3

Giving the Treaty a Purpose: Comparing the Durability of Treaties and Executive Agreements

3.1 Introduction

The U.S. is an international anomaly in that it has two commitment devices to conclude agreements with other states, the executive agreement and the treaty.(Hathaway, 2008) The existence of two seemingly parallel instruments has drawn much academic interest. While the legality of the executive agreement has traditionally been challenged for its lack of textual support in the Constitution, the treaty has always had a difficult time justifying its existence as a U.S. policy instrument for practical reasons.

Indeed, as early as in the1940s, both the scholarly and the public debate wrestled with the question of whether there still is a place for the treaty in a day and age where a nation's success increasingly relies on its ability to cooperate with other states (McClure, 1941; Borchard, 1942, 1944). More recently, the treaty has re-emerged into the crosshair of its critics. During the Obama administration, only 19 treaties have been approved by the Senate, the lowest number of approvals during a presidential term since president Ford. Criticized mainly for the high hurdles its ratification poses, some argue that it is time to fully replace the treaty with the executive agreement, a supfosedly more flexible and easier to conclude commitment device(Hathaway, 2008). This demand is based on the assumption that treaties and executive agreements are interchangeable commitment devices. If the domestic law and the international community view treaties and executive agreements as interchangeable, the arguments goes, then there is no reason to try to overcome all the institutional hurdles created through the Advice and Consent process in favor of an instrument that offers no advantages in return.

But as important as the fundamental assumption of substitutability of treaties and executive agreements is, there have been few attempts to verify it. Instead, much of the discourse is dominated by doctrinal arguments, examining whether the Constitution limits the use of executive agreements to certain issue areas and makes the treaty the exclusive instrument in others. However, the past has taught us that neither the courts[1] nor the State Department[2] show much concern for delineating both instruments based on constitutional grounds, calling into question whether doctrine alone can provide a strong justification in favor of or against preserving the treaty as a commitment tool.

[1] *See* e.g. United States v. Belmont, 301 U.S. 324 (1937); United States v. Pink, 315 U.S. 203 (1942); Dames & Moore v. Regan, 453 U.S. 654 (1981).

[2] Former U.S. State Department Legal Adviser Harold H. Koh provides two reasons for why the state department uses treaty, namely comity towards congress and the "powerful message" that is sent to the world through the treaty ratification process. He considers the question of legal substitutability as the "long-dominant" view (Koh, 2012, 91–92).

This chapter takes a different approach. Instead of asking whether treaties and executive agreements are de jure interchangeable, it examines whether both instruments are de facto interchangeable. It does so by considering whether treaties and executive agreements lead to different outcomes. If each commitment device leads to a different result, then this implies that the devices are qualitatively different from one another and abolishing one of the instruments cannot be done without incurring adverse consequences. If instead the use of both instruments leads to identical outcomes, there is no reason to preserve the treaty as a policy instrument, as the executive agreement should be able to fully perform the treaties' functions.

Motivated by considerations in relevant literature, the outcome measure of choice is agreement reliability, measured as the duration for which an agreement is in force. Based on all 7,966 treaties and executive agreements that have been reported in the Treaties in Force Series from 1982 to 2012, this chapter is the first to demonstrate that agreements concluded in the form of the treaty last significantly longer than agreements concluded as executive agreements. The result holds even after controlling for a number of covariates that could influence the durability of the agreement. The findings imply that treaties are a more reliable commitment device than the executive agreement. By using a treaty, a president signals a higher level of commitment to the underlying promise than through the use of an executive agreement. As a consequence, negotiation partners will put more trust in promises concluded as treaties than in those concluded as executive agreements. Giving in to demands of abolishing the treaty would make it impossible for presidents to indicate how dedicated they are to the underlying promise, in turn hampering the conclusion of agreements which require particularly high levels of commitment.

The rest of the chapter proceeds as follows: Section 3.2 lays out the institutional foundation of the different commitment devices and presents theories on how treaties differ from executive agreements. Section 3.3 describes the data and methodology used in this study and presents summary statistics. Section 3.4 presents the results of a formal test of instrument durability, while Section 3.5 discusses them. A last section concludes.

3.2 Theory

In order to conclude an agreement that is recognized as a binding international obligation, it is now recognized that the U.S. constitution provides two different mechanisms. The first option is the traditional treaty. Treaties follow Article

II's Advice and Consent procedure, which implies that, while a treaty can be signed by the executive, it still requires a two-third majority in the Senate in order to be ratified and become binding international law.[3]

The second option to conclude international contracts is the executive agreement. Among the executive agreements, there are again different types. *Congressional*–executive agreements require a simple majority in both the House of Representatives and the Senate. They are used in subject areas in which the executive does not have sole competences. Congressional approval can be obtained after the agreement was negotiated, as was the case with the North American Free Trade Agreement or the Uruguay Round Agreements of the General Agreement on Tariffs and Trade. However, it is much more common for Congress to provide broad authorization to the president *ex ante* through broader statutory authorization (Hathaway, 2008).

If the executive has the competence to make policy without referring to Congress, the president may use *sole* executive agreements. Such areas encompass, among others, issues under the president's general executive authority or the function as commander-in-chief of the armed forces. Sole executive agreements do not require congressional approval, but, like congressional-executive agreements, need to be reported to Congress subject to the "Case-Zablocki Act".[4]

2.1 Legal Substitutability

From an international legal viewpoint, treaties and executive agreements are perfect substitutes. Indeed, international law does not recognize the term "executive agreement". The term "treaty" is more broadly defined than in the domestic context of the U.S. The Vienna Convention on the Law of Treaties Art. II (1) (a) states that any written agreement between states governed by international law qualifies as a "treaty" and thus, creates a binding legal commitment. Since both U.S. treaties and executive agreements meet this definition, there is no legal difference between either of those commitment devices

[3]U.S. Const. art. II, § 2, cl. 2.

[4]It is important not to "fetishize" this triptych of treaties, congressional-executive agreements and sole executive agreements. Indeed, most recent scholarship has called attention to its unsuitability in categorizing two very recent agreements, namely the Paris Climate Change Agreement and the Iran Nuclear Deal (Galbraith, 2017; Koh, 2017). However, in the author's view, complexity is a cost that needs to be justified and since this chapter is interested in the *substantive* difference between executive agreements and treaties concluded between 1982 and 2012 and does not seek to discuss or illegitimize novel forms of international agreements, there is little use in moving beyond this traditional distinction.

from the perspective of international law.[5] Domestically, the issue of legal substitutability has traditionally been more controversial. To be sure, there is a broad consensus that Congressional participation cannot fully be removed by substituting the treaty for the sole executive agreement. However, views on the interchangeability of treaties and congressional-executive agreements are less harmonious. The Constitution does not expressly mention the existence of an instrument that resembles today's congressional-executive agreement, resulting in a debate about how to interpret the silence. To early proponents, it was largely sufficient to show that interchangeability offers flexibility and best describes the practice of U.S. foreign policy to argue that treaties and congressional-executive agreements should act as legal substitutes (McClure, 1941; Wright, 1944; McDougal and Lans, 1945a,b). Later arguments rested on the idea of the existence of "constitutional moments" that would allow constitutional interpretation to be informed by consistent practice of the president, Congress and the Supreme Court. Such moments, particularly formed through practice in the 1940s, are believed to have transformed the meaning of the Treaty Clause, providing a constitutional basis to the congressional-executive agreement (Ackerman and Golove, 1995).

In contrast, opponents of legal substitutability highlight the lack of clear textual support. An extreme view holds that the Treaty Clause is clear in making Advise and Consent the exclusive method for the approval of international agreements (Borchard, 1945; Berger, 1972; Tribe, 1995). A more moderate view suggests that treaties and congressional-executive agreements both have their own and exclusive areas of applicability. The argument rests on the idea that the U.S. constitution has conferred limited powers upon Congress and the executive and that executive agreements can only be used within this limited scope. Treaties as the default tool for matters in foreign affairs are not similarly constrained. Thus, if a matter of foreign policy falls outside of the competences that have been conferred upon Congress, the treaty is held to be the exclusive instrument through which legally binding commitments can be made (Yoo, 2001).

Even though one might find appeal in the rationale underlying the analysis of those arguing against substitutability, the predominant view has long been that treaties and congressional-executive agreements are perfect legal substitutes under domestic law (Koh, 2012). This view is not only supported by several court decisions,[6] but is also reflected in Restatement (Third) of the

[5]While the U.S. is not a party to the VCLT, the State Department effectively views both treaties and executive agreements as meeting the VCLT's definition (Rovine, 1974; Frankowska, 1987).

[6]A general overview of the treatment of the executive agreement by the Supreme Court

Foreign Relations Law of the United States § 303 comment (e), in which the American Law Institute states:

> *The prevailing view is that the Congressional-Executive agreement can be used as an alternative to the treaty method in every instance. Which procedure should be used is a political judgment, made in the first instance by the president, subject to the possibility that the Senate might refuse to consider a joint resolution of Congress to approve an agreement, insisting that the president submit the agreement as a treaty.*[7]

2.2 Differences in Reliability

The view that treaties and executive agreements can be considered perfect legal substitutes naturally raises the question if and why the United States needs two legal instruments that regulate the same types of international relationships. Indeed, scholarly writings have repeatedly left readers with doubt as to why the U.S. should not abolish the treaty in favor of the congressional-executive agreement. In the late 2000's, these doubts transformed into strong normative claims. With an article fittingly titled "Treaties' End", Hathaway characterizes the existence of two conflicting commitment tools as an international anomaly that ultimately undermines the legitimacy and reliability of agreements the U.S. concludes with other nations (Hathaway, 2008). According to her analysis, treaties are less reliable commitments than executive agreements because the treaty makes it difficult for presidents to credibly tie their hands. In particular, even after ratification, the treaty would offer the president two additional possibilities to renege on his promise, in turn making it difficult for negotiation partners to rely on promises concluded in the form of the treaty.

The first of these two opportunities to renege is rooted in the fact that non-self-executing treaties have to follow a two-step process to become enforcable U.S. law. That is, after ratification, non-self-executing treaties require additional implementation through a legislative act for which a simple majority in both the House and the Senate is required. Compare this to the executive

and numerous further references is provided by Van Alstine (2011).

[7] Restatement (Third) of the Foreign Relations Law of the United States § 303 cmt. e (1987). While it should be noted that the approved draft of the Restatement (Fourth) is conspicuously silent on the matter of interchangeability, so far there is no indication that this silence can provide new wind to those arguing against interchangeability.

agreement, which is self-executing by default and for which otherwise the implementing legislation can be conducted in the same step as the ratification. It is argued that the treaties' two-step process makes it possible for the president to renege on his promise after ratification, whether intentionally or because the domestic political costs are too high. The second argument is that treaties, unlike congressional agreements, can be more easily withdrawn from by the president unilaterally, whereas the withdrawal from congressional-executive agreements requires congressional participation. Again, this would allow presidents to renege on their promise even after a treaty has gone through the Advice and Consent process.

To Hathaway, the consequence of a difference in reliability is that the treaty as the less reliable instrument should be abandoned in favor of the executive agreement.[8] This claim resonated with some scholars of international law (Bradley, 2015) and even sparked vivid reactions in public publishing outlets. For example, in 2014, under the title The End of Treaties?, the online companion of the American Journal of International Law published several essays by prominent international legal scholars and officials in the State department, discussing whether the treaty will have any place in the future of U.S. foreign policy.[9] By today, it might be fair to describe it as the predominant view that treaties have (almost) no relevant function as an international policy tool that could not be similarly fulfilled by the executive agreement.

An opposing view assumes treaties not to be the less reliable, but the more reliable instrument. The argument is based on the assumption that the higher legislative hurdle to conclude a treaty imposes additional political costs on the president that are the consequence of having to assure a two-thirds majority in the Senate. Hence the constitutional requirements would make treaties a more costly commitment device that, in many instances, only those presidents are willing to incur that have the intent to follow through on their promise. Ultimately, the availability of the treaty allows presidents to indicate the seriousness of their commitment, and negotiation partners to distinguish between those presidents that are strongly committed to follow through on their promise (and thus use a treaty), and those who have weaker levels of commitment and thus are only willing to incur the lower costs of the congressional-executive agreements (Martin, 2000, 2005; Setear, 2002).

[8]It should be noted that this conclusion is partially at odds with signaling theory. Even under the assumption that treaties are the less reliable commitment device, there may still be value in retaining both instruments in order to be able to distinguish between player types. See the discussion on signaling costs in the next paragraph.

[9]AJIL Unbound 108 (2014), available at `https://www.cambridge.org/core/journals/american-journal-of-international-law/ajil-unbound`.

To embed this line of reasoning in the more established game theoretical vocabulary, the argument is that the availability of two signaling devices with differential costs gives rise to separating equilibria in which only the "reliable" players use the treaty, whereas other types rely on the congressional-executive agreements.[10] While not explicitly addressed by Martin, it is worthwhile noting that an implication of this model is that the difference in signaling costs is especially pronounced between treaties and executive agreements that Congress has authorized *ex ante*. As pointed out, *ex ante* authorization is by far the most common form of statutory authorization and conditional on statutory authorization having been granted, the political costs of securing additional votes would then be zero.[11] In addition to supposedly higher signaling value of treaties, some scholars also cast doubt on the importance of the rationale that treaties can be withdrawn from more easily by the president. As Koremenos (2016) and Galbraith (2017) point out, many agreements in the UN Treaty Collection have escape clauses and withdrawal provisions that would allow a president to legally exit an agreement, regardless of the form in which it has been concluded. As such, it is alleged that the significance of unilateral withdrawal for the reliability of an agreement might be overstated.

A third view advanced by Yoo (2011) sees differences between treaties and executive agreements not only in the costs of signaling, but also in the type of information that is signaled. Yoo argues that treaties are a tool to remove information asymmetries regarding a state's utility function. His leading example is a potential military conflict between the U.S. and China over a territory and negotiations surrounding how this territory would be divided up. In Yoo's view, the domestic struggle for approval of a treaty leads negotiators to reveal information on their true beliefs about the probability with which they could win the war. Note that this is different from signaling the U.S. intention to comply with an agreement dividing up the territory. In this latter regard, Yoo agrees with Hathaway that executive agreements are more difficult to terminate than treaties, in turn arguing that the use of the executive agreement constitutes a more durable commitment.

A fourth view that is prevalent in the writings of political scientists makes the oftentimes implicit assumption that treaties and executive agreements are *de facto* interchangeable. Also labeled "evasion hypothesis", this view assumes that the president's main motivation for choosing one instrument over the other is presidential support in the Senate. If legislation is easy to push through the

[10]The term "reliable type" refers to a president that intends to comply with the agreement in the long term.

[11]However, other political costs could of course still be imposed.

Senate, the argument goes, presidents will rely on the treaty. If, however, securing a two-thirds majority poses difficult, the president can simply switch to the executive agreement without any significant consequences (Crenson and Ginsberg, 2007; Howell, 2015).

2.3 Prior Empirical Work

Empirical evidence on whether treaties are more or less reliable than executive agreements is limited and comes to vastly different conclusions. Margolis (1986) analyzes all international agreements concluded from 1943 to 1977 and argues that the choice between treaties and executive agreements is simply a function of the seat map in the House and Senate. A president who lacks support in the Senate would conclude congressional-executive agreements instead.

Martin (2005) conducts an analysis of 4,953 international agreements concluded between 1980 and 1999 and finds that not the seat map, but the value of the underlying relationship governed by the agreement is determinative for the choice of whether a president uses a treaty or an executive agreement. Here, value is proxied using an indicator for whether the agreement is multilateral, the GNP per capita of the contractual partner, as well as the total GNP. Her conclusions find further anecdotal support by Bradley and Morrison (2012), who recount instances in which important agreements such as SALT II and a nuclear reduction agreement with Russia were originally intended as executive agreements but have later been changed to treaties under pressure by the Senate.

A third study by Hathaway (2008) analyzes 3,119 agreements concluded between 1980 and 2000 and argues that the instrument choice is largely the product of historical path-dependence. Under this view, the prevalence of the executive agreement is the result of Congress' desire to reduce trade barriers in the post-WW II era, which necessitated giving the president more flexibility and authority in negotiating trade agreements. This has then lead to the conventional use of the executive agreements in trade (and financial) matters. In other subject areas such as human rights, the debate was highly politicized and Congress had no desire to give up what was perceived as the nation's sovereignty subject to the lower legislative bar set by the executive agreement. It is argued that these and similar historical events lead to the conventional use of executive agreements in some areas, while others remained dominated by treaties. These conventions established patterns that persist today, even though the underlying events that lead to their formation are no longer relevant or applicable.

All these studies follow a similar approach. The researcher analyzes the

environment in which an international agreement has been concluded and tries to identify patterns which are predictive of the instrument type that has been used. By uncovering choice patterns, the hope is to understand the motivation that drives the president's choice between executive agreements and treaties. If the choice pattern is reflective of a motivation that assigns different significance to treaties and executive agreements, that is taken as evidence that both instruments differ in their quality. However, note that the focus on choice patterns is a very indirect approach to identifying *de facto* differences in policy instruments that rests on a number of strong assumptions, such as a correct model specification and a causal relationship between identified patterns and hypothesized motives. Without making these assumptions, observed actions can be the result of a great number of different motivations, making it impossible to infer which instrument is more reliable.

This paper takes a more direct approach that does not require equally strong assumptions. At the heart of the inquiry into the political differences between treaties and executive agreements lies the question whether each instrument is associated with different results. It is thus instructive to shift the empirical focus from the analysis of choices to the examination of differences in outcomes. Outcomes of international agreements can be compared on a number of dimensions. One possible measure is the level of compliance with an agreement. However, comparing agreements based on compliance rates has several disadvantages in this context. Not only is "compliance" difficult to define. It is also notoriously hard to measure and verify in most contexts. Even if it was possible to accurately measure compliance, it would still leave open the question of how to compare levels of compliance across different agreements in different subject areas.

Motivated by the theoretical work previously discussed, this study instead compares treaties and executive agreements based on their reliability. The reliability of an agreement is measured in the form of its durability. Using durability as a proxy for reliability is justified for three reasons.

First, consider an alternative concept of reliability that one might have in mind, which is the ability for an agreement to withstand shocks in the political or economic environment (Downs and Jones, 2002). It is evident that the probability for shocks to occur increases with time and that agreements which are more resistant to changing circumstances are also those that last longer. Hence, durability is positively correlated even with this alternative concept of reliability. Second, from a purely practical perspective, the duration of a treaty can be measured objectively, whereas the competing concept of reliance would require the investigator to make a number of subjective decisions, such as about the severity of the shock and the extent to which the agreement did or

did not withstand the external pressures.[12] Third, even the theoretical debate uses the concepts of reliability and durability interchangeably, suggesting that both concepts are viewed as substitutes (Martin, 2005, 448; Yoo, 2011, 41; Hathaway, 2008, 1316).

If executive agreements are more reliable commitments than treaties, e.g. because treaties can easily be withdrawn from by the president, then it should be the case that a promise concluded as an executive agreement is more durable than a promise concluded as a treaty. If, on the other hand, treaties are the more reliable instrument because of the high legislative costs that only truly committed negotiators would incur, then the average treaty should outlast the average executive agreement.

3.3 Data Description and Methodology

3.1 The Data

The data set consists of all agreements that have been reported in the *Treaties in Force* (TIF) series that were signed and ratified between 1982 and 2012. TIF is the official collection of international agreements in force maintained by the U.S. Department of State. It includes information on the signing date, the parties, the subject area of the agreement as well as on when the agreement went into force. The agreements in TIF appear in the *Kavass' Guide of Treaties in Force* ("The Guide"). The Guide is an annual publication accompanying TIF. It was first published in 1982 and contains further information useful for researching treaties, such as the treaty subject matter, a short description as well as the parties to the agreement. TIF uses an elaborate but partially incoherent system to categorize agreements by subject area. In total, there are 197 different subjects in the data set, many with single-digit observations. I reduce the dimension of these subject areas into 38 thematically coherent categories. The grouping is detailed in the Appendix.

Of primary relevance to this analysis is the fact that the Guide contains a list of treaties which were indexed in TIF in the year preceding the year of publication, but are not indexed in the publication year's TIF any longer. Based on the Guide, it is thus possible to determine which agreement has been deleted from the TIF publication and in which year the deletion took place. An agreement that was listed in TIF in the previous year but is not listed

[12] For one attempt at codifying the propensity for shocks to occur by issue area, as well as for a discussion of the downsides of this approach, *see* Koremenos (2005).

in the current year is considered to be no longer in force by the U.S. State Department.

For each agreement, the Guide further reports a "Senate Treaty Document Number". This number is assigned to any treaty submitted to the Senate under the Advice and Consent procedure. Regular executive agreements do not receive a Senate Treaty Document Number. The number can thus be used to identify which agreement in the data base is a treaty and which agreement was concluded as an executive agreement.

At this point, it is important to address a possible limitation of this data set. While the TIF is the most comprehensive collection of international agreements to date, there is no data set listing without omission all international agreements the United States has concluded in the past. Researchers could try and complement TIF with other treaty collections in hopes to create a more comprehensive list of agreements. However, this is neither advisable nor practical for several reasons.

First, and most importantly, the only known bias in TIF is the omission of secret agreements which, if publicized, could threaten national security.[13] However, since these secret agreements are not publicly known by definition, it is likely that they are also missing in other databases. Second, the agreements in TIF all follow one comprehensible selection process: They are agreements submitted to Congress pursuant to the Case Act and are considered to be in force by the State Department. Combining these agreements with other databases introduces the possibility for unknown selection biases, threatening the interpretability of any findings. Third, TIF uses its own index system, such that agreements in TIF cannot easily be compared to those from other sources. And fourth, previous attempts to combine data sets have resulted not in more, but substantially fever agreements than contained in the data set used here.[14] For these reasons, it is suggested here that a single data set based on TIF is preferable to a combination of different sources. While being conscious that any results cannot be extrapolated to secret agreements without making further assumptions, there are no known biases introduced by cabining the data in this way.

The data set on international agreements was further complemented with

[13]The Case Act provides that these agreements only need to be transmitted "to the Committee on Foreign Relations of the Senate and the Committee on International Relations of the House of Representatives under an appropriate injunction of secrecy to be removed only upon due notice from the President", 1 U.S.C. 112b(a) (1979).

[14]Hathaway (2008) combines multiple sources, leading to a total number of 3,119 agreements in the period of 1980-2000. In contrast, the data set used here contains 6,148 agreements in the same period.

publicly available information on the president under which an agreement was signed, Senate compositions by party, as well as "legislative potential for policy change" (LPPC) scores for the Senate as used in Martin (2005). LPPC scores reflect how difficult it is for a president to push legislation through. A higher LPPC score indicates lower political costs to implement legislation. The LPPC score is constructed according to the formula:

$$LPPC = Seats_{President} * Unity_{President} - Seats_{Opposition} * Unity_{Opposition}$$

Here, *Unity* refers to voting unity scores published by *Congressional Quarterly*. Higher unity scores indicate more uniform voting patterns.

Overall, the data set contains 7,966 agreements. In longitudinal form, each agreement is observed once per year while it is in force and once when it goes out of force, leading to a total of 129,518 per-year-per-agreement observations.

3.2 Methodology

With each observation in the data set being an agreement-year, it is now of interest how the durability between different types of agreements vary, holding other characteristics constant. Differences in durability, or survival times, can be estimated using survival time analysis. In the social sciences, these methods are also referred to as event history studies (Box-Steffensmeier and Jones, 2004). It is helpful to define a few key terms in order to prevent confusion. Survival time analysis is primarily used in the medical sciences and as such, the terminology is characterized by terms encountered most often in clinical trials. A "subject" is a unit of observation, here an agreement. An "event", "death" or "failure" are synonyms for the occurrence of the incident of interest, here the going-out-of-force of an agreement. The "survival time" is the time period between the start of the observation and the occurrence of the incident, here the period in which an agreement is in force. Agreements that are mentioned in the last period of observation are considered "right-censored", i.e. with a survival time that has a known lower bound and an unknown upper bound.[15] Finally, a "hazard rate" is the probability for an event to occur.

Survival time analysis offers different models to estimate the longevity of an observed subject, each with their individual advantages and disadvantages. The model choice is primarily governed by whether the survival times of the analyzed subjects are continuous or discrete and how they are observed.

[15]The agreement is in force at least until 2014, possibly longer.

To begin the discussion, note that international agreements can go out of force at any point in time and that survival times are thus continuous in nature. However, as described above, survival times are measured only once per year through the publication of TIF. Hence the data can best be described as continuous data that is grouped by year. For truly continuous data in which an event can happen at any point in time, the Cox proportional hazard model (Cox, 1972) has established itself as the preferred choice by researchers, as it is a semi-parametric model that only relies on few assumptions (Lin et al., 1993; Tian et al., 2005).[16] The Cox model is of the form

$$h_i(t|x_i) = h_0(t)e^{x'_i\beta}$$

where i is the individual agreement, t is a period in time, x denotes a set of covariates and h denotes the hazard rate, i.e. the probability for an event to occur. The popularity of this model stems from the fact that it can be estimated without making any parametric assumptions about the baseline hazard rate, $h_0(t)$. However, the Cox model assumes that there are no ties in the data, meaning that no two observations have the exact same survival time. This is due to the fact that ties cannot occur if survival times are measured on a truly continuous scale. Researchers have developed several techniques to deal with ties. The most precise approach is the "exact method" developed by Kalbfleisch and Prentice (1973). Intuitively, if two subjects i and k survive exactly n periods, the exact method considers the alternative that i survived longer than k and the alternative that k survived longer than i and opts for the one that maximizes the associated likelihood function. However, in data sets with many subjects, periods and ties, the exact method is not feasible as computationally it is very intensive. The "Efron method" (Efron, 1977) provides an approximation to the exact method that does not suffer from comparable resource constraints but is less precise.

An alternative to the Cox model is a parametric survival model. Among the parametric models, the complementary log-log discrete model is the uniquely appropriate model for grouped continuous data (Kalbfleisch and Prentice, 2002).[17] It is of the form

$$h_i(t|x_i) = 1 - (1 - h_0(t_i))^{e^{x'_i\beta}}$$

[16]For examples specifically in international relations, *see* Elkins et al. (2006); Simmons (2000).

[17]The statement refers to the continuous-time proportional-hazards model, where observations have been grouped by time. McCullagh (1980) shows that this model is identical to the complementary log-log discrete model.

or, if linearized,

$$log(-log(1 - h(t))) = \alpha_j + x_i'\beta$$

where j denotes grouped time intervals. Note that $\alpha_j = log(-log(1 - h_0(t_j)))$ is an interval-specific complementary log-log transformation of the baseline hazard rate, $h_0(t_j)$. This means that the baseline hazard rate is allowed to vary with each interval, thus imposing only mild parametric assumptions.

Whether to prefer the Cox model in combination with an Efron approximation over the complementary log-log discrete model cannot be answered in a general way. Simulations show that even with heavily tied data sets, the Efron approximation often achieves very accurate results (Hertz-Picciotto and Rockhill, 1997; Chalita et al., 2002; Borucka, 2014). As a rule of thumb, Chalita et al. (2002) propose to compute the quantity

$$pt = \frac{nf - r}{n}$$

where nf is the number of events (here, agreements that went out of force), r is the number of unique survival times and n is the number of agreements. For $0 \leq pt < 0.2$, Chalita et al. suggest a continuous model with likelihood approximation; for $0.2 \leq pt \leq 0.25$, both discrete and continuous models can be used; for $pt > 0.25$, a discrete model is preferred. Here, $pt = 0.19$, which is why a Cox proportional hazard model with Efron approximation is used in the primary model specifications. The complementary log-log model serves as a robustness check.

Both the Cox and the complementary log-log model rely on the assumption that the hazard is proportional to the baseline hazard ratio. This assumption can be tested using the Grambsch and Therneau (1994) method which plots the Schoenfeld residuals against the rank of the time intervals. If the proportionality assumption holds, then there should be no systematic pattern. In a formal test of non-proportionality, 31 of 264 (or 11%) covariates yield significant p-values implying a violation of non-proportionality. Reassuringly, the covariates of interest are not among them. However, even for the remaining covariates, the disproportionality is of little concern for two reasons.

First, note that the Grambsch and Therneau test was developed in the medical context where sample sizes are typically smaller than 100, making the test insensitive to minor disproportionalities. For sample sizes as large as in this study, small confidence intervals lead to significant p-values even

if the data reveals negligible disproportionalities. In addition to the formal tests, visual examination of the Schoenfeld residuals is thus recommended (Vittinghoff et al., 2012). Such a visual examination yields no significant violations of the proportional hazards assumption for any of the subject matter covariates, and a violation only for a handful of countries, typically those with whom the U.S. has only few agreements, such as Burma, Ecuador or New Caledonia. The corresponding graphs are included in the Appendix.

Second, note that the concern for a violation of the proportional hazard assumption stems from the medical sciences, where it is of great importance whether a drug has an inverse or possibly a reverse effect on a subset of patients. However, in the social sciences, researchers are typically interested in average covariate effects across the entire sample. As Allison (2010) highlights, even in cases where the proportionality assumption is violated, estimates can still be interpreted as average covariate effects. Violations of the proportionality assumption thus do not present a threat to the interpretability of the coefficients for most social scientific studies such as the present one.

3.3 Summary Statistics

Table 4.1 reports summary statistics. As can be seen, 5% of all agreements between 1982 and 2012 were concluded in the form of a treaty, making the use of the treaty an exception. 20% of the agreements went out of force. The average agreement was observed to be in force for 15.26 years. Among the agreements that are no longer in force, the average durability is 7.3 years. LPPC scores range from -17 to 17 with an average of -0.13. On average, 50% of the seats in the Senate were held by the president's party at the time the agreement was signed. For 71% of agreements, the Government was divided, with the White House being held by one party and either the Senate, the House or both being held by the other. Together, these numbers indicate that the average agreement could not have been passed in the form of a treaty absent a bipartisan effort, making the treaty a potentially costly instrument. 6% of agreements are multilateral while 1% is concluded with an international organization.

Figures 3.1 and 3.2 depict histograms indicating the number of executive agreements and treaties split by year and by the signing president. What can be seen is that the total number of agreements peaked in 1985 and declined since then. The relative share of treaties among all agreements was greatest in 2010, with 28% of agreements being concluded in the form of a treaty. However, most of these treaties were signed prior to the Obama presidency. Indeed, President Obama has concluded fewer agreements as treaties than any

other president during the period of observation, a finding that has previously been observed by other scholars (Peake, 2015). Meanwhile, agreements signed under President Clinton include the highest share of treaties with 7.6%. Together, this implies that the use of a treaty varies with the President, though executive agreements are by far the more prevalent instrument throughout.

Table 3.2 depicts a list of selected subject areas and the prevalence of treaties and executive agreements in them. The only subject area in which treaties are more prevalent than executive agreements is extradition, where 94% of agreements are concluded as treaties. A likely explanation for this phenomenon is the legal uncertainty surrounding the use of executive agreements to surrender individuals to foreign nations. Whether an individual can be extradited pursuant to a congressional-executive agreement was specifically considered in Ntakirutimana v. Reno,[18] a decision by the 5th Circuit from 1999. Elizaphan Ntakirutimana was to be extradited to be tried before the International Criminal Tribunal for Rwanda pursuant to an executive agreement between the U.S. and the tribunal. The majority opinion held that the extradition pursuant to an executive agreement is constitutional, relying heavily on *Valentine v. U.S.*[19] *Valentine*, a case from 1936, is the most recent Supreme Court decision that arguably could be construed to speak to the question of the constitutionality of congressional authorizations of extraditions. Here, the Supreme Court held that extraditions need to be authorized "by act of Congress or by the terms of a treaty." While today's reading of the ruling might suggest that this is an explicit authorization of congressional-executive agreement, Judge DeMoss, in a minority opinion of *Ntakirutimana*, notes that the court in Valentine dealt with a scenario of extradition under a treaty. The mentioning of Congressional authorization may thus have been "pure dicta".

Like the 5th Circuit in *Ntakirutimana*, academics are split on the question of whether extraditions can be authorized by executive agreement, with some emphasizing a lack of congressional authorization (Yoo, 2011; Hathaway, 2008) while others interpreting *Valentine* as an explicit authorization by the Supreme Court. In an environment of such legal uncertainty, it might be reasonable for the president to rely on the treaty to guarantee the enforceability of the agreement.

Other areas in which treaties are very prevalent encompass 'judicial assistance', which includes agreements to prosecute cross-border crime such as drug trafficking or money laundering, but also stolen passports; 'taxation', which

[18]Ntakirutimana v. Reno, 184 F.3d 419, 437 (5th Cir. 1999).
[19]Valentine v. U.S. ex rel. Neidecker, 299 U.S. 5 (1936).

primarily includes double taxation and taxation information agreements; and 'property', including agreements on the return of stolen vehicles and the transfer of real estate. Considering only subject areas, it seems difficult to explain the use of the treaty along one coherent dimension. For instance, if we think that treaties are especially prevalent among important agreements, we might expect them to be used frequently in agreements relating to national security and defense. However, only 1% of defense agreements are concluded in the form of a treaty. Meanwhile crime prevention, which is often thought of as having a lower priority than national security, includes a much larger share of treaties.

The data also shows that the narrative that treaty use is the result of historical convention at least leaves many subject areas unexplained. For instance, whereas it was previously argued that path-dependence would have lead to treaties being particularly common in human rights law and absent in trade, Table 3.2 shows that neither subject area presents a particularly striking outlier that would make an interesting test of the theory. While in the area of human rights, treaties are somewhat prevalent with 17%, treaty use in this area is still the rare exception rather than a norm. Similarly, the use of treaties in economic areas such as trade, commerce and finance is close to the average of 5%, raising questions as to whether the rarity of treaties in these areas really is best explained by historical shocks or whether it is just a reflection of a differently motivated aversion to the treaty that affects other subject areas as well.

Overall, it seems difficult to explain the wide variety of treaty prevalence in the different subject areas using conventional theories. A full list of agreement use by subject area is included in the Appendix.

The agreements in the data set have been concluded between the U.S. and one or more of 215 countries and 52 international organizations. Table 3.3 depicts the 20 countries with the most agreements in the data set. Interestingly, the three most frequent users of treaties are all Western European countries, namely France, Italy and Germany. In agreements that are multilateral, 20% are concluded in the form of a treaty, far exceeding the share in any bilateral relationship.

3.4 Results

Table 3.4 presents results for the cox proportional hazard model. Model (1) only includes the treaty indicator. Model (2) includes president and sub-

ject area fixed effects. Model (3) additionally includes country fixed effects.[20] Model (4) further controls for the president's share of seats in the Senate, as well as for a divided government. Model (5) does not control for the share of seats, but for LPPC scores, which are arguably a better proxy for the costs of pushing legislation through the Senate. If the instrument use was merely a function of the seat map in the Senate, then the inclusion of either of these covariates should render the coefficient on $Treaty$ insignificant. The standard errors for all models are clustered by agreement.

What can be seen is that in each model specification, the coefficient on the treaty indicator is negative and significantly different from 0. Note that coefficients in survival models express changes in the probability for an event to occur. Here, the event is defined as an agreement going out of force. Hence a negative coefficient indicates a decrease in the probability for an agreement to go out of force if it is concluded in the form of a treaty. The results imply that treaties last significantly longer than executive agreements and that the difference in durability is neither the result of arbitrary subject-matter conventions, nor a by-product of a decision-making process that is primarily driven by the seat map in the Senate.

Table 3.5 runs the same model specifications using a complementary loglog model. Again, the results consistently show that agreements concluded as treaties outlast those concluded as executive agreements.

Having found that treaties outlast executive agreements, consider now the plausibility of the mechanism proposed by Martin (2005) that differences in signaling costs lead to differences in reliability. The costs of the signal are determined by how difficult it is for the president to secure the required votes in the Senate. Hence in a setting where the senatorial support for the president is low, the treaty should send an especially strong signal of commitment. Meanwhile, if the president has a lot of support in the Senate, the differences in costs between executive agreements and treaties are lower and the use of the treaty sends less of a strong signal, which should lead to smaller differences in agreement durability.

Table 3.6 analyzes the validity of this mechanism. Model "LPPC Low" includes only agreements that have been concluded when the LPPC scores for the president were less than 0, such that the use of the treaty is especially costly. In contrast, Model "LPPC High" includes only agreements in which LPPC scores are greater than 0 and the use of the treaty is less costly. Consistent with the signaling mechanism suggested by Martin, the difference

[20] Due to data sparsity, not all country fixed effects can be accurately estimated, which is why this specification is included separately.

between treaties and executive agreements is more pronounced when the use of the treaty is costly, where the coefficient on $Treaty$ is -1.315. In contrast, when LPPC scores are high, the coefficient is -1.171. However, it is cautioned here that this evidence is merely preliminary. In particular, the costs of obtaining a two-thirds majority were high throughout the period of analysis, with the majority of agreements concluded while the president's party held 45 to 55 seats in the Senate. No agreement could be concluded as a treaty without bipartisan support. Hence, there is only limited variance to test cost differentials convincingly.

Statistical significance does not imply substantive relevance and with a large number of observations such as in this study, it is important to complement the statistical findings with evidence for substantive significance of the results. Differences in survival times are best illustrated by comparing estimated survival curves or cumulative hazard curves. A survival curve at time t depicts the probability that a subject survives in t, conditional on having survived up until t. The cumulative hazard in time t is the probability that an event occurs in or prior to t. Figure 3.3 depicts estimated survival and cumulative hazard curves for the preferred Model (5), one corresponding to a treaty and one corresponding to an executive agreement. Numerical covariates have been centered around their mean. For categorical variables, the most prevalent value is used. The survival curves can thus be thought of as corresponding to a "typical" agreement.[21] What can be seen is that for the typical agreement, there is a probability of 0.14 that it breaks down at the end of the period of observation, conditional on having held until then. For executive agreements, that probability is 0.4, more than twice as high. Similarly, there is a 0.15 probability that a treaty breaks down within the window of observation, whereas that probability is 0.5 for executive agreements.

Recall that there are two different types of executive agreements, namely congressional executive agreements and sole executive agreements. So far, the analysis has not distinguished between different types of executive agreements, even though it can be argued that the differentiation is essential. After all, the question of substitutability is only raised with regards to differences between congressional executive agreements and treaties, while it is generally acknowledged that sole executive agreements are very different policy instruments that fall entirely into the president's power and do not require legislative participation. TIF does not distinguish between sole and congressional executive agreements and indeed, to distinguish between the two would require the painstaking effort of searching for authorizing legislation regarding each exec-

[21]The country is Mexico, the president is Reagan and the subject is Defense.

utive agreement in the Statute at Large, a process that cannot be automated easily (Hathaway, 2008). Prior studies have found that the proportion of sole executive agreements is minimal, with an estimated share between 5 and 6% of all agreements (McLaughlin, 1958).[22] To accommodate that some agreements might be sole executive agreements, this study takes the following approach:

It sorts agreements by their durability and assumes that the x quantile are sole executive agreements, where $x \in [0, 0.1]$.[23] It then omits these agreements from the analysis, runs the preferred Model (5) and collects the estimated coefficient on the treaty indicator and its standard error. Note that the assumption that the least durable agreements are sole executive agreements is extremely restrictive. In reality, it is much more likely that some sole executive agreements outlast congressional executive agreements. It can thus be expected that this approach biases the survivability of congressional executive agreements upwards, making it harder to detect a difference between the durability of treaties and executive agreements. If it can be shown that even under these restrictive assumptions, treaties survive executive agreements, this can be regarded as particularly strong evidence for the longer durability of treaties.

Figure 3.4 reports the estimated coefficients and 95% confidence intervals for all x over the range $[0, 0.1]$. What can be seen is that even under the strictest assumption that the 10% shortest-lasting executive agreements are sole executive agreements, there still is a substantial difference between treaties and congressional executive agreements that is statistically different from 0.

3.5 Discussion

This study is motivated by the question of whether the treaty serves a purpose as a modern policy tool. The analysis suggests that it does. There is a statistical and substantive difference between the durability of treaties and executive agreements. Throughout all model specifications, treaties are estimated to have substantially longer survival times. This finding holds, even when it is assumed that the 0.1 quantile of the executive agreements with the shortest survival time are sole executive agreements to which interchangeability does not apply. Together, the findings provide strong evidence that treaties outlast executive agreements. While it cannot be ruled out that differences in survival times are also influenced by presidential preferences, the

[22] See also Cong. Research Serv., 95TH Cong., International Agreements: An Analysis of Executive Regulations and Practices.

[23] For instance, $x = 0.05$ assumes that the 5% least durable agreements are sole executive agreements.

Congressional seat map and historical path-dependence, it was demonstrated that even after controlling for all these characteristics, agreements concluded as treaties last longer than those concluded in the form of an executive agreements. The results imply that the treaty is a more reliable commitment device than an executive agreement. Being able to signal different commitment levels can lead to separating equilibria in which only those with a stronger intent to perform rely on the treaty, whereas others rely on an executive agreement. Abolishing the treaty would lock negotiators out of the possibility to signal the seriousness of their promise, effectively turning separating equilibria into pooling equilibria in which all presidents use the same instrument.

It is worth noting that finding the treaty to serve a different purpose than the executive agreement does not necessarily imply that having two signaling devices is normatively desirable. Indeed, there might be reasons to argue for a reduction of international commitment devices, that, to my knowledge, have so far evaded the attention of international legal scholars.[24] These reasons originate from the economic literature on signaling, most importantly the seminal work of Spence (1973) of signaling in the job market. Spence shows that, under certain conditions, the possibility to signal ones' commitment level can lead to separating equilibria that are pareto-inferior to the pooling equilibria when signaling is impossible. This result is best demonstrated formally, but to nonetheless provide some intuition, compare an Arrow–Debreu (Arrow and Debreu, 1954) world of perfect information to a world with imperfect information with and without signaling devices. In the Arrow–Debreu world, every mutually beneficial contract will be concluded and every unbeneficial contract will not, providing a benchmark for optimality. In a world with imperfect information and without signaling, some international agreements will not be concluded because of uncertainties about the president's level of commitment. In particular, agreements that require a high level of commitment of the president may be foregone due to concerns that the president may not be dedicated enough. The availability of different signaling instruments reveals information about the president's commitment level, thus moving us closer to the Arrow–Debreu benchmark. At a first glance, it might then be suggested that the separating equilibria achieved under signaling are superior to the pooling equilibria if signaling is not possible. However, note that signaling comes at a cost. In particular, presidents that do not want to be perceived as having a low level of commitment need to incur the higher costs of the treaty instru-

[24] To be sure, this is a known result in the literature on contracts, *see* Aghion and Hermalin (1990).

ment.[25] In contrast, if signaling is not possible, there are no signaling costs and presidents are always perceived as having an average level of commitment. It is then possible for the signaling costs to outweigh the benefits achieved from the additional contracts that are concluded under signaling, hence leading to a loss in overall welfare.

Whether the availability of signaling devices has welfare enhancing effects depends to a large extent on the costs of the signal, as well as the distribution of potential agreements. For instance, if most international agreements that the United States could potentially conclude promise to yield very high payoffs, compared to the costs of concluding an agreement as a treaty, then signaling is more likely to yield overall welfare gains. However, if this is not the case, the availability of signaling devices can reduce mutual gains. It would go beyond the scope of this study to formally analyze and discuss these conditions, to predict whether the existence of two parallel signaling devices as an anomaly of the United States should ultimately be preserved and if not, which instrument should be abolished. The remarks on Spence's signaling model are merely intended to highlight a fallacy in the current scholarship that seems to equate the positive question of whether treaties and executive agreements are distinct with the normative question of whether they are desirable.

With regards to the mechanism that is responsible for treaties outlasting executive agreements, Martin (2005) suggests that differences in reliability are the consequence of increased political costs imposed by the required two-thirds majority. The evidence is at least consistent with this mechanism, as the difference between executive agreements and treaties is especially pronounced when the president has low senatorial support, making the conclusion of an agreement as a treaty particularly costly. Nonetheless, it is important to note that high political costs are not necessarily the exclusive driver of the results. Indeed, even under the assumption that treaties and executive agreements are associated with identical (or no) costs and produce the same information, repeated interaction can result in outcomes in which only those who intend to comply over the long term rely on the treaty instrument. That is because reputational concerns in repeated interactions can turn what would otherwise be considered as "cheap talk" into credible commitments (Kim, 1996). Once a president starts using treaties for agreements intended to last for a long time and preserves the executive agreements for short-term agreements, negotiation partners will form the expectation that this pattern persists in the future. The president then has an incentive to act consistent with these expectations, in order to be able to indicate his level of commitment in future interactions

[25] Whether these are reputational costs or higher costs of concluding the agreement.

when it matters. In other words, even if it was purely out of convention or by chance that the treaty established itself as the more serious commitment, both promissor and promissee can benefit from the possibility to be able to signal differing levels of commitment, providing incentives to preserve differentiating signaling mechanisms, even if there is no difference in the underlying costs (Guzman, 2008a).

In addition to speaking to the narrower question of the treaty's purpose, this study also contributes to a broader strand of literature analyzing choices in the face of different political instruments. For example, scholars have raised questions as to why many appointments of the president follow the Advice and Consent procedure in the Senate if there is the possibility to appoint nominees unilaterally through recess appointments. Similar to much of the political science literature on the treaty, it has been argued that the choice is a determined by the seat map in the Senate(Corley, 2006; Carrier, 1994). This study suggests that a different line of inquiry may lead to fruitful discoveries as well. In particular, it may be the case that the president's inclination to appoint by means of Advice and Consent presents a particularly high level of commitment towards the candidate, in turn increasing the appointee's perceived legitimacy and making her more likely to endure political turmoil or criticism. A similar rationale focused on differences in signaling costs may further help explain the presidential choice between executive orders and statute (Mayer, 1999), as well as motivations for abstaining from amending the meaning of statutory provisions through signing statements (Cooper, 2005). Both executive orders and signing statements can be characterized as policy tools the president can use unilaterally at a relatively low cost to circumvent the more costly process of enacting policy preferences through formal legislation. The results of this study may help explain the constraints under which these unilateral tools can be used, as well as their potential disadvantages in the form of low-cost signaling.[26]

3.6 Conclusion

Relying on survival time analysis, this inquiry revealed that treaties are more durable commitments than executive agreements. There was a 0.15 probability that a typical agreement concluded as a treaty in 1982 broke down by 2012, compared to a 0.5 probability that it broke down when concluded as an executive agreement. In contrast to recent arguments advanced by both

[26]For a thorough formal and empirical treatment of unilateral presidential powers, *see* Howell (2015).

legal scholars and political scientists, treaties are neither solely a reflection of the seat map in the Senate, nor is their use merely a result of historical path-dependence. Instead, the results of this chapter imply that treaties are qualitatively different instruments than executive agreements that, on average, signal a more serious commitment related to the terms of the underlying agreement. Abolishing the treaty would make it difficult for presidents to signal their intended level of commitment, in turn impacting the kinds of international agreements other states are willing to conclude with the U.S.

3.7 Figures and Tables

Table 3.1: Summary Statistics

	Mean	SD	Min	Max	Median	IQR
In Force Year	1996	8.59	1982	2012	1995	15
Treaty	0.05	0.22	0	1	0	0
Event	0.20	0.40	0	1	0	0
Out of Force Year*	1998	7.63	1983	2012	2000	11
Durability	15.26	9.03	0	32	15	16
Durability*	7.30	5.68	0	30	6	8
LPPC	-0.13	9.46	-17	17	0	17
Share Senate	0.50	0.04	0	1	0.50	0
Divided Government	0.71	0.45	0	1	1	1
Multilateral	0.06	0.24	0	1	0	0
Intl Organization	0.01	0.12	0	1	0	0

Summary Statistics for the variables used in this data set. An asterisk indicates that the statistics only include treaties that have gone out of force in the period of observation.

Table 3.2: Agreement Use by Subject Area

Subject	# EAs	# Treaties	Mean Treaty
Agriculture	454	1	0.00
Education	64	0	0.00
Postal Matters	239	0	0.00
Defense	1433	9	0.01
Other	138	2	0.01
Labor	131	3	0.02
Finance	500	22	0.04
Trade and Commerce	748	35	0.04
US Boundaries	52	4	0.07
IP	23	2	0.08
Environment	196	20	0.09
Fisheries	83	9	0.10
Human and Fundamental Rights	15	3	0.17
Property	8	5	0.38
Taxation	103	75	0.42
Judicial Assistance	93	80	0.46
Extradition	5	75	0.94

The table depicts the prevalence of treaties and executive agreements for selected subject areas. Statistics for all subjects are included in the Appendix.

Table 3.3: Agreement Use by Partner Country

Country	# EAs	# Treaties	Mean Treaty
Mexico	247	6	0.02
Japan	250	2	0.01
Russia	219	4	0.02
United Kingdom	195	10	0.05
Canada	190	10	0.05
Egypt	188	2	0.01
South Korea	139	2	0.01
Germany	116	7	0.06
Philippines	116	2	0.02
France	106	10	0.09
Australia	102	4	0.04
China, Republic	104	1	0.01
Indonesia	100	2	0.02
Israel	97	3	0.03
Brazil	98	1	0.01
Ukraine	92	4	0.04
Pakistan	95	0	0.00
Peru	92	1	0.01
Italy	82	6	0.07
Jordan	85	2	0.02

The table depicts the prevalence of treaties and executive agreements for the 20 most frequent partner countries in the data set. Statistics for all countries are included in the Appendix.

Table 3.4: Cox Proportional Hazard Model

	Dependent Variable:				
	Survival Time				
	(1)	(2)	(3)	(4)	(5)
Treaty	-1.324^{***}	-1.313^{***}	-1.164^{***}	-1.176^{***}	-1.180^{***}
	(0.237)	(0.254)	(0.262)	(0.262)	(0.262)
Divided				−0.099	−0.086
				(0.135)	(0.129)
Senate Share				1.423	
				(0.949)	
LPPC					0.008
					(0.004)
President FEs		✓	✓	✓	✓
Subject FEs		✓	✓	✓	✓
Country FEs			✓	✓	✓
Observations	129,518	129,518	129,518	129,518	129,518
Log Likelihood	−12,790	−12,143	−11,905	−11,901	−11,900
Wald Test	31^{***}	$23{,}684^{***}$	$149{,}155^{***}$	$146{,}607^{***}$	$146{,}560^{***}$

Note: *p<0.05; **p<0.01; ***p<0.001

The results of a cox proportional hazard regression of survival time on a treaty indicator and several covariates. Standard errors are clustered by agreement.

Table 3.5: Complementary Log-Log Model

	Dependent Variable:				
	Event Occurrence				
	(1)	(2)	(3)	(4)	(5)
Treaty	-1.324^{***}	-1.314^{***}	-1.164^{***}	-1.176^{***}	-1.180^{***}
	(0.238)	(0.270)	(0.283)	(0.283)	(0.283)
Divided				−0.099	−0.086
				(0.128)	(0.123)
Senate Share				1.426	
				(0.939)	
LPPC					0.008
					(0.004)
Constant	−20.473	−22.219	-37.915^{*}	−38.682	−37.993
	(26.011)	(18.997)	(25.876)	(19.619)	(32.866)
President FEs		✓	✓	✓	✓
Subject FEs		✓	✓	✓	✓
Country FEs			✓	✓	✓
Interval FEs	✓	✓	✓	✓	✓
Observations	129,518	129,518	129,518	129,518	129,518
Log Likelihood	−7,708	−7,061	−6,822	−6,818.152	−6,818
Akaike Inf. Crit.	15,484	14,269	14,224	14,220	14,219

Note: *p<0.05; **p<0.01; ***p<0.001

The results of a generalized linear model with a complementary log-log link function regressing survival time on a treaty indicator and several covariates. Standard errors are clustered by agreement.

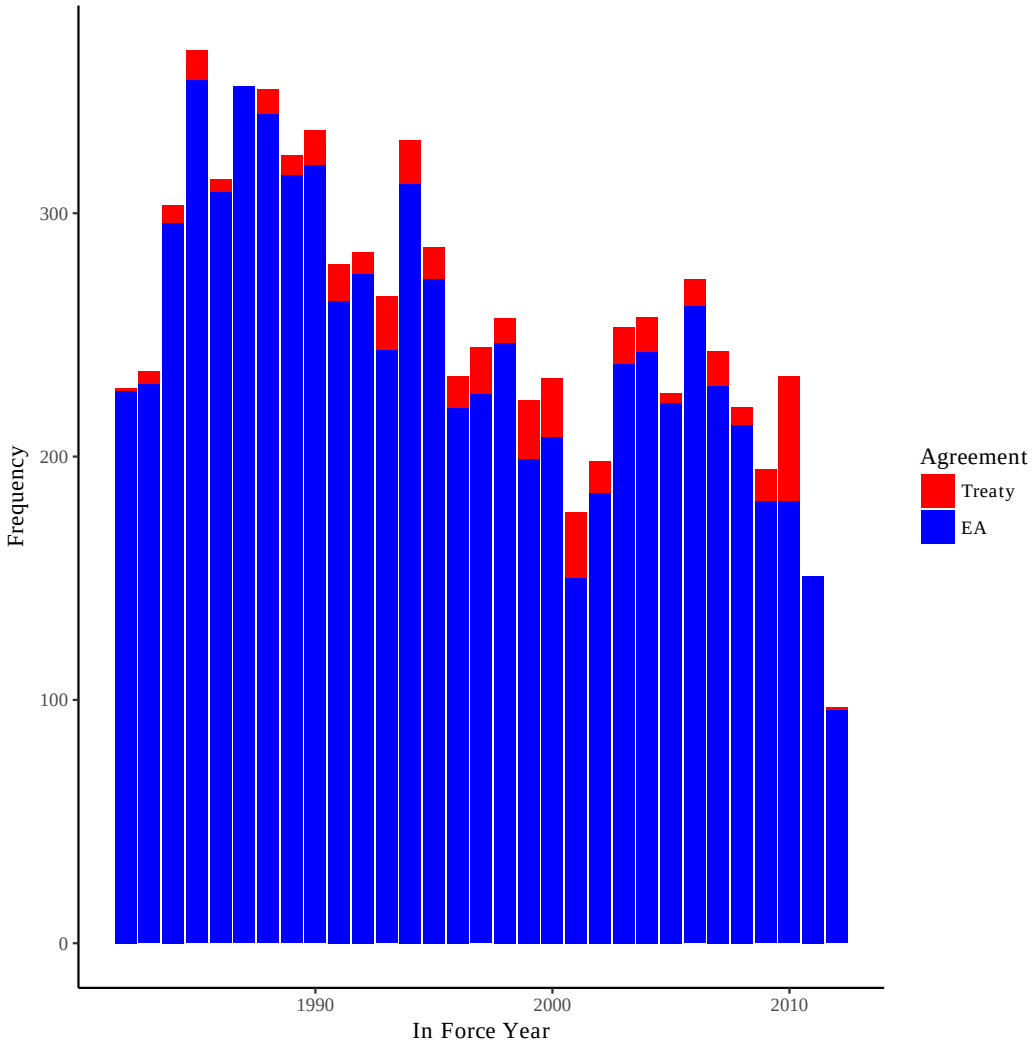
Table 3.6: Cox Model by Senatorial Support

	Dependent Variable:	
	Survival Time	
	LPPC High	LPPC Low
Treaty	-1.170^{**}	-1.315^{**}
	(0.366)	(0.393)
President FEs	✓	✓
Subject FEs	✓	✓
Country FEs	✓	✓
Observations	59,619	67,972
Log Likelihood	−5,316	−5,400
Wald Test	$70{,}308^{***}$	$124{,}366^{***}$

Note: *p<0.05; **p<0.01; ***p<0.001

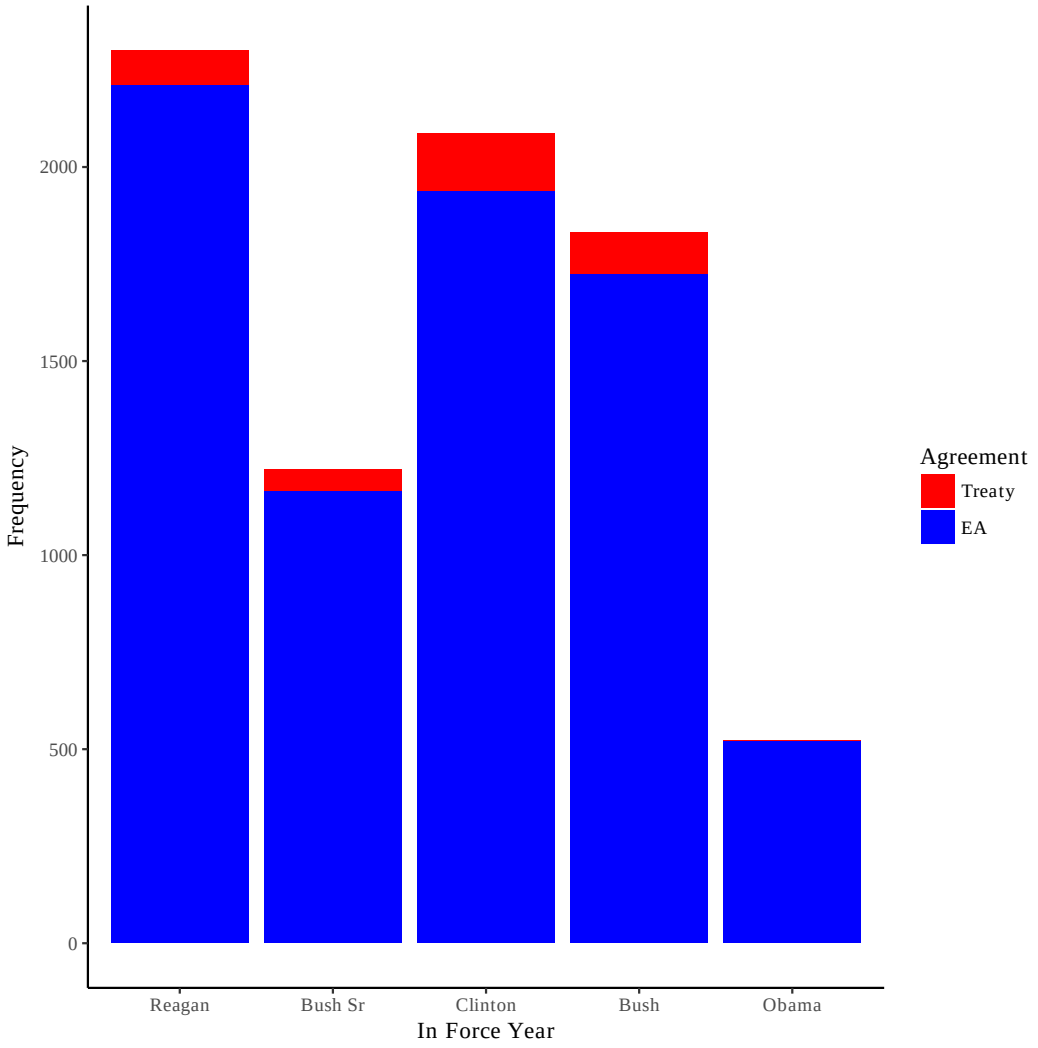
The results of a cox proportional hazard regression of survival time on a treaty indicator and several covariates. Model LPPC High includes only agreements concluded when LPPC scores were greater than 0. LPPC Low includes only agreements concluded when LPPC scores were less than 0. Standard errors clustered by agreement.

Figure 3.1: Agreement Types over Time



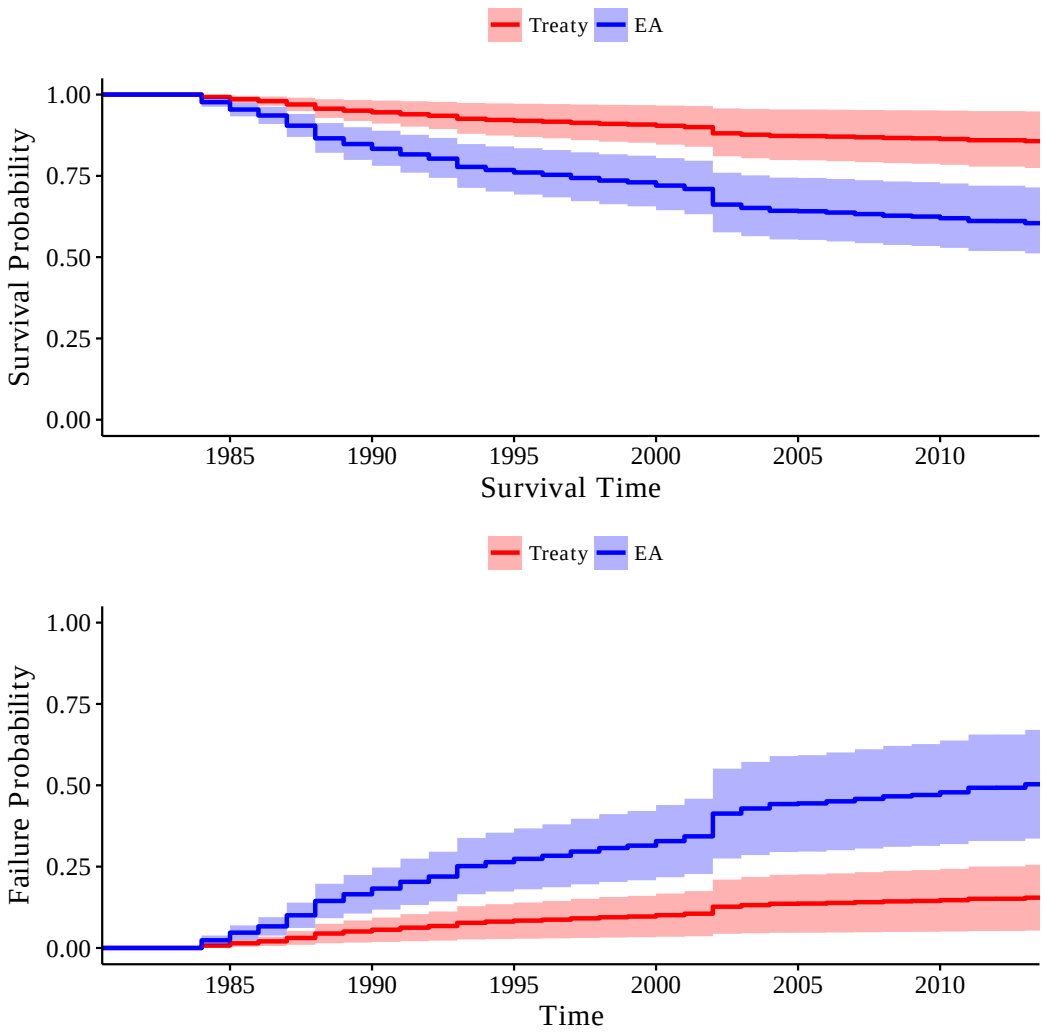
This graph depicts the use of executive agreements and treaties over time.

Figure 3.2: Governing Law Clause Usage over Time



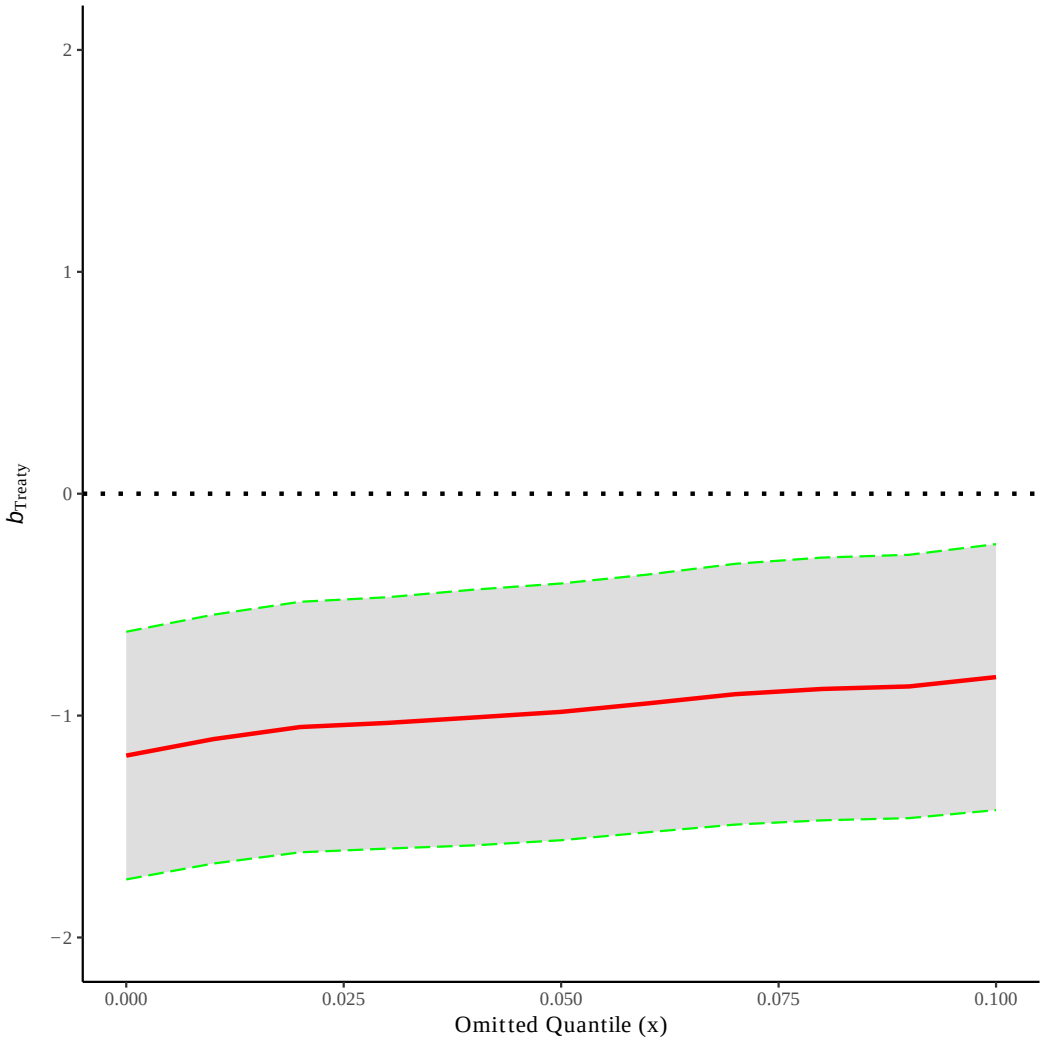
This graph depicts the use of executive agreements and treaties by the different presidents.

Figure 3.3: Governing Law Clause Usage over Time



This graph depicts estimated survival curves (top) and estimates hazard curves (bottom) for treaties and agreements over the period of observation. Shaded areas are 95% confidence intervals.

Figure 3.4: Omitting Sole Executive Agreements



This graph depicts the coefficient on the treaty indicator of Model (3) under the assumption that the x quantile of agreements are sole executive agreements and should thus be omitted from the analysis.

Chapter 4

Conforming Against Expectations: The Formalism of Non-Lawyers at the WTO

co-authored with Jerome Hsiang

4.1 Introduction

The dispute settlement system under the General Agreement on Tariffs and Trade (GATT) was notoriously dysfunctional. Commentators blamed much of its failures on the prevalence of dispute settlement panelists who have never received legal training. Many of these non-lawyer panelists were diplomats, and conventional wisdom held that their lack of legal training mired them in consensus-based negotiations, remaining too deferential to myopic state interests and "the ethos of diplomats" (Weiler, 2002). As a result, the argument went, the GATT dispute settlement system rendered *ad hoc* decisions and gave little thought to jurisprudential coherence or precedents (Hudec, 1970; Shaffer et al., 2016). To fix this quagmire, the World Trade Organization (WTO) Dispute Settlement Body (DSB) was designed specifically to encourage legalization and a rule by lawyers (Pauwelyn, 2005).

Even today, nearly two and half decades after the WTO came online, commentators have not forgotten the GATT's experience with non-lawyer panelists. Many continue to view them with skepticism, worrying about the possibility of backslide towards a negotiation-based system if too many non-lawyers serve on WTO panels and the Appellate Body (Appleton, 2016). The thinking goes that lawyers are a bulwark against this kind of backslide because they are more loyal to formalist rules when conducting legal analysis, therefore making them better suited to further professionalize the dispute settlement process, increase predictability, and compensate for the increasing complexity of international disputes (Young, 1995; Hudec, 1999; Davey, 2002).

Not everyone agrees that it is a good idea today to promote more legalization and further entrench the rule by lawyers, however. These scholars question the wisdom of additional legalization, implicating its attendant coterie of lawyers and the rising costs of accessing the WTO dispute settlement body (Goldstein and Martin, 2000; Pauwelyn, 2005; Busch and Pelc, 2009). They suspect that non-lawyers are more instrumentalist, better attuned to state welfare needs, and introduce a certain amount of much-needed flexibility to the dispute settlement process (Pauwelyn, 2015). Such flexibility is thought to be important because states are understandably wary about the emergence of a technocratic organization willing to privilege jurisprudence over state interests.

The debate surrounding the normative desirability of non-lawyers in adjudicatory positions has profound implications for various areas of international trade law. For instance, commentators have noted with some concern the increasingly politicized environment surrounding panelist and Appellate Body member appointments (Elsig and Pollack, 2014; Shaffer et al., 2016). These

concerns are partially fueled by recent American efforts to discipline the Appellate Body for "'mak[ing] law' outside the context of resolving a dispute."[1] By blocking the reappointment of South Korean lawyer Seung Wha Chang to the Appellate Body, the United States is paving the way for more non-lawyers to become Appellate Body members in the future.

Looking beyond the WTO, one can find evidence of systemic ambivalence towards lawyers as adjudicators. Tellingly, the newest wave of bilateral trade agreements are split on whether lawyers are required for dispute settlement. Both the Transatlantic Trade and Investment Partnership's proposed Investment Court System and the Comprehensive Economic and Trade Agreement between Canada and the European Union are reasonably read to require its tribunal members to have law degrees.[2] But the Trans Pacific Partnership takes a softer stance, articulating no direct or indirect requirement for law degrees.[3] Similarly, the recently concluded EU-Japan Economic Partnership Agreement and the US-Korea Free Trade Agreement do not mandate their arbitrators have law degrees.[4]

Behind the entire debate is a key assumption that lawyers take a more formalist approach to adjudication, whereas non-lawyers are more instrumentalist. However, this assumption has not yet been empirically verified and there are reasons to doubt its validity. While it may be true that international dispute adjudication is an inherently legal task that is evaluated in a legal context, it does not then follow that lawyers are necessarily more formalist. The reverse could be true as well. After all, adjudicators work in social settings, and non-lawyers may be incentivized to be more rigorously formalist in their analysis in order to compensate for the fact that their opinions will be evaluated by the same legal audience and on the same legal standards as opinions produced by lawyers.

We empirically explore the hypothesis that lawyers are more formalist

[1]Statement by the United States at the Meeting of the WTO Dispute Settlement Body Geneva, May 23, 2016, p.2.

[2]See TTIP Draft Art. 3(4)(9)(4) and CETA Draft Article 8.27(4): "The Members of the Tribunal shall possess the qualifications required in their respective countries for appointment to judicial office, or be jurists of recognized competence. They shall have demonstrated expertise in public international law." Without stating it outright, these article effectively makes it a de facto requirement that tribunal members have law degrees and legal education.

[3]TPP draft articles 28.10(1)(a): [panelists will] "have expertise or experience in law, international trade, other matters covered by this Agreement or the resolution of disputes arising under international trade agreements."

[4]See the EU-Japan Free Trade Agreement Chapter XX, Article 10(a) and the U.S.-Korea Free Trade Agreement Art. 22.9(4)(b).

by comparing panel reports produced by lawyers to those produced by non-lawyers at the WTO. Our measure for formalism in panel reports is the precedent citation rate. The more an opinion cites previous cases, the more adjudicators are signaling that they are constrained and guided by jurisprudence.

Our findings show that non-lawyers cite precedent at higher rates than lawyers, providing evidence against the assumption that non-lawyers are less formalist. This difference in precedent citation rates of lawyers and non-lawyers decreases with experience, indicating that non-lawyers without an established reputation as competent panelists may be using higher citation rates to compensate for the lack of legal credentials and the negative associations that come with it. Our findings further show that the increased use of citations for inexperienced non-lawyers comes with a decrease in the probability that a panel report will be reversed by the WTO Appellate Body, suggesting that the increased rate of precedent citations has substantive consequences to the panelists.

Our finding that non-lawyers are more likely to signal adherence to formalist rules than lawyers challenges basic assumptions regarding their respective adjudication methods. Considering that there is a robust debate surrounding the value of further legalization of the WTO dispute settlement process, these findings should be taken into consideration by scholars and policymakers alike. More generally, the virtues attached to non-lawyer adjudicators—such as outside disciplinary expertise—may not come with a tradeoff of less legal rigor. Appointing non-lawyer adjudicators may thus enhance the quality of judicial decision, especially in areas where judges struggle with the evaluation of scientific or technical evidence.

The rest of the chapter proceeds in four parts. In section 4.2, we expand on our theoretical motivations and set forth background information relevant to our study. Section 4.3 describes our empirical methods and data. We present and interpret our results in section 4.4. A last section discussed the implications of our findings and concludes.

4.2 Theory

2.1 Formalism in Legal Opinions

The broad philosophical contours that define formalism and realism are well established. Formalism argues that the legal process is rationally determinate, involves mechanical adjudication, and demands that legal conclusions be found without recourse to non-legal reasoning (Leiter, 1999). It is based on deduc-

tive logic, with discoverable "correct" answers to every legal question (Posner, 1986). Some have even described it as a giant syllogism machine, where the adjudicator is merely a skillful mechanic (Neuborne, 1992). By contrast, realism argues that legal reasoning is rationally and causally indeterminate. Legal reasoning neither requires a specific outcome, nor is it sufficient to explain that outcome. Instead, decisions are reached with reference to the unique facts of each case and the real world consequences of any given legal conclusion. In its most ideal form, realist adjudication looks a lot like policy analysis (Dahl, 1957; Leiter, 2005). For simplicity's sake, we adopt a broad understanding of realist accounts that includes ideology-driven models of judicial behavior, such as the attitudinal model (Rohde and Spaeth, 1976; Segal and Cover, 1989; Segal and Spaeth, 1993; Sunstein et al., 2004), as well as the broader category of ideal point models (Martin and Quinn, 2002).[5]

It is a well-known characteristic of judicial systems that opinions need to be written in ways that reflect formalist legal reasoning (Chemerinsky, 2001; Wardlaw, 2010). In order to deflect criticism, even the most realist adjudicators cannot openly admit that their legal opinions are merely expressions of their individual preferences. Instead, they are required to craft "good" legal arguments that follow legal logic and canon, thereby embedding their opinions in the lineage of prior jurisprudence. Such requirements are costly in time and resources to adjudicators, but they are potentially rewarded. By signaling adherence to formalist standards, adjudicators can win the respect of their peers, prominent lawyers, politicians, and scholars. In some cases, career prospects are directly linked to their ability to signal formalism, and especially talented adjudicators are expected to creatively use precedents to show that they are suitable for higher judicial posts.[6] One point of note here is that it might not be required for adjudicators to actually feel inherently constrained by formalist rules; it may be sufficient that they act like artists or craftsmen, demonstrating laudable finesse when wielding these legal tools (Posner, 2008).

It has previously been demonstrated that the signaling of formalist tendencies is an important and strategic part of judicial behavior, both at the domestic and the international level. Choi and Gulati (2008) show that fed-

[5]Some authors further differentiate between realist and skeptic accounts, with the attitudinal model belonging to the latter, see e.g. Stephenson (2009). However, a broad understanding of realist behavior is more common, see e.g. Ho and Quinn (2010); Friedman, Martin, Clark, Lemos, and Larsen (forthcoming). A more nuanced differentiation does not provide any benefits to our study.

[6]Expectations about how much formalism should be signaled may be U shaped. Neophytes and masters may be expected to cite more, in order to demonstrate competence and brilliance, respectively.

eral circuit court judges are more likely to cite favorable out-of-circuit cases in both high-stakes scenarios and if their opinion is accompanied by a written opposing opinion. This indicates that the use of precedent, while being costly, lends legitimacy and authority to the judge's decision and deflects criticism. Supreme Court justices display similar strategic behavior in their use of citations, increasing the use of references to the Federalist Papers (Corley et al., 2005), rhetorical sources (Hume, 2006) and legal precedent (Lupu and Fowler, 2013) as a reaction to dissensus. Lupu and Voeten (2012) show that judges on the European Court of Human Rights are most likely to signal reliance on precedent if the value of persuading those domestic courts responsible for implementing the opinion is particularly high.

We assume (and later show) that the value of signaling through precedent citation is just as high, if not higher at the WTO. Because the WTO exists at the mercy of the states (Guzman, 2008a), there is an increased need for panelists to convey to governments that their hands are tied and that their decision is a necessary conclusion based on the applicable provisions and jurisprudence. Commentators have remarked that WTO panelists are "anxious to demonstrate to governments and the wider public that their decisions are not based on subjective political values."[7] The attempt to achieve "legitimacy through technicality" at the WTO lead to a remarkably dry and highly technical legal writing style in panel reports that is reminiscent of the legal opinions written in some civil law jurisdictions; one that is often hard to understand for those outside of the profession (Schwartz, 2001).

Our query focuses on differences in tendencies between lawyers and non-lawyers to signal formalism via precedent citation. In accordance with much of the literature on strategic judicial behavior, we assume the writing of judicial opinions to be costly and that these costs extend to the use of precedent (Landes and Posner, 1976; Choi and Gulati, 2008; Lupu and Voeten, 2012; Lupu and Fowler, 2013; Badawi and Baker, 2015). Among the different signaling mechanisms, we focus on the use of precedent as a way to indicate that the panelist can engage with notoriously complex WTO case law and embed their decisions competently in its jurisprudence. Indeed, precedent citation is implicated both philosophically and practically with formalist ideals. Philosophically, formalist ideals stress that the only sources of legitimate legal conclusions are statutes, regulations, and prior judicial decisions (Stephenson, 2009). Thus, an adjudicator must engage in close and disinterested analyses of such sources to tease out any legal answers. In order for adjudicators to signal that they are doing a good job at being formalist, they must employ ci-

[7]Schwartz (2001, 7)

tations to demonstrate the provenance and importance of their legal theories, arguments, and conclusions (Solum, 2006). In addition, precedent citation is indispensable to making formalist adjudication methods work, as the cost of ascertaining whether legal conclusions are "correct" would skyrocket without sufficient road signs from previous adjudicators. On a practical level, precedent citations are easy to quantify and thus compare in an objective way. Translating other measures of formalist tendencies, such as whether dissenters in a prior decisions rule in conformance with precedent in a subsequent decision, necessarily increases the amount of subjective interpretations the researcher has to make.

2.2 Lawyers versus Non-Lawyers

International adjudicators are not straightforward analogs of their domestic counterparts; the anarchical backdrop of international relations makes simple comparisons problematic. Unlike domestic judges, international adjudicators have to balance individual state party's interest in submitting to adjudication against a systemic interest in ensuring the existence of an independent and impartial dispute settlement organization (Guzman, 2008b).

Given this backdrop, the lawyer/non-lawyer dichotomy is of particular concern to states. Lawyers are thought to be more protective of judicial independence, more willing to follow precedent, and more interested in promoting legalization (Young, 1995; Hudec, 1999; Davey, 2002). Non-lawyers, by contrast, are allegedly more prone to negotiated solutions, even at the cost of decreasing overall legalization levels at an international organization (Weiler, 2002; Appleton, 2016).

Since many commentators forcefully argued that increasing legalization is essential to the WTO's success as a dispute settlement body, the collective profile of present-day WTO panelists is a little curious. Although a little more than half have legal backgrounds, very few panelists have private law experience and nearly none have domestic judicial experience. Instead, the overwhelming majority of panelists have significant experience serving in government, and many have been diplomats based in Geneva (Pauwelyn, 2015).[8] These ratios are somewhat surprising, suggesting that—for states anyway—the debate over the value of legalization, lawyers, and non-lawyers is far from over. In fact, some commentators have recently noted that states increasingly favor panelists that have policy and WTO experience over international law

[8] 88% of WTO panelists have significant governmental service experience, 57% have been Geneva-based diplomats, 56% have law degrees, 15% have a background in private law, and only 3% have judicial experience.

backgrounds and litigation experience (Elsig and Pollack, 2014).

For the purposes or our study, we define lawyer panelists as panelists who have received at least one postsecondary law degree. As pointed out above, the inclusion of precedent is assumed to be costly. At the same time, lawyer panelists, being better versed in prior WTO case law and its legal relevance, can more easily identify and marshal precedential decisions to support their own arguments. Because the cost of identifying and using precedent is lower to lawyer panelists, one could reasonably expect lawyer panelists to cite more precedent than their non-lawyer counterparts.

Hypothesis 1: Lawyer panelists cite precedents at a higher rate than non-lawyer panelists.

But now consider another plausible model, where observers treat non-lawyer panelists differently than lawyer panelists. Observers may assume that non-lawyers are less acquainted with WTO case law, and therefore demand punctilious work from them. Such disparate treatment would then change the incentives presented to non-lawyers and induce them to cite more than lawyers to appease these outside observers.

There are several potential sources for this disparity in treatment. First, lawyers may question non-lawyer's ability to produce good legal analysis, incentivizing non-lawyers to cite precedent more frequently to show that they understand the law. Lawyers may display this kind of unease regarding non-lawyers because WTO panel proceedings are trial-like affairs, where parties submit sophisticated written arguments, attend oral hearings, and appeal decisions that turn on highly technical readings of legal text. Indeed, Hudec (1999) and Davey (2002), both of whom were prominent participants in the WTO dispute settlement procedure, displayed skepticism towards panelists without legal training and have cautioned that the increasing complexity of WTO cases requires more lawyers to be appointed in the future. In addition, panel reports are scrutinized largely by lawyers. The Appellate Body, which consists mostly of lawyers, regularly reviews panel reports.[9] Moreover, the decision to appeal a panel report is made by governmental legal departments staffed by lawyers, sometimes working in conjunction with major international law firms.

Second, the WTO Secretariat, primarily through its Legal Affairs Division, may provide non-lawyer panelists with more information about prior case law.

[9] As of December 12, 2016, five out of seven current members and eighteen out of twenty former members of the Appellate Body are lawyers.

Whether consciously or not, the WTO Secretariat staff may assume that non-lawyer panelists are not as well versed in WTO law as lawyer panelists. They may also more actively advocate for certain interpretations of the law when presenting research to non-lawyers.

Third, state parties may hold similar negative assumptions about non-lawyers, prompting them to submit briefs with more precedent citations in the text. Since judicial economy concerns may cause adjudicators to adopt the arguments and language of parties, extra information presented as precedent citations could ultimately affect the rate at which precedents are cited in final panel reports(Stephenson, 2009).

All three sources of disparate treatment stem from a common skepticism towards non-lawyers, questioning whether they are sufficiently acquainted with WTO jurisprudence. Such skepticism could then lead to higher citation rates in decisions written by non-lawyers. A potential corollary is that citation rates may decrease as non-lawyers become more experienced and establish a reputation as competent WTO panelists.

Hypothesis 2.1: Lawyer panelists cite precedents at a lower rate than non-lawyer panelists.

Hypothesis 2.2: The difference in precedent citation rates decreases as non-lawyers become more experienced.

To be sure, one might contend that precedent citation is of less importance at the WTO than in some domestic court systems. Indeed, the WTO's official stance on precedent has long been considered curious. While neither the Dispute Settlement Understanding (DSU) nor the WTO Agreement explicitly rule out precedent being binding at the WTO, these two instruments are reasonably read to preclude it. However, we consider it unlikely that this results in precedent citations being irrelevant in WTO panel reports. After all, the Appellate Body has made it abundantly clear that panelists of the first instance are expected to take precedents very seriously. In fact, when panelists write reports that directly contradict Appellate Body precedents, the Appellate Body often singles out those reports for criticism.[10] As for the panelists themselves, there are strong incentives to avoid being publicly shamed by the

[10]"It is well settled that Appellate Body reports are not binding, except with respect to resolving the particular dispute between the parties . . . [t]his, however, does not mean that subsequent panels are free to disregard the legal interpretations and the ratio decidendi contained in previous Appellate Body reports that have been adopted by the DSB.", Appellate Body Report, *United States – Final Anti-Dumping Measures on Stainless Steel from Mexico*, WT/DS344/AB, adopted 30 April 2008.

Appellate Body, even if most WTO panelists are drawn from the ranks of government bureaucrats (Pauwelyn, 2015).[11].

Nonetheless, we take seriously the need to demonstrate that precedent citation rates are practically important to the outcome of a WTO decision. WTO decisions have previously been distinguished based on their quality, where the quality is measured as the probability for reversal, conditional on appeal (Busch and Pelc, 2009).[12] We adopt this measure of judicial opinion quality and analyze whether the citation rates are associated with a difference in reversal probability.[13]

Hypothesis 3: Changes in citation rates are associated with changes in the probability of reversal.

2.3 How WTO Reports are Drafted

Before moving on to the data, it is useful to briefly sketch out the process of how WTO reports are drafted, as this process differs from the writing of judicial opinions in a domestic court system. The following description is based on interviews conducted by one of the authors with members of the WTO Secretariat.

WTO panel reports are collaborative efforts that involve substantial coordination between panelists, chairs, and the Secretariat. From a practical standpoint, the panel decision-making process has to accommodate geographical realities and the busy schedules of panelists drawn from across the world. Despite these obstacles, panelists will typically travel to Geneva three times over the course of any given dispute in order to deliberate and work face-to-face.

Panel chairs oversee the proceedings to ensure that the final report meets his or her preferences. Some chairs take a more active drafting role, making

[11] While reappointment is usually not in the cards for WTO panelists, the panelists stature at home may be diminished by being singled out for criticism. The panelist's credibility as an authority on the WTO DSB may also be damaged. Finally, specific criticism can also be detrimental to a panelist's sense of professionalism.

[12] Also see Busch and Pelc (2009) for a discussion on why it is justified to analyze the subset of appealed cases only.

[13] Note that we do not make the causal claim that precedent citation rates influence reversal probability. Indeed, we think it unreasonable to assume that just adding citations at more or less arbitrary places in a decision decreases the probability for reversal. However, as citing precedent systematically can signal support by prior case law, we think that a meaningful correlational relationship may exist.

their preferences regarding precedents known to other panelists and the Secretariat. Others chairs, however, see themselves more as facilitators, promoting more consensus-based approaches and smoothing over disputes between panelists and the Secretariat.

No matter how chairs see their role, the WTO Secretariat plays a major part in the dispute settlement process. Its job is to facilitate and guide the panels.[14] When a panel is first composed, the Secretariat will assign a legal team to assist the panelists with all aspects of the case. It typically falls on the Secretariat legal team to prepare case documents, liaise with state parties, and arrange for the chair and panelists to work in Geneva. The legal team will also research and draft panel reports in close conjunction with their assigned chairs and panelists.[15] Once drafting is complete, the chair and panelists will review the reports line by line, making changes as they see fit. As a result, panel proceedings are best described as collaborative endeavors between the panelists and the Secretariat.

The panel reports themselves tend to follow established patterns. The findings section–the focus of this study–will typically begin by identifying legal rules relevant to the dispute at hand, drawn from previous WTO and General Agreement on Tariffs and Trade (GATT) reports. By doing this, the panel is carefully setting up the contours of the legal environment under which it will conduct subsequent analysis; citing precedents here constrains the universe of reasonable arguments that the panel can make later in the report.[16] After presenting the legal rules, the report will analyze each separate issue raised by the complainant and present the panel's determination.

4.3 Data Description and Methodology

3.1 Population

The relevant unit of observation in this study is the panel report, produced by WTO dispute settlement panels of the first instance. Our data set contains every panel report that has been written by WTO panels from its inception

[14] Article 27 DSU.

[15] Occasionally a panelist will prefer to personally draft a section of a report, but this is considered rare.

[16] Although it is conceivable that panels cite without relying on what they cite, we think this is unlikely. For one thing, it suggests that panels are spending substantial amounts of energy in a frivolous endeavor. Moreover, the overwhelming percentage of precedent is cited in support of a rule from another case, meaning that the drafters are deliberately choosing to cite those rules.

until 2015.[17] By definition, we have no information on disputes in which the parties settled at an early enough stage so that a panel did not issue a complete report. We exclude GATT panel reports, as the transition from the GATT to the WTO was accompanied by significant structural changes. We also exclude reports produced by the WTO Appellate Body and arbitration panels, as their functions are fundamentally different from those of regular dispute settlement panels. The Appellate Body is composed of seven fixed members, each serving four-year terms. As such, they do not provide a comparable variance in panelist characteristics. Functionally, they lack their own fact-finding ability and only consider questions of law. Arbitration panels follow their own idiosyncratic rules and are thus not comparably embedded into the WTO system. Having thus cabined the scope of our study, our data set consists of the population of all 194 regular panel reports written between 1995 and 2015.

Since this is the study of the population of all reports, rather than a sample of all WTO disputes, we caution that all of our results should be interpreted as estimates for disputes that result in panel reports and that extrapolation to any broader population of disputes requires additional assumptions. In particular, our results do not directly speak to the many WTO disputes that settle or are not pursued by the complainant anymore prior to the decision of a panel.

3.2 Data

The outcome of primary interest is the rate at which panelists cite precedent from cases produced by the Appellate Body of the WTO. We limit ourselves to citations to Appellate Body decisions because Appellate Body reports are the only kind of precedent which is clearly recognized as being of legally persuasive quality. Previous studies have identified precedent using an often adopted codified notation, taking the form of "WT/DS***/AB", where *** stands for the number assigned by the WTO to a particular dispute.[18] However, we found that this method is prone to significant error, because different panel reports use different citation styles. In about 60% of reports, most references to precedent take the above form. But in the remaining 40% of cases, panels cite precedent either by adding the abbreviated dispute number only the first time a report is mentioned, or not at all.[19] As such, relying only on the codified form of dispute identifiers leads to an underinclusive data set, so we complement

[17] The panel reports are available at https://www.wto.org/english/tratop_e/dispu_e/dispu_status_e.htm.

[18] See the definition in Pelc (2014).

[19] If no dispute numbers are used throughout the text, it is typical for a report to be preceded by a table that indicates which short-hand notation references any given case, e.g.:

this approach with an alternative. We exploit the fact that, if not accompanied by a dispute number, references to other cases are always written in italics and start with the respondent, followed by a space, a hyphen, another space, and a short description of the case, e.g. *US – Softwood Lumber V*, *Argentina – Textiles and Apparel* or *EC – Bananas III*.

Since manually counting references to precedent is labor intensive and prone to error, we rely on Python to conduct automated text processing. The clear advantage of this approach is its objectivity and reproducibility, as human judgment does not factor into the identification process. However, there is a potential caveat: not every reference to precedent indicates support for a past decision. Instead, a panel might cite a previous report as a way to distinguish its current decision. To check the plausibility of this concern, we took a random sample of 10% of our disputes and manually counted precedent citations, noting whether each citation was in favor of or against the findings and conclusions in previous decisions. Based on this sample, 99.5% of references to prior reports are in favor, which is why we discount the possibility that our approach is overinclusive to any relevant degree.

A typical panel report is divided into eight sections: 1) Introduction, 2) Factual Aspects, 3) Findings and Recommendations, 4) Arguments of the Parties, 5) Arguments of Third Parties, 6) Interim Review, 7) Findings of the Panel, 8) Conclusions and Recommendations Requested by the Parties. It is not rare for sections four and five to be the longest. However, we do not analyze the citations in these two sections because they are not part of the legal argument of the panel. Instead, these sections are restatements of the parties' arguments. Because we are only interested in the panelists' own probability of citing precedent, we exclusively consider a citation if it appears in section seven, "Findings of the Panel." As alternative outcome measures, we also include the variety of Appellate Body reports cited in a given decision as well as citations to regular panel reports.

Based on the hypotheses presented in the introduction, our main variable of interest is whether the chair is a lawyer, holding a law degree in her respective home country. To code this variable, we rely on data provided by Pauwelyn (2015). Pauwelyn's data set includes substantial background information on each panelist at the WTO, collected by analyzing their CVs and conducting extensive internet searches and personal interviews. Among the variables in the data set is a binary variable indicating whether the chair holds a law degree, which we use as our treatment indicator.

Japan – Alcoholic Beverages; Appellate Body Report, *Japan – Taxes on Alcoholic Beverages*, WT/DS11/AB, adopted 4 October 1996.

We are interested in the effect that assigning a lawyer chair to a dispute has on the frequency of citing precedent. Defining treatment as the assignment of a lawyer chair guarantees that our treatment indicator T is well defined and does not rely on the alteration of immutable characteristics. Treatment status for observation i can be changed from 0 to 1 simply by assigning a different chair to the dispute. Figure 4.1 provides a descriptive overview over the distribution of citation practice across time for lawyer chairs and non-lawyer chairs. It demonstrates no significant patterns, safe for the fact that the four panel reports that cite precedent most frequently are written by non-lawyers. A closer examination reveals that these decisions have all been written under the same chair, Alberto Juan Dumont, a finding is taken into account when performing robustness checks.[20]

3.3 Methodology

If complaints were randomly assigned to panels, we would be able to treat WTO decision-making as a natural experiment and identifying the causal effects of interest would be fairly simple. But this is not the case. According to Article 8.6 of the DSU, the Secretariat first proposes nominations to the panel, which have to meet certain criteria formulated in Articles 8.1, 8.2 and 8.10 DSU.[21] Further, Article 8.10 requires that at least one panelist is from a developing country if the dispute is between a developing and a developed country. The parties then have the opportunity to reject panelists for "compelling reasons."[22] Subsequently, the Secretariat proposes new names. If the parties still cannot agree on a set of panelists, they are determinatively selected by the Director-General of the WTO. According to our data, this occurs in about 60% of disputes. It is at least possible that certain types of decisions are more (or less) likely to be assigned to panels with lawyer/common-law chairs and that these decisions require more (or less) reference to precedent.

Because treatment is not randomly assigned, identification of the treatment effects relies on the selection on observables assumption. That is, treatment assignment is required to be independent of the potential outcome Y_{it} for disputes with similar covariates X_i. Formally, we assume that

$$E[Y_{it}|X_i, T_i] = E[Y_{it}|X_i]$$

[20] These decisions are *US – Shrimp (Ecuador)*, *US – Carrier Bags*, *US – Zeroing (Korea)* and *US – Shrimp and Sawblades*.

[21] For example, they have to be sufficiently experienced and diverse.

[22] *Understanding on Rules and Procedures Governing the Settlement of Disputes*, Apr. 15, 1994, art. 8.6, in *Marrakesh Agreement Establishing the World Trade Organization*, Annex 2 (1994).

We argue that there are three main dimensions which could govern both the choice of which chair to appoint to the dispute and the propensity to cite precedent.

First, we control for panel-specific characteristics. As mentioned above, roughly 60% of panels are appointed by the Director-General and we control for this appointment procedure. We further combine data for the chair's home countries with the JuriGlobe database on legal origins to determine whether the chair comes from a common law country or a civil law country.[23] We also control for the chair's experience, measured as the number of panels she previously participated in. Our data further includes indicator variables for the sum of law degrees among the other two panelists, as well as the sum of common-law trained panelists.[24]

Second, there are dispute-specific covariates that need to be controlled for. For example, the central *National Treatment* doctrine at the WTO, formulated in GATT Article III, requires states to not discriminate between national and foreign "like products". Whether two products are "like" has been and continues to be subject of great legal debate, with some emphasizing that like products have to have similar external attributes, such as shape and form, and others requiring that like products have to be of similar use to the consumer (which translates into a high cross-price elasticity). Consequently, panelists have much precedent to draw from when considering whether two products are like, such as the famous case on *Japan – Alcoholic Beverages*.[25] At the same time, GATT Article III cases are often legally complex and might require the appointment of a chair who is a lawyer. If left unaccounted for, the dispute type could potentially be correlated with both the propensity for treatment

[23]JuriGlobe data available at http://www.juriglobe.ca/eng/sys-juri/index-alpha.php. The JuriGlobe database uses five different categories: Civil, Common, Muslim, Jewish and Customary. When a country has a mix of multiple legal systems, we code it 1 if its legal system is based on the common law and other systems which are not the civil-law system. For example, India has common law, customary law and Muslim law and is thus coded 1 in our data set. Similarly, we code a country 0 if its legal system is based on the civil law and other legal systems, but not the common law. For example, Japan uses both civil and customary law and is coded 0 in our data base. This leaves us with three states home to panelists with legal systems that possess aspects of both civil and common law: Israel, Mauritius and South Africa. We code these three states 0, indicating that they are not predominantly common law systems. However, whether these are coded 0 or 1 has no significant effect on our results.

[24]From Peresie (2005), we know that adding a female judge to a panel might have varying effects, depending on the gender of the remaining judges. Those effects do not have to follow any specific functional form, which is why we choose to factorize our measure.

[25]Appellate Body Report, *Japan - Taxes on Alcoholic Beverages*, WT/DS8/AB/R (Oct. 4, 1996).

assignment and our outcome variable.

If one were to take the view that each dispute is unique, this would call for the inclusion of dispute fixed-effects and statistical analyses would become impossible. Instead, we classify disputes based on a number of covariates, drawing from the WTO Dispute Settlement Database by Horn and Mavroidis.[26] This database codes the GATT articles and other codes that the complaining party alleges were violated.[27] We include indicators for when the moving party complained of violations of GATT Art. I, Art. II, Art. III, the Anti-Dumping Agreement (ADA), the Safeguards Agreement (SG), the Agreement on Subsidies and Countervailing Measures (SCM), and the Sanitary and Phytosanitary Agreement (SPS) in connection with GATT Art. XII.

Further, disputes might be of differing economic importance to the parties, which can translate into strategic behavior both during the process of selecting the panelists and when writing an opinion. In order to control for the economic stakes of a dispute, we include the volume of bilateral trade between complainant and respondent for the goods mentioned in the dispute. The corresponding data comes from Kucik and Pelc (2015), and we complemented it using the WTO Dispute Settlement Database by Horn and Mavroidis and the analytical data base of the World Bank's World Integrated Trade Solution.[28]

The third dimension we control for is the party dimension. Since Article 8.10 requires at least one panelist from a developing country on the panel if the dispute is between a developing and a developed country, we include an indicator for such disputes. Furthermore, the legal system of the parties may matter. It could be the case that common-law countries have a higher propensity to nominate common-law panelists, as they speak the same "legal language." We thus control for the number of common-law countries that are part of the dispute.

Lastly, we include time fixed effects by year of panel constitution to control for the fact that later cases simply have more precedent to cite. We also control for the length of the report prior to the panel's analysis, as it is possible that a panel may present an especially elaborate recitation of the parties' arguments, decreasing the panel's need to lay out its own legal argument in such detail.

Table 4.1 summarizes the variables we use in our analysis. Our estimand of interest is the average treatment effect on the treated (ATT), defined as

$$\tau_{ATT} = E[Y_{i1} - Y_{i0}|T_i = 1]$$

[26]Database available at http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/0,„conte
SitePK:469382,00.html.

[27]Note that the initial complaint is filed before the panel is established, thus making it a pre-treatment covariate.

[28]Available at http://wits.worldbank.org.

We obtain estimates for this quantity by combining multiple imputation, matching, and post-matching regression analysis.

Multiple Imputation

As Table 4.1 shows, our only covariate with missing values is the economic volume of the disputes. The typical approach to this type of missing data is to delete observations listwise. However, deleting listwise can introduce severe biases as missingness could occur non-randomly (Honaker and King, 2010). Relying on the Amelia II package for R (Honaker et al., 2011), we use multiple imputation based on the EMB algorithm (Honaker and King, 2010) to impute missing economic volume data. Multiple imputation is a procedure by which m values are imputed for each missing value in the data set. It is assumed that the data follows a multivariate normal distribution and that unobserved data is missing at random, conditional on the observed data. The m values are then drawn from the distribution of unobserved data and are imputed, creating m data sets that keep the same observed data and vary the unobserved data. The researcher can then perform the analysis on all m data sets and combine the results.

While it is generally recommended to use at least 5 to 10 imputations for each missing data point, we find that our estimates vary for such small numbers of imputations across multiple iterations with different starting values.[29] We increase the number of imputations to 100, because we only have a few missing data points in a single covariate and the EMB algorithm drastically increases computational speed when compared to more traditional algorithms. By increasing the number of imputations this way, there is virtually no variance in repeated iterations. Figure 4.2 depicts the difference in coefficient estimates and confidence intervals obtained by listwise deletion and multiple imputation for our full model.

Matching

After imputation, we use propensity scores to create a matched data set on which to conduct the analysis. Matching is an increasingly popular analytical

[29]The algorithm is a deterministic algorithm that attempts to find the global maximum of the likelihood surface for the complete-data parameters θ. As such, it is sensitive to the starting value. Further, the distribution of the unobserved data from which the imputed values are drawn may require the researcher to increase the number of imputations in order to obtain consistent estimates.

method in international legal scholarship.[30] It seeks to mitigate a major problem in small-sample observational and experimental studies, where the treatment and control groups might differ on a number of covariates and where observed differences in the outcome between the two groups might not be caused by the treatment variable of interest, but by a third covariate in which both groups differ significantly. In addition, matching addresses concerns stemming from a lack of common support in the covariates of treated and control units. In very large samples, exact one-to-one matching can yield the causal effect of interest simply by comparing the difference in means between two groups.[31] However, when exact one-to-one matching is not possible, it is recommended to supplement the matching algorithm with post-matching regression analysis (Ho et al., 2007). Our matching algorithm employs the most common method in the social sciences, nearest-neighbor propensity score matching without replacement. Our link function to estimate propensity scores is a probit link function.

The intent to apply both methods of multiple imputation and of propensity-score matching to the data raises a question about how these two methods should be combined. In general, there are three ways in which researchers have used multiple imputation and propensity-score matching together. Table 4.2 summarizes the three methods. The primary difference between them is the point at which they combine the estimates across multiple imputations. Method I combines the m imputed values into one imputation and then obtains propensity scores for a single data set using those combined imputed values. Method II combines the m estimated propensity scores and implements the matching algorithm on a single data set using the combined propensity scores. Method III implements the entire procedure for each of the m imputed data sets and only combines the estimates at the very end.

Method I is a procedure that can be found in the literature on international institutions (Chilton and Versteeg, 2016), though its effectiveness in reducing bias has not been formally studied. However, note that this procedure is equivalent to creating only one data set where missing values are obtained from a single draw from the unobserved data. Thus, many advantages of multiple imputation are effectively lost and the estimates no longer reflect uncertainty about the unobserved values. Whether Method II or Method III is more suited to minimizing bias in the estimates for the quantity of interest has not yet been formally demonstrated. In simulations, both methods are able to reduce bias

[30] Two recent examples include Chilton and Versteeg (2016) and Nielsen and Simmons (2015).

[31] Provided that the assumption of selection on observables holds.

significantly when compared to listwise deletion, but there is no consensus on which method is more effective when compared to one another (see Mitra and Reiter, 2011, 2016; Hill, 2004). But note that Method II is designed for a setting in which the quantity of interest can be directly obtained from the matched data, e.g. by computing differences in means. If it is intended to complement the matching algorithm with post-regression analysis, Method II is not suitable, as it leaves the researcher with a single data set. For this data set, the unobserved data is again missing, reintroducing the same problem that multiple imputation is intended to solve.[32] Method III is the only procedure that is consistent throughout in its treatment of the missing data problem by retaining the m imputed data sets until the analysis is completed.

Therefore, we believe that Method III is the most appropriate procedure for our study. The effect of the implementation of a matching procedure on the balance can be assessed by comparing the mean improvement in difference of propensity scores between treated and control units (Chilton and Versteeg, 2016). We thus collect the difference in the means of the propensity score for treated and control units before and after matching for each of the 100 imputations. Figure 4.3 plots this mean difference across all 100 imputations, together with 95% confidence intervals for Models (2), (3) and (4). The matching algorithm successfully and significantly increases the balance of our propensity scores by over 50%, in turn significantly ameliorating concerns that any results are driven by differences in observed characteristics or by extrapolation to a covariate range that lacks common support.

Post-Matching Regression Analysis

Our primary outcome of interest is the frequency with which precedent is cited in a given panel report. Since precedent citations are of the set of non-negative integers, count data models are suitable for the post-matching regression analysis. However, the decisions are of varying lengths. Finding 20 references to precedent in a section of 1,000 words is qualitatively different from finding 20 references in a section of 50,000 words. In the first case, precedent likely played a substantial part in the analysis, whereas in the second case, the citing of precedent is likely to play only a minor part in a larger legal analysis.[33] We are thus not interested in the absolute number of precedent references, but in the rate of reference, as this is our most accurate measure

[32]One can think of ways in which the researcher uses multiple imputation a second time after obtaining the combined propensity scores. However, we are not aware of any study that has attempted to implement such a procedure; its effectiveness remains untested and unknown.

[33]To be sure, it is possible that panels discuss precedent in great length, thus increasing

of the degree to which an opinion claims to rest on past opinions. The length of the section on the panel's findings can be understood as the exposure, i.e. the opportunities to cite precedent. The more words in this section, the more exposure and the more opportunities to cite precedent there are.

Let μ denote the expected count, n be the exposure (length of the section on the panel's findings in words) and $\boldsymbol{x}$ be the covariate matrix. Then our model is

$$log(\tfrac{\mu}{n}) = \alpha + \beta'\boldsymbol{x}$$

which can be rewritten as

$$log(\mu) = \alpha + \beta'\boldsymbol{x} + log(n)$$

This model can be estimated like any regular count model, where the logged exposure is included as an offset, i.e. the coefficient is fixed at 1.

Of the count data models, a deviance test suggests that the negative binomial model is more appropriate than a Poisson model. Since we have only few observation without Appellate Body citations, we prefer the negative-binomial model over the zero-inflated negative-binomial model. Our model of choice thus is the negative-binomial model. However, in order to rule out the possibility that our findings rely on the specific parametric assumptions of the negative-binomial model, we also present results based on a Poisson and a Beta regression as a robustness check. Because we run the post-matching regression analysis on each imputed data set separately, we combine the results of all 100 imputations using Rubin's rules (Rubin, 1988).

Hypothesis 3, that precedent citation rates are associated with a change in the probability for reversal, requires the definition of another outcome measure. We rely on the aforementioned database on WTO disputes by Horn and Mavroidis to code a panel report's outcome on appeal. Because data on the outcome on appeal is only available until 2011, we limit this part of the analysis to panel reports that have been circulated between 1995 and 2011. Note that the average appeal includes 12.39 claims of the panel erring in its findings or conclusions. The Appellate Body can accept or reverse a claim or take an action defined as 'other' in the data set.[34] We use a relative measure of dispute outcomes as proposed by Hoekman et al. (2008) that is the number of

the word count and lowering the share of precedent to word count, giving the impression that precedent is not of great relevance to an opinion. However, we find that a lengthy discussion of precedent is typically accompanied by recurring references to the same case. Our procedure should thus lead to a reasonable measure even in those latter cases.

[34]This includes, for example, cases in which the Appellate Body feels a determination is not possible due to a lack of evidence.

claims that have been reversed by the Appellate Body, divided by the number of claims it examined, as we believe this is the most accurate proxy of a decision's quality from a formalist perspective. Alternatively, one could conceive of a binary outcome measure that indicates whether a panel decision has been substantively overturned. However, "substantive overturn" is difficult to define in the context of the WTO. Busch and Pelc (2009) propose that a reversal on a single claim is sufficient to constitute "substantive overturn." However, there are a myriad of disputes where this is not the case. To illustrate, consider *EC – Chicken Cuts*,[35] a dispute about measures by the European Communities (EC) seeking to reclassify frozen boneless chicken cuts impregnated with salt from *salted meat* to *frozen meat* in order to impose higher tariffs. The panel held that the EC measures resulted in duties in excess of the EC schedule. The Appellate Body considered 11 different claims made by the parties. It reversed the panel on only one finding regarding the question of whether an importing state's conduct can unilaterally constitute "subsequent practice" under the VCLT 31(3)(b), a question that had no relevance for the panel's final conclusions. Examples such as these are plentiful and indeed, a cursory analysis suggests that a high reversal rate on individual claims correlates positively with substantive reversal. We thus prefer a relatively over a binary measure.

The association between citation rates and reversal probability can then be estimated using a negative-binomial regression, where the number of reversals is the outcome variable and where the number of claims is included as an offset.[36]

4.4 Results

4.1 Primary Results

Table 4.3 depicts our results for the effect of appointing a lawyer chair. Model (1) omits economic volume as a proxy for the stake of the disputes and Model (2) depicts the combined results of our analysis after imputing 100 values for each missing observation in volume, nearest-neighbor matching and the subsequent regression. Based on both models, we find that, holding everything

[35] *European Communities – Customs Classification of Frozen Boneless Chicken Cuts*, WT/DS269/AB/R (Sep. 12, 2005)

[36] Note that, while missing values are again imputed 100 times, the absence of a binary treatment indicator does not allow us to perform matching in combination with post-matching regression analysis. We thus omit the matching procedure.

else constant, the coefficient on our indicator for lawyer chairs is negative and significant at the 5% and 1% level. Therefore, the analysis provides support for hypothesis 2.1, namely that assigning a dispute to a lawyer chair decreases the propensity to cite precedent, and against hypothesis 1. In Model (3), we add two interaction terms. First, we interact *Lawyer Chair* with *Common Law Chair*, as the legal tradition might have an effect on precedent citation rates only if the chair was exposed to a thorough legal education. Second, we interact *Lawyer Chair* with *Chair Experience* in order to assess the plausibility of the causal mechanism we propose to explain for why non-lawyer chairs cite precedent at higher rates. We hypothesize that non-lawyers might cite more due to a disparate treatment of non-lawyers, compared to lawyers. This disparity could originate from the WTO Secretariat or the parties' briefs, supplying lawyers with different information than non-lawyers, or from the legal community at large, skeptical about the performance of non-lawyers in a quasi-judicial role. What all these rationales have in common is an assumption that non-lawyer could lack critical information required to draft legally compelling reports.

If this is indeed the case, then we expect the difference in citation rates between non-lawyers and lawyers to diminish as non-lawyers become more experienced and establish a reputation as competent WTO panelists (hypothesis 2.2). And indeed, this is what we find based on Model (3). The coefficient on *ChairExperience* is negative and significant for non-lawyer chairs, meaning that non-lawyer chairs tend to cite less when they become more experienced. The significant positive interaction effect indicates that experience has less of a negative effect on the citation frequency for lawyer chairs. Indeed, lawyer chairs cite precedent at slightly higher rates as they become more experienced.

We have previously highlighted the importance of the WTO Secretariat in drafting panel decisions. This implies that changes in the personnel of the Secretariat could lead to changes in the use of precedent in WTO panel reports. While our models include time fixed effects by year, thus effectively controlling for overall time trends including those caused by changes in personnel, it can be insightful to make the influence of the personnel more explicit. Even though information for lower level personnel involved in drafting the WTO opinions is not available, in Model (4) we include factor variables for the Director of the Secretariat's Legal Affairs Division. The Director of the Legal Affairs Division is responsible for the day-to-day planning and organizing of the division's workflow. He or she assists and supports dispute settlement panels and panelists, prepares legal opinions, reviews legal work, and builds relationships with panelists, and thus is the highest ranking official directly involved in the drafting of WTO opinions. The results indicate that, while there

is possibly some variance introduced by the different Directors, our findings remain substantively the same when controlling for them.[37]

Our results indicate that non-lawyer chairs cite significantly more precedent than lawyer chairs. This raises the question whether the additional citations are substantively meaningful or whether the differences are mere artifacts of divergent preferences in writing style, redundancies or due to a greater ability of lawyers to use precedent citations with precision. To investigate these questions, we re-estimate our preferred model using two other outcome variables. We reason that, if the increased number of citations is used to appear more formalist, then the difference we find between non-lawyers and lawyers should be limited to precedent that matters for the dispute at hand. Adversely, if the increased rate of precedent usage is a reflection of an inability to use precedent with precision or a mere stylistic preference, then we should see a difference between lawyers and non-lawyers not only for precedent that matters in the present dispute, but also for less relevant prior decisions. As stated above, our outcome of primary interest are citations to decisions of the Appellate Body, because vertical precedent is the only type of precedent that the Appellate Body clearly recognizes as legally persuasive. We first investigate whether the increased citation rate could be rooted in a tendency to redundantly cite the same precedent repeatedly or whether there is also a difference in the breadth of prior Appellate Body decisions that the panels cite. To that end, our dependent variable is the number of unique Appellate Body decisions referenced in a given report. We then examine whether there is a difference in citation behavior not only for Appellate Body reports, but also for regular panel reports for which a persuasive force is not similarly recognized at the WTO. The results of our analyses are presented in Table 4.4. They indicate that non-lawyers not only cite Appellate Body precedent more frequently than lawyers, but that they also cite a wider range of reports. We find the same pattern of a decrease in differences as non-lawyer chairs become more experienced. Further, the difference in citation frequencies is limited to Appellate Body decisions and does not extend to regular panel reports. Together, these analyses suggest non-lawyers' additional usage of precedent citations does not consist of mere redundancies or stylistic artifacts, but consists of what is considered substantively meaningful by WTO Appellate Body standards.

[37]Note that we do not use this model specification as our preferred model because it does not allow to simultaneously estimate time fixed effects for all years in the data set.

4.2 Robustness Checks

In order to asses the robustness and sensitivity of our results to our specific estimation strategy, we take four additional steps. First, in order to test for whether our results are driven by the parametric assumptions of the negative-binomial model, we reestimate the effects of our base model using a Poisson regression and a Beta-regression. In addition, we use the logged number of words of the "Findings" section not as an offset, but as a covariate, allowing its coefficient to change. This effectively relaxes the assumption that the rate of precedent citations per word is constant. As can be seen from Tables 4.5 and 4.5, the results for all model specifications are qualitatively similar.

Second, in Figure 4.1, we saw that four decisions written by Alberto Juan Dumont were the ones most likely to cite precedent. It is possible that these four decisions constitute outliers that drive our results. In Table 4.5, we thus re-estimate our preferred model while omitting these four decisions. As can be seen, while the outliers slightly increase our estimated base effect, all results remain qualitatively unchanged.

Third, we examine whether our findings rely on the particularities of the matching procedure. In matching, a caliper determines the range of propensity scores from which a control unit can be drawn and matched to each treated unit.[38] A narrow caliper yields a higher similarity between treated and control observations, decreasing the bias in the estimation. But omitting data points from a sample that is not particularly large can lead to known problems of small sample sizes and makes extrapolation more difficult. To address the potential consequences of this trade-off, we examine the effects of interest using calipers from 0.1 to 1.3. For calipers greater or equal to 1.3, all treated units are matched. In the Appendix, we report the number of units that are matched under each caliper as well as the reduction in bias.[39] Figure 4.4 plots our main estimates with 95% confidence intervals. As can be seen, both results are qualitatively insensitive to the caliper size, indicating that the particularities of the matching procedure are not the main drivers behind our results.

Fourth, we consider the possibility that our identification assumption of selection on observables is violated. As in any observational study, the possibility exists that the significance of the coefficient on *Lawyer Chair* is driven by an unobserved, omitted variable introducing endogeneity. In order to assess the sensitivity of the coefficients to the inclusion of additional covariates, it is

[38] Measured in standard deviations of the propensity score. If, for a given treated unit, multiple control units are within the caliper, one of them is randomly selected as the match.

[39] Under the narrowest caliper of 0.1, 61 treated units are matched to 61 control units, reducing bias in observed covariates by 97%.

common to observe how the coefficient behaves if only a subset of covariates is added to the model (Chiappori et al., 2012; Lacetera et al., 2012). The success of this procedure in identifying bias depends on the degree to which observed and unobserved covariates share the same covariance properties (Oster, 2017). In Table A.4.3, we include each dimension of control variables separately. What can be seen is that in each model specification, the coefficient remains significant and qualitatively the same. We acknowledge that this procedure does not rule out with certainty the existence of omitted variable bias. In particular, the establishment of a panel and with it our treatment is preceded by consultations (Art. IV DSU). The initial request for consultations is public information and we control for the relevant information contained in these requests.[40] But other than this initial request, consultations are confidential and it is possible that legal arguments involving precedent are exchanged at the consultation stage which affect both the subsequent panelist selection and the rate of precedent citations. We cannot be certain that these unobservable consultations do not introduce omitted variable bias. Similarly, Pelc (2014) has previously demonstrated that some countries strategically introduce economically insignificant test cases into the WTO in order to establish favorable precedent. If parties are aware of the relevance of panelist characteristics for the establishment of precedent, it is possible that they supplement their efforts of introducing test cases by strategically favoring certain panelist characteristics over others. While we implicitly control for this behavior by controlling for the volume of the dispute, it is still possible that the parametric structure of the generalized linear models used here inaccurately captures the true relationship. Nonetheless, we highlight again that any potential omitted variable must run largely orthogonal to the extensive list of covariates we include. The fact that the coefficient is qualitatively consistent across all covariate specifications, model specifications, and caliper specifications reassures us that it is improbable that there is a violation of the selection on observables assumption that is serious enough to pose a threat to our results.

4.3 Precedent and Reversal

We have found robust evidence that non-lawyer chairs cite more relevant precedent in their panel reports than lawyer chairs and consider now whether this difference leads to a difference in reversal probability on appeal. Table 4.7 depicts our results for hypothesis 3. The variable of interest is the relative

[40]The request for consultations typically does not include citations to precedent, but specifies which articles of the GATT are allegedly violated, which we control for.

occurrence of precedent citations. We include the same covariates as in our full Model (3), with missing values for volume imputed 100 times and results combined using Rubin's rule.[41] Model (5) considers if precedent citation rates are associated with a change in reversal probability. It indicates that higher citation rates decrease the probability of reversal. To further examine for what types of panelists precedent citations decrease reversal probability, in Model (6) we include a three-way interaction term between the citation rate, our indicator for whether the chair holds a law degree and the chair's experience. The model allows us to examine whether the decrease in reversal probability applies to both lawyers and non-lawyers and how this difference is related to the chair's experience. Indicated by the negative significant coefficient on the citation rate, Model (6) implies that inexperienced non-lawyers who frequently cite precedent get reversed less than inexperienced non-lawyers that cite infrequently. The significant positive coefficient on the interaction between citation rates and experience indicates that this difference decreases as non-lawyers become more experienced. At the same time, precedent citations appear to be disassociated from reversal rates for inexperienced lawyers.[42] As we show in the Appendix, the results are robust to alternative model specifications, but are somewhat sensitive to the inclusion of dispute-specific characteristics.[43]

[41]Instead of the year in which the panel was constituted, we include fixed effects for the year in which the panel report was circulated. This is suitable especially for disputes that spanned multiple years and in which the Appellate Body might have issued multiple new reports that the panel could potentially cite.

[42]The relevant coefficient can be obtained by adding up the coefficients on citation rates and the interaction of citation rates and chair lawyer. It is 0.2 and statistically insignificant. We also note that experienced lawyers seem to benefit from embedding their decisions in precedent, which is consistent with our previous finding that lawyers tend to cite more as they become more experienced. However, while we noted above that precedent citation usage might be U-shaped with respect to experience, we are not aware of well-established theoretical models that support this result and thus defer its interpretation to future research.

[43]A Beta specification cannot be estimated due to the occurrence of 0s and 1s in the proportion of reversed claims. A Poisson specification as well as a specification allowing the coefficient on the logged number of claims to vary yield statistically and substantively similar results. An analysis in which only a subset of covariates is included also yields results consistent with our main analysis. However, the size of the coefficients on precedent usage are moderately sensitive to the inclusion of covariates controlling for dispute characteristics. This sensitivity makes the presence of omitted variable bias somewhat more likely than in our analysis of differences in citation rates. In particular, since we group disputes based on the provision that was allegedly violated in the initial complaint, we cannot rule out the possibility that the coefficient size changes for a more granular measure of dispute characteristics. However, in light of the substantive consistency of our findings, we note that even alternative measures of dispute characteristics are unlikely to result in a reversal

4.4 Summary of Results

To summarize our results, we find that: 1) non-lawyer chairs cite more Appellate Body precedent than lawyer chairs; 2) that this difference decreases as non-lawyers become more experienced; 3) evidence consistent with the conjecture that additional citations by inexperienced non-lawyers are associated with a decrease in reversal probability.

Because the output of a negative-binomial regression is difficult to interpret, Table 4.8 translates all findings into the ATT, reported as a change in citation rates across treated observations. It shows that, on average, lawyers cite precedent about 30% less than non-lawyers, implying that the results are both statistically and substantively significant. This difference corresponds to an average difference in reversal rates of 6-7%.[44]

4.5 Implications and Conclusion

The normative debates surrounding the desirability of non-lawyers as adjudicators are framed around the supposed lack of formalist constraints imposed on non-lawyers. Fueled by the negative GATT experience, many scholars thus support decreasing the number of non-lawyer panelists in order to increase legalization of the dispute settlement body. Our findings contradict this assumption. Relying on both parametric and non-parametric estimation techniques, we find strong evidence that panels with legally trained chairs cite precedent at a substantively-and-statistically-significant lower rate than those panels on which the chair has not received formal legal training. The results hold even after controlling for a number of dispute- and panel-specific covariates across multiple dimensions. We also show that the difference in citation rates is driven by differential treatment, particularly of inexperienced non-lawyers. The Appellate Body rewards non-lawyers who increase citation rates with a decrease in reversal probability. If a highly legalized and formal dispute settlement body is deemed desirable, it might then be advisable to appoint more non-lawyers to the WTO, rather than fewer.[45]

---

of the sign on any of the relevant coefficients.

[44] These estimates are based on Model (5). We note again that Model (6) suggests the relationship between precedent citation rates and reversal rates to be heterogeneous. For instance, for the inexperienced non-lawyer, increasing the citation rate by 28% corresponds to a decrease in reversal rates by 10%, whereas for the experienced non-lawyer, it corresponds to a difference of 0%.

[45] We are making this statement with an appropriate degree of caution and cabin it to marginal changes in the proportion of non-lawyer to lawyer-panelists. Permanently changing

With that said, we do not wish our results to be interpreted as speaking directly to the optimal number of lawyer and non-lawyer panelists at the WTO. Our primary contribution is introducing much-needed facts to an important debate that international organizations face. On one hand, international organizations exist and operate to serve state parties and their myopic interests. On the other hand, they must guarantee a measure of neutrality and impartiality in order to act as arbiters of disputes. While it is unclear whether there is an optimal balance between state interest and neutrality, it is clear that the current narrative equating lawyers with more formalist decision-making process that promotes impartiality and non-lawyers with politicized jurisprudence is empirically unfounded. Rather than making a definitive normative claim, our results are meant to encourage a re-evaluation of the debate that corrects for the inaccurate assumption of non-lawyers being less formalist.

In addition to speaking to the narrower question of the optimal design of the WTO dispute settlement system, our findings also have implications for the broader literature on judicial decision making. Commentators have long contemplated the criteria that distinguish "good" from "bad" legal opinions, and whether writing "good" legal opinions is idiosyncratic to those trained in the law. The results of this study imply that receiving a legal education might be a sufficient condition, but it is not a necessary one. At least with respect to citing precedents, non-lawyers are able to produce work that successfully matches (and perhaps even exceeds) the standards expected by the legal community. Our findings are thus consistent with previous studies proposing an "organizational theory" of adjudication. These theories emphasize the organizational and professional context in which the role of the adjudicator is embedded, suggesting that socialization and homogeneity in standards by which to evaluate judicial output may offset the effect differences in pre-adjudicatory characteristics (Posner, 1993, 2008; Steffensmeier and Britt, 2001; Cohen, 2002). Especially in areas where detailed technical or scientific knowledge is required, the appointment of non-judicial adjudicators may thus add valuable expertise to a decision making body without necessarily sacrificing the legalized nature of the process.[46]

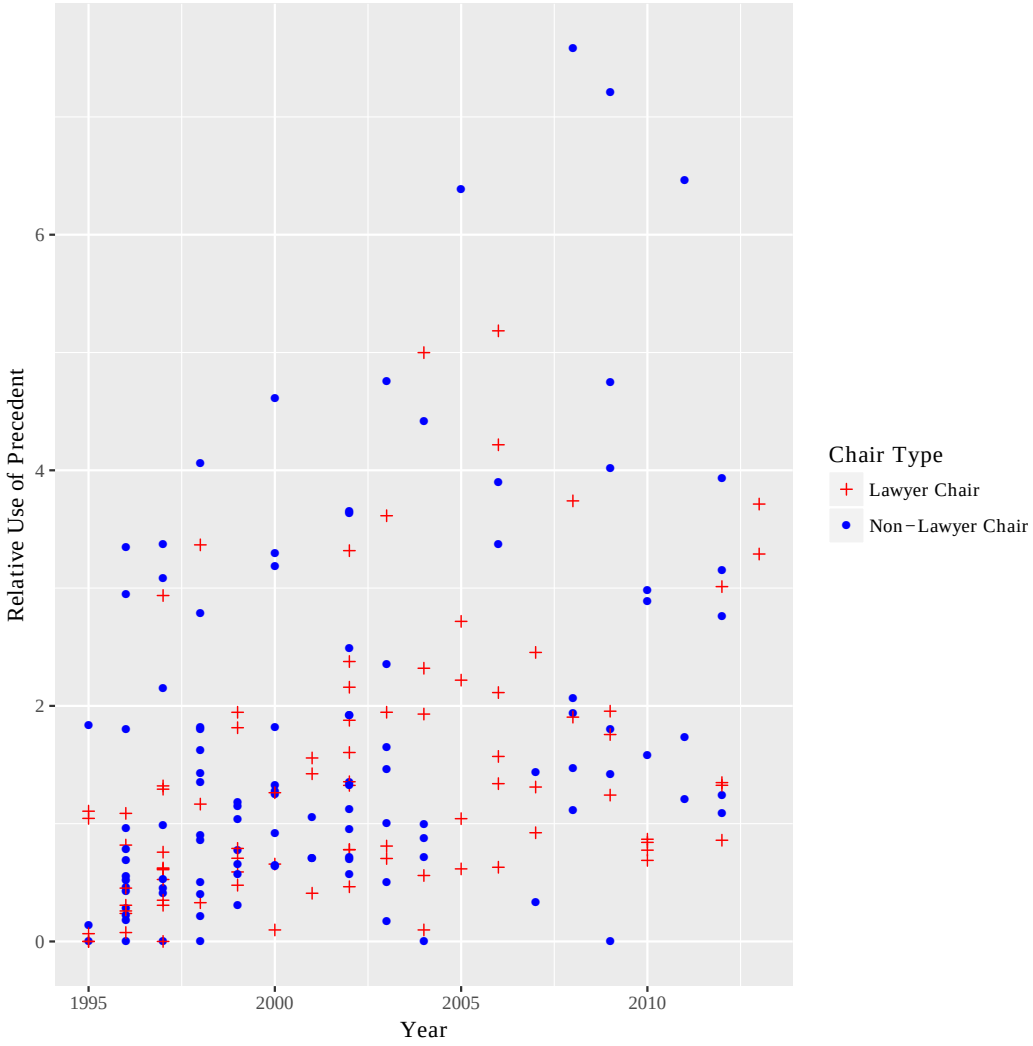
---

this proportion may have more profound, possibly adverse effects. For instance, as more non-lawyers are appointed, it may be the case that the standard by which panel reports are scrutinized changes, and with it the incentives to cite precedent.

[46]On an international level, these may in particular include environmental disputes or those touching upon concerns for public health, as well as investment arbitration in relation to large construction projects, see (Fach Gomez, 2016). While one may be inclined to suggest that expertise can be introduced through expert witness testimony, there is ample literature detailing the insufficiencies of this alternative, see e.g. Beecher-Monas (1998); Vidmar and

4.6 Figures and Tables

Figure 4.1: Citation Frequency by Panelist Type



This graph depicts the citation frequency of Appellate Body precedent per 1000 words for lawyer chairs and non-lawyer chairs, grouped by the year. Each dot represents one report.

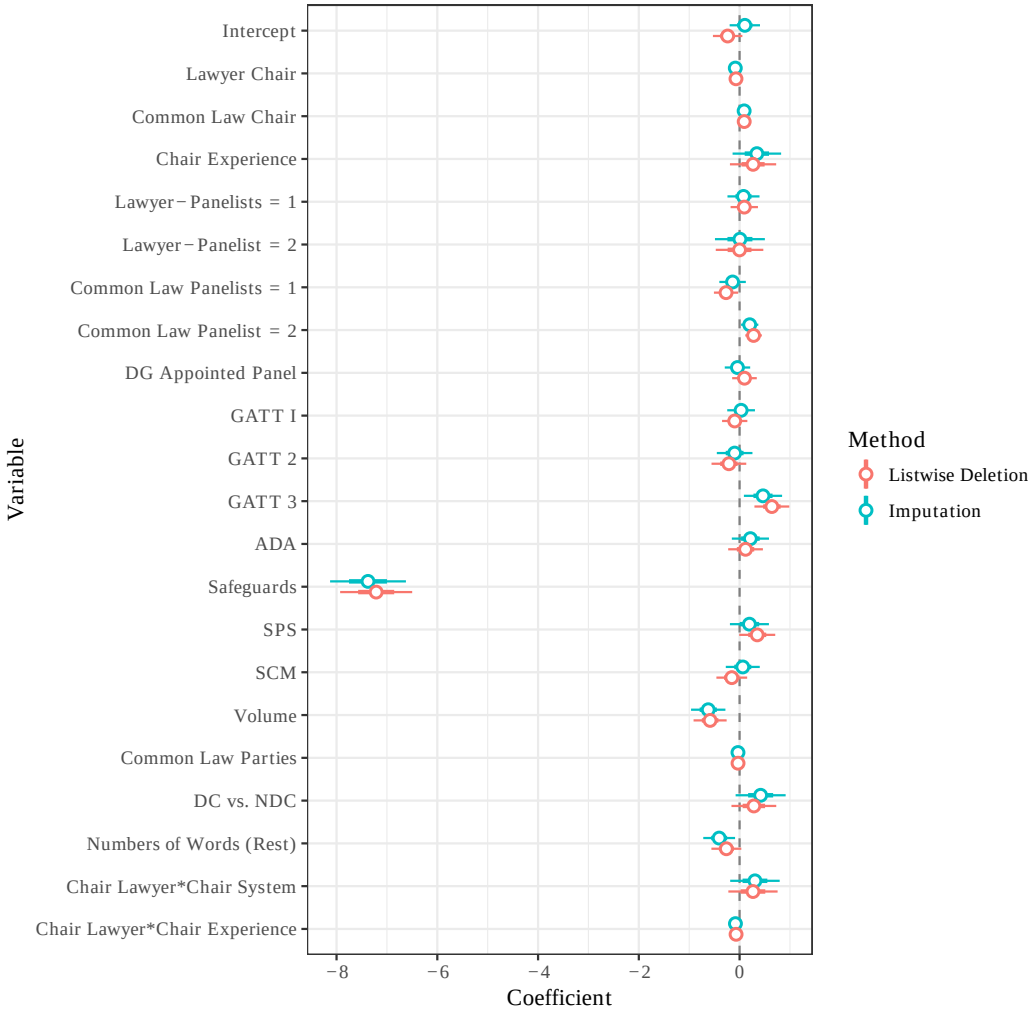
Diamond (2001); Gatowski et al. (2001); Moreno (2003).

Table 4.1: Summary Statistics

	Min	Max	Mean	SD	Med	IQR	NAs
AB Citations	0	488	81.37	96.88	48	81	0
AB Citation Variety	0	48	13.65	10.66	11	13	0
Panel Citations	0	216	30.06	41.28	13	33	0
Chair Lawyer	0	1	0.44	0.50	0	1	0
Chair Common Law	0	1	0.40	0.49	0	1	0
Chair Experience	0	13	2.88	2.86	2	3	0
Panelist Lawyer	0	2	1.21	0.69	1	1	0
Panelist Common Law	0	2	0.49	0.62	0	1	0
DG Appointed Panel	0	1	0.63	0.48	1	1	0
GATT I	0	1	0.12	0.32	0	0	0
GATT II	0	1	0.10	0.30	0	0	0
GATT III	0	1	0.15	0.36	0	0	0
ADA	0	1	0.28	0.45	0	1	0
SG	0	1	0.09	0.28	0	0	0
SPS	0	1	0.06	0.24	0	0	0
SCM	0	1	0.19	0.39	0	0	0
Common Law Parties	0	2	0.90	0.66	1	1	0
DC vs NDC	0	1	0.45	0.50	0	1	0
Year	1995	2013	2002	5.05	2002	8	0
Number of Words (Findings)	0	489	55.53	66.11	39	45	0
Number of Words (Rest)	0	418	57.75	65.98	43	75	0
Volume	0	14	0.93	2.14	0	0	31

Summary Statistics for the variables used in the analysis. Number of words x 1000, volume in billion $. "Year" refers to the year in which the panel was constituted.

Figure 4.2: Listwise Deletion vs. Multiple-Imputation



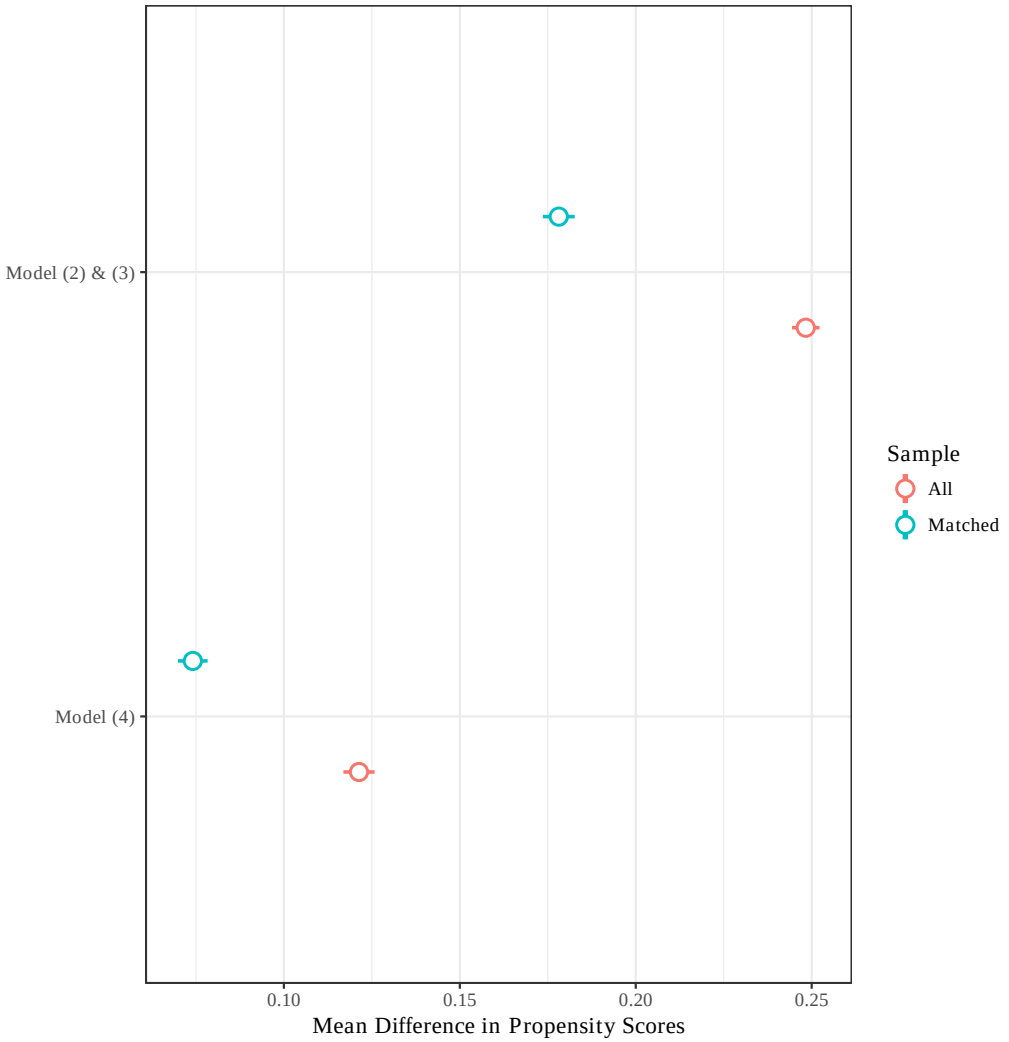
This graph depicts differences between listwise deletion and multiple-imputation in the coefficient estimates and confidence intervals of our analysis for Model (3).

Table 4.2: Combining Multiple Imputation and Matching

	Method I	Method II	Method III
Step 1	Create Imputed Data Set	Create Imputed Data Set	Create Imputed Data Set
Step 2	Repeat m Times	Compute Propensity Scores	Compute Propensity Scores
Step 3	Combine Imputed Unobserved Values	Repeat m Times	Implement Matching Algorithm
Step 4	Compute Propensity Scores	Combine Propensity Scores	Compute Quantity of Interest
Step 5	Implement Matching Algorithm	Implement Matching Algorithm	Repeat m Times
Step 6	Compute Quantity of Interest	Compute Quantity of Interest	Combine Results

A description of different procedures to combine multiple imputation and matching analysis.

Figure 4.3: Balance Improvement Through Matching



This graph depicts differences in the means of propensity scores between treated and control units for Models (2), (3) and (4).

Table 4.3: Post-Matching Regression Analysis (Lawyer Chair)

	Dependent Variable:			
	Precedent Citation			
	Model (1)	Model (2)	Model (3)	Model (4)
Lawyer Chair	-0.265^{*}	-0.248^{**}	-0.622^{***}	-0.470^{**}
	(0.112)	(0.100)	(0.171)	(0.169)
CL Chair	0.198	0.290^{**}	0.077	0.117
	(0.108)	(0.120)	(0.159)	(0.163)
Chair Experience	-0.023	-0.023	-0.085^{***}	-0.088^{**}
	(0.020)	(0.020)	(0.031)	(0.031)
Lawyer-Panelists = 1	0.003	0.052	0.063	-0.037
	(0.163)	(0.171)	(0.169)	(0.165)
Lawyer-Panelists = 2	0.157	0.191	0.195	-0.152
	(0.184)	(0.197)	(0.193)	(0.193)
CL Panelists = 1	-0.262^{*}	-0.071	-0.139	-0.346^{*}
	(0.120)	(0.131)	(0.131)	(0.141)
CL Panelists = 2	-0.296	-0.048	-0.007	-0.064
	(0.229)	(0.253)	(0.248)	(0.233)
DG Appointed Panel	-0.094	-0.037	0.030	-0.041
	(0.122)	(0.134)	(0.138)	(0.131)
GATT I	0.178	0.168	0.215	0.129
	(0.175)	(0.188)	(0.184)	(0.182)
GATT II	-0.149	-0.133	-0.098	-0.077
	(0.175)	(0.179)	(0.177)	(0.184)
GATT III	0.532^{**}	0.456^{**}	0.464^{**}	0.319
	(0.180)	(0.194)	(0.190)	(0.198)

Continued on next page

Table 4.3 – continued from previous page

	Model (1)	Model (2)	Model (3)	Model (4)
ADA	0.026	0.138	0.104	0.002
	(0.144)	(0.154)	(0.151)	(0.154)
Safeguards	0.386	0.253	0.418	0.292
	(0.226)	(0.251)	(0.248)	(0.222)
SPS	−0.359	−0.344	0.304	0.431
	(0.246)	(0.246)	(0.245)	(0.241)
SCM	−0.413**	−0.364**	−0.405**	−0.271
	(0.146)	(0.161)	(0.158)	(0.161)
Volume		−0.078**	−0.082**	−0.062*
		(0.031)	(0.030)	(0.031)
CL Parties	0.240**	0.200**	0.201**	0.253**
	(0.082)	(0.087)	(0.086)	(0.087)
DC vs. NDC	0.039	−0.006	−0.042	−0.129
	(0.120)	(0.128)	(0.126)	(0.126)
# Words (Rest)	−0.021*	−0.029***	−0.032***	−0.031**
	(0.010)	(0.010)	(0.010)	(0.010)
LC * CLC			0.343	0.182
			(0.241)	(0.258)
LC * Chair Experience			0.090**	0.089*
			(0.039)	(0.040)
Pieter Jan Kuijper (Director)				0.251
				(0.323)
Valerie Hughes (Director)				0.900
				(0.494)

Continued on next page

Table 4.3 – continued from previous page

	Model (1)	Model (2)	Model (3)	Model (4)
William Davey (Director)				-1.189^{*}
				(0.475)
Constant	-7.562^{***}	-7.616^{***}	-7.376^{***}	-6.326^{***}
	(0.374)	(0.375)	(0.376)	(0.356)
Time Fixed Effects	✓	✓	✓	✓
Observations	194	170	170	170
Pseudo-R^2	0.368	0.408	0.433	0.466
Log Likelihood	−907	−789	−785	−779
Akaike Inf. Crit.	1,664	1,653	1,650	1,640
Note:				*p<0.05; **p<0.01; ***p<0.001

The output of a negative binomial regression. (1) omits Volume, which contains missing values. For (2), missing values for Volume were imputed through multiple-imputation. Then, the data was matched using nearest-neighbor propensity score matching, where treatment is defined as having a lawyer chair on the panel. (3) adds interaction terms. (4) includes fixed effects for the Director-General of the WTO Secretariat. Standard errors in parentheses. AIC and Log-Likelihood estimates are averages across iterations. Volume in bio$, words in units of 10,000.

Table 4.4: Regression Analysis Using Alternative Outcome Measures

	Dependent Variable:	
	Unique Citations	Panel Citations
Lawyer Chair	-0.470^{**}	-0.199
	(0.131)	(0.230)
CL Chair	0.222	0.245
	(0.123)	(0.215)
Chair Experience	-0.076^{**}	-0.075
	(0.024)	(0.042)
Lawyer-Panelists = 1	0.215	0.014
	(0.136)	(0.227)
Lawyer-Panelists = 2	0.191	-0.031
	(0.153)	(0.258)
CL Panelists = 1	-0.166	0.320
	(0.103)	(0.179)
CL Panelists = 2	0.196	-0.252
	(0.186)	(0.345)
DG Appointed Panel	-0.119	0.019
	(0.106)	(0.187)
GATT I	0.078	-0.034
	(0.140)	(0.245)
GATT II	-0.197	-0.158
	(0.140)	(0.242)
GATT III	0.191	0.502^{*}
	(0.145)	(0.252)

Continued on next page

Table 4.4 – continued from previous page

	Unique Citations	Panel Citations
ADA	0.079	0.144
	(0.121)	(0.206)
Safeguards	0.202	−0.334
	(0.193)	(0.251)
SPS	0.112	−0.324
	(0.200)	(0.331)
SCM	−0.159	0.300
	(0.122)	(0.212)
Volume	−0.034	0.023
	(0.025)	(0.040)
CL Parties	0.146*	0.148
	(0.067)	(0.118)
DC vs. NDC	−0.010	−0.098
	(0.099)	(0.178)
# Words (Rest)	−0.036***	−0.027*
	(0.008)	(0.014)
LC * CLC	0.151	0.296
	(0.190)	(0.325)
LC * Chair Experience	0.071*	0.038
	(0.031)	(0.053)
Constant	−8.874***	−7.261***
	(0.370)	(0.475)
Time Fixed Effects	✓	✓

Continued on next page

Table 4.4 – continued from previous page

	Unique Citations	Panel Citations
Observations	170	170
Pseudo-R^2	0.405	0.297
Log Likelihood	−505	−645
Akaike Inf. Crit.	1,091	1,370
Note:		*p<0.05; **p<0.01; ***p<0.001

The output of a negative binomial regression of our preferred Model (3) with alternative outcome measures. "Unique Citations" regresses the count of unique Appellate Body reports cited in a decision on the specified covariates. "Panel Citations" regresses the count of citations to regular panel decisions on the specified covariates. Standard errors in parentheses. AIC and Log-Likelihood estimates are averages across iterations. Volume in bio$, words in units of 10,000.

Table 4.5: Alternative Model Specifications

	Dependent Variable:			
	Precedent Citation			
	Poisson	Beta	¬ Offset	¬ Outliers
Lawyer Chair	-0.762^{***}	-0.468^{***}	-0.604^{***}	-0.548^{**}
	(0.150)	(0.182)	(0.170)	(0.168)
CL Chair	0.079	0.157	0.088	0.103
	(0.162)	(0.142)	(0.159)	(0.152)
Chair Experience	-0.094^{***}	-0.054^{*}	-0.080^{**}	-0.076^{*}
	(0.032)	(0.026)	(0.023)	(0.030)
Lawyer-Panelists = 1	−0.128	−0.112	0.088	0.083
	(0.180)	(0.152)	(0.171)	(0.165)
Lawyer-Panelists = 2	−0.034	0.072	0.214	0.160
	(0.206)	(0.172)	(0.195)	(0.188)
CL Panelists = 1	−0.015	−0.077	−0.124	−0.084
	(0.119)	(0.117)	(0.195)	(0.131)
CL Panelists = 2	0.253	0.066	0.009	0.010
	(0.185)	(0.228)	(0.251)	(0.247)
DG Appointed Panel	−0.004	0.016	0.026	0.113
	(0.147)	(0.121)	(0.138)	(0.137)
GATT I	0.367	−0.028	0.242	0.269
	(0.184)	(0.171)	(0.181)	(0.176)
GATT II	−0.308	−0.177	−0.115	−0.089
	(0.201)	(0.162)	(0.175)	(0.176)
GATT III	0.376^{***}	0.395^{*}	0.471^{*}	0.387^{*}
	(0.170)	(0.173)	(0.187)	(0.181)

Continued on next page

Table 4.5 – continued from previous page

	Poisson	Beta	¬ Offset	¬ Outliers
ADA	−0.005	0.101	0.115	−0.027
	(0.173)	(0.132)	(0.154)	(0.149)
Safeguards	0.054	0.378	0.406	0.365
	(0.245)	(0.224)	(0.248)	(0.245)
SPS	0.148	0.458*	0.309	0.260
	(0.238)	(0.221)	(0.248)	(0.243)
SCM	−0.449	−0.276	−0.410**	−0.434**
	(0.140)	(0.141)	(0.157)	(0.153)
Volume	−0.101**	−0.072*	−0.089**	−0.089**
	(0.036)	(0.021)	(0.030)	(0.031)
CL Parties	0.090	0.156*	0.189*	0.170*
	(0.086)	(0.073)	(0.087)	(0.083)
DC vs. NDC	0.021	−0.034	−0.011	−0.044
	(0.110)	(0.112)	(0.127)	(0.124)
Log(# Words Findings)			0.972***	
			(0.077)	
# Words (Rest)	−0.034**	−0.035**	−0.029**	−0.031**
	(0.011)	(0.010)	(0.010)	(0.010)
LC * CLC	0.471*	0.157	0.337	0.306
	(0.224)	(0.250)	(0.241)	(0.237)
LC * Chair Experience	0.120**	0.073*	0.085*	0.088*
	(0.037)	(0.034)	(0.039)	(0.039)
Constant	−6.759***	−6.930***	−7.088**	−7.400***
	(0.367)	(0.387)	(0.727)	(0.371)

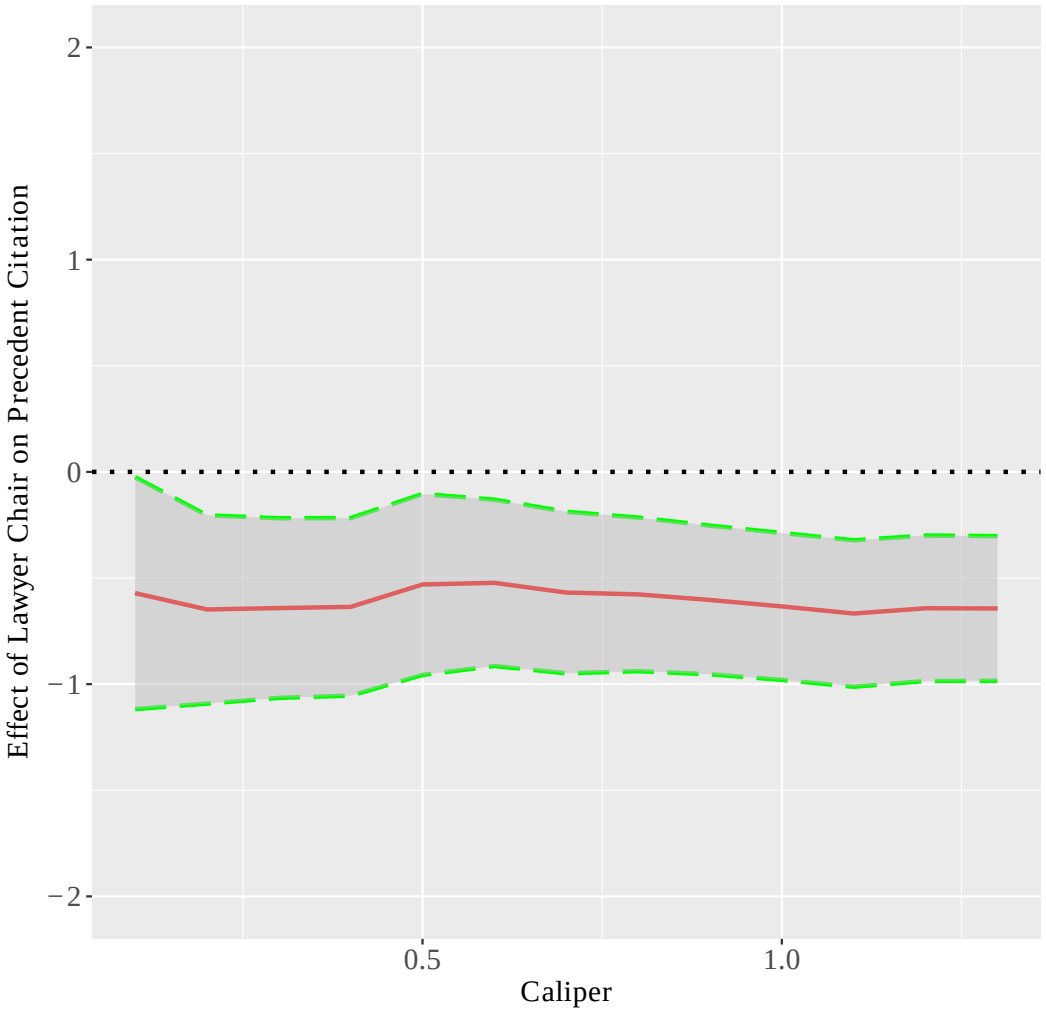
Continued on next page

Table 4.5 – continued from previous page

	Poisson	Beta	¬ Offset	¬ Outliers
Time Fixed Effects	✓	✓	✓	✓
Observations	170	170	170	170
Pseudo-R^2	0.516	0.448	0.754	0.413
Log Likelihood	-2,215	942	−783	−781
Akaike Inf. Crit.	4,509	−1,805	−1,641	1,643
Note:				*p<0.05; **p<0.01; ***p<0.001

Alternative parametric assumptions for the full model. Multiple-imputation with subsequent nearest-neighbor matching and regression analysis. AIC and Log-Likelihood estimates are averages across iterations. Volume in bio$, words in units of 10,000.

Figure 4.4: The Effect of Lawyer Chair Across Calipers



This graph depicts the treatment effect of assigning a lawyer chair for multiple calipers used in the matching procedure.

Table 4.6: Sensitivity Analysis

	Dependent Variable:			
	Precedent Citation			
	Minimal	Panel	Dispute	Party
Lawyer Chair	−0.275*	−0.280*	−0.269*	−0.291*
	(0.113)	(0.114)	(0.109)	(0.114)
CL Chair		0.201		
		(0.123)		
Chair Experience		−0.015		
		(0.021)		
Lawyer-Panelists = 1		0.032		
		(0.175)		
Lawyer-Panelists = 2		−0.064		
		(0.192)		
CL Panelists = 1		−0.035		
		(0.129)		
CL Panelists = 2		−0.094		
		(0.246)		
DG Appointed Panel		−0.090		
		(0.126)		
GATT I			0.233	
			(0.178)	
GATT II			−0.198	
			(0.173)	
GATT III			0.315	
			(0.175)	

Continued on next page

Table 4.6 – continued from previous page

	Minimal	Panel	Dispute	Party
ADA			0.161	
			(0.146)	
Safeguards			0.291	
			(0.246)	
SPS			0.322	
			(0.228)	
SCM			-0.441^{**}	
			(0.156)	
CL Parties				0.056
				(0.087)
DC vs. NDC				−0.001
				(0.123)
# Words (Rest)	-0.025^{**}	-0.023^{*}	-0.034^{***}	-0.025^{**}
	(0.009)	(0.009)	(0.009)	(0.009)
Constant	-6.973^{***}	-7.160^{***}	-7.186^{***}	-7.004^{***}
	(0.291)	(0.354)	(0.302)	(0.300)
Time Fixed Effects	✓	✓	✓	✓
Observations	170	170	170	170
Pseudo-R^2	0.276	0.291	0.344	0.278
Log Likelihood	−809.404	−807.508	−800.409	−809.208
Akaike Inf. Crit.	1,658.807	1,669.015	1,654.818	1,662.416
Note:				*p<0.05; **p<0.01; ***p<0.001

The output of a negative binomial regression. Each model adds a different set of controls (minimum, panel characteristics, dispute characteristics, party characteristics) in order to investigate how the coefficient on Lawyer Chair changes with inclusion. Standard errors in parentheses. Volume in bio$, words in units of 10,000.

Table 4.7: Regression of Reversal Rate on Citation Frequency

	Dependent Variable:	
	Reversal Rate	
	Model (5)	Model (6)
Precedent	-0.210^{*}	-0.437^{**}
	(0.094)	(0.152)
Lawyer Chair	0.114	−0.686
	(0.327)	(0.407)
CL Chair	0.114^{*}	0.006
	(0.057)	(0.075)
Chair Experience	0.544	0.449
	(0.296)	(0.254)
Lawyer-Panelists = 1	−0.397	−0.308
	(0.312)	(0.281)
Lawyer-Panelists = 2	−0.260	0.034
	(0.358)	(0.333)
CL Panelists = 1	−0.058	0.118
	(0.238)	(0.216)
CL Panelists = 2	−1.009	−0.655
	(0.567)	(0.508)
DG Appointed Panel	−0.283	−0.080
	(0.256)	(0.240)
GATT I	0.297	0.220
	(0.321)	(0.282)
GATT II	−0.232	−0.281
	(0.326)	(0.303)

Continued on next page

Table 4.7 – continued from previous page

	Model (5)	Model (6)
GATT III	−0.394	−0.579
	(0.350)	(0.340)
ADA	0.273	0.267
	(0.291)	(0.265)
Safeguards	−0.008	0.028
	(0.430)	(0.382)
SPS	−0.006	0.001
	(0.485)	(0.431)
SCM	−1.042**	−1.192***
	(0.327)	(0.296)
Volume	−0.030	−0.035
	(0.056)	(0.050)
CL Parties	0.671***	0.613***
	(0.175)	(0.155)
DS vs. NDC	−0.221	−0.360
	(0.245)	(0.218)
# Words (Rest)	0.016	0.013
	(0.018)	(0.016)
LC*CLC	−0.693	−0.759
	(0.473)	(0.411)
LC*Experience	−0.202*	0.195
	(0.082)	(0.111)
Precedent*LC		0.639**
		(0.200)

Continued on next page

Table 4.7 – continued from previous page

	Model (5)	Model (6)
Precedent*Experience		0.106^{*}
		(0.048)
Precedent*LC*Experience		-0.277^{***}
		(0.067)
Constant	-0.646	-0.324
	(0.746)	(0.695)
Time Fixed Effects	✓	✓
Observations	108	108
Pseudo-R^2	0.445	0.561
Log Likelihood	-187	-176
Akaike Inf. Crit.	450	435
Note:	*p<0.05; **p<0.01; ***p<0.001	

The output of a negative binomial regression of reversal rates on precedent citation frequencies. Model (5) adds a three-way interaction of citation frequency, the indicator for whether the chair is a lawyer and the chair's experience. Standard errors in parentheses. AIC and Log-Likelihood estimates are averages across iterations. Volume in bio$, words in units of 10,000.

Table 4.8: Summary of Results

	Model (1)	Model (2)	Model (3)	Model (4)
Precedent	$-30\%^{**}$	$-28\%^{**}$	$-28\%^{***}$	$-33\%^{**}$
Reversal	$-8\%^{*}$	$-7\%^{*}$	$-7\%^{*}$	$-9\%^{*}$

$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

The estimated ATT for our analysis as a percent change relative to the citation rate of the control group, as well as estimated reductions in reversal probabilities based on Model (5).

Conclusion

The essays in this dissertation have revisited a number of widely held beliefs about the enforcement of contracts on both the domestic and the international level. Through the use of a series of original data sets, it was shown that many of these beliefs do not withstand empirical scrutiny. Arbitration does not play a dominant role in commercial contracting, sophisticated parties often fail to draft optimal agreements, a promise made as a treaty is substantively different from an executive agreement and non-lawyers may demonstrate a particularly strong adherence to formalism when enforcing trade agreements in an adjudicatory position. Together, the essays suggest that the issue of contract enforcement is often complex and empirical scrutiny is an indispensable part in our understanding of its intricacies.

At the same time, it should be noted that each essay presents a narrowly defined, theoretically motivated inquiry, and neither individual chapters nor the dissertation as a whole can hope to develop a holistic model of how contracts are enforced in practice. This implies that subsequent empirical research is both needed and promising to further our understanding. In this regard, recent advancements in computational methods such as text analysis, web scraping and machine learning offer a unique opportunity to make valuable contributions to the literature and it is with great excitement that I am looking forward to the insights we may gain from the implementation of these methods in the near future.

References

Ackerman, B. and D. Golove (1995). Is NAFTA Constitutional? *Harvard Law Review 108*(4), 799–929.

Aghion, P. and B. Hermalin (1990). Legal Restrictions on Private Contracts Can Enhance Efficiency. *Journal of Law, Economics, & Organization 6*(2), 381–409.

Allison, P. D. (2010). *Survival Analysis using SAS: A Practical Guide* (2 ed.). Cary, North Carolina: Sas Institute.

Anderson, R. and J. Manns (2016). The Inefficient Evolution of Merger Agreements. *(unpublished manuscript)*.

Appleton, A. E. (2016). Judging the Judges or Judging the Members? In L. Choukroune (Ed.), *Judging the State in International Trade and Investment Law*, International Law and the Global South, pp. 11–32. New York City, New York: Springer.

Armour, J., B. Black, and B. Cheffins (2012). Is Delaware Losing Its Cases? *Journal of Empirical Legal Studies 9*(4), 605–656.

Arrow, K. J. and G. Debreu (1954). Existence of an Equilibrium for a Competitive Economy. *Econometrica: Journal of the Econometric Society 22*(3), 265–290.

Badawi, A. B. and S. Baker (2015). Appellate Lawmaking in a Judicial Hierarchy. *The Journal of Law and Economics 58*(1), 139–172.

Baeza-Yates, R. and B. Ribeiro-Neto (1999). *Modern Information Retrieval*, Volume 463. New York City, New York: ACM Press New York.

Basri, C. L. and I. Kagan (2004). *Corporate Legal Departments* (3 ed.). New York City, New York: Practising Law Institute.

Beecher-Monas, E. (1998). Blinded by Science: How Judges Avoid the Science in Scientific Evidence. *Temple Law Review 71*(1), 55–102.

Bennett, S. C. (2002). *Arbitration: Essential Concepts*. New York City, New York: ALM Media, LLC.

Berger, R. (1972). The Presidential Monopoly of Foreign Relations. *Michigan Law Review 71*(1), 1–58.

Berggren, N. and C. Bjørnskov (2013). Does Religiosity Promote Property Rights and the Rule of Law? *Journal of Institutional Economics 9*(2), 161–185.

Bühring-Uhle, C., L. Kirchhoff, and G. Scherer (2006). *Arbitration and Mediation in International Business*. Alphen aan den Rijn, Netherlands: Kluwer Law International.

Bird, S., E. Loper, and E. Klein (2009). *Natural Language Processing with Python*. Sebastopol, California: O'Reilly Media Inc.

Bolton, P. and A. Faure-Grimaud (2010). Satisficing Contracts. *The Review of Economic Studies 77*(3), 937–971.

Borchard, E. (1942). Book Review: International Executive Agreements: Democratic Procedure Under the Constitution of the United States. *Columbia Law Review 42*(5), 887–891.

Borchard, E. (1944). Shall the Executive Agreement Replace the Treaty? *The Yale Law Journal 53*(4), 664–683.

Borchard, E. (1945). The Proposed Constitutional Amendment on Treaty-Making. *The American Journal of International Law 39*(3), 537–541.

Born, G. (2014). *International Commercial Arbitration: International and USA Spectscommentary and Materials* (2 ed.). Alphen aan den Rijn, Netherlands: Kluwer Law International.

Borucka, J. (2014). Methods of Handling Tied Events in the Cox Proportional Hazard Model. *Studia Oeconomica Posnaniensia 2*(2), 91–106.

Box, G. E., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day.

Box-Steffensmeier, J. M. and B. S. Jones (2004). *Event History Modeling: A Guide for Social Scientists*. Cambridge, UK: Cambridge University Press.

Bradley, C. A. (2015, March). *International Law in the U.S. Legal System*. Oxford, UK: Oxford University Press.

Bradley, C. A. and T. W. Morrison (2012). Historical Gloss and the Separation of Powers. *Harvard Law Review 126*(2), 411–485.

Brodersen, K. H., F. Gallusser, J. Koehler, N. Remy, and S. L. Scott (2015). Inferring Causal Impact Using Bayesian Structural Time-Series Models. *The Annals of Applied Statistics 9*(1), 247–274.

Busch, M. L. and K. J. Pelc (2009). Does the WTO Need a Permanent Body of Panelists? *Journal of International Economic Law 12*(3), 579–594.

Buys, C. G. (2003). The Tensions between Confidentiality and Transparency in International Arbitration. *American Review of International Arbitration 14*, 121–138.

Cain, M. D. and S. M. Davidoff (2012). Delaware's Competitive Reach. *Journal of Empirical Legal Studies 9*(1), 92–128.

Carrier, M. A. (1994). When Is the Senate in Recess for Purposes of the Recess Appointments Clause? *Michigan Law Review 92*(7), 2204–2247.

Casella, A. (1996). On Market Integration and the Development of Institutions: The Case of International Commercial Arbitration. *European Economic Review 40*(1), 155–186.

Chalita, L. V. A. S., E. A. Colosimo, and C. G. B. Demétrio (2002). Likelihood Approximations and Discrete Models for Tied Survival Data. *Communications in Statistics-Theory and Methods 31*(7), 1215–1229.

Chayes, A. and A. H. Chayes (1985). Corporate Counsel and the Elite Law Firm. *Stanford Law Review 37*(2), 277–300.

Chemerinsky, E. (2001). Getting Beyond Formalism in Constitutional Law: Constitutional Theory Matters. *Oklahoma Law Review 54*(1), 1–16.

Chiappori, P.-A., S. Oreffice, and C. Quintana-Domeque (2012). Fatter Attraction: Anthropometric and Socioeconomic Matching on the Marriage Market. *Journal of Political Economy 120*(4), 659–695.

Chilton, A. S. and M. Versteeg (2016). Do Constitutional Rights Make a Difference? *American Journal of Political Science 60*(3), 575–589.

Choi, S. and G. Gulati (2008). Bias in Judicial Citations: A Window into the Behavior of Judges? *The Journal of Legal Studies 37*(1), 87–130.

Choi, S. J., M. Gulati, and E. A. Posner (2013). The Dynamics of Contract Evolution. *New York University Law Review 88*(1), 1–50.

Coates, J. C. I. (2012). Managing Disputes through Contract: Evidence from M&A. *Harvard Business Law Review 2*(2), 295–344.

Cohen, J. M. (2002). *Inside Appellate Courts: The Impact of Court Organization on Judicial Decision Making in the United States Courts of Appeals*. Ann Arbor, Michigan: University of Michigan Press.

Cooper, P. J. (2005). George W. Bush, Edgar Allan Poe, and the Use and Abuse of Presidential Signing Statements. *Presidnetial Studies Quarterly 35*(3), 515–532.

Cooter, R. D. and D. L. Rubinfeld (1994). An Economic Model of Legal Discovery. *The Journal of Legal Studies 23*(S1), 435–463.

Corley, P. C. (2006). Avoiding Advice and Consent: Recess Appointments and Presidential Power. *Presidential Studies Quarterly 36*(4), 670–680.

Corley, P. C., R. M. Howard, and D. C. Nixon (2005). The Supreme Court and Opinion Content: The Use of the Federalist Papers. *Political Research Quarterly 58*(2), 329–340.

Cox, D. R. (1972). Regression Models and Life Tables. *Journal of the Royal Statistical Society 34*(2), 187–220.

Craig, W. L. (2010). The Arbitrator's Mission and the Application of Law in International Commercial Arbitration. *American Review of International Arbitration 21*, 243–293.

Crenson, M. A. and B. Ginsberg (2007). *Presidential Power: Unchecked and Unbalanced*. New York City, New York: WW Norton & Company.

Dahl, R. A. (1957). Decision-Making in a Democracy: The Supreme Court as a National Policy-Maker. *Journal of Public Law 6*(2), 279–295.

Dammann, J. (2016). Business Courts and Firm Performance. *(unpublished manuscript)*. available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2889898.

Dammann, J. and H. Hansmann (2009). Globalizing Commercial Litigation. *Cornell Law Review 94*(1), 1–72.

Davey, W. J. (2002). A Permanent Panel Body for WTO Dispute Settlement: Desirable or Practical. In D. L. M. Kennedy and J. D. Southwick (Eds.), *The Political Economy of International Trade Law: Essays in Honor of Robert E. Hudec*. Cambridge, UK: Cambridge University Press.

DeMott, D. A. (2005). The Discrete Roles of General Counsel. *Fordham Law Review 74*(3), 955–982.

Dodge, J. (2011). The Limits of Procedural Private Ordering. *Virginia Law Review 97*(4), 723–799.

Downs, G. W. and M. A. Jones (2002). Reputation, Compliance, and International Law. *The Journal of Legal Studies 31*(S1), S95–S114.

Downs, G. W., D. M. Rocke, and P. N. Barsoom (1996). Is the Good News about Compliance Good News about Cooperation? *International Organization 50*(3), 379–406.

Drahozal, C. R. (1999). Privatizing Civil Justice: Commercial Arbitration and the Civil Justice System. *Kansas Journal of Law & Public Policy 9*(3), 578–591.

Drahozal, C. R. (2016). Empirical Findings on International Arbitration: An Overview. *(unpublished manuscript)*. available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2888552.

Drahozal, C. R. and K. N. Hylton (2003). The Economics of Litigation and Arbitration: An Application to Franchise Contracts. *The Journal of Legal Studies 32*(2), 549–584.

Drahozal, C. R. and R. W. Naimark (2005). *Towards a Science of International Arbitration*. Alphen aan den Rijn, Netherlands: Kluwer Law International.

Drahozal, C. R. and S. J. Ware (2010). Why Do Businesses Use (or Not Use) Arbitration Clauses? *Ohio State Journal on Dispute Resolution 25*(2), 433–476.

Dyreng, S. D. and B. P. Lindsey (2009). Using Financial Accounting Data to Examine the Effect of Foreign Operations Located in Tax Havens and Other Countries on U. S. Multinational Firms' Tax Rates. *Journal of Accounting Research 47*(5), 1283–1316.

Efron, B. (1977). The Efficiency of Cox's Likelihood Function for Censored Data. *Journal of the American Statistical Association 72*(359), 557–565.

Eisenberg, T. and G. P. Miller (2007). Flight from Arbitration: An Empirical Study of Ex Ante Arbitration Clauses in the Contracts of Publicly Held Companies, The. *DePaul Law Review 56*(2), 335–374.

Elkins, Z., A. T. Guzman, and B. A. Simmons (2006). Competing for Capital: The Diffusion of Bilateral Investment Treaties, 1960-2000. *International Organization 60*(4), 811–846.

Elsig, M. and M. A. Pollack (2014). Agents, Trustees, and International Courts: The Politics of Judicial Appointment at the World Trade Organization. *European Journal of International Relations 20*(2), 391–415.

Fach Gomez, K. (2016). The US-EU Transatlantic Trade and Investment Partnership: Should It Leave a Door Open for Non-Legal Arbitrators? *Conflict Resolution Quarterly 34*(2), 189–223.

Farber, H. S. and M. Bazerman (1984). The General Basis of Arbitrator Behavior: An Empirical Analysis of Conventional and Final-Offer Arbitration. Working Paper 1488, National Bureau of Economic Research.

Fich, E. M. and A. Shivdasani (2007). Financial Fraud, Director Reputation, and Shareholder Wealth. *Journal of Financial Economics 86*(2), 306–336.

Fisch, J. E. (1999). The Peculiar Role of the Delaware Courts in the Competition for Corporate Charters. *University of Cincinnati Law Review 68*(4), 1061–1100.

Fisher, R. D. and R. S. Haydock (1995). International Commercial Disputes Drafting an Enforceable Arbitration Agreement. *William Mitchell Law Review 21*, 941–988.

Franck, S. (2006). The Role of International Arbitrators. *ILSA Journal of International & Comparative Law 12*(2), 499–522.

Frankowska, M. (1987). The Vienna Convention on the Law of Treaties before United States Courts. *Virginia Journal of International Law 28*(2), 281–392.

Friedland, P. and L. Mistelis (2010). International Arbitration Survey: Choices in International Arbitration. Survey, Queen Mary University of London and White & Case LLP.

Friedland, P. and L. Mistelis (2015). International Arbitration Survey: Improvements and Innovations in International Arbitration. Survey, Queen Mary University of London and White & Case LLP.

Friedman, B., A. D. Martin, T. Clark, M. Lemos, and A. Larsen. *Judicial Decision Making*.

Galbraith, J. (2017). From Treaties to International Commitments: The Changing Landscape of Foreign Relations Law. *University of Chicago Law Review 84*, 1675–1745.

Garth, B. G. and J. Martin (1993). Law Schools and the Construction of Competence. *Journal of Legal Education 43*(4), 469–509.

Gatowski, S. I., S. A. Dobbin, J. T. Richardson, G. P. Ginsburg, M. L. Merlino, and V. Dahir (2001). Asking the Gatekeepers: A National Survey of Judges on Judging Expert Evidence in a Post-Daubert World. *Law and Human Behavior 25*(5), 433–458.

Gilson, R. J. (1984). Value Creation by Business Lawyers: Legal Skills and Asset Pricing. *Yale Law Journal 94*(2), 239–314.

Gilson, R. J., C. F. Sabel, and R. E. Scott (2014). Text and Context: Contract Interpretation as Contract Design. *Cornell Law Review 100*(1), 23–98.

Ginsburg, T. (2005). International Substitutes for Domestic Institutions: Bilateral Investment Treaties and Governance. *International Review of Law and Economics 25*(1), 107–123.

Goldstein, J. and L. L. Martin (2000). Legalization, Trade Liberalization, and Domestic Politics: A Cautionary Note. *International Organization 54*(3), 603–632.

Grambsch, P. M. and T. M. Therneau (1994). Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. *Biometrika 81*(3), 515–526.

Gruner, D. M. (2003). Accounting for the Public Interest in International Arbitration: The Need for Procedural and Structural Reform. *Columbia Journal of Transnational Law 41*(3), 923–964.

Gulati, M. and R. E. Scott (2012). *The Three and a Half Minute Transaction: Boilerplate and the Limits of Contract Design*. Chicago, Illinois: University of Chicago Press.

Guzman, A. T. (2008a). *How International Law Works: A Rational Choice Theory*. Oxford, UK: Oxford University Press.

Guzman, A. T. (2008b). International Tribunals: A Rational Choice Analysis. *University of Pennsylvania Law Review 157*(1), 171–235.

Hart, O. D. and B. Holmstrom (1986). *The Theory of Contracts*. Cambridge, Massachusetts: MIT Press.

Hathaway, O. A. (2008). Treaties' End: The Past, Present, and Future of International Lawmaking in the United States. *The Yale Law Journal 117*(7), 1236–1372.

Hausman, J., B. H. Hall, and Z. Griliches (1984). Econometric Models for Count Data with an Application to the Patents-R & D Relationship. *Econometrica 52*(4), 909–938.

Hertz-Picciotto, I. and B. Rockhill (1997). Validity and Efficiency of Approximation Methods for Tied Survival Times in Cox Regression. *Biometrics 53*(3), 1151–1156.

Hill, J. (2004). Reducing Bias in Treatment Effect Estimation in Observational Studies Suffering from Missing Data. Institute for Social and Economic Research and Policy Working Paper 04-01.

Ho, D. E., K. Imai, G. King, and E. A. Stuart (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis 15*(3), 199–236.

Ho, D. E. and K. M. Quinn (2010). How Not to Lie with Judicial Votes: Misconceptions, Measurement, and Models. *California Law Review 98*(3), 813–876.

Hoekman, B., H. Horn, and P. C. Mavroidis (2008). *Winners and Losers in the Panel Stage of the WTO Dispute Settlement System*. Working Paper, Research Institute of Industrial Economics.

Honaker, J. and G. King (2010). What to do about Missing Values in Time-Series Cross-section Data. *American Journal of Political Science 54*(2), 561–581.

Honaker, J., G. King, M. Blackwell, and others (2011). Amelia II: a Program for Missing Data. *Journal of Statistical Software 45*(7), 1–47.

Howell, W. G. (2015). *Power without Persuasion*. Princeton, New Jersey: Princeton University Press.

Hudec, R. E. (1970). The GATT Legal System: A Diplomat's Jurisprudence. *Journal of World Trade Law 4*(5), 615–666.

Hudec, R. E. (1999). The New WTO Dispute Settlement Procedure: An Overview of the First Three Years. *Minnesota Journal of Global Trade 8*(1), 1–54.

Hume, R. J. (2006). The Use of Rhetorical Sources by the U.S. Supreme Court. *Law & Society Review 40*(4), 817–844.

Hurlock, J. B., J. Kaye, J. T. McLaughlin, and S. Younger (2011). Task Force on New York Law in International Matters. Survey, New York State Bar Association.

Hylton, K. N. (2000). Agreements to Waive or to Arbitrate Legal Claims: An Economic Analysis. *The Supreme Court Economic Review 8*, 209–263.

Jones, D. and H. Lloyd (2011). Techniques for Controlling Time and Costs in Arbitration. Survey 843, Chartered Institute of Arbitrators.

Kahan, M. (1995). Anti-Dilution Provisions in Convertible Securities Features. *Stanford Journal of Law, Business & Finance 2*(1), 147–166.

Kahan, M. and M. Klausner (1997). Standardization and Innovation in Corporate Contracting (or" The Economics of Boilerplate"). *Virginia Law Review 83*(4), 713–770.

Kakalik, J. S. and R. L. Ross (1983). Costs of the Civil Justice System. Technical report, RAND Corporation, Santa Monica, California.

Kalbfleisch, J. D. and R. L. Prentice (1973). Marginal Likelihoods Based on Cox's Regression and Life Model. *Biometrika 60*(2), 267–278.

Kalbfleisch, J. D. and R. L. Prentice (2002). *The Statistical Analysis of Failure Time Data* (2 ed.). Hoboken, New Jersey: Wiley-Interscience Publication.

Kaufmann, D., A. Kraay, and M. Mastruzzi (2009). Governance Matters VIII: Aggregate and Individual Governance Indicators, 1996-2008. *World Bank Policy Research Paper* (4978).

Kennedy, J. C. (1961). *Corporate Trust Administration and Management* (2 ed.). New York City, New York: New York University Press.

Kim, J.-Y. (1996). Cheap Talk and Reputation in Repeated Pretrial Negotiation. *The RAND Journal of Economics 27*(4), 787–802.

Klein, W. A., C. D. Anderson, and K. G. McGuinness (1993). The Call Provision of Corporate Bonds: A Standard Form in Need of Change. *Journal of Corporation Law 18*(4), 653–696.

Knull III, W. H. and N. D. Rubins (2000). Betting the Farm on International Arbitration: Is It Time to Offer an Appeal Option? *American Review of International Arbitration 11*, 531–607.

Koh, H. H. (2012). Treaties and Agreements as part of Twenty-first Century International Lawmaking. In C. D. Guymon (Ed.), *Digest of United States Practice in International Law*, pp. 91–94. Washington, D.C.: The Office of the Legal Adviser, United States Department of State.

Koh, H. H. (2017). Triptych's End: A Better Framework To Evaluate 21st Century International Lawmaking. *The Yale Law Journal Forum 126*, 338–347.

Koremenos, B. (2005). Contracting Around International Uncertainty. *American Political Science Review 99*(4), 549–565.

Koremenos, B. (2016). *The Continent of International Law: Explaining Agreement Design*. Cambridge, UK: Cambridge University Press.

Kucik, J. and K. J. Pelc (2015). Measuring the Cost of Privacy: a Look at the Distributional Effects of Private Bargaining. *British Journal of Political Science 46*(4), 861–889.

Lacetera, N., D. G. Pope, and J. R. Sydnor (2012). Heuristic Thinking and Limited Attention in the Car Market. *The American Economic Review 102*(5), 2206–2236.

Lagerberg, G. and L. Mistelis (2008). International Arbitration: Corporate Attitudes and Practices. Survey, Queen Mary University of London and PricewaterhouseCoopers.

Landes, W. M. and R. Posner (1979). Adjudication as a Private Good. *Journal of Legal Studies 8*(2), 235–284.

Landes, W. M. and R. A. Posner (1976). Legal Precedent: a Theoretical and Empirical Analysis. *The Journal of Law and Economics 19*(2), 249–307.

Leiter, B. (1999). Positivism, Formalism, Realism. *Columbia Law Review 99*(4), 1138–1164.

Leiter, B. (2005). American Legal Realism. In *The Blackwell Guide to the Philosophy of Law and Legal Theory*, pp. 50–66. Hoboken, New Jersey: Blackwell Publoishing.

Levenshtein, V. I. (1966). Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. *Soviet Physics–Doklady 10*(8), 707–710.

Lew, J. D. (1982). The Case for the Publication of Arbitration Awards. In A. J. van den Berg and J. C. Schultsz (Eds.), *The Art of Arbitration: Essays on International Arbitration Liber Amicorum*. New York City, New York: Kluwer Law and Taxation.

Licht, A. N., C. Goldschmidt, and S. H. Schwartz (2007). Culture Rules: The Foundations of the Rule of Law and other Norms of Governance. *Journal of Comparative Economics 35*(4), 659–688.

Lin, D. Y., L.-J. Wei, and Z. Ying (1993). Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals. *Biometrika 80*(3), 557–572.

Linzer, D. A. and J. K. Staton (2015). A Global Measure of Judicial Independence, 1948–2012. *Journal of Law and Courts 3*(2), 223–256.

Lodhi, H., C. Saunders, J. Shawe-Taylor, N. Cristianini, and C. Watkins (2002). Text Classification Using String Kernels. *Journal of Machine Learning Research 2*, 419–444.

Lupu, Y. and J. H. Fowler (2013). Strategic Citations to Precedent on the U.S. Supreme Court. *The Journal of Legal Studies 42*(1), 151–186.

Lupu, Y. and E. Voeten (2012). Precedent in International Courts: A Network Analysis of Case Citations by the European Court of Human Rights. *British Journal of Political Science 42*(2), 413–439.

Manning, C. D., P. Raghavan, and H. Schütze (2008). *Introduction to Information Retrieval*. Cambridge, UK: Cambridge University Press.

Margolis, L. (1986). *Executive Agreements and Presidential Power in Foreign Policy*. Santa Barbara, California: Praeger Publisher.

Martin, A. D. and K. M. Quinn (2002). Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999. *Political Analysis 10*(2), 134–153.

Martin, L. L. (2000). *Democratic Commitments: Legislatures and International Cooperation*. Princeton, New Jersey: Princeton University Press.

Martin, L. L. (2005). The President and International Commitments: Treaties as Signaling Devices. *Presidential Studies Quarterly 35*(3), 440–465.

Mayer, K. R. (1999). Executive Orders and Presidential Power. *The Journal of Politics 61*(2), 445–466.

McClendon, T. T. (2012). The Power of a Suggestion: The Use of Forum Selection Clauses by Delaware Corporations. *Washington & Lee Law Review 69*(4), 2067–2124.

McClure, W. (1941). *International Executive Agreements: Democtratic Procedure under the Constitution of the United States*. New York City, New York: Columbia University Press.

McCullagh, P. (1980). Regression Models for Ordinal Data. *Journal of the Royal Statistical Society. Series B (Methodological) 42*(2), 109–142.

McDougal, M. S. and A. Lans (1945a). Treaties and Congressional-Executive Agreements or Presidential Agreements: Interchangeable Instruments of National Policy: I. *The Yale Law Journal 54*(2), 181–351.

McDougal, M. S. and A. Lans (1945b). Treaties and Congressional-Executive Agreements or Presidential Agreements: Interchangeable Instruments of National Policy: II. *The Yale Law Journal 54*(3), 534–615.

McLaughlin, C. (1958). The Scope of the Treaty Power in the United States II. *Minnesota Law Review 43*(4), 651–726.

Menon, S. (2014). The Transnational Protection of Private Rights: Issues, Challenges, and Possible Solutions. *Proceedings of the Annual Meeting, published by the American Society of International Law 108*, 219–242.

Micheletti, E. B. and J. Parker (2012). Multi-Jurisdictional Litigation: Who Caused This Problem, and Can It Be Fixed? *Delaware Journal of Corporate Law 37*(1), 1–48.

Miller, G. P. and T. Eisenberg (2009). The Market for Contracts. *Cardozo Law Review 30*(5), 2073–2098.

Mitra, R. and J. P. Reiter (2011). Estimating propensity scores with missing covariate data using general location mixture models. *Statistics in Medicine 30*(6), 627–641.

Mitra, R. and J. P. Reiter (2016). A Comparison of Two Methods of Estimating Propensity Scores After Multiple Imputation. *Statistical Methods in Medical Research 25*(1), 188–204.

Moreno, J. A. (2003). Einstein on the Bench: Exposing What Judges Do Not Know About Science and Using Child Abuse Cases to Improve How Courts Evaluate Scientific Evidence. *Ohio State Law Journal 64*(2), 531–584.

Mosteller, F. and D. L. Wallace (1964). *Applied Bayesian and Classical Inference: The Case of the Federalist Papers*. New York City, New York: Springer Verlag.

Neuborne, B. (1992). Of Sausage Factories and Syllogism Machines: Formalism, Realism, and Exclusionary Selection Techniques. *New York University Law Review 67*(2), 419–450.

Nielsen, R. A. and B. A. Simmons (2015). Rewards for Ratification: Payoffs for Participating in the International Human Rights Regime? *International Studies Quarterly 59*(2), 197–208.

O'Connor, E. O. and C. R. Drahozal (2014). Essential Role of Courts for Supporting Innovation. *Texas Law Review 92*(7), 2177–2210.

Oster, E. (2017). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business and Economic Statistics*, 1–18.

Pauwelyn, J. (2005). The Transformation of World Trade. *Michigan Law Review 104*(1), 1–66.

Pauwelyn, J. (2015). The Rule of Law without the Rule of Lawyers? Why Investment Arbitrators are from Mars, Trade Adjudicators are from Venus. *Americal Journal of International Law 109*(4), 761–805.

Peake, J. S. (2015, April). Executive Agreements as a Foreign Policy Tool During the Bush and Obama Administrations. *unpublished manuscript*.

Pecht, G. (2015). Annual Litigation Trends Survey. Survey, Norton Rose Fulbright.

Pelc, K. J. (2014). The Politics of Precedent in International Law: A Social Network Application. *American Political Science Review 108*(3), 547–564.

Peresie, J. L. (2005). Female Judges Matter: Gender and Collegial Decision-making in the Federal Appellate Courts. *The Yale Law Journal 114*(7), 1759–1790.

Posner, R. A. (1986). Legal Formalism, Legal Realism, and the Interpretation of Statutes and the Constitution. *Case Western Reserve Law Review 37*(2), 179–217.

Posner, R. A. (1993). The Material Basis of Jurisprudence. *Indiana Law Journal 69*(1), 1–38.

Posner, R. A. (2004). The Law and Economics of Contract Interpretation. *Texas Law Review 83*(6), 1581–1614.

Posner, R. A. (2008). *How Judges Think*. Cambridge, Massachusetts: Harvard University Press.

Posner, R. A. (2014). *Economic Analysis of Law* (9 ed.). New York City, New York: Wolters Kluwer Law & Business.

Priest, G. L. and B. Klein (1984). The Selection of Disputes for Litigation. *The Journal of Legal Studies 13*(1), 1–55.

Ribstein, L. E. and E. A. O'Hara (2009). *The Law Market*. Oxford, UK: Oxford University Press.

Richman, B. (2011). Contracts Meet Henry Ford. *Hofstra Law Review 40*(1), 77–86.

Riger, M. (1991). The Trust Indenture as Bargained Contract: The Persistence of Myth. *Journal of Corporation Law 16*(2), 211–248.

Rigobon, R. and D. Rodrik (2004). Rule of Law, Democracy, Openness, and Income: Estimating the Interrelationships. Working Paper, National Bureau of Economic Research.

Rish, I. (2001). An Empirical Study of the Naive Bayes Classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, Volume 3, pp. 41–46. IBM.

Rohde, D. W. and H. J. Spaeth (1976). *Supreme Court Decision Making*. New York City, New York: W. H. Freeman.

Romano, R. (1987). The State Competition Debate in Corporate Law. *Cardozo Law Review 8*(4), 709–758.

Romano, R. and S. Sanga (2017). The Private Ordering Solution to Multiforum Shareholder Litigation. *Journal of Empirical Legal Studies 14*(1), 31–78.

Rovine, A. W. (1974). Digest of United States Practice in International Law. Technical report, Office of the Legal Adviser, Department of State.

Rubin, D. B. (1973). Matching to Remove Bias in Observational Studies. *Biometrics 29*(1), 159–183.

Rubin, D. B. (1988). An Overview of Multiple Imputation. In *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp. 79–84.

Sanga, S. (2014). Choice of law: an empirical analysis. *Journal of Empirical Legal Studies 11*(4), 894–928.

Savola, M. (2016). Interim Measures and Emergency Arbitrator Proceedings. *Croation Arbitration Yearbook 23*, 73–97.

Schwartz, A. and R. E. Scott (2003). Contract Theory and the Limits of Contract Law. *Yale Law Journal 113*(3), 541–619.

Schwartz, B. (2001). Lawyers and the Emerging World Constitution. *Asper Review of International Business and Trade Law 1*, 1–32.

Scott, R. E. and G. G. Triantis (2006). Anticipating Litigation in Contract Design. *The Yale Law Journal 115*(4), 814–879.

Segal, J. A. and A. D. Cover (1989). Ideological Values and the Votes of U.S. Supreme Court Justices. *The American Political Science Review 83*(2), 557–565.

Segal, J. A. and H. J. Spaeth (1993). *The Supreme Court and the Attidunial Model*. Cambridge, Massachusetts: Harvard University Press.

Setear, J. K. (2002, January). The President's Rational Choice of a Treaty's Preratification Pathway: Article II, Congressional-Executive Agreement, or Executive Agreement? *The Journal of Legal Studies 31*(S1), S5–S39.

Shaffer, G., M. Elsig, and S. Puig (2016). The Extensive (but Fragile) Authority of the WTO Appellate Body The Variable Authority of International Courts. *Law and Contemporary Problems 79*(1), 237–274.

Shavell, S. (1995). Alternative Dispute Resolution: An Economic Analysis. *The Journal of Legal Studies 24*(1), 1–28.

Simkin, G. (2005). Litigation Trends Survey. Survey, Fulbright & Jaworski LLP.

Simmons, B. A. (2000). International Law and State Behavior: Commitment and Compliance in International Monetary Affairs. *American Political Science Review 94*(4), 819–835.

Sipes, D. A. and M. E. Oram (1988). *On Trial: The Length of Civil and Criminal Trials*. Williamsburg, Virginia: National Center for State Courts.

Slottje, D. (2006). *Economic Damages in Intellectual Property: A Hands-On Guide to Litigation*. Hoboken, New Jersey: John Wiley & Sons. Google-Books-ID: TbPBw6I5hA8C.

Smith, C. W. and J. B. Warner (1979). On Financial Contracting: An Analysis of Bond Covenants. *Journal of financial economics 7*(2), 117–161.

Solum, L. B. (2006). The Supreme Court in Bondage: Constitutional Stare Decisis, Legal Formalism, and the Future of Unenumerated Rights. *University of Pennsylvania Journal of Constitutional Law 9*(1), 155–208.

Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics 87*(3), 355–374.

Steffensmeier, D. and C. L. Britt (2001). Judges' Race and Judicial Decision Making: Do Black Judges Sentence Differently? *Social Science Quarterly 82*(4), 749–764.

Stephenson, M. C. (2009). Legal Realism for Economists. *Journal of Economic Perspectives 23*(2), 191–211.

Stipanowich, T. (2009). Contract and Conflict Management. *Wisconsin Law Review 2001*(3), 831–918.

Sunstein, C. R., D. Schkade, and L. M. Ellman (2004). Ideological Voting on Federal Courts of Appeals: A Preliminary Investigation. *Virginia Law Review 90*(1), 301–354.

Thies, D. (2010). Rethinking Legal Education in Hard Times: The Recession, Practical Legal Education, and the New Job Market. *Journal of Legal Education 59*(4), 598–622.

Thorner, A. W. (1987). Legal Education in the Recruitment Marketplace: Decades of Change. *Duke Law Journal 1987*(2), 276–291.

Tian, L., D. Zucker, and L. J. Wei (2005). On the Cox Model with Time-Varying Regression Coefficients. *Journal of the American statistical Association 100*(469), 172–183.

Tribe, L. H. (1995). Taking Text and Structure Seriously: Reflections on Free-Form Method in Constitutional Interpretation. *Harvard Law Review 108*(6), 1221–1303.

Van Alstine, M. P. (2011). Treaties in the Supreme Court, 1901–1945. In D. L. Sloss, M. D. Ramsey, and W. S. Dodge (Eds.), *International Law in the U.S. Supreme Court: Continuity and Change*, pp. 191–224. Cambridge, UK: Cambridge University Press.

Van den Berg, A. J. (1981). *The New York Arbitration Convention of 1958: Towards a Uniform Judicial Interpretation*. Alphen aan den Rijn, Netherlands: Kluwer Law International.

Vidmar, N. and S. S. Diamond (2001). Juries and Expert Evidence The Jury in the Twenty-First Century: An Interdisciplinary Conference. *Brooklyn Law Review 66*(4), 1121–1180.

Vittinghoff, E., D. V. Glidden, S. C. Shiboski, and C. E. McCulloch (2012). Survival Analysis. In *Regression Methods in Biostatistics*, Statistics for Biology and Health, pp. 203–259. Boston, Massachusetts: Springer.

Wagner, G. (2014). The Dispute Resolution Market. *Buffalow Law Review 62*(5), 1085–1158.

Wardlaw, K. M. (2010). Umpires, Empathy, and Activism: Lessons from Judge Cardozo. *Notre Dame Law Review 85*(4), 1629–1662.

Ware, S. J. (2013). Is Adjudication a Public Good? "Overcrowded Courts" and the Private Sector Alternative of Arbitration. *Cardozo Journal of Conflict Resolution 14*(3), 899–921.

Weber, A. C., C. A. Pascuzzo S, G. d. S. Pastore, and R. D. Marques (2014). Challenging the "Splitting the Baby" Myth in International Arbitration. *Journal of International Arbitration 31*(6), 719–734.

Weidemaier, W. M. C. (2015). Customized Procedure in Theory and Reality. *Washington & Lee Law Review 72*(4), 1865–1946.

Weiler, J. H. (2002). The Rule of Lawyers and the Ethos of Diplomats: Reflections on the Internal and External Legitimacy of WTO Dispute Settlement. *American Review of International Arbitration 13*, 177–191.

Williamson, O. E. (1983). Credible Commitments: Using Hostages to Support Exchange. *The American Economic Review 73*(4), 519–540.

Wolrich, P. M. (2011). CIArb Costs of International Arbitration Survey. Survey 843, International Chamber of Commerce Commission on Arbitration.

Wright, Q. (1944). The United States and International Agreements. *The American Journal of International Law 38*(3), 341–355.

Yoo, J. (2011). Rational Treaties: Article II, Congressional-Executive Agreements, and International Bargaining. *Cornell L. Rev. 97*(1), 1–44.

Yoo, J. C. (2001, February). Laws as Treaties?: The Constitutionality of Congressional-Executive Agreements. *Michigan Law Review 99*(4), 757–852.

Young, M. K. (1995). Dispute Resolution in the Uruguay Round: Lawyers Triumph over Diplomats. *The International Lawyer 29*(2), 389–409.

Appendix

A.1.1 Detailed Description of Text Analysis Procedures

The textual analysis of the contracts was conducted in Python 2.7, relying to a great extent on the Natural Language Toolkit (NLTK). Most of the techniques used are described in detail in Bird et al. (2009).[47] Due to the large number of contracts and the associated computational intensity, the program was executed on the Savio Institutional Cluster of UC Berkeley's BRC High Performance Computing.

Identification of Parties

In order to identify which agreements are international, I scan each agreement for party names. However, scanning the entire contract for party names is computationally intensive and leads to many false matches, as companies that are not party to the agreement might be mentioned later in the text. I thus first identify the paragraph containing the parties to the agreement.

Virtually all contracts begin by naming the parties and then specifying how the contracts refers to them. The term by which the parties are referenced is specified in quotation marks contained in parentheses. For example, an agreement might begin stating

This purchasing agreement (this "Agreement") is entered into by and between company A and B (together referred to as "the parties").

I use the following regular expression to identify the first paragraph that contains quotation marks encapsulated within parentheses:

$$\backslash(.*\backslash"(.+?)\backslash"$$

I include the first matching paragraph into the list of paragraphs containing party information. In addition, I add the two paragraphs preceding the match and all consecutive paragraphs that also contain quotation marks within parentheses. That is because the party information is sometimes broken up across multiple paragraphs, even though these cases are the rare exception.

I then define a list of 632,442 companies and individuals that have disclosed information through filings with the SEC. These parties are included in lowercase and in different forms to take into account that parties might write company names differently. For example, the algorithm identifies with the company "PT Holdings, Inc." all mentions of "pt holdings, inc.", "pt holdings

[47] The most current version of this book is accessible online at http://www.nltk.org/book/.

inc", "pt holdings incorporated" and "pt holdings". Versions that exclusively include lemmatized words mentioned in a collection of 234,377 words of the English vocabulary are dropped. This is necessary as there are company names such as "Hungary" which lead to many false hits. In total, the final list includes 630,106 companies with their place of incorporation and their economic headquarters.

The program scans the defined paragraphs for the mentioning of these companies. If multiple company names are included in a paragraph but one company name is fully included in another company name, only the longest company name is regarded a party to the contract. This is done because some company names are so generic that they are often included in other company names. For instance, the company "Energy Inc." is fully included in "Hawaii Energy Inc." but is certainly not a party if the company name "Hawaii Energy Inc." is mentioned, so "Energy Inc." is then dropped.

The paragraphs are then scanned for the mentioning of countries in their noun and adjective form. For any given country i, if the list of companies does not yet include a company from country i but i is mentioned in the paragraph, an unidentified company from country i is added to the list of parties.

The program then simply counts the number of companies registered in the U.S. and those registered outside of the U.S. to determine whether the contract is domestic, international or foreign. If information on the place of registration is not available, the location on file with the SEC is used instead.

Identification of Contract Format and Type

In order to identify the contract format, I scan the text for the first mentioning of one of the following words: *agreement, plan, note, policy, guideline, program or contract*. The format of the contract corresponds to the word that appears first. For example, if a contract has the heading "*Purchasing Agreement*", the format will be "Agreement", whereas a document entitled "*Note Exchange*" will be considered a "Note".

In order to identify the contract type, I first define terms that are indicative of the type of contract. The following is a breakdown of agreement types and corresponding terms.

Table A.1.1 : Agreement Types and their Terms

Type	Terms
Consulting	consulting
Employment	employer, employee, employment, severance, non competition, termination, management continuity, transition, appointment
Incentives	pension, stock unit, award, incentive, compensation, management stability, stock option, restricted stock, tax deferred savings, reimbursement, retention, separation allowance, retirement, bonus, dsu, medical plan, benefit, indemnification, health plan, executive plan, savings and investment, stock ownership, restoration plan, performance share, stock retainer, performance plan, management stockholders, indemnity, director stock, directors stock, change in control, change of control
Joint Venture	joint venture
Lease	lease, line access, sublease, tenant, landlord
Legal	settlement, tolling, waiver
Licensing	license, licensing
Loan	credit, loan, subordination, borrow, lender, commitment
M&A	merger, separation and distribution, share exchange, earnout, earn out
Neg. Instrument	promissory
Sales	purchase, sale, purchasing, sell, distribution
Security	security, mortgage, collateral
Transportation	transportation, precedent

Terms which determine agreement types.

I then extract from the contract all text up to the first occurrence of one of the words defining the format. Typically, this results in a string that contains only the title of the agreement, such as "*Employment Agreement*" or "*Licensing Agreement*". In most other cases, the string contains all text up the point where the agreement is defined in the contract. For instance, in the above example where a contract begins with

This purchasing agreement (this "Agreement") is entered into by and between company A and B (together referred to as "the parties").

the matched string would contain the words "*This purchasing agreement*", and possibly a preceding table of contents.

The matched string (in lowercase) is scanned for all the terms listed in Table A.1.1 . If a term is included in the text, an internal "score" of the corresponding contract type is increased by 1. The type of the contract is the type with highest score, though typically, only one of the types receives a score greater than 0.

Identification of Clauses

In order to identify governing law and forum selection clauses, I first preprocess the text. The preprocessing consists of the following steps:

1. Break up text into paragraphs
2. Convert paragraph to lowercase
3. Remove punctuation and special characters
4. Remove stop words
5. Tokenization
6. Stemming

Step 1-3 are self-explanatory. Removing stop words such as "the", "is", "at" and "which" is a common procedure in natural language processing, because stop words are typically not meaningful in determining the content of a text (Lodhi et al., 2002).[48] To define the stop words that are to be removed, I rely on the "stopwords" corpus of NLTK.

Text tokenization is essentially the process of breaking up a string of characters into analyzable pieces. A unit of analysis can be words, word combinations, sentences or entire paragraphs. Here, the goal is to use tokens to identify whether a clause is a forum selection clause. A useful unit of analysis is each word. I thus tokenize each paragraph into words.

Text stemming is the process of removing morphological affixes from words, leaving only the word stem. The idea is that words originating from the same

[48]Note that the removal of stop words should depend on the goal of the analysis. For instance, stop words can be useful in identifying the author of a text, because patterns in the use of stop words can vary strongly and consistently from one author to the next. For instance, stop words have been used to identify the original author of disputed federalist papers (Mosteller and Wallace, 1964).

word stem should be treated the same, as morphological affixes are only the product of grammatical rules and conventions which are disassociated from the actual meaning of the word. Stemming is an algorithm-based process that differs from one language to the other. I rely on the popular Snowball algorithm for the English language, included in NLTK. I complement this stemming algorithm with additional rules useful in text classification. For instance, I do not stem the word "arbitration" into its word stem "arbitr", because the word "arbitration" is less predictive of an arbitration clause than words such as "arbitrator".

The following example illustrates the output of the preprocessing procedure:

Before preprocessing: *This is an arbitration clause between two companies that defines the appointment process, the seat and the location of the arbitral proceeding. It serves as an example.*

After preprocessing: *arbitration claus two compani defin appoint process seat locat arbitr proceed serv exampl*

After preprocessing, I manually define a set of text features indicative of whether a clause is a forum selection clause. In essence, a feature is information about the text. Among others, features can help the researcher predict whether the document is of a relevant class or not. In theory, anything about a token can be a feature, such as information about the first or last letter of the token, the occurrence of a particular word within the token, the last letter of a word in the token or a combination of multiple tokens. In document classification, features should be defined to maximize the accuracy of a document's class. I start by allowing every word in a hand-coded sample of 5,226 paragraphs to be its on feature and create a list of the words most predictive of forum selection clauses. I then complement this list using an initial set of words typically used in forum selection clauses, based on my reading of these clauses. I then again repeatedly test the performance of each word feature, keeping highly predictive features and dropping those that are not predictive. I also add certain combinations of words to the list of features. The final list includes the following words and word combinations:

court, forum, irrevoc, proceed, venu, action, jurisdict, brought, district, inconveni, object, placeholderst, sit, lay, southern, suit, waiv, uncondit, bring, appel, submit, exclus, process, fullest, state, heard, recognit, plead, herebi, appointe, nonexclus, judgment, arbitration, aris, hereaft, borough, convenien, counti, suprem, summon, disput, hereto, law, lack, manhattan, parti, settl,

(jurisdict, submit), (exclus, jurisdict), (jurisdict, disput), (jurisdict, nonexclus), (jurisdict, resolv), (jurisdict, venu), (jurisdict, litig), (jurisdict, controversi), (jurisdict, referr), (jurisdict, suit), (jurisdict, proceed), (jurisdict, forum), (jurisdict, submiss), (arbitr, resolv), (submit, exclus), (submit, court), (compet, jurisdict), (disput, parti), (take, place), (consent, jurisdict), (irrevoc, submit),(unit, state, district, court), (exclus, forum), (person, jurisdict), (irrevoc, uncondit), (govern, law), (trial, juri, waiv), (legal, proceed), (agreement, arbitr), (placeholderst, jurisdict), (placeholderst, court), (disput, resolut), (fullest, extent, permit, law), (inconveni, forum), (aforement, court), (aforesaid, court), (final, judgment), (such, court), (govern, author), (waiv, right), (disput, arbitration), (trial, juri, waiver), (settl, arbitration), (resolv, arbitration), (determin, arbitration)[49]

Using these features, I train two different naive Bayes classifiers, one for court selection clauses and one for arbitration clauses. Both classifiers are supplemented with additional manual rules designed to increase accuracy. For instance, when a known arbitration organization is mentioned in the clause, it will automatically be considered an arbitration clause. Similarly, if the word "*arbitration*" is part of an enumaration of words of at least 3 items, the others of which do not contain a word starting with "*arb*", then the clause is deemed not an arbitration clause. That is because the word "*arbitration*" is often mentioned in a list of legal actions for which a specific consequence is defined in the contract (e.g. "*In the event of any litigation, arbitration, mediation or government action, (...)*").

Choice-of-law clauses are identified using the word count of a clause as well as a set of manual rules based on the occurrence of the following strings in the unstemmed text:

governing law, law governing, shall be governed by, interpret, construe, govern , governed, governing, the laws, the law

[49]The feature *placeholderst* is a place holder included for state names.

A.1.2 Arbitration Clauses Illustrating Cosine Similarity

Clauses with a similarity of 0.6

Clause 1:

15. ARBITRATION. ANY CONTROVERSY OR CLAIM BETWEEN OR AMONG THE PARTIES HERETO INCLUDING BUT NOT LIMITED TO THOSE ARISING OUT OF OR RELATING TO THIS INSTRUMENT, AGREEMENT OR DOCUMENT OR ANY RELATED INSTRUMENTS, AGREEMENTS OR DOCUMENTS, INCLUDING ANY CLAIM BASED ON OR ARISING FROM AN ALLEGED TORT, SHALL BE DETERMINED BY BINDING ARBITRATION IN ACCORDANCE WITH THE FEDERAL ARBITRATION ACT (OR IF NOT APPLICABLE, THE APPLICABLE STATE LAW), THE RULES OF PRACTICE AND PROCEDURE FOR THE ARBITRATION OF COMMERCIAL DISPUTES OF J.A.M.S. ENDISPUTE OR ANY SUCCESSOR THEREOF ("J.A.M.S."), AND THE "SPECIAL RULES" SET FORTH BELOW. IN THE EVENT OF ANY INCONSISTENCY, THE SPECIAL RULES SHALL CONTROL. JUDGMENT UPON ANY ARBITRATION AWARD MAY BE ENTERED IN ANY COURT HAVING JURISDICTION. ANY PARTY TO THIS INSTRUMENT, AGREEMENT OR DOCUMENT MAY BRING AN ACTION, INCLUDING A SUMMARY OR EXPEDITED PROCEEDING, TO COMPEL ARBITRATION OF ANY CONTROVERSY OR CLAIM TO WHICH THIS AGREEMENT APPLIES IN ANY COURT HAVING JURISDICTION OVER SUCH ACTION. A. SPECIAL RULES. THE ARBITRATION SHALL BE CONDUCTED IN THE COUNTY OF ANY BORROWERŚ DOMICILE AT THE TIME OF THE EXECUTION OF THIS INSTRUMENT, AGREEMENT OR DOCUMENT AND ADMINISTERED BY J.A.M.S. WHO WILL APPOINT AN ARBITRATOR; IF J.A.M.S. IS UNABLE OR LEGALLY PRECLUDED FROM ADMINISTERING THE ARBITRATION, THEN THE AMERICAN ARBITRATION ASSOCIATION WILL SERVE. ALL ARBITRATION HEARINGS WILL BE COMMENCED WITHIN 90 DAYS OF THE DEMAND FOR ARBITRATION; FURTHER, THE ARBITRATOR SHALL ONLY, UPON A SHOWING OF CAUSE, BE PERMITTED TO EXTEND THE COMMENCEMENT OF SUCH HEARING FOR UP TO AN ADDITIONAL 60 DAYS. B. RESERVATION OF RIGHTS. NOTHING IN THIS ARBITRATION PROVISION SHALL BE DEEMED TO (I) LIMIT THE APPLICABILITY OF ANY OTHERWISE APPLICABLE STATUTES OF LIMITA-

TION OR REPOSE AND ANY WAIVERS CONTAINED IN THIS INSTRUMENT, AGREEMENT OR DOCUMENT; OR (II) BE A WAIVER BY BANK OF THE PROTECTION AFFORDED TO IT BY 12 U.S.C. SEC. 91 OR ANY SUBSTANTIALLY EQUIVALENT STATE LAW; OR (III) LIMIT THE RIGHT OF BANK HERETO (A) TO EXERCISE SELF HELP REMEDIES SUCH AS (BUT NOT LIMITED TO) SETOFF, OR (B) TO FORECLOSE AGAINST ANY REAL OR PERSONAL PROPERTY COLLATERAL, OR (C) TO OBTAIN FROM A COURT PROVISIONAL OR ANCILLARY REMEDIES SUCH AS (BUT NOT LIMITED TO) INJUNCTIVE RELIEF, WRIT OF POSSESSION OR THE APPOINTMENT OF A RECEIVER. BANK MAY EXERCISE SUCH SELF HELP RIGHTS, FORECLOSE UPON SUCH PROPERTY, OR OBTAIN SUCH PROVISIONAL OR ANCILLARY REMEDIES BEFORE, DURING OR AFTER THE PENDENCY OF ANY ARBITRATION PROCEEDING BROUGHT PURSUANT TO THIS INSTRUMENT, AGREEMENT OR DOCUMENT. NEITHER THIS EXERCISE OF SELF HELP REMEDIES NOR THE INSTITUTION OR MAINTENANCE OF AN ACTION FOR FORECLOSURE OR PROVISIONAL OR ANCILLARY REMEDIES SHALL CONSTITUTE A WAIVER OF THE RIGHT OF ANY PARTY, INCLUDING THE CLAIMANT IN ANY SUCH ACTION, TO ARBITRATE THE MERITS OF THE CONTROVERSY OR CLAIM OCCASIONING RESORT TO SUCH REMEDIES.

Clause 2:

THIS LETTER AGREEMENT SHALL BE GOVERNED BY AND CONSTRUED IN ACCORDANCE WITH THE INTERNAL LAWS OF THE STATE OF NEW YORK WITHOUT GIVING EFFECT TO PRINCIPLES OF CONFLICTS OF LAW. THE UNDERSIGNED AND GLOBAL CROSSING HOLDINGS LTD. EACH WAIVES THE RIGHT TO A JURY TRIAL OR COURT TRIAL. THE SOLE AND EXCLUSIVE METHOD TO RESOLVE ANY CLAIM IS BINDING ARBITRATION UNDER THE RULES OF THE AMERICAN ARBITRATION ASSOCIATION. The parties each waive his/her right to commence an action in any court to resolve any claim arising out of or related to this letter agreement, except for an action for injunctive relief pending resolution of a claim through binding arbitration.

Clauses with a similarity of almost 1 (difference underlined):

Clause 1:

"(A) If a dispute or controversy arises out of or in connection with this Agreement, the parties shall first attempt in good faith to settle the dispute or

controversy by mediation under the Commercial Mediation Rules of the American Arbitration Association before resorting to arbitration or litigation. (...) The Executive shall pay all costs and expenses, including attorneys' fees and disbursements, of the Company and the Executive in connection with any legal proceeding (including arbitration), whether or not instituted by the Company or the Executive, relating to the interpretation or enforcement of any provision of this Agreement, that is resolved in favor of the Company pursuant to a final, unappealable judgment. The non-prevailing party, as set forth above, shall pay prejudgment interest on any money judgment obtained by the prevailing party as a result of such proceeding, calculated at the rate provided in Section 1274(b)(2)(B) of the Code."

Clause 2:

"(A) If a dispute or controversy arises out of or in connection with this Agreement, the parties shall first attempt in good faith to settle the dispute or controversy by mediation under the Commercial Mediation Rules of the American Arbitration Association before resorting to arbitration or litigation. (...) (C) The Company shall pay all costs and expenses, including attorneys' fees and disbursements, of the Company and the Executive in connection with any legal proceeding (including arbitration), whether or not instituted by the Company or the Executive, relating to the interpretation or enforcement of any provision of this Agreement, that is resolved in favor of the Executive pursuant to a final, unappealable judgment. The non-prevailing party, as set forth above, shall pay prejudgment interest on any money judgment obtained by the prevailing party as a result of such proceeding, calculated at the rate provided in Section 1274(b)(2)(B) of the Code."

A.1.3 Published ICC Award No. 10947

Interim Award in Case No. 10947 in 2002 (June)

Insurance Companies, Subrogated Insurers (France) v State-owned Company (Ecuador)

Industry: Not Available

Case Type: International

Award Amount: Unknown

Claimant's Attorney: Not Available

Respondent's Attorney: Not Available

Award Date: June 2002

Arbitrator: Robert Lawson (Chairman); Charles Poncet; Sean Gates

Country: Switzerland

Place: Geneva

Language: English

Source:
Bulletin de l'Association Suisse d'Arbitrage, 2004, pp. 308-332 (ASA Bulletin 2/2004)

Commentary citations:

Cited documents:

Cited Court Decisions
Ecuadorian Supreme Court, 7 February 1994, "Ecuadorian State v. Empresa Eléctrica del Ecuador Inc."
Swiss Federal Tribunal, ATF II 229-233 (1979), "Hafinag AG v. Modernbau Klier and Rabe KG"
Swiss Federal Tribunal, ATF 118 II 353-358 (1992), "Fincantieri"
Swiss Federal Tribunal, ATF 118 II 193-198 (1992), "G. v. V. SpA."
Swiss Federal Tribunal, ATF 127 III 279, "Fomento de Construcciones y Contratas SA v. Colon Containers Terminal SA"

Cited Awards
Award, "Benteler et al. v. Belgian State"
Cited ICC Rules

Art. 4
Art. 6
Art. 35

Cited Legislation
Ecuadorian Constitution, Art. 14 (ex Art. 16)
Ecuadorian Civil Code, Arts. 7, 1505, 1726
Ecuadorian Code of Civil Procedure, Art. 101
Ecuadorian Law on Arbitration and Mediation, September 1997
Swiss Federal Law on Private International Law, 1987, Arts. 9-177.2

Cited Treaties
Inter-American Convention on International Commercial Arbitration, 1975 (Panama Convention)
Inter-American Convention on Extraterritorial Validity of Foreign Judgments and Arbitral Awards
New York Convention on the Recognition and Enforcement of Foreign Arbitral Awards, 10 June 1958

ICC Award No. 10947

Interim Award in Case No. 10947 in 2002 (June), ICC Award No. 10947

A.1.4 Names and Abbreviations of Arbitration Organizations

Table A.1.2 : Arbitration Organizations

Abbreviation	Full Name
AAA	American Arbitration Association
JAMS	Judicial Arbitration and Mediation Services
ICC	International Chamber of Commerce
CIETAC	China International Economic and Trade Arbitration Commission
LCIA	London Court of International Arbitration
HKIAC	Hong Kong International Arbitration Centre
SIAC	Singapore International Arbitration Centre
SCC	Stockholm Chamber of Commerce

Abbreviations and corresponding names of arbitration organizations used throughout the article.

A.1.5 Detailed Regression Results for Arbitration and Court Clauses

Table A.1.3 : Logit-Regression on Arbitration Clause Usage

	Dependent variable:			
	Arbitration Clause			
	(1)	(2)	(3)	(4)
International	0.297***	0.317***	0.337***	0.345***
	(0.011)	(0.012)	(0.046)	(0.052)
Year				−0.005***
				(0.001)
Year*International				−0.001
				(0.003)
	Agreement Type			
Employment		0.993***	1.018***	1.019***
		(0.011)	(0.011)	(0.011)
Joint Venture		1.202***	0.898***	0.899***
		(0.056)	(0.075)	(0.075)
Lease		0.651***	0.658***	0.659***
		(0.020)	(0.021)	(0.021)
Legal		−0.498***	−0.466***	−0.465***
		(0.033)	(0.036)	(0.036)
Licensing		1.031***	0.977***	0.979***
		(0.024)	(0.027)	(0.027)
Loan		−0.662***	−0.577***	−0.577***

Continued on next page

Table A.1.3 – continued from previous page

	(1)	(2)	(3)	(4)
		(0.017)	(0.019)	(0.019)
M&A		0.056**	0.078***	0.082***
		(0.021)	(0.022)	(0.022)
Neg. Instrument		−0.533***	−0.481***	−0.482***
		(0.047)	(0.050)	(0.050)
Sales		0.283***	0.216***	0.217***
		(0.013)	(0.013)	(0.013)
Security		−0.767***	−0.761***	−0.761***
		(0.026)	(0.029)	(0.029)
Transportation		0.445***	0.285***	0.283***
		(0.067)	(0.077)	(0.077)
Other		−0.619***	−0.602***	−0.599***
		(0.032)	(0.034)	(0.034)
		Industry		
Agriculture		0.424	0.320	0.338
		(0.593)	(0.656)	(0.656)
Construction		−0.127**	−0.143***	−0.144***
		(0.041)	(0.043)	(0.043)
Finance		0.016	0.032**	0.031**
		(0.011)	(0.011)	(0.011)
Mining		−0.058***	−0.019	−0.022
		(0.017)	(0.018)	(0.018)
Services		0.059***	0.072***	0.072***

Continued on next page

Table A.1.3 – continued from previous page

	(1)	(2)	(3)	(4)
		(0.010)	(0.011)	(0.011)
Trade		−0.037*	−0.027	−0.027
		(0.015)	(0.016)	(0.016)
Transportation		−0.029*	−0.009	−0.010
		(0.014)	(0.015)	(0.015)
Other		−0.213***	−0.203***	−0.202***
		(0.029)	(0.031)	(0.031)
		Document Format		
Contract		0.044*	0.014	0.008
		(0.022)	(0.024)	(0.024)
Guideline		−1.865***	−1.821***	−1.825***
		(0.211)	(0.211)	(0.211)
Note		−1.029***	−1.069***	−1.068***
		(0.029)	(0.032)	(0.032)
Plan		−1.337***	−1.378***	−1.378***
		(0.015)	(0.016)	(0.016)
Policy		−1.310***	−1.351***	−1.351***
		(0.062)	(0.063)	(0.063)
Program		−1.543***	−1.568***	−1.568***
		(0.049)	(0.050)	(0.050)
Other		−5.370***	−5.346***	−5.349***
		(0.377)	(0.378)	(0.378)

Continued on next page

Table A.1.3 – continued from previous page

	(1)	(2)	(3)	(4)
Constant	-1.318^{***}	-1.767^{***}	-1.786^{***}	5.281^{**}
	(0.022)	(0.026)	(0.026)	(1.752)
Time-Fixed Effects	✓	✓	✓	
Interactions			✓	✓
Observations	504,119	504,119	504,119	504,119
Log Likelihood	−244,467	−220,155	−219,517	−219,550
Akaike Inf. Crit.	488,970	440,401	439,181	439,219
Note:				*p<0.05; **p<0.01; ***p<0.001

The table depicts the estimates for a logit regression of a dummy indicating whether a contract includes an arbitration clause on a dummy indicating whether a contract is an international contract. Standard errors in parentheses. Model (1) includes year-fixed effects. Model (2) additionally controls for type, industry and form of the agreement. Model (3) includes interaction effects between the dummy for international contracts and the type of agreement, as well as the industry. Model (4) imposes a linear time trend and interacts it with the dummy for international contracts. Other interaction effects are omitted to increase readability. The reference categories for categorical variables are the most prevalent categories. For type, that is Incentives; for industry, it is Manufacturing; for format, it is agreement.

Table A.1.4 : Logit-Regression on Court Clause Usage

	Dependent variable:			
	Court Clause			
	(1)	(2)	(3)	(4)
International	0.696***	0.220***	0.592***	0.818***
	(0.010)	(0.010)	(0.042)	(0.047)
Year				0.060***
				(0.001)
Year*International				−0.026***
				(0.002)
	Agreement Type			
Consulting		0.328***	0.433***	0.430***
		(0.028)	(0.029)	(0.029)
Employment		0.418***	0.432***	0.432***
		(0.012)	(0.012)	(0.012)
Joint Venture		0.405***	0.730***	0.732***
		(0.063)	(0.078)	(0.078)
Lease		0.336***	0.357***	0.358***
		(0.022)	(0.023)	(0.023)
Legal		0.655***	0.703***	0.703***
		(0.024)	(0.026)	(0.026)
Licensing		0.667***	0.761***	0.762***
		(0.025)	(0.028)	(0.028)

Continued on next page

Table A.1.4 – continued from previous page

	(1)	(2)	(3)	(4)
Loan		1.643***	1.627***	1.628***
		(0.013)	(0.014)	(0.014)
M&A		1.789***	1.824***	1.828***
		(0.018)	(0.019)	(0.019)
Neg. Instrument		1.043***	1.051***	1.047***
		(0.025)	(0.026)	(0.026)
Sales		1.258***	1.291***	1.292***
		(0.012)	(0.012)	(0.012)
Security		1.285***	1.305***	1.306***
		(0.017)	(0.018)	(0.018)
Transportation		0.746***	0.694***	0.692***
		(0.063)	(0.071)	(0.071)
Other		1.235***	1.289***	1.292***
		(0.021)	(0.022)	(0.022)
		Industry		
Agriculture		−0.047	−1.010	−1.006
		(0.684)	(1.067)	(1.066)
Construction		−0.074*	−0.051	−0.051
		(0.034)	(0.035)	(0.035)
Finance		−0.177***	−0.185***	−0.186***
		(0.010)	(0.010)	(0.010)
Mining		−0.117***	−0.114***	−0.118***
		(0.014)	(0.016)	(0.016)

Continued on next page

Table A.1.4 – continued from previous page

	(1)	(2)	(3)	(4)
Services		0.066***	0.071***	0.073***
		(0.009)	(0.010)	(0.010)
Trade		0.026*	0.032*	0.034*
		(0.013)	(0.014)	(0.014)
Transportation		0.053***	0.029*	0.029*
		(0.013)	(0.014)	(0.014)
Other		−0.038	−0.069**	−0.059*
		(0.024)	(0.026)	(0.026)
Document Format				
Contract		−0.907***	−0.867***	−0.862***
		(0.024)	(0.026)	(0.026)
Guideline		−2.237***	−2.425***	−2.425***
		(0.231)	(0.264)	(0.264)
Note		−0.203***	−0.202***	−0.203***
		(0.016)	(0.017)	(0.017)
Plan		−1.405***	−1.448***	−1.450***
		(0.015)	(0.015)	(0.015)
Policy		−2.200***	−2.273***	−2.276***
		(0.089)	(0.094)	(0.094)
Program		−1.792***	−1.842***	−1.846***
		(0.051)	(0.053)	(0.053)
Other		−3.636***	−3.651***	−3.652***
		(0.129)	(0.133)	(0.133)

Continued on next page

Table A.1.4 – continued from previous page

	(1)	(2)	(3)	(4)
Constant	-1.126^{***}	-1.901^{***}	-1.916^{***}	-1.939^{***}
	(0.021)	(0.024)	(0.024)	(0.013)
Time-Fixed Effects	✓	✓	✓	
Interactions			✓	✓
Observations	504,119	504,119	504,119	504,119
Log Likelihood	−304,258	−270,089	−269,666	−219,550
Akaike Inf. Crit.	608,552	540,270	539,479	439,219
Note:				*p<0.05; **p<0.01; ***p<0.001

The table depicts the estimates for a logit regression of a dummy indicating whether a contract includes an arbitration clause on a dummy indicating whether a contract is an international contract. Standard errors in parentheses. Model (1) includes year-fixed effects. Model (2) additionally controls for type, industry and form of the agreement. Model (3) includes interaction effects between the dummy for international contracts and the type of agreement, as well as the industry. Model (4) imposes a linear time trend and interacts it with the dummy for international contracts. Other interaction effects are omitted to increase readability. The reference categories for categorical variables are the most prevalent categories. For type, that is Incentives; for industry, it is Manufacturing; for format, it is agreement.

A.2.1 Formalizing Incentive Costs

The following formalizes the theory of incentive costs introduced by leaving the forum unspecified.

‘’ Assume a contract between potential plaintiff P and potential defendant D. D is contemplating whether to breach the contractual terms, harming P for an amount of v, or whether to incur forbearance costs γ and comply. It is socially optimal for D to breach if

$$\gamma > v$$

Without legal recourse, D will not invest in forbearance. If P has the option to seek legal recourse, she can sue D for damages in the amount of d, producing litigation costs of $\delta_P + \delta_D = \delta$. Now consider the possibility for parties to specify a court forum *ex ante*, where the forum is denoted $x \in X$. Different forums have different dispute settlement costs and award different damages. Incorporating the possibility of forum selection, we can state that D will breach if

$$\gamma > d_x + \delta_x$$

In order to maximize the joint surplus of the contract, D has to be incentivized to breach only when it is efficient. This condition is satisfied if

$$v = d_x + \delta_x$$

Assuming for simplicity that overdeterrence is as harmful as underdeterrence, parties thus maximize their joint utility and overall welfare if they choose from the set of forums that minimize the difference between harm and the sum of damages and litigation costs, formally

$$X^* \equiv \{x | x \in X, \operatorname{argmin}_x f(x) = d_x + \delta_x - v\}$$

What about parties that do not choose a forum? They leave the choice where to sue up to P, who can choose from all forums that accept jurisdiction under the default rule. Since P has an incentive to maximize her own as opposed to the joint surplus of the contract, there is no guarantee that the forum chosen by P is the one closest to the social optimum. In particular, let X' be the set of forums that P can sue in. Under the default rule, at the minimum this includes D's state of incorporation and economic headquarters. P will choose to sue in the state in which the difference between her damage award and her litigation costs is maximized. This set is defined by

$$X'^* \equiv \{x | x \in X', \operatorname{argmax}_x f(x) = d_x - \delta_{x,P}\}$$

A.2.2 Demonstrating Heterogeneity in Predictability

Here, I demonstrate that jurisdictions are heterogeneous in their predictability. In order to do so, I construct a measure of the relative use of federal courts by state. This measure is an analogue to the measure of relative use of law proposed by Sanga (2014). In particular, let L denote the forum in which parties litigate and let C denote the forum in which parties agree to litigate in *ex ante*. Then the relative use of courts in state s, $R(s)$, is defined as

$$R(s) \equiv \frac{p(L=s)}{p(C=s)}$$

The quantity is estimated using empirical averages. $R(s) > 1$ indicates that a state experiences more litigation than would be expected given the number of contracts that refer to litigation in said state. Conversely, $R(s) < 1$ indicates that litigation occurs less than would be expected solely based on the share of contracts. The numerator is estimated using all federal material lawsuits reported to the SEC by registered issuers between 2000 and 2016. The denominator is based on court selection clauses referring parties to federal courts in all material contracts reported to the SEC between 2000 and 2016. Restricting the analysis to federal litigation has the obvious shortcoming of being limited in scope and thus, does not allow the researcher to draw inferences on the substantial number of filings issued in state courts without making additional assumptions. However, due to the lack of comprehensive data or reporting requirements for state court litigation, any analysis that relies on convenience samples obtained through third party offers such as Westlaw or LexisNexis is subject to potential sampling bias. I thus prefer comprehensive data on the smaller subset of material federal litigation over less reliable data on all litigation, albeit that the latter is broader in scope.

Column $R(s)$ in Table A.2.1 presents the results for each of the states with more than 50 disputes.[50] It shows significant heterogeneity between jurisdictions in the number of lawsuits filed. In particular, with a ratio of 0.015, Delaware has by far the lowest relative litigation rate, followed by New York with 0.47. This is consistent with aforementioned theoretical considerations that often depict New York and Delaware as the most predictable jurisdictions in the nation. All other jurisdictions have a litigation rate greater than 1.

Even though the relative litigation rate provides a first insight, it has several potential shortcomings. In particular, it is possible that parties strategically choose to litigate certain types of disputes in some courts and other types of

[50]States with less than 50 disputes are omitted due to the lack of informational quality that can be obtained from such few observations.

disputes in other courts. If that is the case, then litigation rates could be a mere reflection of a difference in contract types that are governed by the different forums analyzed. In addition, since a substantial proportion of contracts does not specify a court forum at all, it is possible that the observed litigation originates not from contracts that specified a given forum, but from contracts that failed to specify a forum *ex ante*. In order to address these concerns, I estimate the following equation through negative binomial regression:

$$log(Disputes)_{s^*} = \alpha + \beta_1 type_{s^*} + \beta_2 HQ_{s*} + \beta_3 Incorp_{s^*} + \beta_4 log(contracts)_{s^*}$$

Here, $s^* \in S \setminus \{s\}$, i.e. for each state, the coefficients are estimated using only data from the other states. The coefficients obtained are then used to compute $\widehat{Disputes}_s$, the predicted number of disputes in state s. The procedure can be thought of as the best guess for the number of disputes in state s, based only on observed litigation rates in all other states. It controls for the distribution of contract types. In addition, the estimation approach controls for the economic headquarters and states of incorporation, because these are the two primary forums that parties can file a lawsuit in if the contract does not include a dispute resolution clause. The final quantity of interest is the relative over- / underutilization of courts, denoted as $U(s) = \frac{Disputes_s}{\widehat{Disputes}_s}$. If $U(s) > 1$, this means that courts are used more often than would be expected, based on litigation frequency in other states. Conversely, if $U(s) < 1$, litigation occurs less frequent than would be expected. Column $U(s)$ in Table A.2.1 presents the outcome of this quantity. Again, Delaware courts seem to deter litigation most frequently, followed by Colorado, Maryland and New York, all of which have roughly the same rate of 0.65. Interestingly, California courts, which are the most prominent example of a contextualist jurisdiction rejecting the notion that contractual language has an unambiguous meaning (Gilson 2014), are among the five most frequently overused court jurisdictions. Insofar, the analysis is consistent with theoretical considerations by which contextualist jurisdictions are significantly less predictable than textualist jurisdictions.

Table A.2.1: Relative Use of Courts

State	Disputes	Contracts	$R(s)$	$U(s)$
Delaware	74	12,742	0.14	0.61
Colorado	80	1,601	1.24	0.64
Maryland	75	1,123	1.66	0.65
New York	1002	51,863	0.48	0.67
Arizona	65	697	2.32	0.72
Virginia	74	1,331	1.38	0.74
Utah	59	685	2.14	0.82
Ohio	117	1,731	1.68	0.90
North Carolina	84	1049	1.99	0.94
Illinois	165	3,862	1.06	0.97
Texas	286	5,501	1.29	1.01
Massachusetts	134	2448	1.36	1.05
Nevada	104	1,922	1.35	1.08
New Jersey	140	2,638	1.32	1.08
Connecticut	69	652	2.63	1.12
Missouri	68	728	2.32	1.16
Georgia	109	1,482	1.83	1.17
Michigan	79	777	2.53	1.19
Minnesota	107	1,548	1.72	1.21
Pennsylvania	170	2,230	1.90	1.35
Louisiana	74	316	5.83	1.35
Florida	259	3,794	1.70	1.36
California	521	9,140	1.42	1.38
South Carolina	66	167	9.83	1.79
Washington	71	969	1.82	1.89
Wisconsin	70	473	3.68	1.90
Alabama	55	164	8.34	3.08

This table depicts the relative use of district courts in the different states. Disputes are the number of all federal material lawsuits reported to the SEC, contracts are the number of contracts with forum selection clauses referring parties to federal courts. $R(s)$ is the court usage rate relative to the share of contracts referring parties to the courts, $U(s)$ the ratio of actual usage rate and predicted usage rate, where predictions are derived from a negative binomial regression. States with fever than 50 disputes are omitted to ensure reliability of the results.

A.3.1 Groupings of Treaties by Subject Area

Agriculture: Agriculture, Agricultural Commodities, Poplar Commission
Amity: Amity, Friendship, General Relations, Relations
Arms Limitations: Arms Limitations
Aviation: Aviation, Aerospace Disturbanes
Claims: Claims, Arbitration, Occupation Costs
Crime: Crime, Computer Crime, Smuggling, Corruption, Bribery, War Criminals, Prisoner Transfer, Trafficking in Women and Children, Organized Crime
Culture: Culture, UNESCO, World Heritage, Cultural Heritage, Cultural and Educational Relations, Cultural Property, Cultural Relations, Cultural Relations: Inter-American
Defense: Defense, Economic and Military Cooperation, Evacuation, Naval Vessels, Open Skies, NATO, Missions Military
Diplomacy: Diplomacy, Diplomatic Properties, Diplomatic Relations, Consuls, Properties: Diplomatic, Embassy Sites
Economic and Technical Cooperation: Economic and Technical Cooperation, Sewage Disposal System, Lend-Lease, Economic and Technical Cooperation and Development, Relief Supplies and Packages, Economic Assistance, Economic and Technological Cooperation and Development
Education: Education
Energy: Energy, Petroleum, Pipelines, Solar Energy, Fuels and Energy
Environment: Environment, Forestry, Seals, Whaling, Pollution, Climate, Conservation, Desertification, Chemical Safety, Environmental Cooperation, Environmental Modification
Extradition: Extradition
Finance: Finance, Multilateral Funds, Financial Institutions, Finance: World War II Related
Fisheries: Fisheries, Shellfish
Health: Health, Health and Sanitation
Human and Fundamental Rights: Human and Fundamental Rights, Human Rights, Slavery, Torture, Women - Political Rights, Children, Prisoners of War, Racial Discrimination, Red Cross Conventions, Refugees, Rules of Warfare
IP: IP, Intellectual Property, Trademark, Copyrights, Phonograms
Judicial Assistance: Judicial Assistance, Judicial Assistance and Procedure
Labor: Labor, Employment
Maritime Matters: Maritime Matters, Maritime Interdiction, Seabeds, International Maritime Organization

Nuclear Energy: Nuclear Energy, Nuclear Accidents, Atomic Energy
Other: Antarctica, Arctic, Assistance, Automotive Traffic, Cambodia, Canals, Cemeteries, Civil Emergency Planning, Commissary Facilities, Compact of Free Association, Drivers Licenses, Emergency Management, Emergency Preparedness, Fire Protection, Headquarters, Highways, Humanitarian Assistance, Immigration, Interests Sections, Judicial Assistance and Procedure, Judicial Procedure, Judicial Procedure, Hague Conventions, Maintenance, Medical Assistance, Migration, Nationality, Organization of American States, Passports, Privileges and Immunities, Publications, Regional Commission, Sanctions, Social Security, Termination, Tourism, Tracking Stations, Treaty Succession, UN
Peacekeeping: Peacekeeping, Peace Corps, Peace Treaties, Renunciation of War
Postal Matters: Postal Matters, Postal Arrangements
Property: Property. Industrial Property, Stolen Property, Property Transfer
Satellites: Satellites, Satellite Communications Systems, Remote Sensing
Scientific Cooperation: Scientific Cooperation, Navigation, Weather Modification, Weather Stations, World Meteorological Association, Mapping, Technical Assistance, Technical Cooperation, Technical Assistance and Cooperation, Technological Cooperation, Technology Transfer, Seismic Observations, Seismological Research, Scientific Assistance and Cooperation, Scientific and Technical Cooperation, Oceanography, Oceanographic Research, Missions: Technical, Meteorology, Meteorological Cooperation, Marine Science, Geodetic Survey, Hydrography
Space: Space, Astronauts, Space Cooperation, Space Research
Taxation: Taxation
Telecommunication: Telecommunication, Telecommunication - Inter-American Agreements, Telecommunication - International Telecommunication Union
Trade and Commerce: Trade and Commerce, Coffee, Commerce, Containers, Copper, Cotton, Customs, GATT, Grains, Industrial Cooperation, Investment Disputes, Investments, Jute, Law: Private International, Liquor, Rubber, Schedules, Shipping, Sugar, Timber, Trade, Trade and Commerce: GATT-Related Agreements Trade and Investment, Transportation, Transportation-Foodstuffs, Wheat, Wine
US Boundaries: US Boundaries, Boundaries, Boundary Waters
Visas: Visas
Weapons: Weapons, Chemical Weapons, Chemicals, Nuclear Risk Reduction, Nuclear Test Limitation, Nuclear War

WW II Aftermath: WW II Aftermath, Germany, Holocaust Memorial, International Tracing Service, Reparations

A.3.2 Complete List of Agreement Type by Subject Area

Table A.3.1: Agreement Type by Subject Area

Subject	# EAs	# Treaties	Mean Treaty
Agriculture	454	1	0.00
Claims	28	0	0.00
Education	64	0	0.00
Energy	72	0	0.00
Health	45	0	0.00
Peacekeeping	73	0	0.00
Postal Matters	239	0	0.00
Satellites	40	0	0.00
Scientific Cooperation	533	1	0.00
Space	140	0	0.00
Visas	11	0	0.00
Aviation	538	3	0.01
Defense	1433	9	0.01
Econ. and Techn. Cooperation	674	4	0.01
Nuclear Energy	355	4	0.01
Other	138	2	0.01
Culture	63	1	0.02
Labor	131	3	0.02
Weapons	169	4	0.02
Crime	260	12	0.04
Finance	500	22	0.04
Trade and Commerce	748	35	0.04
Maritime Matters	87	5	0.05
Telecommunication	111	7	0.06
US Boundaries	52	4	0.07
IP	23	2	0.08
Environment	196	20	0.09
WW II Aftermath	10	1	0.09
Fisheries	83	9	0.10
Diplomacy	32	4	0.11
Arms Limitations	34	5	0.13
Human and Fundamental Rights	15	3	0.17
Amity	7	3	0.30
Property	8	5	0.38
Taxation	103	75	0.42
Judicial Assistance	93	80	0.46
Extradition	5	75	0.94

The table depicts the prevalence of treaties and executive agreements for all subject areas.

A.3.3 Complete List of Agreements by Country

Table A.3.2: Agreement Use by Partner Country

Country	# EAs	# Treaties	Mean Treaty
Mexico	247	6	0.02
Japan	250	2	0.01
Russia	219	4	0.02
United Kingdom	195	10	0.05
Canada	190	10	0.05
Egypt	188	2	0.01
South Korea	139	2	0.01
Germany	116	7	0.06
Philippines	116	2	0.02
France	106	10	0.09
Australia	102	4	0.04
China, Republic	104	1	0.01
Indonesia	100	2	0.02
Israel	97	3	0.03
Brazil	98	1	0.01
Ukraine	92	4	0.04
Pakistan	95	0	0.00
Peru	92	1	0.01
Italy	82	6	0.07
Jordan	85	2	0.02
EU	84	2	0.02
Bolivia	83	2	0.02
Hungary	81	4	0.05
Soviet Union	80	4	0.05
Colombia	82	0	0.00
Poland	75	5	0.06
Jamaica	75	2	0.03
Honduras	73	2	0.03
Spain	65	6	0.08
Dominican Republic	70	0	0.00
Romania	63	5	0.07
Argentina	63	3	0.05
Afghanistan	65	0	0.00
India	62	3	0.05
Netherlands	58	6	0.09
Kazakhstan	60	3	0.05
Sri Lanka	57	4	0.07
Panama	56	4	0.07
Turkey	54	2	0.04
Greece	53	2	0.04
El Salvador	54	0	0.00
Ecuador	52	1	0.02
Sweden	46	7	0.13
Norway	50	0	0.00
Chile	49	0	0.00
Finland	44	5	0.10
Morocco	47	2	0.04
Thailand	44	4	0.08
Bangladesh	43	2	0.04
Costa Rica	44	1	0.02
Guatemala	44	1	0.02
Senegal	44	1	0.02
Mongolia	42	2	0.05

Continued on next page

Table A.3.2 – continued from previous page

Country	# EAs	# Treaties	Mean Treaty
Venezuela	42	2	0.05
Singapore	43	0	0.00
Switzerland	41	2	0.05
Bulgaria	37	4	0.10
Congo, Republic	40	1	0.02
South Africa	38	3	0.07
Denmark	31	6	0.16
Latvia	31	6	0.16
Liberia	37	0	0.00
Tanzania	35	0	0.00
Uruguay	33	2	0.06
Nicaragua	34	0	0.00
China, Peoples	30	3	0.09
Czech Republic	27	6	0.18
Georgia	32	1	0.03
Haiti	33	0	0.00
Kenya	33	0	0.00
Lithuania	26	7	0.21
Malaysia	31	2	0.06
Mozambique	32	1	0.03
Ghana	32	0	0.00
Portugal	28	3	0.10
Tunisia	27	4	0.13
Czechoslovakia	30	0	0.00
Sierra Leone	30	0	0.00
Belgium	23	6	0.21
Estonia	23	6	0.21
Guyana	29	0	0.00
Vietnam	29	0	0.00
Madagascar	28	0	0.00
Zambia	27	0	0.00
Austria	20	6	0.23
Marshall Islands	26	0	0.00
Ethiopia	25	0	0.00
Micronesia	25	0	0.00
Uganda	25	0	0.00
Bahamas	22	2	0.08
Nigeria	23	1	0.04
Ireland	17	6	0.26
Croatia	21	1	0.05
Iceland	20	2	0.09
Sudan	22	0	0.00
Trinidad and Tobago	19	3	0.14
Yugoslavia	22	0	0.00
Antigua and Barbuda	19	2	0.10
Armenia	20	1	0.05
Belarus	21	0	0.00
United Arab Emirates	21	0	0.00
Yemen	21	0	0.00
Albania	19	1	0.05
Belize	17	3	0.15
Grenada	17	3	0.15
Luxembourg	15	5	0.25
New Zealand	18	2	0.10
Slovenia	17	3	0.15
Uzbekistan	20	0	0.00
Congo	18	1	0.05

Continued on next page

Table A.3.2 – continued from previous page

Country	# EAs	# Treaties	Mean Treaty
Guinea	19	0	0.00
Maldives	19	0	0.00
Mauritius	19	0	0.00
Azerbaijan	17	1	0.06
Botswana	18	0	0.00
Cote d Ivoire	18	0	0.00
Cyprus	13	5	0.28
Rwanda	17	1	0.06
Slovak Republic	16	2	0.11
Bahrain	16	1	0.06
Gabon	17	0	0.00
Macedonia	17	0	0.00
Niger	17	0	0.00
Oman	17	0	0.00
Saudi Arabia	17	0	0.00
Cameroon	15	1	0.06
Kyrgyzstan	15	1	0.06
Mali	16	0	0.00
Moldova	15	1	0.06
Nepal	16	0	0.00
Paraguay	15	1	0.06
Cuba	15	0	0.00
Barbados	9	5	0.36
Bosnia Herzegovina	14	0	0.00
Cape Verde	14	0	0.00
Central African Republic	14	0	0.00
Djibouti	14	0	0.00
Saint Kitts and Nevis	12	2	0.14
Cambodia	13	0	0.00
Chad	13	0	0.00
Fiji	13	0	0.00
Malta	10	3	0.23
Saint Vincent and the Grenadines	11	2	0.15
Benin	12	0	0.00
Dominica	10	2	0.17
Algeria	10	1	0.09
Iraq	11	0	0.00
Kuwait	11	0	0.00
Laos	11	0	0.00
Mauritania	11	0	0.00
Saint Lucia	9	2	0.18
Gambia	10	0	0.00
Namibia	10	0	0.00
Papua New Guinea	10	0	0.00
Somalia	10	0	0.00
Zimbabwe	9	1	0.10
Lebanon	9	0	0.00
Suriname	9	0	0.00
Tajikistan	9	0	0.00
Malawi	8	0	0.00
Palau	8	0	0.00
Turkmenistan	8	0	0.00
Burkina Faso	7	0	0.00
Burundi	7	0	0.00
Guinea Bissau	7	0	0.00
Kosovo	7	0	0.00
Montenegro	7	0	0.00

Continued on next page

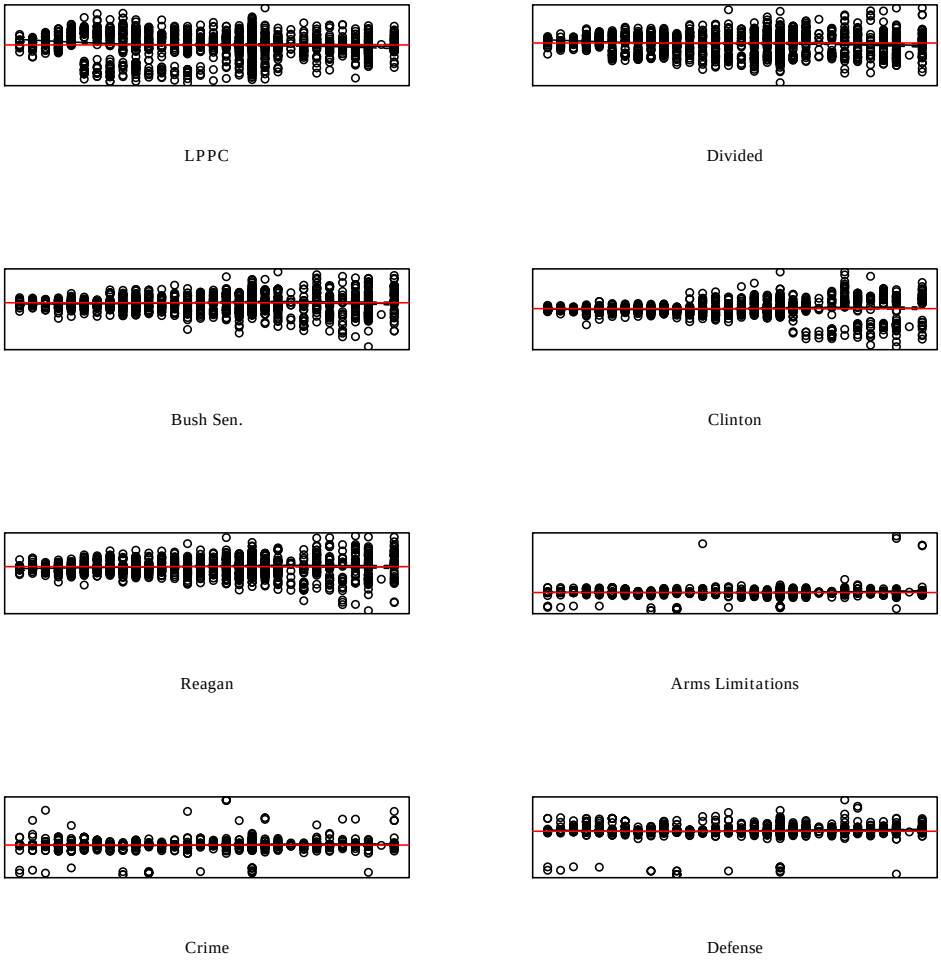
Table A.3.2 – continued from previous page

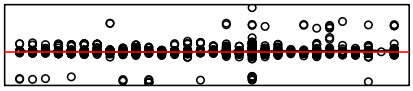
Country	# EAs	# Treaties	Mean Treaty
Qatar	7	0	0.00
Seychelles	7	0	0.00
Equatorial Guinea	6	0	0.00
Samoa	6	0	0.00
Sao Tome and Principe	6	0	0.00
Timor Leste	6	0	0.00
Togo	6	0	0.00
Tonga	6	0	0.00
Angola	5	0	0.00
Brunei	5	0	0.00
Comoros	5	0	0.00
Eritrea	5	0	0.00
North Korea	5	0	0.00
Lesotho	5	0	0.00
Solomon Islands	5	0	0.00
Swaziland	5	0	0.00
Kiribati	4	0	0.00
Libya	4	0	0.00
Liechtenstein	3	1	0.25
Nauru	4	0	0.00
Serbia	4	0	0.00
Serbia and Montenegro	4	0	0.00
Slovakia	2	2	0.50
Bhutan	3	0	0.00
Cook Islands	3	0	0.00
Iran	3	0	0.00
Tuvalu	3	0	0.00
Turks and Caicos Islands	3	0	0.00
Vanuatu	3	0	0.00
Aruba	2	0	0.00
Burma	2	0	0.00
Monaco	2	0	0.00
United Kingdom Anguilla	2	0	0.00
African Union	1	0	0.00
Andorra	1	0	0.00
Hong Kong	1	0	0.00
East Timor	1	0	0.00
Gibraltar	1	0	0.00
New Caledonia	1	0	0.00
Niue	1	0	0.00
Bermuda	1	0	0.00
British Virgin Islands	1	0	0.00
Guernsey	1	0	0.00
Isle of Man	1	0	0.00
Jersey	1	0	0.00

The table depicts the prevalence of treaties and executive agreements for all partner countries in the dataset.

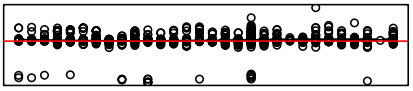
A.3.4 Graphical Representaiton of Schoenfeld Residuals

Figure A.3.1: Schoenfeld Residual Plots

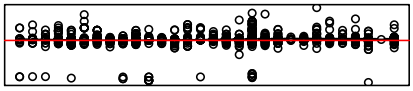




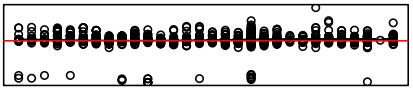
Education



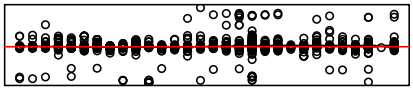
Energy



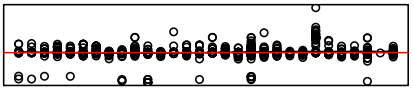
Health



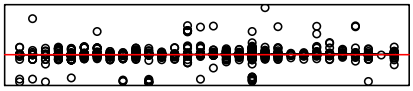
Nuclear Energy



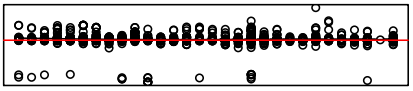
Other



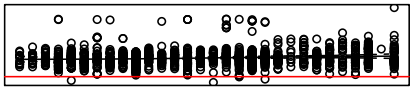
Postal Matters



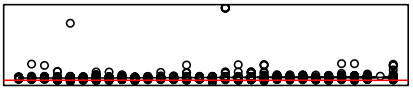
Telecommunication



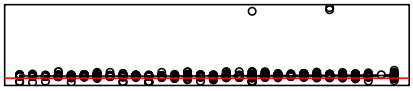
Trade and Commerce



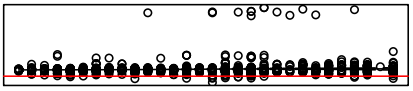
Burma



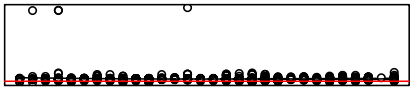
China



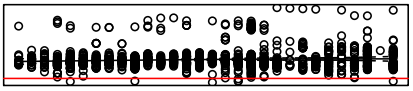
Congo



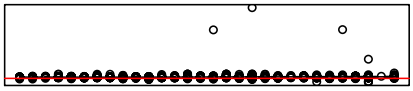
Ecuador



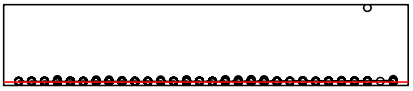
Haiti



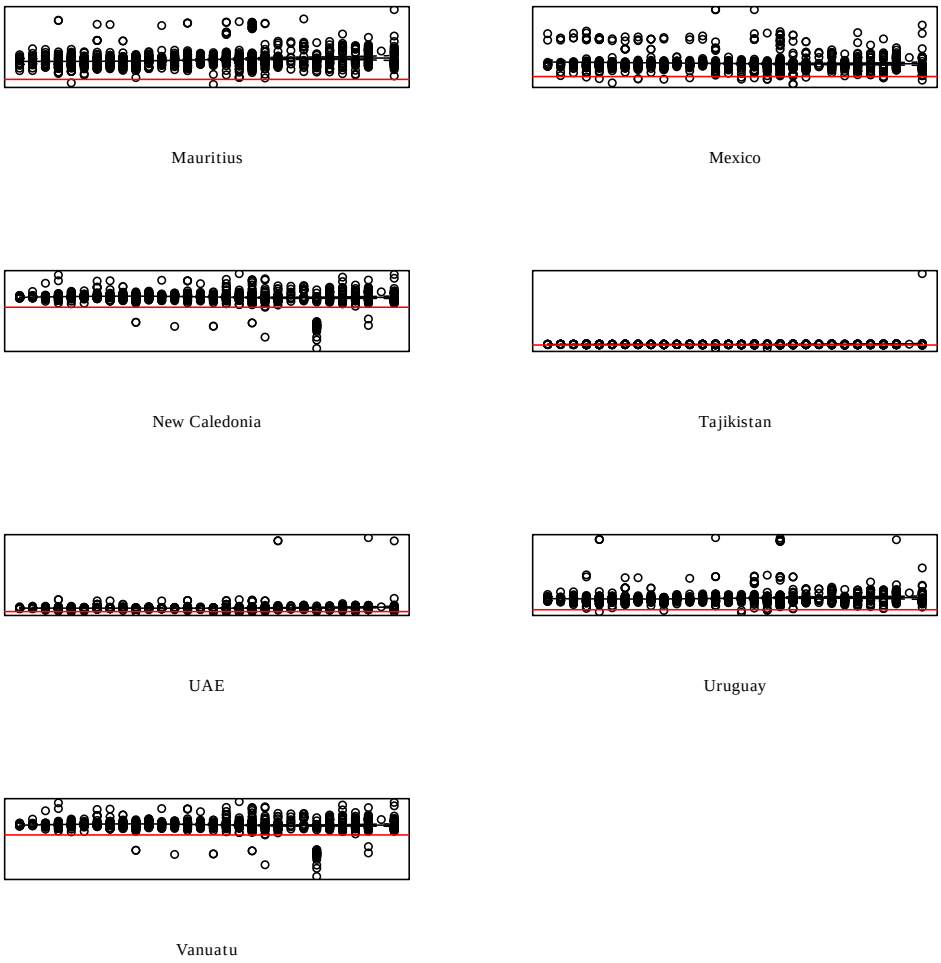
Hungary



Jamaica



Maldives



Schoenfeld residual plots for all covariates that yield significant p-values when testing for non-proportionality.

A.4.1 Reduction in Bias Obtained by Matching with Different Calipers

Table A.4.1: Bias Reduction by Caliper

Caliper	# Units	Bias
0.1	102	−97%
0.2	118	−91%
0.3	118	−86%
0.4	120	−80%
0.5	130	−73%
0.6	136	−62%
0.7	148	−51%
0.8	158	−43%
0.9	160	−36%
1.0	164	−28%
1.1	168	−24%
1.2	168	−21%
1.3	170	−20%

The table depicts the number of units and the reduction in mean differences of propensity scores between panels with lawyer chairs and those with non-lawyer chairs for different calipers in the matching algorithm.

A.4.2 Robustness of Results for the Analysis of Reversal Rates

Table A.4.2: Robustness Test for Reversal Analysis

	Dependent Variable:	
	Reversal Rate	
	Poisson	¬ Offset
Precedent	-0.447^{*}	-0.399^{**}
	(0.208)	(0.147)
Lawyer Chair	-0.731	-0.770^{+}
	(0.494)	(0.394)
Chair Experience	0.008	0.006
	(0.088)	(0.073)
Chair System	0.449^{+}	0.480^{*}
	(0.258)	(0.242)
Lawyer-Panelists = 1	-0.246	-0.243
	(0.313)	(0.270)
Lawyer-Panelists = 2	0.096	0.025
	(0.404)	(0.320)
CL Panelists = 1	0.125	0.093
	(0.258)	(0.210)
CL Panelists = 2	-0.665	-0.649
	(0.438)	(0.484)
DG Appointed Panel	-0.049	-0.037
	(0.256)	(0.229)
GATT I	0.236	0.344
	(0.278)	(0.276)

Continued on next page

Table A.4.2 – continued from previous page

	Poisson	¬ Offset
GATT II	−0.279	−0.324
	(0.316)	(0.296)
GATT III	−0.618	-0.555^{+}
	(0.379)	(0.325)
ADA	0.264	0.386
	(0.264)	(0.265)
Safeguards	0.070	0.167
	(0.399)	(0.373)
SPS	0.0004	0.130
	(0.423)	(0.419)
SCM	-1.206^{***}	-1.118^{***}
	(0.340)	(0.286)
Volume	−0.039	−0.020
	(0.064)	(0.048)
CL Parties	0.622^{***}	0.533^{***}
	(0.175)	(0.155)
DS vs. NDC	−0.362	−0.271
	(0.198)	(0.213)
# Words (Rest)	0.015	0.018
	(0.016)	(0.016)
Log(# Claims)		0.730^{***}
		(0.154)
LC*Experience	0.216^{*}	0.198^{+}
	(0.105)	(0.108)

Continued on next page

Table A.4.2 – continued from previous page

	Poisson	¬ Offset
LC*CLC	-0.896^{+}	-0.800^{*}
	(0.493)	(0.394)
Precedent*LC	0.681^{**}	0.629^{***}
	(0.249)	(0.191)
Precedent*Experience	0.110^{+}	0.090^{+}
	(0.057)	(0.047)
Precedent*LC*Experience	-0.285^{***}	-0.245^{***}
	(0.061)	(0.067)
Constant	−0.386	−0.195
	(0.836)	(0.676)
Time Fixed Effects	✓	✓
Observations	108	108
Pseudo-R^2	0.589	0.637
Log Likelihood	−178	−176
Akaike Inf. Crit.	437	437
Note:	$^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001	

Alternative parametric assumptions for the full model estimating the association between precedent usage and reversal rates. Missing values are imputed 100 times and results. AIC and Log-Likelihood estimates are averages across iterations. Volume in bio$, words in units of 10,000.

A.4.3 Sensitivity of Results for the Analysis of Reversal Rates

Table A.4.3: Sensitivity Analysis for Reversal Rates

	Dependent Variable:			
	Reversal Rate			
	Minimal	Panel	Dispute	Party
Precedent	−0.223	−0.220	−0.378*	−0.293+
	(0.160)	(0.166)	(0.163)	(0.156)
Lawyer Chair	−0.375	−0.381	−0.762+	−0.481
	(0.399)	(0.453)	(0.428)	(0.390)
Chair Experience	−0.010	−0.0003	−0.018	−0.011
	(0.085)	(0.089)	(0.083)	(0.081)
CL Chair		0.359		
		(0.281)		
Lawyer-Panelists = 1		−0.473		
		(0.294)		
Lawyer-Panelists = 2		−0.238		
		(0.320)		
CL Panelists = 1		0.015		
		(0.242)		
CL Panelists = 2		−0.713		
		(0.508)		
DG Appointed Panel		−0.471+		
		(0.266)		
GATT I			0.149	
			(0.275)	

Continued on next page

Table A.4.3 – continued from previous page

	Minimal	Panel	Dispute	Party
GATT II			−0.148	
			(0.307)	
GATT III			−0.734*	
			(0.334)	
ADA			0.158	
			(0.254)	
Safeguards			−0.167	
			(0.370)	
SPS			−0.097	
			(0.457)	
SCM			−0.841**	
			(0.299)	
CL Parties				0.444**
				(0.156)
DC vs. NDC				−0.345+
				(0.200)
# Words (Rest)		0.004	0.016	−0.0001
		(0.016)	(0.016)	(0.014)
LC*Experience	0.103	0.077	0.155	0.081
	(0.121)	(0.132)	(0.117)	(0.114)
LC*CLC		−0.165		
		(0.477)		
Precedent*LC	0.413+	0.458*	0.598*	0.473*
	(0.227)	(0.234)	(0.233)	(0.219)

Continued on next page

Table A.4.3 – continued from previous page

	Minimal	Panel	Dispute	Party
Precedent*Experience	0.079	0.060	0.101 $^{+}$	0.095^{+}
	(0.053)	(0.057)	(0.052)	(0.051)
Precedent*LC*Experience	-0.212^{**}	-0.203^{**}	-0.264^{***}	-0.217^{**}
	(0.076)	(0.079)	(0.075)	(0.071)
Constant	-0.468	-0.199	-0.121	-0.510
	(0.514)	(0.655)	(0.570)	(0.524)
Time Fixed Effects	✓	✓	✓	✓
Observations	108	108	108	108
Pseudo-R^2	0.295	0.350	0.382	0.370
Log Likelihood	-200	-196	-194	-195
Akaike Inf. Crit.	445	452	448	439
Note:				$^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001

The output of a negative binomial regression. Each model adds a different set of controls (minimum, panel characteristics, dispute characteristics, party characteristics) in order to investigate how the coefficient on Precedent changes with inclusion. Standard errors in parentheses. Volume in bio$, words in units of 10,000.