

Appendix for Predictability Versus Flexibility: Secrecy in International Investment Arbitration

Emilie M. Hafner-Burton*, Zachary C. Steinert-Threlkeld+, David G. Victor**

DESCRIPTIVE STATISTICS	2
ASSIGNING CATEGORY CODES.....	4
CODING LONG LIVED.....	5
ARBITRATION FLOW CHART.....	6
CODEBOOK.....	7
VARIABLE DEFINITION.....	7
<i>Claimant Variables.....</i>	<i>7</i>
<i>Compustat Variables.....</i>	<i>15</i>
<i>Secondary Claimant Variables.....</i>	<i>18</i>
NOT ICSID OR COMPUSTAT	21
NOTES TO CODING	27
NOTES TO MERGING WITH COMPUSTAT	34
INTERCODER RELIABILITY.....	36
<i>Shu Shang.....</i>	<i>36</i>
<i>Longhao Wang.....</i>	<i>36</i>
<i>Catherine Nicklen.....</i>	<i>37</i>

*: ehafner@ucsd.edu

+: zst@luskin.ucla.edu

**+: david.victor@ucsd.edu

World Politics, Volume 68, No. 3

Descriptive Statistics

TABLE 1: Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Max
<i>Private</i>	375	.387	.487	0	1
<i>LongLived</i>	370	.527	.500	0	1
<i>Losses_R</i>	364	.997	1.82	0	9
<i>PublicCases_R</i>	364	1.51	2.42	0	13
<i>Reform</i>	375	3.67	3.12	0	11
<i>Additional Facility</i>	375	.093	.291	0	1
<i>Bribery_R</i>	375	.211	.408	0	1
<i>Bribery_C</i>	375	.717	.451	0	1
<i>GDP_R (Log)</i>	373	7.91	1.25	4.70	10.7
<i>GDP_C (Log)</i>	371	10.2	.689	5.57	10.9
<i>Polity_R</i>	361	3.48	6.15	-10	10
<i>Polity_C</i>	372	9.18	3.09	-10	10
<i>FDI_R (Log)</i>	368	.887	1.14	-2.32	3.97

TABLE 2: Variable Correlations

	<i>Private</i>	<i>LongLived</i>	<i>Losses_R</i>	<i>PublicCases_R</i>	<i>Reform</i>	<i>Additional Facility</i>	<i>Bribery_R</i>	<i>Bribery_C</i>	<i>GDP_R (Log)</i>	<i>GDP_C (Log)</i>	<i>Polity_R</i>	<i>Polity_C</i>	<i>FDI_R (Log)</i>
<i>Private</i>	1.00												
<i>LongLived</i>	.211	1.00											
<i>Losses_R</i>	.150	-.003	1.00										
<i>PublicCases_R</i>	.100	.009	.955	1.00									
<i>Reform</i>	.105	.179	.145	.218	1.00								
<i>Additional Facility</i>	-.206	-.062	-.084	-.052	.047	1.00							
<i>Bribery_R</i>	.003	.067	.470	.473	.013	.061	1.00						
<i>Bribery_C</i>	-.082	.045	.276	.316	.549	.023	.205	1.00					
<i>GDP_R (Log)</i>	.049	.040	.241	.242	.026	.263	.544	.128	1.00				
<i>GDP_C (Log)</i>	.086	.084	.054	.062	.115	-.010	.019	.329	-.005	1.00			
<i>Polity_R</i>	.029	.145	.126	.187	.163	.257	.420	.152	.506	.037	1.00		
<i>Polity_C</i>	-.102	.043	.046	.047	-.059	.053	.064	.309	.046	.521	.125	1.00	
<i>FDI_R (Log)</i>	-.058	.077	-.079	-.007	.313	.069	.050	.234	.172	.020	.224	-.017	1.00

Assigning Category Codes

To determine the category code to which a case belongs, we consulted the Product variable, the verbatim description ICSID provides of the claimants' industry. (As of July 2015, ICSID labels this description as the claimant's "Economic Sector".) Because ICSID is very precise with these descriptions, there are almost as many descriptions as there are cases. To facilitate the identification of long-lived relationships, we therefore manually mapped each ICSID description to one of thirteen category codes we created. For example, ICSID describes ARB/05/14 as concerning as "oil exploration contract", which we assign to the "hydrocarbon supply and infrastructure" category; ARB/03/16 is an "airport project" and is assigned to the "ports and airports" category.

For a complete mapping of each ICSID description to the category code, please consult the dataset accompanying this paper. The ICSID description is the "Product" variable, and "Cat_code" provides our industry coding. The Codebook section gives the 13 industry codes we used.

The next section details the mapping from category codes to *LongLived*.

Coding Long Lived

We determined whether a claimant is in a long-lived industry based upon the coding we assigned from ICSID's description of industry. A long-lived industry is one in which the time horizon for the investment to be substantially profitable is relatively long (a decade or longer) and for which the firm cannot readily enter or exit from its investment. The assignment of *LongLived* (1, 0) was done by David Victor based on extensive experience analyzing actual investments in most of these industries. The category includes investments in the fields of electricity, hydrocarbons, mining, ports and airports and roads, railroads and transport infrastructure. We exclude telecommunication investment from *LongLived* because most of those investments were for equipment and wireless systems that, unlike wired telecoms infrastructures, have much shorter time horizons.

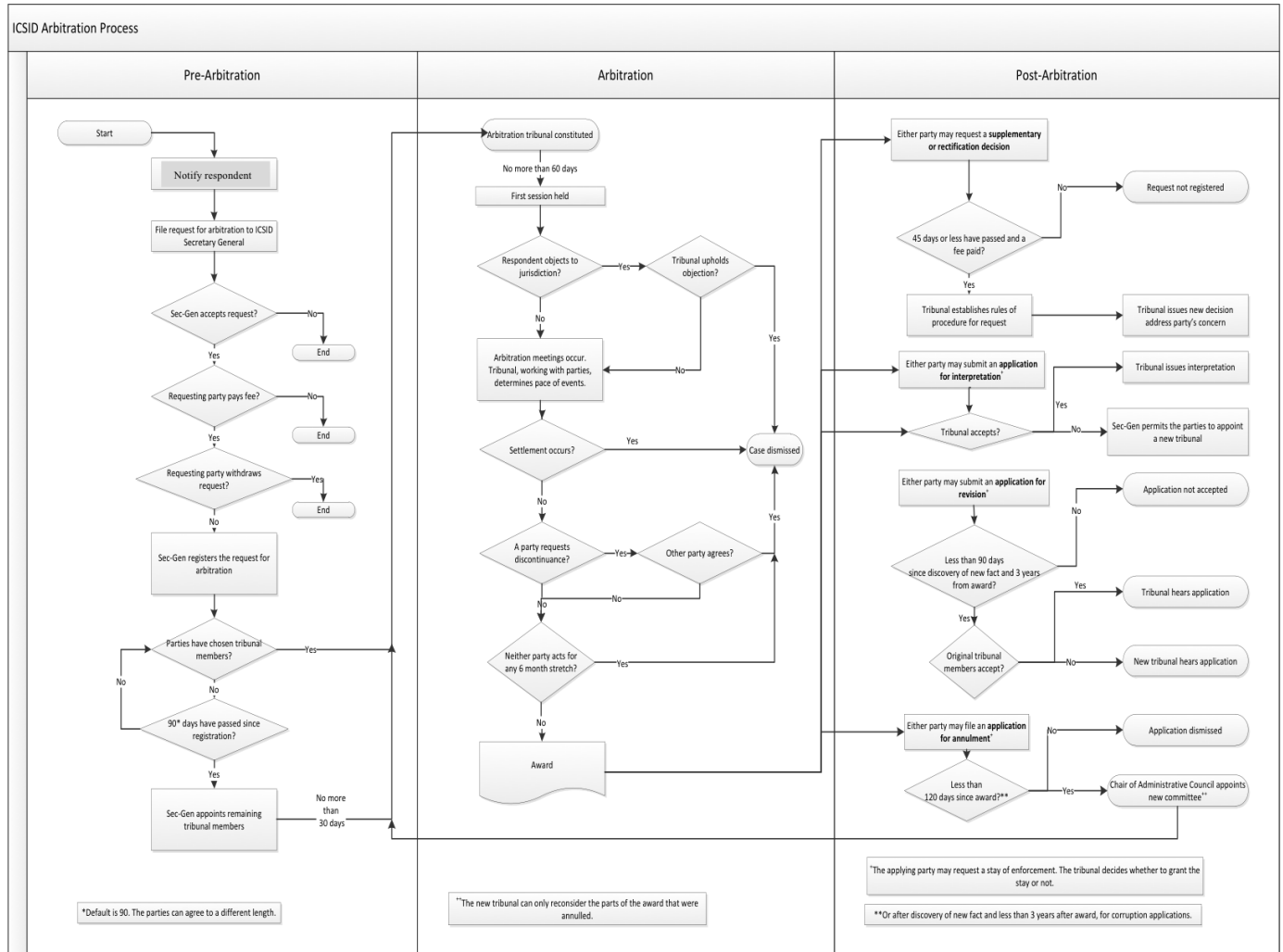
Our coding here aligns with the illustrations of investments that have inspired the theory and practical policy concerns surrounding the “obsolescing bargain” described in the main paper.. As a legal matter, many investments of this type are managed as concessions precisely because they are long-lived and thus exposed to the vagaries of government intervention.

We also tried a different approach to coding this concept by looking just at investments that have the word “concession” in the title as reported by ICSID. However, that approach was flawed because many investments of this type do not actually use the word “concession” and many are made with legal structures that are not strictly concessions. Thus we adopted, for our statistical tests, the industry-by-industry categorization.

Arbitration Flow Chart

Figure 1 here expands on Figure 1 from the main paper.

Figure 1: Detailed Arbitration Flowchart



Codebook

Variable Definition

Claimant Variables

Claimant Variables				
Number	Name	Description	Sample Value	Note
1	Merge	Count variable for tracking cases during coding	1 - 584	
2	Case	Case ID	ICSID ARB/02/16	
3	Product	ICSID's description of claimant industry	Glass production enterprise	
4	Claim_firm	Claimant firm name	Enron	See Notes 4 & 5
5	Category	User-created category label	Financial	
6	Cat_code	Our categories based on the substantive focus of the investment, based on variable 3	<ol style="list-style-type: none">1. Agriculture, fisheries, food, drinks, forestry2. Electricity, electric infrastructure3. Financial4. General industry5. General infrastructure6. Hydrocarbon supply and infrastructure7. Mining8. Other9. Ports and airports10. Roads, railroads and transport infrastructure11. Telecommunications, radio, and television12. Tourism13. Water and sewer	See Note 45

Claimant Variables

Number	Name	Description	Sample Value	Note
			infrastructure	
7	Stock_1	Primary stock ticker symbol	REP.MC (Repsol YPF's Madrid listing)	See Note 24
8	Stock_2	Secondary stock ticker symbol	YPF (Repsol YPF's NYSE listing)	See Note 24
9	Stock_3	Third stock ticker symbol	REP.BA (Repsol YPF's Buenos Aires listing)	See Note 24
10	Count_parties	Number of claimants filing	1, 2, 5	See Notes 3 & 26
11	Claim_state	Home state of claimant	United States	See Notes 25 & 27
12	Claim_ISO	Claimant's state's ISO number	752 (Sweden), 792 (Turkey)	See Note 31
13	Claim_assoc	Is the claimant an industry association?	0. No 1. Yes	
14	Resp_state	Host state	Ukraine, Thailand, etc.	See Note 25
15	Resp_ISO	What is the respondent's state's ISO number?	188 (Costa Rica), 348 (Costa Rica)	See Note 31
16	Resp_OECD19	Is the respondent an original member of the OECD and is not Turkey?	0. No 1. Yes	
17	Resp_OECD20	Is the respondent an original member of the OECD?	0. No 1. Yes	
18	Resp_OECD24	Did the respondent join the OECD before 1991?	0. No 1. Yes	
19	States_match	Are the claimant and respondent states the same?	0. No 1. Yes	
20	Date_register	Date claimant registered the case	05.06.2007	
21	Year_register	Year claimant registered the case	2007	
22	Month_register	Month claimant registered the case	05	
23	Day_register	Day claimant registered the case	06	
24	Date_const	Date tribunal constituted	06.07.2008	See Note 9
25	Year_const	Year tribunal constituted	2008	
26	Month_const	Month tribunal constituted	06	

Claimant Variables

Number	Name	Description	Sample Value	Note
27	Day_const	Day tribunal constituted	07	
28	Reg2const	Number of days passed from case registration to tribunal constitution	252	
29	Arbitration	Did arbitration begin?	0. No 1. Yes	See Note 32
30	Arb_pres	Arbitration president	Yves Fortier	
31	Arb_2	Second tribunal member	Michael Hwang	
32	Arb_3	Third tribunal member	David A.R. Williams	
33	Arb_prescountry	Country of arbitration president	Canada	
34	Arb_2country	Country of second arbitration member	Singapore	
35	Arb_3country	Country of third arbitration member	New Zealand	
36	Jurisdiction_obj	Did (has) the respondent challenged the panel's jurisdiction?	0. No 1. Yes	
37	Jrsdctn_res	Did (has) the panel uphold the jurisdiction challenge?	0. No 1. Yes	See Note 20
38	Investment	Is the dispute over an investment or contract?	0. Contract 1. Investment	
39	NatTreat	Did the claimant accuse the respondent of violating the standard of national treatment?	0. No 1. Yes	See Note 11
40	MFN	Did the claimant accuse the respondent of not according it most-favored nation status?	0. No 1. Yes	See Note 11
41	FPS	Did the claimant accuse the respondent of not granting its investment full protection and security?	0. No 1. Yes	See Note 11

Claimant Variables

Number	Name	Description	Sample Value	Note
42	DirExp	Did the claimant accuse the respondent of a direct expropriation?	0. No 1. Yes	See Notes 11 & 16
43	IndExp	Did the claimant accuse the respondent of an indirect expropriation?	0. No 1. Yes	See Note 11
44	Currency	Did the claimant accuse the respondent of inhibiting international currency transfers?	0. No 1. Yes	See Note 11
45	StabLeg	Did the claimant accuse the respondent of not maintaining a stable legal environment?	0. No 1. Yes	See Notes 11 & 17
46	ObsObl	Did the claimant accuse the respondent of not observing its obligations? This claim often invokes an umbrella clause or a specific breach of contract.	0. No 1. Yes	See Notes 11 & 12
47	Arbitrary	Did the claimant accuse the respondent of engaging in arbitrary, but non-discriminatory, conduct?	0. No 1. Yes	See Note 11
48	Arb_presvote	Vote or president	0. For claimant 1. Against claimant	
49	Arb_2vote	Vote of 2 nd member	0. For claimant 1. Against claimant	
50	Arb_3vote	Vote of 3 rd member	0. For claimant 1. Against claimant	
51	Winner	Did the claimant win?	0. No 1. Yes	See Note 21
52	Judge_rule	Panel's ruling	1. For claimant	See Note 15

Claimant Variables

Number	Name	Description	Sample Value	Note
			2. Against claimant 3. Affirmed/remanded in part 4. Embodied settlement	
53	Money_seek	The amount of damages claimant seeks	\$500,000,000	See Notes 6 & 7
54	Money_seek2011	Money_seek converted to 2011 dollars	\$550,000,000	
55	Interest_sought	Interest calculation sought	LIBOR + 2%	
56	Money_receive	Damages the tribunal awards	\$75,000,000	See Note 28
57	Interest_award	Are interest charges part of the award?	0. No 1. Yes	
58	Interest_calc	How is interest calculated?	LIBOR + 2%	
59	Money_receiveYY YY	The year the award or settled is issued	1999	
60	Money_receive20 11	The value of the award or settlement in 2011 dollars	\$200,000,000	
61	Rule43	Did the case end according to ICSID's Rule 43?	0. No 1. Yes	See Rule 29
62	Settled	Has the case been settled?	0. No 1. Yes	See Notes 21 & 22
63	Settle_amt	Amount of settlement	\$50,000,000	See Note 28
64	Spoils	Did the claimant receive 100% of Money_seek?	0. No 1. Yes	
65	Annul	Did the losing party try to annul the results?	0. No 1. Yes	See Note 13
66	Annul_dec	Was the annulment upheld?	0. No 1. Yes	See Note 13
67	Supplementary	Did the winning party seek a supplementary decision?	0. No 1. Yes	

Claimant Variables

Number	Name	Description	Sample Value	Note
68	Supplement_dec	Did the supplementary decision reverse the initial one?	0. No 1. Yes	
69	Resubmission	Did either party file for resubmission after an award?	0. No 1. Yes	See Note 18
70	Resub_dec	Did the tribunal rule in favor of the party filing the resubmission?	0. No 1. Yes	See Note 18
71	Revision	Did either party ask for a revision after the award?	0. No 1. Yes	See Note 19
72	Revis_dec	Did the tribunal find in favor of the party seeking revision?	0. No 1. Yes	See Note 19
73	Claim_ind	Is the claimant an individual?	0. No 1. Yes	See Note 4
74	Ind_name	If the claimant is an individual, what is its name?	Joseph Lemire	See Note 4
75	CONC	Is the event an ICSID Conciliation proceeding?	0. No 1. Yes	See Note 30
76	ICSID_AF	Is the arbitration using ICSID Additional Facility rules?	0. No 1. Yes	See Note 30
77	UNCITRAL	Is the arbitration using UNCITRAL rules?	0. No 1. Yes	
78	Status	What is the panel's status?	1. Ongoing 2. Concluded	
79	Date_ended	Date the proceedings end	10.05.2010	See Note 14
80	Year_ended	Year the proceedings ended	2010	
81	Reg2end	Date_ended minus Date_registered	560	
82	Indiv_opinion_p res	Did the president issue a separate opinion?	0. No 1. Yes	See Note 8
83	Pres_support	Did the president support the	0. No	See Notes 8 & 23

Claimant Variables

Number	Name	Description	Sample Value	Note
		tribunal's decision?	1. Yes	
84	Pres_rsn	Reason president gives	Disagrees with finding on full protection and security	See Note 8
85	Indiv_opinion2	Did the panelist issue a separate opinion?	0. No 1. Yes	See Note 8
86	lo2_support	Did the panelist support the tribunal's decision?	0. No 1. Yes	See Notes 8 & 23
87	lo2_rsn	Reason given	Disagrees with finding on expropriation	See Note 8
88	Indiv_opinion3	Did the panelist issue a separate opinion?	0. No 1. Yes	See Note 8
89	lo3_support	Did the panelist support the tribunal's decision?	0. No 1. 1. Yes	See Notes 8 & 23
90	lo3_rsn	Reason given	Disagrees with finding on jurisdiction	See Note 8
91	Trade	Does jurisdiction come from a non-NAFTA trade agreement?	0. No 1. Yes	
92	Contract	Does jurisdiction come from a contract between the parties?	0. No 1. Yes	
93	NAFTA	Does jurisdiction come from NAFTA?	0. No 1. Yes	
94	ECT	Does jurisdiction come from the Energy Charter Treaty?	0. No 1. Yes	
95	Notes	Any notes about the case or coding decision	0. No 1. Yes	
96	Litigiousness	Number of times claimant has brought cases at ICSID	2	
97	Claimant_experience	Number of previous cases the claimant's state has been in	20	

Claimant Variables

Number	Name	Description	Sample Value	Note
98	Claimant_prior_losses	Number of previous cases the claimant has lost.	1	Will only know for public cases
99	Claimant_public_cases	Number of previous public cases the claimant has filed	3	
100	MatchKey	Concatenate Case and Claim_firm	ARB/12/12/VattenfallAB and others	Used when updating at case-level

Compustat Variables

COMPUSTAT VARIABLES				
Number	Name	Description	Sample Value	Notes
101	GlobalMatch	Is the claimant in the Compustat Global database?	0. No 1. Yes	
102	NAMatch	Is the claimant firm in the Compustat North America database?	0. No 1. Yes	
103	Public	Is the claimant in either Computer database?	0. No 1. Yes	
104	Notes	Any notes from the merging process		
105	MergeKey	Firm-specific key plus year filed	1221542000 (ADF Group)	
106	Firm-specifickey	ID number assigned by Compustat	122154	
107	Filingdate	Date of info in Compustat	20010131	
108	FiscalYear	Year of Compustat data	2000	
109	CompanyName	Company name	ADF Group Inc.	
110	CompanyLegalName	Legal company name	ADF Group Inc.	
111	StockTicker	Stock ticker ID	ES01670509	
112	IntlSecurityID	International security ID	B01FLQ6	
113	StockExchangeDailyOfficialListCode	Stock exchange daily official list code	ADFJF	
114	Exchange	ID of exchange listed on	19	
115	CountryofIncorporation	Where company is incorporated	CAN	
116	HeadquarterLocation	Where company is headquartered	CAN	
117	Reportedcurrency	Currency company reports its figures in	CAD	
118	CurrencyTranslationRate	How many dollars per unit of currency	1	
119	NativeCurrency	Native currency	1	
120	NAICS	North American Industry	331111	

COMPUSTAT VARIABLES

Number	Name	Description	Sample Value	Notes
		Classification System number		
121	SIC	Standard International Classification number	3312	
122	StockOwnershipCode	Stock ownership code	3	
123	CompanyStatus	Company status code	A	
124	ResearchCompanyDeletionCode	1 if the company no longer exists in Compustat	1	
125	DeletionDate	Date company was removed from Compustat	20070129	
126	IntlDomesticBothFlag	International, domestic, or both flag	B	
127	IndustryFormat	Industry format	INDL	
128	AssetsCurrentTotal	Current total assets (millions)	228.202	
129	AssetsTotal	Total assets (millions)	298.799	
130	CommonEquity	Common equity (millions)	135.815	
131	ComprehensiveIncome	Comprehensive income (millions)	-15.357	
132	CostofGoodsSold	Cost of goods sold (millions)	273.997	
133	CommonStockCapital	Common stock capital (millions)	77.311	
134	EarningsBeforeTaxes	Earnings before taxes (millions)	41.325	
135	Employees	Number of employees (thousands)	1.4	
136	ExchangeRateEffect	Exchange rate effect (millions)	.206	
137	SourcesoffundsfromOperations	Sources of funds from operations (millions)	5729	
138	GrossProfit	Gross profit (millions)	60.388	
139	IncomeBeforeExtraordinaryItem	Income before extraordinary items (millions)	24.215	
140	InvestedCapital	Invested capital (millions)	184.679	
141	Inventories	Inventories (millions)	19.718	
142	LoansClaimsAdvances	Loans/claims/advances (millions)	None in dataset	

COMPUSTAT VARIABLES

Number	Name	Description	Sample Value	Notes
143	TotalLiabilities	Total liabilities (millions)	160.203	
144	NetIncome	Net income (millions)	24.215	
145	NonperformingAssets	Non-performing assets (millions)	None in dataset	
146	OperatingIncomeAfterDe preciat	Operating income after depreciation (millions)	41.324	
147	OperatingIncomeBeforeD eprecia	Operating income before depreciation (millions)	45.723	
148	OperatingProfits	Operating profits (millions)	535012.152	
149	PretaxIncome	Pretax income (millions)	35.447	
150	PretaxIncomeDomestic	Pretax income, domestic (millions)	35.447	
151	PretaxIncomeForeign	Pretax income, foreign (millions)	136	
152	GrossPlantsPropertyEquip men	Gross plants, property, and equipment (millions)	76.394	
153	NetPlantsPropertyEquipm ent	Net plants, property, equipment (millions)	62.225	
154	RDinProgress	Research and development in progress (millions)	-4	
155	TotalRevenue	Total revenue (millions)	334.385	
156	NetSales	Net sales (millions)	334.385	
157	StakeholdersEquity	Stakeholders' equity (millions)	135.815	
158	TotalBorrowings	Total borrowings (millions)	0	
159	TotalCurrentOperatingEx penses	Total current operating expenses (millions)	0	
160	TotalCurrentOperatingRe venue	Total current operating revenue (millions)	0	
161	TotalFairValueAssets	Total fair value assets (millions)	2.9	
162	GrossIncome	Gross income (millions)	898	
163	StaffExpenses	Staff expenses (millions)	11845	

COMPUSTAT VARIABLES

Number	Name	Description	Sample Value	Notes
164	OperatingExpenses	Operating expenses (millions)	102448.935	
165	ExportSales	Export sales	None in dataset	
166	StockMatch	Do the stock tickers we identified match those in Compustat?	0. No 1. Yes	
167	YearMatch	Does year_register equal FiscalYear?	0. No 1. Yes	

Secondary Claimant Variables

Claimant Variables, Second Coding Round

Number	Name	Description	Sample Value	Notes
168	Decision	Has the panel issued a decision?	0. No 1. Yes	See Note 34
169	Claim_OECD19	Is the claimant an original member of the OECD and is not Turkey?	0. No 1. Yes	
170	Claim_OECD20	Is the claimant an original member of the OECD?	0. No 1. Yes	
171	Claim_OECD24	Did the claimant join the OECD	0. No	

Claimant Variables, Second Coding Round

Number	Name	Description	Sample Value	Notes
		before 1991?	1. Yes	
172	Consolidated	Was the case combined with any other case?	0. No 1. Yes	See Note 38
173	Claim_US	Is the claimant headquartered in the United States?	0. No 1. Yes	
174	Resp_US	Is the respondent the United States?	0. No 1. Yes	
175	Resp_Arg	Is the respondent Argentina?	0. No 1. Yes	
176	Winner_multinom	Did the claimant win its case?	- 1. No 0. No information 1. Yes	
177	Capital_club	Is the claimant from Belgium, Canada, France, Germany, Greece, Italy, Luxembourg, Netherlands, Norway, Italy, Switzerland, the United Kingdom, or the United States?	0. No 1. Yes	
178	Claim_capitalclub_win	Number of public cases that all Capital_club states have won, up to 2 years before the case's constitution.	5	
179	Claim_capitalclub_tot	Total number of public cases that all Capital_club states have appeared as claimants, up to 2 years before arbitration's constitution.	9	
180	Claim_capitalclub_winrate	Claim_capitalclub_win/Claim_capitalclub_tot	5/9	
181	Capitalclub_exp_2	Total number of public and private cases in which claimant is from a Capital_club state	15	
182	Resp_public_lose_nolag	How many public cases has the	3	

Claimant Variables, Second Coding Round

Number	Name	Description	Sample Value	Notes
		respondent lost before the case's constitution?		
183	Resp_public_lose	How many public cases has the respondent lost, up to 2 years before the case's constitution?	4	
184	Resp_public_win	How many public cases has the respondent won, up to 2 years before the case's constitution?	2	
185	Resp_public_tot	Up to 2 years before the case's constitution, in how many public cases has the respondent participated?	5	
186	Resp_public_tot_nolag	In how many cases has the respondent participated?	8	
187	Resp_totexperience_nolag	In how many cases has the respondent participated?	10	
188	Source_legal	Does case information come from ICSID, italaw.com, a law firm, the Transnational Dispute Management journal, the International Institute for Sustainable Development, or UNCTAD?	0. No 1. Yes	See Note 1
189	Doc_Length	Number of pages in the .pdf used to code the lawyer variables	96	
190	Doc_Type	Case document used to code the lawyer variables	1. Award 2. Jurisdiction 3. Annulment 4. Other	
191	C_LawyerTot	Number of lawyers representing the	4	

Claimant Variables, Second Coding Round

Number	Name	Description	Sample Value	Notes
		claimant		
192	R_LawyerTot	Number of lawyers representing the respondent	5	
193	C_LawyerGov	Number of claimant lawyers from the government	0	
194	R_LawyerGov	Number of respondent lawyers from the government	3	
195	Notes	Notes from coding the lawyer variables		
196	Source_Public	For public awards, is the source of information ICSID, italaw.com, or somewhere else?	0. ICSID 1. italaw.com 2. Other	
197	Concession	Does the Product variable contain the word "concession"?	0. No 1. Yes	

Not ICSID or Compustat

Not ICSID or Compustat Variables

Number	Name	Description	Sample Value	Notes
198	PolityMatch_C	Match key for getting claimant's Polity2 score	United States2000	Stops in 2010

Not ICSID or Compustat Variables

Number	Name	Description	Sample Value	Notes
199	PolityMatch_R	Match key for importing Polity2 score for respondent	Venezuela, Bolivarian Republic Of2005	Stops in 2010
200	Polity2_C	Claimant's Polity2 score	8	See Note 35
201	Polity2_R	Respondent's Polity2 score	2	See Note 35
202	IMF_Grp_C	IMF country classification for claimant	1. High income 2. Middle income 3. Low income	See Note 36
203	IMF_Grp_R	IMF country classification for respondent	1. High income 2. Middle income 3. Low income	See Note 36
204	DoingBusiness_Year	Year used to match respondent to World Bank's Doing Business data	2005	See Note 37
205	Enforcement_Time	Number of calendar days to complete enforcement procedures	300	See Note 37
206	Enforcement_Cost	Cost of courts, enforcement, and attorneys as a fraction of 200% of country's per capita income	.45	See Note 37
207	Enforcement_Procedures	Number of steps required to use respondent's domestic courts	40	See Note 37
208	Claim_2011GDP	Gross domestic product of claimant's home country in 2011 dollars	\$30,000	See Notes 39 & 41
209	Claim_Trade_GDP	(Imports + Exports)/GDP for claimant state	.85	See Notes 39 & 41
210	Claim_FDI_Inward_GDP	(Inward FDI)/GDP for claimant state	.2	See Notes 39 & 41
211	Claim_FDI_Outward_GDP	(Outward FDI)/GDP for claimant state	1.2	See Notes 39 & 41
212	Resp_2011GDP	Gross domestic product of respondent's home country in 2011	\$12,000	See Notes 39 & 41

Not ICSID or Compustat Variables

Number	Name	Description	Sample Value	Notes
		dollars		
213	Resp_Trade_GDP	(Imports + Exports)/GDP for respondent state	.9	See Notes 39 & 41
214	Resp_FDI_Inward_GDP	(Inward FDI)/GDP for respondent state	.3	See Notes 39 & 41
215	Resp_FDI_Outward_GDP	(Outward FDI)/GDP for respondent state	.6	See Notes 39 & 41
216	Claim_Dem	Does the claimant have a Polity2 score greater than or equal to 6?	0. No 1. Yes	See Note 35
217	Resp_Dem	Does the respondent have a Polity2 score greater than or equal to 6?	0. No 1. Yes	See Note 35
218	Public_provision	Does the treaty or contract require an outcome be made public?	0. No 1. Yes	
219	Arb_pres_GDP2000	GDP of arbitration president's country in 2000 dollars	\$25,000	See Notes 39 & 41
220	Arb_pres_GDP2011	GDP of arbitration president's country in 2011 dollars	\$28,000	See Notes 39 & 41
221	Arb_pres_FDIin	(Inward FDI)/GDP for arbitration president's state	.6	See Notes 39 & 41
222	Arb_pres_Polity	Polity2 score of country of arbitration president	8	See Note 35
223	Arb_pres_IMFGroup	IMF country classification for arbitration president's country	1. High income 2. Middle income 3. Low income	See Note 36
224	Arb_pres_experience	Number of prior cases on which arbitration president has served	10	
225	Arb_2_experience	Number of prior cases on which second panelist has served	5	
226	Arb_3_experience	Number of prior cases on which third	11	

Not ICSID or Compustat Variables

Number	Name	Description	Sample Value	Notes
		panelist has served		
227	Claim_year_created	Year the claimant was founded	1921	Not applicable if individual
228	Claim_age	Years claimant existed before filing claim	83	Not applicable if individual
229	Resp_CPI	Respondent's Corruption Perceptions Index in year case is constituted	80	0-100 scale
230	Resp_CPI_Rank	Respondent's rank on CPI	25	Ties allowed
231	Claim_Bribe	Has the claimant signed the OECD's Convention on Combating Bribery of Foreign Public Officials in International Business Transactions?	0. No 1. Yes	Compare entry into force of the Convention to date case registered
232	Resp_Bribe	Has the respondent signed the OECD's Convention on Combating Bribery of Foreign Public Officials in International Business Transactions?	0. No 1. Yes	Compare entry into force of the Convention to date case registered
233	Claim_FOTPKey	Key used to match claimant for Freedom of the Press, based on country and year arbitration completes	UnitedStates1990	No values before 1993
234	Claim_FOTP	Freedom of the Press score in claimant country when arbitration completes	0-100	No values before 1993
235	Resp_FOTPKey	Key used to match respondent for Freedom of the Press, based on country and year arbitration	New Zealand1990	No values before 1993

Not ICSID or Compustat Variables

Number	Name	Description	Sample Value	Notes
		completes		
236	Resp_FOTP	Freedom of the Press score in respondent country when arbitration completes	0-100	No values before 1993
237	Industry_PublicLosses_Total	Total number of public losses ever for the claimant's industry	3	
238	Industry_PublicWins_Total	Total number of public wins ever for the claimant's industry	2	
239	Industry_CasesFiled_Total	Total number of cases ever for the claimant's industry	12	
240	Industry_PublicLosses_Running	Total number of cases filed before the case was registered	25	
241	INGO_Claim	The number of international NGOs in the claimant state in the year the case is filed	2658	
242	INGO_Resp	The number of international NGOs in the respondent state in the year the case is filed	1254	
243	FreedomFromCorruption_Claim	Heritage Foundation's freedom from corruption score in year case is registered, claimant	82	
244	FreedomFromCorruption_Resp	Heritage Foundation's freedom from corruption score in year case is registered, resp	50	
245	WGI_CC_Claim	World Governance Indicators' Control of Corruption for claimant state in year case is registered	1.34	
246	WGI_CC_Resp	World Governance Indicators' Control of Corruption for claimant	-.29	

Not ICSID or Compustat Variables

Number	Name	Description	Sample Value	Notes
		state in year case is registered		
247	Inflation_C	World Bank GDP Deflator annual inflation percent of the claimant the year case is filed	3.23	
248	Inflation_R	World Bank GDP Deflator annual inflation percent of respondent the year case is filed	1.23	
249	FDI_NatDiversity_C	Diversity of claimant's FDI in year claim is filed	3.61	From Wellhausen 2015; coverage is 1990-2008
250	FDI_NatDiversity_R	Diversity of respondent's FDI in year claim is filed	1.93	From Wellhausen 2015; coverage is 1990-2008

Notes to Coding

1. There are two primary sources for arbitration documents. The first source to check is ICSID's website. It publishes .pdfs of tribunal documents that are made public during or after a tribunal's proceedings. ICSID does not always publish all documents that are available. The coder should always check www.italaw.com as well, as that website often has documents ICSID does not. If a document cannot be found at either website, one can assume it is not in the public domain.
2. Some cases are thrown out on jurisdiction grounds, meaning many of the variables should be coded as missing. For example, ICSID ARB/05/9 was thrown out for lack of jurisdiction, so Awarded, Spoils, Interest_sought, and Interest_calc are all coded as missing. A case thrown out on jurisdiction grounds is coded as the claimant having lost.
3. There are many arbitration cases in which more than one party files. In other words, there are cases with multiple claimants. Each claimant receives its own line of coding in the dataset, though most of the variables will be the same. For example, ICSID/ARB/02/9 has 5 claimants. It therefore receives 5 lines in the dataset. Each line has the same data except for claim_ind and ind_name.
4. If an arbitration case has individuals as claimants, the Claim_firm variable is their name. Claim_ind and Ind_name are filled out as well so that Claim_firm equals ind_name.
5. The Count_parties variable needs to count firms and individuals. For example, ICSID ARB/02/9 should be coded a 5, not a 2 (there are 3 individuals, and 2 companies). Though each row only represents one claimant, Count_parties will still equal 5 (or whatever value is required for the case in question).
6. The sought variable should code what the claimant is seeking, not what the tribunal decides is a valid measurement to seek. For example, in ICSID ARB/93/1 the claimant seeks \$10.8 million but the tribunal says that only \$4.45 million of that is with merit. The value of sought in this case is nonetheless the claimant's \$10.8 million.
7. Claimants are not always precise with the amount of money they seek. Sometimes they offer a range of possible valuations, often based on different calculation procedures. In cases where there is no clear dollar figure, the sought value should be the average of the given figures and the coder should make a note of this ambiguity. For example, in ICSID/ARB/99/2, the claimant seeks \$50-\$70 million dollars, so sought is coded at \$60 million and there is a note explaining this coding decision.
8. The 9 variables Indiv_opinion_p through Io3_reason should only have values in very specific instances. The _opinion parts should only receive a 1 if there is a separate document with their opinion. If that document agrees with the tribunal's decision - it is a concurring opinion - then the _support variable is a 1. If it does not - a dissenting opinion

- then the `_support` variable is a 0. The `_rsn` variable is for the coder to go into some detail about what the panelist's separate opinion actually says. If there are no separate documents, which is most often the case, then the proper coding is 0--0--0--. If the case is still pending, it is not possible for the panelists to have released any opinions, so the final nine values should be .-.-.-.-.-.-.-.-.-.

9. Many panels are reconstituted, which means that the original 3 panelists are not the same as the 3 panelists who issue opinions. This often arises due to death or objections of one of the parties. In these cases, the `Date_const.` variable should be the very first date the panel was constituted.

10. If there is a panelist from the Netherlands, their country is coded as Holland.

11. To code variables 39-47, keyword searches were performed. If the document was an image, the entire document was read. For example, searching for "expropriation" will show you where in the document, if at all, claims about (in)direct expropriation are made. For each document, searches for "expropriation", "indirect", "full protection", "arbitrary", "discrim", "currency", "stable", "national", "umbrella", and "transfer" were performed. The results were then double-checked by finding the part of the document where the claimant lays out specific claims.

12. `ObsObl` is only coded for if the claimant specifically references a umbrella clause from the BIT in question.

13. An annulment is when the respondent seeks to vacate the tribunal's decision. A supplementary decision is when the claimant is not happy with the tribunal's decision and wants a new decision. For example, ICSID/ARB/01/03 has annulment, supplementary, and resubmission stages. The resubmission stage is when either party is not happy with any of the appeals and wants a completely new tribunal process. There is nowhere in the codebook to code for resubmission. If a resubmission occurs, the coder should make note of it in the Notes column of the workbook.

14. `Date_ended` is the date all proceedings have finished, not the date that one stage happens to have ended. For example, ICSID/ARB/93/1 has an award rendered on February 21st, 1997. A revision proceeding occurred almost two years later, however, and that process did not end until a settlement on July 26th, 2000. The proper code for `Date_ended` here is therefore July 26th, 2000. If there is still a stage ongoing, this variable's value is missing.

15. If the tribunal's award embodies a settlement the parties reached, then `Judge_rule` is always code a 4, even if the settlement clearly favors one party over the other.

16. Direct expropriation occurs when the state seizes the assets of the claimant. The state becomes the new holder of the property. Indirect expropriation is when the nominal property rights holder remains an individual or firm, but the individual or firm cannot

utilize its property as originally intended because of changes in the state's laws. Indirect expropriation can take the form of new regulations, tax increases, or forcing the individual or firm to use the property in a certain way.

17. Stable legal environment means that the respondent state has changed the laws or contracts surrounding an investment without a level tantamount to expropriation. For example, in ICSID/ARB/11/28, the state of Turkey unilaterally voided the claimant's contract, awarded new work to the claimant's subsidiary but refused to pay the agreed upon rates, and insisted the claimant complete work faster than the timeframe asked for in the contract. Because these actions seek to change the terms of the contract without actually seizing the claimant's assets, they should be coded StabLeg = 1.

18. A resubmission occurs when either party disagrees with a tribunal's full or partial annulment of an award. That party files a resubmission to seek to overturn all or part of the annulment. Resubmissions occur in front of a new tribunal.

19. A revision is when either party seeks to change an award upon discovery of a new fact that was unknown to the Tribunal and filing party at the time of the award.

20. Jrsdctn_res should only be coded as 1 if the case is thrown out on jurisdiction grounds. If a jurisdiction objection was raised and the case is still ongoing, code Jrsdctn_res as a 0. The presumption is the case would have ended already if ICSID lacked jurisdiction.

21. If settled is coded as a 1 - a settlement was reached - then winner is coded as missing unless the case is a conciliation proceeding (in which case winner="N/A". The logic is that a settlement will have favored one side over the other, so without knowing the settlement's text we cannot judge which party actually won even though one did.

22. If a settlement occurs at any point during a proceeding, then Settled = 1. For example, ICSID/ARB/03/11 had an award issued but annulment proceedings started. During annulment, a settlement was reached. Settle therefore equals 1.

23. If an arbitrator issues a dissenting opinion but still signs the award document in favor of the claimant, s/he is coded a 1 (voting for the claimant). If the panel rules against the claimant and the arbitrator dissents but has signed the award, then the panelist is still coded as 0 (against claimant).

24. Winner is coded as 1 (the claimant wins) when a majority of the panel (panels are almost always 3 arbitrators) supports the claimant at the end of the merit phase. Support is determined by reading the decision on the merits of a case's award.

25. Stock_1 codes a company's primary stock symbol. This corresponds to the symbol for the exchange where the company is headquartered. If the company is not listed in the

same jurisdiction as its headquarters, choose the exchange where it first listed. Stock_2 is for a company's second listing, Stock_3 for the third.

26. Belgium and Luxembourg sign BITs together. This means that any company registered in either country has a Claim_state value of "Belgium-Luxembourg".

27. 12 entries have a Count_parties value of "Unknown". These are cases where ICSID has listed the claimant as "[NAME] and others" and no other documentation is available that provides the others' names. For example, ARB/12/12 is "Vattenfall AB and others", ARB/01/2 is "Antoine Goetz and others".

28. In two cases, the Claim_state value is also "Unknown". In ARB/09/18, the claimant is Cambodia Power Company. The company was created as part of a World Bank project, and the amount of private sector participation is unclear; there is very little case information available, so it is unclear who actually owns the CPC. It is possible that a contract dispute is what allows the company, nominally Cambodian, to sue Cambodia. The second case is ARB/04/11. Russell Resources International Limited is the named claimant. The case has very little published information, and Googling does not find a correct Russell Resources. The only companies Google returns are small American ones not involved in mining. While the claim_state is most like the United States because the respondent is the D.R.C., no confirmatory evidence is available.

29. If the case is settled, the tribunal issues an award embodying the settlement, and that award lists the amount of money transferred, that amount is coded in the Money_receive and Settle_amt variables.

30. If settled is coded as 1 (settlement occurred), then Rule43, the proceeding variable, is usually 1 as well. (Rule 43 stipulates that a proceeding is discontinued if both parties agree in writing, usually because of a settlement.) If Rule43 is 0 while settled is still 1, then Rule 44 was invoked (it allows for discontinuance if one party requests it and the other does not respond after a set time limit). If Rule43 is a 1 while settled is a 0, that means ICSID's documentation does not explicitly say "settlement agreed by the parties"; it will say something along the lines of, "The Tribunal issues an order taking note of the discontinuance of the proceeding pursuant to ICSID Arbitration Rule 43(1)."

31. If an ICSID case is labeled "CONC(AF)/YY/##", then ICSID_AF=0 and CONC=1. The logic is that this is first and foremost a conciliation proceeding, not an arbitration, and ICSID_AF is designed to capture just arbitrations.

32. The ISO Codes are the International Standard Organization's country codes. They are included in order to make plots easier to produce and merging across datasets easier. Some country names, like Russia and Taiwan, were changed to match the ISO Code name. For Belgium-Luxembourg, Belgium's ISO code was used because it is the larger country.

33. If the tribunal is new and so has not been constituted or the case was closed before a tribunal was constituted (so Date_const is missing), then arbitration is also missing.

34. All data is current as of 06.01.2012.

35. Decision equals 1 if winner equals 0 or 1, 0 otherwise.

35. As of 10.31.2011, there are not Polity2 scores for 2012. Values for 2011 were used instead. ICSID recognizes Belgium and Luxembourg as the same state during arbitration, so the Polity2 score for Belgium was used. Barbados, Grenada, and Hong Kong are not in Polity's dataset. For pre-reunification Germany, Germany West's scores were used.

37. IMF country classifications come from *Classifications of Countries Based on Their Level of Development: How it is Done and How it Could be Done* by Lynge Nielsen, IMF Working Paper WP/11/31. Taiwan and Hong Kong are not in the IMF data.

38. Turkmenistan is not in the DoingBusiness data (used for the Enforcement_) variables. Gabon, the Seychelles, Trinidad and Tobago, Guyana, Gambia, Liberia, Grenada, Saint Kitts and Nevis, Gabon, Liberia, and Iceland are in the DB data but do not have values for the Enforcement_ variables. The DoingBusiness data only goes back to 2003; for cases filed before then, Doing Business values from 2003 were used. See <http://www.doingbusiness.org/Methodology/enforcing-contracts#cost> for the World Bank's notes on its Enforcing_ measures.

39. Consolidated equals zero if the case was not merged or is the one that was merged into. When two cases are consolidated, the newest case is coded as 1, the "parent" case as a 0. If two cases were registered on the same date and are then consolidated, the one with the later ICSID case ID is coded as 1. See ARB(AF)/04/3 and ARB(AF)/04/4 for examples.

40. The economic variables in "Not ICSID or Compustat Variables" coding come from the World Bank's World Development Indicators unless otherwise noted (the categorical income variables come from the IMF).

41. ARB/04/11 is Russell Resources International vs. the Democratic Republic of Congo. Very little information exists for RRI. It was coded as being based in Malta given an address found through Google.

42. All national level investment and trade data (Resp_FDI_Inward etc.) comes from the World Bank's World Development Indicators and is for the year in which the case was registered.

43. We coded CONC/83/1, ARB(AF)/98/1, and CONC/03/1 as Rule43 == 1 even though ICSID's website did not specifically reference this rule. For CONC/83/1, ICSID refers to Conciliation Rule 33, which discusses communicating conciliation outcomes. CONC/03/1 was settled before the commission started, so no rule is invoked. According to ICSID, the outcome of ARB(AF)/98/1 was an "Award embodying the parties' settlement agreement

rendered on September 18, 2000". This is the only instance of ICSID not invoking Rule 43 or 44, so we interpret it as an oversight on ICSID's part.

44. ARB/06/15 concerns a settlement that occurred before the claimants filed the ICSID case. The respondent objected to jurisdiction because of the existence of a settlement between the respondent and claimants. ICSID ruled in favor of the respondent. Settled and Rule43 were originally coded as 1 but updated to 0.

45. ARB(AF)/04/2 is coded as ongoing because it quickly left ICSID. The tribunal was constituted on November 2, 2004 and settled on April 4, 2005. The settlement moves the parties over to using UNCITRAL's arbitration rules, and ICSID records no further procedural details. The case was originally coded as having concluded on April 4, but we feel the case does not fit in the dataset since it quickly left ICSID.

46. ICSID identifies 4 cases as being settled but does not mention Rule43 in the "Outcome of Proceeding" section. We coded these cases as Rule43==1 because we are unaware of any instance where a case could be settled outside of the Rule43 channel. The cases are ICSID/ARB/97/5, ICSID/97/8, ICSID/99/8, and CONC/82/1

47. Cat_code is based on Product, where Product is how ICSID describes the claimant's industry on the case information page. We consolidated ICSID's coding because each ICSID description is clearly a specific instantiation of a more general class, e.g. there are both "electricity concessions" and "electricity distribution concession", both of which are coded as "Electricity and electric infrastructure" in our data. See the "Product_CatCode" sheet in the data workbook to see which ICSID product categories fall into which of our category codes.

48. Before 2012, the Corruption Perceptions Index was on a 0-10 scale. Starting in 2012, scores are shown on a 0-100 scale. Our data multiplies all data before 2012 by 10 to make comparable to 2012 data. There is no CPI data before 1995, and our dataset has no cases constituted in 1995, so the first year of coverage is 1996. Countries missing from the CPI are:

- Grenada (2010, 2005, 1998)
- Togo (2005)
- Seychelles (2003)
- Burundi (2002, 1996)
- Guyana (2002)
- United Arab Emirates (2002)
- Congo, The Democratic Republic of (2001, 2000, 1999)
- Mali (2001)
- United States of America (2001, 2000, 1999)
- Georgia (2000)
- Honduras (2000)
- Sri Lanka (2000)
- Liberia (1999)
- Gambia (1999)

- Burkina Faso (1997)
- Slovakia (1997)
- Albania (1996)
- Saint Kitts and Nevis (1996)
- Papua New Guinea (1996)

49. For measures with Belgium and Luxembourg, the values for Belgium were used to match to ICSID's "Belgium-Luxembourg".

50. For the Freedom of the Press measure, Cyprus(Greece) was used for ICSID's "Cyprus".

51. Freedom of the Press measures come from Freedom House.

Notes to Merging with Compustat

Once each firm-claimant was coded from ICSID (n=583), Zachary merged the data with information available from Compustat's North America Fundamentals Annual (hereforth known as Compustat North America) as well as the Global Fundamentals Annual (Compustat Global) database. Many variables that Compustat covers were not pulled from their database; for a complete list of Compustat variables, please visit http://wrds-web.wharton.upenn.edu/wrds/tools/variable.cfm?library_id=7. The merging proceeded as follows.

First, Zachary used Google to see if each claimant was a publicly listed company. This searching was used to complete variables `stock_1`, `stock_2`, and `stock_3`. When a claimant used. For example, Mobil Cerro Negro Holding, Ltd (ARB/07/27) was coded as `stock_1=XOM`, the symbol for Exxon Mobil (the parent company). www.investing.businessweek.com proved essential to tracking corporate ownership chains.

Second, Zachary used Compustat data from 1974-2011 (Compustat North America) and 1987-2011 (Compustat Global) to verify that the public companies from step 1 were found in Compustat. If a match was found in Compustat Global, then `GlobalMatch?` = 1. If the match was found in Compustat North America, then `NAMatch?` = 1. A few public companies were initially coded as .5 to indicate many possible matches. Zachary then undertook more detailed research on each company before deciding on a 1 or 0; for more detail, see the second Notes column (column CR). The authors can also provide earlier versions of the data to show the coding steps where `GlobalMatch?` or `NAMatch?` equaled .5. `GlobalMatch?` and `NAMatch?` are never both 1, though they can both equal 0. 216 firm-arbitration events were matched with one of the Compustat datasets.

Third, Zachary copied and pasted from Compustat the data for the matching firm-year. For example, the claimant in ARB/09/14 is Maersk Olie, a subsidiary of A.P. Moeller – Maersk A/S; from Compustat Global, Zachary copied the A.P. Moeller - Maersk A/S values for 2009. This process was completed for all 216 matches.

Fourth, not every firm-arbitration event had an exact match. No Compustat annual data exists for 2012, so the newest claimants do not have exact matches. In addition, some companies do not enter Compustat when they file for arbitration or have left Compustat by the time they file arbitration. In both cases, the closest year was used. To easily know which companies did not have year matches, Zachary created the variable `YearMatch`; of the 216 matches, 189 have a year match.

Fifth, Zachary created the `StockMatch` variable to easily separate the Compustat Global from Compustat North America firm-arbitration events. Compustat North America provides a ticker symbol whereas Compustat Global does not, so if `StockMatch` = 1 then the company is in the North America dataset.

Sixth, there is some missing data among the publicly traded companies. Standard Chartered is only in Compustat Global Security Daily, so it has no Compustat information even though

it is a publicly traded company; Standard Chartered is the only instance in the dataset for which this happened. Moreover, Compustat does not cover TSX.V, a Canadian stock exchange for junior (aka "venture") companies, those too small to list on Toronto's TSX. This lack of coverage means four firm-arbitration events for public companies do not have Compustat matches: Niko Resources (x2), Vanessa Ventures Ltd., and TG World Petroleum limited.

Seventh, the pulls from Compustat Global and Compustat North America did not cover exactly the same data. Zachary added missing variables to each dataset so the North American and Global companies would have the same number of Compustat variables. To Compustat Global, Zachary added tic, CURRTR, CI, COGS, EBIT, GP, LCAT, NI, NPAT, OIADP, OIBDP, PIDOM, PIFO, RDIP, UGI, XLR, XOPR, SALE, SEQ, TB, TCOE, and TCOR. To Compustat North America, Zachary added ISIN, SEDOL, FSRCOPT, EXRE, and OPPRFT. Added variables are coded with "N/A", e.g. Vivendi Universal (ARB/97/3) is in Compustat Global, so Non-performing Assets (PIFO) = N/A.

Eighth, Zachary combined the Compustat Global and Compustat North America firm-arbitration events into one dataset. Before this step, they were two worksheets because of the different variables in each.

Ninth, Zachary created the MatchKey variable in preparation for combining the 216 Compustat firm-arbitration events with the rest of the firm-arbitration events. MatchKey was created for all 583 firm-arbitration events and 216 firm-arbitration events with Compustat matches; the overlap of 216 will be the 216 Compustat matches.

Tenth, Zachary used the MatchKey and Excel's VLOOKUP() formula to merge the Compustat data into the larger ICSID data. Excel returned "#N/A" for all firm-arbitration events which did not have matches in Compustat; Zachary found all these entries and replaced them with "NotinCompustat".

Eleventh, Zachary added metadata to rows 588-590 to make it easy to understand data availability. Count Available counts the number of data points for each column of data; its formula is $=583-\text{COUNTIF}(D5:D587,"")-\text{COUNTIF}(D5:D587,"N/A")-\text{COUNTIF}(D5:D587,"Unknown")$. 583 is the number of firm-arbitration events; the first count if subtracts out blank values, the second N/A, and the third Unknown. Percent Available then divides Count Available by 583. For the Compustat columns, the Count Available formula is $583-\text{COUNTIF}(CW5:CW587,"")-\text{COUNTIF}(CW5:CW587,"N/A")-\text{COUNTIF}(CW5:CW587,"Unknown")-\text{COUNTIF}(CW5:CW587,"NotinCompustat")$. The fourth countif() computation was added to subtract out all the non-matches. A third row of metadata, Percent of Max, was added to these columns to correct the low Percent Available amounts. The largest value of Percent Available is 37.0497%, i.e. there are only 216 matches ($216/583=.370497$). Percent of Max divides each column's percent available by 37.0497; the resulting values range from 1% to 100% and represent how much data is available for those firm-arbitration events with Compustat matches.

Intercoder Reliability

Shu Shang

The first intercoder reliability session occurred on Friday, January 20th, 2012. Shu Shang, the new coder, was asked to code 10 cases that Zachary, the original coder, had already coded. After reviewing the first four cases, it became clear that coding patterns were significantly different. Much of the difference owed to coding rules that Zachary had internalized but not recorded in the codebook, and the codebook was updated accordingly. Shu Shang was a law student at the University of Southern California who responded to an advertisement placed there.

The second session occurred on Friday, February 3rd. Of the remaining 6 cases, the coding match rate was 76%; of 294 cells of information required, disagreement existed on 70. The variables for which the two coders disagree at least 50% of the time were Jrsdctn_res, Sought, Interest_sought, Awarded, Interest_awarded, Spoils, and Supplementary.

The third session occurred on Friday, February 10th. On the first pass, the coders disagreed on 59 out of 250 possible values, for a total match rate of 76.4%. After discussing responses, it was decided that the original coder had miscoded 17 responses. Correcting those led to a success rate of 83%. 25 disagreements came from ICSID ARB/02/9 because the second coder did not realize there was an award document, resulting in many "Unknown" values. The "Unknown" values would have been right had the document actually not been available. The second coder also consistently confused the coding of the settled variable, coding a "1" for every case even though none were actually settled. Had these been coded correctly, the match rate would have been 90%.

To avoid the possibility of significantly different interpretations of key variables, the two coders coded different parts of each case. Shu coded Case, Product, Claim_firm, Stock, Count_parties, Claim_state, Claim_assoc, Resp_state, Date_register, Date_const., Arbitration, Arb_pres, Arb_2, Arb_3, Arb_prescountry, Arb_2country, Arb_3country, Settled, Settle_amt, Compl_ind, Ind_name, ICSID_AF, UNCITRAL, Private, Status, and Date_ended. Zachary coded the remaining variables.

Longhao Wang

On May 10th, 2012, Longhao Wang started to help Zachary with coding metadata for ICSID cases. Longhao coded the following variables:

1. Case
2. Product
3. Claim_firm
4. Count_parties
5. Claim_assoc
6. Resp_state

7. Date_register
8. Date_const
9. Arbitration
10. Arb_pres
11. Arb_2
12. Arb_3
13. Arb_prescountry
14. Arb_2country
15. Arb_3country
16. Settled
17. Compl_ind
18. Ind_name
19. ICSID_AF
20. Status
21. Date_ended

Zachary reviewed Longhao's first five cases. They were completed with 100% accuracy. Longhao coded cases ICSID/ARB/06/02 through ICSID/ARB/01/10.

Longhao Wang was a law student at the University of California – Irvine and responded to an advertisement for a part-time research assistant.

Catherine Nicklen

Eight cases had material available only in Spanish. Since Zachary, David, and Emilie cannot read Spanish, they hired Catherine Nicklen, a graduate student at UCSD's School of International Relations and Pacific Studies, to code these cases. Zachary trained her on ARB/06/01, a case in English. Catherine coded cases ARB/01/10, ARB/02/10, ARB/03/07, ARB/04/07, ARB/06/19, ARB/07/06, ARB/07/32, and ARB/08/10. We were not able to undertake interceder reliability on these cases, but Catherine and Zachary were in constant communication. Zachary reviewed the coded cases to make sure they were internally consistent. After two rounds of review, the cases were completed to satisfaction.

Catherine Nicklen was a 2nd-year masters student at the School of International Relations and Pacific Studies at the University of California – San Diego. She responded to an advertisement for a native Spanish speaker to assist translating and coding arbitration documents.