Is More Better or Worse?

New Empirics on Nuclear Proliferation and Interstate Conflict by Random Forests

Online Appendix

Akisato Suzuki

School of Law and Government

Dublin City University

Dublin 9, Ireland

akisato.suzuki@gmail.com

How Random Forests works

Random Forests uses an ensemble of categorization or regression trees for prediction (Strobl, Malley, and Tutz, 2009: 325). Categorization is for categorical dependent variables while regression is for continuous dependent variables. Each tree analyzes a bootstrap sample from the data and randomly chooses a specified number of regressors out of all regressors in the equation; then all results are aggregated (by voting for classification and by averaging for regression) to make a prediction (Liaw and Wiener, 2002: 18; Strobl, Malley, and Tutz, 2009: 331-334).

I use 1,000 bootstrap samples and set the number of regressors to be used in each tree, *mtry*, by the default setting for regression (the total number of regressors / 3 = 2). For robustness checks, I also set *mtry* at 3, 4, 5, and 6 – i.e. all regressors whereby the model is equivalent to Bagging (Breiman 1996). Only unipolarity and the number of democratic states are significantly sensitive to these different setting (the results available below). See also Strobl, Malley, and Tutz (2009: 333, 343).

Table A-1: Conditional variable importance (fatal dispute-state ratio)

	trial 1	trial 2	trial 3	trial 4	trial 5
# of nuclear states	yes	yes	yes	yes	yes
Nuclear year counter	no	no	no	no	no
# of democratic states	yes	yes	yes	yes	yes
Gross world products	yes	yes	yes	yes	yes
Unipolarity	yes	yes	yes	yes	yes
lagged dispute-state ratio	yes	yes	yes	yes	yes

Figure A-1: Partial dependence plot (fatal dispute-state ratio)

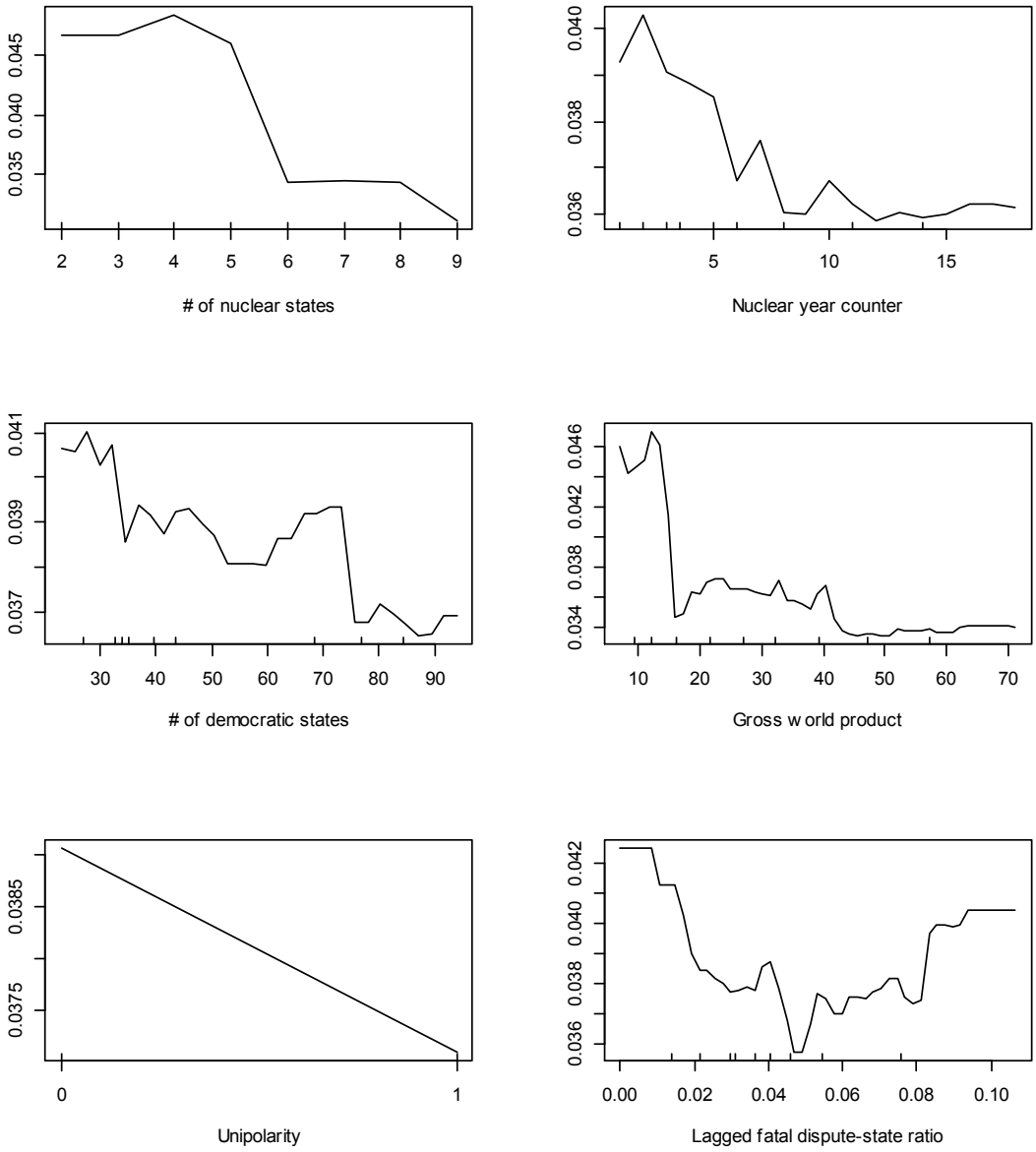

Table A-2: Conditional variable importance (war-state ratio)

	trial 1	trial 2	trial 3	trial 4	trial 5
# of nuclear states	yes	yes	yes	yes	yes
Nuclear year counter	no	no	no	no	no
# of democratic states	yes	yes	yes	yes	yes
Gross world products	yes	yes	yes	yes	yes
Unipolarity	yes	yes	yes	yes	yes
lagged dispute-state ratio	no	no	no	no	no

Figure A-2: Partial dependence plot (war-state ratio)

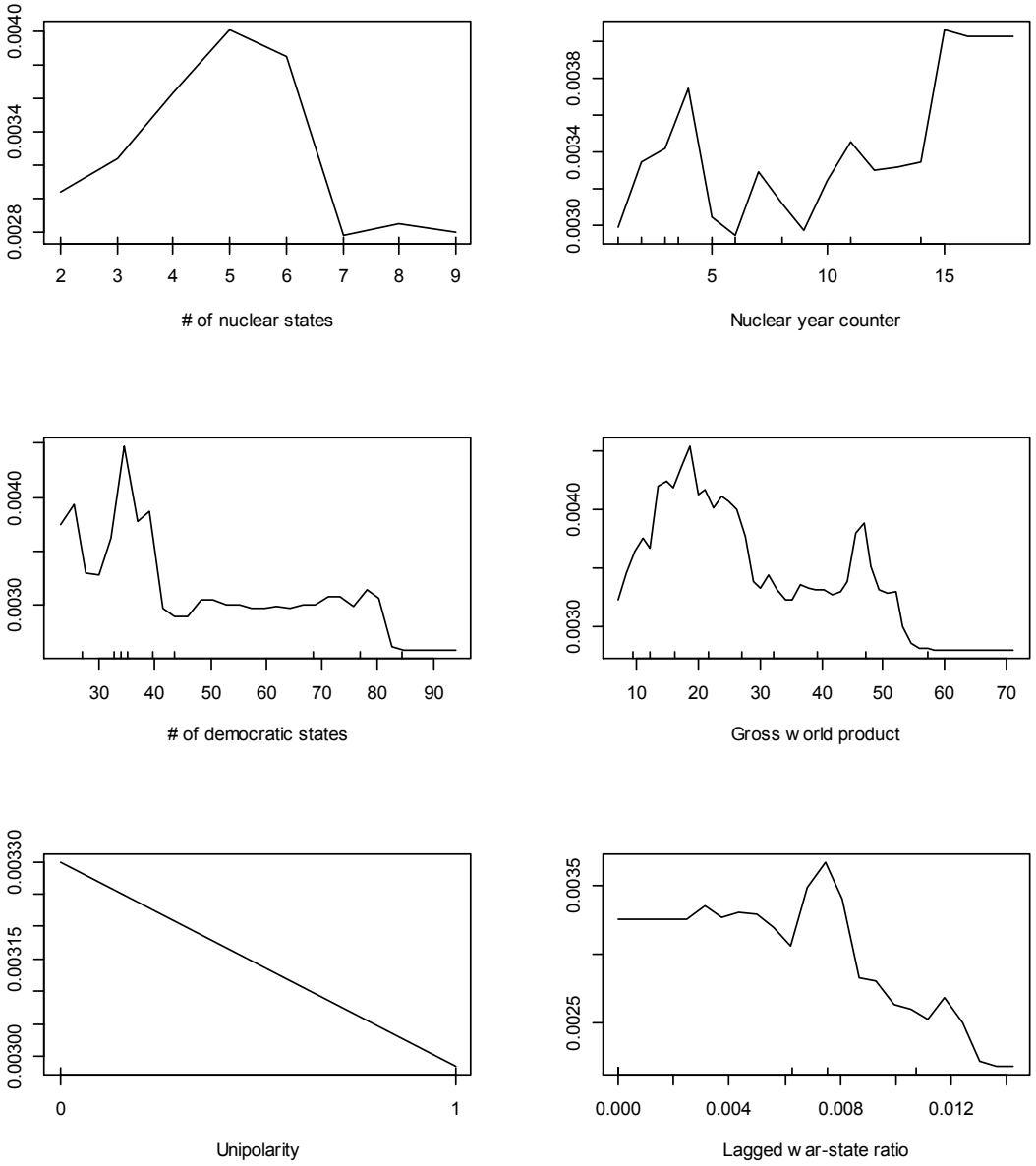

Table A-3: Conditional variable importance (North Korea not coded as a nuclear state)

	trial 1	trial 2	trial 3	trial 4	trial 5
# of nuclear states	yes	yes	yes	yes	yes
Nuclear year counter	no	no	no	no	no
# of democratic states	yes	yes	yes	yes	yes
Gross world products	yes	yes	yes	yes	yes
Unipolarity	yes	yes	yes	yes	yes
lagged dispute-state ratio	yes	yes	yes	yes	yes

Figure A-3: Partial dependence plot (North Korea not coded as a nuclear state)

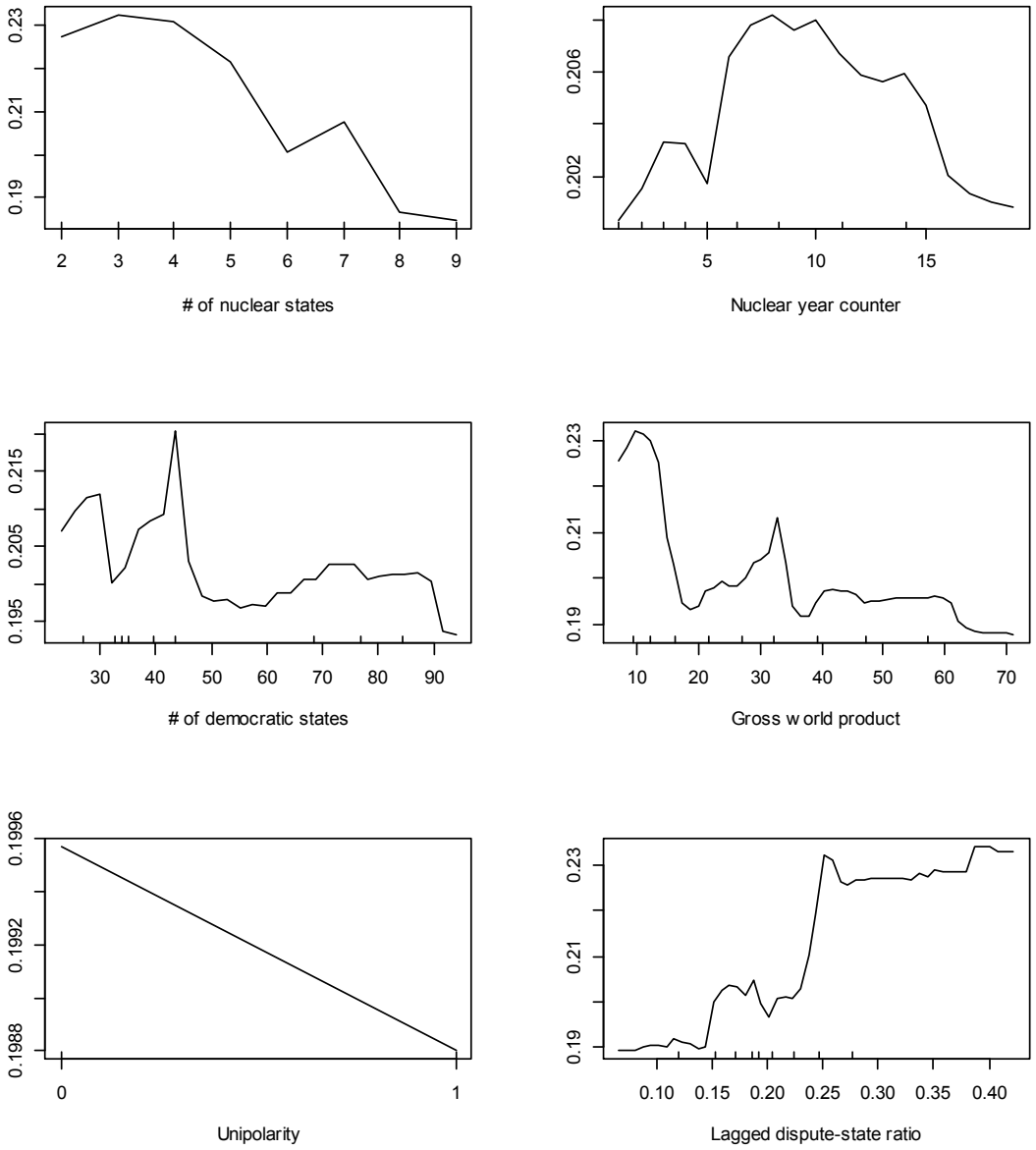

Table A-4: Conditional variable importance (using Way's [2012] coding of nuclear states)

	trial 1	trial 2	trial 3	trial 4	trial 5
# of nuclear states	yes	yes	yes	yes	yes
Nuclear year counter	no	no	no	no	no
# of democratic states	yes	yes	yes	yes	yes
Gross world products	yes	yes	yes	yes	yes
Unipolarity	yes	yes	yes	yes	yes
lagged dispute-state ratio	yes	yes	yes	yes	yes

Way (2012) codes Israel from 1969, South Africa from 1979-1991, and Pakistan from 1987 as nuclear states.

Figure A-4: Partial dependence plot (using Way's [2012] coding of nuclear states)

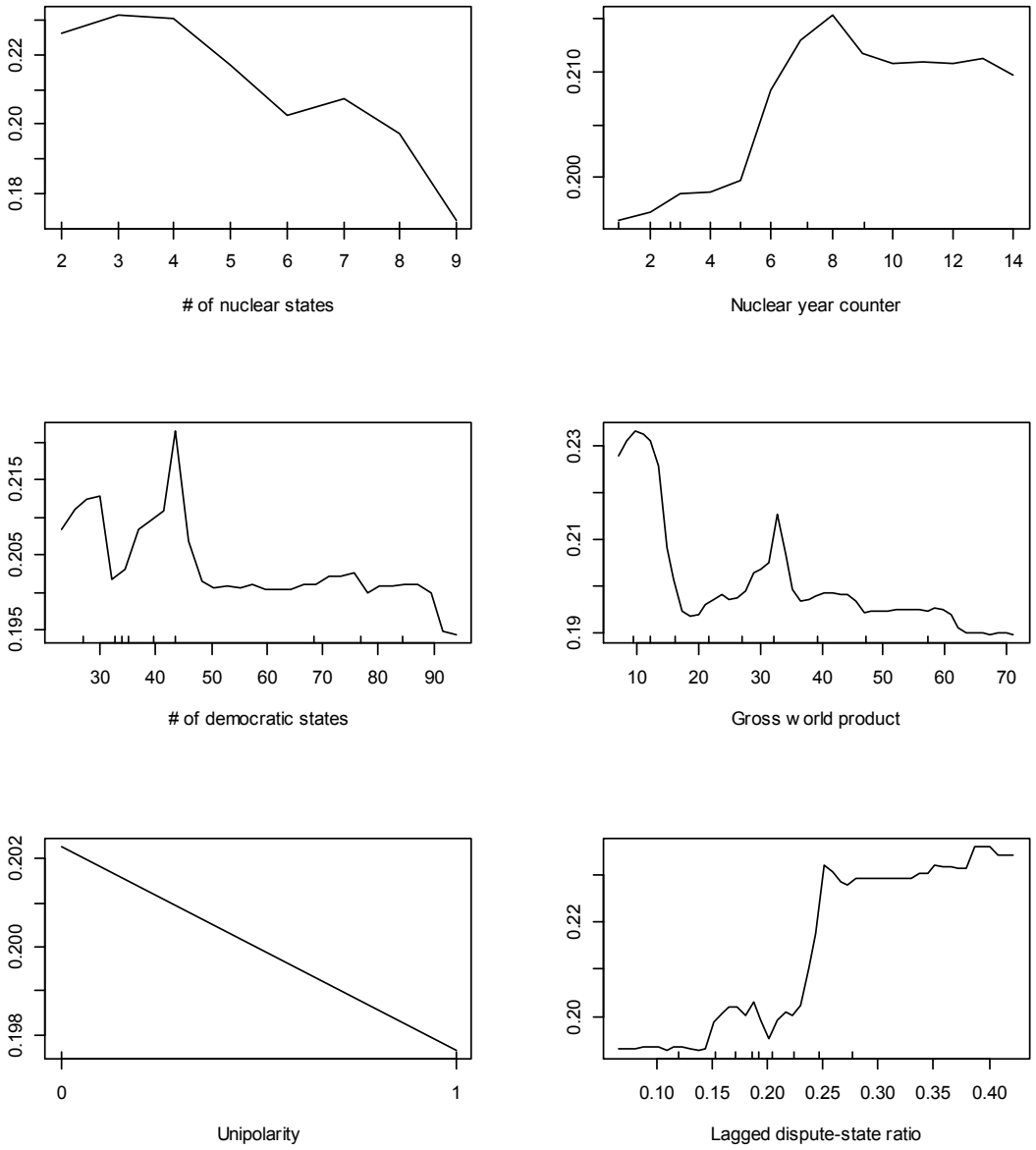

Table A-5: Conditional variable importance (*mtry* = 3)

	trial 1	trial 2	trial 3	trial 4	trial 5
# of nuclear states	yes	yes	yes	yes	yes
Nuclear year counter	no	no	no	no	no
# of democratic states	yes	yes	yes	yes	yes
Gross world products	yes	yes	yes	yes	yes
Unipolarity	yes	yes	yes	yes	yes
lagged dispute-state ratio	yes	yes	yes	yes	yes

Figure A-5: Partial dependence plot (*mtry* = 3)

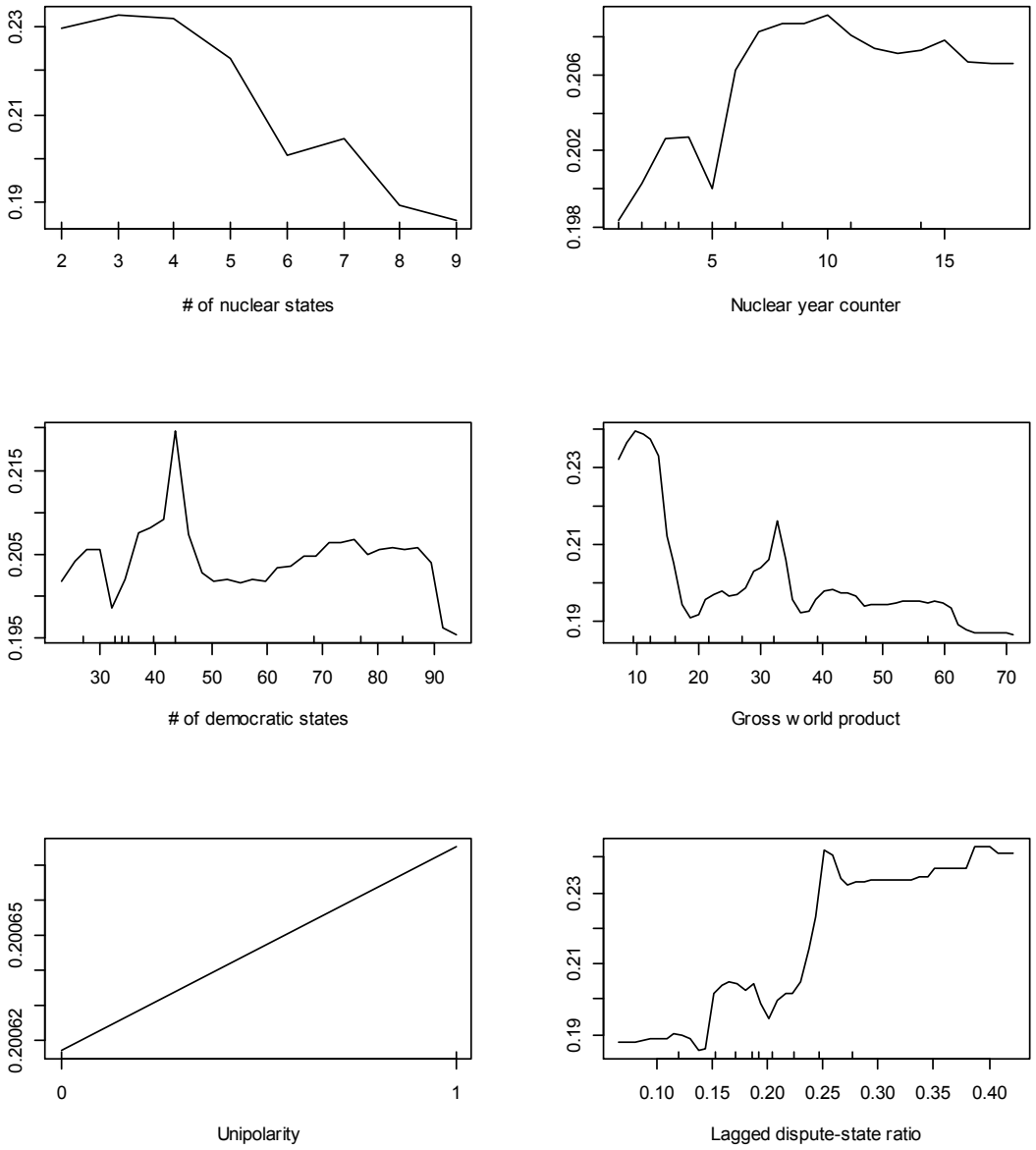

Table A-6: Conditional variable importance (*mtry* = 4)

	trial 1	trial 2	trial 3	trial 4	trial 5
# of nuclear states	yes	yes	yes	yes	yes
Nuclear year counter	no	no	no	no	no
# of democratic states	yes	yes	yes	yes	yes
Gross world products	yes	yes	yes	yes	yes
Unipolarity	yes	yes	yes	yes	yes
lagged dispute-state ratio	yes	yes	yes	yes	yes

Figure A-6: Partial dependence plot (*mtry* = 4)

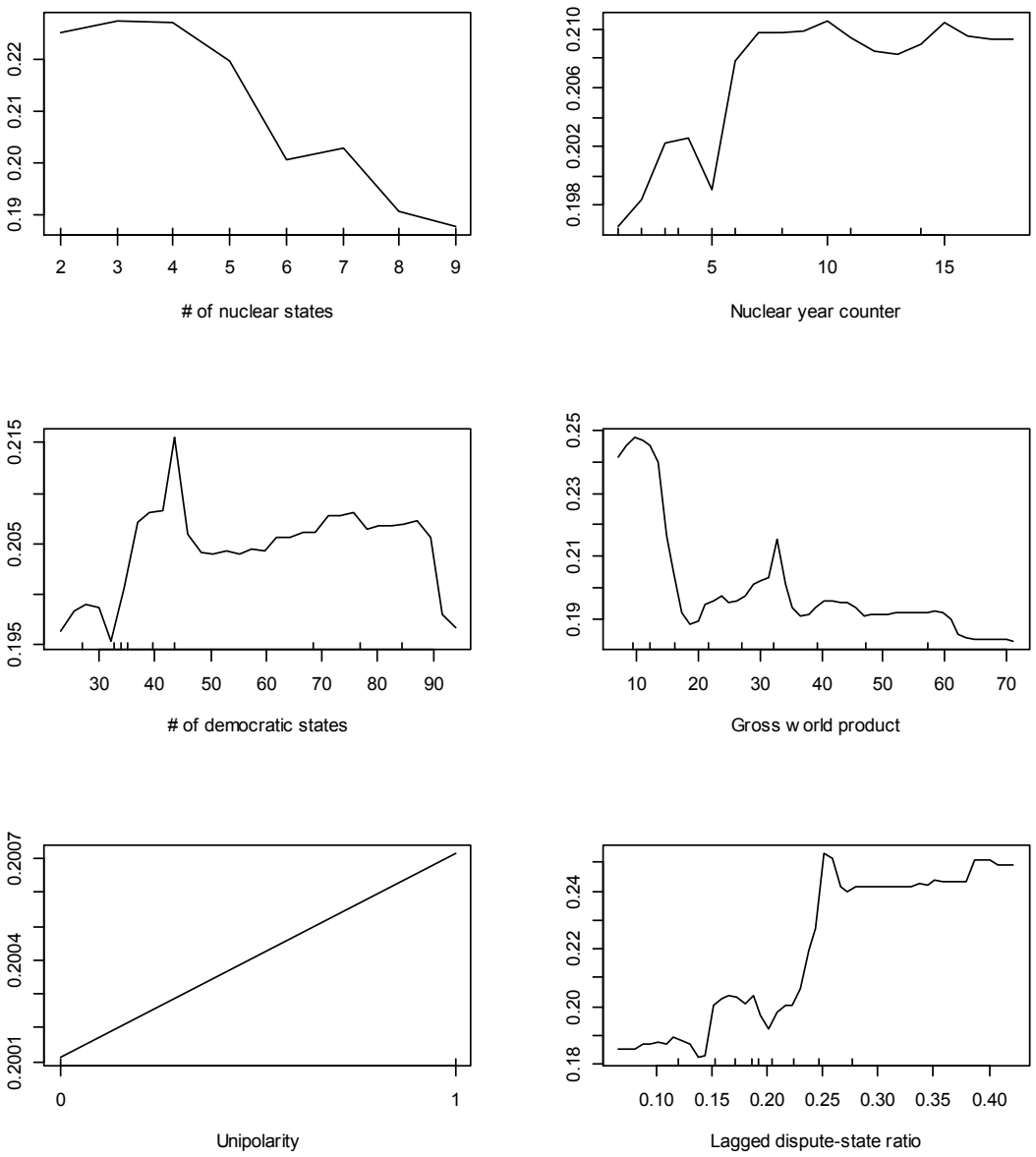

Table A-7: Conditional variable importance (*mtry* = 5)

	trial 1	trial 2	trial 3	trial 4	trial 5
# of nuclear states	yes	yes	yes	yes	yes
Nuclear year counter	no	no	no	no	no
# of democratic states	no	yes	yes	yes	no
Gross world products	yes	yes	yes	yes	yes
Unipolarity	yes	yes	yes	yes	yes
lagged dispute-state ratio	yes	yes	yes	yes	yes

Figure A-7: Partial dependence plot (*mtry* = 5)

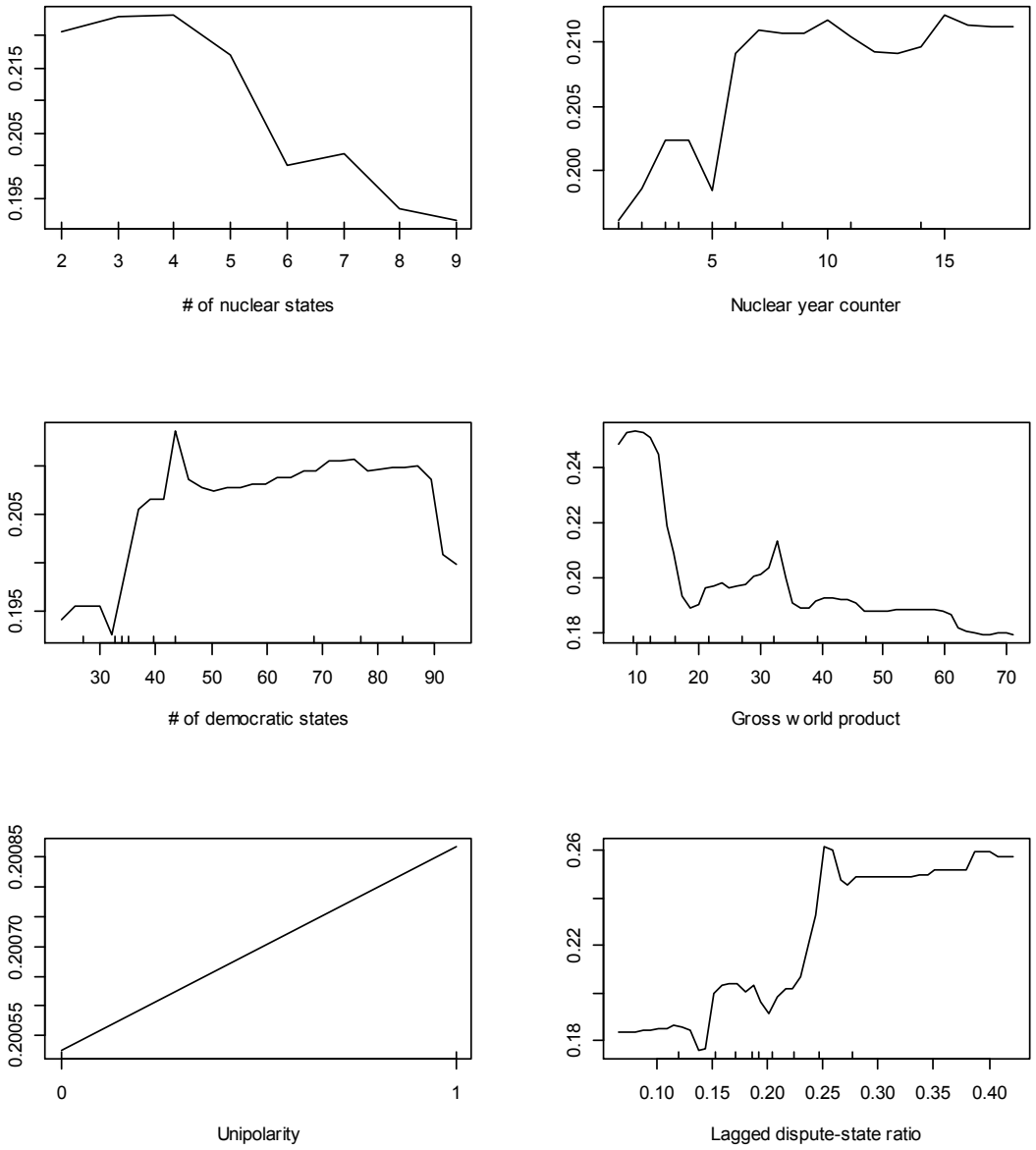

Table A-8: Conditional variable importance (*mtry* = 6)

	trial 1	trial 2	trial 3	trial 4	trial 5
# of nuclear states	yes	yes	yes	yes	yes
Nuclear year counter	no	no	no	no	no
# of democratic states	no	no	no	no	no
Gross world products	yes	yes	yes	yes	yes
Unipolarity	yes	yes	yes	no	yes
lagged dispute-state ratio	yes	yes	yes	yes	yes

Figure A-8: Partial dependence plot (*mtry* = 6)

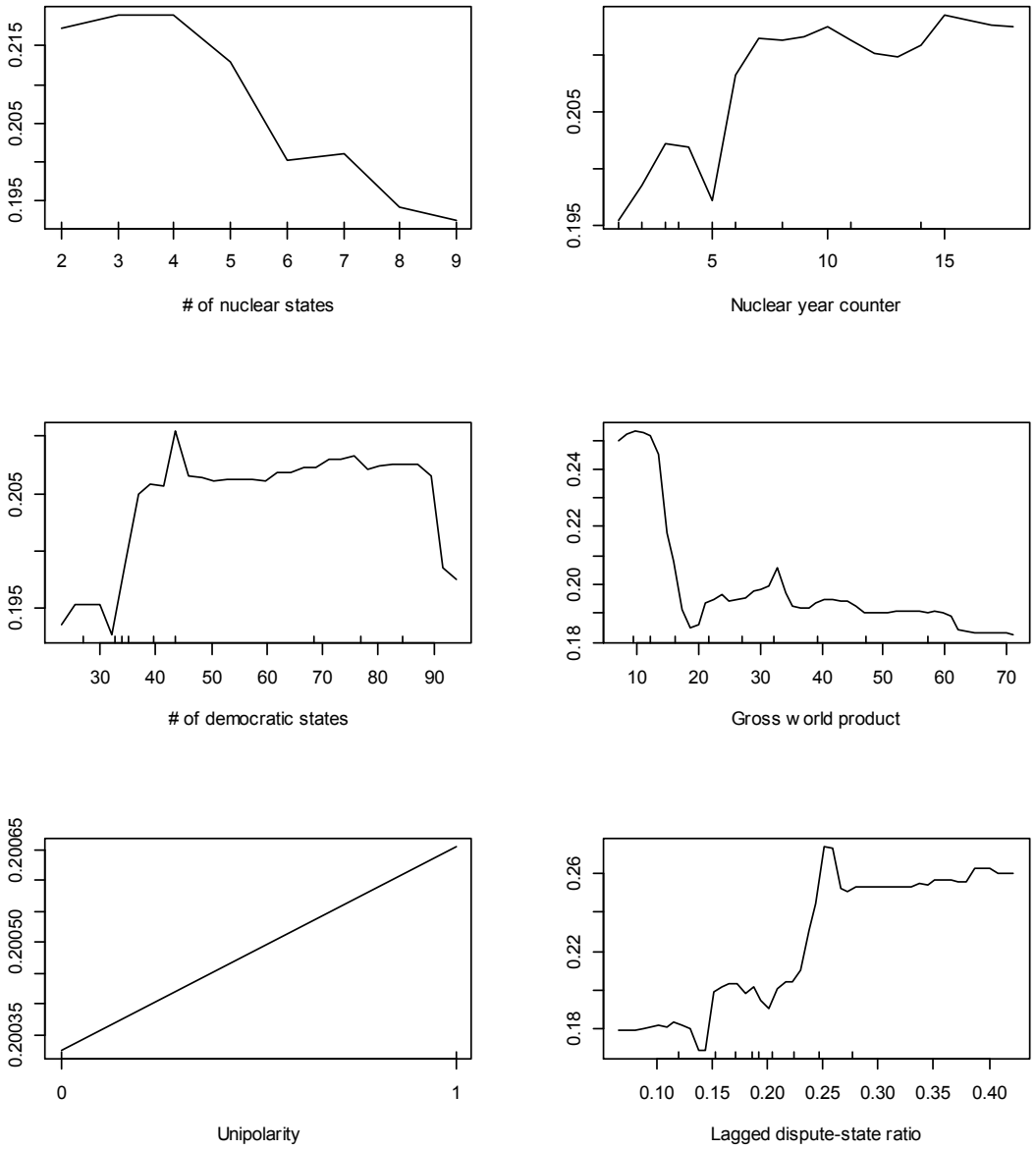

Table A-9: Correlation matrix of the regressors

	# of nuclear states	# of nuclear states (exc. N. Korea)	nuclear year counter	nuclear year counter (exc. N. Korea)	# of democratic states	growth world product	unipolarity
# of nuclear states	1.000	-	-	-	-	-	-
# of nuclear states (exc. N. Korea)	0.998	1.000	-	-	-	-	-
nuclear year counter	0.296	0.309	1.000	-	-	-	-
nuclear year counter (exc. N. Korea)	0.371	0.353	0.890	1.000	-	-	-
# of democratic states	0.808	0.801	0.479	0.568	1.000	-	-
growth world product	0.865	0.855	0.547	0.655	0.963	1.000	-
unipolarity	0.711	0.706	0.357	0.426	0.944	0.855	1.000

Table A-10: ARMA regression of the dispute-state ratio

	Model 1	Model 2	Model 3	Model 4	Model 5
# of nuclear states	-0.0197***				
	(0.00639)				
Nuclear year counter	-0.000680				
	(0.00139)				
# of nuclear states		-0.0185***			
(exc. N. Korea)		(0.00687)			
Nuclear year counter		-0.00184			
(exc. N. Korea)		(0.00149)			
# of democratic states			-0.00138***		
			(0.000413)		
Growth world product				-0.00205***	
				(0.000519)	
Unipolarity					-0.0559***
					(0.0199)
L.ar	0.217	0.152	0.338	0.297	0.383
	(0.380)	(0.346)	(0.284)	(0.336)	(0.494)
L.ma	0.0916	0.162	0.0504	0.0665	0.0260
	(0.380)	(0.361)	(0.334)	(0.373)	(0.576)
L2.ma	0.160	0.173	0.180	0.169	0.158
	(0.154)	(0.159)	(0.189)	(0.186)	(0.269)
Constant	0.325***	0.325***	0.269***	0.262***	0.218***
	(0.0430)	(0.0429)	(0.0278)	(0.0224)	(0.0170)
Observations	60	60	60	60	60

Semi-robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

The first order of autoregressive process and the second order of moving average are specified based on the results of a partial correlogram and a correlogram. The Phillips-Perron unit-root test rejects the null hypothesis that a unit root is present. The Portmanteau (Q) test for white noise indicates that there is no autocorrelation in the models.

References

Breiman L (1996) Bagging Predictors. *Machine learning* 24 (2): 123-40.

Liaw A and Wiener M (2002) Classification and Regression by randomForest. *R News* 2 (3): 18-22.

Strobl C, Malley J, and Tutz G (2009) An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests. *Psychological methods* 14 (4): 323-48.

Way C (2012) *Nuclear Proliferation Dates*. Available at: http://www.arts.cornell.edu/crw12/ (accessed 23 January 2015).