Supplementary material
Convergence and robustness checks

Francesco S. Bellelli, Ashar Aftab, Riccardo Scarpa*

23rd June 2021

In this appendix, we assess the robustness of the results by checking fundamental assumptions of the model. As a first point, we provide evidence that the sampled distribution has converged while in the second half we experiment with a diverse range of specifications to test the stability and validity of our findings.

1 Convergence

In this section we discuss the convergence of the reference model (Model I). A theoretical property of MCMC is that — independently from the starting values or complexity of the model — the distribution of MCMC chains are bound to converge on the target distribution after an infinite number of iterations. However, as the number of feasible iterations is finite, the question of whether or not the estimation has converged is fundamental to statistical inference.

Table 1 reports a series of statistics that are used to assess the convergence of the chains. The first three columns present the mean, median and mode of the marginal posterior distributions. In MCMC estimation all of these measures are commonly used for inference. The matching of median, mean and mode hints to the fact that the generated samples are normally distributed and suggest that the Markow Chains have converged. For all of the parameters of the model the mean, median and mode converged toward the same value lending a case in favour of convergence. The only exceptions being the variance parameters of the treaty and country effects. As a matter of fact, the posterior distribution of variance parameters is typically right skewed as illustrated by the histograms in figure 1. The value of the median is comprised between the mean and the mode due to the light skewness in the distribution.

The effective sample size (ESS) is one of the most popular statistics to assess the efficiency of the sampling algorithm. It quantifies the number of independent samples generated during estimation. A higher ESS indicates that the samples are less correlated. Unfortunately, multilevel models for survival data are complicated to estimate, as chains are highly correlated and typically do not mix well (Steele et al., 2004). We follow the advice of Browne et al. (2009) and use orthogonal parametrisation which considerably improves chain mixing. Given the type of data, we obtain a reasonable ESS. *Regional*

*Centre for Environmental and Energy Economics, Durham University Business School, University of Durham, Durham DH1 3LB, United Kingdom. Dr. Bellelli: (e-mail: f.s.bellelli@dunelm.org.uk); Dr. Aftab (e-mail: ashar.aftab@durham.ac.uk); Prof. Scarpa (e-mail: riccardo.scarpa@durham.ac.uk).

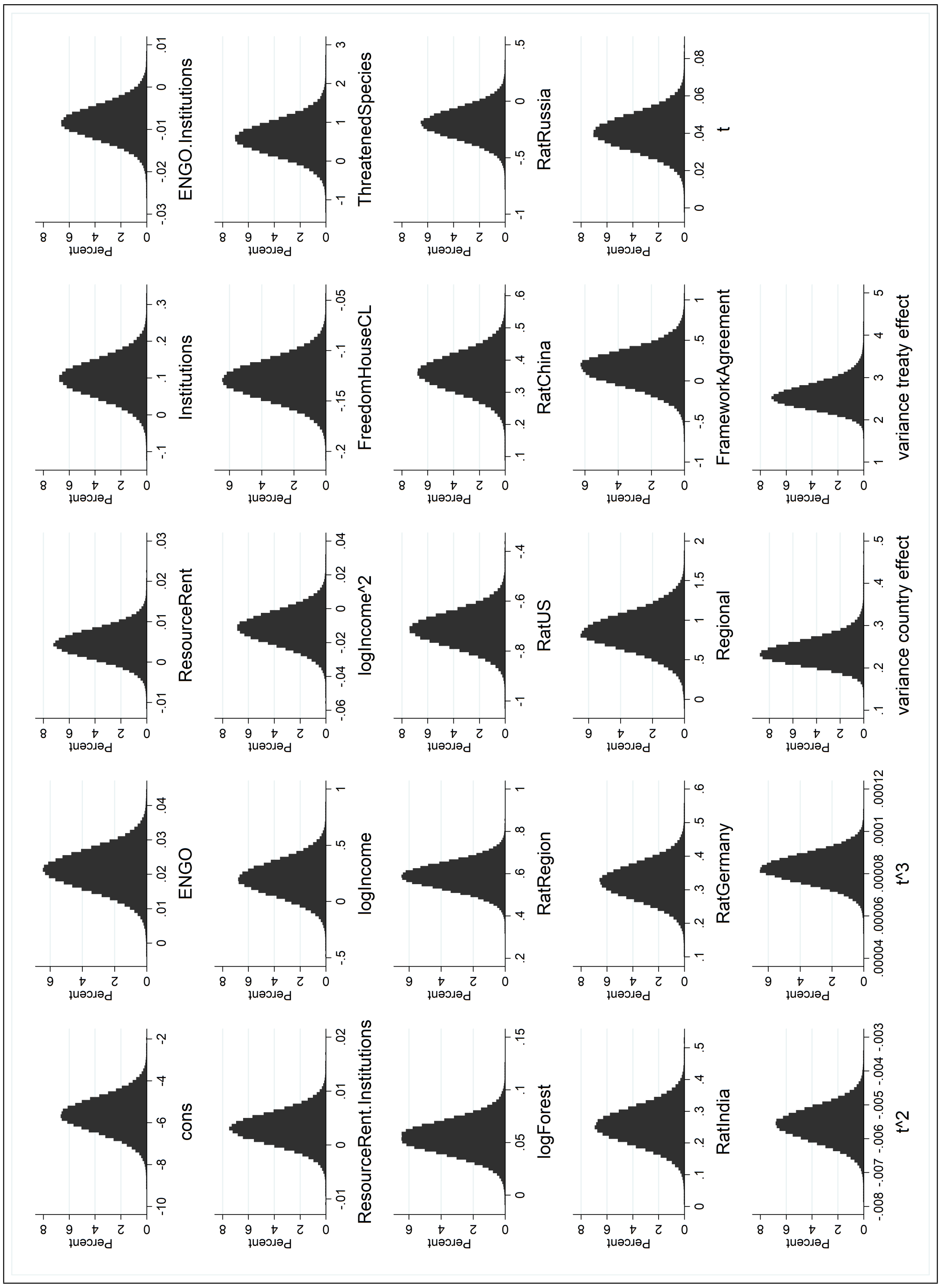

Figure 1: *Marginal posterior distributions*

Notes: Histograms for the distributions of the 24 chains of model I. The densities of MCMC chains are estimates of the marginal posterior distributions. As expected, the parameters of the model are normally distributed while the variance parameters are slightly skewed, with a longer right-tale.

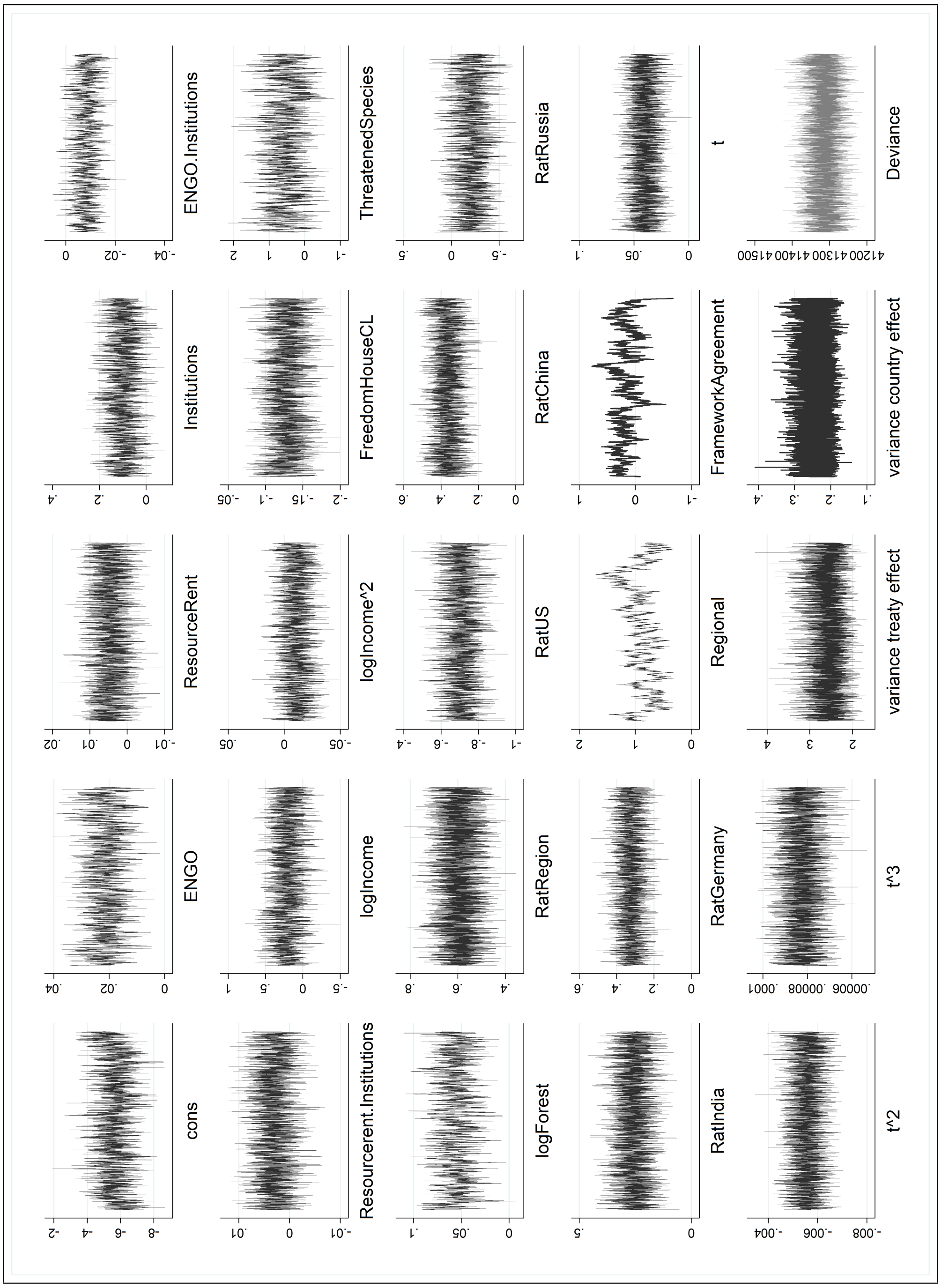

Figure 2: *Parameter traces for model I*

Notes: The traces illustrate the last 20,000 iterations of the MCMC estimation. They are commonly used to assess convergence and mixing of the chains. Traces that look like white noise processes indicate that the sampling algorithm moves efficiently through the distribution and that the chains have converged around a high probability region.

Table 1: Convergence statistics for model I

	Mean	Median	Mode	ESS	RL 2.5%	RL 97.5%	BD
$ENGO$	0.021	0.021	0.021	7712	34 890	36 648	12 792
$ResourceRent$	0.004	0.004	0.005	22 193	40 824	40 880	180 932
$Institutions$	0.098	0.098	0.098	18 287	42 696	45 664	366 395
$ENGO.Institutions$	−0.008	−0.008	−0.008	8870	34 570	36 038	535 418
$Resourcerent.Institutions$	0.003	0.003	0.003	25 073	41 272	39 856	108 403
$logIncome$	0.207	0.207	0.208	22 457	41 736	41 968	32 477
$logIncome^2$	−0.011	−0.011	−0.011	22 714	42 664	42 824	12 551
$FreedomHouseCL$	−0.129	−0.129	−0.129	24 221	24 342	23 462	480
$ThreatenedSpecies$	0.590	0.589	0.587	8147	36 246	37 898	675 411
$logForest$	0.055	0.055	0.055	7842	41 578	40 772	92 843
$RatRegion$	0.590	0.590	0.591	47 753	33 800	33 808	2790
$RatUS$	−0.714	−0.714	−0.714	18 307	43 784	43 368	6558
$RatChina$	0.361	0.361	0.361	33 000	37 656	34 504	3479
$RatRussia$	−0.199	−0.199	−0.198	18 056	25 222	25 426	30 591
$RatIndia$	0.254	0.254	0.253	39 846	34 608	34 784	2810
$RatGermany$	0.325	0.325	0.324	33 227	38 056	36 792	2732
$Regional$	0.863	0.862	0.864	533	92 810	119 068	1 231 190
$FrameworkAgreement$	0.171	0.171	0.169	901	121 388	101 160	1 599 008
t	0.041	0.041	0.041	37 539	35 488	35 608	7458
t^2	−0.006	−0.006	−0.006	34 421	37 144	35 936	2262
t^3	0.000	0.000	0.000	36 620	36 728	37 272	4393
$cons$	−5.643	−5.642	−5.644	8968	49 464	50 312	12 949
Variance *treaty* level ($\sigma_{u_{j_1}}$)	2.584	2.563	2.537	79 637	31 656	31 856	303
Variance *country* level ($\sigma_{u_{j_2}}$)	0.239	0.237	0.235	98 590	30 448	31 136	247

Total iterations	800000	Burnin	300000
Stored Chain length	250000	Thinning	2

Notes: ESS is the Effective Sample Size statistic. ESS assesses the chains on the base of their correlation. "RL 2.5%" and "RL 97.5%" are the Raftery-Lewis statistics (Raftery & Lewis, 1992) for the 2.5% and 97.5% percentiles. The Raftery-Lewis statistcs are estimated for a margin of error of 0.005 with a probability of 95%. "BD" is the Brooks-Draper statistic; it is calculated for $k = 2$ significant figures and a significance level of $\alpha = 0.05$.

and $FrameworkAgreement$ exhibit the lowest ESS because of the small amount of independent observations compared to variables at the country and ratification level. For $Regional$ and $FrameworkAgreement$ the effective variability is much lower than for other variables. Since they are not time-varying, the only variability is across treaties. The actual number of observations is around 250 independent observations. This is much less than country variables which have approximately 5000 country-year combinations for the post-1990 period, and even less than variables at the ratification level which can rely on observations from more than 200000 separate treaty-country-year dyads. Given the limitations associated with the type of data and the available information on treaties, we postulate that the ESS results are satisfactory.

Estimation of multilevel survival models with MCMC notoriously yields highly correlated chains (Browne et al., 2009). For this reason we opt for a very high number of iterations. In total we perform almost one million iterations, out of which we discard one every two samples, for a total of 550000 samples generated. This practice is called "thinning" and is used to reduce the autocorrelation in the chains. We also choose a very long burn-in period. In fact, we discard the initial 300000 out of 550000 samples to make sure the inference is based on a chain that has converged. Likewise, the number of iterations has been selected in an overly prudent fashion in order to base inference on a number of samples that is as large as possible. The large number of samples allows more precise estimates for the coefficients of the model.

As a supplementary graphical check, in figure 2 we present traces of the last 20000 iterations of the MCMC estimation. These traces show that the chains seem to have converged around a mean and that they explore efficiently the joint distribution. Healthy trace plots resemble a white noise process, with a constant mean and variance. We also notice that *Regional* and *FrameworkAgreement* mix less efficiently than other chains, but well enough considered the available information on treaties. Overall, the traces seem to indicate that the chains have converged and that the algorithm is mixing well.

Additional diagnostic statistics are reported in the last columns of table 1. The Raftery & Lewis (1992) statistic gives information on the number of iterations needed to yield estimates of the 2.5 and 97.5 percentile, which together form an interval containing 95% of the distribution. This statistic is used as a diagnostic to assess convergence and also measures the precision of quantile estimates from the posterior distribution. The Raftery-Lewis statistic is known to be conservative and usually suggests more iterations than necessary (Browne, 2004). If the statistic is satisfied the actual quantile distribution (0.025,0.975) of the parameter should be less than 1% different from the estimated probability. All our chains satisfy the Raftery-Lewis diagnostic.

Finally, the Brooks-Draper diagnostic is a statistic for the mean of the posterior distribution. It estimates the iterations required to achieve estimates of the mean with a given level of significance and a desired number k of significant figures. In the last column of table 1 we report the number of iterations needed to quote mean estimates with a precision of 2 significant figures and a confidence level of 95%. For example, the Brooks-Draper statistic for the parameter of *ENGO* implies that 12792 iterations are needed to express the mean estimate “0.02” with 95% of confidence; we run a total of 800000 iterations, well above the recommended number. The only two parameters for which the recommended level is not reached are *Regional* and *FrameworkAgreement*; again, given the lower number of independent observations, it is harder to obtain high levels of precision for these parameters. Nevertheless, if the test is run with $k = 1$ the chains satisfy the requirements of the Brooks-Draper diagnostic; in fact, the Brooks-Draper statistic for *FrameworkAgreement* is 15991 and 12312 for *Regional* when $k = 1$.

To further test the convergence of the chains we follow Gelman & Rubin (1992) who suggest starting estimation from several different points in order to ensure the algorithm explores the entire joint distribution. We experiment with 5 different starting points. The starting values are obtained by multiplying the vector of starting values of Model I by a scalar taking the values of -1, 0, 2, 3, and 4. The starting values of Model I were obtained by fitting a simplified version of the model that does not take into account the cross-classification in random effects with a maximum likelihood method (IGLS)[2]. The results are reported in table 2, they display consistent estimates even with different starting values. Using different starting points allows us to rule out pseudo-convergence — that is to say, the convergence towards a local point of high probability. This is

[2]The starting values for the fixed part of Model I are: 0.0171443, 0.0048971, 0.1881271, -0.0116787, 0.0027066, 0.3452093, -0.0196376, -0.0938002, 0.3799708, 0.0635263, 0.432282, -0.7225903, 0.3376975, -0.299422, 0.1899845, 0.2431447, 0.5060046, 0.2540413, 0.0199538, -0.0041608, 0.0000641, -5.096715.

Table 2: *Convergence from different starting points*

	Model C1		Model C2		Model C3		Model C4		Model C5	
	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
ENGO	0.021***	(0.006)	0.021***	(0.006)	0.021***	(0.006)	0.021***	(0.006)	0.021***	(0.006)
ResourceRent	0.004	(0.004)	0.005	(0.004)	0.004	(0.004)	0.005	(0.004)	0.005	(0.004)
Institutions	0.100**	(0.050)	0.097**	(0.051)	0.098**	(0.051)	0.098**	(0.050)	0.098**	(0.050)
ENGO.Institutions	-0.008**	(0.004)	-0.008**	(0.004)	-0.008**	(0.004)	-0.008**	(0.004)	-0.008**	(0.004)
Resourcerent.Institutions	0.003	(0.003)	0.003	(0.003)	0.003	(0.003)	0.003	(0.003)	0.003	(0.003)
logIncome	0.205*	(0.154)	0.207*	(0.155)	0.203*	(0.152)	0.210*	(0.155)	0.203*	(0.152)
$logIncome^2$	-0.011	(0.010)	-0.011	(0.010)	-0.011	(0.010)	-0.012	(0.010)	-0.011	(0.010)
FreedomHouseCL	-0.129***	(0.019)	-0.129***	(0.019)	-0.129***	(0.020)	-0.129***	(0.019)	-0.129***	(0.019)
ThreatenedSpecies	0.597*	(0.437)	0.590*	(0.440)	0.588*	(0.448)	0.601*	(0.438)	0.590*	(0.437)
logForest	0.055***	(0.015)	0.055***	(0.015)	0.055***	(0.015)	0.055***	(0.016)	0.055***	(0.016)
RatRegion	0.591***	(0.063)	0.590***	(0.063)	0.590***	(0.063)	0.590***	(0.063)	0.591***	(0.063)
RatUS	-0.711***	(0.069)	-0.711***	(0.069)	-0.712***	(0.069)	-0.713***	(0.069)	-0.714***	(0.069)
RatChina	0.361***	(0.058)	0.361***	(0.058)	0.361***	(0.057)	0.360***	(0.058)	0.361***	(0.058)
RatRussia	-0.197*	(0.134)	-0.196*	(0.133)	-0.198*	(0.133)	-0.199*	(0.132)	-0.199*	(0.133)
RatIndia	0.254***	(0.058)	0.253***	(0.058)	0.253***	(0.058)	0.255***	(0.058)	0.254***	(0.057)
RatGermany	0.323***	(0.052)	0.324***	(0.053)	0.325***	(0.052)	0.324***	(0.053)	0.324***	(0.052)
Regional	0.892***	(0.220)	0.907***	(0.236)	0.889***	(0.220)	0.842***	(0.239)	0.906***	(0.236)
FrameworkAgreement	0.158	(0.231)	0.170	(0.224)	0.141	(0.235)	0.165	(0.226)	0.150	(0.229)
t	0.041***	(0.010)	0.041***	(0.010)	0.040***	(0.010)	0.041***	(0.010)	0.041***	(0.009)
t^2	-0.006***	(0.000)	-0.006***	(0.000)	-0.006***	(0.000)	-0.006***	(0.000)	-0.006***	(0.000)
t^3	0.000***	(0.000)	0.000***	(0.000)	0.000***	(0.000)	0.000***	(0.000)	0.000***	(0.000)
cons	-5.660***	(0.806)	-5.689***	(0.815)	-5.629***	(0.815)	-5.650***	(0.835)	-5.652***	(0.800)
Random part										
Variance *treaty* level	2.577	(0.304)	2.587	(0.305)	2.584	(0.306)	2.587	(0.305)	2.589	(0.306)
Variance *country* level	0.239	(0.031)	0.239	(0.030)	0.239	(0.031)	0.239	(0.030)	0.239	(0.030)
Units: *treaty*	257		257		257		257		257	
Units: *country*	190		190		190		190		190	
Obs: *ratification*	219266		219266		219266		219266		219266	
k	-1		0		2		3		4	
DIC:	41707.30		41707.34		41706.98		41707.58		41707.47	
Burnin:	200000		200000		200000		200000		200000	
Chain Length:	150000		150000		150000		150000		150000	
Thinning:	2		2		2		2		2	

Notes: ***, ** and * indicate one-tailed Bayesian p-values respectively lower than 0.01, 0.05 and 0.10. Estimates for the same model with different starting values. The starting values are obtained by multiplying the vector of starting values of Model I by the scalar k. The starting values of Model I are obtained by fitting with IGLS a simplified version of the model that does not take into account the cross-classification in random effects.

particularly dangerous in multimodal distributions.

Overall the Raftery-Lewis and Brooks-Draper diagnostics suggest that we have run the MCMC simulation for long enough to achieve a stable convergence. Nonetheless, the estimates for $FrameworkAgreement$ and $Regional$ should be quoted with a lower level of precision to guarantee the same confidence level. The traces of the chains suggest that the algorithm is mixing well and explores the distribution with sufficient efficiency. ESS statistics confirm this conclusion and indicate that the chains have generated a sample large enough to make a reliable inference. Finally, the graphical representation of the marginal posterior distributions, as well as the values of mean, median and mode suggest that the distribution of the chain values has converged to the target distribution. By using different starting values we rule out the possibility of pseudo-convergence and ensure that the chains have converged on the point of highest density.

2 Robustness checks

Firstly, we assess the sensitivity of the estimates for the other main variables in the same way it was done for industrial lobbying. The first four models of table 3 employ different proxies for environmental lobbying and the quality of institutions in order to evaluate the consistency of the results when different measurements are used.

The first model replaces the proxy $ENGO$ with $ProtectedArea$. $ProtectedArea$ is defined as the percentage of territory that is designated as protected area. We postulate that when environmental lobbying is more influential it succeeds in providing protection to a greater territorial area. Like $ENGO$, $ProtectedArea$ exhibits a positive and significant relationship with the hazard of ratification. Countries with higher $ProtectedArea$ have a higher probability of ratifying environmental agreements. The second proxy for environmental lobbying is $EnvConcern$. $EnvConcern$ is the percentage of the population that reports being concerned about climate change in a survey conducted in 2008 by Tien et al. (2015) on respondents from 119 different countries. The number of respondents in each countries varies between 500 and 8000. The assumption behind the use of this variable is that environmental lobbying correlates with the public's concern for the environment. This relationship could be affected by cultural and political factors but, in general, a higher environmental concern should result in stronger environmental pressure. The estimate for $EnvConcern$ is positive and statistically significant at the 10% level.

Both $EnvConcern$ and $ProtectedArea$ indicate that stronger environmental lobbying increases the chances of ratifying environmental agreements. These results are consistent with those of model I.

We also experiment with two additional indices for the quality of institutions. $Institutions2$ is the "Government Effectiveness indicator" from the World Governance Indicators (World Bank, 2017) and $Institutions3$ is the "Economic Freedom index" by Fraser Institute (2017). These indicators aggregate several aspects of institutional quality into one measure and are consistently calculated for a large number of countries and years. Both indicators are similarly constructed but have been compiled by different organ-

Table 3: *Robustness checks*

	ProtectedArea		**EnvConcern**		**Institutions2**		**Institutions3**		**Logit**		**Non-parametric**		**No country RE**	
	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
$ResourceRent$	-0.007	(0.005)	0.003	(0.004)	0.004	(0.004)	0.004	(0.004)	0.005	(0.004)	0.007*	(0.004)	0.005***	(0.002)
$ENGO$					0.024***	(0.006)	0.025***	(0.006)	0.023***	(0.006)	0.021***	(0.006)	0.018***	(0.002)
$EnvConcern$	0.005*	(0.003)												
$ProtectedArea$			0.018***	(0.003)										
$Institutions$	0.305**	(0.174)	0.127**	(0.058)					0.107**	(0.054)	0.061*	(0.046)	0.200***	(0.028)
$Institutions2$					0.223***	(0.050)								
$Institutions3$							0.236***	(0.050)						
$EnvConcern.Instit$	-0.004**	(0.002)												
$ProtectedArea.Instit$			-0.004*	(0.003)										
$ENGO.Institutions$									-0.008**	(0.004)	-0.009**	(0.004)	-0.013***	(0.001)
$ENGO.Institutions2$					-0.010***	(0.004)								
$ENGO.Institutions3$							-0.010***	(0.004)						
$ResourceRent.Instit$	-0.003	(0.004)	0.003	(0.003)					0.003	(0.003)	0.005*	(0.003)	0.003***	(0.002)
$ResourceRent.Instit2$					0.002	(0.003)								
$Resourcerent.Instit3$							0.002	(0.003)						
$logIncome$	0.023	(0.187)	0.232*	(0.164)	0.184	(0.152)	0.213*	(0.152)	0.261*	(0.165)	0.221*	(0.157)	0.353***	(0.091)
$logIncome^2$	0.002	(0.012)	-0.014*	(0.010)	-0.010	(0.010)	-0.010*	(0.010)	-0.015*	(0.010)	-0.010	(0.010)	-0.020***	(0.006)
$FreedomHouseCL$	-0.129***	(0.025)	-0.121***	(0.020)	-0.110***	(0.019)	-0.110***	(0.019)	-0.142***	(0.021)	-0.135***	(0.020)	-0.097***	(0.012)
$ThreatenedSpecies$	1.101**	(0.624)	0.605*	(0.464)	0.536	(0.426)	0.546	(0.435)	0.689*	(0.481)	0.475	(0.441)	0.343***	(0.178)
$logForest$	0.061***	(0.021)	0.056***	(0.015)	0.054***	(0.015)	0.053***	(0.015)	0.061***	(0.016)	0.058***	(0.016)	0.066***	(0.006)
$RatRegion$	0.481***	(0.080)	0.594***	(0.063)	0.635***	(0.063)	0.599***	(0.063)	0.658***	(0.069)	0.764***	(0.070)	0.433***	(0.059)
$RatUS$	-0.689***	(0.082)	-0.717***	(0.069)	-0.600***	(0.070)	-0.710***	(0.069)	-0.770***	(0.074)	-1.065***	(0.073)	-0.738***	(0.067)
$RatChina$	0.379***	(0.070)	0.357***	(0.058)	0.495***	(0.059)	0.360***	(0.058)	0.395***	(0.062)	0.394***	(0.059)	0.357***	(0.058)
$RatRussia$	-0.199	(0.165)	-0.206*	(0.135)	-0.230**	(0.136)	-0.190*	(0.134)	-0.225*	(0.145)	-0.148	(0.133)	-0.286***	(0.129)
$RatIndia$	0.229***	(0.071)	0.252***	(0.058)	0.390***	(0.057)	0.256***	(0.057)	0.273***	(0.062)	0.217***	(0.058)	0.196***	(0.057)
$RatGermany$	0.313***	(0.063)	0.332***	(0.053)	-0.250***	(0.054)	0.327***	(0.052)	0.353***	(0.056)	0.151***	(0.053)	0.256***	(0.052)
$Regional$	0.764***	(0.231)	0.862***	(0.227)	0.819***	(0.237)	0.882***	(0.227)	0.914***	(0.237)	0.697***	(0.242)	0.830***	(0.233)
$FrameworkAgreement$									0.186	(0.225)	0.199	(0.233)		
t	0.059***	(0.012)	0.038***	(0.010)	0.073***	(0.009)	0.041***	(0.010)	0.039***	(0.010)			0.025***	(0.009)
t^2	-0.007***	(0.001)	-0.006***	(0.000)	-0.006***	(0.000)	-0.006***	(0.001)	-0.006***	(0.001)			-0.005***	(0.000)
t^3	0.000***	(0.000)	0.000***	(0.000)	0.000***	(0.000)	0.000***	(0.000)	0.000***	(0.000)			0.000***	(0.000)
$cons$	-5.501***	(1.070)	-5.670***	(0.846)	-5.280***	(0.791)	-5.420***	(0.794)	-5.985***	(0.878)	-19.418***	(9.331)	-5.937***	(0.465)

Robustness checks (continued)

	ProtectedArea		**EnvConcern**		**Institutions2**		**Institutions3**		**Logit**		**Non-parametric**		**No country RE**	
	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
Random part														
Variance *treaty* level	2.384	-0.288	2.575	-0.302	2.57	-0.302	2.591	(0.304)	2.885	(0.343)	2.978	(0.347)	2.359	(0.278)
Variance *country* level	0.216	-0.036	0.28	-0.036	0.218	-0.028	0.223	(0.029)	0.281	(0.036)	0.243	(0.031)		
Units: *treaty*	254		257		257		257		257		257		257	
Units: *country*	118		189		190		190		190		190		190	
Obs: *ratification*	132756		217848		219266		219266		219266		219266		219266	
DIC:	39814.16		55767.59		56002.29		55975.69		55900.92		40792.95		57746.34	
Burnin:	200000		200000		200000		200000		200000		250000		100000	
Chain Length:	250000		250000		250000		250000		250000		300000		200000	
Thinning:	2		2		2		2		2		2		2	

Notes: ***, ** and * indicate one-tailed Bayesian p-values respectively lower than 0.01, 0.05 and 0.10.

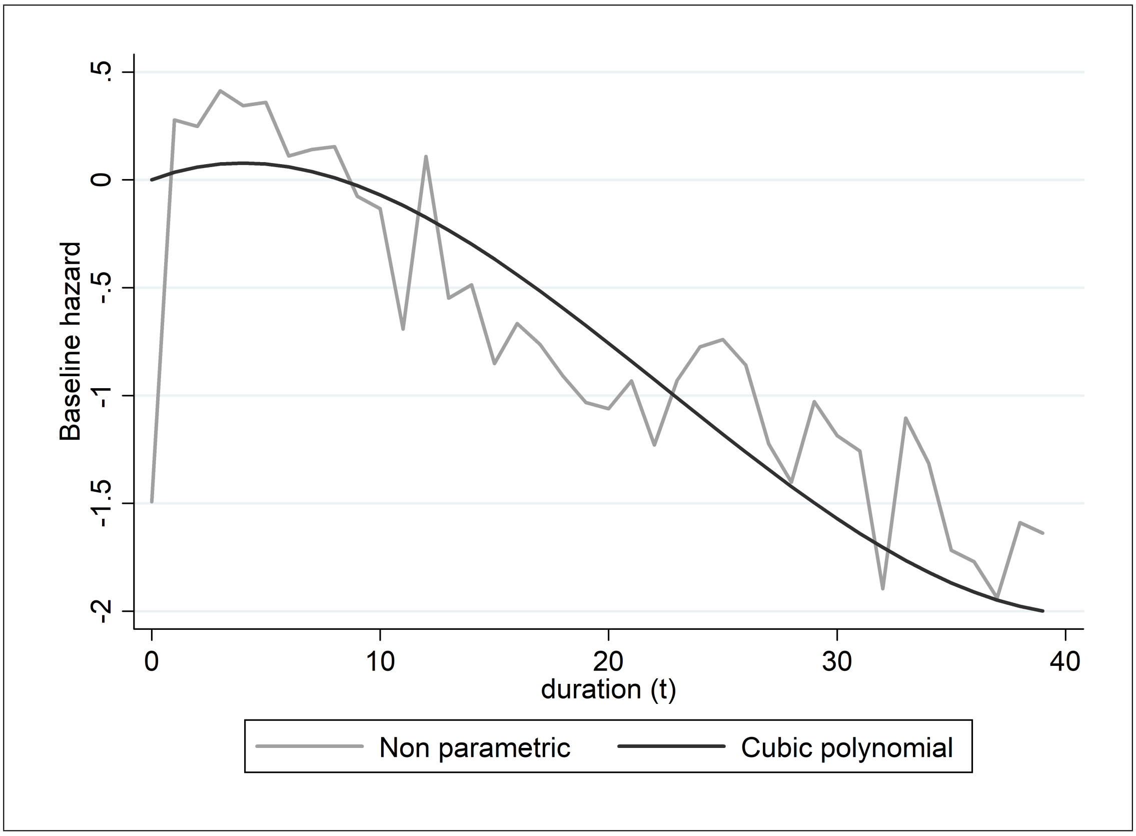

Figure 3: *Comparison of the baseline hazard of model I with a non-parametric definition*

Notes: In discrete survival models a way of obtaining non-parametric baseline hazard function is to use dummy variables for each duration interval. Here, we compare the non-parametric definition to the cubic polynomial we use in Model I.

isations. They capture several aspects related to the quality of institutions, including: the efficiency of the bureaucracy, rule of law, protection of property rights, quality of economic legislation and the extent of corruption in business practices. The results for these two variables are positive and strongly significant; countries with high scores in $Institutions2$ and $Institutions3$ seem to engage more in international cooperation. We obtain this result after controlling for other political and economic factors such as the level of income and the state of democracy. Institutions appear to be a crucial determinant of ratification.

As a further validation of our model, in table 3 we report the estimates of the model when a different link function and baseline hazard definition are used. Model *Logit* shows that the estimates obtained with a logit link are essentially identical to the ones obtained with a cloglog link function. The next model is estimated with a non-parametric baseline function instead of a cubic polynomial. In a discrete setting the non-parametric baseline is derived using dummies for the individual duration periods. As a result, this approach implies the estimation of a much larger number of parameters, heavily

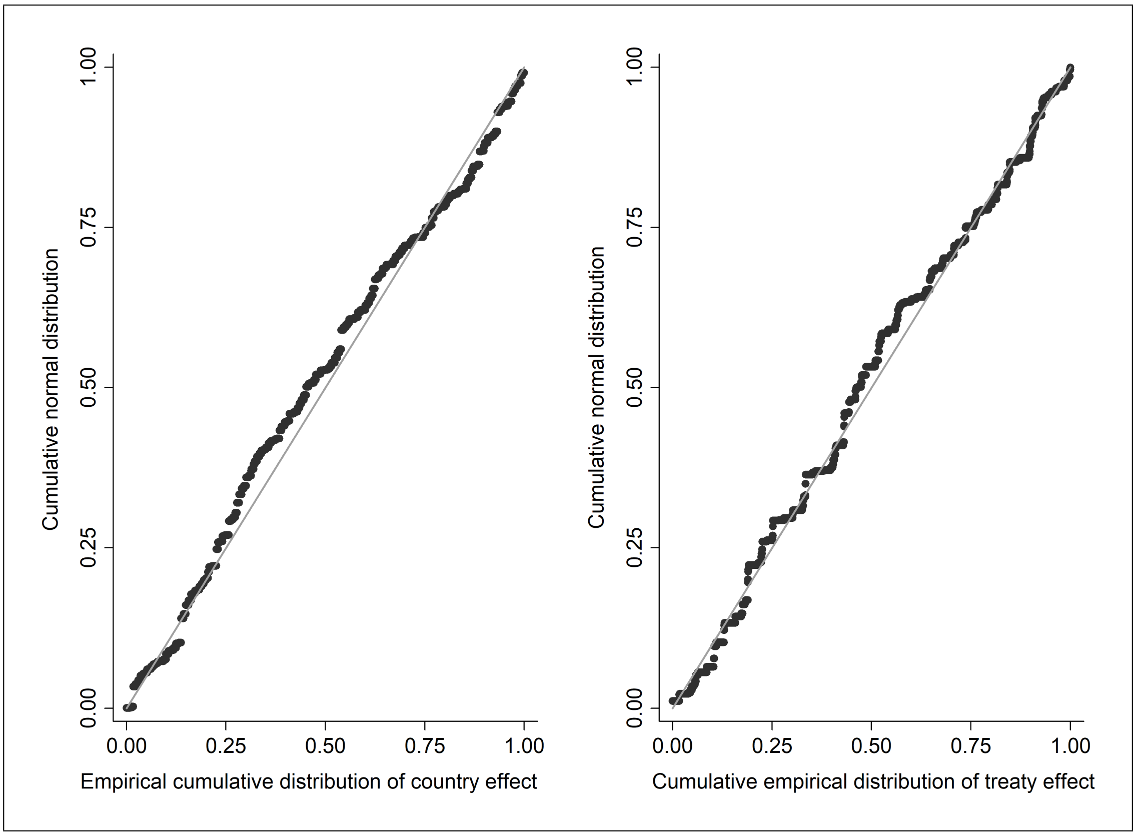

Figure 4: *P-P plot for the distribution of the treaty and country effect*

Notes: The empirical distribution of the treaty and country effects are plotted against a cumulative normal distribution to check for normality. Both distributions follow very closely the diagonal line, suggesting normality in both treaty and country random effects.

affecting both the estimation time and convergence speed of the parameter chains. The estimates of the model have all the same sign and are very close to the ones in model I. The only differences are a slightly lower coefficient for $Institutions$ and a slightly higher $ResourceRent$. The increase in $Resourcerent$ makes the variable significant at the 10% level of significance. The other estimates do not substantially change from the results in other models. In figure 3 we compare the non-parametric baseline hazard with the baseline hazard in model I. The cubic polynomial seems to be a reasonable approximation of the non-parametric version and does not seem to distort the final results. Hence, In model I we opt for the more parsimonious cubic polynomial. The vast gains in estimation time make it a worthwhile simplification and — despite being less versatile than the non-parametric definition — we deem the cubic polynomial is sufficiently accurate.

Finally, in the last column we present the results for a simplified version of model I in which the country random effect is omitted. The estimates are very close to the ones of model I. However, the standard errors are consistently biased downward, leading to erroneous conclusions on the significance of the parameters. This result highlights the importance of using a multilevel strategy to model the clustering of ratifications within the same country and also shows that, despite most of the heterogeneity lies at the treaty level, the country effect needs to be included in the analysis. Our model assumes normality at the treaty and country levels. This property is inspected in Figure 4 where the cumulative distribution of the treaty and country residuals are plotted against a cumulative normal distribution. If the residuals were distributed as a perfect normal distribution the plots would lie along the diagonal line. We observe that both the country and the treaty effects follow very closely the diagonal line suggesting that they are approximately normally distributed.

Overall the results are stable and consistent with what is found in the main results of our study. The findings are unaffected by changes in the definitions of lobbying and quality of institutions. Different proxies are experimented leading to similar conclusions. We also experiment with different specifications of the model. The results obtained with a non-parametric baseline hazard, a different link function and a different multilevel specification, are all similar to those in Model I.

References

Browne, W. (2004). An illustration of the use of reparametrisation methods for improving MCMC efficiency in crossed random effects models. *Multilevel Modelling Newsletter*, 16, 13–25.

Browne, W., Steele, F., Golalizadeh, M., & Green, M. (2009). The use of simple reparameterizations to improve the efficiency of Markov chain Monte Carlo estimation for multilevel models with applications to discrete time survival models. *Journal of the Royal Statistical Society*, 172(Part 3), 579–598.

Fraser Institute (accessed 2017). Economic freedom rankings. *The Fraser Institute*. (Available at: https://www.fraserinstitute.org/).

Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457–472.

Raftery, A. & Lewis, S. (1992). How many iterations in the Gibbs sampler? In J. Bernado, J. Berfer, A. Dawid, & A. Smith (Eds.), *Bayesian Statistics* chapter 4, (pp. 765–776). Oxford: Oxford University Press.

Steele, F., Goldstein, H., & Browne, W. (2004). A general multilevel multistate competing risks model for event history data, with an application to a study of contraceptive use dynamics. *Statistical Modelling*, 4, 145–149.

Tien, M., Markowitz, E., Howe, P., Ko, C.-Y., & Leiserowitz, A. (2015). Predictors of public climate change awareness and risk perception around the world. *Nature Climate Change*, 5, 1014–1020.

World Bank (2017). World Governance Indicators. *The World Bank*. (Available at: http://info.worldbank.org/governance/wgi/index.aspx).