Supplemental Information for Measuring Peace Agreement Strength in Civil War

Rob Williams*

August 25, 2021

Contents

1 Descriptive statistics . . . **2**
1.1 Temporal coverage . . . 6
1.2 PA-X and PAM covariates . . . 7
1.3 Multi-conflict agreements . . . 8

2 Model parameters . . . **10**

3 Cross-validation . . . **12**

4 Sensitivity analyses . . . **13**
4.1 Provision selection . . . 13
4.2 Baseline model compared to full model . . . 16

5 Advantages over additive index . . . **17**

6 Duration . . . **18**

7 Model estimation with Stan . . . **19**

8 MCMC diagnostics . . . **20**

9 Computing environment . . . **22**

References . . . **24**

*Postdoctoral Research Fellow, Washington University in St. Louis, rob.williams@wustl.edu, jayrobwilliams.com

1 Descriptive statistics

Figure 1 displays the frequency of all agreement provisions **Y**.

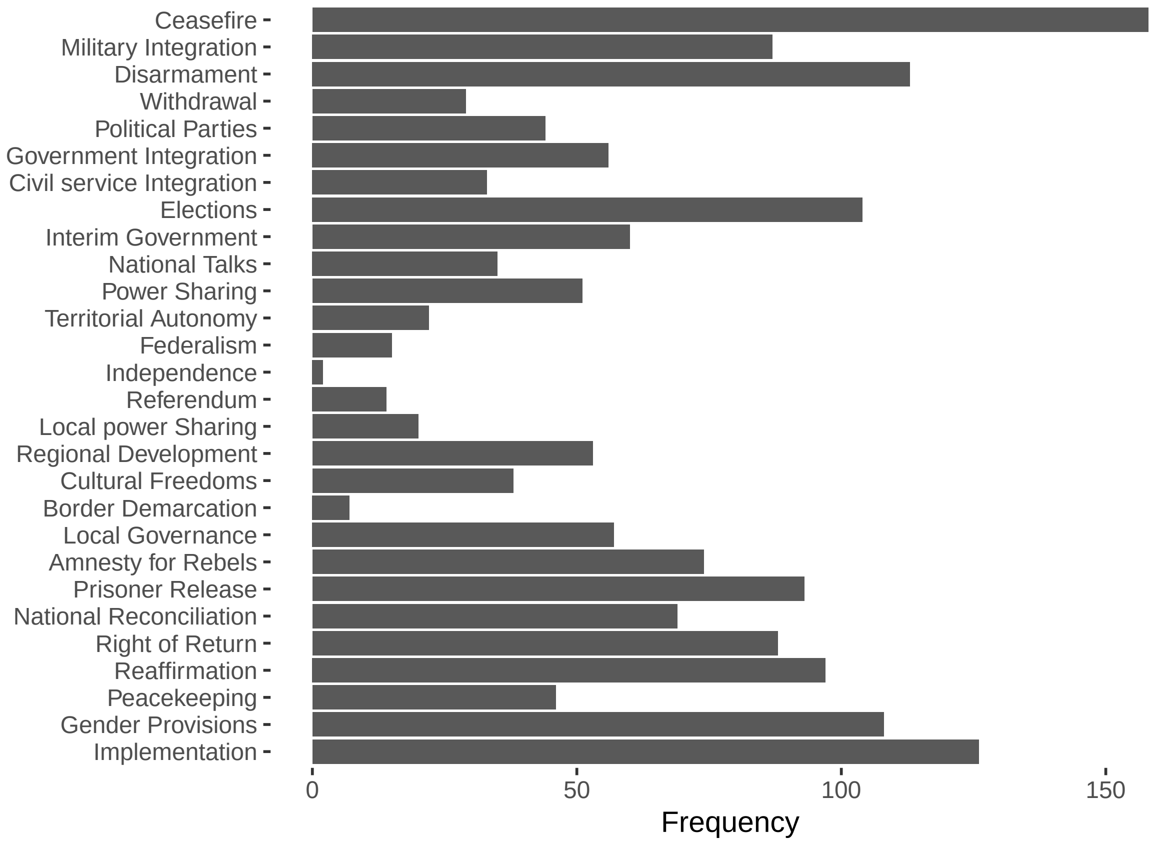

Figure 1: Count of provisions in the data.

Figure 2 displays the correlation matrix for agreement provisions **Y**.

Figures 3 and 4 display the correlation matrix for PA-X and PAM provisions, respectively, contained in **X**.

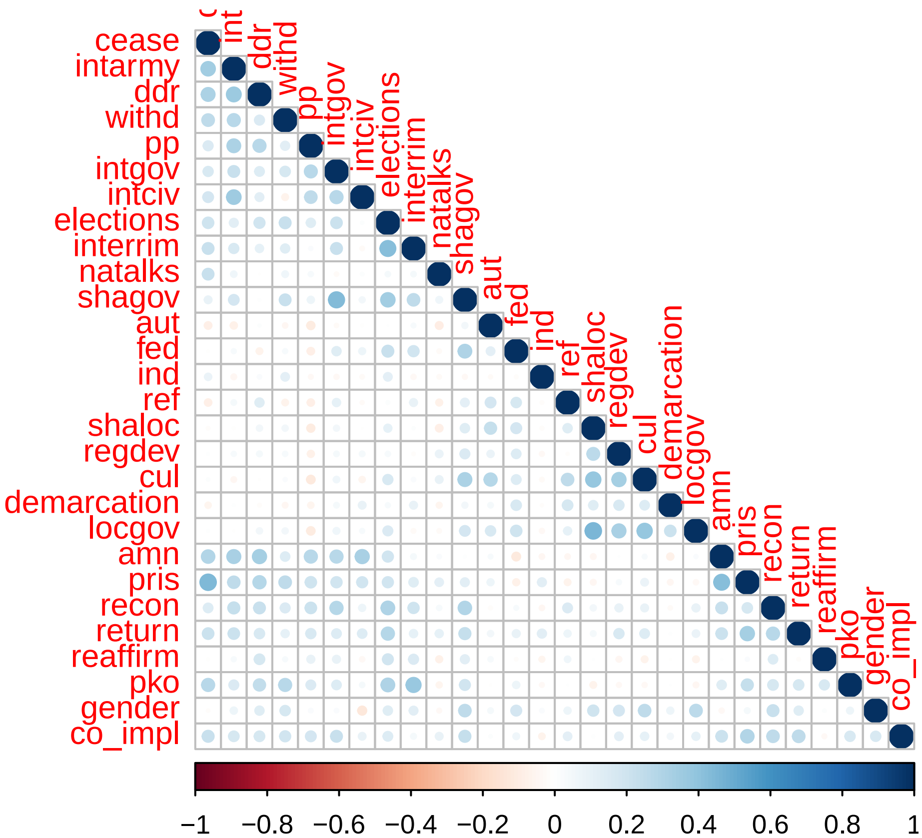

Figure 2: Correlation matrix for agreement provisions

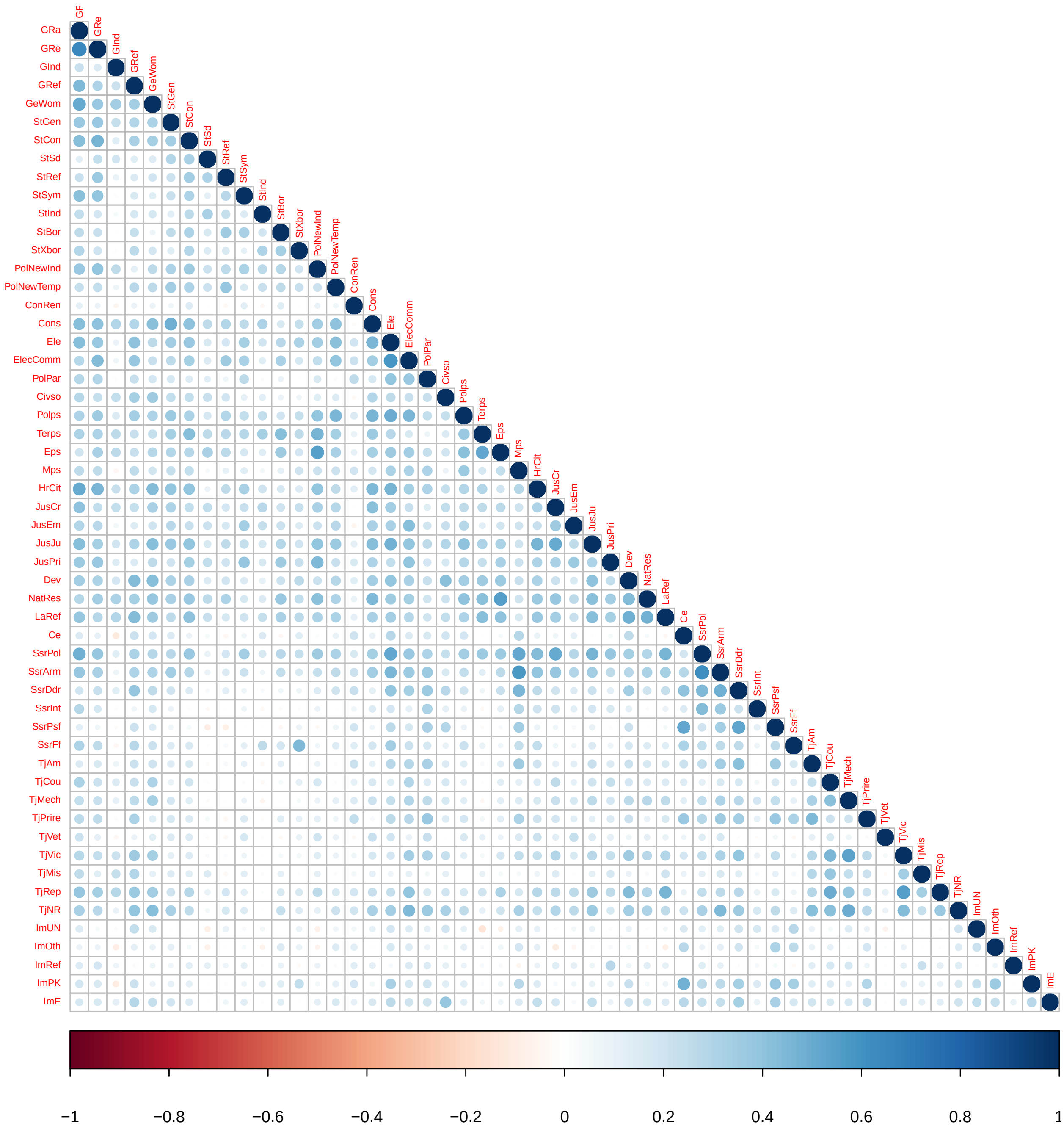

Figure 3: Correlation matrix for PA-X covariates on θ prior components

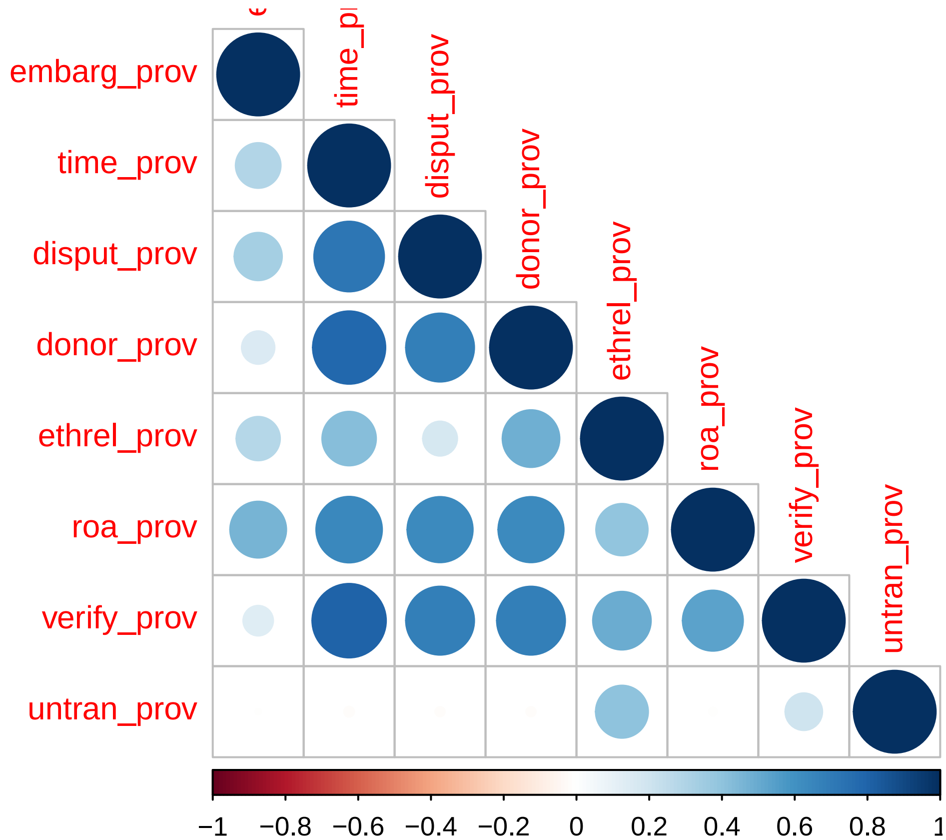

Figure 4: Correlation matrix for PAM covariates on θ prior components

1.1 Temporal coverage

Figure 5 illustrates the data coverage across the three data sources used in the full, conflict-level, robust, and differential models.

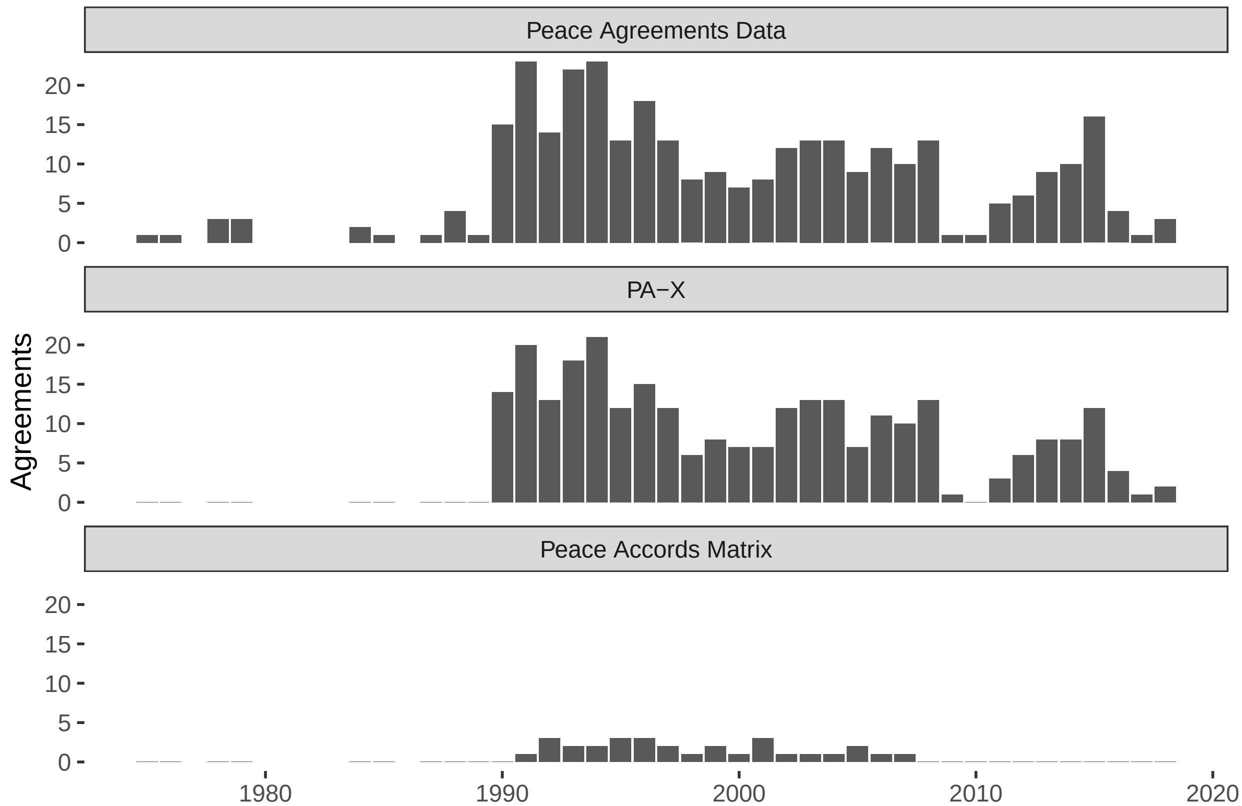

Figure 5: Coverage across data sources

1.2 PA-X and PAM covariates

Figures 1 and 2 present the provisions from PA-X and PAM, respectively, included in the full model in Section 2.2 of the paper.

Ethnic Groups	Party Reform	Opposition Forces
Religious Groups	Civil Society	Withdrawal of Foreign Forces
Indigenous Groups	Political Powersharing	Amnesty
Refugees	Territorial Powersharing	Judicial Accountability
Gender	Economic Powersharing	Transitional Justice
Nature of State	Military Powersharing	Prisoner Release
State Configuration	Citizenship	Vetting and Lustration
Self Determination	Criminal Justice	Victims
State Status Referendum	State of Emergency	Missing
State Symbols	Judiciary and Courts	Reparations
Independence	Prisons	Reconciliation
Border Delimitation	Development	UN Signatory
Cross-border Provisions	Natural Resources	Other Signatory
New Institutions	Land Reform	Implementation Referendum
Temporary Institutions	Ceasefire	Peacekeeping Mission
Constitutional Reaffirmation	Police	Enforcement Mechanism
Constitutional Reform	Armed Forces	
Elections	Disarmament, Demobilization and Reintegration	
Electoral Commission	Intelligence Services	

Table 1: Agreement-level covariates from PA-X

Arms Embargo	Ethnic Relations Council
Implementation Timeline	External Review
Dispute Resolution	Verification Mechanism
Implementation Support	UN Authority

Table 2: Agreement-level covariates from Peace Accords Matrix

1.3 Multi-conflict agreements

An agreement can be signed to terminate multiple separate conflicts, and the UCDP Peace Agreements Data contain 3 such agreements that are signed in more than one conflict. Table 3 presents these agreements. I deal with these cases by splitting the agreements, creating one observation per agreement-conflict pair (e.g., the Vance-Owen Plan in the Bosnia-Herzegovina and Serb conflict is separated from the Vance-Owen Plan in the Bosnia-Herzegovina and Croat conflict). The Vance-Owen Plan and and Deed of Commitment thus become two separate agreements, while the Nationwide Ceasefire in Myanmar becomes three. This splitting is necessary because the same agreement may be stronger or weaker in different conflicts due to different underlying issues driving the violence or different drivers of post-conflict instability.

Agreement	State	Year	Conflicts
Vance-Owen Plan	Bosnia-Herzegovina	1993	2
Nationwide Ceasefire	Myanmar (Burma)	2015	3
Deed of Commitment	Myanmar (Burma)	2015	2

Table 3: Multiple conflict agreements

An agreement with the same conflict resolution provisions may be stronger in one conflict because it addresses more of the rebels' grievances and weaker in another because of a mismatch between the provisions and the second group's grievances. Similarly, the same conflict prevention provision may be varyingly effective in different dyads involved in the same conflict. A group that has an external ally that can deter the government from reneging on an agreement will benefit less from detailed enforcement mechanisms than a group without such an ally, meaning that the contribution of detailed enforcement mechanisms to agreement strength will be lower in the former case.

Splitting agreement signed in multiple conflicts also makes empirical sense because an agreement signed between a government and multiple rebel groups in multiple conflicts does not automatically fail when one conflict restarts. While the violence introduced by the recurrence of one conflict may destabilize relationships between the state and other signatories, there is no systematic evidence that the resumption of hostilities between two signatories to a multiparty agreement will undermine the peace between the other signatories (Nilsson 2008).

Splitting the multi-conflict agreements introduces 3 sets of agreements with identical provisions. The model will give each set of disaggregated agreements identical θ values as the data used to estimate them will be identical. While it may seem problematic that agreements will have identical strength estimates even though they address different contexts, this is actually desirable. Because the measurement model uses only the content of agreements themselves, an agreement signed to terminate two different conflicts will have the same strength in both conflicts. To assess the independent effect of peace agreements on post-conflict outcomes, we must account for all other relevant factors, but doing so requires a measure of agreement strength that does not draw on outside information.

2 Model parameters

Figure 6 presents the item characteristic curves and observed values for all provisions in the full model, replicating Figure 3 for all provisions.

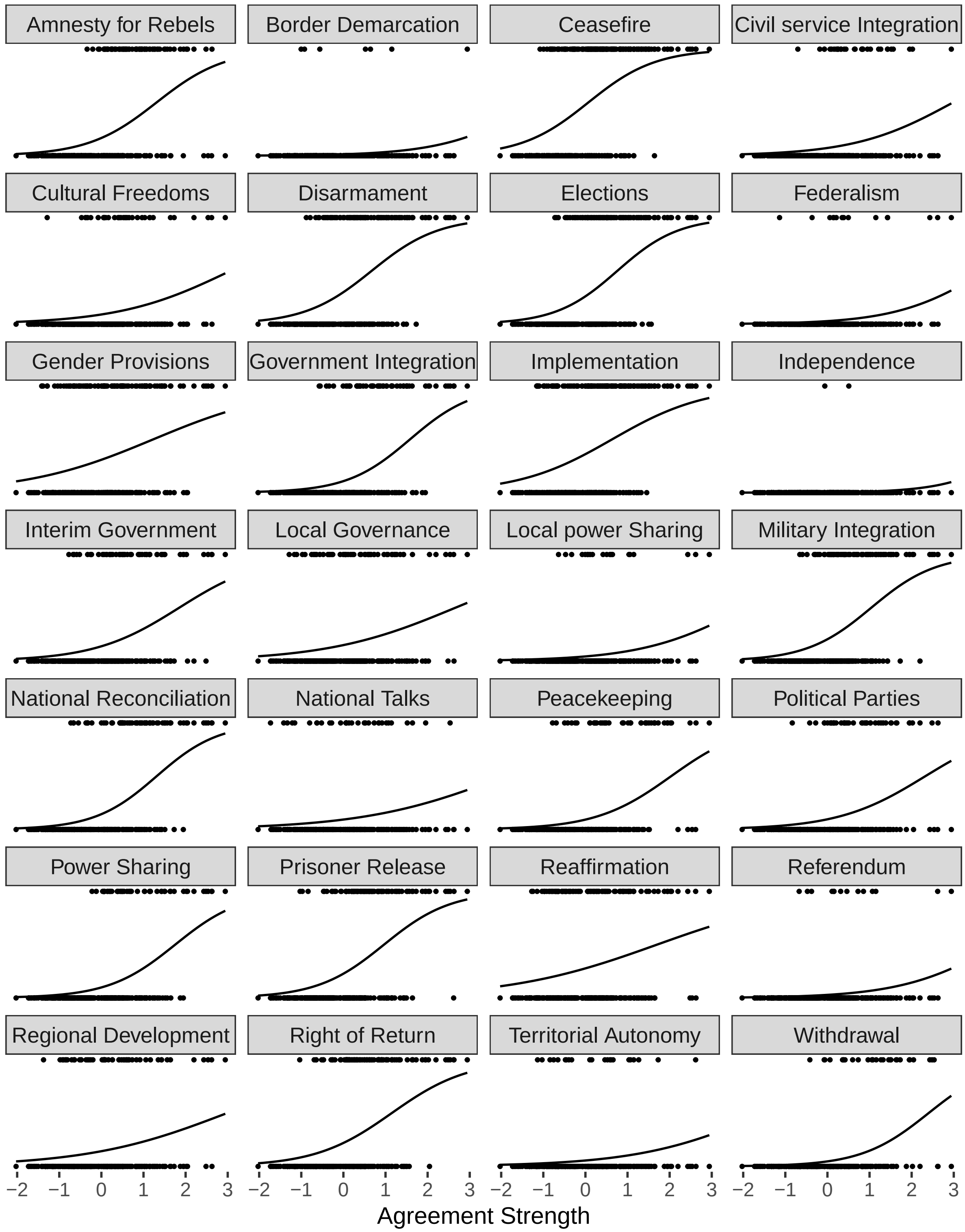

Figure 6: Distribution of observed provisions and item characteristic curves

3 Cross-validation

The 3-fold cross-validation uses fits two different types of models to outcome data. For agreement outcome (continuing or failed) it uses logistic regression and for agreement duration it uses Cox proportional hazard regression. In both cases, it includes a dummy variable to account for whether an agreement was signed during the *Cold War* or not.

4 Sensitivity analyses

4.1 Provision selection

The identification restriction that $\beta_{type} > 0$ requires evaluating whether any provisions should be excluded due to being related to a different latent construct. This is done by examining whether any γ_j estimates have posterior distributions close to 0 (Bafumi et al. 2005). Figure 7 displays the posterior distributions for all γ estimates.

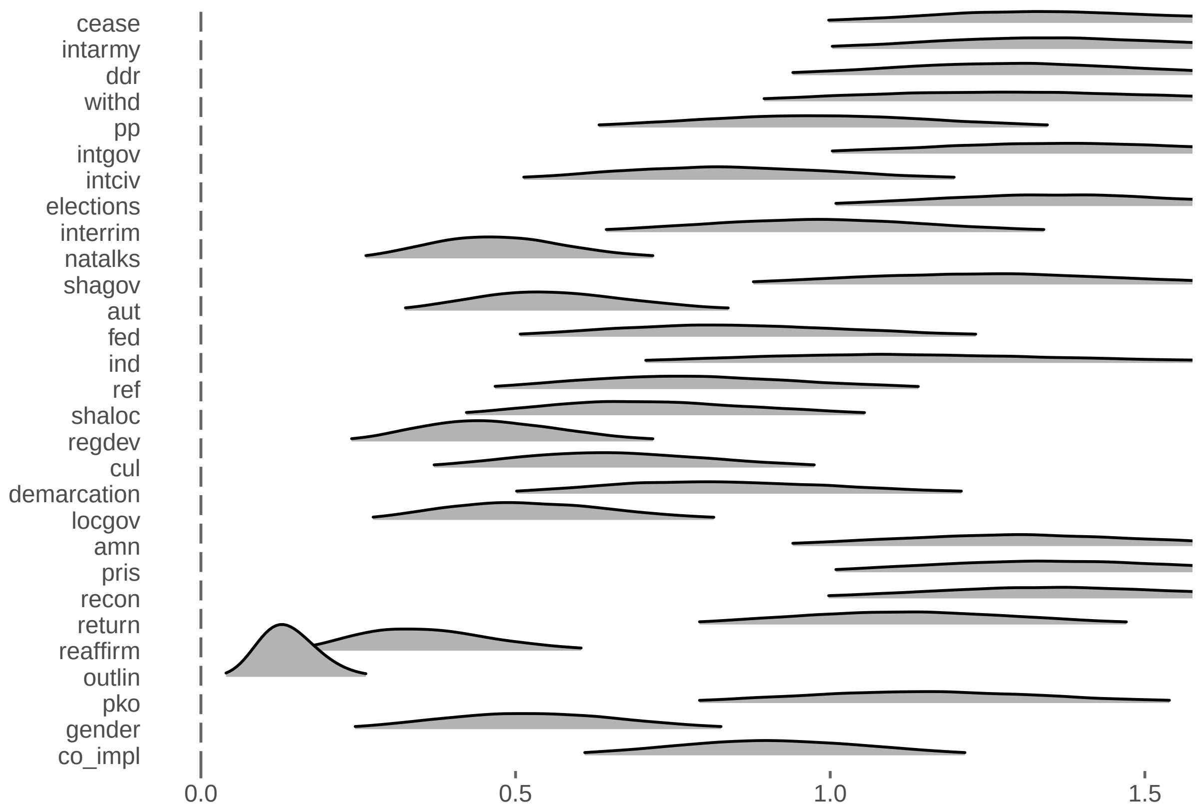

Figure 7: Posterior densities for all discrimination parameters

The only provision with a density closer to 0 is outlining. The group of provisions that had posterior densities close to 0 and were ommitted from Williams et al. (2021) (autonomy, federalism, independence, referendum, local power sharing, regional development, cultural freedoms, and local governance) are discernable as a group of provisions with a lower average γ value than the retained provisions, with the exception of independence which is much higher than in Williams et al. (2021). However, none of these provisions have distributions suggesting their exclusion. When including the outlining provision, only 6 agreements have no provisions. This is in contrast to 25 agreements with no provisions in the paper when outlining is excluded.

Figure 8 plots the rank ordering of agreement strength for the full model presented in the paper as well as one that includes the outlining provision.

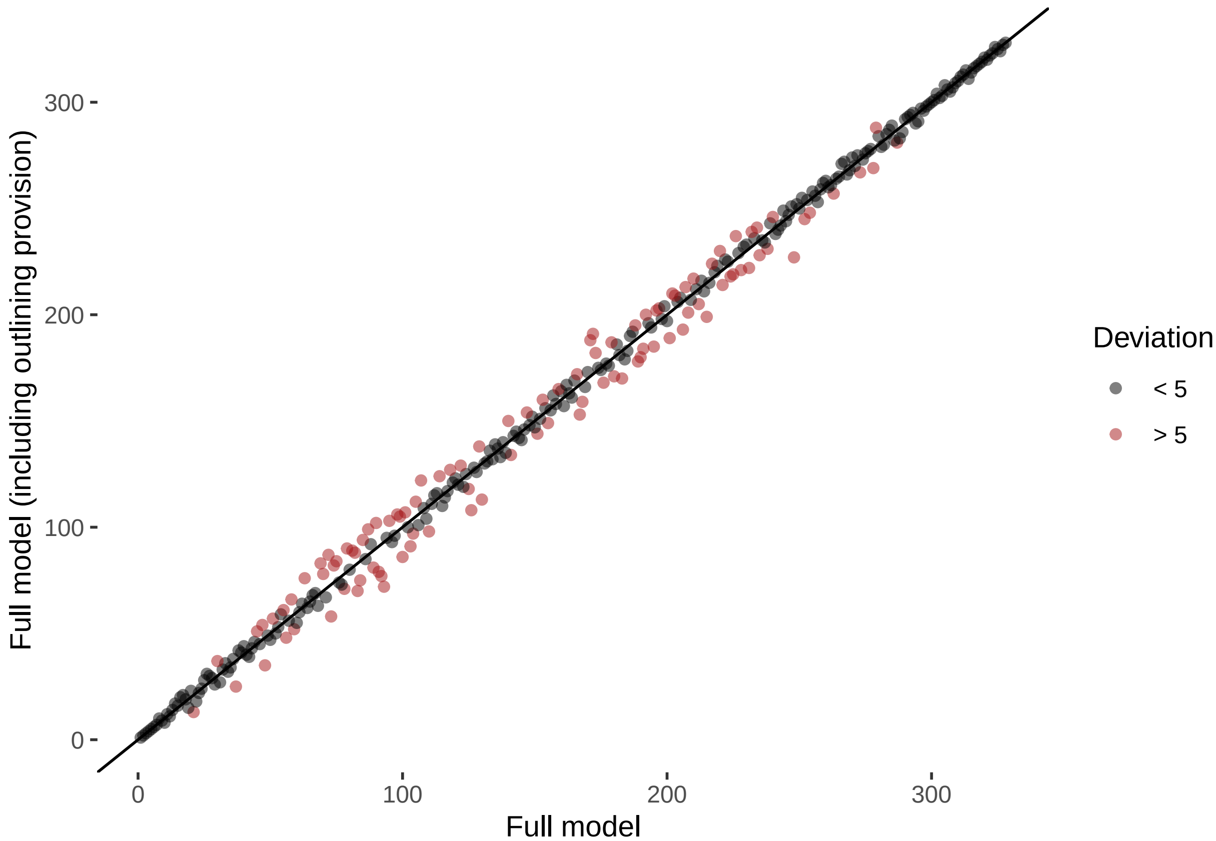

Figure 8: Shift in rank ordering of agreement strengths between the full model and one using all provisions

Figure 9 replicates Figure 6 in the paper but includes the outlining provision.

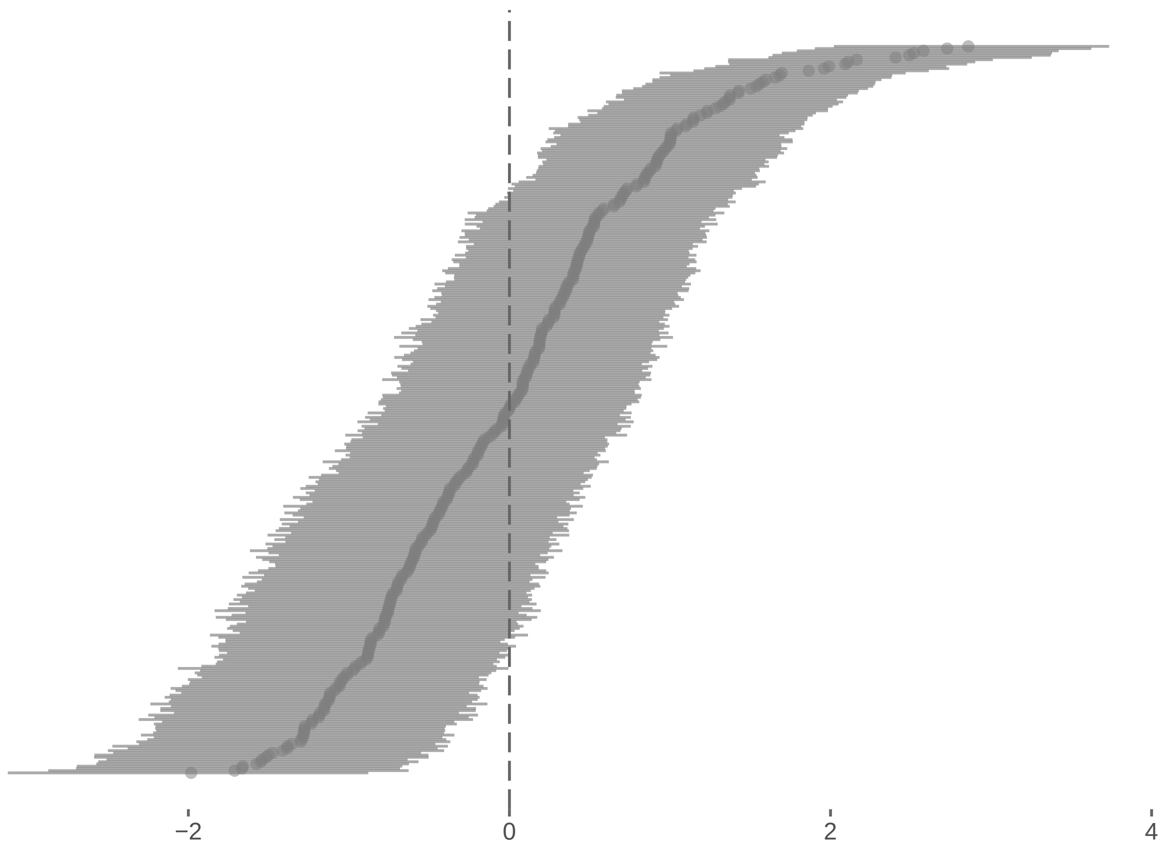

Figure 9: Distribution of agreement strengths with all provisions

4.2 Baseline model compared to full model

Figure 10 plots the rank ordering of agreements in the baseline and full models against one another. Any agreement whose rank order position shifts more than five places between the two scores is plotted in red.

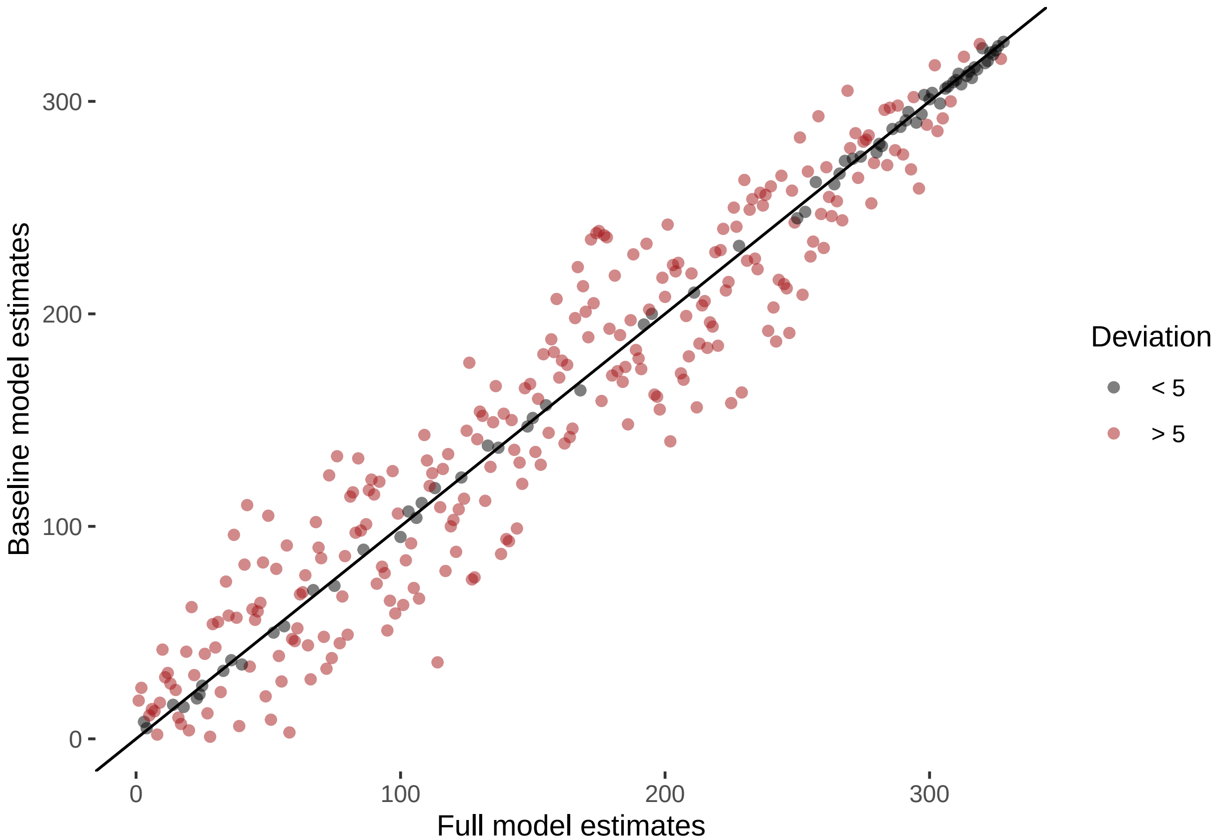

Figure 10: Rank odering of agreement strengths between the full scores and those with no information from PA-X or PAM

5 Advantages over additive index

To illustrate the prevalence of ties that an additive index would yield, Figure 11 presents a histogram of additive index values for all 328 agreements in the data.

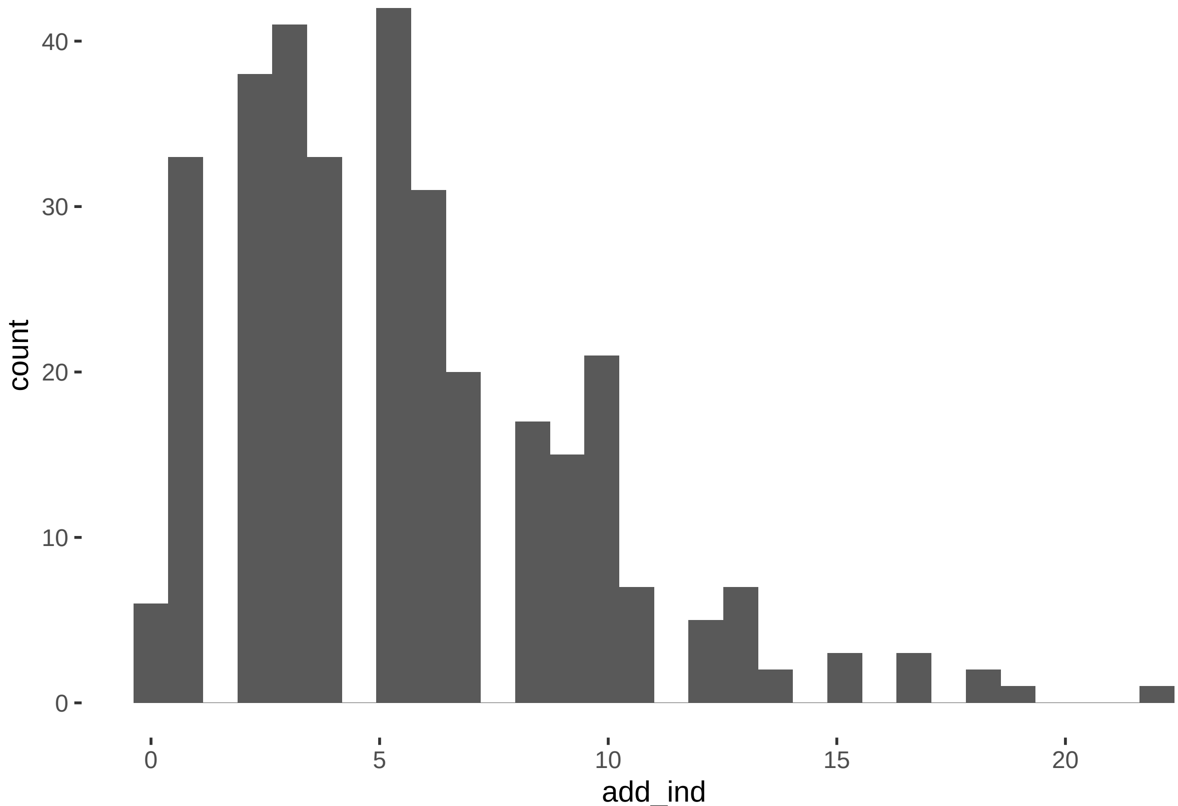

Figure 11: Histogram of additive index values

Although Williams et al. (2021) find that their latent measure of agreement strength is highly correlated with a simple additive index of provisions, the model has many advantages over an additive index. With these updated data, no agreement has more than 22 of 29 provisions. However, many agreements have the same number of provisions, so a latent variable approach to measuring agreement strength solves the problem of ties in the additive index. The most common number of provisions, 5, occurs in 42 agreements.

Any analysis that explains changes in the strengths of peace agreements over the duration of a conflict can incorporate uncertainty about agreement strength in a way that an additive index cannot.

6 Duration

The Cox proprotional hazard models mentioned in the conclusion are presented in Table 4. The first two columns use time-invariant covariates, while the third includes the time-varying covariate of aggregate implementation. All three fail to find a significant relationship between agreement strength and duration.

	Full Sample	PAM Only	PAM Only
Agreement Strength	-0.30^{*}	0.24	0.37
	(0.10)	(0.43)	(0.39)
Aggregate Implementation			-0.02
			(0.02)
AIC	1663.00	66.36	55.49
Num. events	152	10	10
Num. obs.	328	30	272
PH test	0.93	0.63	0.12

$^{*}p < 0.05$

Table 4: Cox proportional hazards models of agreement failure

7 Model estimation with Stan

The IRT parameters θ, α, and γ are reparameterized after estimation in terms of the mean and standard deviation of θ following (Bafumi et al. 2005) to reduce correlation among the IRT parameters and speed up sampling.

$$\theta_i^{adj} = \frac{(\theta_i - \bar{\theta})}{sd(\theta)} \tag{1}$$

$$\alpha_j^{adj} = \frac{(\alpha_j - \bar{\theta})}{sd(\theta)} \tag{2}$$

$$\gamma_j^{adj} = \gamma_j \, sd(\theta) \tag{3}$$

The parameters δ, α, and γ are further reparameterized during estimation with a non-centered parameterization to speed up sampling:

```
data {
  int<lower=1> C;
}
parameters {
  vector[O] alpha_raw;
  vector<lower=.001>[O] gamma_raw;
  vector[M] theta_raw
  vector[C] delta_raw;
}
transformed parameters {
  vector[O] alpha_reparam;
  vector<lower=.001>[O] gamma_reparam;
  vector[C] delta;
  alpha_reparam = mu_alpha + sigma_alpha * alpha_raw;
  gamma_reparam = mu_gamma + sigma_gamma * gamma_raw;
  delta = mu_delta + sigma_delta * delta_raw;
}
model {
  alpha_raw ~ std_normal();
  gamma_raw ~ std_normal();
  delta_raw ~ std_normal();
}
```

8 MCMC diagnostics

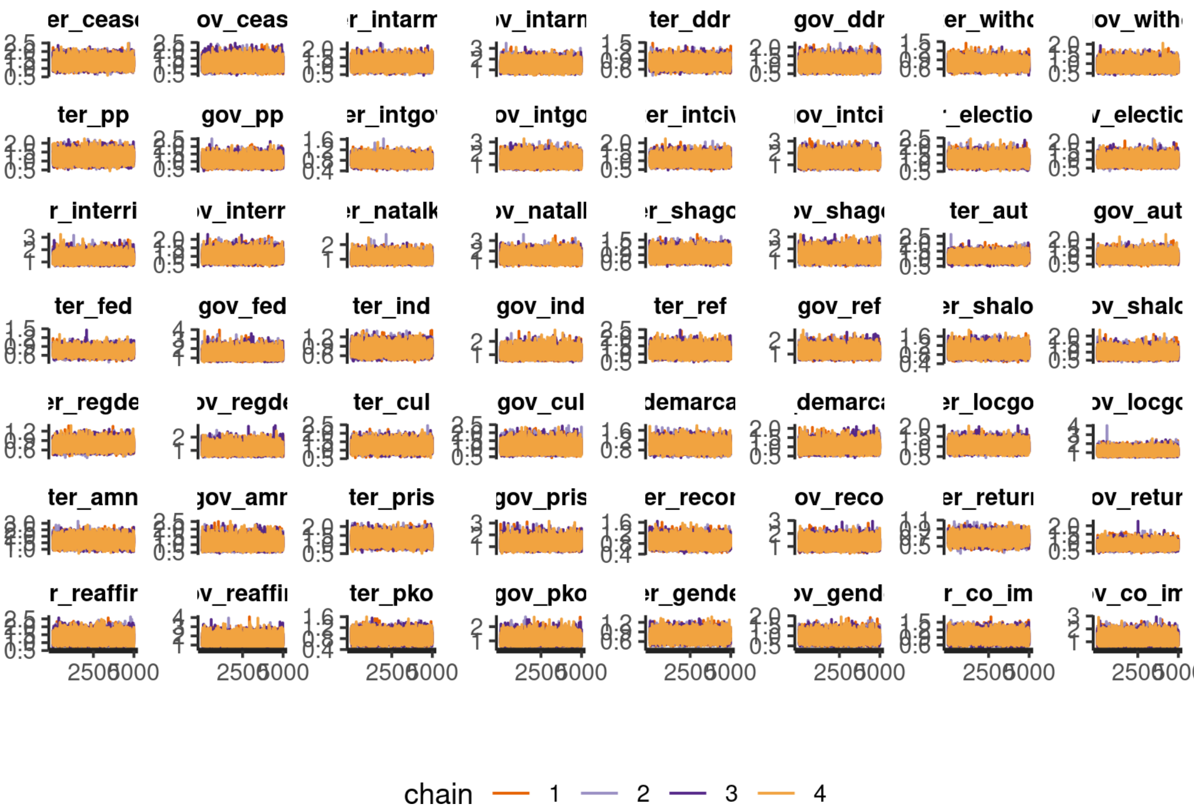

Figure 12: Discrimination parameters

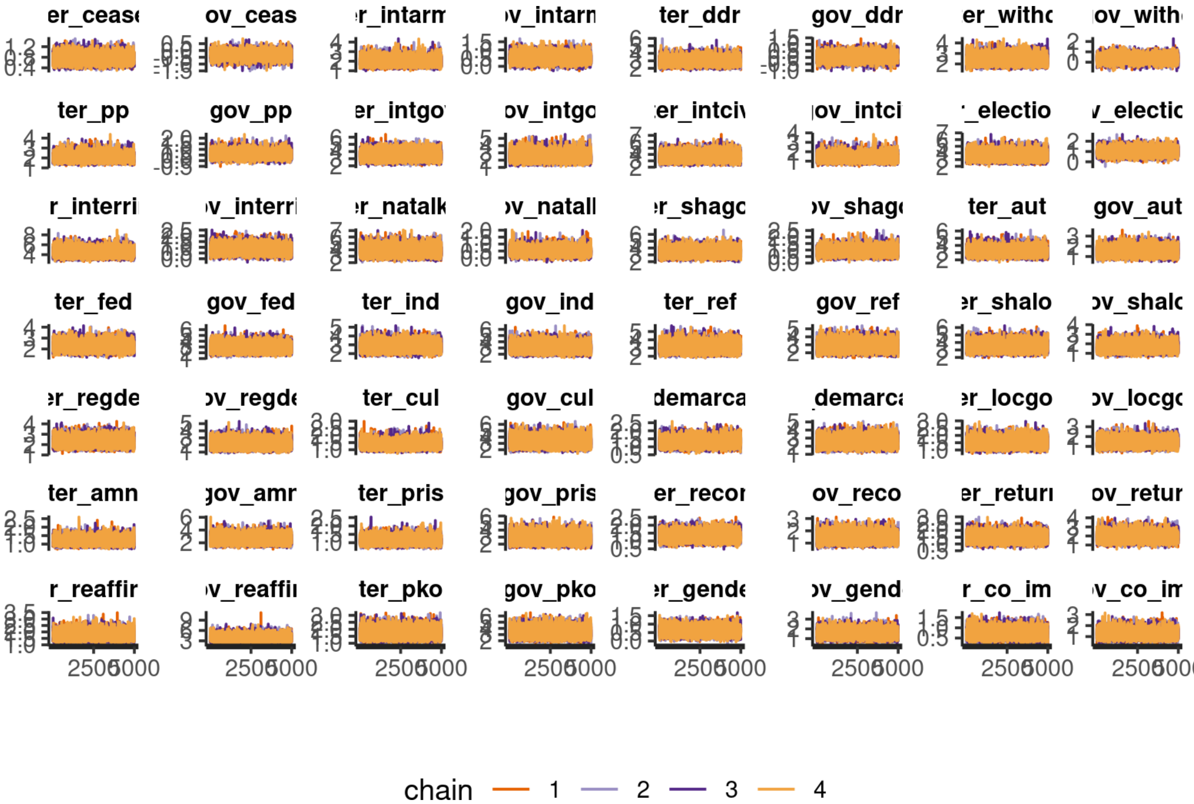

Figure 13: Difficulty Parameters

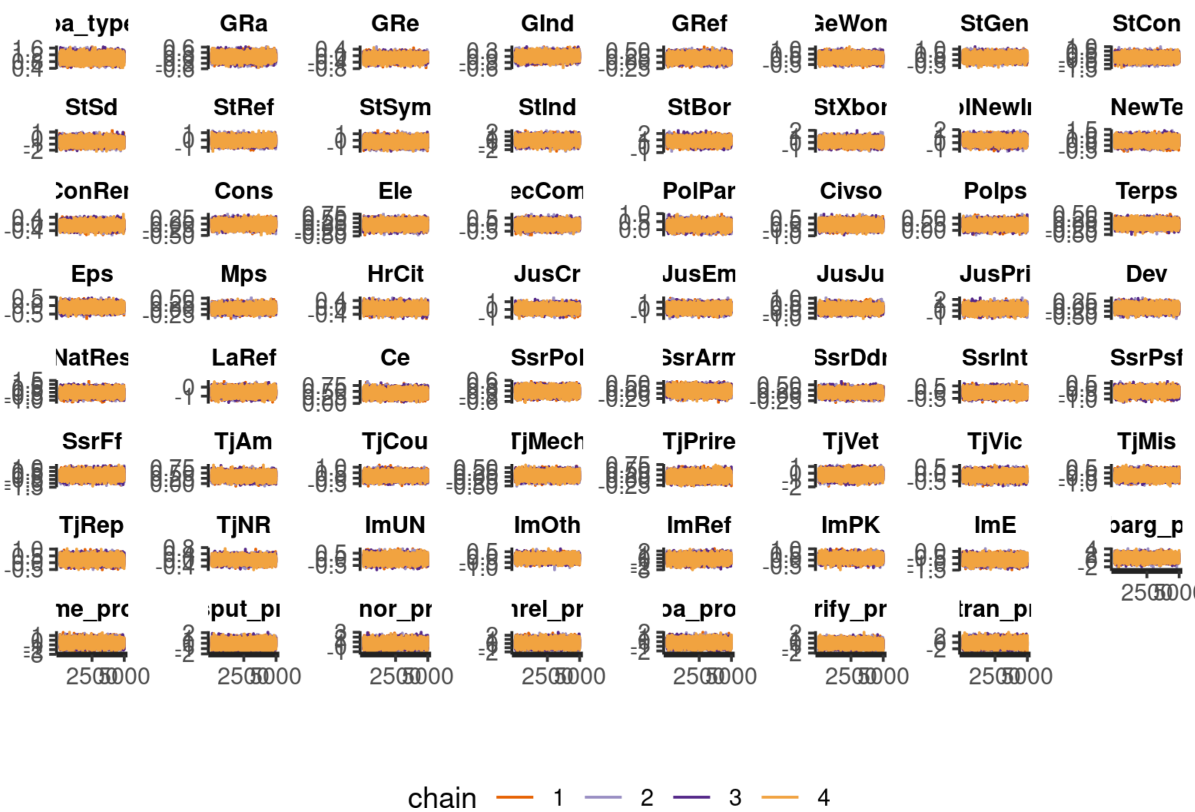

Figure 14: Beta parameters

9 Computing environment

- R version 4.1.1 (2021-08-10), `x86_64-pc-linux-gnu`
- Locale: `LC_CTYPE=en_US.UTF-8`, `LC_NUMERIC=C`, `LC_TIME=en_US.UTF-8`, `LC_COLLATE=en_US.UTF-8`, `LC_MONETARY=en_US.UTF-8`, `LC_MESSAGES=en_US.UTF-8`, `LC_PAPER=en_US.UTF-8`, `LC_NAME=C`, `LC_ADDRESS=C`, `LC_TELEPHONE=C`, `LC_MEASUREMENT=en_US.UTF-8`, `LC_IDENTIFICATION=C`
- Running under: `Ubuntu 20.04.3 LTS`
- Matrix products: default
- BLAS: `/usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3`
- LAPACK: `/usr/lib/x86_64-linux-gnu/openblas-pthread/liblapack.so.3`
- Base packages: base, datasets, graphics, grDevices, methods, stats, utils
- Other packages: corrplot 0.88, dplyr 1.0.6, english 1.2-5, forcats 0.5.1, ggplot2 3.3.5, ggrepel 0.9.1, ggridges 0.5.3, purrr 0.3.4, readr 1.4.0, rstan 2.21.2, StanHeaders 2.21.0-7, stringr 1.4.0, survival 3.2-11, texreg 1.37.5, tibble 3.1.2, tidyr 1.1.3, tidyverse 1.3.1, xtable 1.8-4
- Loaded via a namespace (and not attached): abind 1.4-5, arm 1.11-2, assertthat 0.2.1, backports 1.2.1, base64enc 0.1-3, boot 1.3-28, broom 0.7.6, callr 3.7.0, cellranger 1.1.0, checkmate 2.0.0, cli 2.5.0, cluster 2.1.2, coda 0.19-4, codetools 0.2-18, colorspace 2.0-1, compiler 4.1.1, crayon 1.4.1, curl 4.3.1, data.table 1.14.0, DBI 1.1.1, dbplyr 2.1.1, digest 0.6.27, ellipsis 0.3.2, evaluate 0.14, fansi 0.5.0, farver 2.1.0, foreign 0.8-81, Formula 1.2-4, fs 1.5.0, generics 0.1.0, GGally 2.1.2, ggmcmc 1.5.1.1, glue 1.4.2, grid 4.1.1, gridExtra 2.3, gtable 0.3.0, haven 2.4.1, Hmisc 4.5-0, hms 1.1.0, htmlTable 2.2.1, htmltools 0.5.1.1, htmlwidgets 1.5.3, httr 1.4.2, inline 0.3.19, jpeg 0.1-8.1, jsonlite 1.7.2, knitr 1.33, labeling 0.4.2, lattice 0.20-44, latticeExtra 0.6-29, lifecycle 1.0.0, lme4 1.1-27, loo 2.4.1, lubridate 1.7.10, magick 2.7.2, magrittr 2.0.1, MASS 7.3-54, Matrix 1.3-3, matrixStats 0.58.0, minqa 1.2.4, modelr 0.1.8, munsell 0.5.0, nlme 3.1-152, nloptr 1.2.2.2, nnet 7.3-16, openxlsx 4.2.3, parallel 4.1.1, pillar 1.6.1, pkgbuild 1.2.0, pkgconfig 2.0.3, plyr 1.8.6, png 0.1-7, prettyunits 1.1.1, processx 3.5.2, ps 1.6.0, R6 2.5.0,

RColorBrewer 1.1-2, Rcpp 1.0.6, RcppParallel 5.1.4, readxl 1.3.1, reprex 2.0.0, reshape 0.8.8, rio 0.5.26, rlang 0.4.11, rmarkdown 2.8, rpart 4.1-15, rstudioapi 0.13, rvest 1.0.0, scales 1.1.1, splines 4.1.1, stats4 4.1.1, stringi 1.6.2, tidyselect 1.1.1, tools 4.1.1, utf8 1.2.1, V8 3.4.2, vctrs 0.3.8, withr 2.4.2, xfun 0.23, xml2 1.3.2, yaml 2.2.1, zip 2.2.0

References

Bafumi, Joseph, Andrew Gelman, David K. Park, and Noah Kaplan. 2005. "Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation." *Political Analysis* 13 (2): 171–87. https://doi.org/10.1093/pan/mpi010.

Nilsson, Desirée. 2008. "Partial Peace: Rebel Groups Inside and Outside of Civil War Settlements." *Journal of Peace Research* 45 (4): 479–95. https://doi.org/10.1177/0022343308091357.

Williams, Rob, Daniel J. Gustafson, Stephen E. Gent, and Mark J. C. Crescenzi. 2021. "A Latent Variable Approach to Measuring and Explaining Peace Agreement Strength." *Political Science Research and Methods* 9 (1): 89–105. https://doi.org/10.1017/psrm.2019.23.