

Supplemental Appendix For

The Politics of Scrutiny in Human Rights Monitoring: Evidence from Structural Topic Models of U.S. State Department Human Rights Reports

by

Benjamin E. Bagozzi and Daniel Berliner

This supplemental appendix proceeds in eight parts. First, we provide summary statistics and source information for our independent and control variables. In the second section, we report a range of additional figures to supplement the results presented in the main paper, including maps of the spatial variation of each topic, and results for each covariate included in our main model. Following this presentation, we discuss the topics identified within a series of 10 and 20-topic STM's, and their similarities to our primary 15-topic model results. Fourth, we present a variety of secondary model fit statistics that bolster our choice of a 15-topic model as our primary STM. Fifth, we evaluate the stability of our topics across a far larger (100-topic) STM. Sixth, we interpret a set of comparable topics and topwords that were obtained via the application of (i) a latent Dirichlet allocation (LDA) model and (ii) k-means.

Seventh, we present a number of 15-topic STM robustness specifications, which demonstrate that our primary conclusions with respect to topics and covariate effects are robust to (i) the inclusion of a PTS measure of human rights abuses derived from Amnesty International reports in place of our primary latent human rights protection control, (ii) the inclusion of a document-indexed word-count control that records the number of final words in our preprocessed documents, and (iii) the omission our nonlinear year time trend control. Building on this analysis, section eight evaluates the robustness of our primary covariate effects under three alternate conditions (i) the 10-topic STM referenced above, (ii) the aforementioned 20-topic STM, and (iii) a 15-topic STM drawn from a single initialization. We find stable covariate effects across all three cases. Finally, we present a set of time series plots for our proportion based measures of Topical Attention.

Explanatory Variables

Table A.1: Summary Statistics for Independent and Control Variables

	Median	Mean	Std. Dev.	Min	Max	Source
Year	1997	1997	9.78	1977	2012	Naturally Coded
U.S. Ally	0.00	0.30	0.46	0.00	1.00	Gibler (2009)
Polity 2	4.00	1.44	7.09	-10.00	10.00	Marshall, Jaggers and Gurr (2010)
Ln GDP pc	7.52	7.67	1.58	3.91	11.09	Bank (2013)
Ln U.S. Aid	16.79	14.24	6.46	0.00	22.02	USAID (2014)
Civil Conflict	0.00	0.26	0.55	0.00	2.00	Gleditsch et al. (2002)
OECD Member	0.00	0.15	0.35	0.00	1.00	Naturally Coded
Ln U.S. Trade	6.63	6.68	2.60	0.00	13.33	Barbieri and Keshk (2012); Census.gov (2015)
PTS (Amnesty)	3.00	2.71	1.08	1.00	5.00	Gibney et al. (2014)
Latent HR Protection	0.01	0.05	1.18	-2.84	3.92	Fariss (2014)
Years to Election	1.00	1.47	1.12	0.00	3.00	Naturally Coded
Rep. President	0.55	1.00	0.50	0.00	1.00	Naturally Coded
Pres. Chambers	0.91	1.00	0.86	0.00	2.00	Naturally Coded

Supplementary Figures

Figure A.1: Average Topic Prevalence by Country (1977-2012): Topic 1 (“Killings/Disappearances”)

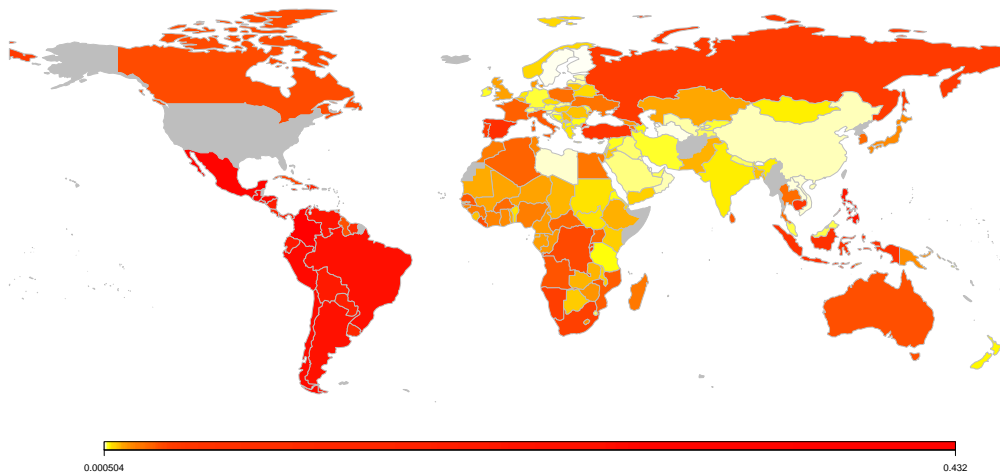


Figure A.2: Average Topic Prevalence by Country (1977-2012): Topic 2 (“Labor Rights”)

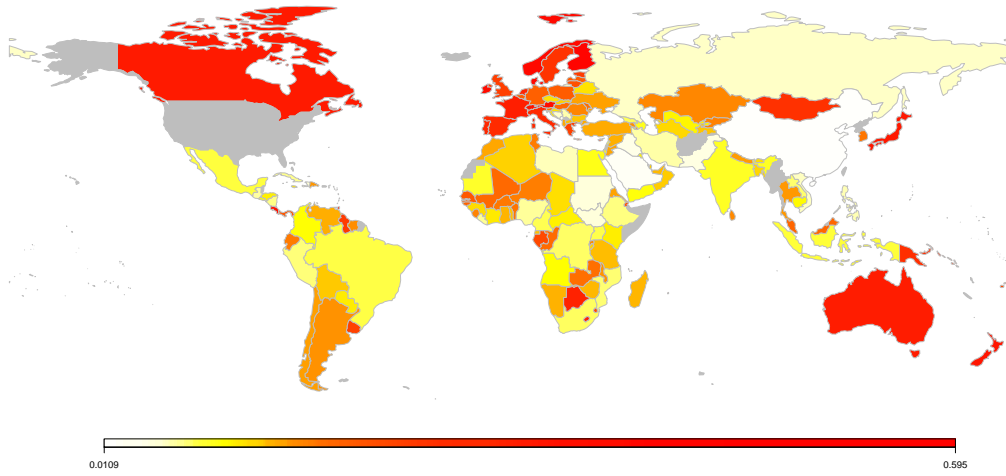


Figure A.3: Average Topic Prevalence by Country (1977-2012): Topic 3 (“Terrorism/Torture”)

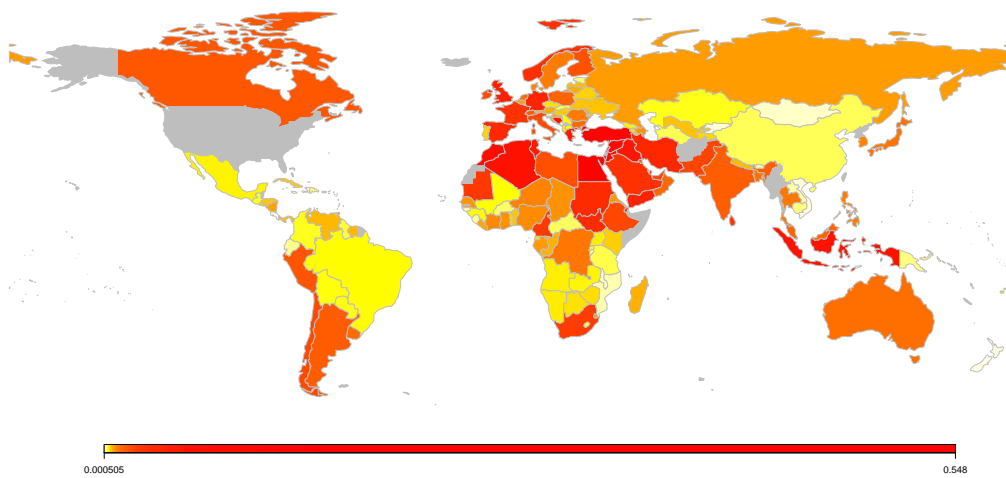


Figure A.4: Average Topic Prevalence by Country (1977-2012): Topic 4 (“State/Non-State Violence”)

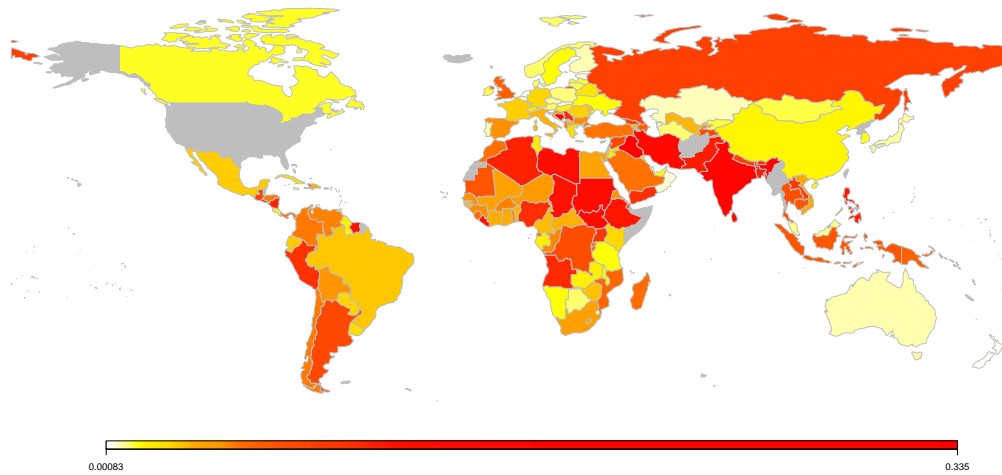


Figure A.5: Average Topic Prevalence by Country (1977-2012): Topic 5 (“Elections/Parties”)

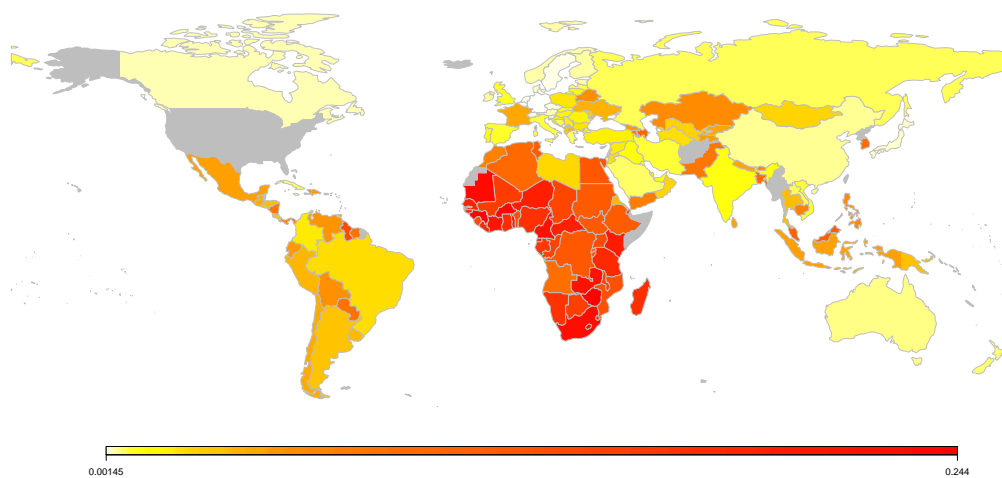


Figure A.6: Average Topic Prevalence by Country (1977-2012): Topic 6 (“Human Trafficking”)

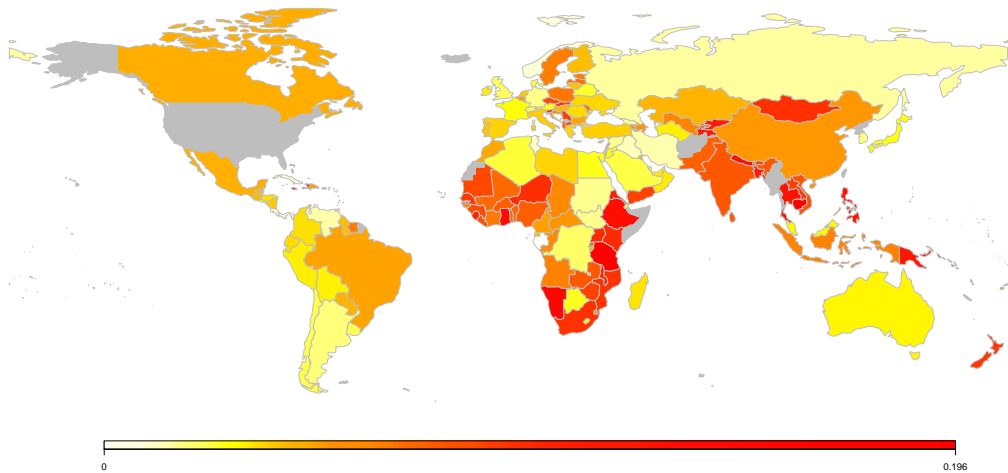


Figure A.7: Average Topic Prevalence by Country (1977-2012): Topic 7 (“Prison Conditions”)

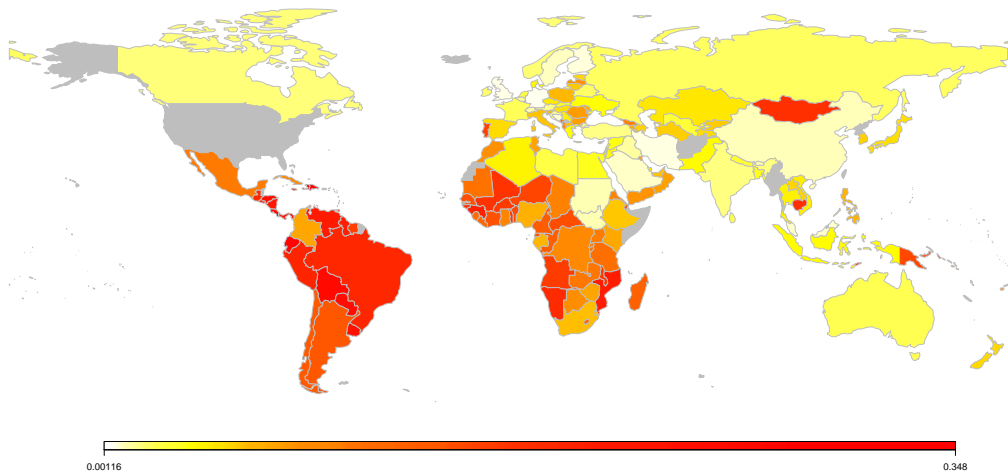


Figure A.8: Average Topic Prevalence by Country (1977-2012): Topic 8 (“Economic Systems”)

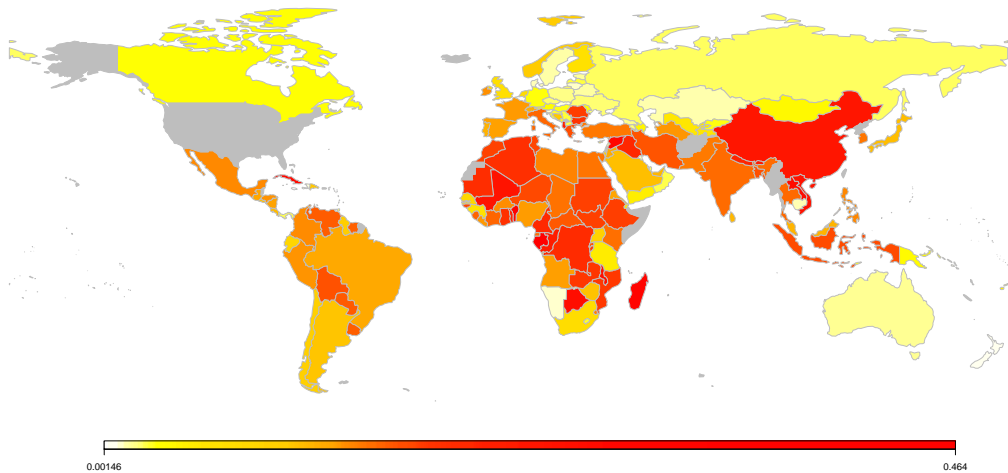


Figure A.9: Average Topic Prevalence by Country (1977-2012): Topic 9 (“Freedom of Movement”)

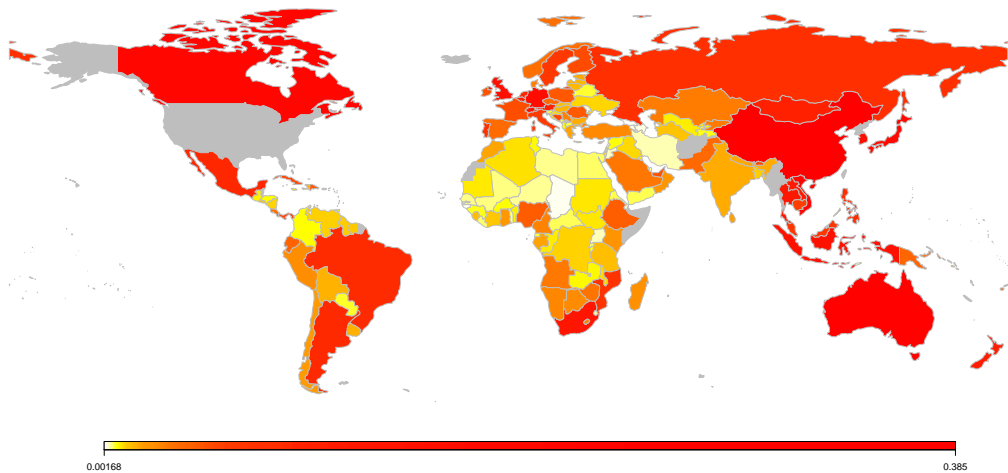


Figure A.10: Average Topic Prevalence by Country (1977-2012): Topic 10 (“Freedom of Expression”)

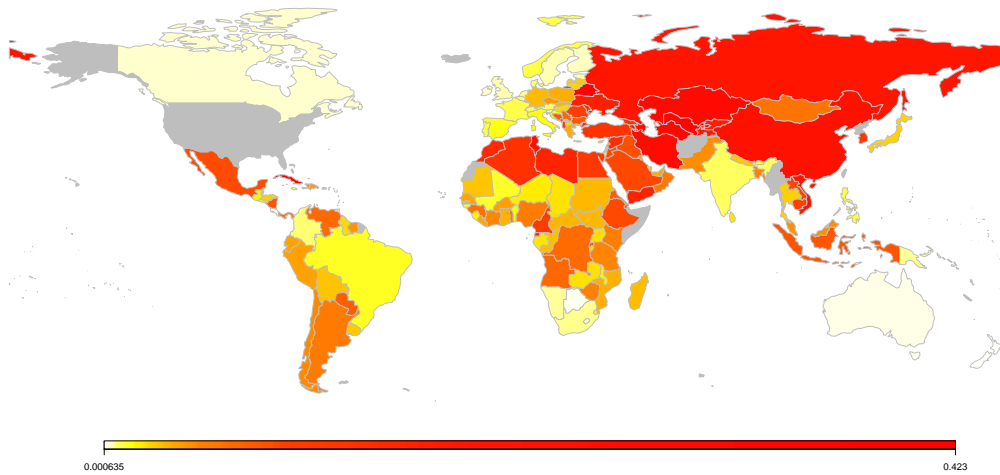


Figure A.11: Average Topic Prevalence by Country (1977-2012): Topic 11 (“Justice System”)

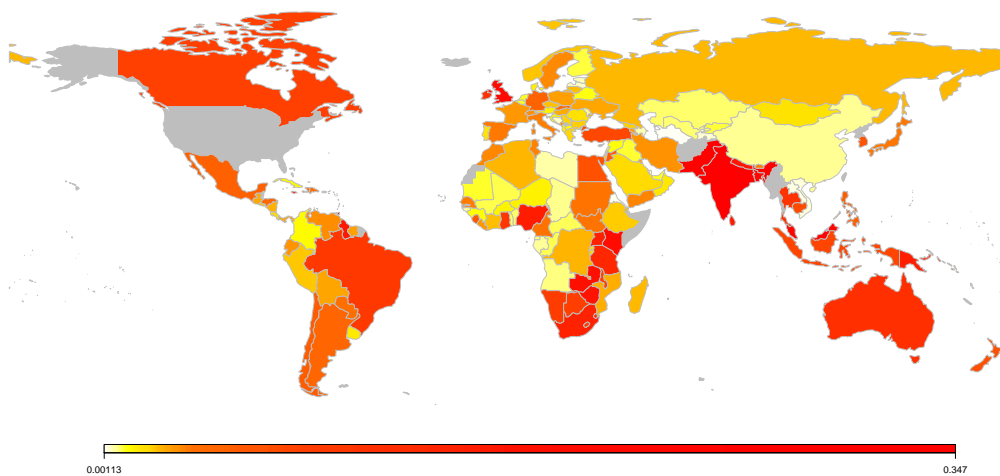


Figure A.12: Average Topic Prevalence by Country (1977-2012): Topic 12 (“Religion/Citizenship”)

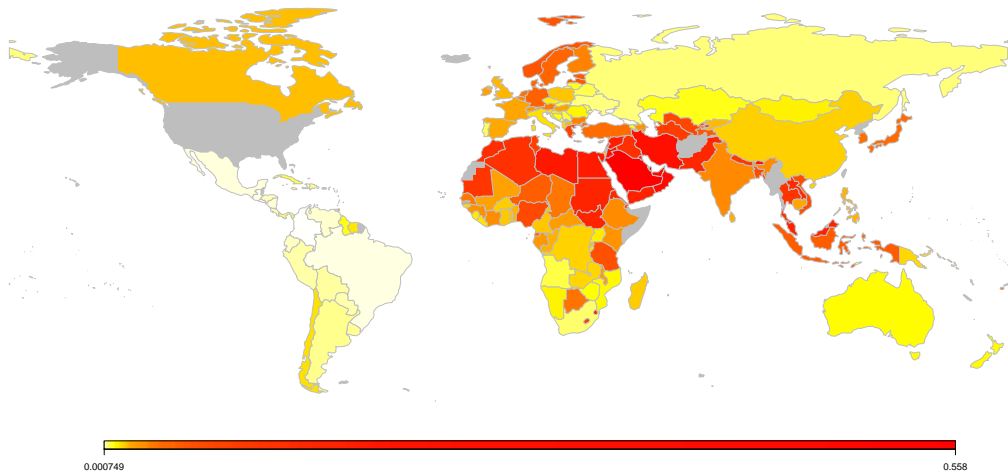


Figure A.13: Average Topic Prevalence by Country (1977-2012): Topic 13 (“New Rights Issues”)

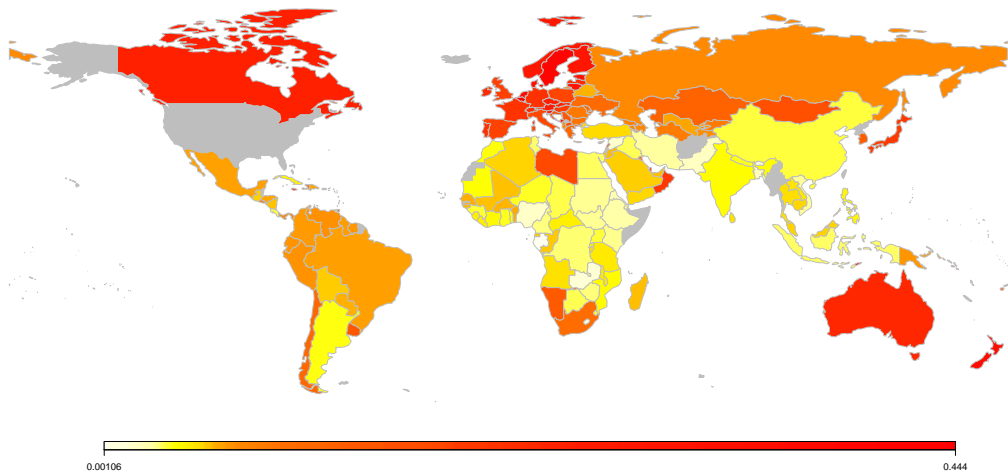


Figure A.14: Average Prevalence by Country (1977-2012): Topic 14 (“Ethnic/Minority Rights”)

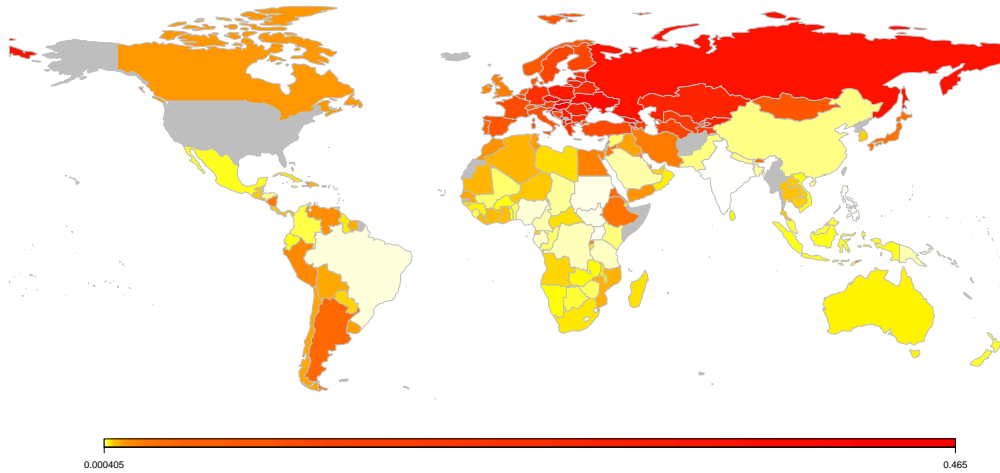


Figure A.15: Average Topic Prevalence by Country (1977-2012): Topic 14 (“Humanitarian Crises”)

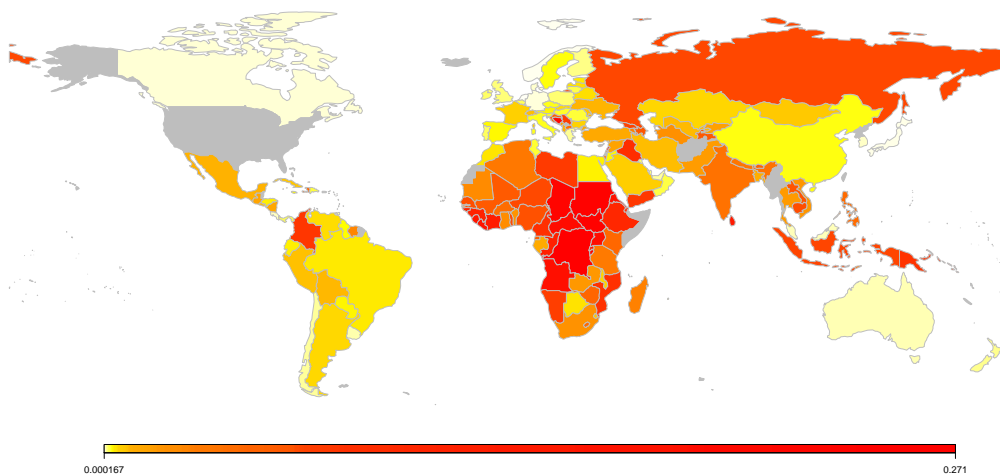


Figure A.16: Covariate relationship with each topic, based on the estimated effect of an increase in the covariate on the expected prevalence of each topic. Increases are from 0 to 1 for dichotomous variables, and from one standard deviation below the mean to one standard deviation above for others. Horizontal lines show 95 percent confidence intervals.

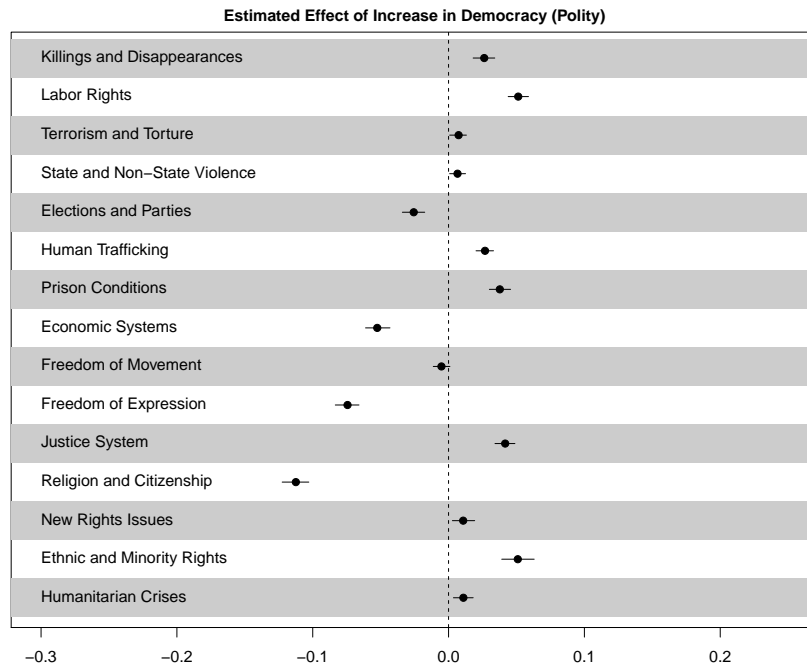


Figure A.17: Covariate relationship with each topic, based on the estimated effect of an increase in the covariate on the expected prevalence of each topic. Increases are from 0 to 1 for dichotomous variables, and from one standard deviation below the mean to one standard deviation above for others. Horizontal lines show 95 percent confidence intervals.

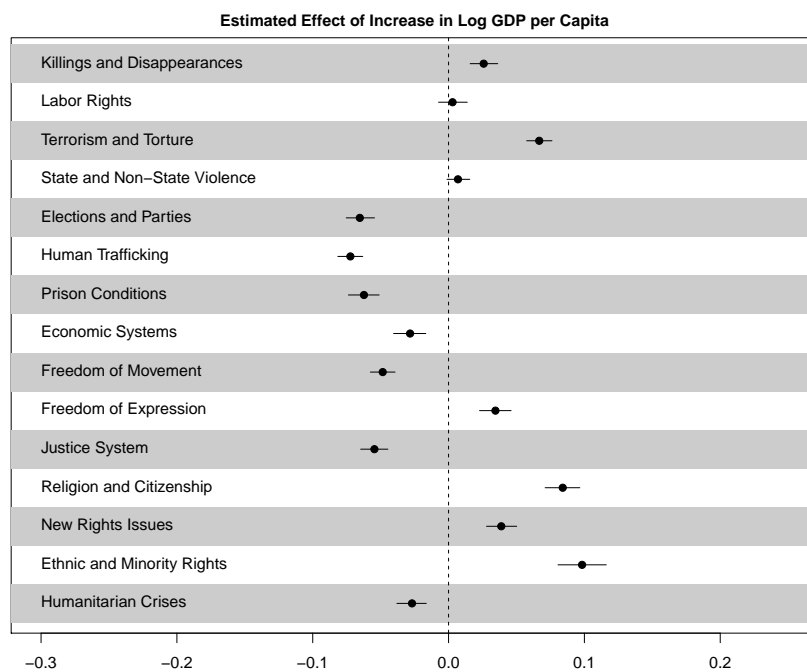


Figure A.18: Covariate relationship with each topic, based on the estimated effect of an increase in the covariate on the expected prevalence of each topic. Increases are from 0 to 1 for dichotomous variables, and from one standard deviation below the mean to one standard deviation above for others. Horizontal lines show 95 percent confidence intervals.

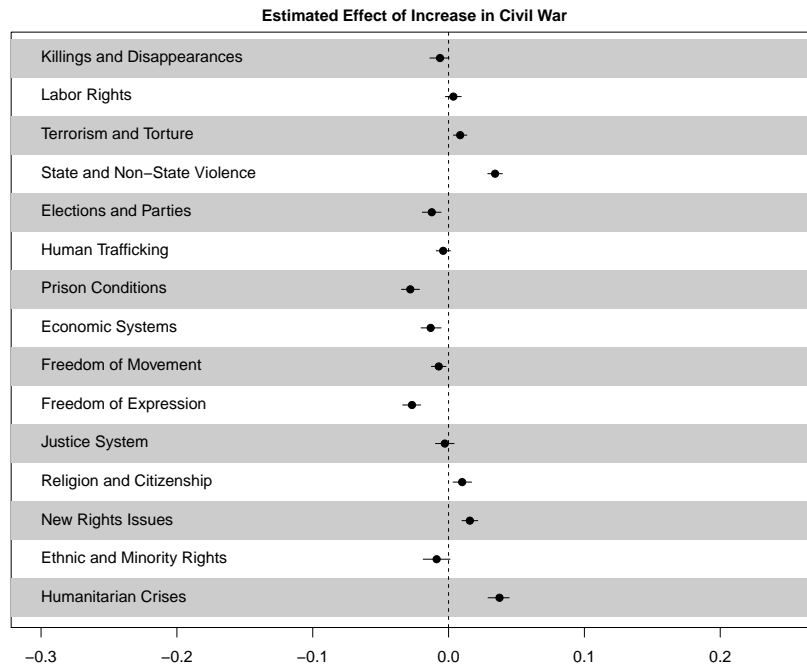


Figure A.19: Covariate relationship with each topic, based on the estimated effect of an increase in the covariate on the expected prevalence of each topic. Increases are from 0 to 1 for dichotomous variables, and from one standard deviation below the mean to one standard deviation above for others. Horizontal lines show 95 percent confidence intervals.

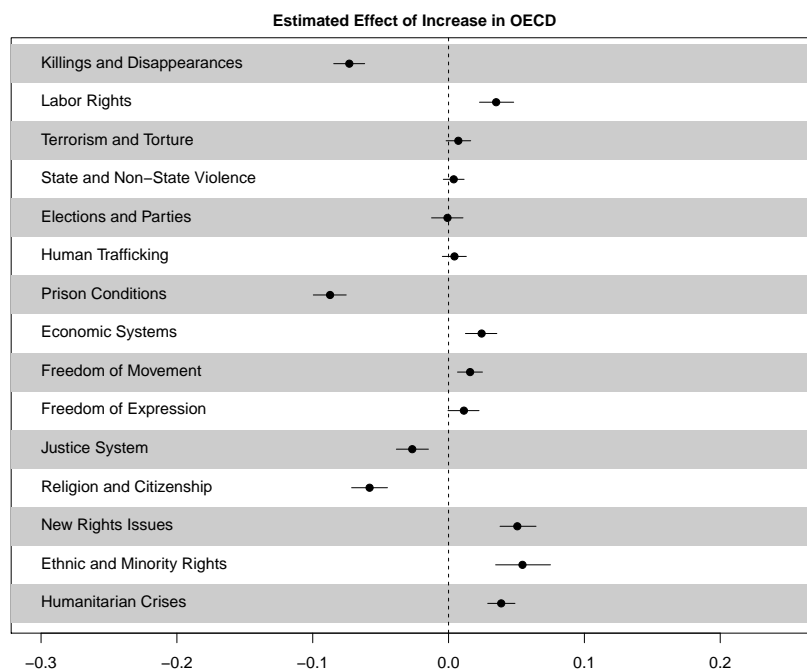


Figure A.20: Covariate relationship with each topic, based on the estimated effect of an increase in the covariate on the expected prevalence of each topic. Increases are from 0 to 1 for dichotomous variables, and from one standard deviation below the mean to one standard deviation above for others. Horizontal lines show 95 percent confidence intervals.

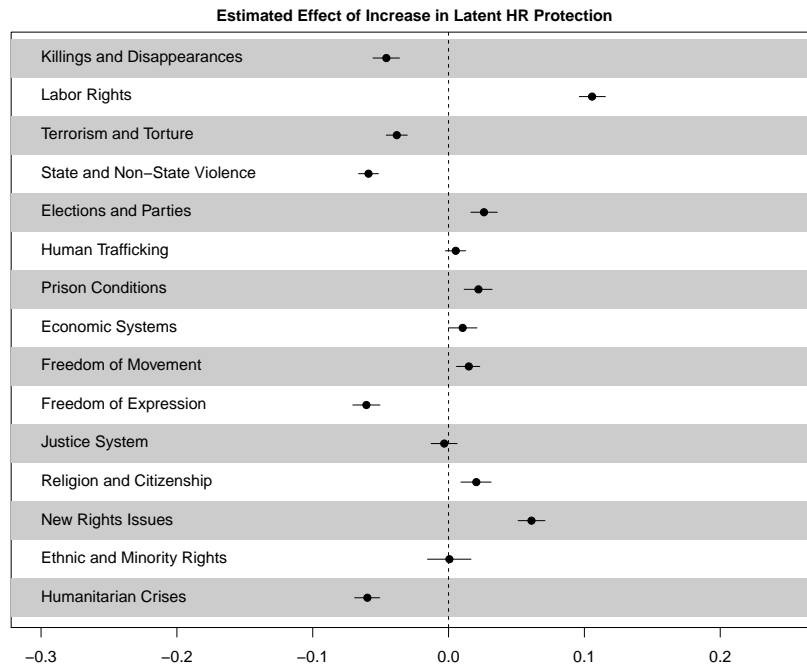


Figure A.21: Covariate relationship with each topic, based on the estimated effect of an increase in the covariate on the expected prevalence of each topic. Increases are from 0 to 1 for dichotomous variables, and from one standard deviation below the mean to one standard deviation above for others. Horizontal lines show 95 percent confidence intervals.

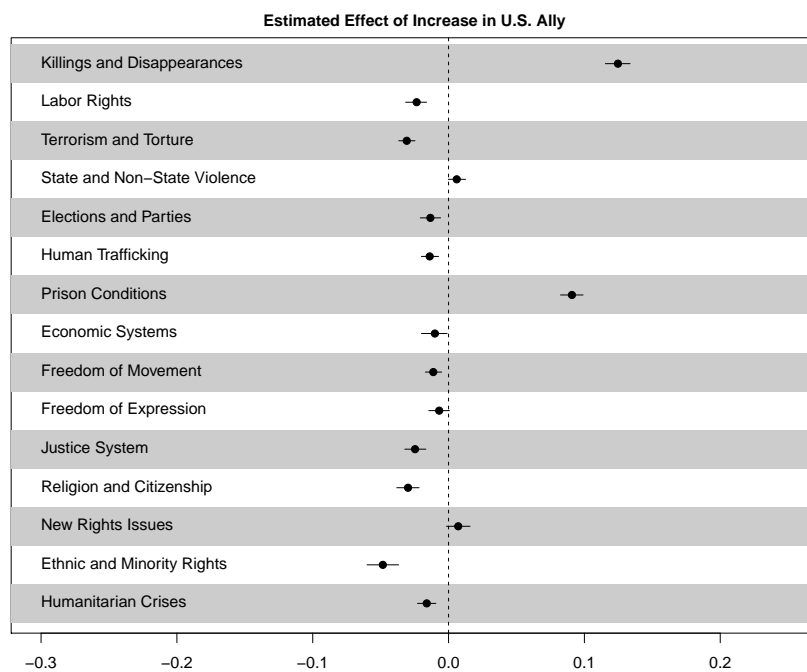


Figure A.22: Covariate relationship with each topic, based on the estimated effect of an increase in the covariate on the expected prevalence of each topic. Increases are from 0 to 1 for dichotomous variables, and from one standard deviation below the mean to one standard deviation above for others. Horizontal lines show 95 percent confidence intervals.

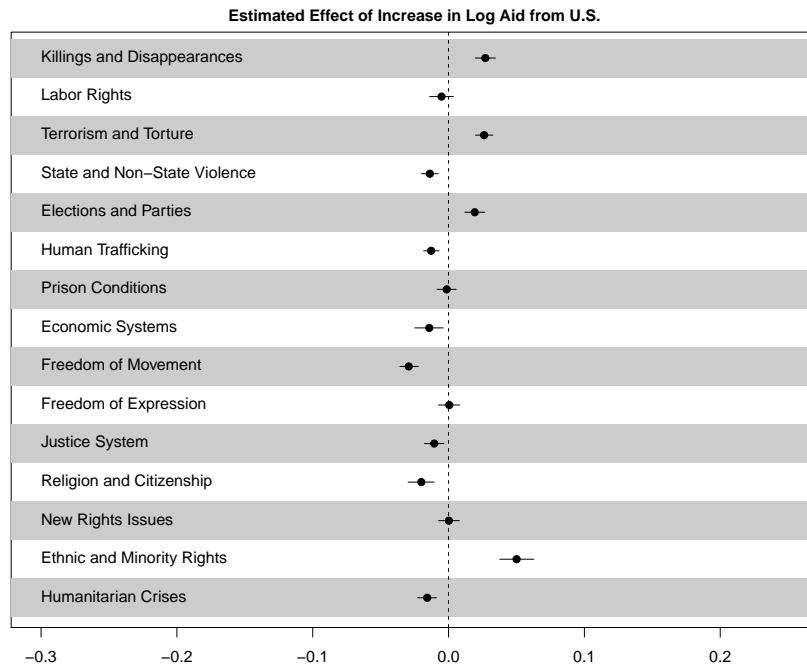


Figure A.23: Covariate relationship with each topic, based on the estimated effect of an increase in the covariate on the expected prevalence of each topic. Increases are from 0 to 1 for dichotomous variables, and from one standard deviation below the mean to one standard deviation above for others. Horizontal lines show 95 percent confidence intervals.

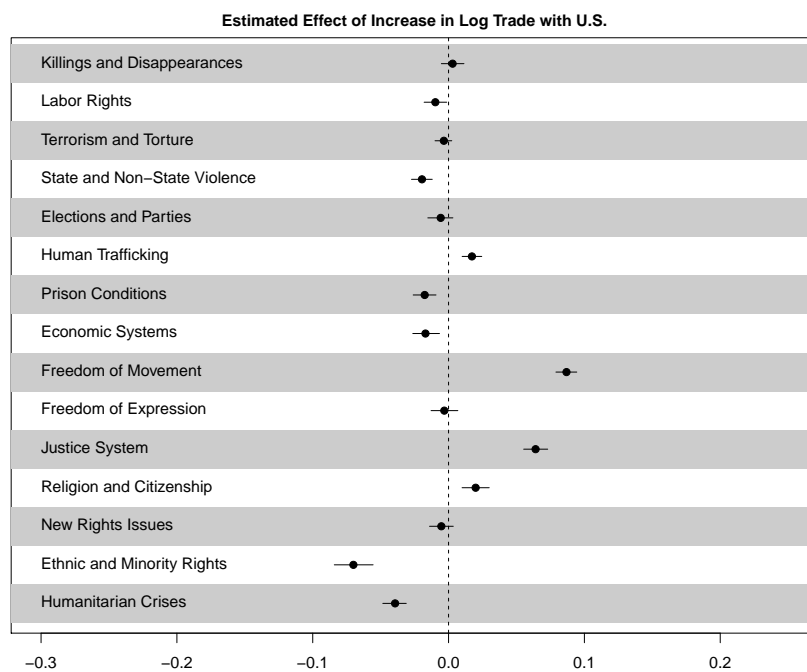


Figure A.24: Covariate relationship with each topic, based on the estimated effect of an increase in the covariate on the expected prevalence of each topic. Increases are from 0 to 1 for dichotomous variables, and from one standard deviation below the mean to one standard deviation above for others. Horizontal lines show 95 percent confidence intervals.

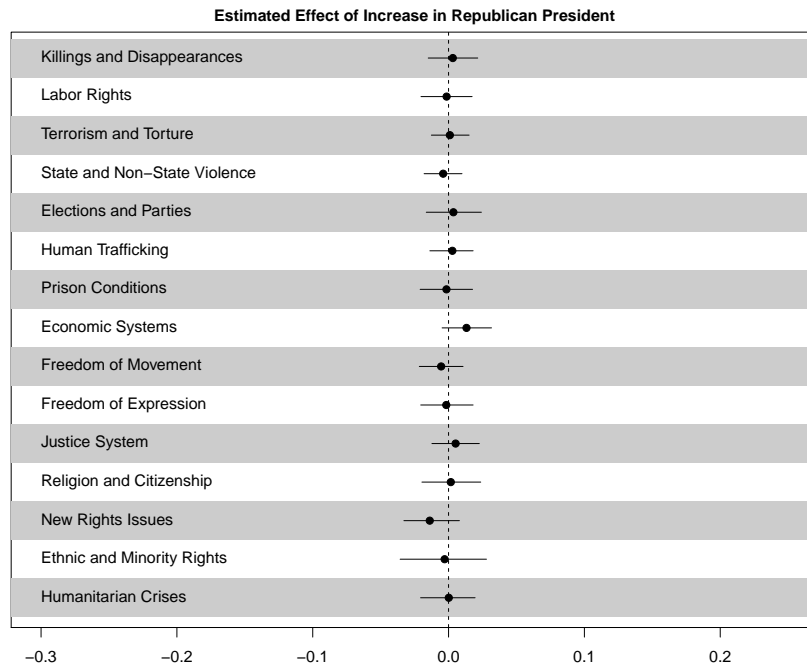


Figure A.25: Covariate relationship with each topic, based on the estimated effect of an increase in the covariate on the expected prevalence of each topic. Increases are from 0 to 1 for dichotomous variables, and from one standard deviation below the mean to one standard deviation above for others. Horizontal lines show 95 percent confidence intervals.

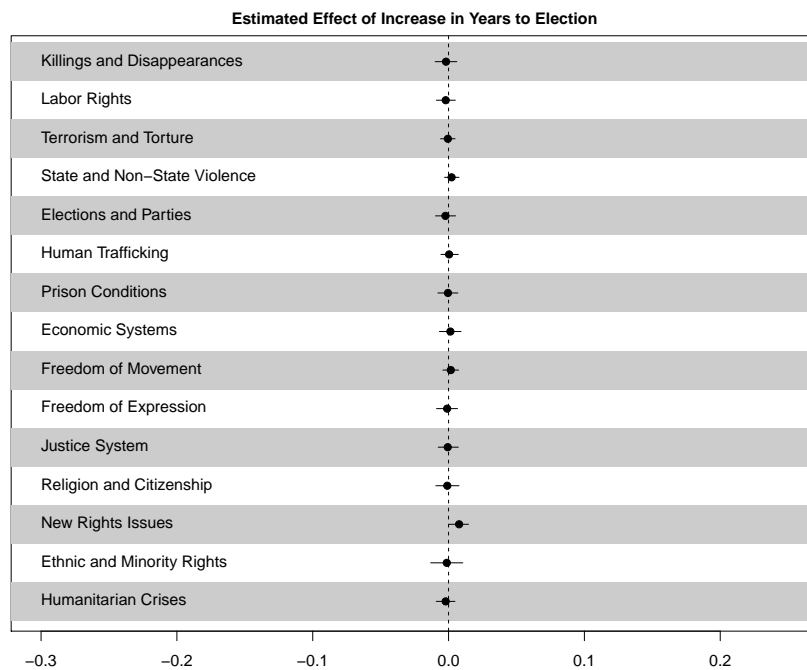
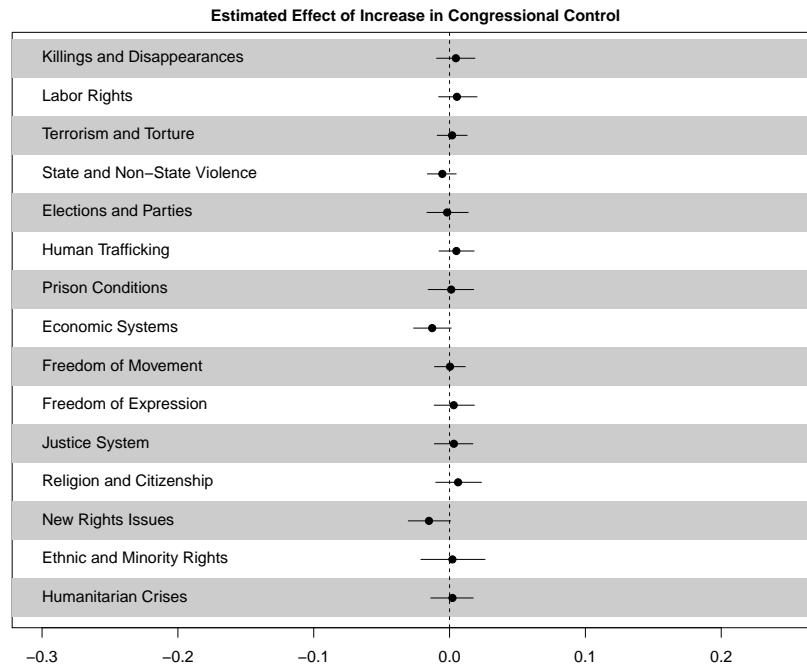


Figure A.26: Covariate relationship with each topic, based on the estimated effect of an increase in the covariate on the expected prevalence of each topic. Increases are from 0 to 1 for dichotomous variables, and from one standard deviation below the mean to one standard deviation above for others. Horizontal lines show 95 percent confidence intervals.



10 and 20-Topic Models

Our main paper’s primary analysis focused exclusively on a 15 topic STM. Recall that because researchers must choose the topic number to be evaluated within STMs, this 15 topic STM represents only one of many plausible STMs for our corpus. As discussed in our *Modeling Approach* section, we selected a 15 topic STM based upon the guidance offered by STM researchers as well as past applications of topic models to international relations corpora. Even so, because there is no single correct answer to one’s choice of topic numbers within STMs, it is important to evaluate the stability of our identified topics across a wider range of topic numbers. This section does so qualitatively by examining the topics identified within comparable 10 and 20 topic STMs, and comparing these topics’ topwords to the 15 topic STM reported in our main paper. We estimate these 10 and 20 topic STMs using the same approach that was employed in generating our primary 15 topic STM: we employ different starting parameter values to estimate a set of 50 STM initializations for both our 10 and 20 topic models, and select a best performing “final” STM for each topic number from these initializations based jointly

upon exclusivity and semantic coherence statistics.

We report the top 20 wordstems, as based upon frequency exclusivity scoring metrics, for our 10 and 20 topic STMs within Table A.27 and Table A.46 below. Based upon these reported topwords, we generally find that the topics identified by our 10 and 20 topic STMs are highly consistent with the 15 topics identified and discussed within our primary STM, and label the former sets of topics accordingly. These similarities suggest that we are indeed identifying stable and coherent topics in our primary empirical analysis. Yet, our qualitative assessments of the 10 and 20 topic STM topwords also suggest that these topic models yield suboptimal topics relative to our 15 topic model. Specifically, we find that the 10 topic model tends to conflate a number of the most coherent topics identified by the 15 topics STM, whereas the 20 topic model appears to overly separate a number of meaningful topics—in some cases yielding topics with high degrees of conceptual overlap with each other.

For example, comparing the results of the 10-topic model to the main results presented in the paper, one can see that several topics no longer appear, or appear ‘merged’ together. For example, no Labor Rights topic is present, and many of the topwords associated with that topic in the main results now appear associated with various other topics instead, but always subsidiary to different main concepts. The same could be said for the State and Non-State Violence topic. Human Trafficking and New Rights Issues each now overlap with one another, across Topic 3 and 6, rather than appearing as two distinct topics. Words reflecting Freedom of Expression appear largely split between Topic 5, which now primarily reflects elections, and Topic 8, where they are combined with words reflecting protest, dissidents, and Freedom of Movement. Based on these differences, we conclude that the 15-topic model offers more coherent and distinct estimated topics than the 10-topic model.

Comparing the results of the 20-topic model to the main results presented in the paper, one can see that several new topics now appear, and several topics from the 15-topic model now seem ‘split apart’ into multiple separate topics. A close consideration of many of these, however, leads us to conclude that the additional topics in the 20-topic model do not offer substantial gains in offering coherent, distinct, and relevant topics for interpretation and analysis. Topics 5 and 6, for example, overlap considerably in the concepts they capture, as do Topics

4 and 20. Importantly, the three topics from the 15-topic model results that primarily reflected violence, conflict, and violations of physical integrity rights, remain largely identical in the 20-topic model results, appearing here as Topics 15, 16, and 19.

We believe that these patterns reinforce our decision to focus on a 15 topic model, as opposed to a 10 or 20 topic model, in our primary analysis. This conclusion is also supported by the results of model-fit statistics discussed below.

Figure A.27: Topwords for 10-Topic STM

Topic	Top 20 Words	Labels
1	roma, prosecutor, parliament, russian, ethnic, traffick, orthodox, region, church, minor, parlamentari, european, properti, ombudsman, ngos, soviet, regist, languag, registr, pretrial	Ethnic and Minority Rights
2	soldier, rebel, civilian, arm, displac, north, conflict, guerrilla, war, idp, armi, paramilitari, kill, militari, militia, forc, fight, humanitarian, attack, icrc	Humanitarian Crises
3	sexual, discrimin, societ, access, traffick, prohibit, asylum, internet, protect, exploit, child, disabl, gender, corrupt, provid, victim, penalti, complaint, law, freedom	New Rights Issues
4	guarante, econom, emigr, recent, african, peopl, economi, rate, now, amnesti, current, will, black, polici, communist, growth, popul, freeli, degrad, coup	Economic Systems
5	opposit, presid, code, presidenti, newspaper, parti, elect, decre, demonstr, candid, independ, interior, associ, constitut, elector, assembl, polit, branch, vote, televis	Elections and Parties
6	girl, child, station, beat, rape, rural, fgm, magistr, ngos, district, tradit, end, farm, radio, howev, unlik, children, boy, broadcast, see	Human Trafficking
7	terrorist, israel, west, bank, arab, emerg, territori, milit, terror, northern, state, occupi, custodi, bomb, act, tribal, christian, jewish, order, east	Terrorism and Torture
8	activist, provinc, chines, china, sentenc, offici, protest, site, religi, internet, imprison, dissid, provinci, block, migrant, citi, foreign, author, detain, lawyer	Freedom of Movement
9	indigen, congress, offic, land, feder, investig, murder, judg, complaint, inmat, labor, prosecutor, compani, senat, crime, victim, worker, commiss, drug, polic	Killings and Disappearances
10	islam, muslim, king, employ, foreign, sharia, male, citizenship, prime, worker, servant, christian, non-muslim, divorc, citizen, must, appoint, amend, deport, religion	Religion and Citizenship

Figure A.28: Topwords for 20-Topic STM

Topic	Top 20 Words	Labels
1	prosecutor, ethnic, roma, parliament, russian, orthodox, church, soviet, parlamentari, properti, languag, region, minor, televis, russia, pretrial, citizenship, republ, former, wit	Ethnic and Minority Rights
2	chines, church, dissid, china, religi, communist, cathol, polici, peopl, control, abroad, enterpris, provinci, foreign, ethnic, approv, central, minor, cultur, committe	Freedom of Movement
3	opposit, presidenti, independ, newspaper, registr, activist, candid, presid, decre, deni, harass, journalist, refus, regist, ralli, print, demonstr, regim, passport, meet	Political Supression
4	minimum, respect, wage, constitut, labor, sector, hour, age, bargain, enforc, union, compulsori, worker, strike, nongovernment, practic, disabl, prohibit, collect, privat	Labor Rights
5	islam, christian, tortur, interior, muslim, bahai, articl, sentenc, penal, ban, governor, terrorist, ali, moham, emerg, code, book, lawyer, univers, non-muslim	Religion
6	king, sharia, foreign, citizenship, islam, citizen, male, noncitizen, royal, passport, non-muslim, deport, servant, divorc, sponsor, marri, employ, sunni, permiss, worker	Citizenship
7	asylum, seeker, immigr, traffick, victim, european, applic, ombudsman, anti-semit, roma, racial, sexual, deport, complaint, jewish, disabl, discrimin, compar, countri, effect	Asylum and Immigration
8	district, polic, injur, girl, shot, state, custodi, kill, violenc, -year-old, death, mob, rape, see, child, bond, man, bail, cast, riot	Justice System
9	provinci, provinc, judg, inmat, senat, reform, polic, jail, street, prison, justic, cell, common, air, fire, await, branch, poor, yet, less	Prison Conditions
10	fgm, tradit, region, ethnic, presid, coup, presidenti, franc, radio, farm, opposit, african, customari, rural, gendarmeri, elector, formal, urban, french, girl	Elections and Parties
11	garante, amnesti, regim, econom, now, emigr, current, tion, rate, popul, ment, recent, will, economi, polit, cruel, growth, per, degrad, inhuman	Economic Systems
12	activist, migrant, site, provinc, internet, compani, local, protest, accord, web, regist, administr, regul, birth, media, block, agenc, confess, face, sentenc	Freedom of Expression
13	traffick, prostitut, victim, ngos, problem, child, children, exploit, approxim, girl, sexual, train, howev, ngo, provid, year, see, domest, assist, program	Human Trafficking
14	act, parliament, magistr, prime, industri, minist, amend, parlamentari, opposit, bail, seat, power, trade, chief, rule, order, commission, politician, ralli, board	British Legacy
15	soldier, rebel, civilian, idp, war, displac, conflict, arm, humanitarian, militia, fight, abduct, north, see, camp, forc, provinc, attack, numer, rape	Humanitarian Crises
16	indigen, guerrilla, congress, paramilitari, kidnap, prosecutor, murder, san, militari, threat, jose, kill, municip, armi, depart, ombudsman, carlo, arm, drug, investig	Killings and Disappearances
17	israel, west, black, bank, south, arab, territori, occupi, african, white, africa, jewish, strip, east, resid, detaine, closur, interrog, secur, land	Occupation
18	corrupt, sexual, societ, access, child, gender, discrimin, penalti, stateless, birth, share, internet, pretrial, ngos, violenc, rape, orient, lack, protect, transpar	New Rights Issues
19	bomb, insurg, milit, armi, east, terrorist, muslim, kill, disappear, extrajudici, tribal, iraq, emerg, villag, special, paramilitari, attack, personnel, civilian, separatist	State and Non-State Violence
20	feder, legisl, sex, northern, safeti, equal, race, bargain, commiss, wage, industri, democrat, trade, benefit, agreement, provis, minimum, respect, review, bill	Law and Equality

Topic Number Validation

The conclusions discussed above strongly suggest that our main paper’s 15 topic STM is an ideal model for our analyses, as based upon a qualitative assessment of the topwords and topics obtained under 10, 15, and 20-topic STMs. The present section further validates this 15 topic-number choice via a more objective selection of model fit diagnostics. While we agree with Roberts et al. (2014) that the choice of topic number should be primarily based on theoretical grounds, and on qualitative readings of topics and documents, we view model selection diagnostics as a compelling piece of secondary evidence in this respect—given their objectivity. Here, we focus on the exclusivity and semantic cohesion metrics discussed in our main paper (and below), and use these model fit diagnostics to evaluate the three STMs discussed above. Recall that these specific STMs were each identified as being the best performing model for a chosen k , according to the starting parameter values that yielded a Pareto optimal set of semantic cohesion and exclusivity scores for that k across 50 initializations. Hence, the 10, 15, and 20-topic models evaluated above (and below) represent the *best performing* model for a chosen k . We accordingly calculate each final model’s mean and median semantic cohesion and exclusivity measures—across that model’s k topics—and then compare our models based upon these model fit statistics.

This exercise provides us with two semantic cohesion metrics and two exclusivity metrics for each comparison model. Semantic cohesion captures the external consistency of a given topic, via the co-occurrence of topwords within one’s identified topics (Roberts et al., 2014), where more positive values represent higher semantic cohesion and thus better model fit. Exclusivity instead measures how exclusive one’s top words are to each topic based upon each top word’s relative probabilities of association across topics. Thus, unlike semantic cohesion, exclusivity penalizes topics that are too much alike (1070 Roberts et al., 2014). As was the case for semantic cohesion, larger (positive) exclusivity values again correspond to better performing models. For each comparison model, we choose to report and evaluate both mean and median model fit metrics¹ so as to ensure that our comparisons are not overly influenced by

¹As mentioned above, here we are averaging (or taking medians) across the individual semantic cohesion and exclusivity measures recorded for each topic within each model.

any single strong or weak performing outlier topic. These quantities are reported in Table A.2, where in addition to our mean and median exclusivity and semantic cohesion metrics, we also rank each STM from “1 = best” to “3 = worst” according to a given metric, and provide our three STM’s averaged ranking (i.e., across all four metrics) in the final column to this table.

Table A.2: Model Fit Diagnostics for $k = \{10, 15, 20\}$ STMs

Topic	Mean	Median	Mean	Median	Average
Number	Exclusivity	Exclusivity	Semantic Cohesion	Semantic Cohesion	Ranking
10	8.62	8.61	-1.66	-0.93	2.00
15	8.85	8.96	-4.82	-2.04	1.75
20	8.99	8.95	-5.16	-2.26	2.25

Turning to these quantities (Table A.2), we find that our model fit results generally support the qualitative conclusions drawn above. With regards to (mean and median) semantic cohesion, the 15-topic STM yields the second highest semantic cohesion measures among the three k ’s evaluated. By comparison, the 10-topic model consistently ranks first based upon semantic cohesion whereas the 20-topic model performs worst among our three models in terms of both mean and median semantic cohesion. However, with respect to our mean and median exclusivity measures, we instead find that the 10-topic STM is the worst performing model in both cases, whereas our 15-topic model yields the highest overall *median* exclusivity score and the second highest overall *mean* exclusivity score. Taken together, these results indicate that while the 15-topic model does not consistently outperform the 10 and 20-topic models across all metrics, it also never comes in last place for any given metric — something that cannot be said for either the 10 or 20-topic STMs. As a result, our mean placement rankings for each STM—across all four metrics—favor the 15-topic model (with an average ranking of 1.75) over the 10 and 20-topic models (with average rankings of 2.00 and 2.25, respectively). Like the qualitative evidence discussed above, these results imply that the 15 topic model is the single best performing across our k ’s evaluated and all initializations (50 STM initializations in total).

To further assess these findings, we next broaden the range of STMs under evaluation. Specifically, we follow the STM approach discussed earlier to estimate comparable “best performing” 5 and 25-topic STMs by effectively identifying the best performing STM model for each additional k according to 50 different STM initializations. We then derive (mean and median) semantic cohesion and exclusivity metrics for our optimal 5 and 25-topic STMs, and report these model fit diagnostics alongside those for our 10, 15, and 20-topic STMs in Table A.3 below. By and large, these additional STM models perform as expected. The 5-topic STM achieved mean and median semantic cohesion metrics that are each superior to those of any of the STMs discussed thus far, but also yielded what are by far the worst exclusivity metrics across all k ’s evaluated. By contrast, the 25-topic STM outperforms all other models in terms of exclusivity, but comes in dead last in terms of semantic cohesion. Ranking our 5 to 25-topic STMs according to a scale ranging “1 = best” to “4 = worst” for each model fit diagnostic and calculating each STM’s average ranking, we again find in Table A.3 that our 15-topic STM performs the best across all four metrics, with an average ranking (2.75) that is notably lower than a majority of the remaining for STMs (with average rankings falling between 3.00 and 3.25); save for the $k = 10$ model which also scores 2.75 in this context.

Table A.3: Model Fit Diagnostics for $k = \{5, 10, 15, 20, 25\}$ STMs

Topic	Mean	Median	Mean	Median	Average
Number	Exclusivity	Exclusivity	Semantic Cohesion	Semantic Cohesion	Ranking
5	8.22	8.36	-0.85	-0.94	3.25
10	8.62	8.61	-1.66	-0.93	2.75
15	8.85	8.96	-4.82	-2.04	2.75
20	8.99	8.95	-5.16	-2.26	3.25
25	9.06	9.15	-8.60	-5.27	3.00

100-Topic STM

To further evaluate the stability of our primary (15-topic STM) topics, as well as to better highlight the promise of STMs for the identification of salient themes within the State Department’s Country Reports more generally, this section estimates and interprets a far larger (100-topic) STM as applied to the same texts discussed in our primary paper. For each of our 100 identified topics, we present the top five most highly associated words according to frequency-exclusivity scoring metrics in Tables A.4-A.5 below.²

We can begin by noting a number of similarities between the 100-topic STM’s identified topics and those of our primary 15-topic STM. For instance, the Killings and Disappearances topic identified by our 15-topic STM is very similar to topic 70 of the 100-topic STM, with four of the top five words from topic 70 also being found within the Killings and Disappearances topic. In addition, topics 20 (more general killing-oriented events) and 27 (insurgent and atrocity-related killings) of the 100-topic STM also share some similarities with our Killings and Disappearances topic. The second primary STM topic (Labor Rights) can be found within topic 7 of the 100-topic STM; where “union” appears in both topword sets, and where the remaining top five words for topic 7—“worker,” “strike,” “employee,” and “sector,”—each also clearly capture labor rights issues. Some elements of our 15-topic STM’s topic 3 (Terrorism and Torture) have clear overlap with the 100-topic STM’s topic 4—which appears to be specific to the Israeli-Palestinian conflict—as all five topwords can be found within the 20 topwords reported for Terrorism and Torture. The more general elements of our Terrorism and Torture topic are clearly found in topic 16 of the 100-topic STM, whose top five words include “terrorist,” “terror,” “bomb,” and “torture.” Similarly, numerous elements of our State and Non-State Violence topic can be coherently found in the 100-topic STM’s topics 95 and 89.

The elections component to topic 5 from our original 15-topic STM, which we referred to as Elections and Parties can clearly be found in the 100-topic STM’s topic 18, with overlapping topwords of “presid,” “perisidenti,” and “elector”³ whereas topic 65 now appears to be capturing the party-dimension of our Elections and Parties topic, with topwords such as “opposit,”

²Note that we also reviewed the output of our 100-topic STM more closely (e.g., by examining the full 20 topwords for each topic) before reaching the conclusions discussed below.

³And with the latter now also including the words “candid,” and “vote.”

‘parti,’ and ‘seat.’ Topic 6 (from the 15-Topic STM), labeled Human Trafficking, most directly corresponds to topic 75 within the 100-topic STM, with all five topwords for the latter appearing within our Human Trafficking topic’s topwords. Our Prison Conditions and Justice System topics likewise exhibit systematic overlap with a number of the 100-topic STM’s topics, most notably topics 48 and 13, but also topics 42, 57, and 97. Comparably, the Freedom of Movement and Freedom of Expression topics from our 15-topic STM exhibit overlap in topwords with the 100-topic STM’s topics 8 and 10, respectively, as well as with topic 28. Topic 8 from the 15-topic STM (i.e., Economic Systems) shares a striking similarity with topic 5 in the 100-topic STM, wherein we find that four of the latter’s top five words also appear in Economic Systems’ topwords. The same can be said for Religion and Citizenship (topic 12 in the original 15-topic STM) and topic 80 in the 100-topic STM. Our aforementioned New Rights Issues topic can be found most clearly in topics 58 and 61 (which each appear to be about asylum seeking) and topic 67 (a more general new rights issue topic) of the 100-topic STM. These latter topics also overlap with our previously identified Ethnic and Minority Rights topic, which additionally shares elements with topics 87 and (a more European-oriented) topic 64. Finally, we can note that the topwords found within topic 15 (Humanitarian Crises) from our original 15-topic STM encompass four of the five topwords found within the 100-topic STM’s topic 81, and also share similarities with topic 12.

The above evidence suggests that the topics identified within our 15-topic STM are relatively stable, as each can be recovered even when the level of granularity of one’s topics is significantly expanded. Concurrently, our 100-topic STM also identifies a number of additional, highly coherent topics. These additional topics were not found within our more general 15-topic STM, but may be of interest to future researchers seeking to deploy STMs for more fine grained codings of the State Department’s Country Reports, and as such, we highlight a few examples here. Notably, Topic 35 appears to relate to rights abuses associated with child soldiers, with topwords such as “child,” “kidnap,” and “children.” Topic 63 of the 100-topic STM is clearly capturing protests, with its top five words being “demonstr,” “student,” “protest,” “riot,” and “dispers.” Topic 72 appears to relate to internet censorship, and includes “internet,” “site,” “web,” and “block” within its top five words. Interestingly, topic 94 appears to encom-

pass State Department language denoting improvements in a country’s human rights records, with topwords such as “increas,” “improv,” and “effor.” In addition, topic 66 directly pertains to police abuses, and includes topwords such as “beat,” “polic,” and “frequent.” Quite distinctly, we also find separate topics in this larger STM that appear to be related to emigration (topic 19) and immigration (topic 83). Finally, topic 69 clearly captures discussions of monarchies with topwords such as “king,” “royal,” and “monarchi,” whereas topic 84 corresponds to the Catholic Church (“cathol,” “church,” “priest,” “dissid,” “mass”).

At the same time, a number of features of the 100-topic STM make this topic model less useful for our primary interest in identifying the varying attention to *human rights* issues across the State Department’s Country Reports. For instance, the 100-Topic STM now identifies a large number of country and/or region specific topics, rather than human rights topics. Examples include topics exclusively related to Russia/USSR (topic 32), Israel/Palestine (topic 4), the Middle East in general (topic 46), U.S.-Latin American relations (topic 41), Northern Ireland (topic 54), China (topic 100), and France (topic 23). In addition, many of the topics identified by the 100-topic STM lack any coherent interpretation, whether one examines the top five words reported in Tables A.4-A.5 or the larger topwords and associated documents. For example, topic 2 appears to include transition words of little direct meaning, whereas topic 96 encompasses words used to refer back to other sections of a given Country Report, and topic 68 appears to lack any coherent theme of note. Hence, taken together, our assessment of the 100-topic STM suggests that the topics identified by our 15-topic STM are largely coherent and stable, and that these general human rights topics can be more efficiently estimated by a relatively conservative 15-topic STM than by a larger 100-topic STM. Nevertheless, the novel, and theoretically coherent, topics identified by the 100-topic STM suggest that an STM of this size may be preferable for those interested in identifying and coding more fine grained (and region-specific) human rights themes and trends from these text data.

15-Topic Topword Results When Using LDA and K-Means

In this section, we re-estimate the 15 latent topics underlying our Country Reports using (i) a latent Dirichlet allocation (LDA) model and (ii) k-means. For the LDA model, we assign the

Table A.4: 100-Topic STM Topwords (Topics 1-60)

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
asylum	still	respect	israel	guarante	church	union	feder	provinc	district
european	yet	prohibit	bank	amnesti	orthodox	worker	territori	provinci	cast
anti-semit	recent	pretrial	arab	rate	jehovah	strike	indigen	air	bond
applic	late	discrimin	west	econom	wit	employe	nativ	decre	custodi
german	indic	practic	occupi	loan	relig	sector	vari	martial	act
Topic 11	Topic 12	Topic 13	Topic 14	Topic 15	Topic 16	Topic 17	Topic 18	Topic 19	Topic 20
governor	idp	court	idp	ngos	terrorist	sharia	candid	permiss	kill
policemen	soldier	appeal	war	lack	emerg	foreign	presid	travel	attack
citi	arm	convict	republ	corrupt	terror	sponsor	vote	permit	end
murder	north	case	conflict	disabl	bomb	mail	presidenti	abroad	shot
uniform	rape	judge	properti	facil	tortur	noncitizen	elector	past	unknown
Topic 21	Topic 22	Topic 23	Topic 24	Topic 25	Topic 26	Topic 27	Topic 28	Topic 29	Topic 30
christian	film	franc	amend	black	minist	insurg	constitut	prosecutor	parti
articl	broad	french	act	african	prime	civilian	televis	ombusman	communist
convert	can	gendarmeri	magistr	south	general	icrc	independ	jewish	socialist
penal	normal	magistr	appoint	white	administ	kill	radio	pretrial	appar
governor	ground	leagu	meet	africa	part	conflict	citizen	anti-semit	control
Topic 31	Topic 32	Topic 33	Topic 34	Topic 35	Topic 36	Topic 37	Topic 38	Topic 39	Topic 40
assult	russian	bahai	girl	region	farm	parliament	freeli	share	migrant
evict	soviet	ali	rural	child	bail	parliamentari	nongovernment	birth	activist
owner	russia	islam	station	kidnap	magistr	monitor	minimum	stateless	sensit
hivaid	presidenti	cleric	child	children	treason	minor	compulsori	orient	birth
facto	regist	guardian	school	ngos	farmer	languag	extrajudici	gay	bureau
Topic 41	Topic 42	Topic 43	Topic 44	Topic 45	Topic 46	Topic 47	Topic 48	Topic 49	Topic 50
congress	secur	agricultur	citizenship	adolesc	iraq	activist	prison	coup	accord
san	detent	rural	resid	senat	sunni	journalist	accord	decre	district
inter-american	detaine	urban	natur	judg	arab	lawyer	inmat	transit	countri
suprem	releas	subsist	noncitizen	penitentiari	jew	ban	condit	tribun	agenc
peasant	prison	popul	elig	superior	societi	tortur	medic	militari	develop
Topic 51	Topic 52	Topic 53	Topic 54	Topic 55	Topic 56	Topic 57	Topic 58	Topic 59	Topic 60
inmat	customari	nation	northern	state	east	limit	asylum	southern	militari
birth	mob	govern	british	trade	west	system	island	north	civilian
parent	tradi	group	kingdom	commiss	regul	counsel	compliant	militia	armi
director	fgm	program	act	recommend	agreement	trial	seeker	south	disappear
certif	dispers	year	unit	econom	shot	defend	ombudsman	camp	commiss

Table A.5: 100-Topic STM Topwords (Topics 61-100)

Topic 61	Topic 62	Topic 63	Topic 64	Topic 65	Topic 66	Topic 67	Topic 68	Topic 69	Topic 70
asylum	prostitut	demonstr	roma	opposit	beat	corrupt	tion	king	indigen
seeker	enterpris	student	minor	parti	police	law	ment	royal	ombudsman
ombudsman	factori	protest	restitut	ralli	frequent	traffick	martial	western	prosecutor
sexual	child	riot	european	rule	routin	internet	subvers	monarchi	municip
discrimin	shelter	dispers	unemploy	seat	often	child	inter	kingdom	drug
Topic 71	Topic 72	Topic 73	Topic 74	Topic 75	Topic 76	Topic 77	Topic 78	Topic 79	Topic 80
milit	internet	interior	dissid	traffick	southeast	howev	children	fgm	islam
tribal	site	sentenc	enterpris	prostitut	villag	year	disabl	slaveri	sharia
ordin	web	council	congress	victim	separatist	countri	abus	girl	muslim
tribe	media	protest	central	exploit	foundat	previous	women	marriag	christian
extremist	block	author	peopl	sexual	secular	person	child	unhcr	mosqu
Topic 81	Topic 82	Topic 83	Topic 84	Topic 85	Topic 86	Topic 87	Topic 88	Topic 89	Topic 90
rebel	local	immigr	cathol	refus	act	ethnic	ngo	guerilla	media
soldier	confess	deport	church	newspap	amend	return	ngos	paramilitari	television
war	offici	illeg	priest	station	restrict	former	registr	kidnap	pressur
troop	harass	migrant	dissid	seiz	coalit	citizenship	author	depart	station
fight	candid	pend	mass	credibl	bar	languag	observ	massacr	observ
Topic 91	Topic 92	Topic 93	Topic 94	Topic 95	Topic 96	Topic 97	Topic 98	Topic 99	Topic 100
violenc	code	land	new	regim	see	alleg	state	legisl	china
remain	ministri	factori	increas	revolutionari	section	defend	continu	equal	chines
intimid	penal	camp	report	execut	relig	attorney	deni	respect	basic
peac	provis	unhcr	improv	revolut	number	end	refus	bargain	north
littl	judg	border	effort	rappporteur	order	case	individu	protect	ordin

model’s α parameter to the commonly assigned value of 0.1, and use Gibbs sampling with a single chain of 2,000 iterations (treating the first 1,000 iterations as burn in). For k-means, we set our clusters to 15, wherein k-means then seeks to effectively assign each of our documents into a single one of these 15-clusters. The results for the fifteen topic LDA model appear in Figure A.29, where, as was the case for our primary STM, we report the words most associated with each identified topic according to frequency-exclusivity scoring metrics. We derive a comparable set of topwords for our 15-cluster k-means approach in Figure A.30 by transforming each unique word’s vector of cluster centers into proportions⁴ and retaining the highest 20 word proportions for each of our 15 identified document clusters.

Turning first to our LDA topwords, one can first note that topic 1 from this LDA model exhibits substantial overlap with topic 5 (Elections and Parties) from our primary STM, with shared topwords such as “presid,” “radio,” “presidenti,” “magistr,” and “elector.” By contrast,

⁴I.e., by dividing each word-vector indexed cluster center by the sum of the cluster centers for that vector.

topic 2 appears to be fairly unique to the LDA model, largely capturing political, mass-media, and social topics associated with Russia and the former USSR, where one can note that “russian,” “soviet,” and “russia” all appear within this LDA topic’s 20 topwords. Topic 3 bears a striking resemblance to topic 2 (‘Labor Rights’) from our primary STM, with the former’s top six words all appearing within our STM’s Labor Rights topics’ topwords, and 12 overlapping topwords in total. Likewise, topic 4 shares a majority of its topwords with our 15-Topic STM’s Religion and Citizenship topic (i.e., topic 12), while LDA topic 5 shares a great many of its topwords with our STM model’s topic 10 (Freedom of Expression) and LDA topic 6 shares the majority of its topwords with the 15-Topic STM’s Human Trafficking topic (i.e., topic 6). To a slightly lesser degree, LDA topic 7 displays a fair amount of resemblance to topic 9 (Freedom of Movement), with roughly half of these two topics’ topwords exhibiting overlap across the STM and LDA model.

Turning next to the bottom half of Figure A.29, topic 8 is highly comparable to topic 13 (New Rights Issues) from our primary STM, with shared topwords such as “sexual,” “corrupt,” “discrimin,” and “protect.” Topic 9 exhibits a degree of resemblance to STM topic 3 (Terrorism and Torture), although Topic 9’s topwords suggest that this LDA topic may also be more directly capturing the Israeli-Palestinian conflict. The LDA model’s topic 10 bears a striking resemblance to topic 8 (Economic Systems) from our primary STM, including its sharing of the same top three topwords with this STM topic. Similarly, the vast majority of topwords found within LDA topic 11 can be found in STM topic 1, which suggests that this LDA topic is largely capturing the Killings and Disappearances theme discussed in our main paper. Topic 12 in Table A.29 exhibits a close correspondence with topic 6 (Humanitarian Crises) from our primary STM, with shared topwords such as “rebel,” “soldier,” “civilians,” “displace,” and “kill.” The topwords that appear under topic 13 in Table A.29 have a fair amount of overlap with STM topic 14 (Ethnic and Minority Rights), although it would appear that some of the STM model’s topwords in this regard have been assigned to LDA Topic 2 (which, as mentioned above, appears to be capturing Russia and USSR-specific themes). To some degree, LDA topic 14 bears a resemblance with STM topic 11 (Justice System) as both include topwords such as “district,” “custodi,” “polic” and “bond.” Finally, topic 15 in the LDA model does not have a

close match to any of the 15 primary STM topics, though it appears to primarily be capturing British Colonial legacies and racial issues in (South) Africa.

Altogether, the similarities and differences between the LDA-identified topics and our primary STM-identified topics offer several insights. First and foremost, these similarities suggest that the 15 latent topics that we identify and discuss in the main paper are highly robust to estimation strategy, which in turn suggests that these 15 topics are coherent features of the Country Reports texts, rather than idiosyncracies of any particular topic modeling strategy. Furthermore, the fact that we see many of the same topics reflected across our 15 topic STM and LDA specifications also indicates that the conclusions reached in the main paper, with respect to topic interpretation, do not disproportionately rest on our choice of an STM over an LDA model. At the same time, we also note that the LDA model appears to give greater weight to region-specific words in several topics than does our primary STM results that incorporate country-level covariate information. For the purposes of our paper, which endeavors to identify systematic patterns and biases in the State Department's Country Reports, these country or region-specific topics are of less interest, which further justifies our decision to favor the STM for our primary analysis.

Figure A.29: Topwords for 15-topic LDA model

Topic	Top 20 Words
1	fgm; tradit; girl; presid; rural; beat; radio; farm; station; region; presidenti; magistr; ethnic; urban; street; primari; dispers; customari; franc; elector
2	prosecutor; russian; registr; regist; media; region; local; soviet; televis; ethnic; orthodox; parliament; parlamentari; observ; presidenti; wit; station; pressur; pretrial; russia
3	minimum; wage; respect; constitut; bargain; union; strike; collect; free; labor; compulsori; freeli; sector; nongovernment; hour; freedom; trade; interfer; safeti; practic
4	islam; king; sharia; muslim; non-muslim; christian; foreign; male; divorc; sunni; passport; bahai; citizenship; royal; marri; husband; obtain; council; citizen; noncitizen
5	opposit; activist; demonstr; lawyer; decre; releas; deni; articl; journalist; tortur; student; detain; sentenc; newspap; harass; prison; parti; confisc; presid; candid
6	traffick; prostitut; provid; ngos; victim; problem; children; approxim; prohibit; countri; child; practic; domest; sexual; exploit; disabl; ngo; assist; protect; women
7	provinc; chines; china; provinci; dissid; communist; regul; enterpris; offici; church; cathol; peopl; religi; sensit; migrant; central; basic; foreign; protest; citi
8	sexual; corrupt; access; ngos; child; internet; societ; disabl; discrimin; domest; protect; gender; provid; prohibit; birth; penalti; rape; freedom; violenc; exploit
9	israel; west; bank; arab; territori; occupi; iraq; jewish; east; strip; resid; secur; clousur; soldier; detainee; militari; permit; interrog; jew; terrorist
10	guarante; econom; amnesti; emigr; regim; polit; now; recent; popul; assembl; tion; coup; parti; current; ment; economi; freedom; polici; will; rate
11	indigen; congress; guerrilla; murder; prosecutor; kidnap; investig; offic; paramilitari; judg; militari; san; threat; jose; inmat; crime; kill; senat; municip; ombudsman
12	rebel; civilian; soldier; displac; arm; north; idp; war; conflict; kill; armi; humanitarian; forc; militia; fight; attack; abduct; icrc; southern; troop
13	roma; asylum; feder; european; applic; minor; immigr; seeker; ombudsman; anti-semit; equal; racial; interior; victim; europ; complaint; communiti; crime; jewish; percent
14	district; milit; custodi; kill; villag; death; tribal; polic; injur; state; factori; jail; bail; cast; girl; bond; muslim; murder; shot; camp
15	act; south; african; black; parliament; white; africa; amend; industri; magistr; northern; review; british; racial; trade; minist; offens; power; asian; parlamentari

Relative to the LDA case, the results for our k-means approach diverge more significantly from those of the STM, although a number of similarities can be found as well. Before discussing these results, recall that k-means differs from our LDA and STM approaches in that k-means assumes that each document arises from a single distinct topic, rather than from a mixture of topics. Further, because k-means reports cluster centers, rather than probabilities, we cannot as directly derive topwords according to frequency-exclusivity scoring metrics. We instead simply transform each (unique) word's vector of cluster centers to proportions and identify the highest 20 word proportions for each of our 15 identified document clusters. The one drawback to this approach is that in several instances, exceptionally rare words in the corpus make it into some cluster's topwords by virtue of k-means never assigning documents containing these words to a majority of our other clusters. Nevertheless, general themes remain recognizable in many of our 15 clusters in Figure A.30, and it is to these results that we now turn.

We begin by noting a number of similarities between Figure A.30 and our primary 15-topic STM. Cluster 7 shares many of its topwords with our 15-topic STM's Killings and Disappearances topic (Topic 1), including words such as "indigen," "paramilitari," and "indigen." Cluster 4 exhibits a very close correspondence to the Religion and Citizen Topic identified within the 15-topic STM, including shared words such as "king," "islam," "sharia," and "non-muslim." Cluster 9 encompasses many of the same—or similar—topwords to that of our 15-topic STM's Humanitarian Crises topic (Topic 15), with shared words such as "rebel" and "fight." Somewhat more ambiguously, Cluster 15 seems to correspond to our previously identified Ethnic and Minority Rights topic (Topic 14), with the shared highest ranked topword "roma." Finally, the 13th identified cluster in A.30 encompasses several topwords from our 15-topic STM's State and Non-State Violence Topic (e.g., "regim," "iraq"). Hence, taken together, k-means and the STM do appear to be capturing a number of comparable themes—which predominately encompass wartime human rights abuses (or abusers).

The remaining clusters in Figure A.30 cannot be as easily matched to our 15-topic STM's topwords. In some cases this is because our k-means approach appears to be identifying coherent but distinct themes, whereas in other cases k-means seems to be identifying relatively

incoherent sets of topwords. Regarding the former, clusters 2 and 8 are arguably capturing race and ethno-religious rights issues, potentially in the developing world and European contexts respectively. Cluster 3 appears to be clustering Russian and Israeli country reports into a single unique cluster, with topwords “israel,” “russian,” “russia,” “west,” and “strip.” Cluster five may likewise encompass elements of our Freedom of Movement and Freedom of Expression STM-topics, primarily in the Chinese context. At the same time, the topwords reported under Clusters 1, 6, 10, 11, 12, and 14 are much less coherent, often encompassing macro-level social and/or institutional concepts alongside more fine grained geographical or individual-level terms (e.g., Cluster 14 includes “milit” and “sharia” but also “river” and “forest,” as well as “widow”). Given that a full third of our clusters appear to be largely incoherent under k-means, we conclude that this approach—while still potentially useful—may be less ideal for identifying cross-cutting human rights topics than were our previously discussed STM and LDA strategies. This conclusion is consistent with the nature of the Country Reports themselves, wherein each Report typically encompasses multiple disparate themes and issues, rather than the single coherent theme (as k-means presumes).

Figure A.30: Topwords for 15-cluster k-means

Cluster	Top 20 Words
1	southeast; insult; solitari; bahai; hang; monarchi; stateless; pardon; chair; israel; secular; rent; chapter; penitentiari; european; pace; hunger; summon; probat; academi
2	black; color; africa; white; african; tion; south; ment; indian; asian; soviet; though; policeman; bullet; speed; now; southeast; policemen; proscib; probabl
3	israel; russian; russia; west; strip; arab; coloni; east; orthodox; jew; bank; jewish; occupi; hatr; window; cemeteri; synagogu; vandal; barrier; curfew
4	bahai; king; royal; monarchi; moham; embassi; ali; sunni; noncitizen; cleric; exit; convert; sharia; non-muslim; expatri; islam; denomin; soviet; imam; congreg
5	practition; china; chines; anniversari; underground; ordin; dissid; crackdown; ratio; tight; bishop; writer; patriot; subvers; guid; coupl; oversea; pregnanc; autonom; sensit
6	non-government; ment; tion; inter; vital; invas; guarante; inviol; corpus; habea; -parti; french; loan; diet; coup; conscienc; capita; literaci; fulfil; martial
7	guerrilla; paramilitari; carlo; unionist; jose; brigad; massacr; homicid; san; narcot; corp; collabor; peasant; kidnap; lieuten; colonel; forens; indigen; displac; inter-american
8	racist; racism; german; -seat; treati; noncitizen; part-tim; diet; european; anti-semit; chat; asylum; island; hagu; reful; roma; seeker; pornographi; racial; bicamer
9	rebel; militia; loot; gendarmeri; franc; summarili; government-control; fight; unlik; alli; southern; customari; rob; fgm; french; slaveri; progovern; soldier; antigovern; unconfirm
10	adolesc; metropolitan; secretariat; vigilant; sugar; understaf; inter-american; epz; undocu; anticorrupt; mob; restitut; misdemeanor; suscept; penitentiari; transgend; descent; reinstat; finish; firearm
11	ment; tion; countrysid; king; inviol; assassin; proscib; martial; guarante; now; leftist; though; probabl; habea; coup; corpus; adjust; -parti; indian; unusu
12	ment; tion; non-government; -parti; iraq; guarante; inter; martial; diet; now; inviol; invas; probabl; adjust; viewpoint; countrysid; pro; socialist; coup; emigr
13	iraq; gendarmeri; nonpolit; coup; fgm; slaveri; exit; unconfirm; sunni; embassi; defam; franc; expediti; oppon; leagu; opposit; regim; subsist; countrysid; inviol
14	milit; tribal; sharia; cast; bond; epz; river; tribe; forest; sharp; trust; sedit; insurg; koran; customari; bail; vigilant; leagu; widow; communal
15	roma; noncitizen; misdemeanor; potabl; politic; diet; reproduct; bisexu; contracept; restitut; feloni; stateless; racist; beg; hagu; anticorrupt; adolesc; chat; gender; -seat

15-Topic STMs with Alternate Controls

Figure A.31: Topwords for 15-topic STM using PTS (Amnesty) in place of latent human rights protections

Topic	Top 20 Words	Labels
1	indigen, guerrilla, congress, paramilitari, prosecutor, kidnap, murder, militari, threat, investig, kill, san, jose, depart, disappear, offic, armi, municip, arm, drug	Killings and Disappearances
2	respect, minimum, wage, union, bargain, trade, constitut, legisl, parliament, equal, freeli, free, languag, safeti, collect, race, nongovernment, religion, freedom, economi	Labor Rights
3	israel, arab, west, bank, territori, terrorist, occupi, jewish, east, emerg, jew, secur, strip, detainee, terror, closur, militari, southeast, interrog, administr	Terrorism and Torture
4	regim, bomb, insurg, disappear, armi, extrajudici, war, iraq, attack, northern, arm, throughout, late, assassin, milit, earli, special, execut, villag, thousand	State and Non-State Violence
5	opposit, african, presid, south, presidenti, coup, tradit, africa, elect, elector, parti, newspaper, minist, black, magistr, customari, farm, ralli, fgm, radio	Elections and Parties
6	traffick, girl, ngos, prostitut, child, ngo, victim, corrupt, school, approxim, district, exploit, develop, year, train, children, problem, sexual, end, boy	Human Trafficking
7	ministri, prison, lack, children, inmat, poor, street, labor, guard, resourc, code, sector, corrupt, disabl, pretrial, condit, age, overcrowd, domest, child	Prison Conditions
8	garante, econom, amnesti, communist, emigr, polici, recent, popul, tion, ment, growth, polit, now, rate, world, rural, current, will, product, urban	Economic Systems
9	feder, provinc, chines, provinci, china, regul, immigr, migrant, enterpris, northern, basic, north, cathol, peopl, oversea, studi, legisl, compani, racial, minor	Freedom of Movement
10	activist, journalist, author, demonstr, protest, deni, sentenc, articl, detain, regist, lawyer, registr, refus, independ, candid, block, offic, print, confisc, permiss	Freedom of Expression
11	act, polic, district, industri, bail, custodi, amend, parliament, chief, pend, recommend, death, commiss, murder, state, magistr, milit, bond, bill, board	Justice System
12	islam, muslim, king, sharia, foreign, non-muslim, male, christian, citizenship, citizen, passport, divorc, marri, servant, religi, sunni, noncitizen, mosqu, bahai, permiss	Religion and Citizenship
13	sexual, asylum, discrimin, internet, societ, gender, seeker, access, exploit, protect, corrupt, complaint, violenc, victim, penalti, stateless, anti-semit, prohibit, ombudsman, traffick	New Rights Issues
14	roma, prosecutor, ethnic, russian, parliament, church, minor, orthodox, soviet, properti, parlamentari, region, european, ombudsman, wit, languag, pretrial, russia, interior, europ	Ethnic and Minority Rights
15	soldier, rebel, displac, idp, civilian, north, see, conflict, arm, abduct, militia, humanitarian, forc, fight, kill, rape, provinc, attack, numer, war	Humanitarian Crises

Figure A.32: Alternate estimates for changing topic prevalence over time, when using PTS (Amnesty) in place of latent human rights protections

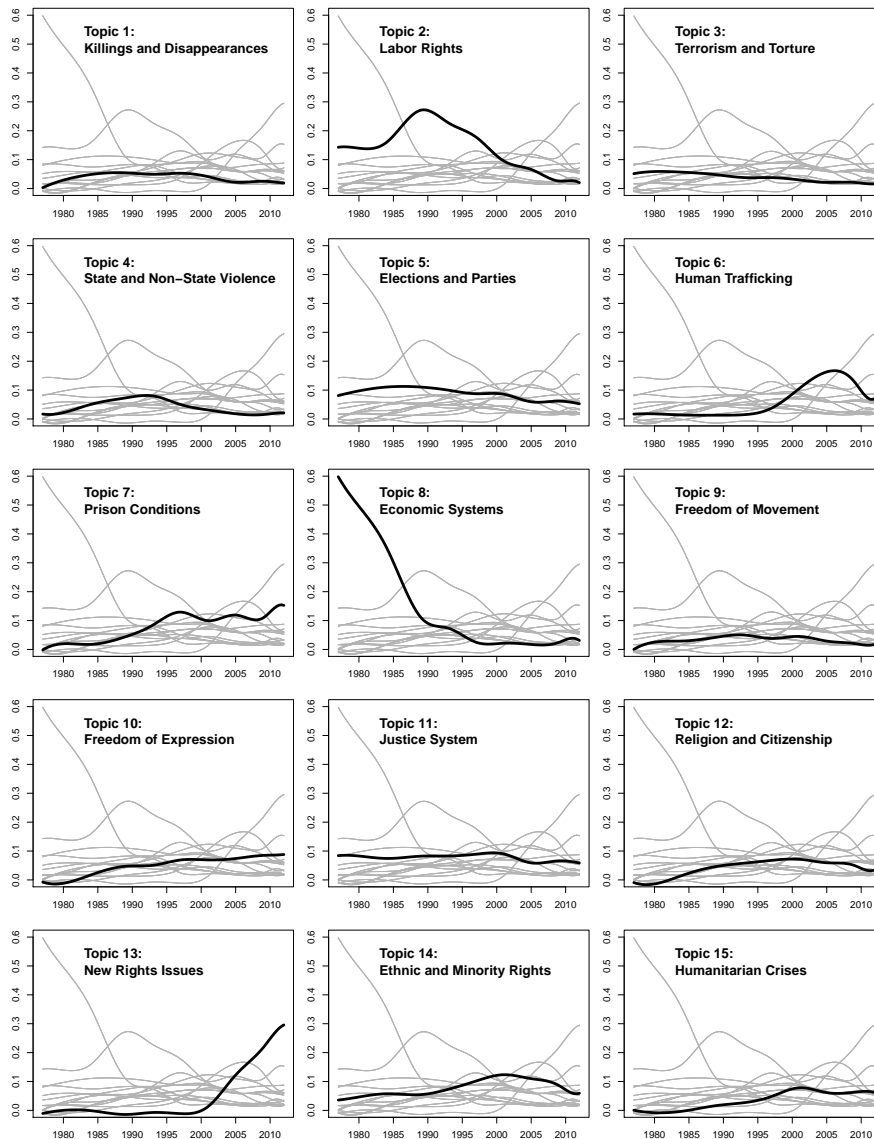


Figure A.33: Alternate estimates for primary covariates of interest, when using PTS (Amnesty) in place of latent human rights protections

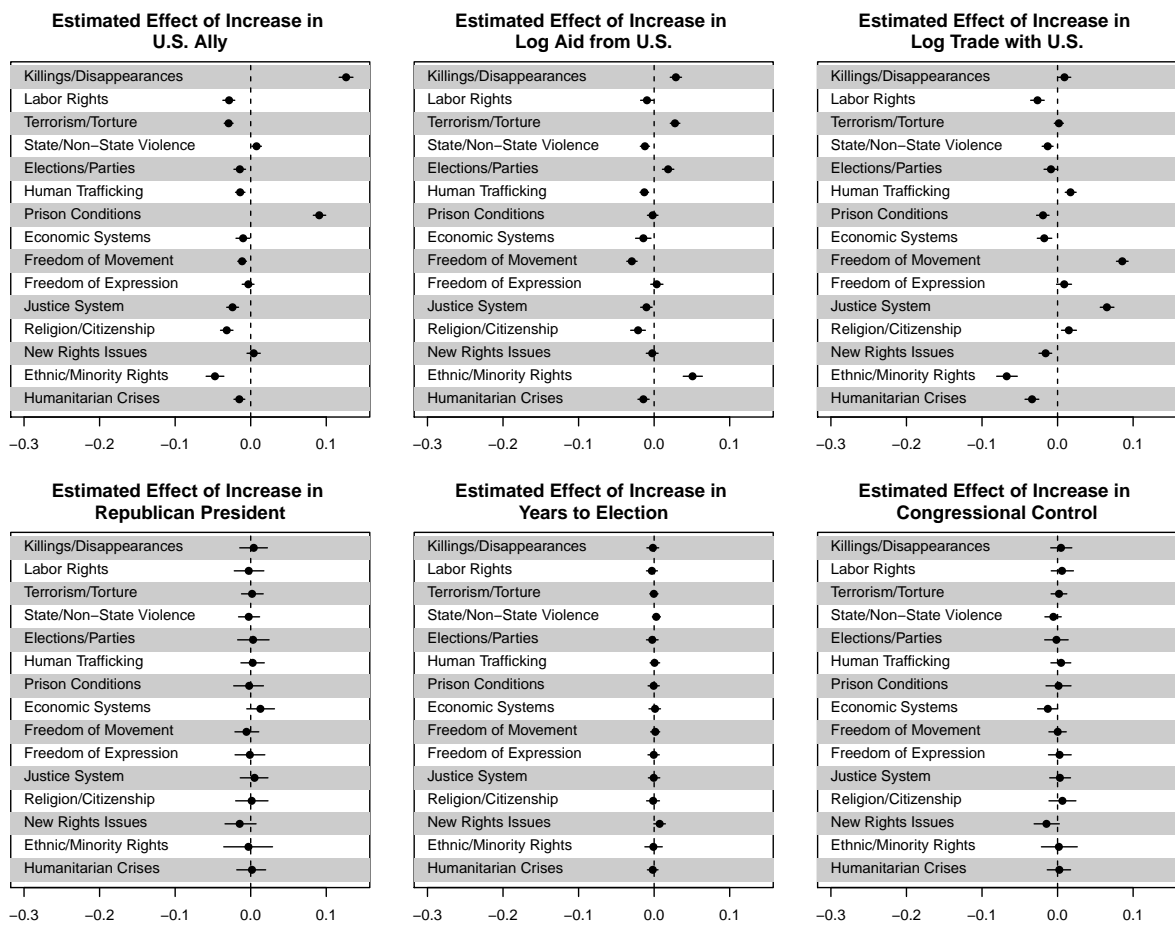


Figure A.34: Topwords for 15-topic STM when adding document word-count control

Topic	Top 20 Words	Labels
1	indigen, guerrilla, congress, paramilitari, prosecutor, kidnap, murder, militari, threat, investig, kill, san, jose, depart, disappear, offic, armi, drug, municip, arm	Killings and Disappearances
2	respect, minimum, wage, bargain, constitut, union, trade, parliament, equal, legisl, freeli, free, languag, safeti, collect, race, nongovernment, religion, freedom, exil	Labor Rights
3	israel, arab, west, bank, territori, terrorist, occupi, jewish, east, emerg, jew, secur, strip, terror, detainee, closur, militari, resid, southeast, permit	Terrorism and Torture
4	regim, bomb, insurg, disappear, armi, extrajudici, war, attack, northern, iraq, milit, arm, throughout, late, assassin, earli, special, execut, villag, thousand	State and Non-State Violence
5	opposit, african, presid, south, presidenti, coup, tradit, africa, elect, elector, parti, newspaper, minist, black, customari, magistr, farm, ralli, radio, fgm	Elections and Parties
6	traffick, girl, ngos, prostitut, child, ngo, victim, corrupt, school, approxim, exploit, district, year, develop, children, train, problem, end, sexual, boy	Human Trafficking
7	ministri, prison, lack, children, inmat, poor, street, guard, resourc, labor, code, sector, disabl, corrupt, pretrial, overcrowd, condit, age, due, child	Prison Conditions
8	guarante, econom, amnesti, communist, emigr, polici, recent, popul, tion, growth, ment, polit, now, rural, rate, world, current, will, product, urban	Economic Systems
9	feder, provinc, chines, provinci, china, regul, immigr, migrant, enterpris, northern, basic, north, cathol, peopl, oversea, legisl, studi, compani, racial, minor	Freedom of Movement
10	activist, journalist, author, protest, demonstr, deni, sentenc, articl, detain, lawyer, registr, regist, refus, independ, candid, block, offic, print, confisc, permiss	Freedom of Expression
11	act, polic, district, industri, amend, bail, custodi, parliament, chief, pend, recommend, death, commiss, state, murder, magistr, bond, bill, board, milit	Justice System
12	islam, muslim, king, sharia, foreign, non-muslim, male, christian, citizenship, passport, citizen, divorc, marri, servant, religi, bahai, noncitizen, mosqu, sunni, royal	Religion and Citizenship
13	sexual, asylum, discrimin, internet, societ, gender, access, seeker, exploit, corrupt, protect, complaint, victim, violenc, penalti, stateless, anti-semit, ombudsman, prohibit, individu	New Rights Issues
14	roma, prosecutor, ethnic, russian, parliament, church, minor, orthodox, soviet, region, properti, parlamentari, european, ombudsman, wit, languag, pretrial, russia, interior, televis	Ethnic and Minority Rights
15	soldier, rebel, displac, idp, civilian, north, conflict, see, arm, abduct, militia, humanitarian, forc, fight, kill, provinc, rape, attack, numer, war	Humanitarian Crises

Figure A.35: Alternate estimates for changing topic prevalence over time, when adding document word-count control

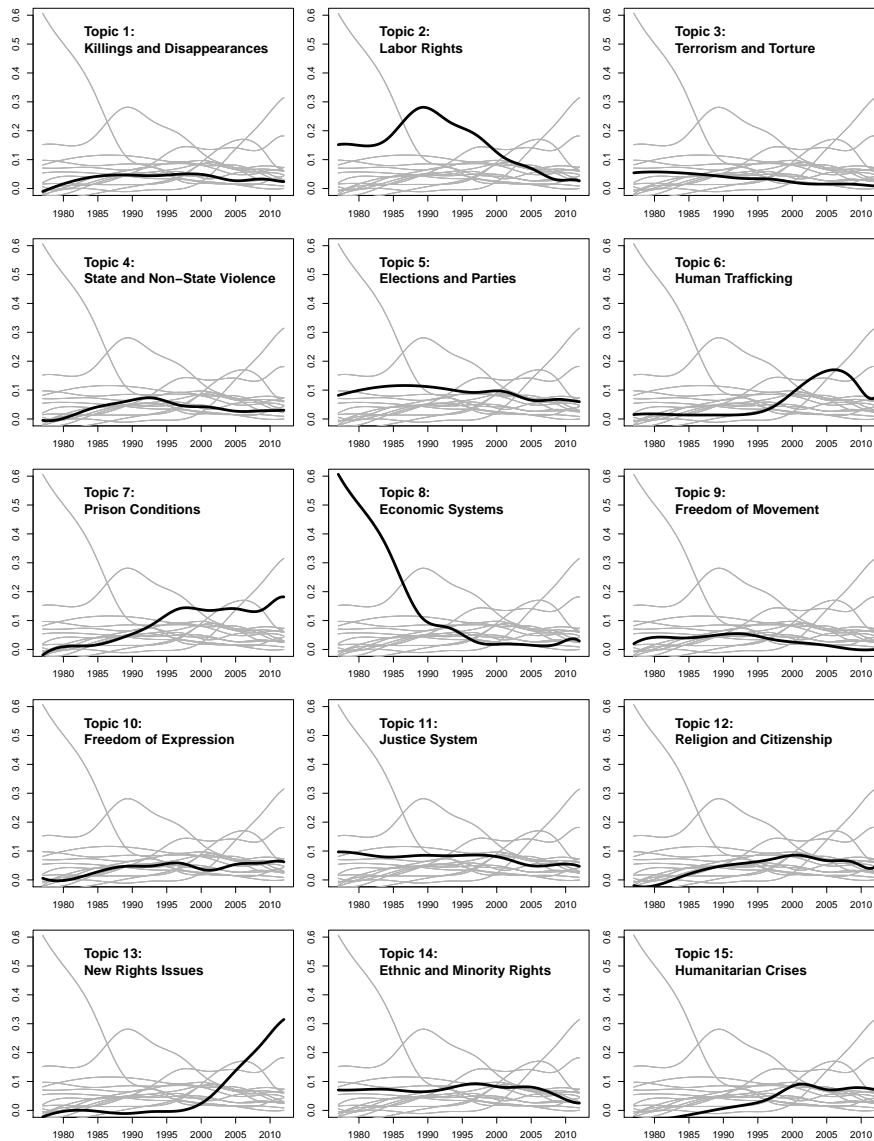


Figure A.36: Alternate estimates for primary covariates of interest, when adding document word-count control

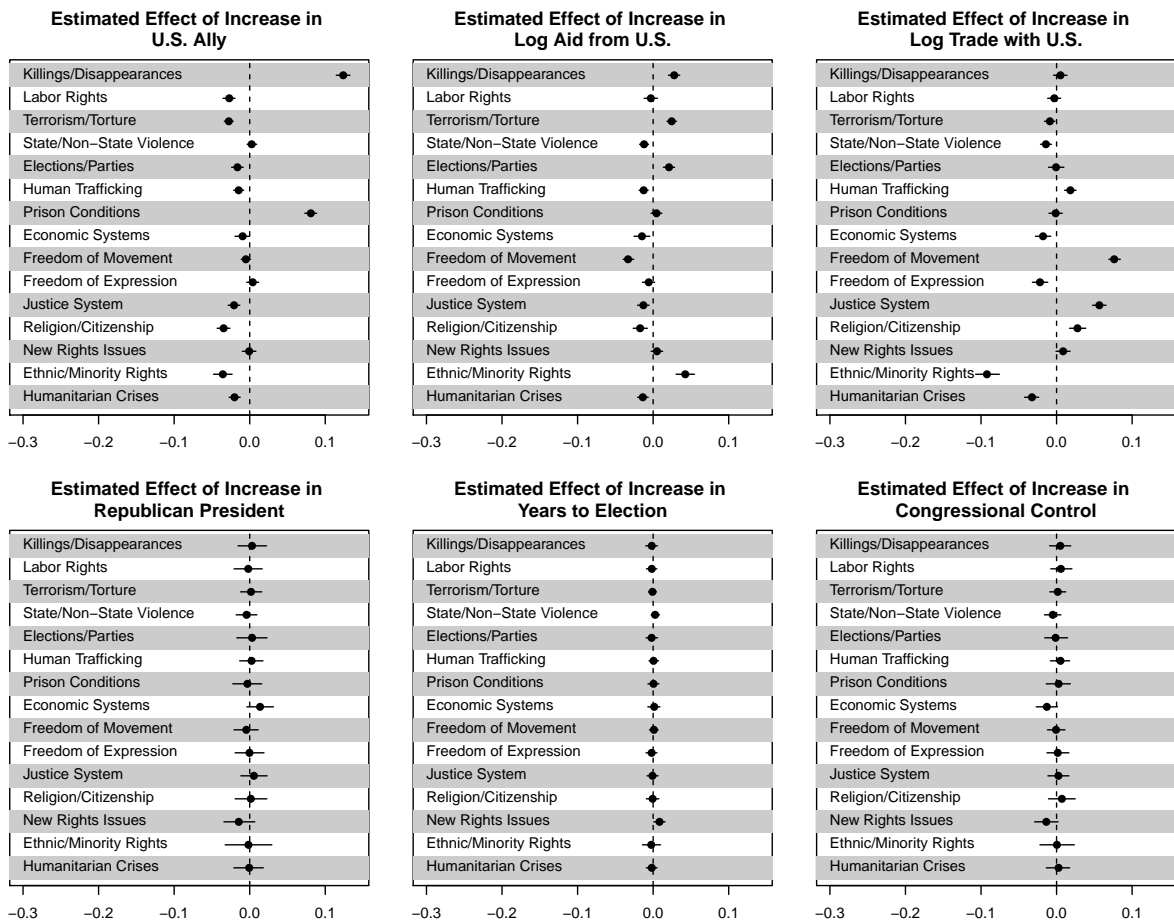
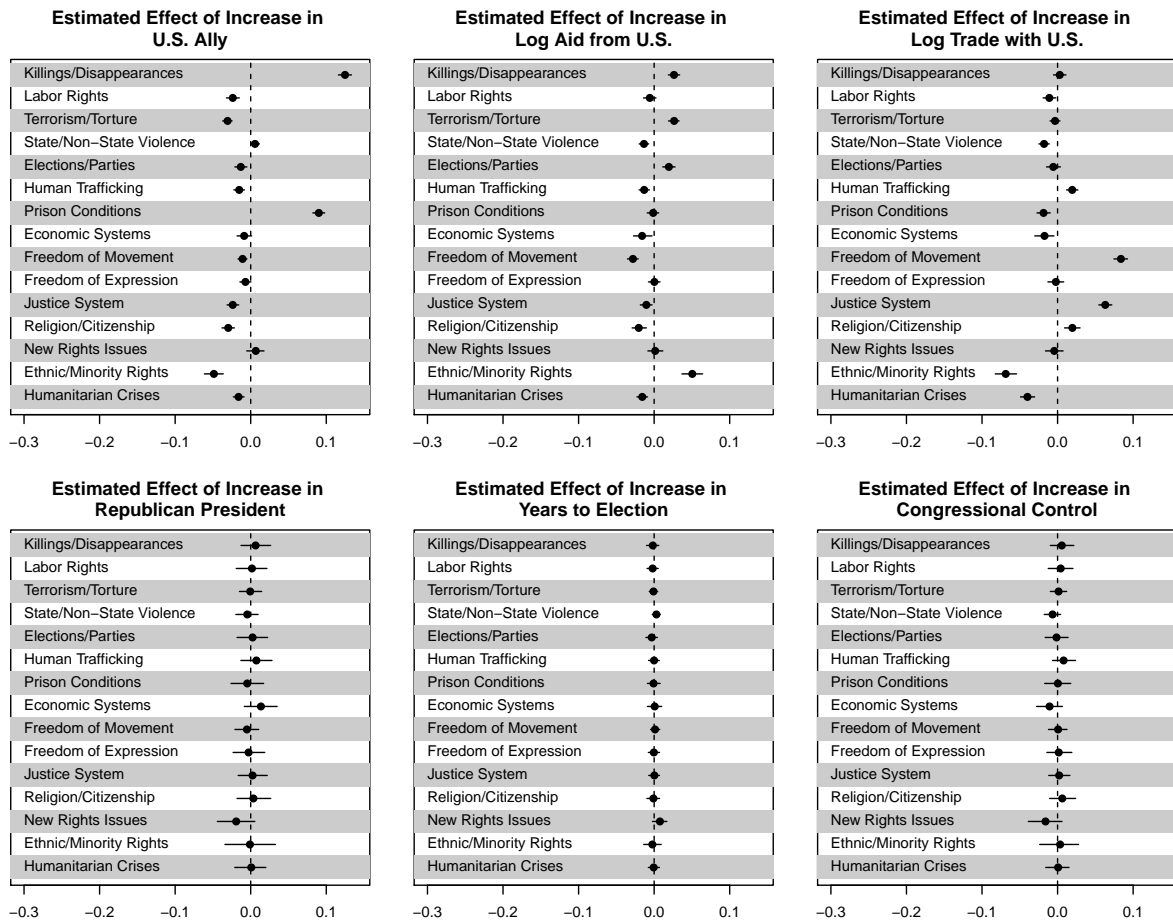


Figure A.37: Topwords for 15-topic STM when excluding year time trend

Topic	Top 20 Words	Labels
1	indigen, guerrilla, congress, paramilitari, prosecutor, kidnap, murder, militari, threat, investig, kill, san, jose, depart, offic, armi, disappear, drug, municip, arm	Killings and Disappearances
2	respect, minimum, wage, union, bargain, constitut, trade, equal, parliament, legisl, freeli, free, languag, safeti, collect, race, nongovernment, religion, freedom, hour	Labor Rights
3	israel, arab, west, bank, territori, occupi, jewish, terrorist, east, emerg, jew, secur, strip, detainee, closur, terror, militari, southeast, permit, resid	Terrorism and Torture
4	regim, bomb, insurg, armi, disappear, extrajudici, northern, war, attack, iraq, terrorist, arm, milit, assassin, throughout, late, earli, emerg, special, execut	State and Non-State Violence
5	opposit, african, presid, south, presidenti, coup, tradit, africa, elect, elector, parti, newspaper, minist, black, magistr, customari, farm, ralli, fgm, radio	Elections and Parties
6	traffick, girl, ngos, prostitut, child, ngo, victim, corrupt, school, district, approxim, develop, exploit, year, train, children, problem, end, sexual, boy	Human Trafficking
7	ministri, prison, lack, inmat, children, poor, street, labor, guard, resourc, code, sector, corrupt, pretrial, disabl, condit, overcrowd, age, domest, due	Prison Conditions
8	guarante, econom, amnesti, communist, emigr, polici, recent, popul, tion, growth, now, ment, polit, rate, world, rural, current, will, product, life	Economic Systems
9	feder, provinc, chines, provinci, china, regul, immigr, migrant, enterpris, basic, northern, north, cathol, oversea, peopl, studi, legisl, compani, racial, minor	Freedom of Movement
10	activist, journalist, author, demonstr, protest, deni, articl, sentenc, detain, lawyer, regist, registr, independ, refus, block, candid, print, offici, confisc, prison	Freedom of Expression
11	act, polic, district, industri, amend, bail, custodi, parliament, chief, pend, recommend, death, commiss, state, murder, bond, magistr, bill, board, activist	Justice System
12	islam, muslim, king, sharia, foreign, non-muslim, male, christian, citizenship, citizen, passport, divorc, marri, servant, religi, mosqu, bahai, sunni, husband, noncitizen	Religion and Citizenship
13	sexual, asylum, discrimin, societ, internet, gender, access, exploit, seeker, protect, corrupt, complaint, victim, violenc, penalti, stateless, anti-semit, traffick, ombudsman, prohibit	New Rights Issues
14	roma, prosecutor, ethnic, russian, parliament, church, minor, orthodox, soviet, properti, region, parlamentari, european, wit, languag, ombudsman, pretrial, interior, russia, media	Ethnic and Minority Rights
15	soldier, rebel, displac, idp, civilian, north, conflict, see, arm, abduct, humanitarian, militia, forc, fight, kill, rape, provinc, attack, numer, war	Humanitarian Crises

Figure A.38: Alternate estimates for primary covariates of interest, when excluding year time trend



Further Assessment of Covariate Effects

This section examines the robustness of our covariate effects with respect to both the non-linear year variable and the six primary covariates evaluated within the main paper (i.e., U.S. ally, log aid from the U.S., log trade with the U.S., Republican president, years to election, and Congressional control). We do so first by extracting the associated covariate effects for the 10 and 20-topic STM models presented in Section 2 above. Recall that these alternate STMs largely recover comparable topics to those of our primary 15-topic STM. Using the topic labels assigned in Section 2 above, we obtain comparable plots for the effect of year, and for our six covariates of interest, within the 10-topic STM, and plot these effects in Figures A.39-A.40. Beginning first with Figure A.39, we find that our 10-topic STM's identified topics vary over

time in similar manners to the 15-topic STM's topics. Most notably within Figure A.39, one can observe the same (i) dramatic increase in New Rights Issues beginning in the year 2000, (ii) continuous decline in Economic Systems throughout the time series, and (iii) rise and fall in the prevalence of Human Trafficking during the final third of our time series—although the latter effect is less pronounced in the 10-topic case. Also akin to the 15-topic STM results, we find in Figure A.40 that our U.S.-domestic controls continue to exhibit little-to-no effect on any of our identified topics, whereas U.S. ally and log aid from the U.S. again exert the largest effects on the Killings and Disappearances and Ethnic and Minority Rights, respectively. For Log Trade with the U.S., we find that a number of previously insignificant effects are now significant, although the direction of each estimated effect remains comparable to that of the 15-topic STM, save for Terrorism and Torture, which was previously slightly negative and non-significant and now is positive and significant.

Turning to the updated topic prevalence estimates for our 20-topic STM, which appear in Figures A.41-A.42, we find that our alternate (20-topic) STM has identified similar temporal variation to that of the main 15-topic STM and the 10-topic STM. Regarding variation in our 20 identified topics over time, for example, one can again observe comparable rises and falls in the Labor Rights and Human Trafficking topics, as well as a sharp rise in New Rights Issues beginning in the year 2000, and a decline in Economic Systems over the entire course of the series. In addition, the covariate plots in Figure A.42 once again suggest that domestic U.S.-based factors have null effects on our identified topics, save perhaps for Economic Systems. U.S. alliances, aid, and trade, each again exhibit comparable effects to the 10-topic and 15-topic estimates, most notably for the Ethnic and Minority Rights, Killings and Disappearances topics. In all, we find little-to-no discrepancies between the 15-topic and 20-topic covariate estimates. Thus, in sum, the effects of our primary covariates of interest appear to be highly stable across one's choice of topics—both in terms of the direction and significance of a covariate's effect.

Finally, this section next re-evaluates the covariate effects for our 15-topic STM when using a single set of starting parameter values. This is in contrast to the primary 15-topic STM approach used in our main paper, which followed Roberts et al. (2014) to initialize 50 15-topic STMs, each with different starting parameter values, and ultimately chose one of these ini-

tializations for our primary 15-topic model based upon Pareto optimal semantic cohesion and exclusivity scores. While the use of a single set of starting parameters can potentially undermine the interpretability of our identified topics, findings in this regard will help us to evaluate the extent to which our conclusions (with respect to both topwords and covariate effects) disproportionately rest on the identification of a primary 15-topic STM through our optimization of semantic cohesion and exclusivity. We hence estimate a single initialization 15-topic STM, and report the topwords,⁵ year-based topic-estimates, and primary covariate estimates from this ‘single initialization 15-topic STM’ in Figures A.43-A.45 below.

Based upon Figure A.43, the single initialization of our 15-topic STM exhibits a number of similarities with our primary (multiple-initialization) 15-topic STM, as well as a few noticeable differences. Regarding similarities, topic 1 shares the vast majority of its topwords with the Economic Systems topic from our primary STM, and the same can be said for topic 2 and our previously discussed Killings and Disappearances topic. For the single initialization model, we can also easily label topics 3-4 as Freedom of Expression and Freedom of Movement, given these topics’ close congruence with these two primary 15-topic STM topics. Likewise, topics 5 and 6 are very clearly capturing our aforementioned Elections and Parties and Human Trafficking topics. Topic 7 is less coherent, but the fact that it shares words such as “corrupt,” “access,” “internet,” and “individu” with our New Rights Issues topic leads us to label it as such. On the other hand, while Topic 8 shares several topwords with our Terrorism and Torture topic, it appears to be most directly focused on general security concerns within the Middle East—leading us to conservatively label this topic as “Middle East” here. Topic 9 appears to overlap fairly well with our previously identified Justice System topic, albeit now with moderately more focus on police abuses, whereas the topwords in topic 10 are nearly identical to those found within our Religion and Citizenship topic. One can similarly observe a great deal of coherent overlap between the topwords in topic 12 and the Ethnic and Minority Rights topics. Topic 11, on the other hand, appears to encompass elements previously identified within the Killings and Disappearances topic, perhaps implying that the single initialization 15-topic STM in this case has split a previously coherent topic into two overlapping topics—topics 2 and

⁵As based upon frequency exclusivity scoring metrics.

11—although we now label the latter as Criminal Violence in order to distinguish it from the former. By contrast, the topwords in Topic 13 match very closely with those of the topic labeled Humanitarian Crises in our main paper, as, for example, nine of its top 10 wordstems also appear in Humanitarian Crises’ top 20 wordstems. Finally, topic 15 can be easily characterized as Labor Rights, while topic 14 not only overlaps with our Prison Conditions topic, but also now encompasses a number of topwords that appear to be more directly related to labor conditions, leading us to now label this topic as Prison and Labor Conditions.

Given that virtually all of the topword vectors in Table A.43 can be closely matched with the topics identified within our primary 15-topic STM, we label them accordingly and now turn to evaluating the effects of our primary covariates on these alternate topic estimates. Beginning first with the estimated temporal changes in topic prevalence presented in Figure A.44, one can clearly observe the previously highlighted temporal changes in Economic Systems (topic 1), Human Trafficking (topic 6), New Rights Issues (topic 7), and Labor Rights (topic 15). At the same time, we do see a more noticeable rise-and-fall in topic 14 (Prison and Labor Conditions) than was the case for the Prison Conditions topic in the main paper, which provides us with a degree of secondary evidence to suggest that this topic is indeed now not only capturing prison conditions, but also labor conditions that may have been previously grouped with our Labor Rights topic. The covariate effects in Figure A.45 reinforce the above conclusions, both in reaffirming the null effects of our domestic-U.S. covariates, and in implying consistent effects for U.S. ally, log aid from the U.S., and log trade with the U.S. While the vast majority of the latter effects correspond quite closely to those of our main 15-topic STM estimates, and hence do not warrant extensive discussion here, one additional point of note is that for U.S. ally, the two largest covariate effects now correspond to *both* the Killings and Disappearances *and* Criminal Violence topics—reaffirming our earlier contentions that this single initialization had largely divided our previous Killings and Disappearances topic into two overlapping topics. As such, the coherence of the single initialization 15-topic STM appears to be slightly suboptimal to our primary, multiple-initialization 15-topic STM, which is in line with our choice of a final 15-topic STM that maximized semantic cohesion and exclusivity. Nevertheless, the primary conclusions discussed with respect to temporal variation and covariate effects in the main paper

appear to be robust to an STM approach that does not select a model based upon this multiple-initialization strategy.

Figure A.39: Alternate estimates for changing topic prevalence over time, 10-topic STM

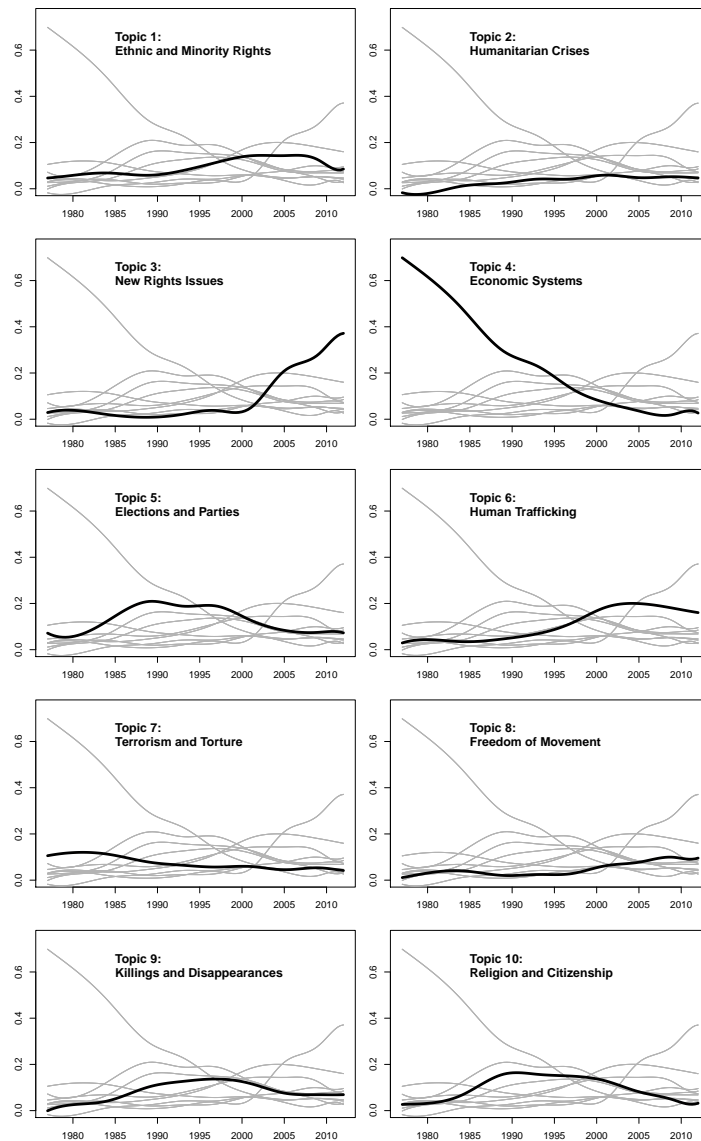


Figure A.40: Alternate estimates for primary covariates of interest, 10-topic STM

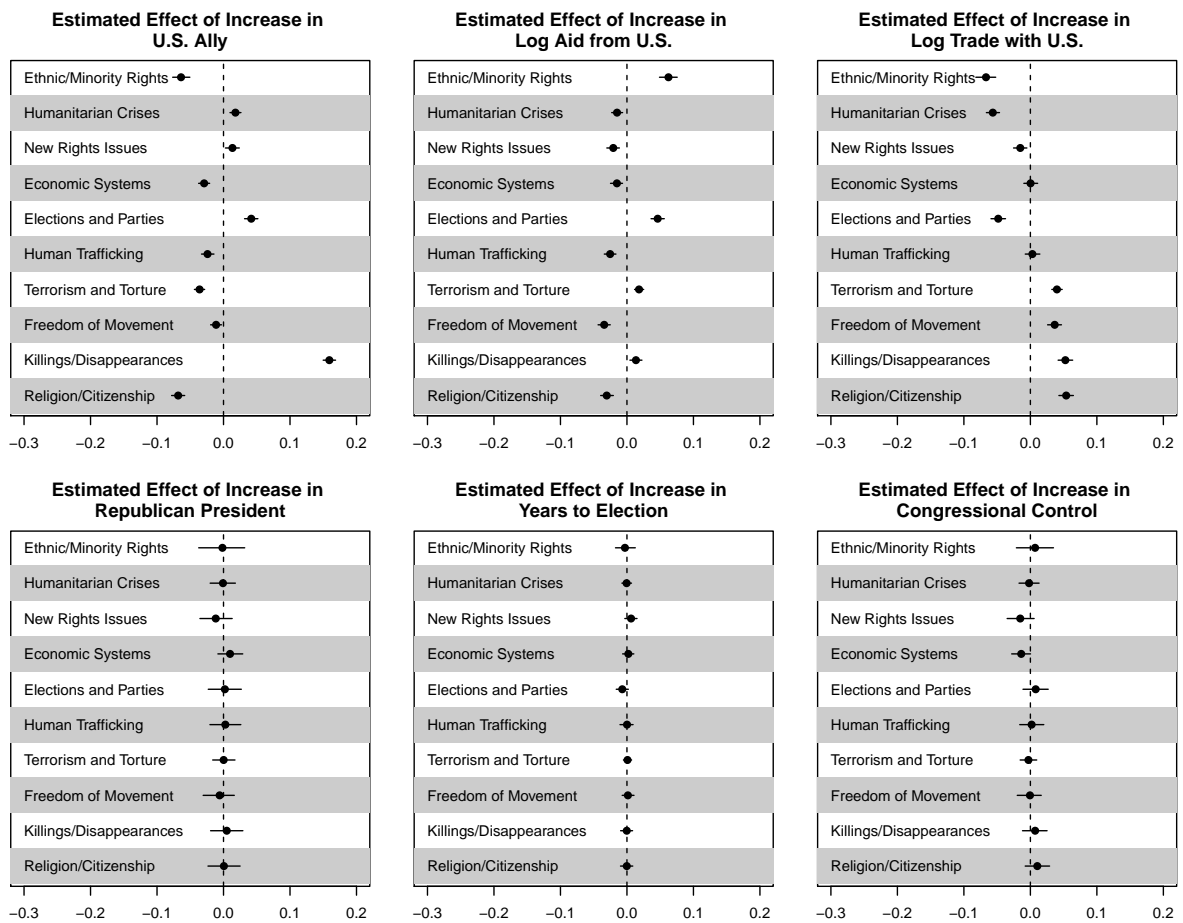


Figure A.41: Alternate estimates for changing topic prevalence over time, 20-topic STM

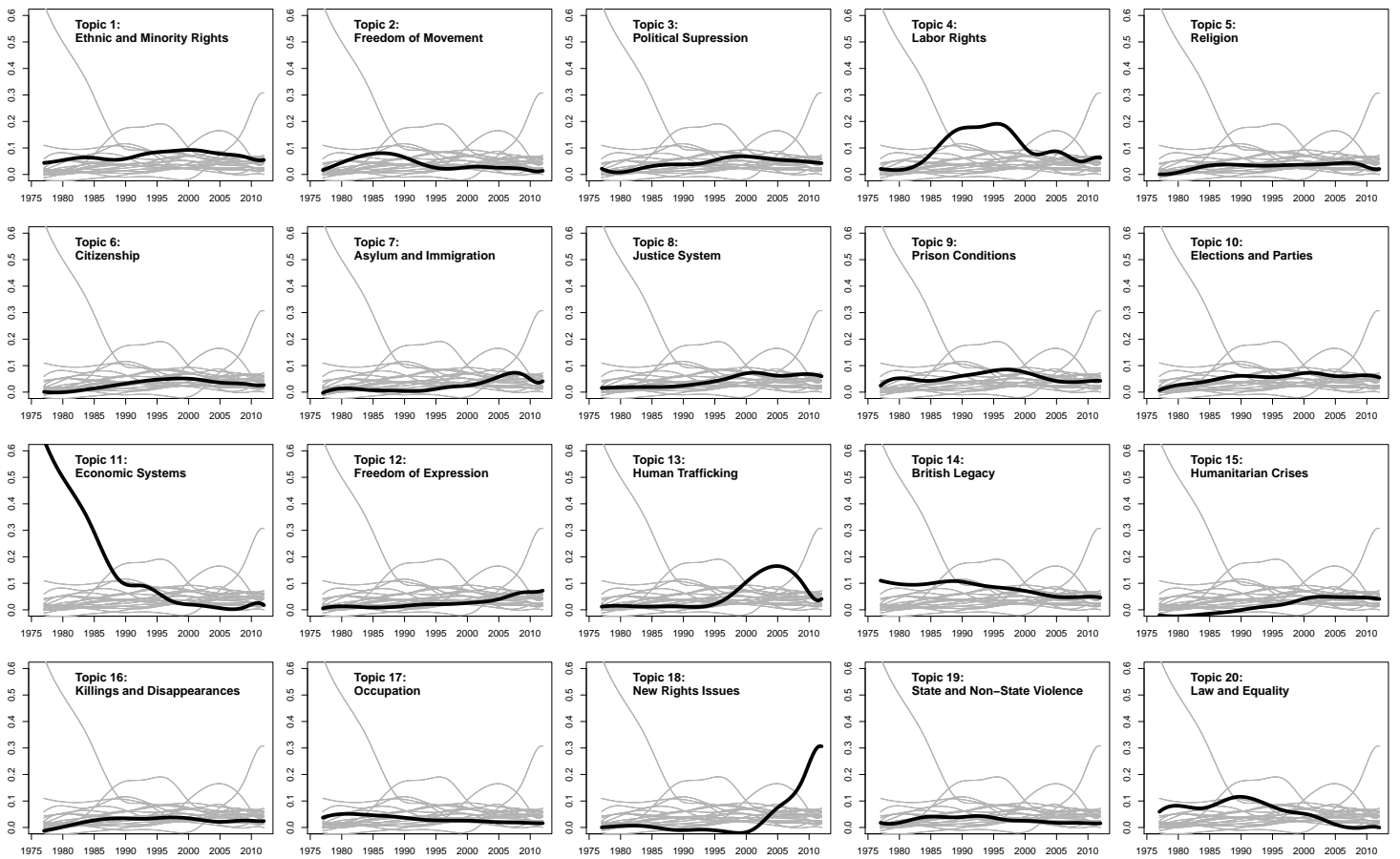


Figure A.42: Alternate estimates for primary covariates of interest, 20-topic STM

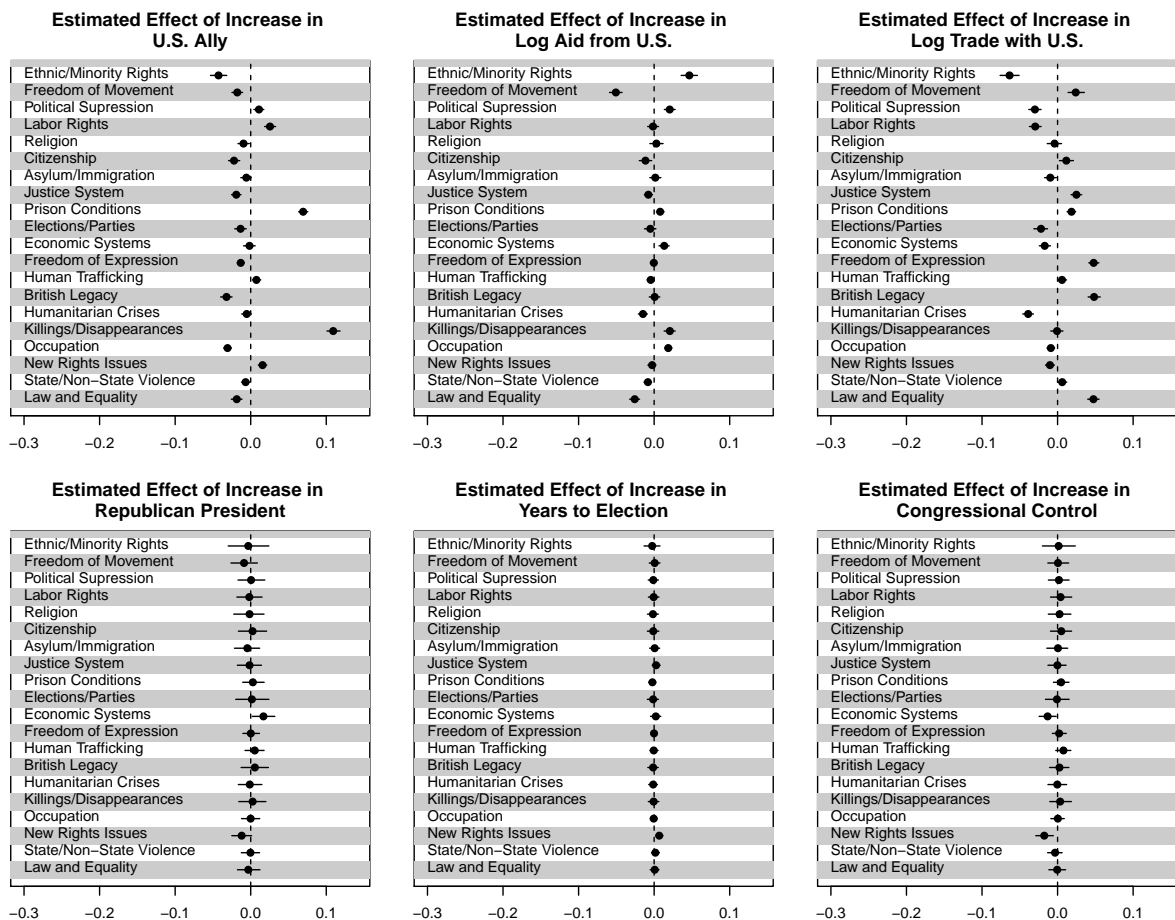


Figure A.43: Topwords for 15-Topic STM, based upon a single initialization

Topic	Top 20 Words	Labels
1	guarante, amnesti, african, econom, black, now, emigr, south, regim, recent, coup, current, growth, will, polit, economi, africa, tion, power, rate	Economic Systems
2	feder, murder, kill, armi, paramilitari, militari, terrorist, guerrilla, kidnap, arm, disappear, bomb, terror, attack, threat, civilian, death, command, extrajudici, governor	Killings and Disappearances
3	fgm, girl, opposit, beat, farm, tradit, radio, region, presid, dispers, unlik, station, customari, magistr, see, ethnic, broadcast, african, presidenti, teacher	Freedom of Expression
4	provinc, chines, china, provinci, regul, dissid, enterpris, communist, religi, sensit, cathol, church, polici, peopl, foreign, offici, basic, central, oversea, approv	Freedom of Movement
5	opposit, activist, newspaper, candid, deni, independ, decre, demonstr, journalist, articl, confisc, sentenc, passport, refus, lawyer, presidenti, meet, print, regim, harass	Elections and Parties
6	traffick, victim, sexual, asylum, prostitut, immigr, discrimin, exploit, societ, internet, prohibit, countri, seeker, protect, provid, disabl, center, person, child, assist	Human Trafficking
7	corrupt, ngos, birth, ngo, access, internet, media, societ, share, fine, lack, sexual, gender, stateless, transpar, individu, note, hivaid, web, inform	New Rights Issues
8	arab, israel, west, bank, territori, iraq, occupi, jewish, resid, bahai, islam, jew, permit, noncitizen, sunni, strip, east, secur, deport, western	Middle East
9	district, polic, prostitut, kill, child, shot, poor, often, children, rape, girl, suspect, factori, jail, provinci, land, -year-old, problem, ngos, injur	Justice System
10	islam, muslim, king, ban, sharia, christian, act, non-muslim, prime, tribal, male, servant, minist, amend, husband, parliament, milit, marri, woman, royal	Religion and Citizenship
11	indigen, prosecutor, congress, san, ombudsman, jose, complaint, land, inmat, drug, judg, municip, compani, senat, carlo, secretariat, offic, crime, director, adolesc	Criminal Violence
12	roma, prosecutor, ethnic, parliament, russian, church, orthodox, minor, languag, parlamentari, soviet, region, properti, televis, regist, local, european, interior, ombudsman, russia	Ethnic and Minority Rights
13	soldier, rebel, civilian, displac, north, war, arm, conflict, idp, attack, fight, humanitarian, militia, camp, abduct, southern, insurg, kill, forc, icrc	Humanitarian Crises
14	ministri, code, sector, assembl, labor, privat, minimum, citizen, interior, associ, practic, constitut, wage, prohibit, nation, famili, judici, formal, compulsori, salari	Prison and Labor Conditions
15	bargain, wage, trade, industri, legisl, act, minimum, union, collect, safeti, parliament, feder, respect, race, provis, employ, sex, equal, strike, freeli	Labor Rights

Figure A.44: Alternate estimates for changing topic prevalence over time, based upon a single initialization

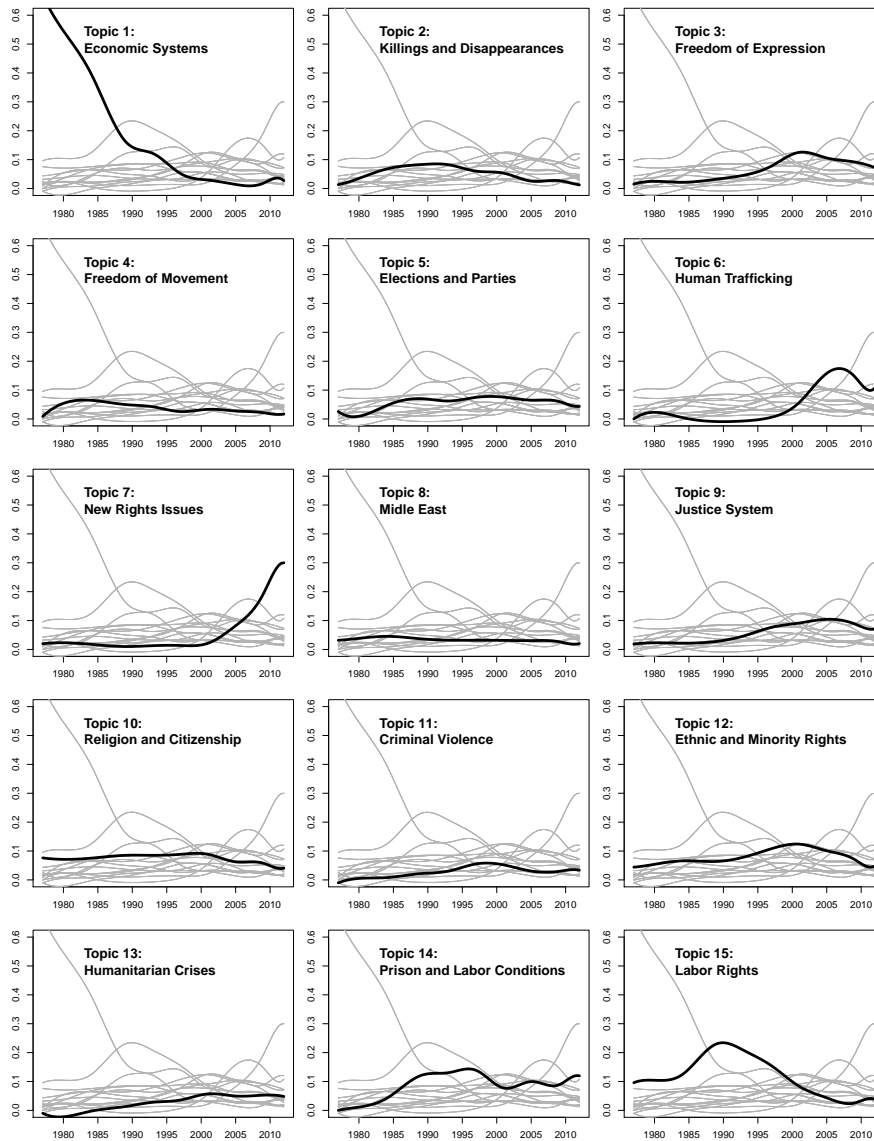
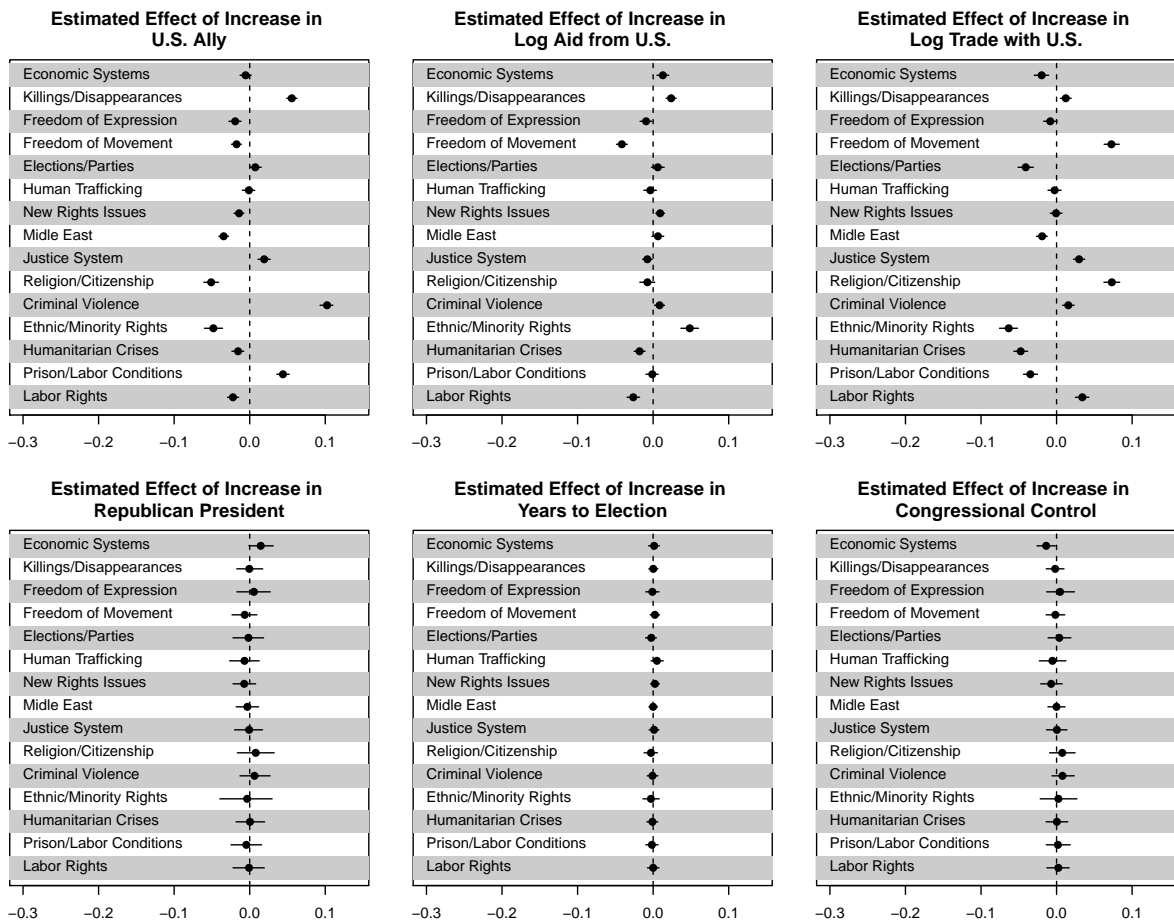
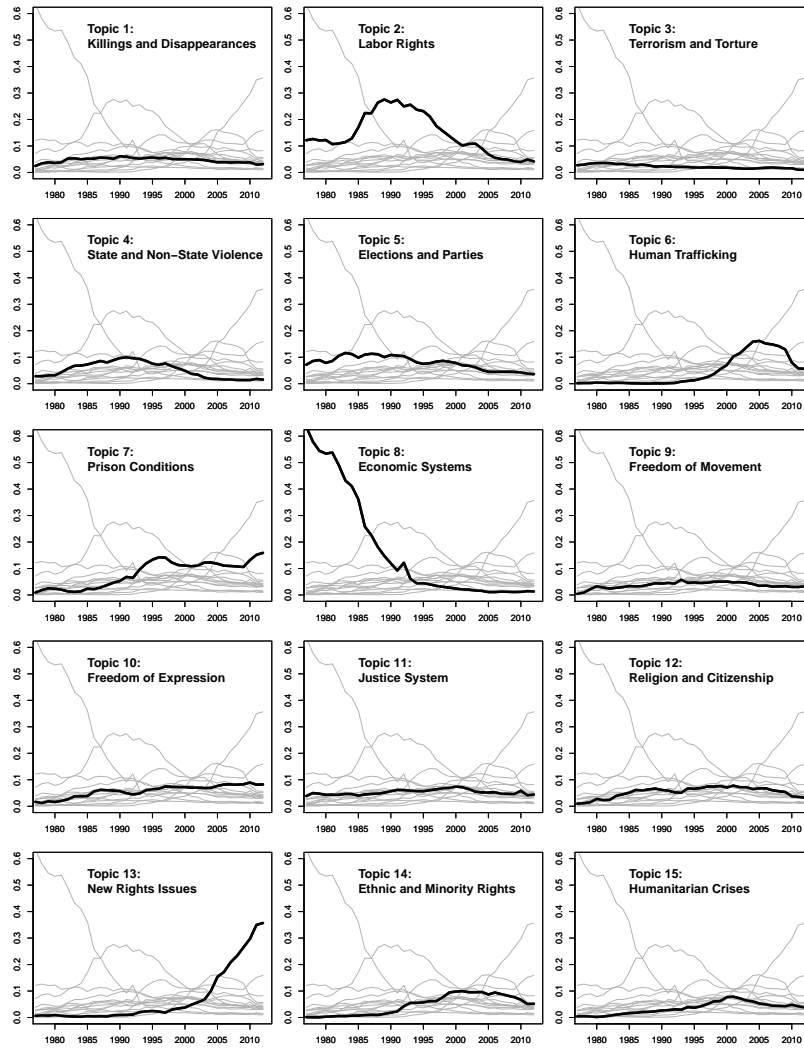


Figure A.45: Alternate estimates for primary covariates of interest, based upon a single initialization



Topical Attention Plots

Figure A.46: Topical attention over time



References

- Bank, World. 2013. "World Development Indicators 2013." The World Bank.
- Barbieri, Katherine and Omar Keshk. 2012. "Correlates of War Project Trade Data Set Codebook, Version 3.0." Online: <http://correlatesofwar.org>.
- Census.gov. 2015. "U.S. Trade in Goods by Country." Retrieved on 1/10/2015, from: <http://www.census.gov/foreign-trade/balance/>.
- Fariss, Christopher J. 2014. "Respect for Human Rights has Improved Over Time: Modeling the Changing Standard of Accountability." American Political Science Review 108(4):297–318.
- Gibler, Douglas M. 2009. International Military Alliances, 1648-2008. CQ Press.
- Gibney, M., L. Cornett, R. Wood and P. Haschke. 2014. "Political Terror Scale 1976-2012." Retrieved on 12/20/2014, from: <http://www.politicalterror scale.org/>.
- Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg and Håvard Strand. 2002. "Armed Conflict 1946-2001: A New Dataset." Journal of Peace Research 39(5):615–637.
- Marshall, Monty G., Keith Jagers and Ted Robert Gurr. 2010. Polity IV Project: Political Regime Characteristics and Transitions, 1800-2010. Center for Systemic Peace.
- Roberts, Margaret E., Brandon M. Steward, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Gadarian, Bethany Albertson and David G. Rand. 2014. "Structural Topic Models for Open Ended Responses." American Journal of Political Science 4(58):1064–1082–442.
- USAID. 2014. "Prepared Foreign Assistance Data Files." Retrieved on 3/1/2015, from: <https://eads.usaid.gov/gbk/data/prepared.cfm>.