Supplemental Information:
Ethnic Diversity in Central Government Cabinets

The Supplemental Information contains a description of ethnic name classification systems and the political relevance of names (SI.1), a comparison of the scope of the analysis (SI.2), details on creating the cabinet minister dataset (SI.3), a list of name communities used by NamePrism (SI.4), details on calculating the cabinet diversity measure (SI.5), simulation results converting cabinet diversity into seat change (SI.6), trends in the cabinet diversity measure (SI.7), reasons for and specifications of control variables and Country ELF (SI.8), the results of stationarity tests (SI.9), full model results with alternate model specifications (SI.10), alternative independent and dependent variables (SI.11), validation of results using hand coded data and specific cases (SI.12), and cabinet seat change simulations (SI.13).

Replication data and code for all empirical analysis is posted on the author's website.

SI.1: Name Classification Systems

Most older name classification methods use dictionaries of name-ethnicity pairs and simply look-up inputted names in these lists (Mateos, 2007). Such studies are necessarily limited to single countries or small samples due to the computational expense of precisely identifying the ethnicity of long lists of individuals. Starting with Mateos, Webber and Longley (2007), geographers and computer scientists have made various attempts to use parts of names to classify them into ethnicities. Mateos, Webber and Longley (2007) take the approach of identifying 185 cultural, linguistic, and ethnic groups in the United Kingdom and classifying names based on geographic proximity. Though their study is limited because of the time and resources required to produce an extensive listing of name-ethnicity pairs, it serves as inspiration for a number of computer science methods.

Ambekar et al. (2009) was the first serious attempt to classify names using fragments or groups of letters into thirteen ethnicities via a Hidden Markov Model and an original corpus of 20 million names from news articles. Chang et al. (2010) use a straightforward probabilistic model to classify Facebook users in the United States into ethnicities. Torvik and Agarwal (2016) match partial fragments of words to a corpus of 5 million scientists on PubMed. Finally, Treeratpituk and Giles (2012) alter the Ambekar et al. (2009) approach by using character sequences from a corpus of Wikipedia articles to train a multinomial logit classifier.

Pool, Stoffman and Yonker (2012) and others have successfully used these algorithmic measures, but the measures themselves are very limited. First, the maximum number of ethnicities in these classifiers is twenty-six, which is not representative of worldwide ethnic variation. Second, previous methods use training datasets with few observations. Further, no one classifier maximizes both the number of ethnicities used and the size of the training dataset. NamePrism combines the largest training dataset with the most name based ethnicity classifications.

Politically Relevant Names

For our purposes, we are interested in names capturing politically salient ethnic distinctions. One way we can show this is by demonstrating that country leaders consider the signal that a cabinet ministers' name will send to citizens when selecting such ministers. From literature on labor market discrimination (e.g., Bertrand and Mullainathan, 2004) we know that humans associate names with ethnic cues and discriminate against individuals whose names indicate that they belong to a certain ethnic group even if they do not identify with said group. This applies to political situations such as voters selecting judges with Hispanic sounding names.[1]

We also know that politicians strategically employ descriptive characteristics like their names in order to appeal to citizens with certain ethnicities. Bobby Jindal,[2] Ted Cruz, and Beto O'Rourke[3] all represent prominent US politicians whose name changes have raised questions about whether they are trying to affiliate with or distance themselves from certain ethnic groups in order to win votes.

What is more, political leaders think strategically about the signal that a politicians' name will send to citizens and try to optimize politician selection and how politicians use their names in order to send clear signals to citizens regarding politicians' ethnicity. To investigate this question, I conducted field interviews in India with civil servants, non-governmental organizations that worked with politicians, and academics studying how politicians manage ethnic identity.[4] Interview subjects brought up and agreed with the idea that political leaders considered the names of potential politicians when deciding whether they should run for office or be appointed to leadership positions.[5] Political leaders use two methods to make sure politicians' names send the correct ethnic signal to constituents. First, political leaders tend to seek out politicians with traditionally ethnic names and encouraged them to run in constituencies that match the ethnic identity their name signals. Second and more commonly, political leaders encourage or force politicians to change their names to clearly signal ethnic identity. For example, party leaders decided in one case to drop a candidate's surname because it incorrectly identified their ethnic identity where in another case party leaders told a candidate to add a surname that reflected their ethnic identity.[6]

[1] Emma Platoff and Alexa Ura, 2018, "In Texas Republican Judicial Primaries, do Hispanic-Sounding Surnames Spell Loss?" `https://www.texastribune.org/2018/03/01/texas-highest-courts-have-few-hispanic-judges-some-attribute-surname-c/`.

[2] Annie Gowen and Tyler Bridges, 2015, "From Piyush to Bobby: How Does Jindal Feel About his Family's Past?" `https://www.washingtonpost.com/politics/from-piyush-to-bobby-how-does-jindal-feel-about-his-familys-past/2015/06/22/7d45a3da-18ec-11e5-ab92-c75ae6ab94b5_story.html`.

[3] W. Gardner Selby, 2018, "Why Rafael Cruz Goes by Ted and Robert O'Rourke by Beto," `https://www.statesman.com/news/20180817/politifact-why-rafael-cruz-goes-by-ted-and-robert-orourke-by-beto`.

[4] University IRB approval 201910066.

[5] Interview, December 13, 2019; Interview, December 30, 2019; Interview, January 14, 2020.

[6] Interview, January 14, 2020.

These are cases where political leaders are clearly thinking both about the ethnic identity of a politician, but also their name and how citizens will react to reading or hearing the name as an indicator of ethnic identity. Few political leaders would openly admit to considering the names of politicians when selecting candidates or political leaders, so secondary source interviews are likely the closest we can get to describing this link.

SI.2: Countries in the Analysis

Table SI.2.1 presents a number of studies that use data on cabinet ministers from various sources. The three papers related to ethnicity or patronage politics (via cabinet seat allocations or dismissals) all focus on a variety of African countries over relatively long time periods. In contrast, three studies on womens' appointments to cabinets make different tradeoffs between time span and country coverage. Although they all use the CIA's Chiefs of State and Cabinet Members of Foreign Governments dataset, none of them utilize both a long time span and many countries.

Table SI.2.1: Countries Used in Cabinet Analyses

Study	Variable	Source	Year	Countries
Francois, Rainer and Trebbi, 2015	Ethnicity	Africa South of the Sahara and The Europa World Year Book	1960-2004	15 Africa
Arriola, 2009, Opalo, 2011	Cabinet Seats	Africa South of the Sahara	1961-2007	43 Africa
Kroeger, 2017	Cabinet Dismissals	Africa South of the Sahara	1976-2010	37 Africa
Escobar-Lemmon and Taylor-Robinson, 2005	Women	COS	1980-2003	18 S. Amer.
Krook and O'Brien, 2012	Women	COS	2009	117
Jacob, Scherpereel and Adams, 2014	Women	COS	5 yr. increments 1979-2009	120
This study	Ethnicity	COS	1967-2017	149

SI.3: Dataset Specifics

The CIA's COS dataset is extensively used by academics and practitioners and is trusted as a reliable source of information on world leaders. The CIA includes leaders from all countries with recognized governments other than the United States in the dataset. These data are only as accurate as the CIA's abilities to collect information on world leaders and countries' abilities to announce the appointment of new leaders. That said, there have thus far not been concerns raised about the validity of this dataset, and any validity concerns would likely be country specific and, therefore, would be captured in country fixed-effects.

The COS dataset includes all cabinet level ministers as well as the country's Ambassador to the United States (since 1994) and representative to the United Nations (since 1996). The CIA does not include data for the United States. Countries occasionally enter and exit the dataset when the CIA determines that no stable government is in place and, thus, no cabinet listing is available.

The dataset of cabinet minister names runs from 1967 to 2017. COS reports prior to 2001 are not available from the CIA, so I collected them from scanned copies of printed reports held at several university libraries. Reports post-2001 are available on the CIA website in PDF format. See `https://www.cia.gov/library/publications/world-leaders-1/`. The COS report is published monthly; I choose the January report for consistency. The data from 1967 to 2001 comes from scans of the COS report from The Ohio State University, Michigan State University, the University of Michigan, and the University of California. A team of software developers assisted in extracting the text of these yearly COS PDF documents and assembling a database of cabinet members. Scans varied considerably in quality, as did the quality of the print. From 1967 to 1971, the reports appear to have been produced on a typewriter, making some characters and words almost indecipherable. Print quality improved in 1972 and again in 2001 when digital versions were produced. The software developers first processed the PDFs through Microsoft Excel Macros, which extracted the data from the PDFs and placed them in appropriate columns. Because of PDF quality, the team then cleaned the data by hand checking each entry for foreign characters and incorrect spelling. For example, sometimes the country name appeared in the minister name field due to computer processing errors.

I then hand checked all 233,582 resulting cabinet member's names to make sure no extraneous information was present due to processing errors. Each name was prepared by removing accent marks, military titles, and foreign characters that could not be recognized by the NamePrism system. Consistent spelling was key. All names were taken from the official CIA translations and were presented in Roman script. Spelling and name translation issues that vary across countries and years are accounted for using fixed effects.

SI.4: Stem and Leaf Ethnicities

NamePrism identifies thirty-nine name communities or groups of names where people with these names frequently interact with each other. The authors of NamePrism then decided to label these name communities with geographic designations representing the dominant name community in a given geographic region. Table SI.4.1 displays the stem and leaf ethnicities representing these geographic categories as defined by Ye et al. (2017). These classifications are geographically based names for the thirty-nine name communities that emerge from the training dataset. Stem ethnicities can contain one or two subgroups of leaves. For example, the Muslim stem contains a Turk leaf which is then further differentiated into Central Asia and Turkey leaves. However, another Muslim leaf is Arabian Peninsula, which is not differentiated into subleaves. South Asia, Celtic English, Israel, and Greece are all stems that do not have leaves. I refer to the ten stem ethnicities as those listed in the stem column and the thirty-nine leaf ethnicities as the smallest subunit of each stem.

Importantly, NamePrism is used to provide an estimate of the diversity of names within a given cabinet, not the actual ethnic identification of any one name. Therefore, lack of coverage for any particular geographic region does not mean that NamePrism will fail to detect variation in names within this region. Country fixed-effects in all model specifications account for varying numbers of name communities present in any given country.

Table SI.4.1: Stem and Leaf Ethnicities

Stem	Supra Leaves	Sub Leaves
Muslim	Turk	Central Asia (Kazakhstan, Azerbaijan, Uzbekistan)
Muslim	Turk	Turkey
Muslim	Arabian Peninsula (UAE, Iraq, Yemen, Bahrain, Syria, Jordan, Oman, Lebanon, Qatar, Kuwait, Saudi Arabia)	
Muslim	Maghreb (Tunisia, Morocco, Algeria, Libya)	
Muslim	Nubian (Egypt, Somalia, Sudan)	
Muslim	Persian (Iran, Afghanistan)	
Muslim	Pakistanis	Bangladesh
Muslim	Panistanis	Pakistan
European	East European (Czech Republic, Slovakia, Hungary, Poland)	
European	South Slav (Bosnia and Herzegovina, Serbia, Slovenia, Montenegro, Bulgaria, Croatia, Macedonia)	
European	German (Germany, Austria, Netherlands, Switzerland)	
European	Baltics (Estonia, Latvia, Lithuania)	
European	French (Belgium, France)	
European	Russian (Russia, Belarus, Ukraine)	
European	Italian	Italy
European	Italian	Romania
Africa	West African (Congo, Liberia, Senegal, Sierra Leone, Guinea, Ghana, Togo, Benin, Nigeria)	
Africa	East African (Zambia, Rwanda, Tanzania, Kenya, South Sudan, Ethiopia, Uganda)	
Africa	South Africa (Botswana, Namibia, Zimbabwe, South Africa)	

Table SI.4.1: Stem and Leaf Ethnicities

Stem	Supra Leaves	Sub Leaves
East Asia	Indochina	Cambodia
East Asia	Indochina	Myanmar
East Asia	Indochina	Thailand
East Asia	Indochina	Vietnam
East Asia	Chinese (Singapore, Hong Kong, China, Taiwan)	
East Asia	Malay	Indonesia
East Asia	Malay	Malaysia
East Asia	South Korea	
East Asia	Japan	
Hispanic	Spanish (Dominican Republic, Uruguay, Guatemala, Colombia, Venezuela, Chile, Panama, Bolivia, El Salvador, Ecuador, Argentina, Honduras, Peru, Costa Rica, Paraguay, Mexico, Spain, Nicaragua)	
Hispanic	Portuguese (Portugal, Brazil, Angola, Mozambique)	
Hispanic	Philippines	
Nordic	Scandinavian	Denmark
Nordic	Scandinavian	Sweden
Nordic	Scandinavian	Norway
Nordic	Finland	
South Asia (Nepal, India, Sri Lanka)		
Celtic English (Ireland, United Kingdom)		
Israel		
Greece		

SI.5: Calculating Cabinet Diversity

To explain in detail how the cabinet diversity calculation process works, I consider an example from the cabinet of Afghanistan in 2017. Afghanistan is an ideal case in which to code ethnicity because comprehensive self-identified ethnicity codings exist for recent cabinet members and because Afghanistan has a diverse mix of ethnic groups that includes tribal groups and different nationalities. These groups are not common outside of Afghanistan, so this case is one of the worst for identifying name communities using NamePrism. Table SI.5.1 shows the names and positions of six cabinet ministers along with their ethnic identification as listed on a database of Afghan leaders.[7] The remaining three columns are three of the thirty-nine name communities.

Table SI.5.1: Calculating Cabinet Diversity

Name	Position	Listed Ethnicity	Muslim-Persian Pct.	Muslim-Paikistanis Pakistan Pct.	South Asian Pct.
Abdul Bari Jahani	Information and Culture	Pashtun	0.934	0.0253	0.0004
Asadullah Hanif Balkhi	Education	Tajik	0.0161	0.8077	0.0331
Faiz Muhammad Osmani	Hajj and Islamic Affairs	Turkmen	0.1388	0.0015	0.2427
Farida Mohmand	Higher Education	Pashtun	0.2065	0.3599	0.014
Ferozuddin Feroz	Public Health	Tajik	0.1366	0.4187	0.1764
Ghulab Nabi Farahi	Parliamentary Affairs	Pashtun	0.8703	0.0005	0.0327
		Sum	2.3023	1.6136	0.4993
		Average	0.3837	0.2689	0.0832

First, it is clear that the listed self-identification of cabinet ministers does not match any of the names of the thirty-nine name communities. Instead, there is a correspondence between a particular ethnic self-identification and a name community. Pashtun leaders' names tend to identify as members of the Muslim-Persian name community, as exemplified by Ministers Jahani and Farahi. Tajik ministers' names tend to identify as Muslim-Pakistanis-Pakistan, as exemplified by Ministers Balkhi and Feroz. Minister Osmani, a Turkmen, has a name that identifies most strongly as South Asian. Sometimes, however, names do not strongly correspond to any particular name community. Minister Mohmand is Pashtun, but NamePrism identifies him as primarily Muslim-Pakistanis-Pakistan.

One way to better address names that do not strongly signal membership in any one name community is to use the predicted probabilities from the thirty-nine name communities to provide information about the ethnic diversity inherent in each Minister's name. In Minister Jahani's case, his name loads almost completely on Muslim-Persian, but other Minister's names belong less distinctly to any one name community.

This technique is also designed to address the common problem that names from different

[7] `http://www.afghan-bios.info/database.html`

ethnicities are all from the same geographic region and, therefore, their name community with the highest predicted probability is the same. By using predicted probabilities, names that have the highest predicted probability in the same name community can vary on the predicted probabilities associated with other name communities. This means that a set of names that are all primarily West African may vary in the predicted probability that the names are South African or Muslim. These other predicted probabilities help to distinguish the ethnic identity of names from the same geographic region. All of the predicted probabilities add information about the diversity of the names in the cabinet that helps to provide an overall measure of cabinet diversity. I demonstrate the validity of this method by comparing it to handed coded ethnicity data in SI.12.

I will illustrate the cabinet diversity score calculation using just these six Ministers. To obtain p, the average predicted probability across ministers in each name community, I take the sum of the predicted probabilities for each name community (the Sum row in Table SI.5.1) and divide by the total number of ministers in the cabinet (in this example six). The result is the Average row in Table SI.5.1, which I then square. I then calculate the cabinet diversity score by taking $1 - \sum_{i=1}^{39} 0.3837^2 + 0.2689^2 + 0.0832^2 + \cdots$ where the averages in the remaining thirty-six name communities are also added. Once calculated using the names of all cabinet ministers, this represents the cabinet diversity score for Afghanistan in 2017.

I do not weight cabinet ministries based on their perceived importance because the importance of any one ministry is country and time period specific; coding important cabinet positions is left for future research.

SI.6: Converting Cabinet Diversity to Seat Change

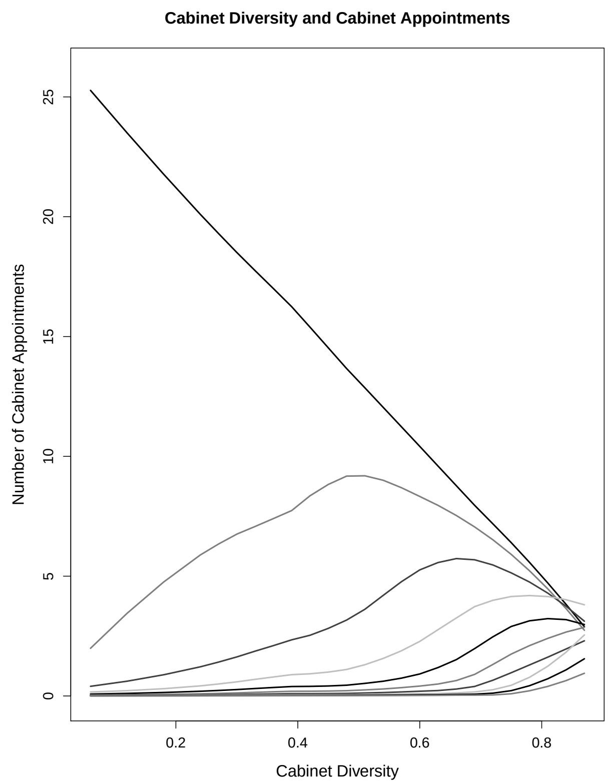
The *Cabinet Diversity* measure is a cabinet-level measure, but country leaders change *Cabinet Diversity* by adding or removing individuals from the cabinet. To convert from *Cabinet Diversity* to cabinet seat changes I simulate possible cabinet configurations given a *Cabinet*

Diversity score. To restrict the analysis a bit, I focus on the mean cabinet size of 28 ministers, and I assume there are ten potential ethnic groups that could enter the cabinet. I simulate 100,000 possible ways that the country leader could allocate cabinet seats to members of the ten ethnic groups.[8] I then calculate the *Cabinet Diversity* score for each possible cabinet seat allocation. For simplicity, in each simulation I assign the ethnic group receiving the most seats to group 1 and the ethnic group receiving the fewest seats to group 10. Figure SI.6.1 displays the results where each curve represents the average number of cabinet seats that a group receives at a given level of *Cabinet Diversity*.[9] When *Cabinet Diversity* is close to zero, one group dominates the cabinet. As *Cabinet Diversity* increases, new groups are introduced into the cabinet, and the number of ministers belonging to the previously dominant group decreases. At a *Cabinet Diversity* score of about 0.9, all ten groups have roughly equal representation in the cabinet. Thus, when a citizen observes the ministers in the cabinet, they will recognize ten ethnic groups with a relatively similar number of ministers belonging to each group. More typically, at a *Cabinet Diversity* level of 0.5, group 1 has 13 seats, group 2 has 10 seats, group 3 has 4 seats, and group 4 has 1 seat.

[8]The simulation is constructed by randomly sampling the number of cabinet seats allocated to each ethnic group. The sampling procedure is ordered; the largest ethnic group is sampled first (from 0 to 28) and each ensuing sample has a maximum size of 28 minus the seats already allocated. I then use these samples to calculate a *Cabinet Diversity* score.

[9]I use a LOWESS smoother with a span of 50% to smooth over artifacts of the simulation.

Figure SI.6.1: Relationship Between Cabinet Diversity and Cabinet Appointments

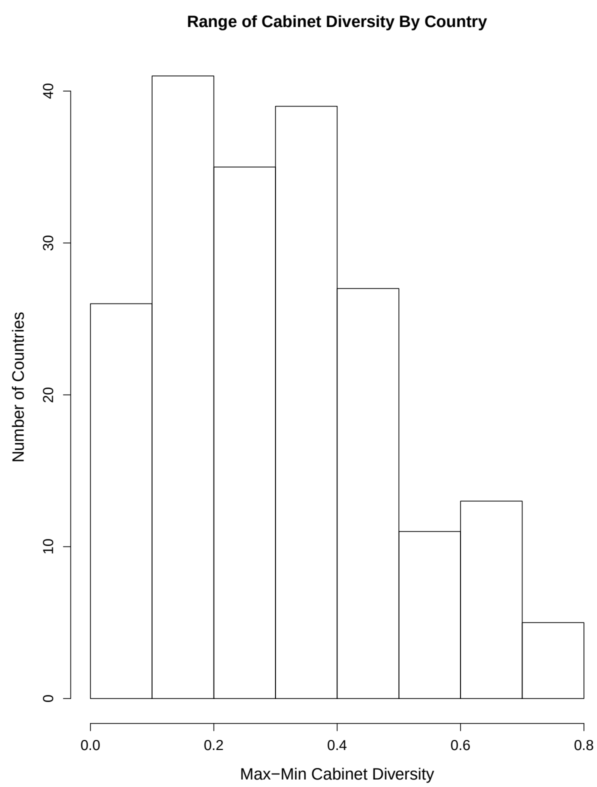


Cabinet membership among ten ethnic groups over varying levels of *Cabinet Diversity*. Each line represents the number of seats allocated to one of the ten ethnic groups. Colors are used to help differentiate between groups.

SI.7: Cabinet Diversity Trends

Figure SI.7.1 displays a histogram of the within country variation of *Cabinet Diversity*. We can see that there is quite a lot of change within countries over time, with many countries dramatically increasing or decreasing their *Cabinet Diversity* during the time series. For example, Zimbabwe has the twelfth highest range, 0.64. This is in line with common knowledge about how radically the cabinet has shifted in Zimbabwe. Countries like Kazakhstan and Tajikistan have shifted the least and maintained a high level of *Cabinet Diversity* throughout the time series.

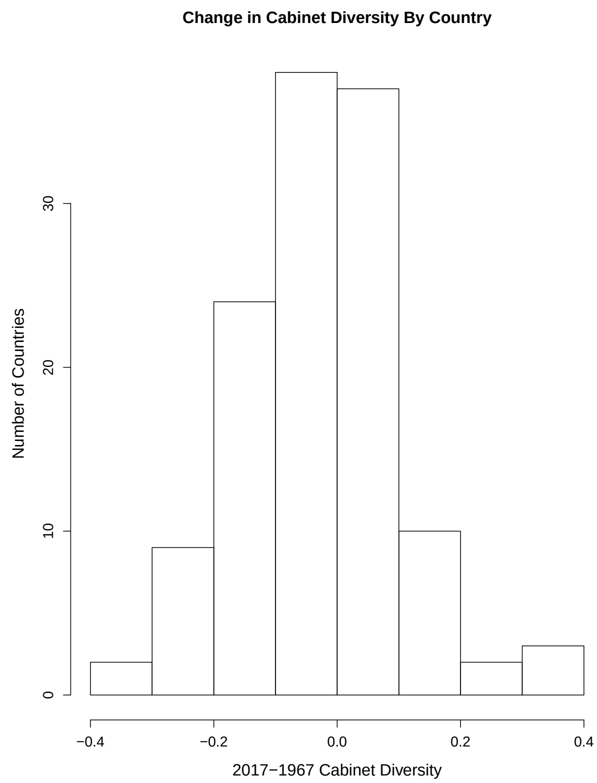
Figure SI.7.1: Within Country Range of Cabinet Diversity



The figure includes all 197 currently recognized countries, though some do not span the entire dataset. Range is the country-year with the maximum cabinet diversity score minus the country-year with the minimum score.

Figure SI.7.2 displays a histogram of the change in *Cabinet Diversity* in 2017 compared to 1967. Unlike Figure SI.7.1, this Figure only accounts for *Cabinet Diversity* in these two years. As should be evident, *Cabinet Diversity* does not always linearly increase in a particular country. This is consistent with the theory that non-programmatic leaders play an important role in increasing *Cabinet Diversity*. The prevalence of such leaders is not dictated by time period.

Figure SI.7.2: Change in Cabinet Diversity from 1967 to 2017



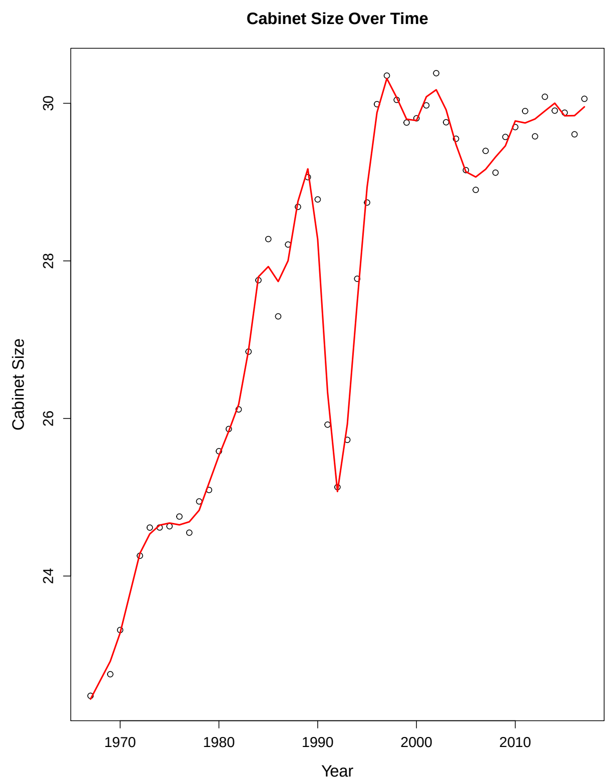
The figure includes 125 countries with observations in both 1967 and 2017.

Figure SI.7.3 confirms the results from previous studies like Opalo (2011) that claim cabinet size is increasing over time. I also confirm the cabinet size trends present in the Cross-National Time Series (Banks and Wilson, 2016) dataset. The COS data is likely more accurate than previous measures of cabinet size because it relies on the same source, the CIA, for all countries and years; other data sources are compiled from myriad primary and secondary materials. In general, cabinet size has increased significantly since the late 1960s.[10] Cabinets have grown from fewer than twenty-four members in 1969 to about thirty in 2017.

[10]Figure SI.7.3 displays a LOWESS smoother with a span of 15% for ease of interpretation.

The collapse of the Soviet Union had a significant impact on cabinet size for several years while new countries were created and cabinets filled.

Figure SI.7.3: Increasing Cabinet Size Over Time



Average number of seats in cabinets worldwide since 1967. Based on COS data.

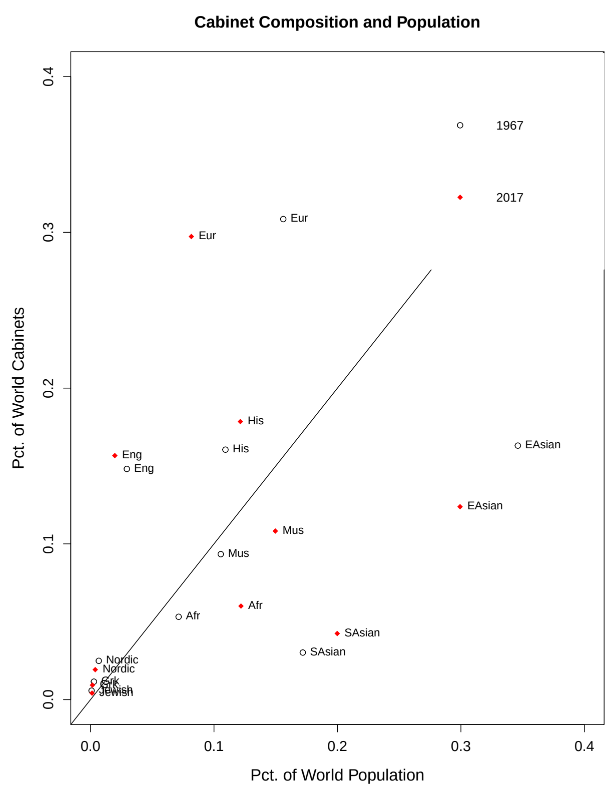
Figure SI.7.4 compares the representation of stem ethnicities in cabinets to their world population.[11] We would expect some correlation between the two, but not perfect correlation because this comparison is between world ethnic population and name based ethnic cues among cabinet ministers. Those ethnicity-years above the line are over-represented in world cabinets compared to their population, while those under the line are under-represented. Europeans and those with European ancestry dominate cabinets, making up about thirty percent of cabinet membership, with English and Hispanic members each taking about fifteen percent. Europeans are also the most over-represented group, consisting of between ten and fifteen percent of the world population, but taking thirty percent of the world's cabinet seats.

[11] World population data is from the World Bank.

English and Hispanic members are also over-represented. These three groups were the main colonizers, so their persistent influence in cabinets is not surprising.[12]

Africans, Muslims, and South and East Asians are significantly under-represented. The percentage of East Asians has declined somewhat since 1967, meaning that East Asians are slightly less under-represented. However, the percentage of Africans, Muslims, and South Asians has grown without their share of cabinet representation increasing. The other stem ethnicities, Nordic, Jewish, and Greek, are not significantly represented in the world population or in cabinets.

Figure SI.7.4: Cabinet Composition and Population Over Time



Diagonal line shows equal representation in population and cabinets. Stem ethnicities above the line are over-represented in cabinets, those below the line are under-represented.

[12]Some of the effect may be because colonized citizens adopted names more closely aligned with their colonizers, but that certainly does not explain this large over-representation.

SI.8: Control Variables and Country ELF

In this section, I describe control variables used in the analysis as well as the *Country ELF* variable. First, it is reasonable to expect that more diverse countries will have more diverse cabinets (Francois, Rainer and Trebbi, 2015). Leaders of diverse countries will naturally select more diverse ministers because the pool of qualified individuals for a ministerial post will be diverse. This means that the level of diversity in a country is one of the most important control variables for this analysis.

I expect that changes in country-level ethnic diversity will be correlated with changes in cabinet-level diversity. Country-level ethnic diversity can change because of immigration, emigration, and differential birth or death rates. Another common way for country-level ethnic diversity to change occurs when individuals alter their location within existing ethnic frameworks (Chandra, 2012, 153-154).[13]

I use a measure of country-level ethnolinguistic fractionalization (ELF) to capture country ethnic diversity. *Country ELF* is defined as $1 - \sum_{i=1}^{n} q^2$ where n is the number of politically relevant ethnic groups at a given point in time in a given country and q is the proportion of the population belonging to each group. Higher values indicate more politically relevant country diversity.[14]

A number of institutional factors may influence cabinet diversity. First, the level of democracy in a country might dictate a leader's ability to make the cabinet appointments they desire. I measure democracy using the standard *Polity* IV score (from -10 to 10). I do not preclude authoritarian regimes from this analysis, as even authoritarian leaders need to maintain public support. Cabinet ministers are typically chosen from the legislature

[13]Country leaders use may also use government resources like census categories to impose new ethnic categories on citizens, but the ethnic identity measure I use excludes government orchestrated semantic changes in ethnic membership (Vogt et al., 2015, 1329).

[14]My measure comes from Cederman, Wimmer and Min (2010). See SI.11 for other measures of *Country ELF*. Politically relevant ethnic groups are those that have at least one organization claiming to represent them at the regional or national level, irrespective of national recognition. Groups that would have such representation were it legally allowed are also considered politically relevant.

(Back et al., 2008). A larger legislature means more opportunities for ethnic minorities to be represented and to catch the attention of the country leader, making their appointment to the cabinet more likely (Escobar-Lemmon and Taylor-Robinson, 2005). Thus, I include an indicator, *Legislature Size*, from CNTS to represent the logged size of the lower chamber of the legislature (ranging from 0 to 7). If a leader just suffered a coup, they might insulate themselves with a less diverse cabinet (Mehler, 2009; Roessler, 2016). *Coups* is a count of the number of coups in a given year from CNTS.

Relatedly, if the country leader is from a minority ethnic group, that leader might not have the political power to appoint diverse members to the cabinet. Instead, they may be stuck appointing ministers of the majority ethnicity in order to counteract their own diversity (O'Brien et al., 2015). I include an indicator *Leader Group Size* that ranges from 0 to 1 based on the leader's group's percentage of the population (Fearon, Kasara and Laitin, 2007).[15]

Leaders' selection of cabinet ministers is often constrained by party dynamics. Leaders managing coalitional governments may be forced to cede some cabinet appointment decisions to coalition partners. I include a dichotomous variable *Coalition* from V-Dem to indicate whether the government is in coalition (Blondel and Thiebault, 1988). The presence of factional or identity based parties in the political system may impact how cabinets are formed and the attention paid to cabinet diversity (Htun, 2004). *Factional* is a dichotomous indicator from V-Dem indicating whether these parties are prevalent in the political system. Finally, large party majorities in parliament may be responsible for appointing more cabinet ministers from that particular political party (Back et al., 2008). *Majority* from CNTS is a logged index of seats held by the largest party in the legislature (ranging from 0 to 8.2).

Past and current violence may also influence cabinet composition. Civil conflict often heightens ethnic tensions, which may lead to the construction of less diverse cabinets. To measure civil conflict, I include *Peace Years* from Cederman, Wimmer and Min (2010) which

[15] See Table SI.10.1 for summary statistics of the control variables.

is a logged count of the number of years since the country was involved in a war.

I measure ongoing unrest using the logged number of *Terrorist Attacks* in a country-year as recorded by the Global Terrorism Database (GTD) (range from 0 to 8.3) (LaFree and Dugan, 2007). GTD uses a broad definition of a terrorist attack, though most are aimed at achieving economic, political, and social goals and must be outside of a warfare context. Of all types of ongoing unrest, terrorist attacks are the most unpredictable and potentially destabilizing (Hunter, J. Bennett and Robbins, 2018). Protests and riots typically involve only minimal violence, whereas terrorist attacks raise questions about the country leader's ability to prevent violent events from occurring in the country (Peek and Sutton, 2003).

Logged *GDP* per capita (in 2000 dollars) from V-Dem measures the financial resources of the leader. I also control for logged *Population* from Cederman and Girardin (2007) and the *Size* of the cabinet calculated from the cabinet diversity data. These indicators could impact the leader's ability or willingness to make diverse cabinet appointments.

SI.9: Stationarity Tests

I use a Fisher-type unit-root test based on Phillips-Perron tests and reject the null hypothesis of containing a unit root (specified in Stata using xtunitroot) (De Boef and Keele, 2008).

SI.10: Model Specifications

Table SI.10.1 shows summary statistics for relevant variables with the transformations applied to each variable listed in the second column.

Table SI.10.2 presents the full results from the model described in the main text. In the alternate specifications, I address endogeneity concerns with this type of model. From a technical standpoint, the lagged structure of the regression means that independent variables all occurred prior to the realization of the cabinet diversity dependent variable. Since fixed effects are included, this means that we are looking at how cabinet diversity changes based on

Table SI.10.1: Summary Statistics

	Scale	Min	Max	Stdev	Mean	Median
Cabinet Diversity	%	0.00	1.00	0.24	0.59	0.62
Country ELF	%	0.00	1.00	0.31	0.42	0.41
Non-Prog. Dist.	%	0.01	0.98	0.30	0.50	0.52
School	Log	1.39	7.62	0.84	6.18	6.45
Population	Log	5.61	14.10	1.37	9.26	9.15
Peace Years	Log	0.00	4.17	1.42	2.34	2.83
Terrorist Attacks	Log	0.00	8.28	1.45	0.89	0.00
Legislature Size	Log	0.00	6.91	2.07	4.02	4.71
Coups	Count	0.00	2.00	0.15	0.02	0.00
Leader Group Size	%	0.01	1.00	0.33	0.52	0.53
Coalition	Binary	0.00	1.00	0.48	0.36	0.00
Factional	Binary	0.00	1.00	0.38	0.17	0.00
Majority	Log	0.00	8.21	2.06	4.25	5.04
GDP	Log	4.90	12.30	1.20	8.67	8.69
Polity	Count	-10.00	10.00	7.75	-0.50	-4.00
Size	Count	1.00	277.00*	16.16	28.22	25.00

*Russia had notoriously large cabinets before the fall of the Soviet Union.

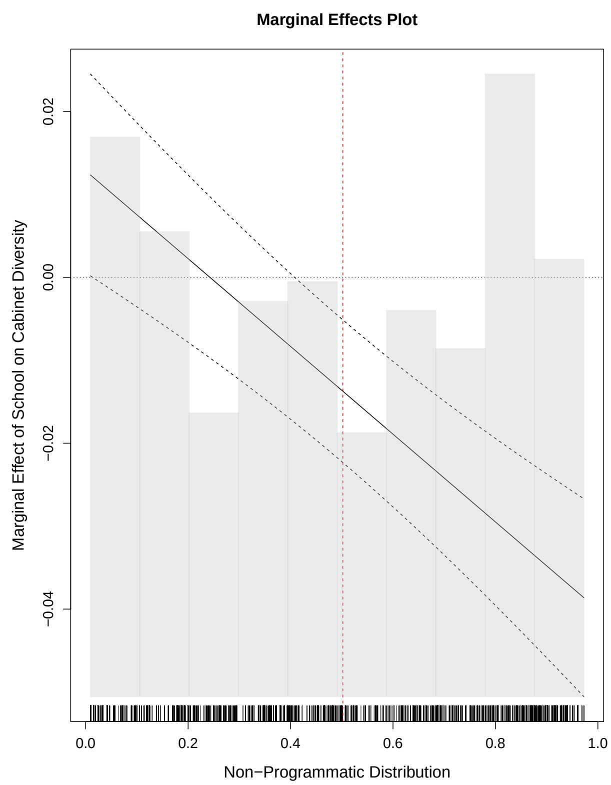
the interaction between non-programmatic distribution and school enrollment in the previous period. There is a robust literature studying the relationship between country-level diversity and good distribution that suggests that these variables may be endogenous. In this case, these two variables are allowed to be endogenous because they are modeled as co-occurring.

Does this endogeneity concern translate over to cabinet diversity and non-programmatic distribution? Such an argument would suggest that a diverse cabinet influences the types of goods the country leader distributes. This is a plausible explanation in cases where the country leader has very little control over her own cabinet and, thus, can only respond to the cabinet that she ends up with. However, most country leaders have much more flexibility even given constraints from coalitions and representational requirements to appoint who they wish to their cabinet. Thus, the country leader chooses they way she wishes to distribute goods and chooses the diversity of her cabinet. As the former is a change in personal management style whereas the latter requires hiring and firing cabinet ministers, it is more plausible that non-programmatic distribution changes first and is followed by a change in cabinet diversity.

This sequence is indeed what is suggested by the case study evidence presented in the main text.

Figure SI.10.1 shows the other side of the interaction in the main text. Though I do not theorize about this side of the interaction, it is plausible for an effect to exist. For especially non-programmatic leaders, increasing *School* somewhat decreases *Cabinet Diversity.* As private goods are reduced, non-programmatic leaders cannot productively use cabinet seats to provide patronage, so there is suggestive evidence that country leaders redirect their usage of cabinet seats and give them to co-ethnics.

Figure SI.10.1: Marginal Effect of School



Vertical line indicates mean value. 90% Confidence Intervals.

Table SI.10.3 presents a number of alternative specifications to check the robustness of the standard fixed effects model with country robust standard errors. Model 1 is a naive pooling model with no fixed effects, but clustered standard errors. Model 2 uses a Cochrane-Orcutt estimation method and includes region level fixed effects. Model 3 is a random effects

Table SI.10.2: Regression on Cabinet Diversity

	Dependent variable:
	Cabinet Diversity
Country ELF_{t-1}	0.083^{***}
	(0.026)
Non-Prog. Dist._{t-1}	0.311^{***}
	(0.062)
School_{t-1}	0.013^{*}
	(0.007)
Non-Prog. Dist.:School	-0.053^{***}
	(0.011)
Population_{t-1}	0.036^{***}
	(0.012)
Peace Years	-0.001
	(0.002)
Terrorist Attacks_{t-1}	-0.004^{**}
	(0.002)
Legislature Size_{t-1}	0.003
	(0.003)
Coups_{t-1}	0.017^{*}
	(0.009)
Leader Group Size_{t-1}	-0.002
	(0.015)
Coalition	0.007
	(0.005)
Factional	0.003
	(0.006)
Majority_{t-1}	-0.003
	(0.004)
GDP_{t-1}	-0.010
	(0.007)
Polity_{t-1}	0.001^{**}
	(0.001)
Size_{t-1}	-0.0004^{**}
	(0.0002)
Observations	3,180
R^2	0.026
Adjusted R^2	-0.021
F Statistic	5.105^{***} (df = 16; 3033)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

OLS with country fixed effects. Cluster robust standard errors in parentheses.

model of the main results.

The pure pooling and Cochrane Orcutt models are generally consistent with the results from the main model. More importantly, the random effects model including year intercepts on top of country fixed effects is consistent with the main model.

Model 4 contains fixed effects and a lag dependent variable. For those concerned with this technique because of Nickell (1981), recall that the bias decreases as the length of the panel increases. Judson and Owen (1999) show that this bias is negligible when using a panel with thirty periods or more. The panel is unbalanced with 197 total panels and an average of 42 periods. Thus, the fixed effects, lagged model should produce unbiased estimates of the true effect. Indeed, the results are the same in size and significance to the main fixed effects model with clustered standard errors.

A reason to specify Model 4 is that it allows us to calculate dynamic trends via the lagged dependent variable. The model in the main text estimates the total effect of cabinet diversity over all periods, whereas Model 4 estimates the one period effect. Using Model 4, the long-term multiplier of the effect of country ELF on cabinet diversity is $\frac{\text{Country ELF}_{t-1}}{1-\text{Cabinet Diversity}_{t-1}} = \frac{0.049}{1-0.410} = 0.08$ (Williams and Whitten, 2012). It is the case here that the coefficient on cabinet diversity in the fixed effects model is 0.083. Similarly, the long-term multiplier of the effect of the coefficient of non-programmatic distribution is 0.337 compared to 0.311. This implies that the long-term effect mostly occurs in the first few periods; substantively, this means that the country leader quickly recognizes changes in country ELF and adjusts his cabinet accordingly (Nymoen, 2004, 14). Because the change appears to happen quickly, dynamic simulations are not particularly interesting (Williams and Whitten, 2012).

Model 6 presents a lagged dependent variable model without fixed effects. Lagged dependent variable models alone tend to under-estimate the size of effects, as most of the variation is consumed by the lagged dependent variable (Angrist and Pischke, 2009, 246). They also equate all countries with similar lagged values of cabinet diversity, which seems implausible in our case. Guryan (2001, 23) suggests bracketing the minimum effect size using the lagged dependent variable model. The signs of the main coefficients are in the correct direction, but some lose significance because of the very strong effect of the lagged dependent variable.

Model 7 is a random effects maximum likelihood estimation wherein countries are specified with random intercepts. This allows for between country comparisons without eliminating all of the country-level variation.

Despite the consistency in findings across many model specifications, this analysis is still limited by the quality of the available data and the methods employed. I address these concerns by employing many robustness checks both here and in the ensuing sections that consider alternative model specifications and independent and dependent variables. It is important to note that none of these variables are particularly easy to measure. Most variables are constructed from government data, which varies widely in quality. Further, many of the main measures employed in the analysis are proxies because of the difficulty of measuring these variables directly. This is why it is especially important to check the robustness of the results in many ways. Though the models use lagged independent variables, a main limitation of this approach is that causal leverage is relatively limited. Telling a causal story in this context would mean interviewing country leaders to precisely understand their thinking when appointing cabinet ministers. Even such interviews are unlikely to reveal much because country leaders have no incentive to reveal how they think about cabinet appointments to anyone who could potentially provide this information to their political rivals. Future research may wish to consider interviewing country leaders to see what they say about ethnicity in the cabinet appointment process.

Table SI.10.3: Alternate Specifications

	Dependent variable:						
	Cabinet Diversity						
	Pooling	*Cochrane Orcutt*	*Random Effects*	*FE w/ Lag*	*RE w/ Lag*	*Lag*	*REML*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cabinet Diversity$_{t-1}$				0.410***	0.412***	0.913***	
				(0.017)	(0.017)	(0.008)	
Country ELF$_{t-1}$	0.207***	0.037*	0.086***	0.049**	0.052**	0.023***	0.080***
	(0.015)	(0.022)	(0.026)	(0.024)	(0.024)	(0.007)	(0.024)
Non-Prog. Dist.$_{t-1}$	−0.849***	0.127	0.307***	0.199***	0.195***	−0.068*	0.264***
	(0.093)	(0.108)	(0.062)	(0.056)	(0.057)	(0.040)	(0.061)
School$_{t-1}$	−0.053***	−0.002	0.013*	0.008	0.008	−0.001	0.011
	(0.011)	(0.013)	(0.008)	(0.007)	(0.007)	(0.005)	(0.008)
Population$_{t-1}$	−0.013***	−0.038***	0.018	0.036***	0.022	0.001	−0.005
	(0.004)	(0.005)	(0.019)	(0.011)	(0.017)	(0.002)	(0.012)
Peace Years	−0.014***	0.006*	−0.001	−0.001	0.0001	−0.001	−0.002
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
Terrorist Attacks$_{t-1}$	−0.025***	−0.002	−0.002	−0.002	−0.0003	−0.003**	−0.004***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Legislature Size$_{t-1}$	0.007	−0.002	0.005	0.004	0.005*	0.004	0.003
	(0.006)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Coups$_{t-1}$	−0.013	0.014*	0.017*	0.004	0.004	−0.016*	0.016*
	(0.022)	(0.008)	(0.009)	(0.009)	(0.009)	(0.010)	(0.009)
Leader Group Size$_{t-1}$	0.189***	−0.003	−0.004	0.008	0.006	0.022***	0.007
	(0.015)	(0.018)	(0.015)	(0.013)	(0.013)	(0.006)	(0.014)
Coalition	0.069***	0.013*	0.004	0.006	0.003	0.009**	0.007
	(0.010)	(0.008)	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)
Factional	0.054***	0.026***	0.003	0.0002	−0.0004	0.004	0.004
	(0.012)	(0.010)	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)
Majority$_{t-1}$	−0.017**	0.004	−0.004	−0.004	−0.005	−0.005	−0.003
	(0.007)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)
GDP$_{t-1}$	0.021***	0.040***	−0.013*	−0.002	−0.004	−0.001	−0.012*
	(0.005)	(0.008)	(0.007)	(0.006)	(0.007)	(0.002)	(0.007)
Polity$_{t-1}$	−0.006***	−0.001	0.001*	0.001	0.0005	−0.001**	0.001*
	(0.001)	(0.001)	(0.001)	(0.0005)	(0.0005)	(0.0004)	(0.001)
Size$_{t-1}$	0.002***	−0.002***	−0.0004**	−0.001***	−0.001***	−0.001***	−0.0004**
	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0001)	(0.0002)
Year$_{t-1}$							0.001***
							(0.0004)
Non-Prog. Dist.:School	0.136***	−0.015	−0.052***	−0.034***	−0.033***	0.010	−0.044***
	(0.016)	(0.018)	(0.011)	(0.010)	(0.010)	(0.007)	(0.011)
Constant	0.662***	0.437***				0.066**	−1.247**
	(0.074)	(0.135)				(0.032)	(0.625)
Observations	3,180	3,180	3,180	3,180	3,180	3,180	3,180
R^2	0.205		0.022	0.188	0.186	0.858	
Adjusted R^2	0.201		−0.036	0.148	0.138	0.857	
Log Likelihood							2,949.788
Akaike Inf. Crit.							−5,859.576
Bayesian Inf. Crit.							−5,738.283
F Statistic	50.976***		4.157***	41.245***	40.357***	1,121.464***	
Note:							*p<0.1; **p<0.05; ***p<0.01

SI.11: Alternative Variables

In this section, I will discuss alternative specifications of cabinet diversity and country ELF.

As a robustness check, I create a revised version of cabinet diversity (called *Weighted Cabinet Diversity*) that downweights the influence of the majority name based ethnic group in a given country in the calculation of cabinet diversity. Ye et al. (2017) define the countries where each of the thirty-nine ethnicities is predominant.[16] Filling the cabinet with members that signal the predominant ethnicity is not an indication of cabinet diversity, so the new measure is calculated by $1 - \sum_{i=1}^{38} p^2 - \alpha q^2$ where p is the predicted probability of all thirty-eight non-majority ethnicities summed across all cabinet ministers and divided by the total number of ministers, q is the predicted probability of the majority ethnicity summed across all cabinet ministers and divided by the total number of ministers, and α is a scaling parameter which I generally vary $1 < \alpha \leq 2$.

Table SI.11.1 displays main model specifications substituting in the weighted measure for the cabinet diversity measure used in the main text. All of the main supply and demand variables retain their sign and most retain their significance. The weight used here is $\alpha = 2$. This suggests that the results are not an artifact of the measure used to calculate the level of diversity in the cabinet.

I check the robustness of the country-level ELF measure using two additional datasets, Annett (2001) and Alesina et al. (2003). One common complaint with the ELF measure is that ethnic, linguistic, and religious fractionalization are combined into a single measure. ELF has traditionally been considered a stagnant measure and was first collected in the Soviet Union publication *Atlas Narodov Mira.* Alesina et al. (2003) splits out ethnic, linguistic, and religious fractionalization based on data from Encyclopedia Britannica, the CIA, national censuses, and previous scholarly research. I use their ethnic fractionalization measure which was collected in whatever year is most proximate to 2003. I match this measure and year to

[16] They do this for 118 countries; I complete the correspondence.

Table SI.11.1: Alternative Dependent Variable

	Dependent variable:		
	Weighted Cabinet Diversity		
	Cochrane Orcutt	*FE*	*FE w/ Lag*
	(1)	(2)	(3)
Cabinet Diversity$_{t-1}$			0.368***
			(0.017)
Country ELF$_{t-1}$	0.017	0.063	0.045
	(0.040)	(0.042)	(0.039)
Non-Prog. Dist.$_{t-1}$	0.009	0.497***	0.325***
	(0.192)	(0.100)	(0.094)
School$_{t-1}$	0.003	0.015	0.007
	(0.023)	(0.012)	(0.011)
Population$_{t-1}$	−0.139***	0.030	0.047**
	(0.010)	(0.020)	(0.018)
Peace Years	0.009*	0.003	0.002
	(0.006)	(0.003)	(0.003)
Terrorist Attacks$_{t-1}$	−0.004	−0.008***	−0.005**
	(0.003)	(0.003)	(0.002)
Legislature Size$_{t-1}$	−0.013*	0.0003	0.002
	(0.007)	(0.006)	(0.005)
Coups$_{t-1}$	0.025*	0.030*	0.007
	(0.014)	(0.015)	(0.014)
Leader Group Size$_{t-1}$	−0.015	0.031	0.038
	(0.031)	(0.026)	(0.024)
Coalition	0.030**	−0.001	0.001
	(0.014)	(0.009)	(0.008)
Factional$_{t-1}$	0.057***	0.016	0.008
	(0.017)	(0.010)	(0.009)
Majority$_{t-1}$	0.016**	0.0004	−0.001
	(0.007)	(0.006)	(0.005)
GDP$_{t-1}$	0.049***	−0.011	−0.0001
	(0.015)	(0.011)	(0.010)
Polity$_{t-1}$	−0.002*	0.002*	0.001
	(0.001)	(0.001)	(0.001)
Size$_{t-1}$	−0.003***	−0.001***	−0.002***
	(0.0003)	(0.0003)	(0.0003)
Non-Prog. Dist.:School	−0.011	−0.089***	−0.057***
	(0.033)	(0.017)	(0.016)
Constant	0.313		
	(0.240)		
Observations	3,116	3,116	3,116
R^2		0.032	0.161
Adjusted R^2		−0.015	0.120
F Statistic		6.138*** (df = 16; 2972)	33.527*** (df = 17; 2971)
Note:			*p<0.1; **p<0.05; ***p<0.01

the country ELF data and find a correlation of 0.68, which is strong. Several other authors also attempt to break down ELF into component parts. In general, I believe that the cabinet ELF measure contains ethnic, religious, and linguistic fractionalization because it is based on people's names which are derived from all three of these factors. Thus, I expect weaker, but consistent results compared to the main model.

Annett (2001) takes a different approach to the issue of calculating ELF by claiming that data from the *Atlas Narodov Mira* is not detailed and is incomplete. He uses the *World Christian Encyclopedia* to obtain a new measure of ethnolinguistic fractionalization in 1982. This allows him to obtain a larger sample size then in the Soviet data. By contrast, Cederman, Wimmer and Min (2010) measure country ELF among all politically relevant ethnic groups as identified by an expert survey. The latter measure gives a much better indication of the relative strength of ethnic groups that might plausibly be appointed to the cabinet, while Annett's measure includes all ethnic groups. The correlation between the two measures is strong, 0.74. Thus, I again expect weaker, but consistent results compared to the main model.

Table SI.11.2 displays the results. Models 1 and 2 use the Alesina measure, while Model 3 uses Annett. Regional Fixed Effects are used. In most cases, the sign and magnitude of the various supply and demand variables are consistent with the main model. The *Non-Programmatic Distribution* variable is now non-positive, but is not significant. In general, because panel data is not available, these regression results rely on between country variation and the effectiveness of regional fixed effects. Neither is anything close to ideal.

The problem with adopting any measure other than Cederman, Wimmer and Min (2010) is that there are no other cross-national time-series datasets measuring country ethnicity. This is mostly due to the incorrect assumption that ELF does not change over time.

As noted previously, Cederman, Wimmer and Min (2010)'s measure of ELF accounts for what they determine to be politically relevant ethnic groups. However, even if an ethnic identity is politically relevant, it is not necessarily politicized. When there are systematic

Table SI.11.2: Alternative Independent Variable

	Dependent variable:		
	Cabinet Diversity		
	Cochrane Orcutt	*FE*	*FE*
	(1)	(2)	(3)
Alesina_{t-1}	0.195^{*}	0.126	
	(0.111)	(0.121)	
Annett_{t-1}			0.146
			(0.152)
Non-Prog. Dist._{t-1}	0.015	0.383	−0.553
	(0.678)	(0.674)	(0.541)
School_{t-1}	−0.010	0.031	−0.015
	(0.077)	(0.079)	(0.068)
Population_{t-1}	−0.005	−0.010	−0.009
	(0.021)	(0.022)	(0.021)
Peace Years	0.010	−0.001	−0.034
	(0.016)	(0.017)	(0.022)
Coups_{t-1}	−0.021	−0.021	-0.043^{**}
	(0.017)	(0.017)	(0.020)
Legislature Size_{t-1}	−0.035	−0.046	−0.025
	(0.033)	(0.036)	(0.052)
Coups_{t-1}	−0.053	−0.073	0.031
	(0.124)	(0.127)	(0.110)
Leader Group Size_{t-1}	0.070	0.008	0.071
	(0.107)	(0.110)	(0.124)
Coalition	0.037	0.013	0.168^{*}
	(0.056)	(0.057)	(0.094)
Factional	0.173^{***}	0.159^{***}	0.061
	(0.047)	(0.049)	(0.095)
Majority_{t-1}	0.002	0.025	0.008
	(0.036)	(0.038)	(0.054)
GDP_{t-1}	−0.024	−0.012	0.026
	(0.035)	(0.034)	(0.038)
Polity_{t-1}	−0.007	-0.009^{*}	−0.007
	(0.005)	(0.005)	(0.007)
Size_{t-1}	0.001	0.002	0.001
	(0.001)	(0.002)	(0.002)
Non-Prog. Dist.:School	0.007	−0.054	0.103
	(0.108)	(0.108)	(0.094)
Constant	0.542	0.209	0.321
	(0.627)	(0.652)	(0.560)
Observations	110	110	97
R^2		0.648	0.638
Adjusted R^2		0.489	0.449
Residual Std. Error		0.174 (df = 75)	0.171 (df = 63)
F Statistic		4.064^{***} (df = 34; 75)	3.370^{***} (df = 33; 63)
Note:			*p<0.1; **p<0.05; ***p<0.01

inequalities in the distribution of political power, incorporating a more ethnically diverse cabinet is even more valuable an endeavor. I account for the potential impact of politicized ethnicity in two ways. First, the main statistical models include a variable *Factional* that represents whether political competition occurs through ethnic political parties. In general, this variable does not significantly influence cabinet diversity by itself. The presence of ethnic parties could plausibly interact with country-level ELF to produce a particularly politicized ethnic environment where the effect on cabinet diversity is pronounced. Model 1 in Table SI.11.3 interacts *Factional* with country-level ELF to determine if ELF in a particularly ethnically polarized environment has an especially large impact on cabinet diversity. This is not the case. ELF continues to increase ethnic cabinet diversity, but the interaction with the presence of ethnic parties does not heighten this effect.

Second, I introduce a new variable measuring how political power is distributed across social groups including ethnicity. This variable, *Power*, runs from 0 to 4 where 0 means that one social group co-opts all political power and 4 means equitable distribution of power across social groups. I interact this variable with country-level ELF in Model 2 of Table SI.11.3. Again, country-level ELF is important, but contexts with equitable access to political power are not more likely to have a diverse cabinet. Interestingly, when re-coding *Power* into a dummy variable where 1 represents equitable distribution of power, a 4 on the original scale, and 0 represents all other power distribution structures (*Power 01*), contexts where power is equitably distributed have less cabinet diversity (Model 3). That is, when the politicization of ethnicity is a non-issue because political power is equally shared, cabinets are less diverse. All measures of ethnic politicization have problems, but the main results regarding both the influence of country-level ELF on cabinet diversity and the significance of the interaction between *Non-Programmatic Distribution* and *School* indicate that these relationships are robust to model specifications.

Table SI.11.3: Alternative Measures of Ethnic Polarization

	Dependent variable:		
	Cabinet Diversity		
	(1)	(2)	(3)
Country ELF_{t-1}	0.084^{***}	0.080^{**}	0.083^{***}
	(0.026)	(0.031)	(0.026)
Factional	0.005		
	(0.011)		
Power_{t-1}		-0.009	
		(0.006)	
Power 01_{t-1}			-0.053^{***}
			(0.014)
Non-Prog Dist_{t-1}	0.312^{***}	0.302^{***}	0.295^{***}
	(0.062)	(0.063)	(0.062)
School_{t-1}	0.013^{*}	0.011	0.011
	(0.007)	(0.008)	(0.008)
Country ELF:Factional	-0.005		
	(0.019)		
Country ELF:Power		0.005	
		(0.011)	
Country ELF:Power 01			-0.002
			(0.050)
Non-Prog Dist:School	-0.053^{***}	-0.052^{***}	-0.050^{***}
	(0.011)	(0.011)	(0.011)
Population	0.036^{***}	0.042^{***}	0.037^{***}
	(0.012)	(0.012)	(0.012)
Peace Years	-0.001	-0.002	-0.002
	(0.002)	(0.002)	(0.002)
Terrorist Attacks_{t-1}	-0.004^{**}	-0.004^{**}	-0.004^{**}
	(0.002)	(0.002)	(0.002)
Legislature Size_{t-1}	0.003	0.004	0.004
	(0.003)	(0.004)	(0.004)
Coups_{t-1}	0.017^{*}	0.018^{*}	0.018^{*}
	(0.009)	(0.009)	(0.009)
Leader Group Size_{t-1}	-0.002	-0.002	-0.0002
	(0.015)	(0.015)	(0.015)
Coalition	0.007	0.007	0.006
	(0.005)	(0.005)	(0.005)
Majority_{t-1}	-0.003	-0.003	-0.003
	(0.004)	(0.004)	(0.004)
GDP_{t-1}	-0.010	-0.010	-0.009
	(0.007)	(0.007)	(0.007)
Polity_{t-1}	0.001^{**}	0.002^{***}	0.002^{***}
	(0.001)	(0.001)	(0.0005)
Size_{t-1}	-0.0004^{**}	-0.0004^{**}	-0.0005^{***}
	(0.0002)	(0.0002)	(0.0002)
Observations	3,180	3,212	3,212
R^2	0.026	0.028	0.034
Adjusted R^2	-0.021	-0.019	-0.012
F Statistic	4.807^{***} (df = 17; 3032)	5.107^{***} (df = 17; 3064)	6.422^{***} (df = 17; 3064)
Note:			*p<0.1; **p<0.05; ***p<0.01

SI.12: Cross-Country Applicability and Validation

NamePrism inevitably works better in some country contexts than others. I acknowledge that this is the nature of trying to develop an ethnic diversity measure across so many countries. To reiterate, NamePrism works not because it can identify precise ethnic categorization, rather NamePrism identifies cabinet-level name variation that is indicative of cabinet diversity. In this section, I present a validation of this approach by comparing NamePrism with hand coded ethnicity data from fifteen African countries. I then explore and discuss some of the issues associated with trying to identify countries where NamePrism may do a particularly good or poor job at identifying ethnic name diversity.

Validation

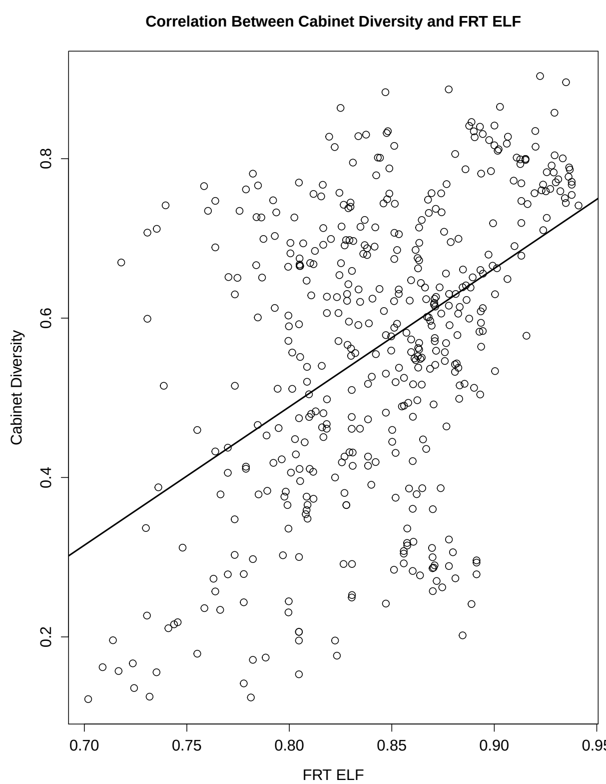
Francois, Rainer and Trebbi (2015) (FRT) build a model predicting ethnic cabinet balancing and test this model using hand coded ethnicities for cabinet members in fifteen African countries. I use their data on ethnic group presence in cabinets to compare with my measure of cabinet diversity. The comparison is inherently imperfect. FRT use all available sources to determine the ethnicity of ministers in these countries. This included employing expert consultants in each country to organize the ethnic identification process that relied heavily on expert classification and primary and secondary documents. Such information is certainly not present in the names of ministers and, consequently, it is unlikely that citizens can accurately identify the ethnicity of ministers. Nevertheless, there should be a moderate and positive correlation between the two measures because both are trying to measure ethnic cabinet diversity.

FRT calculate the proportion of cabinet seats belonging to members of each ethnic group in a given country. I use these data to calculate the Herfindahl-Hirschman index for each country-year's cabinet. The data generally spans from 1960 to 2004 and includes Benin, Cameroon, Cote d'Ivoire, Democratic Republic of Congo, Gabon, Ghana, Guinea, Kenya, Liberia, Nigeria, Republic of Congo (Brazzaville), Sierra Leone, Tanzania, Togo, and Uganda. I merge these data with the cabinet diversity scores which results in 448 country-year observations, an average time-series of 30 years per country.

The correlation between these indices is 0.46. This is surprisingly strong. NamePrism only contains three name communities related to African names, and its coverage of African names in general is quite poor. Additionally, names are a proxy for ethnic self-identification, so the 0.46 correlation meaningfully validates the cabinet diversity measure.

Figure SI.12.1 shows a scatterplot with a bivariate correlation line comparing the two indicies. As is evident, the scale of the two indicies is different. However, because models include country fixed-effects, we are interested in the within country change in diversity. Thus, the positive and relatively strong correlation displayed between NamePrism and the Francois, Rainer and Trebbi (2015) data is an important finding.

Figure SI.12.1: Correlation Between NamePrism and FRT Diversity Indicies



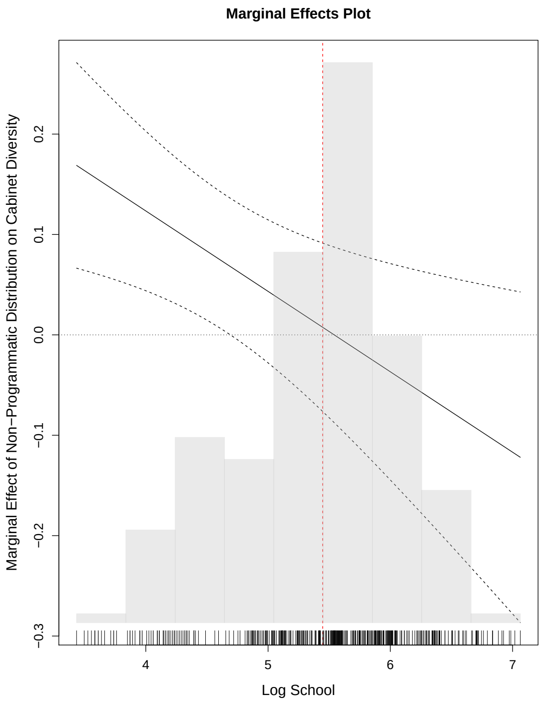
Points are country-year cabinet diversity estimates for fifteen African countries (FRT outliers excluded).

To determine whether the main effects hold in this subset of fifteen countries, I subset my dataset to include just these African countries. The results are in Table SI.12.1 and the marginal effects plots in Figure SI.12.2. Though I use different controls and a different estimation strategy, I find support for Francois, Rainer and Trebbi (2015)'s findings, *Country ELF* does positively correlate with *Cabinet Diversity.* It is very difficult to accurately estimate this model because country ELF is so high in almost all of these countries. Thus, the effect of *Population* on *Cabinet Diversity* may be inconsistently estimated because of Francois, Rainer and Trebbi (2015)'s sample of countries. This is why I include a much wider variety of countries in the main analysis so that country ELF will vary substantially from zero to one and we will not be making biased inferences based on extrapolating beyond the dataset.

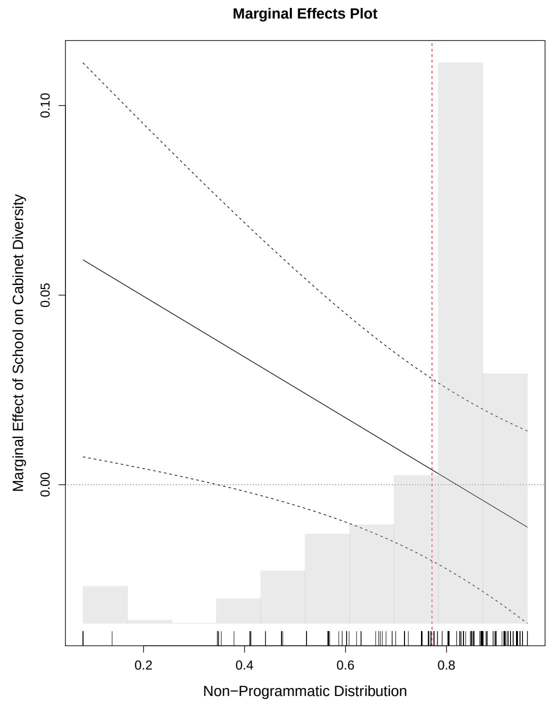
Table SI.12.1: Replicating Francois, Rainer, and Trebbi (2015)

	Dependent variable:	
	Cabinet Diversity	
	FE	*FE w/ Lag*
	(1)	(2)
Cabinet Diversity$_{t-1}$		0.433***
		(0.047)
Country ELF$_{t-1}$	0.212***	0.103***
	(0.040)	(0.038)
Non-Prog. Dist.$_{t-1}$	0.444**	0.329**
	(0.182)	(0.165)
School$_{t-1}$	0.066*	0.053*
	(0.034)	(0.031)
Population$_{t-1}$	−0.079**	−0.035
	(0.031)	(0.029)
Peace Years	−0.002	−0.002
	(0.005)	(0.005)
Terrorist Attacks$_{t-1}$	−0.003	−0.003
	(0.007)	(0.006)
Legislature Size$_{t-1}$	0.014	0.018**
	(0.010)	(0.009)
Coups$_{t-1}$	0.016	−0.006
	(0.021)	(0.019)
Leader Group Size$_{t-1}$	0.021	0.100
	(0.083)	(0.076)
Coalition	0.025	0.017
	(0.017)	(0.015)
Factional	−0.002	0.011
	(0.019)	(0.017)
Majority$_{t-1}$	−0.011	−0.018*
	(0.011)	(0.010)
GDP$_{t-1}$	0.012	0.009
	(0.020)	(0.018)
Polity$_{t-1}$	0.005**	0.003
	(0.002)	(0.002)
Size$_{t-1}$	−0.001***	−0.002***
	(0.0005)	(0.0004)
Non-Prog. Dist.:School	−0.080**	−0.066*
	(0.038)	(0.034)
Observations	410	410
R^2	0.176	0.326
Adjusted R^2	0.110	0.270
F Statistic	5.046*** (df = 16; 379)	10.735*** (df = 17; 378)
Note:		*p<0.1; **p<0.05; ***p<0.01

Figure SI.12.2: Marginal Effects on Cabinet Diversity



(a) Marginal Effect of Non-Programmatic Distribution



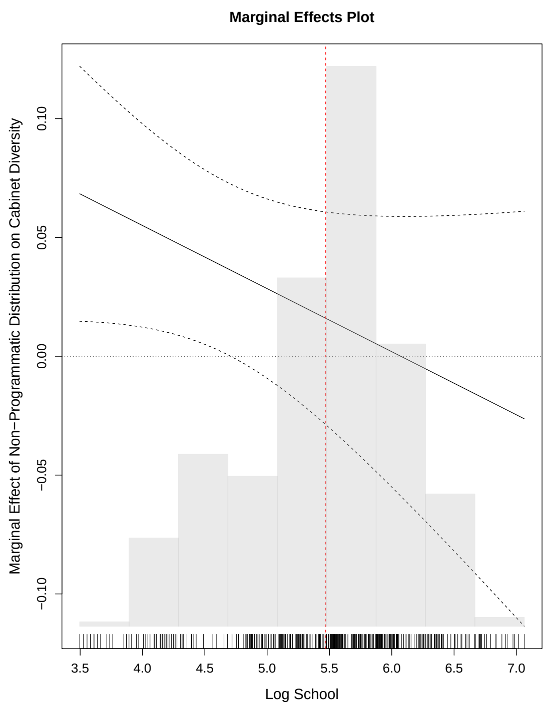
(b) Marginal Effect of School

I also use the Francois, Rainer and Trebbi (2015) data as the main dependent variable (Table SI.12.2). The table and Figure SI.12.3 show that the main results hold when examining only these African countries with Francois, Rainer and Trebbi (2015)'s measure.

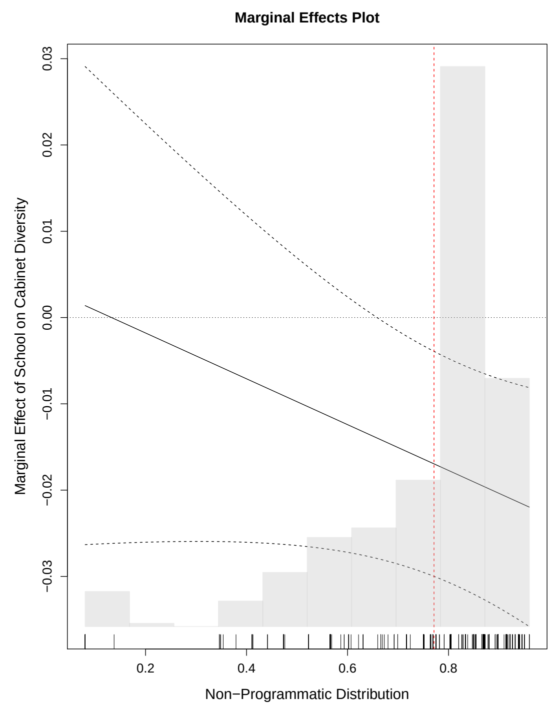
Table SI.12.2: Analysis with Francois, Rainer, and Trebbi (2015)'s Data

	Dependent variable:
	FRT ELF
Country ELF$_{t-1}$	0.347***
	(0.021)
Non-Prog. Dist.$_{t-1}$	0.161*
	(0.098)
School$_{t-1}$	0.004
	(0.018)
Population$_{t-1}$	0.040**
	(0.017)
Peace Years	$-$0.005*
	(0.003)
Terrorist Attacks$_{t-1}$	$-$0.005
	(0.004)
Legislature Size$_{t-1}$	$-$0.003
	(0.005)
Coups$_{t-1}$	0.030***
	(0.011)
Leader Group Size$_{t-1}$	$-$0.186***
	(0.044)
Coalition	0.010
	(0.009)
Factional	$-$0.006
	(0.010)
Majority$_{t-1}$	0.005
	(0.006)
GDP$_{t-1}$	0.010
	(0.011)
Polity$_{t-1}$	0.002*
	(0.001)
Size$_{t-1}$	$-$0.0004*
	(0.0002)
Non-Prog. Dist.:School	$-$0.027
	(0.020)
Observations	396
R^2	0.511
Adjusted R^2	0.471
F Statistic	23.829*** (df = 16; 365)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure SI.12.3: Marginal Effects on Cabinet Diversity



(a) Marginal Effect of Non-Programmatic Distribution



(b) Marginal Effect of School

Cross-Country Applicability

From the above validation, we can see that NamePrism effectively captures cabinet diversity in a context where there are three name communities dedicated to African names. While cabinet ministers in West Africa fall into the West African name community, the variation in the proportions of a name associated with the other name communities provide informative signals about the ethnic diversity present in the cabinet.

We know, however, that the name community method may work better in some countries or regions than others. First, we may be concerned that some countries have undue influence on the results, especially countries where ethnicity is particularly salient or names are a particularly good match with ethnicity. Based on an analysis of residuals, hat values, and Cook's distance, five country-year observations are particularly influential in predicting cabinet diversity. I remove all observations for these five countries — Vietnam, Mexico, Burkina Faso, Angola, and Kyrgyzstan — and re-run the main model without these countries (Model 1 of Table SI.12.3). The main results hold.

Next, we might wish to see if the effect is pronounced in places where power is not equally shared among ethnic groups. Subsetting the dataset to cases where *Power 01* is 0 (non-equal distribution of power), the main results hold (Model 2).

Subsetting by country or region of interest is difficult because of the low number of country observations per region. The confidence intervals on the interaction term in such subsets is extremely large, making it difficult to meaningfully determine how well the effect holds for certain countries of interest. One contributing factor is that similar countries on the *Non-Programmatic Distribution* and *School* measures are generally geographically proximate, making it difficult to achieve the country-level variation required for the interaction term to be estimated reliably. To give a sense of these difficulties, I run the main model on geographically defined country subsets: Africa, Europe, (North and South) America, and Asia. I also group North America, Australia and New Zealand, and Northern and Western

Europe together into "Western" countries. These results are shown in Models 3 through 7 in Table SI.12.3. There are between 17 and 42 countries per subset, making the confidence intervals very large and point estimates not precise. The direction of the interaction holds where there is enough statistical power and support. Because of these issues, the results are perhaps best interpreted as providing no evidence to contradict the main findings and other robustness checks.

One simple way to assess the amount of country-level variation in the results is to examine the sign and significance of country-level fixed effects in the main regression model. I extract the fixed effects, subset the data to only cases where the fixed effects are significant, and sort by magnitude. The model significantly over-predicts cabinet diversity in India, Vietnam, Mali and Senegal. The model significantly under-predicts cabinet diversity in a variety of African, European, and Asian countries. This indicates that the model is not differentially better at fitting some countries than others.

Table SI.12.3: Country Subsets

	Dependent variable:						
	Cabinet Diversity						
	No Influential	*Ethnic Issues*	*Africa*	*Europe*	*America*	*Asia*	*West*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Country ELF_{t-1}	0.083^{***}	0.085^{***}	0.102^{***}	0.803^{*}	-0.048	-0.502^{*}	0.436
	(0.025)	(0.025)	(0.028)	(0.461)	(0.089)	(0.300)	(1.555)
Non-Prog. Dis_{t-1}	0.287^{***}	0.355^{***}	0.685^{***}	-0.464	-0.007	0.119	4.644
	(0.060)	(0.063)	(0.084)	(0.362)	(0.290)	(0.153)	(3.317)
$School_{t-1}$	0.012^{*}	0.020^{***}	0.095^{***}	-0.037	-0.044	-0.013	0.001
	(0.007)	(0.008)	(0.014)	(0.023)	(0.035)	(0.014)	(0.036)
Non-Prog. Dist:School	-0.049^{***}	-0.061^{***}	-0.137^{***}	0.059	0.008	-0.019	-0.605
	(0.010)	(0.011)	(0.017)	(0.055)	(0.046)	(0.025)	(0.474)
Population	0.036^{***}	0.024^{**}	-0.059^{***}	0.183^{*}	0.247^{***}	0.011	-0.019
	(0.012)	(0.012)	(0.019)	(0.105)	(0.037)	(0.020)	(0.140)
Peace Years	-0.001	-0.002	-0.002	0.022^{***}	-0.012^{**}	0.011^{***}	-0.033
	(0.002)	(0.002)	(0.003)	(0.009)	(0.006)	(0.004)	(0.036)
Terrorist $Attacks_{t-1}$	-0.004^{**}	-0.004^{**}	0.0005	0.003	-0.008^{***}	-0.001	0.003
	(0.002)	(0.002)	(0.003)	(0.004)	(0.003)	(0.002)	(0.007)
Legislature $Size_{t-1}$	0.004	0.001	-0.0003	0.016	-0.013	0.008^{*}	-0.188^{**}
	(0.003)	(0.003)	(0.006)	(0.011)	(0.010)	(0.005)	(0.081)
$Coups_{t-1}$	0.018^{**}	0.020^{**}	0.026^{**}	0.006	0.013	0.014	
	(0.009)	(0.009)	(0.013)	(0.066)	(0.022)	(0.015)	
Leader Group $Size_{t-1}$	-0.004	0.001	-0.039	-0.022	0.043^{*}	-0.036	-0.025
	(0.014)	(0.015)	(0.034)	(0.038)	(0.023)	(0.047)	(0.039)
Coalition	0.006	0.006	0.018^{*}	-0.008	-0.008	0.012	-0.006
	(0.005)	(0.006)	(0.010)	(0.012)	(0.014)	(0.009)	(0.016)
Factional	0.003	0.001	0.017	0.0005	-0.00000	0.008	0.036
	(0.006)	(0.006)	(0.012)	(0.020)	(0.012)	(0.010)	(0.095)
$Majority_{t-1}$	-0.003	-0.001	-0.001	-0.022	0.013	-0.006	-0.081^{**}
	(0.003)	(0.004)	(0.006)	(0.014)	(0.009)	(0.005)	(0.034)
GDP_{t-1}	-0.011^{*}	-0.015^{**}	-0.023^{**}	-0.015	-0.034	0.007	0.128^{***}
	(0.007)	(0.007)	(0.011)	(0.022)	(0.025)	(0.010)	(0.044)
$Polity_{t-1}$	0.001^{**}	0.002^{***}	0.003^{***}	0.002^{*}	-0.001	-0.002^{*}	-0.063^{*}
	(0.0005)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.034)
$Size_{t-1}$	-0.0003^{**}	-0.0004^{**}	-0.001^{***}	-0.0004	-0.0005	-0.0003	0.0003
	(0.0002)	(0.0002)	(0.0003)	(0.0004)	(0.001)	(0.0002)	(0.001)
Observations	3,107	2,878	1,131	647	642	745	414
R^2	0.026	0.037	0.117	0.037	0.103	0.041	0.080
Adjusted R^2	-0.021	-0.013	0.070	-0.039	0.048	-0.027	0.005
F Statistic	4.931^{***}	6.480^{***}	8.844^{***}	1.427	4.329^{***}	1.843^{**}	2.206^{***}

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

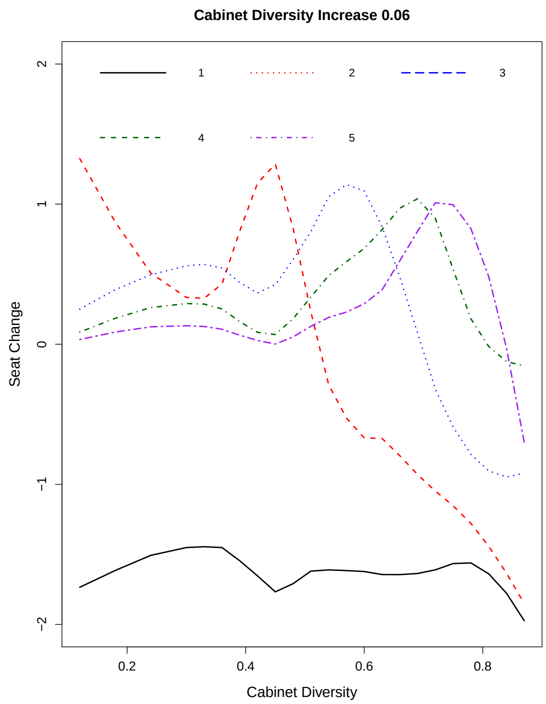
SI.13: Cabinet Seat Change Simulations

Figure SI.13.1 displays a panel for two increases in *Cabinet Diversity* (0.06 and 0.21). The x axis is the *Cabinet Diversity* score after the increase; that is, an x value of 0.42 indicates an increase in *Cabinet Diversity* from 0.36 to 0.42 in Panel A and from 0.21 to 0.42 in Panel B. The y axis represents the change in the number of seats held by each of the five largest ethnic groups in the cabinet. The five smaller groups in the simulation are not represented here to simplify the interpretation.

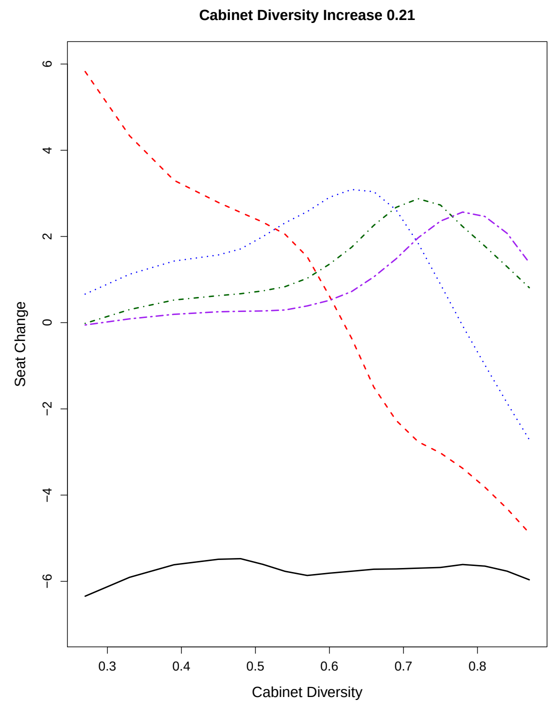
It is evident that the impact of changing *Cabinet Diversity* on seat share depends on the initial and final levels of *Cabinet Diversity.* In Panel A, when *Cabinet Diversity* is low and increases by 0.06, group 2 and group 3 split the two seats that group 1 loses. When *Cabinet Diversity* is high and increases, larger groups start losing seats to smaller groups. For example, when *Cabinet Diversity* increases to 0.80, groups 1, 2, and 3 lose seats and the smaller groups not shown on the plot begin gaining seats. The net seat change of different combinations of a *Cabinet Diversity* increase of 0.06 is between two and three seats.

Similar interpretations are possible with a *Cabinet Diversity* increase of 0.21. The seat change across this range is between six and seven seats. Importantly, in both the 0.06 and 0.21 increase of *Cabinet Diversity* new groups without previous representation often entered into the cabinet. Further, even with a relatively small change of 0.06, at least two ministers of the majority group were replaced by minority group ministers. I argue that these changes in cabinet composition are significant; they represent a change in between seven and twenty-five percent of the cabinet just due to an increase in *Non-Programmatic Distribution.*

Figure SI.13.1: Shifts in Cabinet Diversity



(a) Increase of 0.06



(b) Increase of 0.21

Each line represents one of the five largest ethnic group's change in cabinet membership for the given increase in *Cabinet Diversity* conditional on the level of *Cabinet Diversity*.

References

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat and Romain Wacziarg. 2003. "Fractionalization." *Journal of Economic Growth* 8(2):155–194.

Ambekar, Anurag, Charles Ward, Jahangir Mohammed, Swapna Male and Steven Skiena. 2009. Name-Ethnicity Classification from Open Sources. In *15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM pp. 49–58.

Angrist, Joshua D. and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, New Jersey: Princeton University Press.

Annett, Anthony. 2001. "Social Fractionalization, Political Instability, and the Size of Government." *IMF Staff Papers* 48(3):561–592.

Arriola, Leonardo R. 2009. "Patronage and Political Stability in Africa." *Comparative Political Studies* 42(10):1339–1362.

Back, Hanna, Thomas Persson, Kare Vernby and Helena Wockelberg. 2008. In Tranquil Waters: Swedish Cabinet Ministers in the Postwar Era. In *The Selection of Ministers in Europe.* London: Routledge.

Banks, Arthur S. and Kenneth A. Wilson. 2016. "Cross-National Time-Series Data Archive." https://www.cntsdata.com.

Bertrand, Marianne and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *The American Economic Review* 94(4):23.

Blondel, Jean and Jean-Louis Thiebault. 1988. "The Study of Western European Cabinets." *European Journal of Political Research* 16(2):115–123.

Cederman, Lars-Erik, Andreas Wimmer and Brian Min. 2010. "Why Do Ethnic Groups Rebel? New Data and Analysis." *World Politics* 62(01):87–119.

Cederman, Lars-Erik and Luc Girardin. 2007. "Beyond Fractionalization: Mapping Ethnicity onto Nationalist Insurgencies." *American Political Science Review* 101(1):173–185.

Chandra, Kanchan, ed. 2012. *Constructivist Theories of Ethnic Politics*. New York: Oxford University Press.

Chang, Jonathan, Itamar Rosenn, Lars Backstrom and Cameron Marlow. 2010. ePluribus: Ethnicity on Social Networks. In *4th International AAAI Conference on Weblogs and Social Media*. Association for the Advancement of Artificial Intelligence pp. 18–25.

De Boef, Suzanna and Luke Keele. 2008. "Taking Time Seriously." *American Journal of Political Science* 52(1):184–200.

Escobar-Lemmon, Maria and Michelle M. Taylor-Robinson. 2005. "Women Ministers in Latin American Government: When, Where, and Why?" *American Journal of Political Science* 49(4):829–844.

Fearon, James D., Kimuli Kasara and David D. Laitin. 2007. "Ethnic Minority Rule and Civil War Onset." *American Political Science Review* 101(1):187–193.

Francois, Patrick, Ilia Rainer and Francesco Trebbi. 2015. "How Is Power Shared in Africa?" *Econometrica* 83(2):465–503.

Guryan, Jonathan. 2001. Desegregation and Black Dropout Rates. Working Paper 8345 National Bureau of Economic Research Cambridge, MA: .

Htun, Mala. 2004. "Is Gender like Ethnicity? The Political Representation of Identity Groups." *Perspectives on Politics* 2(3):439–458.

Hunter, Lance Y., David J. Bennett and Joseph W. Robbins. 2018. "Destabilizing Effects of Terrorism on Party System Stability." *Terrorism and Political Violence* 30(3):503–523.

Jacob, Suraj, John A. Scherpereel and Melinda Adams. 2014. "Gender Norms and Women's Political Representation: A Global Analysis of Cabinets, 1979-2009." *Governance* 27(2):321–345.

Judson, Ruth A. and Ann L. Owen. 1999. "Estimating Dynamic Panel Data Models: A Guide for Macroeconomists." *Economics Letters* 65:9–15.

Kroeger, Alex. 2017. "Dominant Party Rule, Elections, and Cabinet Instability in African Autocracies.".

Krook, Mona Lena and Diana Z. O'Brien. 2012. "All the President's Men? The Appointment of Female Cabinet Ministers Worldwide." *The Journal of Politics* 74(3):840–855.

LaFree, Gary and Laura Dugan. 2007. "Introducing the Global Terrorism Database." *Terrorism and Political Violence* 19(2):181–204.

Mateos, Pablo. 2007. "A Review of Name-Based Ethnicity Classification Methods and Their Potential in Population Studies." *Population, Space and Place* 13(4):243–263.

Mateos, Pablo, Richard Webber and P. A. Longley. 2007. The Cultural, Ethnic and Linguistic Classification of Populations and Neighbourhoods Using Personal Names. Technical Report 116 University College London London: .

Mehler, Andreas. 2009. "Peace and Power Sharing in Africa: A Not so Obvious Relationship." *African Affairs* 108(432):453–473.

Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49(6):1417–1426.

Nymoen, Ragnar. 2004. "Dynamic Models.".

O'Brien, Diana Z., Matthew Mendez, Jordan Carr Peterson and Jihyun Shin. 2015. "Letting Down the Ladder or Shutting the Door: Female Prime Ministers, Party Leaders, and Cabinet Ministers." *Politics & Gender* 11(4):689–717.

Opalo, Kennedy O. 2011. "Ethnicity and Elite Coalitions: The Origins of Big Man Presidentialism in Africa.".

Peek, Lori A. and Jeannette N. Sutton. 2003. "An Exploratory Comparison of Disasters, Riots and Terrorist Acts." *Disasters* 27(4):319–335.

Pool, Veronika K., Noah Stoffman and Scott E. Yonker. 2012. "The People in Your Neighborhood: Social Interactions and Mutual Fund Portfolio Choice.".

Roessler, Philip. 2016. *Ethnic Politics and State Power in Africa: The Logic of the Coup-Civil War Trap*. Cambridge: Cambridge University Press.

Torvik, Vetle I. and Sneha Agarwal. 2016. Ethnea–an Instance-Based Ethnicity Classifier Based on Geo-Coded Author Names in a Large-Scale Bibliographic Database. In *International Symposium on Science of Science*. Washington, DC: .

Treeratpituk, Pucktada and C. Lee Giles. 2012. Name-Ethnicity Classification and Ethnicity-Sensitive Name Matching. In *26th AAAI Conference on Artificial Intelligence*. Association for the Advancement of Artificial Intelligence.

Vogt, Manuel, Nils-Christian Bormann, Seraina Ruegger, Lars-Erik Cederman, Philipp Hunziker and Luc Girardin. 2015. "Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Data Set Family." *Journal of Conflict Resolution* 59(7):1327–1342.

Williams, Laron K. and Guy D. Whitten. 2012. "But Wait, There's More! Maximizing Substantive Inferences from TSCS Models." *The Journal of Politics* 74(3):685–693.

Ye, Junting, Shuchu Han, Yifan Hu, Baris Coskun, Meizhu Liu, Hong Qin and Steven Skiena. 2017. Nationality Classification Using Name Embeddings. In *2017 ACM Conference on Information and Knowledge Management*. Washington, D.C.: ACM pp. 1897–1906.