Measuring politically-relevant identity, with and without groups

Kyle L. Marquardt

UNIVERSITY OF GOTHENBURG
DEPT OF POLITICAL SCIENCE

Varieties of Democracy (V-Dem) is a new approach to conceptualization and measurement of democracy. The headquarters – the V-Dem Institute – is based at the University of Gothenburg with 23 staff. The project includes a worldwide team with 5 Principal Investigators, 19 Project Managers, 33 Regional Managers, 134 Country Coordinators, Research Assistants, and 3,500 Country Experts. The V-Dem project is one of the largest ever social science research-oriented data collection programs.

Please address comments and/or queries for information to:

V-Dem Institute

Department of Political Science

University of Gothenburg

Sprängkullsgatan 19, Box 711

405 30 Gothenburg

Sweden

E-mail: contact@v-dem.net

Measuring politically-relevant identity, with and without groups[*]

Kyle L. Marquardt

Department of Politics and Governance

International Center for the Study of Institutions and Development

National Research University Higher School of Economics

Abstract

Quantitative scholarship on civil conflict still largely relies upon the ethnic group as the foundation for measures of politically-relevant diversity and, in particular, identity-based political inclusion. However, ethnicity remains notoriously difficult to measure: even cutting-edge analyses are subject to the issues of intra- and inter-ethnic variation in identity salience that plagued earlier work. Here I propose a new way to measure identity-based exclusion. Specifically, I use latent variable models to combine data from both the Ethnic Power Relations Project, which uses the demographic size of politically-relevant ethnic groups to operationalize inclusion; and the Varieties of Democracy Project, which measures overall identity-based inclusion without directly accounting for demographic group size. The latent variable models combine insights from both measurement approaches, ameliorating concerns about using either strategy in isolation. In addition to providing cross-nationally cohesive data on identity-based exclusion for future work, these models provide a framework for scholars to build their own theoretically-driven models of politically-relevant diversity and inclusion.

A wide body of literature links the exclusion of politically-relevant ethnic groups to important social-scientific outcomes, particularly civil conflict (Cederman, Wimmer & Min 2010, Cederman, Gleditsch & Buhaug 2013, Cederman, Hug, Schädel & Wucherpfennig 2015). In this tradition, scholars first enumerate the universe of politically-relevant ethnic groups present at a given level of analysis, then determine the degree and form in which these groups are excluded or included. However, as Kanchan Chandra (2006, 2012) has detailed, traditional conceptualizations of ethnic identity yield contradictory criteria for membership in an ethnic group, and belie significant intra- and inter-ethnic variation in the attributes that make some identities more salient than others. Exclusion itself may vary substantially within and across groups, often in correlation with the salience of identity. Measures of exclusion based on ethnic group demographics may therefore have difficulty accounting for the actual level of identity-based exclusion present at any given level of analysis.

An alternative approach to measuring identity-based exclusion involves coding its overall prevalence within a society, without first enumerating the groups it concerns. Such an approach avoids making assumptions about the specific social identities relevant to identity-based inclusion, as well as the degree to which members of these identity groups face similar levels of inclusion. By avoiding these assumptions, such data may 1) better capture the extent of exclusion within and across groups and 2) be more sensitive to changes on these metrices. However, such an approach has its own disadvantages. First, the demographic strength of cohesive groups is important in determining whether or not exclusion leads to outcomes of interest. By not explicitly incorporating group demographics into the measurement process, this approach cannot account for this important element of identity politics. Second, the absence of clear anchoring on ethnic demographics exacerbates concerns about cross-national comparability in the coding of this phenomenon.

In this paper, I use Bayesian latent variable modeling techniques to combine these two approaches, leveraging the strengths of both to provide longitudinally and cross-nationally comparable measures of identity-based exclusion. To do so, I first conceptually and empirically illustrate the advantages and disadvantages of both existing approaches, using the current gold standards in the literature: Ethnic Power Relations (EPR) and Varieties of Democracy (V–Dem). The EPR Project measures the demographic size of relevant identity groups in a territory, and explicitly focuses on ethnic groups. In contrast, the V–Dem data collection process does not involve enumerating groups. Instead, expert coders use Likert scales to report the general level of identity-based inclusion; these data are then aggregated using a Bayesian Ordinal Item Response Theory model (Pemstein et al. 2018). As expected, I find that V–Dem and EPR data strongly correlate and generally reveal similar cross-national and within-country trends in identity-based inclusion. However, the analyses also reveal the V–Dem data provide nuanced information about change

within countries, while EPR data provide rougher—but potentially more cross-nationally comparable—information.

I build on insights from these analyses to develop the two latent variable models. The first model directly combines the EPR and V–Dem data on elite-level political inclusion, essentially using the cross-nationally demographically-consistent EPR data to increase the cross-national comparability of the the V–Dem data. The second model expands on this analysis to incorporate data on the ethnic fractionalization and polarization of different country-year observations, as well as the day-to-day discrimination against group members. The end result of these analyses are two cross-nationally comparable and nuanced indicators of this vital concept, both of which evince high levels of content and construct validity in the framework of Adcock & Collier (2001). The models I deploy in this analysis are also readily modifiable, providing scholars the opportunity to build on them as new data become available, or to better fit their theories of identity politics.

1 Measuring identity with groups

Many measures of identity-based diversity and exclusion—particularly in the sphere of conflict research—use the ethnic group as the foundation for their measures of latent identity-based conflict potential (Fearon & Laitin 2003, Montalvo & Reynal-Querol 2005, Esteban, Mayoral & Ray 2012, Vogt, Bormann, Rüegger, Cederman, Hunziker & Girardin 2015). There are important theoretical reasons to do so. For example, researchers associated with the prominent Ethnic Power Relations (EPR) project argue that measuring ethnic groups is important "because the nation-state itself relies on ethnonational principles of political legitimacy." As a result, members of ethnic groups that have power in a society have access to material and symbolic resources denied to members of other groups (Wimmer, Cederman & Min 2009, 321). Members of excluded groups may resort to violence or other forms of conflict to gain access to these resources. Demographics are thus of clear importance in this context: countries and regions in which a high proportion of the population belongs to an excluded group(s) are those most likely to outcomes like civil conflict.

Despite the clear theoretical connection between ethnic groups and political power—and thus ethnic demographics and outcomes of social-scientific interest—there is substantial disagreement about what constitutes ethnic identity, not to mention how best to operationalize this concept cross-nationally. These disagreements indicate that measures of exclusion based on ethnic demographics may not always accurately reflect the level of exclusion within a society, creating problems for analyses that use these data on the right-hand side.

1.1 Heterogeneity between and within groups

A fundamental assumption of datasets that build on ethnic groups is that co-ethnicity imbues all ethnic groups and members of these groups with some degree of common identity that allows them to act collectively. This assumption is problematic: not all ethnic groups are equally cohesive, and not all members of ethnic groups equally identify with the group (Chandra 2006, Chandra 2012). A technique for dealing with this problem is to focus on "politically-relevant" ethnic groups, or groups that evince a sufficient degree of common identity to collectively engage in political behavior (Posner 2004, Wimmer, Cederman & Min 2009, Cederman, Wimmer & Min 2010, Vogt et al. 2015), in essence rendering them politically exchangeable. However, politically relevant groups likely vary in the content of their identity, and members of these groups vary in their attachment to the group. For example, a group defined by inherited physical traits may have a different level of cohesion than a group defined by religious preferences. At the individual-level, members of the religiously-defined group who are not themselves religious may have different political preferences than members who are. In other words, while both groups may have acted collectively to affect politics in the past, there is no guarantee that they will behave similarly—or be statistically exchangeable—at any point in time other than that in which they were engaging in political behavior.

Scholars who use data sets that include politically relevant ethnic groups are aware of this problem, and attempt to mitigate it in a variety of ways. Specifically, cutting-edge datasets incorporate a variety of characteristics of both groups and group members into their analyses. For example, some analyses incorporate the degree to which these dimensions are cross-cutting or shared by members of the group in question (Selway 2011, Bormann, Cederman & Vogt 2015), or weight the groups by different characteristics. However, if traits other than politically-relevant ethnic identity are of importance to political mobilization, a focus on the ethnic group risks 1) artificially dividing populations that face similar forms of inclusion and 2) ignoring members of other groups that may share these traits and thus have similar political proclivities as the members of the delineated groups. For example, consider a case in which there are two equally-sized ethnic groups (A and B) and a religious cleavage (1 and 2) that divides the population into a religious majority and minority: the identities A1, B1 each represent 33% of the population; the identities A2 and B2 each represent 17% of the population. Group A is represented in the executive branch of the government, and thus both A1 and A2 are politically included by the standards of the EPR project; Group B (B1 and B2) has no representation. However, the religious cleavage is in fact the identity characteristic most salient for representation, with Group 1 enjoying the most access to power. Members of the included ethnic group who are of the stigmatized religious group (A2) enjoy no access to government-distributed material resources, while members of the excluded ethnic group who are of

the high-status religious group (B1) have the same access as their co-religionists (A1). In principle, weighting the distance between these groups by religion would add nuance to the largely-misleading ethnic story, but would still be misleading because religion—not ethnicity weighted by religion—is the relevant identity attribute for exclusion.[1]

Equally importantly, these data sets generally assume that the traits they include as covariates related to the group are time-invariant. For example, if a group is linguistically distinct from a polity's dominant group at time t, these datasets generally assume that the group remains linguistically distinct at time $t+50$. Given that processes of linguistic assimilation are relatively common, this assumption is problematic: if linguistic differences lead to exclusion, a group that linguistically assimilates over the course of 50 years will no longer face exclusion.

1.2 Selection of ethnic groups

Scholars constructing datasets that use ethnic groups as their basis must first decide what constitutes an ethnic group. As a large body of work illustrates, there is no clear consensus on either what constitutes an ethnic group or how to consistently measure these groups cross-nationally (Fearon 2003, Chandra 2006, Chandra 2012, Marquardt & Herrera 2015, Hale 2017). Moreover, all identity groups are subject to constant contestation and change (Abdelal, Herrera, Johnston & McDermott 2009, Brubaker 2002).

On one level, this concern is about the inferences researchers can draw from large scale group-based datasets: as Jenne and Bochsler note, "quantitative researchers must bracket questions about the origins, functions or boundaries of of group identies in order to create large-N databases" (Bochsler, Green, Jenne, Mylonas & Wimmer 2021). On another level, cross-national measures of identity-based exclusion that use the ethnic group as their foundation risk misrepresenting the actual level of identity-based exclusion within a society. To build on Fearon's classic example of measuring ethnicity in Somalia (2003), if clan membership provides the basis for inclusion in a country, but clans are not considered "ethnic," then a country with widespread clan-based exclusion would be incorrectly coded as having no identity-based exclusion. As Marquardt & Herrera (2015) note, different conceptualizations of what constitutes "ethnic" can result in drastically different enumerations and thus drastically different country-level statistics.

Adding a temporal component to this argument makes the problem even more difficult. Since the content of ethnic identities can change over time, formerly excluded groups can become included and groups can merge and disaggregate. These processes can also occur concurrently: Irish Americans became "White" as their political power (i.e. level of

---

[1]In this context, religious identity should perhaps determine the relevant ethnic groups in this society. However, this line of reasoning makes analyzing the causal role of identity on any political outcome impossible: if inclusion or exclusion is the primary criterion for identifying ethnicity, then any attempts to investigate the relationship between ethnicity and inclusion are wholly circular.

inclusion) increased (Ignatiev 1995). While an appropriately fine-grained dataset can model these changes, determining precisely when these changes occurred is difficult. Is "White" more politically relevant than "Irish" for all Irish Americans in all political spheres, and in what year did it become so?

Datasets that use politically-relevant ethnic groups face an additional issue: selection bias (Hug 2003, Hug 2013, Birnir, Wilkenfeld, Fearon, Laitin, Gurr, Brancati, Saideman, Pate & Hultquist 2014). Specifically, they explicitly select only identity groups that have become politically salient—not the universe of potential identity groups—and the process of ethnic group boundary making may itself be a function of conflict or exclusion (Brubaker 2002).[2]

2 Measuring identity without groups

An alternative approach for measuring socially relevant identity is to analyze the prevalence of identity-based inclusion in a society. Instead of first enumerating the ethnic groups in a state (politically relevant or irrelevant), then determining whether or not members of each group are excluded, a scholar could determine the degree to which social identity overall is linked to inclusion in a society. A reasonable implementation of this approach would solve—at least to some extent—the problems with measuring exclusion with ethnic groups, though it would raise issues of its own.

2.1 Advantages

2.1.1 Heterogeneity between and within groups

This approach makes no assumptions about the characteristics of groups being excluded or the degree to which exclusion is consistent across group members. As a result, it sidesteps many of these concerns with regard to measuring exclusion at the group level. However, this approach does rely on a strong assumption that the fact of inclusion renders all groups and individuals exchangeable in terms of their political proclivities. In principle, this assumption is problematic. If exclusion is easily remedied by some groups and members of groups, then the distribution of inclusion will potentially be unstable across time as formerly excluded groups (and members of groups) become included. For example, if assimilation is an option for some members of stigmatized groups, but not for others, then those who can assimilate may do so and thus exhibit different political preferences from those who cannot.

---

[2]Technically, even measures of ethnic groups that do not use political relevance as a criteria are subject to this bias, since there is an essentially infinite number of potential ethnic groups and, even using relatively lax criteria, only "relevant" groups are counted.

In practice, an appropriately granular time-series analysis can account for this problem to some extent: those groups and group members that can easily evade inclusion would no longer count toward overall polity-level inclusion, should they assimilate or otherwise cease to exist. Indeed, this potential to vary over time is an advantage of a focus on inclusion writ large, as opposed to the degree to which ethnic groups are included: group-based data sets assume that the political relevance of ethnic groups is largely fixed, and their traits largely constant. In contrast, a correctly-measured overall inclusion variable would account for changes in the relevance of the traits that lead to inclusion, as well as the size of the groups that exhibit these traits.

2.1.2 Selection of ethnic groups

Since there is no need to explicitly enumerate ethnic groups, measuring identity-based inclusion directly drastically ameliorates the concern of omitting possibly relevant populations. Similarly, measuring identity-based inclusion sidesteps the circular reasoning often involved in creating politically-relevant group data sets. As previously discussed, the political inclusion or exclusion of a group can increase the salience of a group's common identity, increasing the likelihood it is considered 1) a cohesive unit and 2) politically relevant. Processes of inclusion thus define the universe of groups, which in turn defines the degree to which inclusion occurs. By not explicity defining the universe of groups, directly measuring identity-based inclusion avoids this issue.

2.2 Disadvantages

2.2.1 Demographics matter

A primary disadvantage of measuring identity-based exclusion without groups is that groups—and demographics—matter. For example, a common ethnic identity facilitates collective action (Hale 2008). As a result, a society with great exclusion of a demographically small population may have lower odds of conflict onset than one with moderate exclusion of demographically large population. A measure of identity-based exclusion that does not account for this distinction would therefore miss a highly important aspect of identity politics.[3]

---

[3]It is worth noting that prominent scholarship argues that researchers should primarily investigate identity-based exclusion and exclusion-related outcomes at the group level—not the region- or country-level (Cederman, Gleditsch & Buhaug 2013). Among work that does analyze conflict at the region- or country-level, there is clear evidence that some measures of demographics correlate with the probability of conflict onset, though a vibrant discussion about the correct way to parameterize these demographics continues (Bochsler et al. 2021). While measures of fractionalization have become less popular, measures of polarization or demographic strength (e.g. the proportion of the population that is excluded, the proportion of the population belonging to the largest excluded group, the number of excluded groups) still enjoy widespread use.

2.2.2 Cross-national comparability

Accurately measuring levels of identity-based exclusion across time requires in-depth substantive and theoretical expertise. Coders measuring the extent of identity-based exclusion must 1) identify individuals likely excluded for identity-based characteristics, 2) identify forms of exclusion, 3) combine these two aspects to determine who is excluded and how, and 4) estimate the overall prevalence of 3) in a coding unit. Performing all four tasks requires great contextual and conceptual knowledge, essentially necessitating expert-coding. Equally importantly, it is unlikely that a single coder could accurately code more than several cases, regardless of their level of expertise. An expert on identity-based exclusion in Kazakhstan is likely unaware of what such exclusion looks like in Ethiopia and, in fact, might inaccurately measure this concept in Ethiopia by using their region-specific identity politics knowledge as a reference point. As a result, this measurement exercise likely requires different sets of experts coding different cases. As Marquardt (2020) details, such a scenario presents substantial concerns for cross-national comparability in the context of measuring identity-based exclusion: different experts may perceive scales differently—a phenomenon known as differential item functioning, or DIF—and this scale perception may systematically vary across countries/regions. As a result, comparing estimates from one country to another could be misleading.

3 The data

The previous discussion yields several expectations about how the two approaches to measuring identity-based exclusion and inclusion should compare to each other. Since both approaches are intended to measure the same concept, they should correlate. However, a group-based approach should provide rougher estimates of changes in the concept than the direct approach for two main reasons. First, the group-based approach is less able to track changes in the relative salience of different identity categories—as well as salience of exclusion itself—than the direct measurement approach. Second, the group-based approach assumes that changes in a group's status in elite politics would affect all group members equally, leading to jumps in levels of inclusion and exclusion as demographically-large groups' status changes.

To compare and contrast the two different approaches to measuring identity-based inclusion, I use measures from the Ethnic Power Relations (EPR) data set and the Varieties of Democracy (V–Dem) data set.

3.1 Group-based exclusion: EPR

The Ethnic Power Relations (EPR) data set uses regional experts to determine both the politically relevant ethnic groups in the society, as well as whether or not they are ex-

cluded; in cases of divergent expert opinions, additional experts are consulted (Cederman, Wimmer & Min 2010, Vogt et al. 2015). The resulting data thus consists of a list of politically relevant groups in a country, with additional data regarding the demographic size of the group and their location in the country's political constellation (i.e. the degree and manner in which they are included or excluded). For the purposes of EPR, political inclusion represents representation at the executive level of politics: if a member of a given group has an executive-level appointment (e.g. as president or prime minister, or cabinet member) then the entire group is coded as included. Otherwise, the group is excluded.[4]

EPR provides several aggregations of the group-year level data to the country-year level. In these analyses I focus on the proportion of the politically-included population over the entire population, since this measure is the conceptually closest to the V–Dem data.[5]

3.2   Exclusion without groups: V–Dem

The V–Dem project uses a network of over 3,000 experts to code a variety of variables related to democracy, including those related to identity group political inclusion (Coppedge, Gerring, Lindberg, Skaaning, Teorell et al. 2017a). Each expert is assigned one of 11 surveys related to their area of substantive expertise; each survey includes a number of questions with Likert-scale or numeric responses. Generally five or more experts code each country; V–Dem policy is to have a majority of local experts code country years except when impossible (e.g. for countries like North Korea (Coppedge, Gerring, Lindberg, Skaaning, Teorell et al. 2017b)). Given the previous discussion about the importance of deep country-knowledge for measuring identity-based exclusion, the presence of many local experts is clearly essential.

To analyze the specific indicator of political inclusion, 1,100 experts provided responses to the question "Is political power distributed according to social groups?" (Figure 1). There are five possible ordinal responses, which correspond to situations ranging from institutionalized and monopolized minority control of the political system, to social identity being largely irrelevant to politics. The question further defines a social group as one defined by "caste, ethnicity, language, race, region, religion, or some combination thereof," and explicitly does not include socioeconomic status or sexual orientation. This definition captures the groups encapsulated by a broad conceptualization of ethnicity. However, it diverges sharply from ethnicity-based measures in that it does not attempt to measure the groups in question, but rather the degree to which political power is

---

[4]The EPR coding schema includes different subcategories of inclusion and exclusion, which I do not consider in this paper in the interest of simplicity.

[5]In countries where ethnicity is "irrelevant" to political power per the EPR coding, I code the entirety of the population as included.

Figure 1: V–Dem identity-based inclusion question

Question: Is political power distributed according to social groups?

Clarification: A social group is differentiated within a country by caste, ethnicity, language, race, region, religion, or some combination thereof. (It does not include identities grounded in sexual orientation or socioeconomic status.) Social group identity is contextually defined and is likely to vary across countries and through time. Social group identities are also likely to cross-cut, so that a given person could be defined in multiple ways, i.e., as part of multiple groups. Nonetheless, at any given point in time there are social groups within a society that are understood - by those residing within that society – to be different, in ways that may be politically relevant.

Responses:

0: Political power is monopolized by one social group comprising a minority of the population. This monopoly is institutionalized, i.e., not subject to frequent change.

1: Political power is monopolized by several social groups comprising a minority of the population. This monopoly is institutionalized, i.e., not subject to frequent change.

2: Political power is monopolized by several social groups comprising a majority of the population. This monopoly is institutionalized, i.e., not subject to frequent change.

3: Either all social groups possess some political power, with some groups having more power than others; or different social groups alternate in power, with one group controlling much of the political power for a period of time, followed by another – but all significant groups have a turn at the seat of power.

4: All social groups have roughly equal political power or there are no strong ethnic, caste, linguistic, racial, religious, or regional differences to speak of. Social group characteristics are not relevant to politics.

equally distributed among groups. In other words, it is directly measuring the degree to which identity-based cleavages are relevant to political power in the society, not the proportion of the population that is politically privileged over others due to their identity.

The broadness of this question is of special importance to the measurement of inclusion: local experts are likely attuned to the identity-based cleavages in their country, and this phrasing allows them to leeway to code exclusion based on these cleavages. While this leeway is helpful for rigorously assessing the level of exclusion in a country, it also increases concerns about cross-national comparability.

The V–Dem Project aggregates coder data using a Bayesian measurement model which takes into account both clustered and expert-specific DIF, though the sparsity of the data (i.e. generally six or fewer coders per observation and incomplete bridging in the form of experts coding either additional cases or anchoring vignettes) potentially limits the efficacy of the approach (Pemstein et al. 2018). In the following analyses I use the

point estimate from this model (the median over the estimate's posterior distribution) as the V–Dem estimate of identity-based political inclusion.

3.3 Descriptive comparison of EPR and V–Dem variables

Figure 2 presents a scatterplot illustrating the relationship between the EPR and V–Dem political inclusion variables across all country-years present in both data sets. Both variables are scaled such that higher values represent greater inclusion, e.g. a score of 1 in the EPR data set indicates that all citizens are included in elite politics, at least insofar as their identity group is concerned. This comparison reveals that while inclusion is correlated across the two data sets, there are significant differences between them. Perhaps most noticeably, there are many country-years which the EPR codes as having perfect inclusion that the V–Dem codes as having very low levels of inclusion. There are four potential main reasons for this difference. First, the V–Dem variables have a less strict criteria for the relevant identity groups included in the measurement. As a result, V–Dem coders may be coding the exclusion of groups that EPR coders do not consider politically relevant (e.g. tribes and clans in Somalia). Second, the V–Dem variable is multidimensional, incorporating both the country's rough demographic level of inclusion, as well as the intensity of the inclusion. As a result, there will be discrepancies between the two data sets if a relatively small proportion of a country's population faces extreme discrimination: EPR would code this case as largely inclusive due to the included groups' demographic dominance, while V–Dem experts would likely code it as having an intermediate or low level of inclusion due to the severity of the exclusion. Finally, highly exclusive societies that are also largely monoethnic (e.g. North Korea) are conceptually difficult for V–Dem experts to code: all North Koreans are equally *excluded* from politics. As a result, V–Dem experts can (and do) interpret such cases as being highly exclusive, whereas the EPR project would code it as highly inclusive because ethnicity is irrelevant to inclusion.

Table 1 investigates the cases with the largest discrepancies in detail, showing country-year observations in which either 1) the observation is in the lowest quantile of EPR observations and highest quantile of V–Dem observations (left column) or 2) the reverse (right column). It is worth noting that there are some countries which V–Dem and EPR coders universally consider to be of different types (highlighted in bold): for example, V–Dem coders consider North Korea, Oman, Qatar and Swaziland to be highly exclusive societies, while EPR considers them to be highly inclusive. These divergences all point to differences in how V–Dem experts and EPR coders conceptualize relevant groups. For example, Swaziland is likely a case of different perceptions of identity (EPR codes it as monoethnic, while V–Dem coders may perceive different groups as being relevant). Oman and Qatar are largely monoethnic countries in which large non-citizen populations
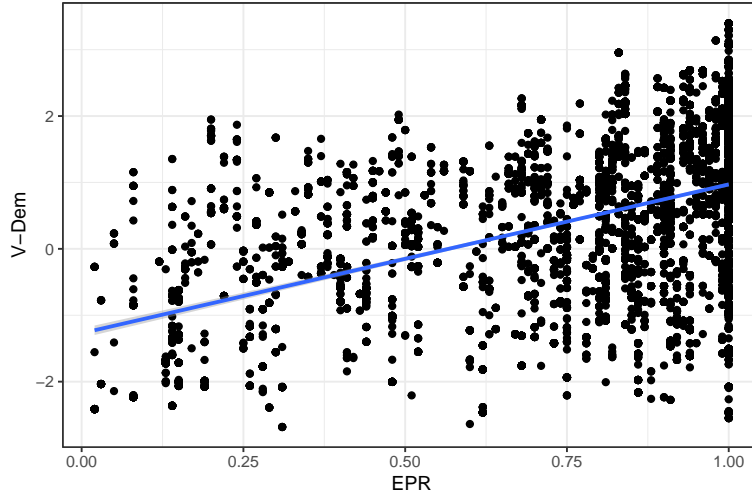
Figure 2: Comparison of measures of inclusion

Table 1: Country-years with large discrepancies between V–Dem and EPR coding

Low EPR, High V–Dem
Benin (1997-2006), Brazil (1989-1995, 2014-2016), Bhutan (2001-2017), Côte D'Ivoire (2001-2003), Comoros (1999-2002), Liberia (2006, 2014-2017), Nepal (2008), Sierra Leone (2003-2006)

High EPR, Low V–Dem
Burundi (1963-1966, 1990-1993), Haiti (1947-1990, 2014-2017), **North Korea** (1949-2017), **Oman** (1972-2017), **Qatar** (1972-2017), Singapore (2017), Somalia (1971-1992, 2013), **Swaziland** (1969-2017), Tunisia (1957-2011), UAE (1972-1994, 2006-2017), Venezuela (1950-1958), Yemen (1996-2011, 2015-2016)

are excluded; while EPR excludes non-citizens from their coding scheme, V–Dem coders likely incorporate these non-citizens into their coding.

Finally, Figure 3 provides a longitudinal analysis of V–Dem and EPR variables at the country-level from 1946-2017, using the BRICS countries and the United States of America as examples.[6] Two things are readily apparent. First, EPR codings vary to a much lesser extent within countries than do the V–Dem variables, evidence that the more fine-grained approach of V–Dem is better able to capture nuanced changes in the severity of exclusion than the blunt instrument of ethnic demographics. For example, EPR codes the United States as being relatively inclusive from the beginning of the time series to 2012 because Whites (the main included group) constituted a large demographic majority. V–Dem data reflect the gradual improvement in minority political power from a very low level in the 1940s to a relatively high level at present (with a decline in recent

---

[6]I convert V–Dem output to a 0-1 scale using the cumulative distribution function of the normal distribution.

Figure 3: Comparison of EPR and V–Dem data
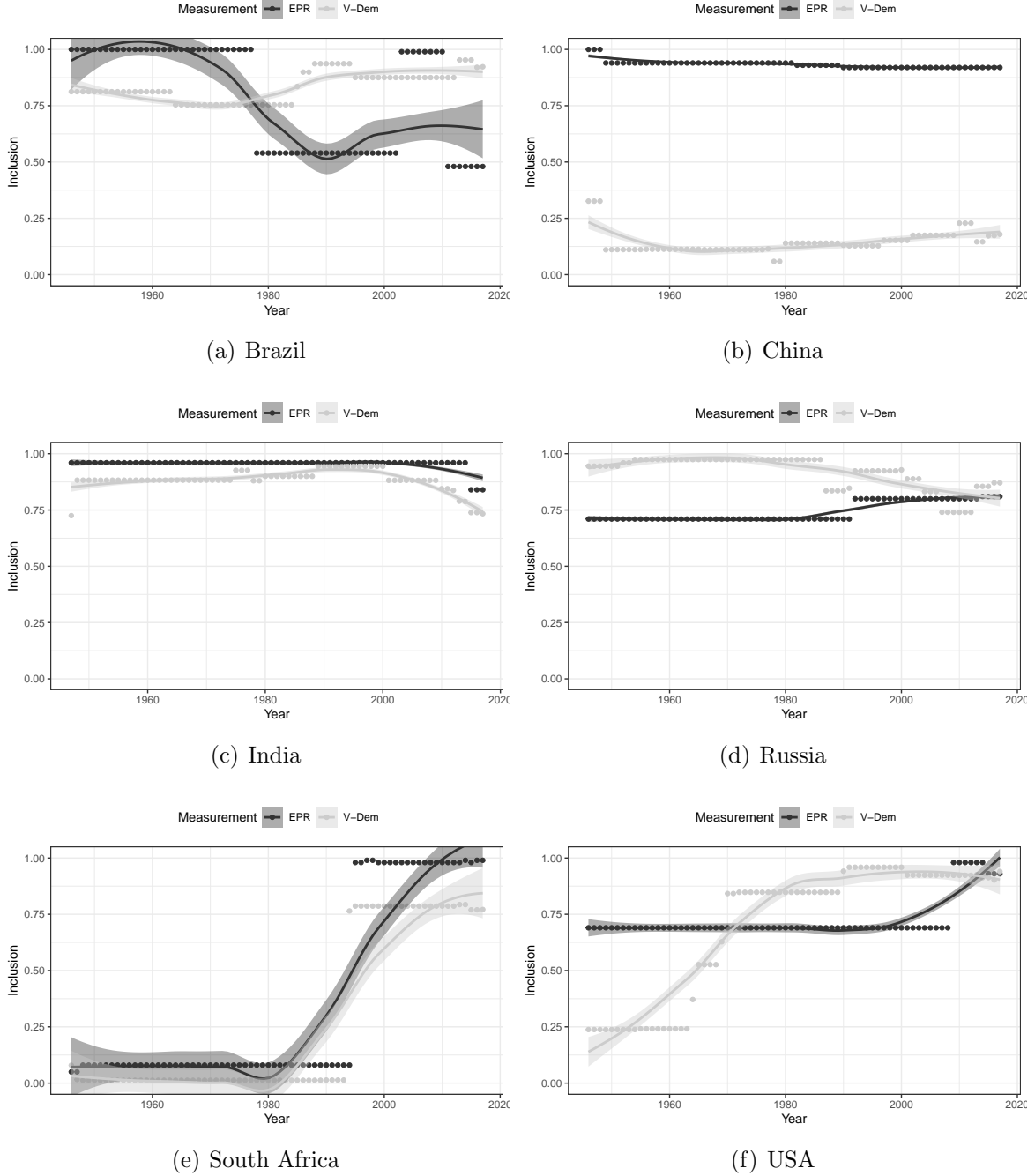
(a) Brazil

(b) China

(c) India

(d) Russia

(e) South Africa

(f) USA

Lines represent local regression estimates and corresponding uncertainty; they do not represent measurement uncertainty about estimates.

years). Similarly, while the election of Barack Obama to the post of president marks the inclusion of African Americans into politial power according to EPR coding criteria—and thus a marked increase in inclusion in the United States—the V–Dem data do not show a large *de facto* shift.

Second, in five cases (Brazil, India, Russia, South Africa and the United States), general trends are roughly similar between the V–Dem and EPR codings, though the V–Dem data tend to show more nuanced changes than the periods of stasis followed by rapid change observable in the EPR data, which are due to changes in the demographic composition of executive-level posts leading to swathes of the population changing from excluded to included (this phenomenon is particularly notable in Brazil). In the final case (China), there are clear differences in coding criteria between EPR and V–Dem. While EPR codes China as being largely inclusive (likely due to the demographic dominance of the included group, ethnic Han Chinese), V–Dem codes it as being exclusive, due to the fact that minority groups have little political representation in national politics in China.

4 Quantitative comparison of EPR and V–Dem variables

To further analyze the relationship between different methods of measuring identity-based inclusion, I conduct exploratory regression analyses comparing the different variables. More specifically, I regress both the V–Dem and EPR variables on their corollaries in the other dataset (e.g. I regress the V–Dem measure of social inclusion on the EPR measure of this concept, and vice versa). In this context, causal claims are clearly unwarranted. Instead, the goal here is to determine how different measures correlate, indicating conceptual convergence. Along those lines, to assess the extent to which other factors unrelated to identity may influence the measurement of identity-based inclusion, I also examine the relationship between these measures and other political and demographic variables.

Specifically, I include measures of ethnic fractionalization and polarization which I estimated using group-level data from the CREG dataset (*The Composition of Religious and Ethnic Groups (CREG) Project* 2014). These measures provide insight into the degree to which 1) the ecological presence of identity-based diversity and 2) the form of this diversity influence the estimation of both the demographic presence of included groups in the case of EPR data, or general inclusion in the case of V–Dem data. I use the standard Herfindahl index to estimate country-level fractionalization; fractionalization thus represents the odds that two randomly-selected individuals in a country would be from different groups. Polarization in this context represents the degree to which a country's demographic situation differs from one in which there are two equally-sized groups (a score of one, with a score of zero representing either a perfectly fractionalized

or monoethnic society).

In addition to the variables related to ethnic identity, I also control for several other factors that may influence coders' perceptions of inclusion. First, measures of social inclusion may be proxies for the degree to which a state generally respects its citizens' rights. All analyses therefore include the V–Dem civil liberties index (Coppedge et al. 2017*a*), which measures the overall degree to which a state respects the civil liberties of its subjects. Second, it is possible that measures of identity-based political inclusion proxy general political inclusion. I therefore control of overall political inclusion using the V–Dem measure Polyarchy, which is an aggregate measure representing the degree to which a polity has achieved the ideal of electoral democracy (Teorell, Coppedge, Skaaning & Lindberg 2016). Finally, I control for both population and GPD per capita, using data from the Clio-Infra Project and the Maddison Project, respectively (Clio-Infra 2013, The Maddison Project 2013). I also include year effects to control for general trends over time with regard to inclusion.

All models use standard Bayesian linear regression, implemented with the statistical program Stan (Stan Development Team 2018). Figure 4 presents the results of these analyses, showing the posterior-predicted effect of changing from a low to high value of given variable (i.e. a value under which 2.5 percent of the observations of the variable lie to 97.5 percent), holding all other variables constant at their median. Points represent the posterior median estimate, while horizontal lines represent 95% credible regions, a Bayesian corollary of confidence intervals. The horizontal axis scale represents the scale of the variable in question, or more precisely the variable's 95% density range. "EPR" and "V–Dem" represent the corollary of a given variable from the other dataset.

The EPR and V–Dem variables strongly correlate with each other, as expected given that they purport to measure the same concepts. In the case of the EPR variable, the V–Dem counterpart is the substantively strongest counterpart. However, the strongest correlates of V–Dem variable are civil liberties and polyarchy, which indicates that V–Dem coders are taking into account the overall level of inclusion in a polity when coding. Interestingly, while identity-based polarization tends to be positively correlated with the EPR variable, and fractionalization negatively correlated with these measures; the opposite is true for the V–Dem variable. That is, more polarized societies tend to have higher levels of inclusion according to the EPR data, while they tend to have lower levels of inclusion according to V–Dem. This result perhaps reflects the role of demographics in the EPR data set: more groups mean that fewer groups are able to be included in elite-level politics.

Figure 4: Correlates of EPR and V–Dem data
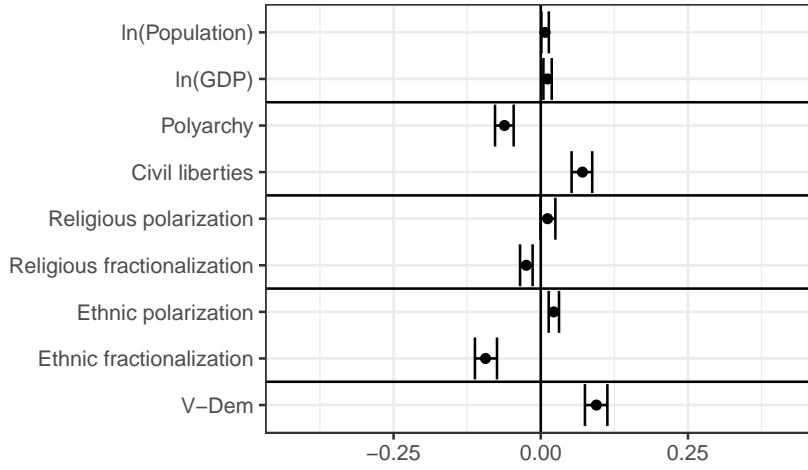
(a) EPR
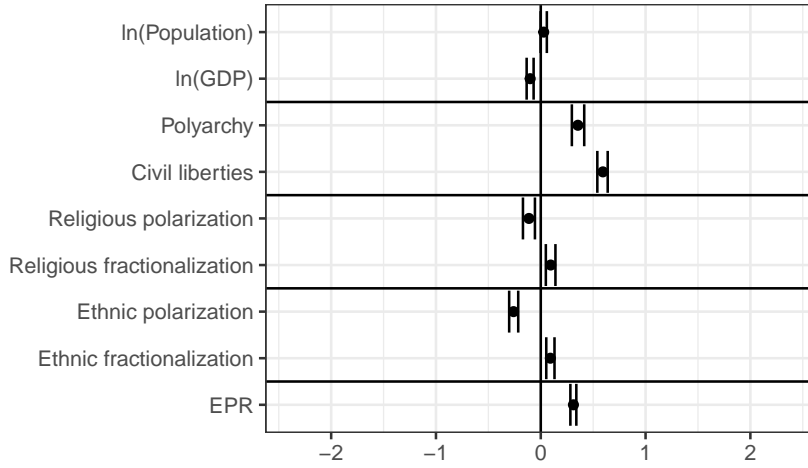
(b) V–Dem

Posterior-predicted effect of changing from low to high value of variable. Points represent posterior-median estimate and lines 95% credible regions.

5 Aggregating V–Dem and EPR data

The analyses I have discussed thus far indicate that measures of politically-relevant identity that use the ethnic group as their unit of measurement (EPR) correlate strongly with those that measure inclusion without groups (V–Dem). However, the relationship between these measures also diverge in substantively important ways that cumulatively indicate that measuring politically-relevant identity requires elements of both. A potential way to square this methodological circle would be to combine the measures using latent variable models.

In recent years latent variable models have been used to measure a variety of important social-scientific phenomena. In addition to extensive applications in the aggregation of expert-coded data; prominent scholarship has used this methodology to measure concepts including democracy (Treier & Jackman 2008, Pemstein, Meserve & Melton 2010), media freedom (Solis & Waggoner 2020), power consolidation (Gandhi & Sumner 2020), and accountability (Lührmann, Marquardt & Mechkova 2020). While some work has discussed the applications of these models to create a "common space" between latent measures of party and public ideology (Bakker, Jolly, Polk & Poole 2014), as far as I am aware this project is the first to extensively use external data (EPR) to bridge an expert-coded data set at the coder level (V–Dem).

Here I provide two different models for doing so. In the first, I use a relatively simple latent variable model to aggregate the V–Dem and EPR data on political inclusion. In essence, this strategy is similar to the standard V–Dem measurement model, which treats every coder as having idiosyncratic reliability and scale perception parameters that weights their contribution to the estimation of the latent concept—in this case, the level of identity-based inclusion in a country year. In this simple model, the EPR variable becomes another coder. However, unlike the V–Dem experts who only code several countries at a maximum, the EPR "coder" codes every single observation in the data set, providing the model with much more data to assess its reliability and scale perception. In doing so, the model would ideally use the EPR data to correct for cross-national differences in scale perception on the part of V–Dem experts, shifting their thresholds to be more in line with the EPR levels based on demographics. As a result, the data would be intrinsically more cross-nationally compatible, but also have an anchor in demographics (which is important in-and-of itself).

The second modeling strategy extends the first, incorporating not just political inclusion data, but also data from V–Dem and EPR on social inclusion — the degree to which residents of a country face identity-based discrimination. Both of these concepts are treated as nodes in a hierarchical latent variable model, where data on overall ethnic and religious diversity provide a prior estimate of exclusion in a society. While this approach is much more complicated than the first model, the greater complexity results in

much more cross-nationally comparable and detailed data that are further aligned with ethnic demographics.[7]

I discuss these approaches in turn. Note that I only estimate values for years in which there are values from both datasets (1946-2017).

5.1 Identity-based inclusion as a latent variable

Figure 5 presents the conceptual framework for the latent variable that aggregates the V–Dem and EPR inclusion variables (Stan code available upon request). Here the circle represents the latent variable being estimated, and squares manifest variables. I represent the V–Dem inclusion variable as an oval to indicate that it represents the scores of many experts.

I enter the expert codings into the model using a modified version of the V–Dem measurement model, which is itself a modified Ordinal Item-Response Theory (IRT) model. Equation 1 presents the partial likelihood for this model.

Figure 5: Latent variable model for identity-based political inclusion
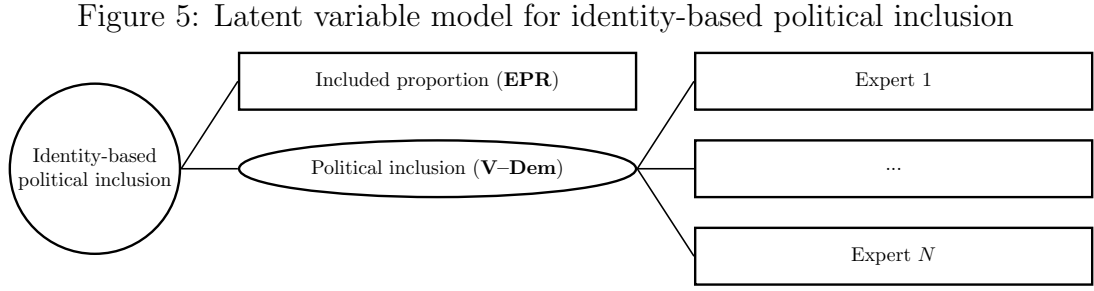
$$\Pr(y_{ctr} = k) = \Phi\left(\tau_{r,k} - \xi_{ct}\beta_r\right) - \Phi\left(\tau_{r,k-1} - \xi_{ct}\beta_r\right) \tag{1}$$

y represents the ordinal coding (values $1, ..., 5$) which expert r provides for country-year ct; ξ is the latent value for this country-year, *a priori* distributed according to a standard normal distribution. Each expert r has a unique reliability parameter $\beta \sim N(1, 1)$, restricted to positive values. This parameter represents the inverse of each expert's stochastic error variance, and in essence weights an expert's contribution to the measurement process based on the degree to which they covary with other experts, conditional on their thresholds (τ). Experts who diverge to a greater extent from other experts are penalized and thus contribute less to the measurement process. The model accounts for DIF—idiosyncratic expert scale perceptions—through τ, $k - 1$ threshold values that are specific to expert r. Thresholds determine the value over which ξ must lie in order for an expert to provide a given scale item over the next lowest. Since τ

---

[7]Insofar as I am aware, this measurement strategy is also the first to pursue a hierarchical modeling strategy at the expert-coding level.

17

varies by experts, it accounts to some extent for variation in scale perception.[8] For computational reasons, I cluster thresholds about universal values, i.e. $\tau_{r,k} \sim N(\gamma_k, .5)$, where $\gamma_k \sim Cauchy(0,2)$.

I add EPR data on inclusion to the model in a very similar fashion, as illustrated by Equation 2:

$$\Pr(y_{ctr} = m) = \Phi\left(\kappa_m - \xi_{ct}\zeta\right) - \Phi\left(\kappa_{m-1} - \xi_{ct}\zeta\right) \tag{2}$$

Specifically, I convert EPR data on inclusion into an ordinal variable based on the proportion of a country's population in a given year that is politically included, with cutpoints at each .125 of the population.[9] In this framework, the EPR data can be conceptualized as an additional coder, who has their own idiosyncratic reliability, $\zeta \sim N(1,1)$, and idiosyncratic thresholds, $\kappa \sim Cauchy(0,2)$. The key difference between the EPR "coder" and a standard V–Dem expert is that the EPR data cover all observations, while V–Dem experts generally code only 1-2 countries.[10] As a result, the model is able to estimate the EPR thresholds and reliability much more precisely than any of the individual V–Dem experts, allowing the EPR data to bridge all observations.

In principle, this approach will allow the EPR data to weight estimates of identity-based political inclusion in a society by bringing the idiosyncratic thresholds of V–Dem experts into alignment with the EPR thresholds. For example, assuming that a V–Dem expert covaries their codings in line with the EPR data, but systematically codes levels of inclusion to be lower than would be suggested by their position relative to the EPR data, the model will adjust their thresholds to be higher than they would be in the absence of the EPR data.

As an additional note, the greater density of V–Dem coders will likely overpower EPR data in terms of within-country trends: e.g. if all V–Dem experts for a country code levels

---

[8]V–Dem uses three different forms of data to gain leverage on idiosyncratic DIF: 1) anchoring vignettes (King & Wand 2007, Pemstein et al. 2018), in which experts code hypothetical scenarios that require no country-specific expertise: 2) asking experts to code multiple years for another country in addition to their main country of focus; and 3) having experts code multiple countries for a single year-observation. The third strategy has documented issues with inducing "jumps" in the data and I therefore remove these codings in this analysis, instead allowing the EPR codings to facilitate bridging. Similarly, while V–Dem proper clusters thresholds by main country coded, I only cluster thresholds about universal values.

[9]These cutpoints are arbitrary, but are intended to be a happy medium between overly-specific cutpoints (which are computationally difficult) and overly-broad cutpoints which would be imprecise. Ordinalization of the data is necessary because the data clearly do not follow a normal distribution, and estimating them as following a Beta distribution substantially downweights their contribution to the estimation process and creates massive computational issues.

[10]Following standard practice (Pemstein et al. 2018), I collapse observations into regimes, i.e. periods in which no manifest variable (expert codings or EPR data) changed for a country. While the main purpose of this approach is to avoid overly confident latent estimates, in practice this reduction strategy reduces the number of observations individual V–Dem experts code, increasing the relative weight of the EPR.

of inclusion increasing, while EPR codes them as decreasing, the trends will likely reflect V–Dem coding. However, in cases of disagreement among individual V–Dem experts, the EPR data should in principle facilitate in adjudicating between the codings: all things being equal, a V–Dem expert who disagrees with the EPR data with regard to country trends should receive a lower reliability score relative to other experts agree with EPR.

5.2 Identity-based inclusion as a hierarchical latent variable

The first latent variable has the advantage of being relatively straightforward. However, Bayesian latent variable models are easily extendable to incorporate additional information about a latent concept of interest. By incorporating additional information, the model will provide more accurate and precise estimates of the concept, in this case identity-based inclusion.

Here I provide a second model that incorporates additional information. Specifically, the simple latent variable model of identity-based political inclusion becomes a node within a larger hierarchical model. This hierarchical model also incorporates data on another form of inclusion ("social inclusion," or the degree to which identity is relevant to state-sponsored discrimination in a country-year), as well as information about broader ethnic demographics (both religious and ethnic fractionalization and polarization, as well as the proportion of the population that belongs to a politically-relevant group). While I discuss the different nodes in turn, Figure 6 provides a conceptual overview of the model.

Figure 6: Hierarchical latent variable model for identity-based inclusion
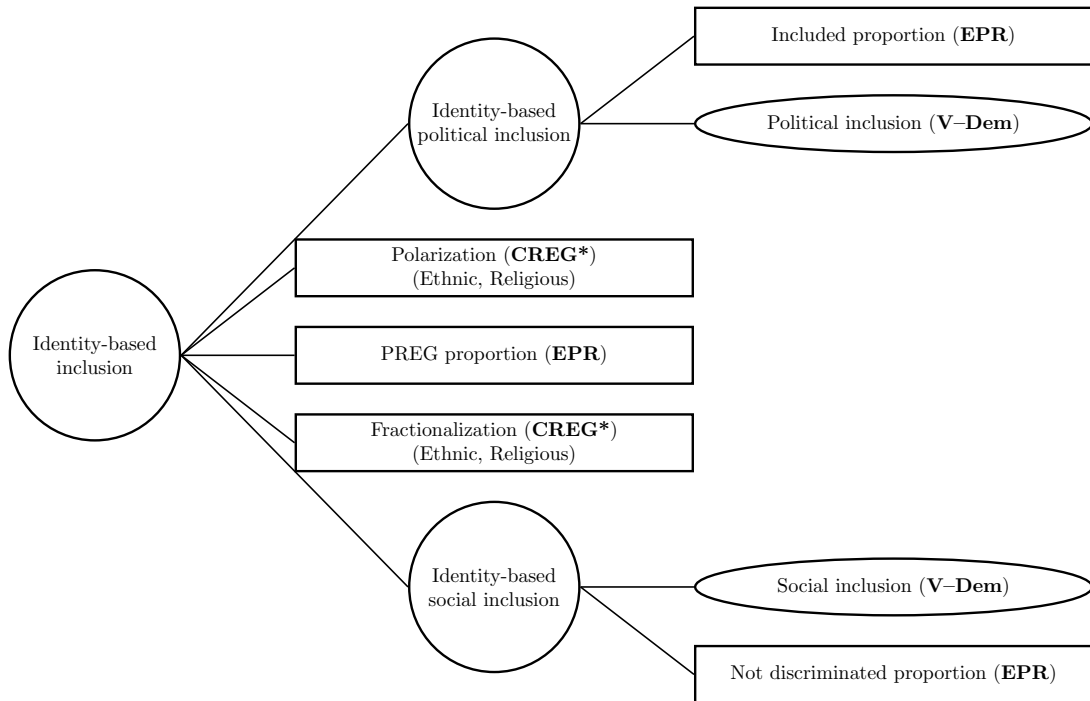
5.2.1 Political inclusion

I model political inclusion exactly as in the the previous model, but instead of modeling each latent value ξ_{ct} as being distributed according to a standard normal distribution, I model them as being normally distributed about a hierarchical prior value (θ_{ct}) with variance $\omega_1 \sim Cauchy(1,1)$ (restricted to positive values), i.e. $\xi_{ct} \sim N(\theta_{ct}, \omega_1)$. In principle, this approach could allow this node to vary substantially from both its prior and the other node; in practice, the high correlation between the different nodes reduces the output to essentially one latent variable.

5.2.2 Social inclusion

While political exclusion provides a strong explanation for elite preferences over conflict outcomes, it relies on a strong instrumentalist approach toward politics to explain why this explanation would hold for the preferences of non-elite group members. That is, these arguments assume that members of ethnic groups feel that their life prospects are linked to those of other group members, and that elite political representation is therefore of great importance—symbolic or otherwise—to individuals not engaged in politics. While classic work on ethnic politics supports this assumption (Bates 1983, Horowitz 2000), another body of literature argues that individual-level experiences are more salient to popular political preferences (Gellner 1983, Chandra 2006, Chandra 2012). More specifically, this body of literature argues that social exclusion, or denying members of a group opportunities available to members of other groups, leads to political conflict. These two forms of exclusion are clearly correlated: Horowitz (2000) notes that political exclusion often presages social exclusion. Given that both social and political inclusion theoretically reflect draws from a similar underlying distribution ("Identity-based inclusion") including them in a hierarchical model is reasonable. In principle social inclusion estimates could diverge substantially from political inclusion estimates, given that it has its own vague variance parameter.

The V–Dem measure for social inclusion is similar to that for political inclusion, but regards civil liberties as opposed to political power. Specifically, 1,145 experts provided an ordinal response to the question "Do all social groups...enjoy the same level of civil liberties, or are some groups generally in a more favorable position?" The five responses to this question range from "members of some social groups enjoy much fewer civil liberties than the general population" to "members of all salient social groups enjoy the same level of civil liberties." As with the question regarding political inclusion, this measure is largely agnostic about the forms of social identities relevant to inclusion, focusing mainly on the degree to which civil liberties are restricted based on social identities. In the EPR data set, social inclusion (the population that is not "Discriminated') represents a subset of the larger political inclusion category in which the state actively persecutes

Figure 7: Relationship between social and political inclusion

(a) V–Dem

(b) EPR
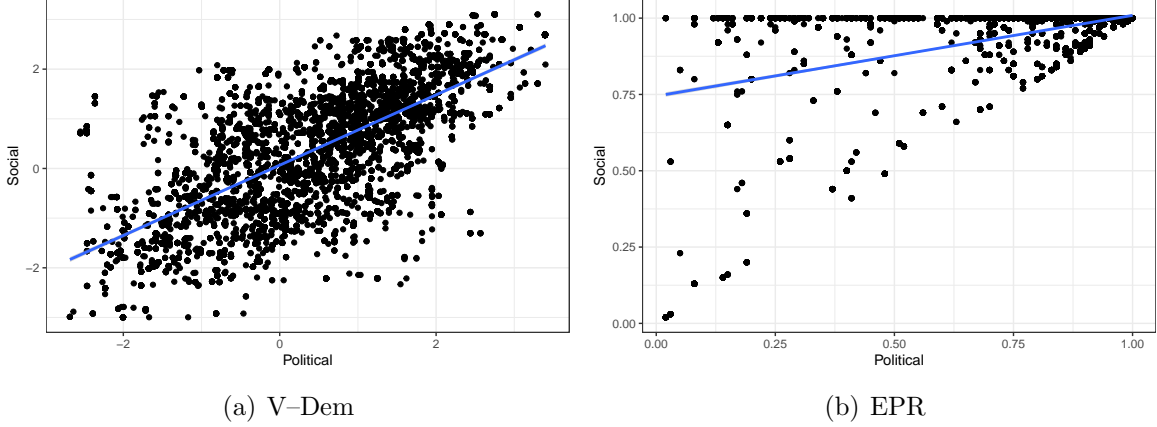
group members, as opposed to group members simply lacking representation.

Both measures are also highly correlated with their political corollaries in their respective datasets, as Figure 7 illustrates. This high correlation is strong evidence that both social and political inclusion are drawn from a common distribution, though perhaps with some variance.

I include social inclusion into the model using the same strategy as with political inclusion: the latent social inclusion latent variable λ_{ct} is distributed $N(\theta_{ct}, \omega_2)$, with with variance $\omega_2 \sim Cauchy(1, 1)$ (restricted to positive values).

5.2.3 Overall identity-based inclusion

Identity-based exclusion can only occur when there are multiple identity groups in a society. As a result, I model the prior for both social and political inclusion as being manifested in different statistics related to identity-based diversity: 1) CREG variables representing both fractionalization and polarization for both religious and ethnic groups, 2) the proportion of a country's population that is politically relevant (EPR). As with the EPR data on inclusion, I ordinalize all of these variables at .125 intervals. I treat each diversity measure p as having reliability parameters $\zeta_p \sim N(-1, 1)$ (restricted to negative values since all diversity variables have a negative relationship with the underlying concept), and k thresholds $\eta_{pk} \sim Cauchy(0, 2)$. Again, note that since both forms of exclusion are allowed to diverge from this prior (which is an overall estimate of politically-relevant identity-based diversity), the model is not assuming that exclusion necessarily follows from identity-based diversity. Instead, the assumption of the model is that exclusion is more likely in certain demographic permutations of identity-based diversity. Due to the iterative nature of the Bayesian algorithm, the relative weight of different demographic statistics– and the level of overall identity salience—is also informed by the estimates of political and social inclusion.

5.3   Modeling

I estimate both models using RStan. They run 50 thousand iterations over eight chains, with a burn-in of 25 thousand and thinning interval of 100. All parameters in both models converged with Gelman-Rubin diagnostic values below 1.1.

In the case of the hierarchical model, the variance parameters reduced to values close to 0. As a result, I only analyze the prior node values since the remaining two latent variables are essentially identical.

6   Validation

I validate the data according to the framework set forth by Adcock & Collier (2001) and McMann, Pemstein, Seim, Teorell & Lindberg (2016), focusing on content and construct validation. In the context of this latent variable model, content validation involves three steps. First, assessing the degree to which the model performs according to the theoretical assumptions that underpin its design by analyzing model parameters. Second, assessing the implications of the model for its output relative to the manifest variables, i.e. determining the degree to which the model diverges from its input variables, and assessing the degree to which these divergences align with theoretical expectations. This latter form of validation overlaps substantially with convergent/divergent analyses, which I do not perform independently. Third, assessing the degree to which model output aligns with theoretical expectations over cases. If the model performs as expected and model output aligns with theoretical expectations and case knowledge, it has high content validity. The analyses indicate that the simple latent variable model performs worse than the hierarchical model in this regard, but both models largely show high content validity.

Construct validity here involves assuming that identity-based inclusion is negatively correlated with conflict onset—as a large body of literature argues—then assessing the relationship between identity-based inclusion and conflict. A valid measure of identity-based inclusion would show the theoretically-expected relationship with conflict onset. The simple latent variable model largely fails this validation exercise, while the more complicated model passes it with flying colors.

6.1   Content validation

To validate the content of the two latent variable models, I first assess the performance of relative model parameters to see if they achieved the modeling objective of using the EPR data to bridge the more nuanced—but potentially less cross-nationally comparable—V–Dem data. Model parameter estimates indicate that both models have realized this objective—the simpler model to a lesser extent than the more complicated model—a conclusion that comparing the model output to the original EPR and V–Dem data confirms.

I then compare both cross-national patterns and within-country trends in the two datasets, finding that output from both models appears to largely dovetail with both modeling expectations and theoretical expectations.

6.1.1 Model performance

The goal of the latent variable models is to use the EPR data to bridge the V–Dem data by anchoring the V–Dem data to politically-relevant ethnic demographics. The modeling process should also allow both data sets to potentially correct errors in the other. To assess whether or not the models work as intended, I focus on both parameter estimates and their substantive influence on latent inclusion values.

In terms of parameter estimates, I focus on reliability and threshold parameters. Reliability parameters influence the degree to which different data sources (EPR and V–Dem coders) affect the latent parameter estimates. Threshold parameters determine the degree to which different levels of a manifest variable differentiate between levels of the latent parameter. Figure 8 illustrates the relationship between V–Dem expert and EPR coding reliability and thresholds, across models and data. The top row illustrates estimates from the simple latent variable that aggregates V–Dem and EPR data on political inclusion. The bottom two rows illustrate this relationship for the hierarchical model, with the middle row illustrating the relationship for political inclusion and the bottom social inclusion. The left column superimposes the posterior median estimate of EPR's reliability as a vertical line over the distribution of V–Dem expert posterior median reliability scores, illustrating the relative strength of the EPR estimate vis-a-vis V–Dem experts in influencing latent parameter estimates. The right column superimposes the posterior median threshold estimates for the EPR manifest variables over the distribution of thresholds for V–Dem experts, illustrating the latent value estimates over which the different sources of data differentiate.

Focusing first on the top row, the EPR posterior median reliability score is substantially lower than the average V–Dem expert score (median estimates of .57 vs. 1.15), which would indicate that the EPR estimates of political inclusion substantively influence country-year estimates of political inclusion to a lesser extent than V–Dem experts. However, the posterior median is a somewhat misleading statistic in this context because there is much higher measurement error in the estimates of V–Dem expert reliability. On average, V–Dem expert reliability scores have a standard deviation of .57 over posterior draws; the EPR reliability has a standard deviation of .05 over posterior draws. As a result, though the EPR data influence the latent variable estimation process to a lesser extent than individual V–Dem experts, the influence is much more consistent over draws from the posterior distribution.

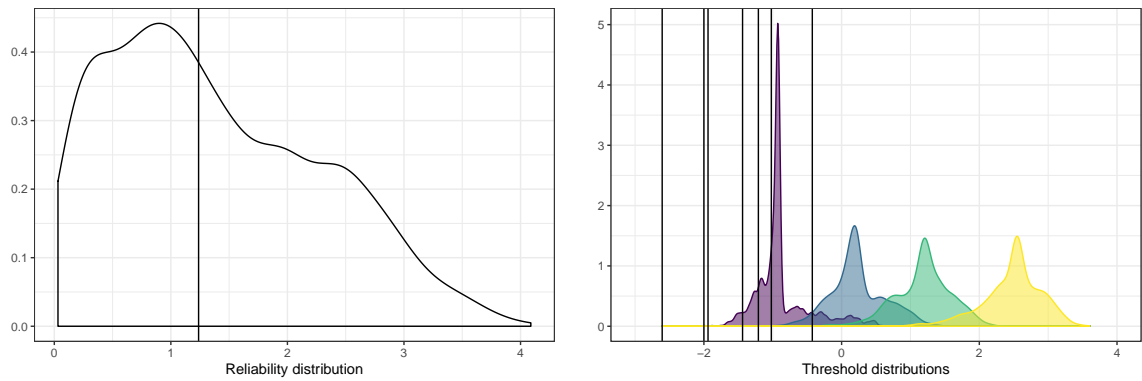In terms of the thresholds, the EPR thresholds are largely concentrated about the

Figure 8: Latent variable reliability and threshold comparison
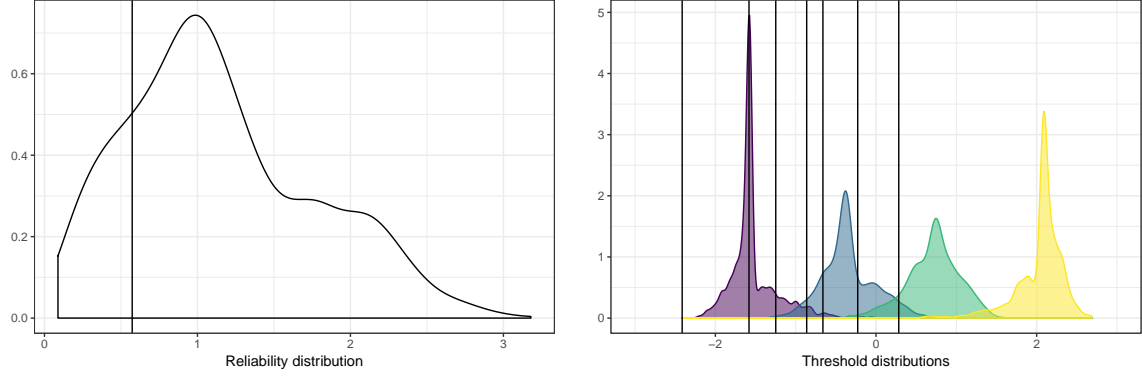
(a) Latent variable
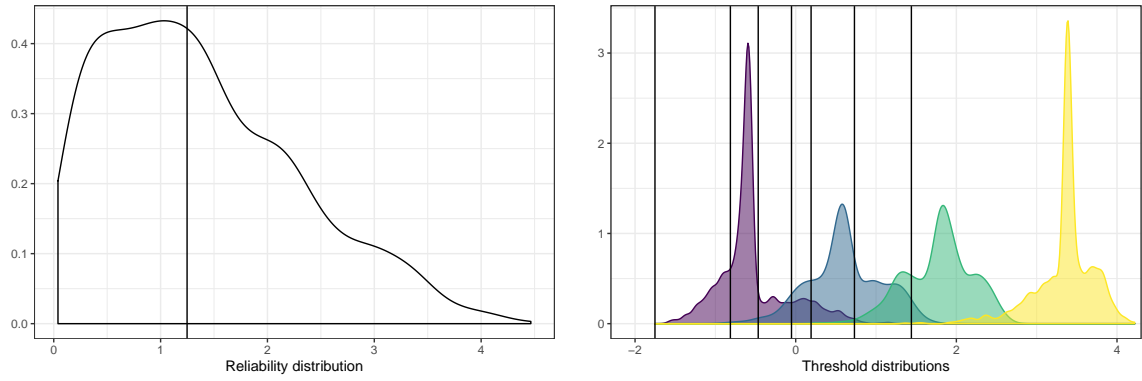
(b) Hierarchical (Political)

(c) Hierarchical (Social)

Densities represent the distribution of V–Dem expert point estimates; vertical lines represent the point estimates for the EPR manifest variables.

first and second thresholds for V–Dem experts. This relative positioning of thresholds indicates that the EPR political inclusion variable faciliates the differentiation of country-years with relatively low inclusion, while poorly differentiating relatively high inclusion. This result stands to reason given the different foci of the V–Dem and EPR data. While the EPR data would code a country where only a small minority is excluded as being relatively inclusive based on the demographic strength of the included population, V–Dem experts might code this case as still being exclusive based on the severity of the exclusion.

The middle row shows a similar pattern in the relationship in the case of the hierarchical model of identity-based inclusion. The EPR data largely do not differentiate between relatively politically-inclusive country years, but do differentiate between relatively politically-exclusive observations. This result is even more pronounced in the case of social inclusion (bottom row), where the EPR data largely differentiates only between very exclusive country-years (about the lowest V–Dem threshold distribution). This result is perhaps understandable given the very strict coding criteria for discrimination in the EPR project (i.e. politically included populations cannot face discrimination by definition), which would allow countries that have high levels of discrimination to still be coded as non-discriminatory.

Finally, in the hierarchical model the EPR data have much higher reliability scores than in the simpler latent variable model. This result is likely due to the increased weight given to ethnic demographics via the prior. Figure 9 illustrates the reliability and threshold parameter estimates for the hierarchical model, excluding estimates for V–Dem experts. I flip the scales on all variables save Political (EPR) and Social (EPR) inclusion to better illustrate their relationship to the latent variable; all variables with flipped scales in fact have negative correlations with overall inclusion. The highest threshold for ethnic fractionalization (E.Frac) thus becomes the value over which the latent variable must fall to belong to the *lowest* fractionalization category.
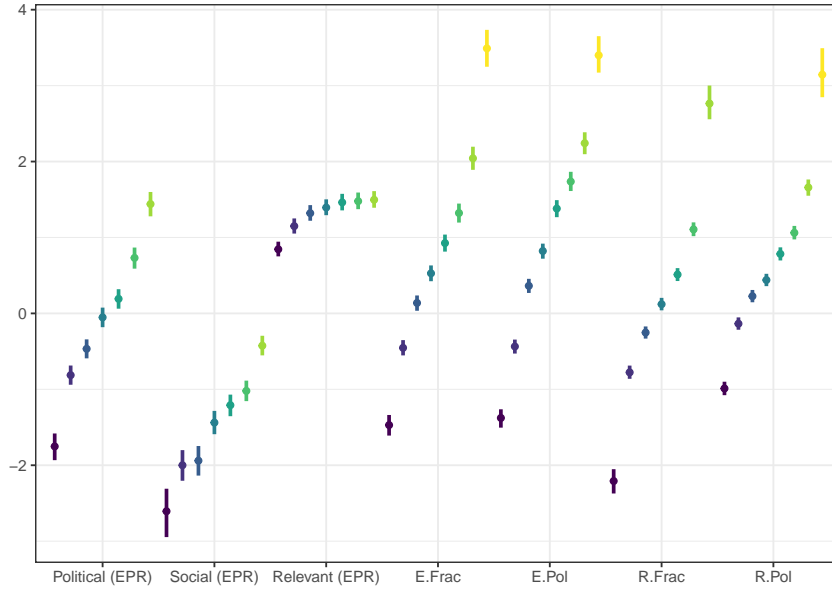
Both political and social inclusion parameters receive the highest reliability values, which follows from the fact that they are the most highly correlated with the concept being estimated (as well as the V–Dem codings). Ethnic fractionalization and polarization (E.Frac and E.Pol) receive high reliability estimates relative to religious fractionalization and polarization (R.Frac and R.Pol) which again follows from the greater focus on ethnic identities relative to religious identities in the data. Finally, the proportion of a population that belongs to a "politically relevant" identity group has a lower reliability than the other ethnic variables, which is perhaps evidence that inclusion and broader demographics (i.e. polarization and fractionalization) are more relevant to the latent concept.

Turning to the threshold estimates, Figure 9 illustrates that measures of fractionalization and polarization — both ethnic and religious — tend to be better at differentiating high levels of the latent concept than the demographic measures of inclusion themselves

Figure 9: Hierarchical model reliability and threshold parameters
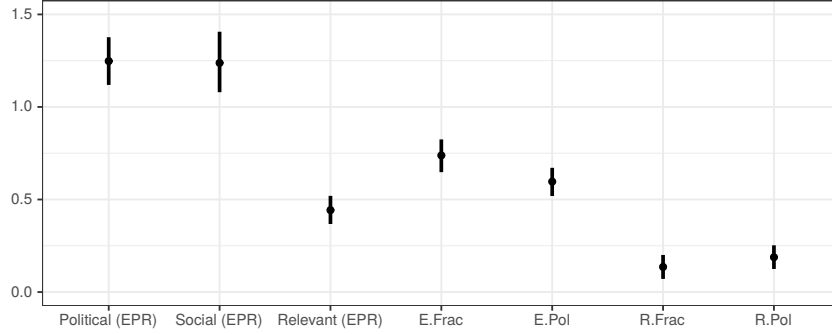
(a) Reliability parameters

(b) Threshold parameters

Points represent posterior medians and vertical lines 95% credible regions. Scales flipped for all variables save Political (EPR) and Social (EPR).

(EPR data on social and political inclusion). This result dovetails with the previous comparisons of the EPR measures and the V–Dem expert thresholds. The V–Dem data indicate that some levels of exclusion are possible even when most of a country's population is represented in elite politics. These analyses indicate that the difference between high and very high levels of inclusion is informed by the degree to which a country is (not) fractionalized or polarized. Countries that have very low polarization and fractionalization are thus the most likely to be inclusive, potentially because they are closest to being monolithic in terms of ethnic-type identities.

Figure 10 illustrates the implications of these model parameters for country-year estimates of inclusion. All four cells show a scatterplot of the posterior median estimates from the original V–Dem political inclusion indicator (vertical axis) and the posterior median estimates from the simple latent varible model (left cells) and the posterior median estimates of the more complex hierarchical model (right cells). The top row divides the estimates by the level of political inclusion from the EPR project, while the bottom row divides the estimates by levels of ethnic fractionalization.

Focusing first on the top row, it is clear that the inclusion of EPR data into the estimation process has altered country-year estimates in both models: data with low values of demographic (EPR) inclusion have systematically lower point estimates relative to the original V–Dem data. However, it is also clear that the hierarchical model allows the EPR data to much more starkly differentiate the estimates of identity-based inclusion, with estimates for each level of demographic inclusion being systematically lower than estimates for the next-highest level. This result confirms the insight from the model parameter analyses: the hierarchical model incorporates the EPR data into the estimation process to a greater extent, in line with the goals of the modeling enterprise.

The bottom row has two functions. In the case of the simple latent variable model of political inclusion (left), ethnic fractionalization plays no role in the estimation process. As a result, there should be no systematic differences based on ethnic fractionalization between the V–Dem and latent variable posterior median estimates, except insofar as ethnic fractionalization is correlated with the EPR inclusion data that are included in the estimation process. In contrast, given the high reliability of the fractionalization variable in the hierarchical model, we would expect there to be strong differentiation between the V–Dem and latent variable posterior median estimates based on levels of ethnic fracitionalization, especially at particularly low levels of fractionalization. Indeed, both of these relationships are visible in the data: there is little systematic difference between the V–Dem measure and the latent estimates from the simple model based on ethnic fractionalization, while there are clearly observable differences in the comparison to the hierarchical variable. This validation exercise confirms that the hierarchical model estimates of identity-based exclusion are informed to a noticeable extent by demographic data in addition to the EPR data.

Figure 10: Relationship between latent variable and input data

(a) Latent (EPR)

(b) Hierarchical (EPR)

(c) Latent (Ethnic fractionalization)

(d) Hierarchical (Ethnic fractionalization)
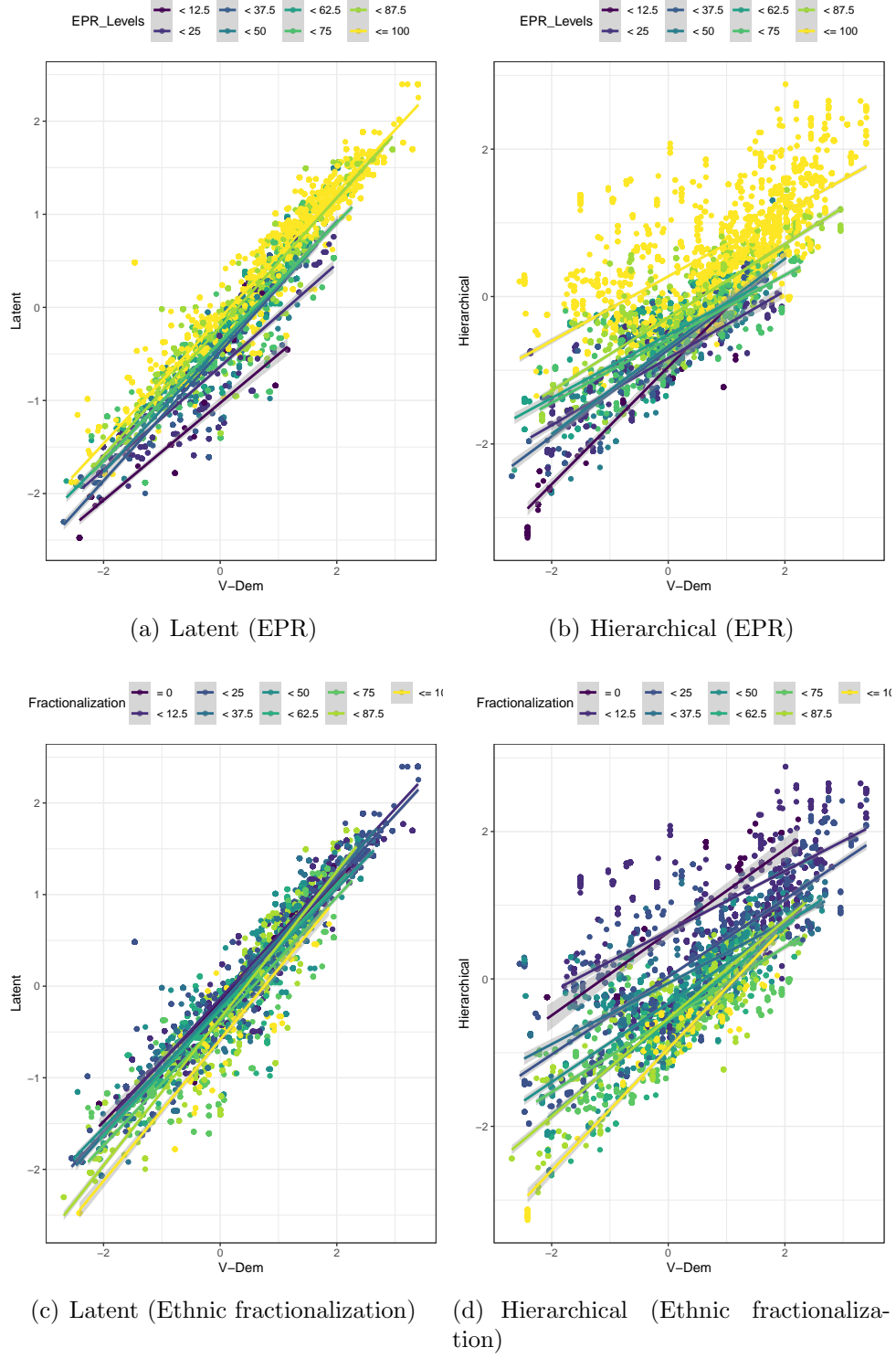
6.2 Across- and within-country comparisons

To analyze whether these modeling strategies are actually of substantive importance, I provide both cross-national analyses of data from 2017 (the most recent year in the dataset) and return to the analysis of longitudinal trends in the BRICS countries and the United States. In principle, these analyses facilitate content validation by facilitating an assessment of the degree to which the model estimates align with substantive knowledge of identity-based exclusion around the world and within countries. In future iterations of the project there will be a more extensive discussion along these lines.

Figure 11 illustrates relative levels of identity-based political inclusion across all countries in the year 2017, focusing on 1) the posterior median estimates from the simple latent variable model, 2) the demographic inclusion estimates from EPR, and 3) the original V–Dem political inclusion estimates. I scale the V–Dem and latent variable estimates to a 0-1 scale using the CDF of a standard normal distribution to facilitate comparison with the EPR data; horizontal lines represent the 90% credible region about the latent variable estimates. Higher values represent greater identity-based political inclusion, lower values the opposite.

Several aspects of this figure warrant discussion. First, the V–Dem indicator and the latent variable estimates track each other closely, with most V–Dem estimates falling within the 90% credible regions from the latent estimate. However, there are cases of divergence: for example, Kazakhstan has a much lower latent estimate of political inclusion than its V–Dem corollary. This result is likely due to the fact that it has a relatively low demographic (EPR) inclusion estimate. Second, while comparing direct estimates from EPR and the latent variable estimate is difficult given the ordinalization of the EPR data in the estimation process, general trends are observable. Most clearly, while there are many cases of perfect demographic inclusion in countries with relatively high levels of estimated latent political inclusion (right cell), there are few in cases with low inclusion. The exceptions to the latter rule are generally confined to cases of great discrepancy between the V–Dem and EPR data (e.g. aforementioned cases like North Korea and Qatar). In general, this analysis yields a conclusion in line with those from the model parameter analyses: the model is allowing the EPR data to weight the V–Dem data to some extent, but the V–Dem data clearly overpower the EPR data in many cases.

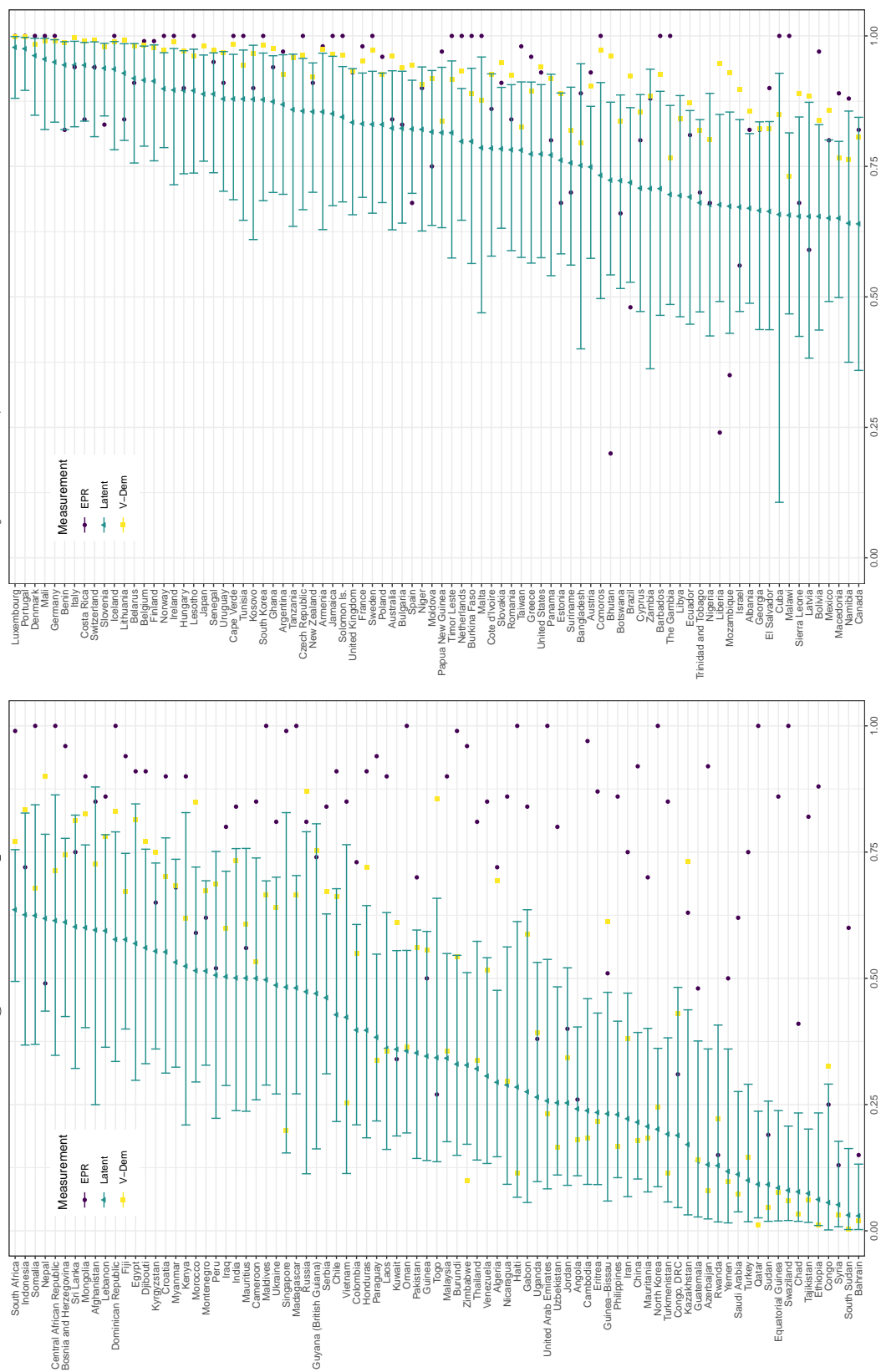Figure 11: Comparison of measures of identity salience, 2017

Figure 12 compares 2017 country-year estimates from both the simple latent variable model of identity-based political inclusion and the more complicated hierarchical model of overall identity-based inclusion. Again, horizontal lines represent 90% credible regions and points posterior medians. Unsurprisingly given the previous analyses, point estimates from the two latent variables are clearly highly correlated. However, there are cases of relatively substantial divergence: for example, Haiti is estimated to have much lower inclusion in the simple model than in the hierarchical model. Such discrepancies may be either due to differences in levels of social and political inclusion in the two countries or, perhaps, the fact that demographics have greater weight in the hierarchical analysis. Haiti has relatively low demographic diversity, which likely increases its score in the hierarchical model.

Also of note are the much tighter credible regions for the hierarchical model; this is likely due to the additional data available to this the model.

For a final step for content validation, I return to longitudinal case studies using the BRICS and USA as examples in Figure 13.[11] Focusing on the latent variable models, it is clear that output from both models generally shows similar trends. There are two important caveats to that statement. First, while the trends are similar, the estimated levels of relative inclusion often diverge substantially, with the hierarchical model generally estimating these cases to be less inclusive than the simpler latent model. These lower values likely reflect both the lower social inclusion in these countries vis-á-vis political inclusion as well their relatively high identity diversity. Second, the case of China shows both dissimilar trends and levels over the two latent variable estimates. Again, this result is likely attributable to China's largely Han Chinese population, which increases the estimated level of inclusion in the hierarchical model; the trends in decreasing inclusion as estimated by the hierarchical model is likely due to increasing social exclusion of minorities in recent years, most prominently the Uighurs.

---

[11]Latent variable estimates converted to a zero to one scale using the CDF of a normal distribution.

Figure 12: Comparison of latent measures of identity salience, 2017

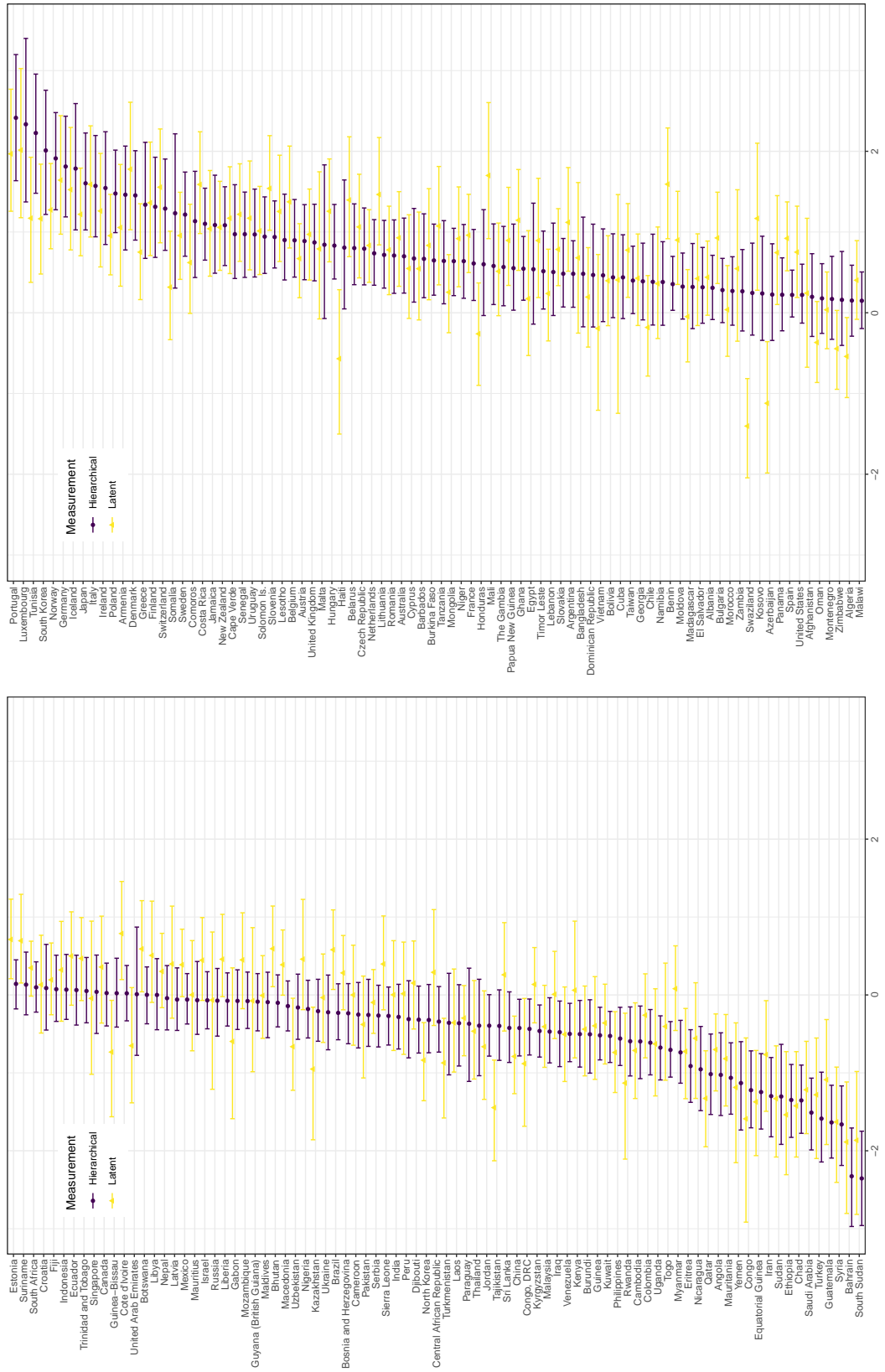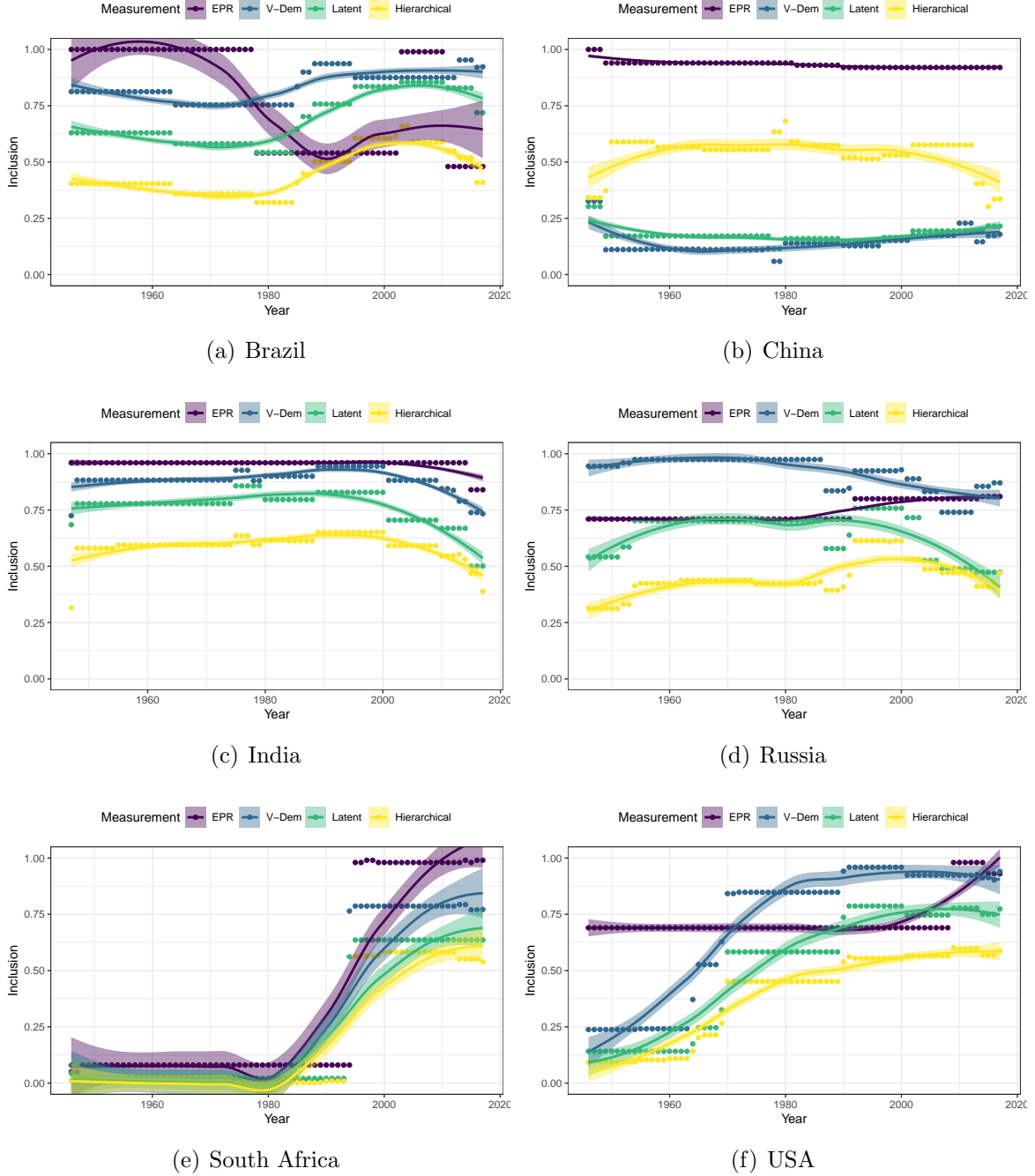Figure 13: Time series for political inclusion

(a) Brazil

(b) China

(c) India

(d) Russia

(e) South Africa

(f) USA

Lines represent local regression estimates and corresponding uncertainty; they do not represent measurement uncertainty about estimates.

6.3 Construct validation: Analyses of civil conflict onset

The content validation in the previous section indicates that the latent-variable models—in particular the hierarchical model—successfully merge the V–Dem and EPR data, with demographic information bridging the V–Dem data. However, the degree to which this merging is worthwhile remains unclear. While the cross-national and within-country analyses provide some face validity for the strategy, a concept as loaded as identity-based inclusion means that readers may very well interpret the validity of the measures differently based on their idiosyncratic understandings of identity and inclusion. Construct validation allows me to assess the degree to which the latent variable estimates evince a theoretically-expected relationship with outcomes of interest. More valid measures show a relationship that is in line with the theoretical expectations. I conduct construct validation by analyzing the relationship between the latent variable posterior medians and civil conflict onset.

As previously discussed, many scholars hold that there is a relationship between exclusion and civil conflict: as identity-based inclusion decreases, the probability of conflict onset increases. Assuming this hypothesis is correct, an accurate measure of identity-based inclusion would predict civil conflict onset. Equally importantly given concerns about the cross-national comparability of data on identity-based inclusion, higher levels of exclusion should predict conflict both across- and within-countries. I therefore analyze data with both country and year random and fixed effects.

Since the goal of this analysis is to validate the data without any pretensions for causal inference, I have endeavored to estimate relatively straightforward and inoffensive models. Specifically, I run linear probability models that have as an outcome the UCDP/PRIO indicator of conflict onset. All models also include a cubic spline for the number of years since a country entered the data set or the last cessation of hostilities. I remove cases of ongoing conflict.

I initially analyze the relationship between four lagged indicators of inclusion independently without controls, including both the EPR and V–Dem indicators of political inclusion for the sake of comparison with the latent variable estimates. I rescale all latent variables from zero to one for the sake of comparing the correlation magnitude. In an additional four models I include controls—V–Dem Polyarchy, ln(GDP per capita) and ln(Population)—to assess robustness. I lag all relevant right-hand side variables by one year. Table 2 presents results from a model with country and year random effects, while Table 3 presents results from a model with country and year fixed effects.

Table 2: Linear probability models of conflict onset with country and year random effects, cubic spline

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Inclusion (EPR)	−0.03** (0.01)				−0.03 (0.02)			
Inclusion (V–Dem)		−0.02* (0.01)				−0.03 (0.02)		
Inclusion (Latent)			−0.02 (0.01)				−0.01 (0.02)	
Inclusion (Hierarchical)				−0.07*** (0.02)				−0.10*** (0.02)
Polyarchy					−0.03* (0.02)	−0.01 (0.02)	−0.02 (0.02)	0.01 (0.02)
ln(GDP per capita)					−0.11** (0.05)	−0.11** (0.05)	−0.11** (0.05)	−0.11** (0.05)
ln(Population)					0.25*** (0.05)	0.25*** (0.05)	0.24*** (0.05)	0.24*** (0.05)
Constant	0.10*** (0.01)	0.09*** (0.01)	0.09*** (0.01)	0.11*** (0.01)	0.44*** (0.06)	0.43*** (0.06)	0.42*** (0.06)	0.45*** (0.06)
Observations	8,609	8,609	8,609	8,609	6,602	6,602	6,602	6,602
Bayesian Inf. Crit.	−5,017	−5,016	−5,015	−5,031	−3,979	−3,979	−3,977	−3,995

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3: Linear probability models of conflict onset with country and year fixed effects, cubic spline

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Inclusion (EPR)	−0.02 (0.02)				0.01 (0.02)			
Inclusion (V−Dem)		−0.02* (0.01)				0.01 (0.02)		
Inclusion (Latent)			−0.01 (0.02)				0.04 (0.02)	
Inclusion (Hierarchical)				−0.08*** (0.02)				−0.05* (0.03)
Polyarchy					0.01 (0.02)	0.002 (0.02)	−0.01 (0.02)	0.02 (0.02)
ln(GDP per capita)					−0.002 (0.07)	0.001 (0.07)	0.01 (0.07)	−0.02 (0.07)
ln(Population)					0.42*** (0.15)	0.42*** (0.15)	0.43*** (0.15)	0.40*** (0.15)
Constant	0.13*** (0.03)	0.14*** (0.03)	0.13*** (0.03)	0.18*** (0.03)	0.43** (0.18)	0.43** (0.18)	0.42** (0.18)	0.45** (0.18)
Observations	8,609	8,609	8,609	8,609	6,602	6,602	6,602	6,602
R²	0.20	0.20	0.20	0.20	0.25	0.25	0.25	0.25

Note: *p<0.1; **p<0.05; ***p<0.01

The results from linear probability models that include country and year random effects (Table 2 presents) provide strong evidence that the hierarchical latent variable model has the highest construct validity among the different measures of identity-based inclusion. When estimated without controls (Models 1-4), all indicators of inclusion save the posterior median from the simple latent variable model show the theoretically-expected significant and negative relationship with conflict onset. The fact that the simple latent variable model of identity-based political inclusion does not show this relationship is evidence that this indicator does not capture this concept as fully as the other indicators. On the other hand, the estimates from the hierarchical latent variable model show the relationship with conflict onset with the highest magnitude, which is strong evidence of its construct validity. Equally importantly, this indicator is the only one that is robust to the addition of controls (Models 5-8). This result indicates that the hiearchical measure of inclusion has the highest construct validity when comparing levels of inclusion cross-nationally.

The results from Table 3 provide further evidence of the construct validity of the hierarchical latent variable model estimates of inclusion. In the context of country and year fixed effects, the hierarchical latent variable model of identity-based inclusion shows the substantively strongest relationship with conflict onset, and is the only indicator that is robust across model specifications.

7   Conclusion

How to best measure politically-relevant identity is an essential question in political science. The most prominent approach—epitomized by the EPR project—is to measure the degree to which politically-relevant ethnic groups are excluded. This approach is a relatively blunt measure of the concept in that it has difficulties accounting for either inter- and intra-ethnic variation in inclusion, as well as fine-grained changes in inclusion over time. An alternative approach—epitomized by the V–Dem project—is to measure the latent level of identity-based inclusion in a society, without regard to specific groups. This approach is perhaps more nuanced, but only indirectly measures ethnic demographics, which are of clear importance to social-scientific outcomes.

In this paper, I compare and contrast these two approaches and argue that, in many ways, they are complementary. Moreover, I propose two latent variable models that aggregate these two types of data. One model is a relatively straighforward ordinal IRT model of identity-based political inclusion that treats the EPR data as fully-bridged and cross-nationally comparable coder. The other model takes a more complicated hierarchical latent variable modeling approach, incorporating additional data regarding both identity-based social inclusion, as well as general identity demographics. In principle, both models would allow the EPR data to anchor V–Dem experts to politically-relevant

ethic demographics, increasing cross-national comparability in a substantively important way. Content validation exercises indicate that, while both models are successful in this regard, the hierarchical model incorporates demographics to a much greater extent. Construct validation indicates that this greater incorporation of demographic data yields estimates that more fully align with important social-scientific theory.

In addition to providing data of potentially great use to scholars interested in measuring identity-based inclusion, I have also designed these models to provide a baseline for future future modifications. Adding additional data regarding identity-based diversity (e.g. from AMAR Birnir et al. (2014) or Livny (ND)) would provide finer-grained measures of identity salience. Incorporating additional data about other forms of identity-based inclusion (e.g. from the exclusion survey that began in V–Dem v9) would only require adding another node to the model. Finally, for the sake of simplicity, I have restricted the time series to dates with perfect coverage across data sources; in principle, using an approach such as that which Fariss (2014) uses to measure human rights across time would allow the time series to extend the full length of the V–Dem data (i.e. to 1900).

References

Abdelal, Rawi, Yoshiko M. Herrera, Alastair Iain Johnston & Rose McDermott. 2009. *Identity as a Variable*. Cambridge: Cambridge University Press pp. 17–32.

Adcock, Robert & David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *The American Political Science Review* 95(3):529–546.

Bakker, Ryan, Seth Jolly, Jonathan Polk & Keith Poole. 2014. "The European Common Space: Extending the Use of Anchoring Vignettes." *The Journal of Politics* 76(4):1089–1101.

Bates, Robert. 1983. *Modernization, Ethnic Competition and the Rationality of Politicis in Contemporary Africa*. Boulder, CO: Westview chapter 9, pp. 152–61.

Birnir, Jóhanna K, Jonathan Wilkenfeld, James D Fearon, David D Laitin, Ted Robert Gurr, Dawn Brancati, Stephen M Saideman, Amy Pate & Agatha S Hultquist. 2014. "Socially relevant ethnic groups, ethnic structure, and AMAR." *Journal of Peace Research* .

Bochsler, Daniel, Elliott Green, Erin Jenne, Harris Mylonas & Andreas Wimmer. 2021. "Exchange on the quantitative measurement of ethnic and national identity." *Nations and Nationalism* pp. 1–19.

Bormann, Nils-Christian, Lars-Erik Cederman & Manuel Vogt. 2015. "Language, Religion, and Ethnic Civil War." *Journal of Conflict Resolution* pp. 1–28.

Brubaker, Rogers. 2002. "Ethnicity without groups." *European Journal of Sociology* 43(02):163–189.

Cederman, Lars-Erik, Andreas Wimmer & Brian Min. 2010. "Why Do Ethnic Groups Rebel? New Data and Analysis." *World Politics* 62(1).

Cederman, Lars-Erik, Kristian Skrede Gleditsch & Halvard Buhaug. 2013. *Inequality, grievances, and civil war.* Cambridge University Press.

Cederman, Lars-Erik, Simon Hug, Andreas Schädel & Julian Wucherpfennig. 2015. "Territorial Autonomy in the Shadow of Conflict: Too Little, Too Late?" *American Political Science Review* 109:354–370.

Chandra, Kanchan. 2006. "What Is Ethnic Identity and Does It Matter?" *Annual Review of Political Science* 9.

Chandra, Kanchan. 2012. Oxford University Press chapter 1-4, pp. 1–178.

Clio-Infra. 2013. "Database." `http://www.clio-infra.eu/`.

Coppedge, Michael, John Gerring, Staffan I. Lindberg, Svend-Erik Skaaning, Jan Teorell et al. 2017*a*. Varieties of Democracy Codebook v7. Technical report Varieties of Democracy Project: Project Documentation Paper Series.

Coppedge, Michael, John Gerring, Staffan I. Lindberg, Svend-Erik Skaaning, Jan Teorell et al. 2017*b*. Varieties of Democracy Methodology v7. Technical report Varieties of Democracy Project: Project Documentation Paper Series.

Esteban, Joan, Laura Mayoral & Debraj Ray. 2012. "Ethnicity and Conflict: Theory and Facts." *Science* 336(6083):858–865.

Fariss, Christopher J. 2014. "Respect for human rights has improved over time: Modeling changing standards of accountability." *American Political Science Review* 108(2):297–318.

Fearon, James D. 2003. "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth* 8(2):195–222.

Fearon, James D. & David D. Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *The American Political Science Review* 97(1):75–90.

Gandhi, Jennifer & Jane Lawrence Sumner. 2020. "Measuring the consolidation of power in nondemocracies." *Journal of Politics* 84(2):1545–1558.

Gellner, Ernest. 1983. *Nations and Nationalism*. Ithaca: Cornell University Press.

Hale, Henry E. 2008. *The Foundations of Ethnic Politics: Separatism of States and Nations in Eurasia and the World*. New York: Cambridge University Press.

Hale, Henry E. 2017. "Focus on the fundamentals: Reflections on the state of ethnic conflict studies." *Ethnopolitics* 16(1):41–47.

Horowitz, Donald L. 2000. *Ethnic Groups in Conflict*. 2 ed. Berkeley: University of California Press.

Hug, Simon. 2003. "Selection Bias in Comparative Research: The Case of Incomplete Data Sets." *Political Analysis* 11(3):255–274.

Hug, Simon. 2013. "The use and misuse of the "Minorities at Risk" project." *Annual Review of Political Science* 16:191–208.

Ignatiev, Noel. 1995. *How the Irish became white*. New York: Routledge.

King, Gary & Jonathan Wand. 2007. "Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes." *Political Analysis* 15(1):46–66.

Livny, Avital. ND. Improving and integrating global diversity estimates using transparent methods. Technical report Cline Center for Democracy, University of Illinois at Urbana-Champaign.

Lührmann, Anna, Kyle L Marquardt & Valeriya Mechkova. 2020. "Constraining governments: New indices of vertical, horizontal and diagonal accountability." *American Political Science Review* 114(3):811–820.

Marquardt, Kyle L. 2020. "How and how much does expert error matter in expert-coded data? Implications for quantitative peace research." *Journal of Peace Research* 57(6):692–700.

Marquardt, Kyle L. & Yoshiko M. Herrera. 2015. "Ethnicity as a Variable: An assessment of measures and datasets of ethnicity and related identities." *Social Science Quarterly* pp. 689–716.

McMann, Kelly M, Daniel Pemstein, Brigitte Seim, Jan Teorell & Staffan I Lindberg. 2016. "Strategies of Validation: Assessing the Varieties of Democracy Corruption Data." *Varieties of Democracy Institute Working Paper* 23(February).

Montalvo, Jose G. & Marta Reynal-Querol. 2005. "Ethnic Polarization, Potential Conflict, and Civil Wars." *The American Economic Review* 95(3):796–816.

Pemstein, Daniel, Stephen A. Meserve & James Melton. 2010. "Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type." *Political Analysis* 18(4):426–449.

Pemstein, Daniel et al. 2018. "The V-Dem Measurement Model: Latent Variable Analysis for Cross-National and Cross-Temporal Expert-Coded Data." *Varieties of Democracy Institute Working Paper* 21(3rd Ed).

Posner, Daniel N. 2004. "Measuring Ethnic Fractionalization in Africa." *American Journal of Political Science* 48(4):849–863.

Selway, Joel Sawat. 2011. "The measurement of cross-cutting cleavages and other multi-dimensional cleavage structures." *Political Analysis* pp. 48–65.

Solis, Jonathan A & Philip D Waggoner. 2020. "Measuring media freedom: An Item Response Theory analysis of existing indicators." *British Journal of Political Science* pp. 1–20.

Stan Development Team. 2018. "RStan: The R interface to Stan.". R package version 2.18.2.
URL: http://mc-stan.org/

Teorell, Jan Teorell, Michael Coppedge, Sven-Erik Skaaning & Staffan I. Lindberg. 2016. "Measuring Electoral Democracy with V-Dem Data: Introducing a New Polyarchy Index." *Varieties of Democracy Institute Working Paper* 25.

The Maddison Project. 2013. "Database." http://www.ggdc.net/maddison/maddison-project/home.htm.

The Composition of Religious and Ethnic Groups (CREG) Project. 2014. Technical report Cline Center for Democracy, University of Illinois at Urbana-Champaign.

Treier, Shawn & Simon Jackman. 2008. "Democracy as a Latent Variable." *American Journal of Political Science* 52(1):201–217.

Vogt, Manuel, Nils-Christian Bormann, Seraina Rüegger, Lars-Erik Cederman, Philipp Hunziker & 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.

Wimmer, Andreas, Lars-Erik Cederman & Brian Min. 2009. "Ethnic Politics and Armed Conflict: A Configurational Analysis of a New Global Data Set." *American Sociological Review* 74(2):316–337.