Using Qualitative Information to Improve Causal Inference*

Adam N. Glynn† Nahomi Ichino‡

July 3, 2014

Forthcoming, *American Journal of Political Science*

Abstract

Using the Rosenbaum (2002; 2009) approach to observational studies, we show how qualitative information can be incorporated into quantitative analyses to improve causal inference in three ways. First, by including qualitative information on outcomes within matched sets, we can ameliorate the consequences of the difficulty of measuring those outcomes, sometimes reducing p-values. Second, additional information across matched sets enables the construction of qualitative confidence intervals on effect size. Third, qualitative information on unmeasured confounders within matched sets reduces the conservativeness of Rosenbaum-style sensitivity analysis. This approach accommodates small to medium sample sizes in a nonparametric framework, and therefore may be particularly useful for analyses of the effects of policies or institutions in a given set of units. We illustrate these methods by examining the effect of using plurality rules in transitional presidential elections on opposition harassment in 1990s sub-Saharan Africa.

*We thank Alberto Abadie, Matthew Blackwell, Maiko Heller, Konstantin Kashin, David Laitin, Evan Lieberman, Nick Weller, Teppei Yamamoto, and seminar participants at Michigan and Yale for helpful comments. Julie Faller, Amanda Pinkston, and Jisu Yoo provided able research assistance. The proposed methods can be implemented with `qualCI` (Kashin, Glynn and Ichino 2014), an R package that is freely available at the Comprehensive R Archive Network (CRAN). All replication files are available on the AJPS Dataverse. We did not receive funding for this research. Earlier versions of this paper were presented at the Institute for Qualitative and Multimethod Research Workshop at Syracuse University, June 23–24, 2012, and the 29th Annual Meeting of the Society for Political Methodology, July 19–21, 2012.

†Associate Professor, Department of Political Science, Emory University, 327 Tarbutton Hall, 1555 Dickey Drive, Atlanta, GA 30322. `aglynn@emory.edu`

‡Assistant Professor, Department of Political Science, University of Michigan, 5700 Haven Hall, 505 S. State St., Ann Arbor, MI 48109. `nichino@umich.edu`

Observational studies in political science are often beset by problems that can lead to fragile and biased estimates of causal effects. Most fundamentally, important confounding variables that affect both the treatment variable and the outcome variable may be unmeasured, and even measured confounding and outcome variables may only be poorly measured. Many of these observational studies are also "medium-n," having fewer observations than is needed for large-sample techniques to provide accurate approximations.

Moreover, this sample size problem afflicts more large-n studies than is generally recognized. Large-n datasets often contain units that are incomparable on measured confounding variables, and this lack of overlap between treatment and control units results in analyses that rely upon extrapolation for causal inference. We may guard against this by restricting a study to a smaller set of similar observations (Brady and Collier 2004) or by removing these incomparable observations by pre-processing the data through matching (Ho et al. 2007). But what often remains after limiting the scope of the analysis in this way is a medium-n study.

We present a set of methods to mitigate these problems and improve causal inferences in medium-n studies through a formal synthesis of qualitative information[1] and quantitative analysis. This synthesis is conducted within the Rosenbaum (2002; 2009) randomization inference-based approach to observational studies, which enables nonparametric inference with small sample sizes. We first demonstrate the basic technique using pairs of units that have been matched on measured confounders, as it simplifies the presentation and allows for an analogy to a repeated use of the comparative method (Lijphart 1975). We then show that these techniques can be extended to some of the more complicated matching strategies in Rosenbaum (2002; 2009).

This approach can integrate qualitative information with a quantitative analysis to improve causal inference in three ways. First, we can ameliorate the effects of difficult-to-measure outcomes by converting qualitative information into ordinal measurement of outcomes *within* the matched sets, which can reduce p-values. Second, additional information on the ranks of the sizes of the absolute within-set differences, as well as information on the difficulty in constructing these ranks and signs, allows us to present *qualitative confidence intervals* – that is, qualitative descriptions of effect sizes that have the same properties as conventional confidence intervals.[2] Third, qualitative information on unmeasured confounders within matched sets facilitates a sensitivity analysis that is less conservative than the typical Rosenbaum-style sensitivity analysis. This approach is feasible because of the medium-n sample size and because results from nonparametric statistics help identify what information will provide the most leverage.

While this approach has many benefits, identifying the information that maximizes statistical power also identifies information that would maximize bias if mismeasured. Because our procedure

[1]By qualitative information, we mean descriptive case summaries that may be converted into ordinal measurements within a small subset of units.

[2]Nonmetric scaling is often not feasible with this amount of information because we rank only the matched sets and not all possible pairs (Kruskal 1964).

partially couples the measurement and analysis processes, it introduces opportunities to corrupt the analysis. We propose in the conclusion to minimize this threat by explicitly separating and outsourcing the measurement stage.

We focus on treatment effects for a binary treatment. It is straightforward, however, to adapt our methods for other causal questions such as treatment effects of continuous treatments or multiple treatments and interactions or for multiple outcomes in this framework. We refer readers to Rosenbaum (2009) for a discussion of these topics, or Caughey, Dafoe and Seawright (2013) for a recent approach to multiple outcomes.

We demonstrate these points through a medium-n study of whether using plurality rules in transitional presidential elections in sub-Saharan Africa in the 1990s increased the severity of opposition harassment in the period leading up to the election. The appendices present the qualitative information from the comparative cases studies that is incorporated into the analysis. We find evidence strongly suggestive of a positive effect of plurality rules on opposition harassment, even after accounting for threats to causal inference. With the full matching implemented in the penultimate section, our approach obtains a one-sided p-value of 4.2% with only 9 units, and a sensitivity analysis accounting for unmeasured confounding demonstrates that this p-value is unlikely to rise above 10%.

This method differs from existing approaches to "mixed methods" for bolstering quantitative analyses with qualitative case studies. In many of these approaches, case studies are used to illustrate an argument and provide a "plausibility check" (Dunning 2012; Fearon and Laitin 2008; George and Bennett 2005). Lieberman (2005) suggests a nested approach in which an unsatisfactory large-n analysis is followed by a model-building small-n analysis. QCA (Ragin 2000) provides a method that accommodates many comparisons and causal factors with a small sample size. Our approach differs from these approaches by formally incorporating qualitative information into a standard statistical framework. Our approach is also more flexible than other formalized procedures for integrating qualitative information, such as Herron and Quinn (2014), which assume binary outcomes or parametric models and often require the elicitation of Bayesian priors.

The paper proceeds as follows. The next section introduces our running example of transitional presidential elections in 1990s sub-Saharan Africa, the formal notation, and randomization inference for pair-matched binary outcome data. Then in each of the following sections, we introduce qualitative information to the analysis to elaborate on our formal mixed method procedure for improving causal inference in medium-n studies. We first incorporate within-pair and between-pair information on the outcome through the signed-rank statistic to generate p-values and qualitative confidence intervals. We then show how full matching and the Quade statistic can further reduce p-values and how qualitative information on unmeasured confounders reduces the conservativeness of Rosenbaum-style sensitivity analysis. The Supplementary Information (SI) presents `R` code for our analyses. The conclusion discusses implications for practice and guidelines for researchers using these methods.

An Illustrative Example and Notation

To demonstrate these methods, we explore the effect of plurality electoral rules on opposition harassment in multi-party presidential elections in sub-Saharan Africa in the 1990s that marked transitions away from authoritarian rule. These transitional elections were watershed events at which citizens of these countries, often for the first time in their lives, had the opportunity to replace an authoritarian incumbent at the ballot box. But they were also precarious moments in which incumbents might employ violence against the opposition in order to stay in power.

Twenty-four sub-Saharan countries held these transitional elections in the 1990s, and 4 of these 24 used plurality rules under which a candidate must obtain more votes than any other candidate in order to be declared the winner.[3] The other countries used some form of runoff rules, which stipulate that should no candidate meet a given vote share threshold (usually 50%) in the first round, weaker candidates are eliminated and the top two finishers compete in a second-round election.[4] This rule and other elements of the election framework were determined by the authoritarian incumbent, with varying degrees of input from opposition representatives and civil society groups through national conferences and constitutional review committees. Foreign constitutional scholars, social scientists, and other experts on democratic institutions were often sponsored by foreign donors' democracy promotion programs to offer advice (Nwajiaku 1994; van Cranenburgh 2011). As we elaborate below, we believe *ex ante* that plurality rules might increase opposition harassment. Our question is therefore whether using plurality rules raised the likelihood and intensity of opposition harassment in these countries' transitional elections.[5]

We begin with an incumbent authoritarian regime that has agreed to hold multi-party presidential elections in the face of pressures for political liberalization. The regime wants to hold onto power by having its favored candidate win the election, and to this end, it allocates its finite resources to a combination of opposition co-optation and harassment. We assume that harassment cannot reliably convert opposition supporters into voters for the regime's favored candidate, and that harassment can suppress voting by some but not all opposition supporters.[6]

[3] Although Nigeria's electoral rules did not have a provision to eliminate any candidates, we have not coded this country as a plurality country because only two political parties were permitted to compete in the elections. Including Nigeria as a plurality country in the analysis increases the statistical significance of all results.

[4] These four countries are Cameroon, Kenya, Malawi, and Tanzania. In Kenya, the winning candidate must also receive a minimum of 25% of the valid votes cast in at least 5 of the 8 provinces of the country.

[5] Our question is related to Shugart and Carey (1992), Jones (1995), Neto and Cox (1997), and Pérez-Liñan (2006), who consider the effect of presidential runoff rules on party system fragmentation and the stability of democracy, but not opposition harassment.

[6] In the wake of riots, strikes, and other costly collective actions that led the regime to accede to multi-party elections, the willingness of a substantial portion of the population to oppose the regime has been demonstrated and is common knowledge. Potential opposition candidates and voters may therefore be willing to endure some

While all are aware of widespread dissatisfaction with the regime, not enough information is available about support for specific challengers to the authoritarian incumbent to ensure Duvergerian coordination in the transitional elections. This means that under plurality rules, a potential challenger who does not have the resources to win a majority but might be able to win a plurality may compete in the election and divide opposition support, reducing the vote margin needed to win the election. For the incumbent authoritarian regime, this makes opposition harassment more likely to be decisive for the outcome of the election and an attractive strategy, particularly if the harassment can be targeted at the supporters of the opposition candidate who is likely to have the most support.

With a runoff provision, the incumbent authoritarian regime could try to place in the top two rather than to win a majority of votes cast in the first round. But this strategy is dangerous because the opposition would gain the opportunity to coordinate behind a single candidate for the second round and the regime's favored candidate may place third and be ineligible for the runoff election. Therefore the incumbent regime's strategy will be to try to win an outright majority in the first round by drawing potential challengers and their supporters into its coalition, which in turn encourages weak challengers to contest the election in order to be co-opted by the regime, even if they do not have the resources to muster a majority.[7] Opposition harassment could help the incumbent by reducing turnout and therefore the number of votes needed to comprise a majority, but resources would need to be diverted from co-optation. Moreover, unlike plurality rule under which harassment can change the threshold for an incumbent win, harassment does not change the requirement of a majority under runoff rules. This means that opposition harassment is relatively less effective than co-optation under runoff rules and is less likely to be decisive. Consequently, we expect plurality rules to lead to greater opposition harassment than runoff rules.

Note that an empirical study of this proposed plurality effect has several difficulties shared by many observational studies. In addition to the small sample size, we are likely to have significant unmeasured confounding because we do not know what information was available to the key actors who set the electoral rule or know how they weighed different considerations. In particular, strong opposition to the incumbent might have increased the amount of opposition harassment under either set of electoral rules and might also have increased the likelihood of using plurality rules. Moreover, and most basically, the outcome variable of opposition harassment is difficult to measure. The remainder of the article tackles these concerns.

Notation and First Analysis

We wish to make causal inferences regarding N_1 treated units ($T = 1$) and a comparable subset of $N_0 \geq N_1$ control units ($T = 0$). For illustrative purposes, we follow Rosenbaum (2002)

harassment to oust the authoritarian incumbent at the ballot box.

[7] Weaker opposition parties may also use the first round to assess and demonstrate their relative strengths before negotiating terms for an alliance in the second round (Arriola 2012; van de Walle 2006).

and initially assume that the N_1 treated units have been pair-matched without replacement to N_1 of the control units. We further assume that the outcome variable has been coded for pairs $s = 1, ..., N_1$ so that the outcome for the first unit in each pair is denoted Y_{s1} and the outcome for the second unit is denoted Y_{s2}. We define T_s to be the treatment condition for the first unit in each pair and $1 - T_s$ to be the treatment condition for the second unit in the pair. We also assume that causal effects are well defined for each individual unit as the difference between two potential outcomes or counterfactuals: the outcome if treatment had been received, $Y(1)$, and the outcome if control had been received, $Y(0)$. We also assume that the observed outcome Y is equal to the potential outcome corresponding to treatment T; the other potential outcome is unknown. Therefore, for pair s, $Y_{s1} = T_s \cdot Y_{s1}(1) + (1 - T_s) \cdot Y_{s1}(0)$ and $Y_{s2} = T_s \cdot Y_{s2}(0) + (1 - T_s) \cdot Y_{s2}(1)$.

For the $2 \cdot N_1$ units in the matching study, the causal effects are written as:

$$\tau_{s1} = Y_{s1}(1) - Y_{s1}(0), \text{ and}$$
$$\tau_{s2} = Y_{s2}(1) - Y_{s2}(0), \text{ for } s = 1, ..., N_1$$

Like many observational studies, we begin our analysis with data from a publicly available dataset. The National Elections across Democracy and Autocracy (NELDA) dataset (Hyde and Marinov 2012) covers our population of interest, and we draw on this dataset to code an outcome variable that takes the value 1 if the opposition is harassed in the run-up to the election, and 0 otherwise. Because weaker incumbents who face strong opposition and are more worried about obtaining a majority are probably less likely to adopt a runoff provision that demands a majority, we pair-match the four plurality countries ($T = 1$) to the four countries with runoff provisions ($T = 0$) that are the most comparable on predictors of this institutional choice. The SI discusses the data and matching details, but we highlight that the plurality countries were exactly matched on the basis of whether the transition follows civil conflict, whether the country had previous experience with military rule, and the level of protest during the transition period. They were also matched on ethnic fractionalization and the log of GDP per capita, two key variables in the democratization literature.[8] These four matched pairs are presented in Table 1, along with their potential outcomes. Note that these countries have been paired in previous comparative studies (Azevedo, 1995, for Cameroon-Gabon; Widner, 1994a, b, c for Kenya–Côte d'Ivoire; Posner, 2004, for Malawi-Zambia; Smith, 2005, for Tanzania–Guinea-Bissau).

First, as discussed above, the potential outcome under treatment is observed for the plurality countries, while the potential outcome under control is unknown. Analogously, the potential outcome under control is observed for the runoff countries, while the potential outcome under treatment is unknown. Second, we inspect the outcome variable only after we match control units to our treated units. Note that information on the outcome variable for the control units that are

[8] We defer discussion of other possible matching variables to the final analysis using full matching (SI).

Treated (Plurality)	$Y(1)$	$Y(0)$	Controls (Runoff)	$Y(1)$	$Y(0)$
Cameroon	1	?	Gabon	?	0
Kenya	1	?	Côte d'Ivoire	?	1
Malawi	1	?	Zambia	?	0
Tanzania	0	?	Guinea-Bissau	?	0

Table 1: *Potential Outcomes for Matched Pairs*

not matched does not contribute to our analysis. This significantly reduces the potential coding burden. If the NELDA dataset had not been available and we had to code the outcomes ourselves for even just the initial analysis, we would only have coded the outcome for these 8 countries in the matched pairs rather than all 24 countries. With the NELDA coding, the difference in outcomes between plurality and runoff countries is positive (2/4, the difference between 3/4 of plurality countries having $Y = 1$ and 1/4 of runoff countries having $Y = 1$), indicating that plurality electoral rules may have caused opposition harassment in these transitional presidential elections.

Because the sample size is small, even if we believe that the matching successfully removed confounding and that in each pair the treated unit and control unit had the same *ex ante* probability of being assigned to treatment, we wonder whether the result could simply be due to chance. A straightforward approach to answering this question is Fisherian randomization inference, which is discussed in detail in Rosenbaum (2002; 2009), and by Bowers and Panagopoulos (2009; 2011), Hansen and Bowers (2008), Ho and Imai (2006), and Keele, McConnaughy and White (2012) in political science. Later, we will consider the assumptions required to use randomization inference in observational studies. For now, consider the following hypothetical question: if we had flipped a coin for each pair to determine which unit would receive treatment and which unit would receive control, then would we find the evidence in the table convincing? This question is typically formalized with a test of the sharp null hypothesis of no effect for any unit:

$$H_0 : \tau_{s1} = \tau_{s2} = 0, \text{ for } s = 1, ..., N_1$$

Under this null hypothesis and an assumption of pairwise randomization, we can generate null distributions and p-values by permuting over all possible pairwise randomizations. For our example with four matched pairs, there are $2^4 = 16$ possible pairwise randomizations. Using McNemar's test for binary outcomes, a special case of a randomization test using a sign score statistic, and with no additional information on the outcome variable, we obtain a one-sided p-value of $4/16 = 0.25$. In the next section, we explain the logic behind these randomization tests with the signed-rank statistic for pair-matched data, and show that this approach allows us to incorporate qualitative information on the outcome variable to improve the analysis.

Using Qualitative Information on the Outcome

When it is not possible to accurately measure a one-dimensional interval variable on an interval scale, this outcome variable may be coded as dichotomous or ordinal. This coarse coding may be necessary when creating a multi-use data set, but it may waste available information and lead to the wrong conclusions in a particular analysis. In this section, we present a method for incorporating additional qualitative information on the outcome to improve inferences about whether a particular treatment has an effect and how large this effect may be. Applying this method decreases the p-value for our analysis.

In our example, opposition harassment may differ in whether the regime targeted opposition leaders or supporters or both, the number of people detained, their treatment, whether violence was used or only threatened, and the extent of any violence. Measurement of this variable may be improved by attending to these components, but we may lack consensus on how much weight each component should be given when constructing an overall measure of opposition harassment. Even if we agreed on the weighting, it may be difficult to obtain the data to construct and place each of these countries on a scale of the severity of opposition harassment. However, because we have matched treated units to control units, even small amounts of this information can increase the power of tests of the sharp null hypothesis.

Incorporating Within and Between-Pair Information

The signed-rank statistic uses the sign of the difference in outcomes for each pair ($sign(Y_{s1}-Y_{s2})$ for $s = 1, ..., N_1$) and the ranks of the absolute values of the within-pair differences in the outcomes ($rank(abs(Y_{s1} - Y_{s2}))$ for $s = 1, ..., N_1$). The pair with the largest absolute difference in outcomes is assigned a rank of N_1, the pair with the smallest absolute difference in outcomes is assigned a rank of 1, and tied pairs are assigned an average of the ranks of those pairs. The statistic is:

$$W = \sum_{s=1}^{N_1} q_s[T_s s_{s1} + (1 - T_s)s_{s2}] \; = \sum_{s=1}^{N_1} W_s$$

where $s_{s1} = 1$ if $Y_{s1} > Y_{s2}$ and $= 0$ otherwise, $s_{s2} = 1$ if $Y_{s2} > Y_{s1}$ and $= 0$ otherwise, and q_s is the rank for each pair.

For our running example, we must delve into the details of eight cases, but only as deeply as necessary to sign the difference in the outcomes for each pair and to rank the absolute differences in outcomes in each pair. Scholars may disagree on how much the numbers of deaths and the extent of violence each contribute to the overall assessment of the severity of opposition harassment, as long as they agree enough to produce the same signs and rankings of the absolute differences. Moreover, debates over the measurement of complex outcome variables need only be settled to the extent that they produce agreement on the signs and ranks, and the sensitivity of the analysis to such disagreements is discussed in the SI. Finally, as we discuss in SI Section C, it is straightforward to conduct a sign test if the ranks cannot be determined.

We have signs for the discordant pairs (pairs with different values of Y), but we need to determine the signs for the concordant pairs (pairs with the same value of Y). For concepts such as opposition harassment, a binary variable coded as zero does not necessarily indicate the complete absence of that phenomenon. There was certainly some opposition harassment in all of the countries coded with $Y = 0$ in Table 1, and not all countries with $Y = 0$ had the same level of opposition harassment. Similarly, two countries coded as $Y = 1$ may not have had similar levels of opposition harassment. By examining the cases in each concordant pair, we may be able to provide enough information to determine a non-zero sign on the pair. This may not be possible for some pairs, in which case the pair remains coded as a tie. However, reducing the number of concordant pairs through these limited comparative studies will improve the power of the test.

Consider the concordant pair with $Y = 0$ in our example (Tanzania–Guinea-Bissau). Tanzania and Guinea-Bissau were both coded as $Y = 0$ with a binary variable from the NELDA dataset. However, closer investigation shows that both had some opposition harassment at a level that often appears in accounts of transitional elections, though less than other countries that were coded $Y = 1$. In Tanzania, several people were killed in fighting between the ruling party and opposition parties, and two newspaper editors were detained on sedition charges after publishing letters critical of the government (U.S. Department of State 1995). Although the opposition could generally hold large public rallies without harassment on the mainland (Commonwealth Observer Group 1995, 16), the ruling party "intimidated and harassed the opposition and did not allow opposition rallies until 2 months prior to elections" on the island of Zanzibar (U.S. Department of State 1995). Election observers also noted reports of harassment and the occasional detention of local opposition supporters, but these were generally fairly minor incidents (Commonwealth Observer Group 1995, 15; AWEPA 1996, 14; U.S. Department of State 1995). In Guinea-Bissau, the incumbent initially resisted the formation of opposition parties by "delaying registration procedures and by police violence" (Rudebeck 2002, 116). Human rights reports note that in February 1992, "five members of an opposition party were beaten and then refused hospital treatment." In addition, "police and security forces harassed opposition forces with detentions and physical mistreatment" (U.S. Department of State 1992, 116). Because of the situation in Zanzibar, we assess Tanzania as having more opposition harassment than Guinea-Bissau, and we code $s_{41} = 1$ for this pair.

Similarly, consider the concordant pair with $Y = 1$ in our example (Kenya–Côte d'Ivoire). Additional information suggests a difference in the severity of opposition harassment, a difference greater than that between Tanzania and Guinea-Bissau which were coded $Y = 0$. The run-up to the 1992 transitional presidential elections in Kenya were marked by "widespread intimidation, kidnapping, robbing and bribing of opposition candidates" (Tordoff 1992, 58), and widespread problems of voters not appearing on the voters register (IRI 1993, 45). In addition, at least 50,000 people

Pair s	Treated (Plurality)	Control (Runoff)	q_s	s_{s1}	s_{s2}	observed W_s	alternate W_s
1	Cameroon	Gabon	3	1	0	3	0
2	Kenya	Côte d'Ivoire	4	1	0	4	0
3	Malawi	Zambia	2	1	0	2	0
4	Tanzania	Guinea-Bissau	1	1	0	1	0

Table 2: *Using Qualitative Information to Rank Differences in Outcomes within Matched Pairs.*

were internally displaced and hundreds killed in violence targeted at ethnic groups that were seen to be supportive of the opposition and making claims to land (Holmquist and Ford 1992, 103). In the run-up to the 1990 elections in Côte d'Ivoire, the ruling *Parti Démocratique de la Côte d'Ivoire* (PDCI) also harassed the opposition organized around Laurent Gbagbo of the *Front Populaire Ivoirien* (FPI), but to a much lesser extent than in Kenya. The PDCI pressured opposition newspapers and journalists, but several opposition newspapers were in circulation (U.S. Department of State 1990, 96). The opposition was able to hold many peaceful pro-democracy demonstrations and opposition meetings, although Gbagbo was at times prevented from making speeches at rallies (Widner 1991, 39). The police also broke up several political rallies with truncheons and tear gas, resulting in several dozen injuries (*Africa Research Bulletin* (*ARB*), Aug 1990 27:7, 9768; *ARB*, Sept 1990 27:8, 9799-800; *ARB*, Sept 1990 27:9, 9826-27).

Moreover, by looking more closely at these cases, we determine that the Kenya–Côte d'Ivoire pair has the largest rank, followed by Cameroon–Gabon, Malawi–Zambia, and finally the Tanzania-Guinea-Bissau pair.[9] Descriptions of these pairs and more details on our rankings are in the Appendix. In each pair, the treated country had more opposition harassment than its paired control country ($Y_{s1} > Y_{s2}$) so that $s_{s1} = 1$ and $s_{s2} = 0$ for these pairs. Table 2 presents the proposed ranks, with observed W_s for the first unit in each pair being treated ($T_s = 1$) and alternate W_s for if the second unit in each pair had been treated ($T_s = 0$).

Table 3 presents the permutation distribution for the signed-rank test for the four pairs under the sharp null hypothesis. The first row corresponds to the observed data, with $W = 10$. No other value of W within the table is as large as the observed $W = 10$, and hence the one-sided p-value is 1/16. Note how much leverage was gained from just these signs and ranks, without full interval measures. And even if interval measures of the outcome were available, we might still use the signed-rank statistic because it provides robust power with non-normal outcomes (Rosenbaum 2002; 2009). Moreover, disagreements regarding the signs and ranks can be accommodated with a sensitivity analysis that calculates p-values for all plausible signs and ranks, and the p-value will be relatively robust to many such disagreements (see SI Section D).

[9] It may seem strange that the largest difference is between countries that were both coded as $Y = 1$, but this merely indicates the severity of opposition harassment in Kenya.

	Pair 1	Pair 2	Pair 3	Pair 4	Pair 1	Pair 2	Pair 3	Pair 4	
					$q_1 = 3$	$q_2 = 4$	$q_3 = 2$	$q_4 = 1$	
					$s_{11} = 1$	$s_{21} = 1$	$s_{31} = 1$	$s_{41} = 1$	
					$s_{12} = 0$	$s_{22} = 0$	$s_{32} = 0$	$s_{42} = 0$	
Permutation	T_1	T_2	T_3	T_4	W_1	W_2	W_3	W_4	W
1	1	1	1	1	3	4	2	1	10
2	1	1	1	0	3	4	2	0	9
3	1	1	0	1	3	4	0	1	8
4	1	1	0	0	3	4	0	0	7
5	1	0	1	1	3	0	2	1	6
6	1	0	1	0	3	0	2	0	5
7	1	0	0	1	3	0	0	1	4
8	1	0	0	0	3	0	0	0	3
9	0	1	1	1	0	4	2	1	7
10	0	1	1	0	0	4	2	0	6
11	0	1	0	1	0	4	0	1	5
12	0	1	0	0	0	4	0	0	4
13	0	0	1	1	0	0	2	1	3
14	0	0	1	0	0	0	2	0	2
15	0	0	0	1	0	0	0	1	1
16	0	0	0	0	0	0	0	0	0

Table 3: *Permutation Distribution for the Signed-Rank Statistic Using Within- and Between-Pair Qualitative Information to Supplement the NELDA Data*

Qualitative Confidence Intervals

Having produced a p-value of .0625, we would like to have a more descriptive representation of plausible sizes for the effect. This is typically a confidence interval in quantitative analyses, and if we had a continuous measure of the outcome variable Y, we could form confidence intervals within the randomization inference framework on the basis of the null hypotheses that we fail to reject (Rosenbaum 2002; 2009). We describe below the procedure for producing such confidence intervals if Y could be measured as a continuous variable, and then discuss forming *qualitative* confidence intervals with qualitative descriptions of the cases.

If Y can be measured, the first step is to alter the null hypothesis by assuming an effect size for each unit in the study. The most straightforward approach is to assume that the effect takes a constant value c for all units, and we use this approach for confidence intervals throughout.

$$H_0 : \tau_{s1} = \tau_{s2} = c, \text{ for } s = 1, ..., N_1$$

We are interested in positive effects for our example, so we start by considering small positive values of c. We can test the adjusted null hypothesis for a fixed value of c at an α level equal to the p-value by subtracting c from Y for the treated units and re-calculating the p-value. For our analysis, this means adjusting Y for the plurality countries such that $Y^*_{Cameroon} = Y_{Cameroon} - c$, for example. We can calculate the p-value as described in the previous section using Y^* for the plurality countries and Y for the runoff countries. We repeat this process until we find the smallest value of c that leads to an increase in the p-value. This value of c, which we denote c_Δ, represents the lower bound of the one-sided $(1 - p\text{-value})\%$ confidence interval. This will be a 93.75% CI for our analysis. Note that this procedure works with statistics other than the signed-rank statistic, although the p-value and the c_Δ will depend on the statistic chosen.

Although we do not have a continuous measure for Y, the signs and ranks will provide enough information to identify the cases that define this value of c_Δ for a simplified version of the signed-rank statistic known as the sign statistic.

$$V = \sum_{s=1}^{N_1} [T_s s_{s1} + (1 - T_s) s_{s2}] \;= \sum_{s=1}^{N_1} V_s$$

where again $s_{s1} = 1$ if $Y_{s1} > Y_{s2}$ and $= 0$ otherwise, $s_{s2} = 1$ if $Y_{s2} > Y_{s1}$ and $= 0$ otherwise. In our study, because all of the treated countries have greater opposition harassment than their paired control countries, the sign statistic produces the same p-value of .0625 as the signed-rank statistic (see the SI for a fuller discussion of the sign statistic).[10]

In general, a tie for any "positive sign" pair (e.g., when $T_s = 1$ and $Y_{s1} > Y_{s2}$ but $Y^*_{s1} = Y_{s2}$) will decrease the sign statistic and increase the p-value. This means that the smallest ranked pair with positive sign from the signed-rank statistic will be the first to tie as we increase c. In our study, the Tanzania–Guinea-Bissau pair has the smallest absolute difference in outcomes among the positive sign pairs, so c_Δ is the c that ties this pair. The lower bound of the 93.75% one-sided confidence interval, c_Δ, is therefore the difference in the severity of opposition harassment between Tanzania and Guinea-Bissau, $Y_{Tanzania} - Y_{Guinea-Bissau}$.

We cannot provide a quantitative description of the difference in harassment intensity between these two countries since quantitative measurements of $Y_{Tanzania}$ and $Y_{Guinea-Bissau}$ are unavailable. However, we have a qualitative description of this difference from the previous section summarized in Table 4. The major difference in opposition harassment between the two countries was that Tanzania banned opposition rallies on Zanzibar, while the opposition in Guinea-Bissau did not face such restrictions.

Characterizing this difference may be more difficult when the units are not countries, but instead provinces or even individuals, for which data is less accessible. Even if it were possible to determine the signs to calculate p-values, descriptions of Y may be less precise for those units and produce less useful and less easily replicable qualitative confidence intervals.

[10] Although they produce the same p-value in our example, in general the p-values may be different because the signed-rank statistic incorporates more information from the data.

Opposition Harassment (Y) $Y_{\text{Tanzania}} - Y_{\text{Guinea-Bissau}}$	93.75% One-Sided Confidence Interval
The major difference in opposition harassment was that Tanzania banned opposition rallies on Zanzibar, while the opposition in Guinea-Bissau did not face such restrictions. (See main text for additional details.)	$[Y_{\text{Tanzania}} - Y_{\text{Guinea-Bissau}}, \infty)$

Table 4: *93.75% One-Sided Qualitative Confidence Interval.* The difference in opposition harassment between Tanzania and Guinea-Bissau represents the lower bound on the 93.75% one-sided confidence interval for $\tau = \tau_{s1} = \tau_{s2}$ for all $s = 1, ..., N_1$.

Using Qualitative Information to Improve Full Matching

Our discussion has so far focused on pair matching, but we may benefit from having a variable number of treated and control units in each matched set through full matching (Hansen 2004). We show that, as discussed in Hansen (2004), we can reduce mismatches and balance on the measured confounders by allowing more general matched sets. We also demonstrate how full matching allows us to include additional units to increase power and reduce sensitivity to unmeasured confounders.

Using Qualitative Information on the Outcome with Full Matching

With pair matching, we matched 4 control units to our 4 treated units, however, we can often improve our matches and the power of our analysis by including additional control units. Table 5 presents a full matching analysis where Madagascar has been included as an additional control unit.[11] Notice that full matching allows us not only to include an additional control unit in the analysis, but to match all former French colonies to other former French colonies, so that we no longer have a former French colony (Côte d'Ivoire) as a control for a former British colony (Kenya). We continue to allow a former British colony (Tanzania) to be matched to a former Portuguese colony (Guinea-Bissau), since this mismatch should only lead to bias against our hypothesis due to poorer overall governance and greater reliance on force in former Portuguese colonies. More generally, if only positive effects are of interest, then mismatches that might produce negative bias can be ignored (Rosenbaum and Silber 2009).

Full matching can reduce mismatches, but it also rules out the use of the signed-rank statistic. We can use Quade's statistic (Quade, 1979; Rosenbaum 2002, 161), a straightforward generalization of the signed-rank statistic that uses both within-set and between-set ranks, in its place. With pair matching, there were $n_s = 2$ units within each set s, but now we allow each set s to have arbitrary n_s units. Within each set s, units $j = 1, ..., n_s$ are assigned ranks from 1 to n_s according to the size of the outcomes Y_{sj}. With $S \leq N_1$ sets, we write these within-set ranks as r_{sj} for $s = 1, ..., S$ and

[11] SI Section E presents the details of the full matching procedure.

Set s	Treated (Plurality)	Control (Runoff)	q_s	r_{s1}	r_{s2}	r_{s3}	r_{s4}	observed Q_s
1	Cameroon	Gabon, Côte d'Ivoire, Madagascar	2	4	3	2	1	8
2	Kenya, Malawi	Zambia	3	3	2	1	NA	15
3	Tanzania	Guinea-Bissau	1	2	1	NA	NA	2

Table 5: *Using Qualitative Information for Full Matching.*

$j = 1, ..., n_s$. For the electoral rule example, Cameroon is listed first ($j = 1$) and has the largest Y of four countries in the first set ($n_1 = 4$), so $r_{11} = 4$. Table 5 presents these within-set ranks for our electoral rule example. See the Appendix for details on the ranking. Note that r_{24}, r_{33}, and r_{34} are not defined because there are only three countries in the $s = 2$ set and two countries in the $s = 3$ set.

As before, the S sets are assigned ranks from 1 to S, which we write as q_s for $s = 1, ..., S$. However, because n_s can now be larger than 2, the between-set ranks q_s are determined by the absolute values of the differences between the maximum and minimum outcomes in the group ($rank(abs(max_j\{Y_{sj}\} - min_j\{Y_{sj}\}))$ for $s = 1, ..., S$). This means that the ranks are determined by $abs(Y_{Cameroon} - Y_{Madagascar})$, $abs(Y_{Kenya} - Y_{Zambia})$, and $abs(Y_{Tanzania} - Y_{Guinea-Bissau})$ for our analysis. Finally, because we allow more than one treated and/or control unit within each group, we define T_{sj} to be a treatment indicator for the jth unit in set s, such that $T_{sj} = 1$ if that unit receives treatment and $T_{sj} = 0$ if not. With these definitions the Quade statistic can be written as:

$$Q = \sum_{s=1}^{S} q_s \sum_{j=1}^{n_s} T_{sj} r_{sj} = \sum_{s=1}^{S} Q_s.$$

where $Q_s = q_s \sum_{j=1}^{n_s} T_{sj} r_{sj}$.

If we define m_s to be the number of treated units in set s ($\sum_{j=1}^{n_s} T_{sj} = m_s$), then conditional on $\{q_s, n_s, m_s, r_{sj}\}$ for $s = 1, ..., S$, the permutation distribution for Quade's statistic can be derived in a manner analogous to the permutation distribution for the signed-rank statistic. Table 6 presents this distribution. The observed data (first row) has the largest value of Quade's statistic, and because there are now 24 rows in the table, the randomization p-value is 1/24.

We can form qualitative confidence intervals like with the signed-rank statistic by using a version of the Quade statistic that does not use between-set ranks. This statistic is known as the stratified rank sum statistic:

$$SRS = \sum_{s=1}^{S} \sum_{j=1}^{n_s} T_{sj} r_{sj}.$$

As with the sign statistic, the stratified rank sum statistic will decrease when c increases to the

	Set 1	Set 2	Set 3	Set 1	Set 2	Set 3	
				$q_1 = 2$ $r_{11} = 4$ $r_{12} = 3$ $r_{13} = 2$ $r_{14} = 1$	$q_2 = 3$ $r_{21} = 3$ $r_{22} = 2$ $r_{23} = 1$	$q_3 = 1$ $r_{31} = 2$ $r_{32} = 1$	
Permutation	$T_{11}, T_{12}, T_{13}, T_{14}$	T_{21}, T_{22}, T_{23}	T_{31}, T_{32}	Q_1	Q_2	Q_3	Q
1	1,0,0,0	1,1,0	1,0	8	15	2	25
2	1,0,0,0	1,1,0	0,1	8	15	1	24
3	1,0,0,0	1,0,1	1,0	8	12	2	22
4	1,0,0,0	1,0,1	0,1	8	12	1	21
5	1,0,0,0	0,1,1	1,0	8	9	2	19
6	1,0,0,0	0,1,1	0,1	8	9	1	18
7	0,1,0,0	1,1,0	1,0	6	15	2	23
8	0,1,0,0	1,1,0	0,1	6	15	1	22
9	0,1,0,0	1,0,1	1,0	6	12	2	20
10	0,1,0,0	1,0,1	0,1	6	12	1	19
11	0,1,0,0	0,1,1	1,0	6	9	2	17
12	0,1,0,0	0,1,1	0,1	6	9	1	16
13	0,0,1,0	1,1,0	1,0	4	15	2	21
14	0,0,1,0	1,1,0	0,1	4	15	1	10
15	0,0,1,0	1,0,1	1,0	4	12	2	18
16	0,0,1,0	1,0,1	0,1	4	12	1	17
17	0,0,1,0	0,1,1	1,0	4	9	2	15
18	0,0,1,0	0,1,1	0,1	4	9	1	14
19	0,0,0,1	1,1,0	1,0	2	15	2	19
20	0,0,0,1	1,1,0	0,1	2	15	1	18
21	0,0,0,1	1,0,1	1,0	2	12	2	16
22	0,0,0,1	1,0,1	0,1	2	12	1	15
23	0,0,0,1	0,1,1	1,0	2	9	2	13
24	0,0,0,1	0,1,1	0,1	2	9	1	12

Table 6: *Permutation Distribution for Quade's Statistic Using Within- and Between-Set Qualitative Information to Supplement the NELDA Data, as well as an additional former French colony as a control unit.*

point that a higher ranked treated unit is tied with a lower ranked control within a set. In our study, we assess the difference in opposition harassment between Tanzania and Guinea-Bissau to be the smallest of any treatment and control comparisons within a set, where the treated unit is ranked higher than the control unit. If they became tied in rank, the p-value would increase. As discussed in the previous section, the Tanzania–Guinea-Bissau difference now defines the lower bound of a one-sided $(1 - p\text{-value})\%$ confidence interval. For this example, the p-value for the stratified rank sum statistic equals the p-value for the Quade statistic so this is a $(1 - 1/24)\% \approx .958\%$ confidence interval, but the p-values from the two statistics will generally not be equal.

Using Qualitative Information on Unmeasured Confounders in Full Matching

The analysis so far has been predicated on hypothetical coin flips or dice rolls within comparable sets of units. However, the units may be somewhat incomparable on a confounder so that our results actually reflect differences in this variable rather than the effect of the treatment. Fortunately, qualitative information can be used in a couple of ways to address this problem. First, qualitative information might uncover the presence of unmeasured confounders. In this example, initial research suggested that a potential confounding factor – whether a prominent opposition figure had long been in exile and might create rifts within the domestic opposition movement upon his return – was more important than originally suspected. Second, if an unmeasured confounder is discovered, then qualitative information can sometimes be used to assess the effects of the confounding. This can be accomplished by changing the thought experiment to employ a weighted coin or die, since the existence of an unmeasured confounder implies that at least one unit in a set may have been *ex ante* more likely to have received treatment than the others.

To formalize the sensitivity analysis with full matching, it is most straightforward to define $\mathbf{T_s}$ as the random vector of treated indexes from set s, drawn from the set Ω_s of possible assignments. The number of such possible assignments is $|\Omega_s| = \binom{n_s}{m_s}$. The parameter $\pi_{s,\{t_1,...,t_{n_s}\}}$ is the *ex ante* probability of realizing the vector of treatments $\mathbf{t_s} = \{t_1, ..., t_{n_s}\}$ in set s. We define $\mathbf{X_s}$ and $\mathbf{U_s}$ to be the matrices of measured and unmeasured confounders, respectively, for all units within set s, such that this *ex ante* probability can be written as the following:[12]

$$\pi_{s,\{t_1,...,t_{n_s}\}} = Pr(\mathbf{T_s} = \{t_1, ..., t_{n_s}\} | n_s, m_s, \mathbf{X_s} = \{X_1, ..., X_{n_s}\}, \mathbf{U_s} = \{U_1, ..., U_{n_s}\})$$

Formally, when the units in set s have the same values of the measured and unmeasured confounders $\{X_1 = ... = X_{n_s}\}$ and $\{U_1 = ... = U_{n_s}\}$, then $\pi_{s,\{t_1,...,t_{n_s}\}} = \frac{1}{\binom{n_s}{m_s}}$. If this holds for our example, then $\frac{1}{\binom{n_1}{m_1}} = \frac{1}{4}$ for $s = 1$, $\frac{1}{\binom{n_2}{m_2}} = \frac{1}{3}$ for $s = 2$, and $\frac{1}{\binom{n_3}{m_3}} = \frac{1}{2}$ for $s = 3$.

[12] In other words, $\mathbf{X_s}$ and $\mathbf{U_s}$ must be sufficient such that $\pi_{s,\{t_1,...,t_{n_s}\}}$ does not also depend on the potential outcomes for the set.

One concern is that the units are not equal on unmeasured confounders (i.e., $\{U_1 \neq ... \neq U_{n_s}\}$). The standard approach to sensitivity analysis in this situation is to propose a range of plausible values for $\pi_{s,\{t_1,\ldots,t_{n_s}\}} \neq \frac{1}{\binom{n_s}{m_s}}$, and to check what happens to the p-values when those probabilities change. Unfortunately, producing a range of plausible values for $\pi_{s,\{t_1,\ldots,t_{n_s}\}}$ may be quite difficult, so researchers often present a series of increasing and decreasing values, leaving the burden of assessing plausibility to the reader. This process may be simplified by using a single sensitivity parameter (Rosenbaum 2002), but at the cost of a conservative analysis. However, we show that qualitative information may be used to make concrete inequality statements about $\pi_{s,\{t_1,\ldots,t_{n_s}\}}$ for some of the sets s, resulting in a less conservative sensitivity analysis. For the sets s where we cannot make specific inequality statements, finding plausible values of $\pi_{s,\{t_1,\ldots,t_{n_s}\}}$ remains as difficult as in a standard Rosenbaum-style sensitivity analysis.

In our example, we may be concerned that the matching variables discussed in SI Section B do not fully capture the strength of opposition, the key variable affecting the outcome that we believe makes an incumbent authoritarian regime more likely to adopt plurality rules.[13] Specifically, we assume that greater opposition strength increases the probability that an individual country will be treated with plurality. Further comparative case studies allow us to assess the relative strength of opposition, and hence give a sense of whether $\pi_{s,\{t_1,\ldots,t_{n_s}\}}$ is greater or less than $\frac{1}{\binom{n_s}{m_s}}$ for each possible treatment allocation in set s. Because the observed data provides the maximum value of $Q = 25$ (row 1 of Table 6), our sensitivity analysis need only focus on that row of the table. This means that we need to consider $\pi_{1,\{1,0,0,0\}}$, $\pi_{2,\{1,1,0\}}$, and $\pi_{3,\{1,0\}}$.

For the former French colonies set ($s = 1$), we judge Cameroon to have greater opposition strength than Gabon or Côte d'Ivoire (see the Appendix). Hence we believe that $1 \geq \pi_{1,\{1,0,0,0\}} \geq \pi_{1,\{0,1,0,0\}} \geq \pi_{1,\{0,0,1,0\}} \geq 0$, or in other words, that Cameroon was more likely to have received treatment than Gabon or Côte d'Ivoire, and $\pi_{1,\{1,0,0,0\}}$ is potentially greater than $\frac{1}{4}$. However, one of the benefits of including Madagascar in the analysis is that we judge Madagascar to have greater opposition strength than Cameroon (see the Appendix) and therefore $\pi_{1,\{0,0,0,1\}} \geq \pi_{1,\{1,0,0,0\}}$. This implies that $\pi_{1,\{1,0,0,0\}}$ can equal its randomization probability of $\frac{1}{4}$ for a variety of different values of $\pi_{1,\{0,1,0,0\}}$, $\pi_{1,\{0,0,1,0\}}$, and $\pi_{1,\{0,0,0,1\}}$, and in particular, for values of $\pi_{1,\{0,0,1,0\}} \leq \pi_{1,\{0,1,0,0\}} < 1/4$. It also means that $\pi_{1,\{1,0,0,0\}} \leq 1/2$, and therefore our sensitivity analysis will be less conservative than an analysis that allows $1/2 < \pi_{1,\{1,0,0,0\}} \leq 1$.

In set $s = 2$, our assessment of opposition strength leaves us unconcerned about any mismatch. We judge Zambia (control) to have greater opposition strength than both Kenya and Malawi (both treated). This implies that $\pi_{2,\{1,1,0\}} \leq \frac{1}{3}$, but to be conservative we set this probability at 1/3. Finally, in set $s = 3$ we allow $\pi_{3,\{1,0\}}$ to take values between 1/2 and 3/4, although recall that we

[13] The sensitivity analysis relies only on changing the probabilities of treatment assignment and does not depend on the assumption used here that key unmeasured confounding filters through a single proximate confounder. However, justifying bounds on the probabilities becomes more complicated when this assumption does not hold.

have already discounted the effects of the mismatch in British-Portuguese colonial background for this pair, so assessment of the likely values of $\pi_{3,\{1,0\}}$ should not consider this difference.[14]

	$\pi_{3,\{1,0\}}$					
$\pi_{1,\{1,0,0,0\}}$	1/2	1.25/2.25	1.5/2.5	2/3	2.5/3.5	3/4
1/4	0.042	0.046	0.050	0.056	0.60	0.063
1.25/4.25	0.049	0.054	0.059	0.065	0.070	0.074
1.5/4.5	0.056	0.062	0.067	0.074	0.079	0.083
2/5	0.067	0.074	0.080	0.089	0.095	0.100
2.5/5.5	0.076	0.084	0.091	0.101	0.108	0.114
3/6	0.083	0.093	0.100	0.111	0.119	0.125

Table 7: *Sensitivity analysis on maximum p-values with qualitative information included on the unmeasured confounder within full matching.* This analysis assumes that because of the assessment of opposition strength, the *ex ante* probability of Kenya and Zambia being the treated units in the second set is at most 1/3 (i.e., $\pi_{2,\{1,1,0\}} \leq 1/3$).

The sensitivity analysis based on these numbers is presented in Table 7, with increasing values of $\pi_{1,\{1,0,0,0\}}$ and $\pi_{3,\{1,0\}}$ corresponding to increasing p-values. Notice that if $\pi_{1,\{1,0,0,0\}} \leq 1/4$, and $\pi_{3,\{1,0\}} \leq 1.5/2.5$, then the p-value is at most 5%. Furthermore, the maximum p-value based on the upper bound for $\pi_{1,\{1,0,0,0\}}$ and a value of $\pi_{3,\{1,0\}} = 3/4$ still provides a p-value of 0.125. Note also that without the additional information on opposition strength we would need to consider values of $\pi_{1,\{1,0,0,0\}} > 1/2$ and $\pi_{2,\{1,1,0\}} > 1/3$, which would increase the p-value. For example, with $\pi_{3,\{1,0\}} = 3/4$ and $\pi_{1,\{1,0,0,0\}} = 1/2$ $\pi_{3,\{1,0\}} = 3/4$, as in the bottom right corner of the table, if $\pi_{2,\{1,1,0\}} = 2/3$ instead of 1/3, the p-value would have been 0.25.

Conclusion

For many questions in political science, researchers face the challenges of difficult-to-measure outcomes, imbalance on measured and unmeasured confounders, and small sample size after removing incomparable units from the study. Analyses of the effects of country-level institutions on large-scale social or political outcomes are particularly vulnerable to these problems, since these institutions are generally chosen endogenously through complex political processes and the pop-

[14] Within the two parameter amplification of the sensitivity analysis (Rosenbaum and Silber 2009), this can be formalized for the Tanzania–Guinea-Bissau pair in two steps. First, we can combine in the parameter λ the positive effects of the mismatch in British-Portuguese colonial background on Tanzania receiving the treatment with the potentially positive effects of an opposition strength mismatch on Tanzania receiving the treatment. Second, we can combine in the parameter δ the negative effects of the mismatch in British-Portuguese colonial background on the outcome difference under the control condition and the potentially positive effects of an opposition strength mismatch on the outcome difference under the control condition. Then we can write $\pi_{3,\{1,0\}} = \frac{exp(\lambda+\delta)+1}{(1+exp(\delta))(1+exp(\lambda))}$. Intuitively, we can make λ relatively large and δ relatively small to incorporate our qualitative knowledge about this mismatch on colonial background.

ulation of units is limited. But, as this paper has demonstrated, the small sample sizes of these observational studies makes feasible the use of qualitative information to improve causal inferences.

In our analysis of the effect of presidential electoral rules on opposition harassment in African countries undergoing regime transition in the 1990s, comparative case studies allowed us to rank within- and between-set differences in the severity of opposition harassment and rank the direction of within-set differences in unmeasured strength of opposition. The techniques described in this paper provide a principled way in which to use the qualitative information we learned from these brief case studies to improve our analysis. We showed that by incorporating case knowledge within the Rosenbaum (2002; 2009) approach, we could improve power and potentially reduce p-values, provide qualitative confidence intervals, and reduce sensitivity to unmeasured confounders.

By showing how and how much additional information can improve causal inference, we offer statistically-grounded guidelines for how mixed methods researchers should direct their efforts in data collection for small- and medium-n studies. The first step is to understand the treatment assignment process and to focus on measurement of the more important matching variables rather than improving measurement of the outcome variable. Then after matching with these variables, researchers should focus on signing and ranking differences in outcomes within concordant pairs or sets in order to, but do no more than, establish within-set and between-set rankings. Existing datasets can be very helpful starting points for both of these steps, and researchers need only to focus on the outcome variable for the cases in the matched sets and not the entire sample. Our methods also point to which set of cases is likely to define the bounds of a qualitative confidence interval for some specified level, so that the researcher can focus on characterizing more precisely the difference in outcomes among those likely cases. Finally, deep knowledge beyond the information encoded in quantitative datasets should be used to assess the relative probabilities of treatment assignment within these sets in order to strengthen and clarify the credibility of a study.

One concern is that the link between the analysis and the partial coding may allow researchers to tailor their analyses to obtain particular results. This hazard can be reduced by outsourcing the partial coding decisions. Experts who do not know the treatment variable of interest could be tasked with measurement of the outcome variable, and other experts who do not know the outcome variable of interest could be assigned to code the unmeasured confounders. This procedure would effectively decouple the analysis from the coding required for the analysis and enable the formal registration of the study as discussed in Humphreys, de la Sierra and van der Windt (2013). Transparency about the matching procedure and careful documentation of the sources used in the comparative case studies will also enable replication and scrutiny of the researcher's coding by others (Lieberman 2010).

More generally, we have shown that even with a small sample size, randomization inference allows qualitative information to be incorporated in a nonparametric statistical framework. Unlike other mixed-methods approaches, our method formally integrates qualitative information with quantitative analysis. Moreover, the formal synthesis does not require parametric assumptions or the elicitation of Bayesian priors. This allowed us to provide evidence suggesting that plurality rules may increase the severity of opposition harassment and to characterize the lower bound for the size of that effect. That we obtained this result with only four countries with plurality rules points to potential gains from expanding the study.

References

Africa Research Bulletin. N.d. Various issues.

Arriola, Leonardo R. 2012. *Multiethnic Coalitions in Africa.* New York: Cambridge University Press.

AWEPA. 1996. *Tanzania's Multi-party Elections.* Amsterdam: African-European Institute.

Azevedo, Mario. 1995. Ethnicity and Democratization: Cameroon and Gabon. In *Ethnic Conflict and Democratization in Africa*, ed. Harvey Glickman. Atlanta, GA: African Studies Association Press pp. 255–88.

Bayalama, Sylvain. 1991. "Pluralism and Political Change in Central Africa." *Africa Today* 38(3):66–71.

Bowers, Jake and Costas Panagopoulos. 2009. ""Probability of What?": A Randomization-based Method for Hypothesis Tests and Confidence Intervals about Treatment Effects." Unpublished manuscript.

Bowers, Jake and Costas Panagopoulos. 2011. "Fisher's randomization mode of statistical inference, then and now." Unpublished manuscript.

Brady, Henry E. and David Collier, eds. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* New York: Rowman & Littlefield.

Cardoso, Carlos. 1994. "A Transição Democrática Na Guiné-Bissau: Um Parto Difícil." *Soronda: Revista Estudos Guineeses* 17:5–30.

Caughey, Devin, Allan Dafoe and Jason Seawright. 2013. "Testing Elaborate Theories in Political Science: Nonparametric Combination of Dependent Tests." Unpublished Manuscript, Presented at the 71st Midwest Political Science Association Annual National Conference, April 13th, 2013.

Commonwealth Observer Group. 1995. The Union Presidential and Parliamentary Elections in Tanzania, 29 October 1995. Technical report.

Dunning, Thad. 2012. *Natural experiments in the social sciences.* New York: Cambridge University Press.

Erdmann, Gero and Neo Simutanyi. 2003. "Transition in Zambia: The Hybridisation of the Third Republic.". Konrad Adenauer Foundation, Occasional Papers.

Fearon, James D. and David D. Laitin. 2008. Integrating Qualitative and Quantitative Methods. In *The Oxford Handbook of Political Methodology*, ed. Janet M. Box-Steffensmeier, Henry E. Brady and David Collier. New York: Oxford University Press pp. 756–76.

Forrest, Joshua B. 2005. Democratization in a Divided Urban Political Culture: Guinea-Bissau. In *The Fate of Africa's Democratic Experiments.* Bloomington: Indiana University Press pp. 246–66.

Gardinier, David E. 1997. Gabon: Limited Reform and Regime Survival. In *Political Reform in Francophone Africa*, ed. John F. Clark and David E. Gardinier. Westview Press chapter 9, pp. 145–61.

George, Alexander L. and Andrew Bennett. 2005. *Case studies and theory development in the social sciences*. Cambridge, MA: MIT Press.

Hansen, Ben B. 2004. "Full matching in an observational study of coaching for the SAT." *Journal of the American Statistical Association* 99(467):609–618.

Hansen, Ben B. and Jake Bowers. 2008. "Attributing Effects to A Clustered Randomized Get-Out-The-Vote Campaign." *Journal of the American Statistical Association* 104(487):873–85.

Herron, Michael C. and Kevin M. Quinn. 2014. "A Careful Look at Modern Qualitative Case Selection Methods." *Sociological Methods and Research* . Forthcoming.

Ho, Daniel E. and Kosuke Imai. 2006. "Randomization Inference with Natural Experiments: An Analysis of Ballot Effects in the 2003 Election." *Journal of the American Statistical Association* 101:888–900.

Ho, Daniel E., Kosuke Imai, Gary King and Elizabeth A. Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15(3):199–236.

Holmquist, Frank and Michael Ford. 1992. "Kenya: Slouching toward Democracy." *Africa Today* 39(3):97–111.

Humphreys, Macartan, Raul Sanchez de la Sierra and Peter van der Windt. 2013. "Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration." *Political Analysis* 21(1):1–20.

Hyde, Susan D. and Nikolay Marinov. 2012. "Which Elections Can Be Lost?" *Political Analysis* 20(2):191–210.

Ihonvbere, Julius O. 1997. "From despotism to democracy: The rise of multiparty politics in Malawi." *Third World Quarterly* 18(2):225–247.

Jones, Mark P. 1995. *Electoral laws and the survival of presidential democracies*. Notre Dame, IN: Notre Dame University Press.

Kashin, Konstantin, Adam N. Glynn and Nahomi Ichino. 2014. *`qualCI` v0.1*. `R` package, `http://cran.r-project.org/web/packages/qualCI/index.html`.

Keele, Luke J., Corrine McConnaughy and Ismail White. 2012. "Strengthening the Experimenter's Toolbox: Statistical Estimation of Internal Validity." *American Journal of Political Science* 56(2):484–99.

Kenya: The December 29, 1992 Election. 1993. IRI.

Khapoya, Vincent. 1988. "Moi and beyond: Towards Peaceful Succession in Kenya?" *Third World Quarterly* 10(1):54–66.

Krieger, Milton. 1994. "Cameroon's Democratic Crossroads, 1990-4." *Journal of Modern African Studies* 32(4):605–628.

Kruskal, J.B. 1964. "Nonmetric Multidimensional Scaling: A Numerical Method." *Psychometrika* 29(2):115–129.

Lieberman, Evan S. 2005. "Nested Analysis as a Mixed-method Strategy for Comparative Research." *American Political Science Review* 99(3):435–52.

Lieberman, Evan S. 2010. "Bridging the Qualitative-Quantitative Divide: Best Practices in the Development of Historically Oriented Replication Databases." *Annual Review of Political Science* 13:37–59.

Lijphart, Arend E. 1975. "II. The Comparable-Cases Strategy in Comparative Research." *Comparative Political Studies* 8(2):158–77.

Malawi. 2002. In *Compendium of elections in Southern Africa*, ed. Tom Lodge, Denis Kadima and David Pottie. Johannesburg: Electoral Institute of Southern Africa.

Marcus, Richard R. 2004. "Political Change in Madagascar: Populist Democracy or Neopatrimonialism by Another Name?" Institute for Security Studies (South Africa) Occasional Paper 89, http://www.iss.co.za/pubs/papers/89/Paper89.htm, accessed July 6, 2012.

Mentan, Tatah. 1998. Cameroon: A Flawed Transition to Democracy. In *Democratization in Late Twentieth-Century Africa: Coping with Uncertainty*, ed. Jean-Germain Gros. Greenwood Press pp. 41–58.

Messone, Nelson N. and Jean-Germain Gros. 1998. The Irony of Wealth: Democratization in Gabon. In *Democratization in Late Twentieth-Century Africa: Coping with Uncertainty*, ed. Jean-Germain Gros. Greenwood Press pp. 41–58.

Mwase, Ngila and Mary Raphael. 1997. The Presidential Elections. In *Multiparty Democracy in Transition: Tanzania's 1995 General Elections*, ed. Samuel S. Mushi and Rwekaza S. Kumandala. Department of Political Science and Public Administration, University of Dar es Salaam pp. 149–84.

NDI and Carter Center. 1992. *The October 31 1991 National Elections in Zambia.* National Democratic Institute for International Affairs.

Neto, Octavio Amorim and Gary W. Cox. 1997. "Electoral Institutions, Cleavage Structures, and the Number of Parties." *American Journal of Political Science* 41(1):149–174.

Newell, Jonathan. 1995. "A Moment of Truth, the Church and Political Change in Malawi, 1992." *Journal of Modern African Studies* 33(2):243–62.

Nwajiaku, Kathryn. 1994. "The National Conferences in Benin and Togo Revisited." *Journal of Modern African Studies* 32(3):429–47.

Pérez-Liñan, Aníbal. 2006. "Evaluating presidential runoff elections." *Electoral Studies* 25:129–46.

Posner, Daniel N. 2004. "The Political Salience of Cultural Difference: Why Chewas and Tumbukas are Allies in Zambia and Adversaries in Malawi." *American Political Science Review* 98(4):529–46.

Quade, Dana. 1979. "Using weighted rankings in the analysis of complete blocks with additive block effects." *Journal of the American Statistical Association* 74(367):680–683.

Ragin, Charles C. 2000. *Fuzzy-set Social Science*. Chicago: University of Chicago Press.

Rosenbaum, Paul R. 2002. "Covariance Adjustment in Randomized Experiments and Observational Studies." *Statistical Science* 17:286–304.

Rosenbaum, Paul R. 2009. *Design of Observational Studies*. New York: Springer.

Rosenbaum, Paul R. and Jeffrey H. Silber. 2009. "Amplification of Sensitivity Analysis in Matched Observational Studies." *Journal of the American Statistical Association* 104(488):1398–1405.

Rudebeck, Lars. 2002. Multi-party Elections in Guinea-Bissau. In *Multi-party Elections in Africa*. Oxford: James Currey pp. 104–27.

Schraeder, Peter J. 1994. "Elites as Facilitators or Impediments to Political Development? Some Lessons from the "Third Wave" of Democratization in Africa." *Journal of Developing Areas* 29(1):69–90.

Shugart, Matthew Soberg and John M. Carey. 1992. *Presidents and Assemblies: Constitutional Design and Electoral Dynamics*. Cambridge: Cambridge University Press.

Smith, Benjamin. 2005. "Life of the Party: The Origins of Regime Breakdown and Persistence under Single-Party Rule." *World Politics* 57(3):421–51.

Takougang, Joseph. 1997. Cameroon: Biya and Incremental Reform. In *Political Reform in Francophone Africa*. Boulder, CO: Westview Press pp. 162–81.

Throup, David and Charles Hornsby. 1998. The Regime in Crisis, January 1990–October 1992. In *Multiparty Politics in Kenya: The Kenyatta and Moi State and the Triumph of the System in the 1992 Elections*. Madison: University of Wisconsin Press chapter 4, pp. 54–91.

Tordoff, Lord. 1992. "Multiparty elections in Kenya December 1992." *Representation* 31(115):57–59.

U.S. Department of State. 1990. *Country Reports on Human Rights Practices for 1990*.

U.S. Department of State. 1991. *Country Reports on Human Rights Practices for 1991*.

U.S. Department of State. 1992. *Country Reports on Human Rights Practices for 1992*.

U.S. Department of State. 1993. *1993 Human Rights Report: Central African Republic*.

U.S. Department of State. 1994. *Gabon Human Rights Practices, 1993*.

U.S. Department of State. 1995. *1995 Human Rights Report: Tanzania*.

van Cranenburgh, Oda. 1996. "Tanzania's 1995 Multi-Party Elections: The Emerging Party System." *Party Politics* 2(4):535–47.

van Cranenburgh, Oda. 2011. "Democracy promotion in Africa: the institutional context." *Democratization* 18(2):443–61.

Van de Walle, Nicolas. 2006. Tipping Games: When Do Opposition Parties Coalesce? In *Electoral Authoritarianism: The Dynamics of Unfree Competition.* Boulder, CO: Lynne Rienner Publishers.

van Donge, Jan Kees. 1995. "Kamuzu's Legacy – the democratization of Malawi." *African Affairs* 94(375):227–57.

VonDoepp, Peter. 1996. "Political transition and civil society: The cases of Kenya and Zambia." *Studies in Comparative International Development* 31(1):24–47.

Widner, Jennifer. 1991. "The 1990 Elections in Côte d'Ivoire." *Issue* 20(1):31–40.

Widner, Jennifer. 1994*a*. "Two Leadership Styles and Patterns of Political Liberalization." *African Studies Review* 37(1):151–174.

Widner, Jennifer A. 1994*b*. Political Reform in Anglophone and Francophone Africa. In *Economic Change and Political Liberalization in Sub-Saharan Africa*, ed. Jennifer A. Widner. Baltimore: Johns Hopkins University Press pp. 49–79.

Widner, Jennifer A. 1994*c*. "Single Party States and Agricultural Policies: The Cases of Ivory Coast and Kenya." *Comparative Politics* 26(2):127–147.

Opposition Harassment

Pair Matching

As noted earlier, the Kenya-Côte d'Ivoire pair has the largest absolute difference in outcomes. Cameroon–Gabon is the pair with next largest absolute difference in outcomes. Heavy intimidation of the opposition, including arrests and torture, preceded the October 1992 presidential elections in Cameroon (Mentan 1998, 44–46). About 300 people were killed in 1991 (Takougang 1997, 169), and at least 400 were killed in the two years leading up to the elections (Schraeder 1994, 81). In Gabon, transitional presidential elections in 1993 followed an extended period of demonstrations and strikes in urban areas and barricaded roads in rural areas (Gardinier 1997; Messone1998). Throughout 1993, government security services intimidated opposition media with electronic jamming and the destruction or confiscation of radio transmission equipment (U.S. Department of State 1994). The National Assembly approved presidential decrees that severely curtailed press freedoms and most opposition newspapers were banned (Gardinier 1997). The U.S. State Department reports that "[Police were] absent from some opposition gatherings which were disrupted by violence attributed to street gangs paid by rival parties," but "[d]uring the 2 weeks prior to the elections, police acted quickly and effectively to assure that demonstrations and confrontations between the opposition and the [ruling] PDG remained peaceful" (U.S. Department of State 1994).

The difference in opposition harassment between Malawi and Zambia was smaller. Malawi had repeated mass arrests of opposition members and opposition leaders were detained and prosecuted (*Africa Research Bulletin* (*ARB*), Apr 1992 29:4, 10548–50; *ARB*, June 1992 29:6, 10618–9; *ARB*, Jul 1992 29:7, 10659; *ARB*, Nov 1992 29:11, 10793–4). Lodge et al. (2002, 130) note that, "Both the UN Joint International Observer Group (JIOG) and the Malawi Electoral Commission (MEC) reported campaign violence and widespread intimidation, bribery and misuse of official positions." The ruling party's paramilitary wing rounded up opposition supporters on such a scale that "there was not enough room, and scores of detainees [had] to be held under guard in tents set up near Blantyre jail" (Ihonvbere 1997, 238, citing *ARB* 1992). Opposition harassment was also serious in Zambia (NDI and Carter Center, 1992, 44-45; *ARB*, Feb 1991 28:2, 10166; *ARB*, Sep 1991 28:9, 10284-5), but to a lesser extent than in Malawi. After political prisoners were released in July 1990, the government did not detain any additional opposition supporters. Ruling party supporters attacked and killed several opposition members, but in many instances the police arrested the attackers and generally allowed opposition rallies to be held (U.S. Department of State 1991, 453).

Tanzania and Guinea-Bissau has the smallest difference in outcomes of all the pairs. We find evidence of harassment over a more extended period in Tanzania than in Guinea-Bissau, but less than in Malawi.

Full Matching

We believe opposition harassment was greatest in Cameroon, second in Gabon, then Côte d'Ivoire, and least in Madagascar. We consider $Y_{Gabon} > Y_{Cote\ d'Ivoire}$ because opposition rallies were disrupted more frequently and opposition media was repressed more violently in Gabon than in Côte d'Ivoire. In Madagascar, once the opposition successfully pressured incumbent President Ratsiraka into holding a constitutional convention and multi-party elections, the election itself proceeded fairly smoothly (Marcus 2004). The only reported incident involves the army, which was loyal to the transitional government, which killed several Ratsiraka supporters while the latter demonstrated in favor of secession by several regions of the country (*ARB*, Oct 1992 29:10, 10759–60).

Opposition harassment was greater in Kenya, where hundreds were killed and tens of thousands displaced, than in Malawi, which had mass arrests and detention of opposition members. As discussed above, Malawi had more opposition harassment than in Zambia, so $Y_{Kenya} > Y_{Malawi} > Y_{Zambia}$. Moreover, because of the great extent of opposition harassment in Kenya, we believe $abs(Y_{Kenya} - Y_{Zambia}) > abs(Y_{Cameroon} - Y_{Madagascar}) > abs(Y_{Cameroon} - Y_{Cote\ d'Ivoire})$ and $abs(Y_{Cameroon} - Y_{Gabon}) > abs(Y_{Tanzania} - Y_{Guinea-Bissau})$. These relationships allow us to provide the ranks in Table 6 and to determine that the Tanzania–Guinea-Bissau pair defines the lower bound of the one-sided confidence interval.

Strength of Opposition

We may be concerned that the matching described in the SI does not sufficiently model treatment assignment. More specifically, the matching variables may not fully capture the strength of opposition which affects the choice of electoral rules, so that the probability of receiving treatment is not equal across the countries in each pair. Further comparative case studies allow us to assess the relative strength of opposition, and hence whether treatment was more likely for one case than the other in each pair.

Within the set of former French colonies, opposition strength was greatest in Madagascar, followed by Cameroon, and then Côte d'Ivoire and Gabon. The opposition in Madagascar organized a strike of nearly 80,000 civil servants and later a march of 400,0000 people on the city center, effectively a general strike. Its leader was able to declare a parallel government, proclaim himself prime minister, and push for a transitional government (Marcus 2004).

The Cameroonian opposition was also capable of organizing extended general strikes, but was less unified than in Malagascar. In Cameroon, the transition began with a security crackdown in 1990 on a large opposition rally in Bamenda, which led to the deaths of six protestors spurring further action. John Fru Ndi, an obscure former member of the governing party who had called for the rally, became a major opposition leader. The umbrella National Coordination of Opposition

Parties and Associations (NCOPA) and the unions began a general strike and implemented their "ghost town" strategy, demanding a national conference (Mentan 1998, 44-46). President Biya eventually conceded to this demand and the ruling CPDM started to fragment, but the opposition also fragmented (Krieger 1994, 608–12).

We believe that the opposition was somewhat stronger in Cameroon, which found a central leader, than in Gabon, which was more divided between a long-standing opposition group in exile and other in-country leaders. Like in Cameroon, the transition in Gabon began with strikes and riots, first among students but which then spread to workers in both the public and private sectors that brought the country to a standstill. President Bongo called for a national conference with advisory status to the president, but the opposition rejected this arrangement and demanded multi-party competition. The Gabonese opposition was divided into at least two groups. One group was the *Parti Gabonais du Progrès*, which had support from the coastal regions and other groups, including workers and professors persecuted by the Bongo regime in the 1970s (Gardinier 1997, 149–52). It was the secretary-general of this party whose assassination in early 1990 provoked major rioting (Bayalama 1991, 68). Another group was *Mouvement de Redressement National* (MORENA), a group of exiles formed in France in the 1980s and originally led by the Catholic priest Paul Mba-Abessolé. It had support in the northern part of the country and from among Catholics and Protestants, who were concerned with Bongo's Islam and Masonry. Mba eventually broke away and formed the *Rassemblement National des Bûcherons*, but the opposition sought to unite for the presidential election (Messone and Gros 1998, 139).

The Cameroonian opposition was also stronger than that in Côte d'Ivoire, although it is difficult to compare the strength of opposition in Côte d'Ivoire and Gabon. The Ivoirien opposition was more unified under Laurent Gbagbo than the Gabonese opposition with its multiple leaders, but it is not clear that the pressure they put on their respective authoritarian incumbents was so different as to affect their relative probabilities of adopting plurality rule. Our best guess is that plurality rule for transitional presidential elections was most likely to have been adopted in Madagascar, followed by Cameroon, then Côte d'Ivoire and Gabon.

Within the set of former British colonies, opposition strength was greater in Zambia than in Kenya, which in turn had a stronger opposition than Malawi. Zambia had a unified opposition led by the head of the Zambian Congress of Trade Unions (ZCTU) and broad support from society. The Zambian economy was heavily dependent upon copper mining, and the trade union movement remained powerful under single-party rule. Beginning in 1989, the regime was challenged by the Movement for Multi-party Democracy (MMD), led by ZCTU leader Frederick Chiluba, with the support of prominent defectors from the ruling party, professionals, business people, and students (VonDoepp 1996, 32–3). Riots led the regime to announce a referendum on the return to multi-party politics, and the postponement of the referendum led to MMD-led mass demonstrations, forcing President Kaunda to accept multi-partyism without a referendum and to schedule elections (Erdmann and Simutanyi 2003, 10-11).

Kenya also had an active civil society and an internal opposition that had the capacity to organize large rallies, but it was not as unified as in Zambia. Following the clearly fraudulent 1988 elections, the ruling Kenya African National Union (KANU) faced increased opposition from churches and the group Mwakenya which had support among farmers and the middle class (Khapoya 1988, 62). The regime's threat to crack down on a major rally led to a split within between opposition politicians and the churches, and the new opposition umbrella group Forum for the Restoration of Democracy (FORD) was also divided between the "Young Turks" and longstanding politicians (Throup and Hornsby 1998).

Malawi had the weakest opposition among this second set. In Malawi, political liberalization was spurred by pressure from foreign governments and international financial institutions, as well as domestic social unrest. After "almost thirty years under the totalitarian control" of the Malawi Congress Party, there was little organized internal opposition to the regime (van Donge 1995, 229). Industrial action and student protests emerged only after the coordinated reading of a Pastoral Letter criticizing the regime in March 1992. Regime opponents in exile had little organization on the ground and formed pressure groups only upon their return (Newell 1995).

For the last set, we are unsure about whether the opposition was stronger in Tanzania or in Guinea-Bissau, although as a former Portuguese colony, Guinea-Bissau was probably more likely to have adopted runoff rules like in Portugal, than Tanzania, a former British colony. The Tanzanian transition was "managed" by the incumbent *Chama cha Mapinduzi* (CCM) party, the successor party to the Tanganyika African National Union (TANU) that brought the country to independence. There were several opposition parties with regional base, but broad coalitions repeatedly collapsed, and the opposition was "fragmented and weak" (van Cranenburgh 1996, 541–2) and "composed of parties of doubtful credibility and leadership" (Mwase and Raphael 1997, 153). Like the CCM in Tanzania, the ruling *Partido Africano da Independência da Guiné e Cabo Verde* (PAICG) in Guinea-Bissau was the party that led the fight for the country's independence. The demand for political liberalization in Guinea-Bissau initially came from younger ruling party members with high positions in the state (Rudebeck 2002, 115). This led to the legalization of the formation of many political parties, and as in Tanzania, the stronger opposition parties were those formed by ruling party defectors with a regional or ethnic base of support after this liberalization (Cardoso, 1994, 26; Forrest, 2005, 252–3; Rudebeck 2002, 115). Neither country had a group or politician that would likely become a strong opposition leader upon political liberalization.