Supplementary Information for
"Using Qualitative Information to Improve Causal Inference"

Adam N. Glynn* Nahomi Ichino†

July 1, 2014

Contents

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

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

A Data

We exclude countries with population less than a half million in 1989 (Cape Verde, Comoros, Djibouti, Equatorial Guinea, Seychelles, São Tomé and Príncipe). We also exclude Djibouti, Sudan, and Mauritania, which the United Nations and many scholars consider to be part of the Middle East/North Africa region rather than sub-Saharan Africa. Botswana, Lesotho, Mauritius, and South Africa are not presidential systems; Eritrea, Rwanda, Sudan, Somalia, Swaziland, Uganda, and Zaire/DRC did not hold transitional multi-party presidential elections in the 1990s. We also exclude Namibia, which held its first post-independence election just before the turn of the decade.

Plurality rule for transitional presidential election (T) = 1 if transitional presidential election used plurality rules, = 0 if it had a run-off provision to eliminate weaker candidates and hold a second round election if no candidate met a vote threshold in the first round. Coded from constitutions available through the African Legislatures Project (www.africanlegislatures project.org), Africa Elections Database (africanelections.tripod.com), Consortium for Elections and Political Process Strengthening (CEPPS) (www.electionguide.org), and Electoral Institute for the Sustainability of Democracy in Africa website (www.eisa.org.za).

Opposition harassment (Y) = 1 if there is any reported incident of opposition harassment before the election, = 0 otherwise. From National Elections across Democracy and Autocracy (NELDA) dataset v3 (Hyde & Marinov 2012).

Civil conflict = 1 if there was a civil conflict leading into the transition, = 0 otherwise.

Frequency of political protests during the transition period, 1988–92 = 0 if none, = 1 if some, = 2 frequent. To be counted, protests had to include explicit demands for political rights or changes in political rulers. Coded by Bratton & van de Walle (1997*b*).

Experience with military rule Share of years from independence to the transition under military rule, calculated using data from Bratton & van de Walle (1997*b*).

Ethnic Fractionalization Probability that two randomly selected people from a given country will belong to different ethnic groups, restricted to groups with at least 1 percent of country population (1990). The fractionalization index is $1 - \sum_{i=1}^{N} p_i^2$, where p_i is the proportion of people in each ethnic group and N is the total number of groups and ranges from 0 (perfectly homogeneous) to 1 (highly fragmented). From Fearon (2003).

Log GDP per capita Log GDP per capita, PPP adjusted. From World Development Indicators (data.worldbank.org/data-catalog), available through Quality of Government Dataset (Teorell et al. 2011).

Former French Colony = 1 if formerly a French colony, = 0 otherwise.

Private Credit Private credit by deposit money banks as share of GDP, averaged over the 5 years before the transition. Calculated using data from Beck et al. (2000).

B Pair Matching

Authoritarian incumbents agree to hold multi-party elections because their right to rule has been challenged. We are concerned that weaker incumbents – those without a strong party organization or face a unified opposition – might be more likely to resort to opposition harassment and also be more likely to adopt plurality rules for these transitional presidential elections. To address this concern, we match a country that had a runoff provision ($T = 0$) to each country that used plurality rules for the transitional presidential elections ($T = 1$). We pair match without replacement, exactly on three variables and on Mahalonobis distance with two variables, using the `optmatch` package version 0.8-1 (Hansen & Klopfer 2006) in `R` version 2.15.2.

The first matching variable is whether there was a civil conflict prior to the transition, since former combatants can be readily mobilized to harass opponents, and electoral rules under these circumstances are part of negotiated peace settlements designed to draw the armed factions into electoral politics (Lyons 2004). The second matching variable is the frequency of political protests in the transition period (coded as none, some, or frequent), as they indicate the public's dissatisfaction with the incumbent authoritarian regime. The third matching variable is experience with military rule, measured as the share of years since independence that the country was under military rule. We include this variable since countries with longer experience with civilian rather than military rule have a greater opportunity to develop political parties, and incumbents with political parties have stronger ties and better information about citizens and have the organization in place for voter mobilization (Bratton & van de Walle 1997*a*). We match exactly on these three variables. We match on Mahalonobis distance with two other variables. The first variable is log GDP per capita, which affects the regime's capacity to co-opt or repress the opposition (Bratton & van de Walle 1997*a*). The second variable is ethnic diversity from Fearon (2003), operationalized as ethnic fractionalization, which affects the complexity and difficulty of forming coalitions behind specific presidential candidates. This matching procedure produces four pairs: Cameroon–Gabon, Kenya–Côte d'Ivoire, Malawi–Zambia, and Tanzania–Guinea-Bissau. The data are presented in Table 1.

Log GDP per capita is larger for the treated unit in the fourth pair (Tanzania) than its matched control country (Guinea-Bissau), but it is smaller for the treated unit in all other pairs. The difference is greatest for the first pair, which includes Gabon, with its high oil revenues and a small population. This imbalance, with a one-sided p-value of $2/16$ using the signed rank statistic, is not too troublesome for our analysis, because it will be improved by the subsequent full matching procedure.

Pair (s)	Country	Treated (T)	Ethnic Frac.	Log GDP per capita	Civil Conflict	Protest into Transition	Military Rule Exp.
1	Cameroon	1	0.887	7.509	0	Frequent	0
1	Gabon	0	0.858	9.585	0	Frequent	0
2	Kenya	1	0.852	7.199	0	Frequent	0
2	Côte d'Ivoire	0	0.784	7.548	0	Frequent	0
3	Malawi	1	0.829	6.364	0	Frequent	0
3	Zambia	0	0.726	7.092	0	Frequent	0
4	Tanzania	1	0.953	6.712	0	Some	0
4	Guinea-Bissau	0	0.818	6.457	0	Some	0

Table 1: Data with Pair Matching

There is also some imbalance on ethnic fractionalization (exact p-value of 1/16), which is lower for the matched control country than the treated country in every pair. In particular, Tanzania has a much greater ethnic fractionalization score than any of the twenty-four countries in our sample, and the within-pair difference on this variable is largest for the Tanzania–Guinea-Bissau pair. There are also a variety of concerns with this and other similar fractionalization measures, including debates about what collection of people should count as an ethnic group, given the fluid nature of ethnicity, inter-ethnic marriages, and the possibility of membership in multiple groups; differing depths of divisions or social and cultural distance between groups; differing levels of salience of ethnicity in politics; whether all possibly enumerated groups are relevant for a particular analysis; and how the population shares of these groups can be best summarized. Tanzanian politics is frequently noted for its lesser emphasis on ethnicity compared to other African countries, sometimes attributed precisely to its large number of very small ethnic groups and to President Nyerere's sustained agenda of socialism and national integration (Kelsall 2003, Miguel 2004). Posner (2004) reviews these concerns and has proposed an alternative fractionalization index of *politically relevant* ethnic groups, but it is not clear how appropriate such a measure is in a highly fluid moment of political transition and mobilization.

Nevertheless, it is important to match on some measure of ethnic diversity, since we believe the coalition possibilities of the ethnic groups in a country affects opposition strength and hence treatment assignment. Because we do not have strong theory for the relationship between ethnic diversity and treatment assignment, we adopt this measure widely used in cross-national analyses for the matching. We address the imbalance through additional qualitative information on opposition strength affected by ethnic diversity.

C The Sign Test

Even when absolute within-pair differences in outcomes cannot be ranked across pairs, we can use the sign statistic. As indicated by its name, the sign statistic uses only the sign of the difference in outcomes for each pair ($sign(Y_{s1} - Y_{s2})$ for $s = 1, ..., N_1$). The statistic is written as:

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

where $s_{s1} = 1$ if $Y_{s1} > Y_{s2}$ and $= 0$ otherwise, and $s_{s2} = 1$ if $Y_{s2} > Y_{s1}$ and $= 0$ otherwise. This is the statistic implicitly used for McNemar's test in the second section of the article.

	Pair 1	Pair 2	Pair 3	Pair 4	Pair 1	Pair 2	Pair 3	Pair 4	
					$s_{11} = 1$	$s_{21} = 0$	$s_{31} = 1$	$s_{41} = 0$	
					$s_{12} = 0$	$s_{22} = 0$	$s_{32} = 0$	$s_{42} = 0$	
Permutation	T_1	T_2	T_3	T_4	V_1	V_2	V_3	V_4	V
1	1	1	1	1	1	0	1	0	2
2	1	1	1	0	1	0	1	0	2
3	1	1	0	1	1	0	0	0	1
4	1	1	0	0	1	0	0	0	1
5	1	0	1	1	1	0	1	0	2
6	1	0	1	0	1	0	1	0	2
7	1	0	0	1	1	0	0	0	1
8	1	0	0	0	1	0	0	0	1
9	0	1	1	1	0	0	1	0	1
10	0	1	1	0	0	0	1	0	1
11	0	1	0	1	0	0	0	0	0
12	0	1	0	0	0	0	0	0	0
13	0	0	1	1	0	0	1	0	1
14	0	0	1	0	0	0	1	0	1
15	0	0	0	1	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0

Table 2: *Permutation Distribution for McNemar's Test Using the NELDA Data*

With this statistic V, we can address the following hypothetical question: how likely is it that we could have observed the result just by chance, that is, if plurality electoral rules had been assigned by a coin flip within each pair? We calculate this p-value using Table 2 which presents the full permutation distribution for the four pairs in our example under the sharp null hypothesis. Each row in the table represents a different possible set of coin flip outcomes, with the first row corresponding to the observed data with $s = 2$. We assume that the coin is fair, so that $Pr(T_s = 1) = 1/2$ for each pair $s = 1, ..., N_1$, and because there are 4 pairs, the probability of observing any individual row of Table 2 is 1/16. Because only rows 1, 2, 5, and 6 of the table have $V \geq 2$, the one-sided p-value is 4/16.

The example demonstrates one of the properties of McNemar's test: concordant pairs, which are pairs with the same value of the outcome variable, can be removed from the dataset without affecting the p-value. In our example, the second pair (Kenya–Côte d'Ivoire) and fourth pair (Tanzania–Guinea-Bissau) do not affect the V statistic in Table 2, since $V_2 = 0$ and $V_4 = 0$ for all permutations. This reveals the limited power of using only a binary coding of the outcome variable and a point for potential improvement with the introduction of qualitative information.

Using only the signs for each pair discussed in Section "Using Qualitative Information on the Outcome" of the paper, the values for the V statistic change. The V_4 column of Table 2 would now take the value 1 for the odd rows of the table but keep the value 0 for the even rows. The V_2 column of Table 2 would now take the value 1 for rows 1–4 and rows 9–12, but keep the value 0 for the other rows. Our observed data in the first row would now produce the value of $V = 4$ while all other rows will still have $V \leq 3$. Hence the V for the observed data is larger than the V values for all other rows, and we have reduced the one-sided p-value to $1/16 = 0.0625$. These changes are reflected in Table 3. Notice that for a binary outcome variable, we only need to conduct comparative studies for the concordant pairs to improve the power of the test. In this example, we only needed to conduct two comparative case studies and determine which unit had the larger outcome value within each pair.

D Sensitivity Analysis on the Signs and Ranks

One concern is how sensitive our analysis is to errors or disagreements about signs and ranks. We can address this concern with a sensitivity analysis that calculates p-values for all possible signs and ranks, but note that some disagreements over ranks will have no effect on the p-value. For example, perhaps we cannot determine or agree on whether the Cameroon-Gabon pair or the Kenya-Côte d'Ivoire pair has a greater absolute difference in outcomes. Suppose contrary to our ranking, the Cameroon-Gabon pair is rank 4 and the Kenya-Côte d'Ivoire pair is rank 3. The p-value we obtain from this adjusted ranking is the same as from our original ranking. More generally, when $T_s s_{s1} + (1 - T_s)s_{s2} = T_{s'} s_{s'1} + (1 - T_{s'})s_{s'2}$ for $s \neq s'$ and $q_s = q_{s'} + 1$, then exchanging the ranks for pairs s and s' does not affect the p-value.

Furthermore, even errors in the signs are likely to have only small effects on the p-value, since such errors are most likely to occur for the pairs where the magnitude of the within-pair difference and the rank are small. For example, we might be concerned that we have missed some incidents of opposition harassment in Guinea-Bissau that were only reported in Portuguese language media. However, even if upon reassessment we find that Guinea-Bissau had more harassment than Tanzania, it is unlikely that the magnitude of this difference will be large, and therefore the rank will remain the same $q_4 = 1$. Because of the small rank, the sign reversal only generates a small

	Pair 1	Pair 2	Pair 3	Pair 4	Pair 1	Pair 2	Pair 3	Pair 4	
					$s_{11}=1$	$s_{21}=1$	$s_{31}=1$	$s_{41}=1$	
					$s_{12}=0$	$s_{22}=0$	$s_{32}=0$	$s_{42}=0$	
Permutation	T_1	T_2	T_3	T_4	V_1	V_2	V_3	V_4	V
1	1	1	1	1	1	1	1	1	4
2	1	1	1	0	1	1	1	0	3
3	1	1	0	1	1	1	0	1	3
4	1	1	0	0	1	1	0	0	2
5	1	0	1	1	1	0	1	1	3
6	1	0	1	0	1	0	1	0	2
7	1	0	0	1	1	0	0	1	2
8	1	0	0	0	1	0	0	0	1
9	0	1	1	1	0	1	1	1	3
10	0	1	1	0	0	1	1	0	2
11	0	1	0	1	0	1	0	1	2
12	0	1	0	0	0	1	0	0	1
13	0	0	1	1	0	0	1	1	2
14	0	0	1	0	0	0	1	0	1
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 Sign Test Using Within-Pair Qualitative Information to Supplement the NELDA Data*

difference in the p-value. Note that in this alternative, $s_{41} = 0$ and $s_{42} = 1$ so that we subtract one from W in the odd rows and add one to W in the even rows of Table 4 (reproduction of Table 3 in the paper). Our observed W would now be 9, and the second row would have $W = 10$, but no other row in the table would have a W as large as 9. Therefore, the p-value would only increase from 1/16 to 2/16.

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

E Full Matching

With full matching, including exact matching on whether the country is a former French colony, our data are grouped into three sets. The first set is former French colonies with “frequent” protests in the transition period – Cameroon, Côte d’Ivoire, and Gabon, and Madagascar. The second set is former British colonies with “frequent” protests in the transition period – Kenya, Malawi, and Zambia. The third set of countries are those with “some” protests in the transition period – Tanzania and Guinea-Bissau. These comprised the fourth pair in the original analysis. Table 5 presents the data.

Set (s)	Country	Treated (T)	Ethnic Frac.	Log GDP per capita	Former French	Civil Conflict	Protest	Military Rule	Private Credit
1	Cameroon	1	0.887	7.509	1	0	2	0	26.440
1	Côte d’Ivoire	0	0.784	7.548	1	0	2	0	35.452
1	Gabon	0	0.858	9.585	1	0	2	0	11.388
1	Madagascar	0	0.861	6.831	1	0	2	0	15.108
2	Kenya	1	0.852	7.199	0	0	2	0	17.806
2	Malawi	1	0.829	6.364	0	0	2	0	9.646
2	Zambia	0	0.726	7.092	0	0	2	0	7.222
3	Tanzania	1	0.953	6.712	0	0	1	0	4.496
3	Guinea-Bissau	0	0.818	6.457	0	0	1	0	3.230

Table 5: Data with Full Matching

Exact matching on civil conflict, protest leading into the transition, being a former French colony, and a slightly weakened military rule variable completely determines the matched sets for this dataset. However, we can still assess balance for other potential matching variables using Quade’s statistic. For log GDP per capita, the one-sided p-value with Quade’s statistic is 11/24. For ethnic fractionalization, the p-value is 1/24. For availability of private credit,[1] measured by private credit by deposit money banks as percentage of GDP, averaged for the five years prior to the transition, the one-sided p-value with Quade’s statistic is 21/24. As before, we address the imbalance on ethnic fractionalization through the overall qualitative assessment of opposition strength.

[1]Arriola (2012) argues that private credit helps keep opposition coalitions financed and unified for transitional and later elections in sub-Saharan Africa.

F Replication Code

The analyses in the article can be replicated with the `R` code below. Some portions are made significantly easier by using the `R` package `qualCI` v0.1 (Kashin, Glynn & Ichino 2014), which includes the data from the Africa example in the article and functions to calculate the p-values and qualitative confidence intervals.

F.1 Paired Design

The first p-value in the article is produced using an exact version of McNemar's test on the data in Table 1.

```
library(exact2x2)
exact2x2(matrix(c(1,0,2,1),nr=2), alternative="greater", paired=TRUE)
# the first 1 in the matrix corresponds to the zero/zero pair
#     (Tanzania/Guinea-Bissau)
# the 0 in the matrix corresponds to the lack of a zero/one pair
# the 2 in the matrix corresponds to the two one/zero pairs
#     (Cameroon/Gabon, Malawi/Zambia)
# the second 1 in the matrix corresponds to the one/one pair (Kenya/Cote d'Ivoire)
```

We produce the second p-value with the signed-rank test on the signs and ranks in Table 2. We can use the standard Wilcoxon function on hypothetical data that correspond to the signs and ranks of table.

```
wilcox.test(c(6,8,4,2),c(3,4,2,1), alternative="greater", paired=TRUE, exact=TRUE)
# Cameroon = 6
# Kenya = 8
# Malawi = 4
# Tanzania = 2
# Gabon = 3
# Cote d'Ivoire = 4
# Zambia = 2
# Guinea-Bissau = 1
```

Alternatively, under the assumption that we flip a fair coin for each pair, each row in the permutation distribution in Table 3 has a $1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2 = 1/16$ chance of occurring under the sharp null hypothesis. Because the observed signed-rank statistic took its maximum value for this example ($W = 10$), the p-value is just the probability for the first row (i.e., 1/16).

We can use `qualCI` to calculate the p-value and qualitative confidence interval:

```
library(qualCI)
data(pluralityPairs) # data from the Africa example
between.ranks <- c(3,4,2,1)
dat <- quade.prep(data=pluralityPairs, set="pair", unit="country", treatment=
"plurality", withinRank="OppHarRank", betweenRank=between.ranks)
dat
qout <- quade(dat)
qout # p-value
qualCI(qout) # qualitative confidence interval
```

F.2 Full Matching Design

We use the Quade statistic to produce the p-value associated with using full matching (Table 6). This can be calculated by creating hypothetical data that correspond to the between-set and within-set ranks of Table 5.

```
# Cameroon = 8
# Gabon = 6
# Cote d'Ivoire = 4
# Madagascar = 2
# Kenya = 10
# Malawi = 6
# Zambia = 3
# Tanzania = 2
# Guinea-Bissau = 1

d<- data.frame(c(1,1,1,1,2,2,2,3,3),c(8,6,4,2,10,6,3,2,1),rep(NA,9))
colnames(d)<-c("set","var","rank")

# within-set rank for units in set 1
d[1,3]<-rank(c(d[1,2], d[2,2], d[3,2], d[4,2]))[1]
d[2,3]<-rank(c(d[1,2], d[2,2], d[3,2], d[4,2]))[2]
d[3,3]<-rank(c(d[1,2], d[2,2], d[3,2], d[4,2]))[3]
d[4,3]<-rank(c(d[1,2], d[2,2], d[3,2], d[4,2]))[4]
# within-set rank for units in set 2
d[5,3]<-rank(c(d[5,2], d[6,2], d[7,2]))[1]
d[6,3]<-rank(c(d[5,2], d[6,2], d[7,2]))[2]
d[7,3]<-rank(c(d[5,2], d[6,2], d[7,2]))[3]
# within-set rank for units in set 3
```

```
d[8,3]<-rank(c(d[8,2], d[9,2]))[1]
d[9,3]<-rank(c(d[8,2], d[9,2]))[2]

# between-set rank for set 1
q1<-rank(c((d$var[d$rank==4&d$set==1]-d$var[d$rank==1&d$set==1]),
    (d$var[d$rank==3&d$set==2]-d$var[d$rank==1&d$set==2]),
    (d$var[d$rank==2&d$set==3]-d$var[d$rank==1&d$set==3])))[1]
# between-set rank for set 2
q2<-rank(c((d$var[d$rank==4&d$set==1]-d$var[d$rank==1&d$set==1]),
    (d$var[d$rank==3&d$set==2]-d$var[d$rank==1&d$set==2]),
    (d$var[d$rank==2&d$set==3]-d$var[d$rank==1&d$set==3])))[2]
# between-set rank for set 3
q3<-rank(c((d$var[d$rank==4&d$set==1]-d$var[d$rank==1&d$set==1]),
    (d$var[d$rank==3&d$set==2]-d$var[d$rank==1&d$set==2]),
    (d$var[d$rank==2&d$set==3]-d$var[d$rank==1&d$set==3])))[3]

# each row is a different permutation of treatment assignment
m<-matrix(NA,ncol=5,nrow=24)
# Q1
m[,1]<-c(rep(q1*d[1,3],6), rep(q1*d[2,3],6), rep(q1*d[3,3],6), rep(q1*d[4,3],6))
# Q2
m[,2]<-c(rep(c(rep(q2*(d[5,3]+d[6,3]),2), rep(q2*(d[5,3]+d[7,3]),2),
    rep(q2*(d[6,3]+d[7,3]),2)),4))
# Q3
m[,3]<-c(rep(c(q3*d[8,3],q3*d[9,3]),12))
# Q
m[,4]<-m[,1]+m[,2]+m[,3]
# note that by construction, observed data is the first row
m[,5]<-(m[,4]>=m[1,4])
mean(m[,5]) # p-value 1/24
```

Again, the observed statistic is the largest possible value of the Quade statistic. Therefore, if the treatment assignment is "as if" random for our example, then $\pi_{1,\{1,0,0,0\}} = \frac{1}{\binom{n_1}{m_1}} = \frac{1}{4}$ for $s = 1$, $\pi_{2,\{1,1,0\}} = \frac{1}{\binom{n_2}{m_2}} = \frac{1}{3}$ for $s = 2$, and $\pi_{3,\{1,0\}} = \frac{1}{\binom{n_3}{m_3}} = \frac{1}{2}$ for $s = 3$, and the p-value would be $\pi_{1,\{1,0,0,0\}} \cdot \pi_{2,\{1,1,0\}} \cdot \pi_{3,\{1,0\}} = 1/4 \cdot 1/3 \cdot 1/2 = 1/24$.

```
pi1.1000 <- 1/4
pi2.110 <- 1/3
```

```
pi3.10 <- 1/2
pi1.1000*pi2.110*pi3.10 # produces the p-value of 1/24
```

We conduct the sensitivity analysis by increasing the values of $\pi_{1,\{1,0,0,0\}}$, $\pi_{2,\{1,1,0\}}$, or $\pi_{3,\{1,0\}}$. For example, the p-values in Table 7 of the paper can be reproduced with the following `R` code:

```
pi1.1000 <- c(1/4,1.25/4.25,1.5/4.5,2/5,2.5/5.5,3/6)
pi2.110 <- 1/3
pi3.10 <- c(1/2,1.25/2.25,1.5/2.5,2/3,2.5/3.5,3/4)
pi2.110*outer(pi1.1000,pi3.10) # produces the numbers before rounding
round(pi2.110*outer(pi1.1000,pi3.10),digits=3) # produces the numbers in the table
```

We can use `qualCI` to produce the p-values for the full matched design with the Quade statistic:

```
data(pluralitySets)
between.ranks <- c(2,3,1)
dat <- quade.prep(data=pluralitySets, set="set", treatment="plurality",
unit="country", withinRank="OppHarRank", betweenRank=between.ranks)
qout <- quade(dat)
qout
```

To produce the qualitative confidence interval,

```
qualCI(qout)
```

will offer the user several pairs of treated and control countries and ask for which pair the between-set rank was most difficult to determine.

The following code will produce the sensitivity analysis in Table 7 using the `qualCI` package.

```
pi1 <- list(c(1/4,1/4,1/4,1/4),
c(1.25/4.25,rep((1-1.25/4.25)/3,3)),
c(1.5/4.5,rep((1-1.5/4.5)/3,3)),
c(2/5,rep(1/5,3)),
c(2.5/5.5,rep((1-2.5/5.5)/3,3)),
c(3/6,rep(1/6,3)))
pi3 <- list(c(1/2,1/2),
c(1.25/2.25,1-1.25/2.25),
c(1.5/2.5, 1-1.5/2.5),
c(2/3, 1/3),
c(2.5/3.5, 1-2.5/3.5),
c(3/4, 1/4))
```

```
table7 <- matrix(data=NA, nrow=length(pi1), ncol=length(pi3))

for(m in 1:nrow(table7)){
for(n in 1:ncol(table7)){
dat[[1]]$prob <- pi1[[m]]
dat[[3]]$prob <- pi3[[n]]
table7[m,n] <- quade(dat)$pval
}
}

table7
```

References

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

Beck, Thorsten, Aslı Demirgüç-Kunt & Ross Levine. 2000. "A New Database on Financial Development and Structure." *World Bank Economic Review* 14:597–605. Dataset available at `http://go.worldbank.org/X23UD9QUX0`.

Bratton, Michael & Nicolas van de Walle. 1997*a*. *Democratic Experiments in Africa: Regime Transitions in Comparative Perspective.* Cambridge: Cambridge University Press.

Bratton, Michael & Nicolas van de Walle. 1997*b*. "Political Regimes and Regime Transitions in Africa, 1910-1994 (ICPSR Dataset 6996).".

Fearon, James D. 2003. "Ethnic Structure and Cultural Diversity Around the World: A Cross-National Data Set on Ethnic Groups." *Journal of Economic Growth* 8(2):191–218.

Hansen, Ben B. & Stephanie Olsen Klopfer. 2006. "Optimal full matching and related designs via network flows." *Journal of Computational and Graphical Statistics* 15(3):609–627.

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

Kashin, Konstantin, Adam N. Glynn & Nahomi Ichino. 2014. "`qualCI` v0.1.". `R` package.

Kelsall, Tim. 2003. "Governance, Democracy and Recent Political Struggles in Mainland Tanzania." *Commonwealth & Comparative Politics* 41(2):55–82.

Lyons, Terrence. 2004. "Post-conflict Elections and the Process of Demilitarizing Politics: The Role of Electoral Administration." *Democratization* 11(3):36–62.

Miguel, Edward. 2004. "Tribe or Nation? Nation-Building and Public Goods in Kenya versus Tanzania." *World Politics* 56(3):327–362.

Posner, Daniel N. 2004. "Measuring Ethnic Fractionalization in Africa." *American Journal of Political Science* 48(4):849–63.

Teorell, Jan, Marcus Samanni, Sören Holmberg & Bo Rothstein. 2011. "The Quality of Government Dataset, version 6 Apr 11." University of Gothenburg: The Quality of Government Institute, `http://www.qog.pol.gu.se`.