What determines the duration of protest events?  
Evidence from Africa

*Online supplemental appendix*

Jacob S. Lewis, *Washington State University*

William Favell, *Washington State University*

Contents

[Descriptive statistics 4](#_Toc184290807)

[Robustness tests 5](#_Toc184290808)

[Hand-coding the size of participation 10](#_Toc184290809)

[Examining variation within durations 15](#_Toc184290810)

[Is repression endogenous with duration? 16](#_Toc184290811)

[Pooled models 17](#_Toc184290812)

[Evaluating pooled versus hierarchical modeling 20](#_Toc184290813)

[Overview 20](#_Toc184290814)

[Hypotheses 20](#_Toc184290815)

[Results 20](#_Toc184290816)

[Spatial autocorrelation 21](#_Toc184290817)

[Alternative datasets 23](#_Toc184290818)

[SCAD Dataset 23](#_Toc184290819)

[Addressing selection issues and endogeneity 25](#_Toc184290820)

[Selecting on repression 25](#_Toc184290821)

[Selecting on student participants 26](#_Toc184290822)

[Selection on labor union participation 27](#_Toc184290823)

[Coarsened Exact Matching – Number of actors 28](#_Toc184290824)

[Coarsened Exact Matching – Students Present 28](#_Toc184290825)

[Coarsened Exact Matching – Labor Present 29](#_Toc184290826)

[Coarsened Exact Matching – In the capital 29](#_Toc184290827)

[Coarsened Exact Matching – Repressed 29](#_Toc184290828)

[Including riots in the analysis 30](#_Toc184290829)

[Holding the day of the event constant 32](#_Toc184290830)

[Additional qualitative evidence 35](#_Toc184290831)

[On the South African student protests 35](#_Toc184290832)

**List of tables**

[Table 1 - Descriptive statistics for continuous and count variables 4](#_Toc169268879)

[Table 2 - Descriptive statistics for dichotomous variables 4](#_Toc169268880)

[Table 3 - Bivariate regressions for Approach 1 (next day) 5](#_Toc169268881)

[Table 4 - Bivariate regressions for Approach 2 (duration) 6](#_Toc169268882)

[Table 5 - SCAD Robustness Test 7](#_Toc169268883)

[Table 6 - Autocorrelation Robustness 8](#_Toc169268884)

[Table 7 - Testing hand-coded protest size variable 9](#_Toc169268885)

[Table 8 - Logistic regression with expanded temporal window 11](#_Toc169268886)

[Table 9 - Neg. Binomial with expanded temporal window 12](#_Toc169268887)

[Table 10 - Descriptive statistics by duration 15](#_Toc169268888)

[Table 11 - Negative binomial regression with lag of repression 16](#_Toc169268889)

[Table 12 - Pooled Day-to-day models 17](#_Toc169268890)

[Table 13 - Pooled Duration Models 18](#_Toc169268891)

[Table 14 – Including spatial autocorrelation 21](#_Toc169268892)

[Table 15 - Effect of urban/rural on protest size 22](#_Toc169268893)

[Table 16 - 2 Stage Selection Model on Repression 25](#_Toc169268894)

[Table 17 - 2 Stage Selection Model on student participation 26](#_Toc169268895)

[Table 18 - 2 Stage Selection Model on labor participation 27](#_Toc169268896)

[Table 19 - Matching data on number of actors 28](#_Toc169268897)

[Table 20 - Causal estimate for number of actors 28](#_Toc169268898)

[Table 21 - Matching data on number of actors 28](#_Toc169268899)

[Table 22 - Causal estimate for number of actors 28](#_Toc169268900)

[Table 23 - Matching data on number of actors 29](#_Toc169268901)

[Table 24 - Causal estimate for number of actors 29](#_Toc169268902)

[Table 25 - Matching data on number of actors 29](#_Toc169268903)

[Table 26 - Causal estimate for number of actors 29](#_Toc169268904)

[Table 27 - Matching data on number of actors 29](#_Toc169268905)

[Table 28 - Causal estimate for number of actors 29](#_Toc169268906)

[Table 29 - Next day (including riots) 30](#_Toc169268907)

[Table 30 - Duration (including riots) 31](#_Toc169268908)

[Table 31 - Likelihood of observing another day 32](#_Toc169268909)

[Table 32 - Likelihood of next day, holding day of event constant 33](#_Toc169268910)

# Descriptive statistics

Table - Descriptive statistics for continuous and count variables

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Min | 1st Q | Median | Mean | 3rd Q | Max |
| Independent variables |  |  |  |  |  |  |
| Log of max population | 2.398 | 12.499 | 13.835 | 13.433 | 14.819 | 16.063 |
| Number of actors | 1 | 1 | 1 | 1.1 | 1 | 14 |
| V-Dem free expression | 0.012 | 0.658 | 0.839 | 0.7329 | 0.878 | 0.9510 |
| V-Dem free association | 0.025 | 0.566 | 0.807 | 0.713 | 0.877 | 0.933 |
|  |  |  |  |  |  |  |
| Dependent variables |  |  |  |  |  |  |
| Duration | 1 | 1 | 1 | 1.376 | 1 | 15 |
|  |  |  |  |  |  |  |

Table - Descriptive statistics for dichotomous variables

|  |  |  |
| --- | --- | --- |
|  | No (0) | Yes (1) |
| **Independent variables** |  |  |
| Student organization | 27,307 | 3,383 |
| Professional organization | 29,346 | 1,344 |
| Labor organization | 26,354 | 4,336 |
| Occurs in capital | 15,875 | 14,815 |
| Occurs in urban area | 11,733 | 18,957 |
| Experienced repression | 25,761 | 4,929 |
|  |  |  |
| **Dependent variable** |  |  |
| Next day? | 27,065 | 3,625 |

# Robustness tests

Table - Bivariate regressions for Approach 1 (next day)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 | Model 8 |
| (Intercept) | -4.736\*\*\* | -3.199\*\*\* | -3.135\*\*\* | -3.105\*\*\* | -3.089\*\*\* | -3.247\*\*\* | -3.174\*\*\* | -3.067\*\*\* |
|  | (0.369) | (0.000) | (0.000) | (0.000) | (0.105) | (0.113) | (0.000) | (0.000) |
| Log of max pop. | 0.137\*\*\* |  |  |  |  |  |  |  |
|  | (0.028) |  |  |  |  |  |  |  |
| Number of actors |  | 0.087\*\*\* |  |  |  |  |  |  |
|  |  | (0.000) |  |  |  |  |  |  |
| Students |  |  | 0.332\*\*\* |  |  |  |  |  |
|  |  |  | (0.000) |  |  |  |  |  |
| Professionals |  |  |  | 0.155\*\*\* |  |  |  |  |
|  |  |  |  | (0.000) |  |  |  |  |
| Labor |  |  |  |  | 0.160\*\* |  |  |  |
|  |  |  |  |  | (0.051) |  |  |  |
| Capital |  |  |  |  |  | 0.550\*\*\* |  |  |
|  |  |  |  |  |  | (0.097) |  |  |
| Urban |  |  |  |  |  |  | 0.242\*\*\* |  |
|  |  |  |  |  |  |  | (0.000) |  |
| Repression |  |  |  |  |  |  |  | -0.188\*\*\* |
|  |  |  |  |  |  |  |  | (0.000) |
| AIC | 20242 | 20617 | 20585.569 | 20619 | 20612 | 20591 | 20610 | 20610 |
| BIC | 20283 | 20658 | 20627 | 20660 | 20654 | 20633 | 2065 | 20652 |
| Log Likelihood | -10116 | -10303 | -10287 | -10304 | -10301 | -10290 | -10300 | -10300 |
| Num. obs. | 30095 | 30690 | 30690 | 30690 | 30690 | 30690 | 30690 | 30690 |
| Num. groups: admin2 | 1978 | 2070 | 2070 | 2070 | 2070 | 2070 | 2070 | 2070 |
| Num. groups: admin1 | 577 | 580 | 580 | 580 | 580 | 580 | 580 | 580 |
| Num. groups: country | 43 | 43 | 43 | 43 | 43 | 43 | 43 | 43 |
| Var: admin2 (Intercept) | 0.524 | 0.580 | 0.584 | 0.577 | 0.559 | 0.541 | 0.568 | 0.581 |
| Var: admin1 (Intercept) | 0.296 | 0.350 | 0.347 | 0.355 | 0.336 | 0.269 | 0.334 | 0.352 |
| Var: country (Intercept) | 0.253 | 0.190 | 0.178 | 0.185 | 0.194 | 0.220 | 0.188 | 0.187 |
| \*\*\*p < 0.001; \*\*p < 0.01; \*p < 0.05 | | | | | | | | |

Table - Bivariate regressions for Approach 2 (duration)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 |
| (Intercept) | 0.071\*\*\* | 0.087\*\*\* | 0.096\*\*\* | 0.086\*\*\* | 0.064\*\*\* | 0.103\*\*\* | 0.095\*\*\* |
|  | (0.010) | (0.010) | (0.011) | (0.011) | (0.013) | (0.011) | (0.010) |
| Pre-event moment. | 0.014\*\*\* |  |  |  |  |  |  |
|  | (0.002) |  |  |  |  |  |  |
| Students |  | 0.120\*\*\* |  |  |  |  |  |
|  |  | (0.017) |  |  |  |  |  |
| Professional |  |  | 0.082\*\* |  |  |  |  |
|  |  |  | (0.025) |  |  |  |  |
| Labor |  |  |  | 0.096\*\*\* |  |  |  |
|  |  |  |  | (0.015) |  |  |  |
| Capital |  |  |  |  | 0.077\*\*\* |  |  |
|  |  |  |  |  | (0.013) |  |  |
| Repress D1 |  |  |  |  |  | -0.012 |  |
|  |  |  |  |  |  | (0.016) |  |
| Repress D2 |  |  |  |  |  |  | 1.118\*\*\* |
|  |  |  |  |  |  |  | (0.040) |
| AIC | 59924.321 | 59946.799 | 59983.359 | 59957.094 | 59961.318 | 59992.361 | 59752.252 |
| BIC | 59965.351 | 59987.829 | 60024.389 | 59998.124 | 60002.348 | 60033.391 | 59793.282 |
| Log Likelihood | -29957.161 | -29968.400 | -29986.680 | -29973.547 | -29975.659 | -29991.181 | -29871.126 |
| Num. obs. | 27065 | 27065 | 27065 | 27065 | 27065 | 27065 | 27065 |
| Num. groups: admin1 | 580 | 580 | 580 | 580 | 580 | 580 | 580 |
| Num. groups: country | 43 | 43 | 43 | 43 | 43 | 43 | 43 |
| Var: admin1 (Intercept) | 0.000 | 0.001 | 0.001 | 0.001 | 0.000 | 0.001 | 0.001 |
| Var: country (Intercept) | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.002 | 0.001 |
| \*\*\*p < 0.001; \*\*p < 0.01; \*p < 0.05 | | | | | | | |

Table - SCAD Robustness Test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Model 1 | Model 2 | Model 3 | Model 4 |
| Intercept | 0.217 | 0.146\*\*\* | 0.063 | 0.475\*\* |
|  | (0.116) | (0.001) | (0.113) | (0.152) |
| Size | 0.067\*\*\* | 0.074\*\*\* | 0.059\*\* | 0.059\*\* |
|  | (0.020) | (0.001) | (0.019) | (0.020) |
| Occurs in capital | 0.402\*\*\* | 0.370\*\*\* | 0.360\*\*\* | 0.356\*\*\* |
|  | (0.076) | (0.001) | (0.075) | (0.075) |
| Student organization |  | 0.757\*\*\* | 0.680\*\*\* | 0.669\*\*\* |
|  |  | (0.001) | (0.070) | (0.070) |
| Professional organization |  | 0.428\*\*\* | 0.459\*\*\* | 0.431\*\*\* |
|  |  | (0.001) | (0.120) | (0.120) |
| Labor organization |  | -0.171\*\*\* | -0.131 | -0.118 |
|  |  | (0.001) | (0.103) | (0.103) |
| Event was repressed |  |  | 0.259\*\*\* | 0.253\*\*\* |
|  |  |  | (0.032) | (0.032) |
| V-Dem free expression |  |  |  | -0.228 |
|  |  |  |  | (0.255) |
| V-Dem free association |  |  |  | -0.429 |
|  |  |  |  | (0.259) |
| AIC | 15009.084 | 14875.493 | 14812.223 | 14798.485 |
| BIC | 15046.208 | 14931.179 | 14874.096 | 14872.732 |
| Log Likelihood | -7498.542 | -7428.747 | -7396.112 | -7387.242 |
| Num. obs. | 3595 | 3595 | 3595 | 3595 |
| Num. groups: Admin 1 | 567 | 567 | 567 | 567 |
| Num. groups: Country | 48 | 48 | 48 | 48 |
| Var: Admin 1 (Intercept) | 0.276 | 0.275 | 0.274 | 0.262 |
| Var: Country (Intercept) | 0.189 | 0.197 | 0.173 | 0.197 |

\*\*\*p < 0.001; \*\*p < 0.01; \*p < 0.05

Table - Autocorrelation Robustness

|  |  |
| --- | --- |
|  | Model 1 |
| Intercept | -0.059\* |
|  | (0.028) |
| Pre-event momentum | 0.011\*\*\* |
|  | (0.002) |
| Occurs in national capital | 0.051\*\*\* |
|  | (0.012) |
| Student organization | 0.119\*\*\* |
|  | (0.018) |
| Professional organization | 0.060\* |
|  | (0.026) |
| Labor organization | 0.094\*\*\* |
|  | (0.016) |
| Repression on day 1 | -0.036\* |
|  | (0.016) |
| Repression on day 2 | 1.069\*\*\* |
|  | (0.056) |
| V-Dem free expression | -0.021 |
|  | (0.058) |
| V-Dem free association | 0.054 |
|  | (0.063) |
| Lag of duration | 0.057\*\*\* |
|  | (0.010) |
| AIC | 58020.798 |
| BIC | 58135.305 |
| Log Likelihood | -28996.399 |
| Num. obs. | 26346 |
| Num. groups: admin1 | 578 |
| Num. groups: country | 43 |
| Var: admin1 (Intercept) | 0.000 |
| Var: country (Intercept) | 0.000 |

Table - Testing hand-coded protest size variable

|  | **Model 1** | **Model 2** | **Model 3** | **Model 4** |
| --- | --- | --- | --- | --- |
| Intercept | -5.274\*\*\* | -3.891\*\*\* | -5.185\*\*\* | -3.595\*\*\* |
|  | (0.849) | (0.976) | (0.871) | (1.011) |
| Log of local population | 0.185\*\* | 0.051 | 0.169\* | 0.061 |
|  | (0.063) | (0.079) | (0.067) | (0.082) |
| Protest size (hand coded) | -0.093 | -0.083 | -0.082 | -0.076 |
|  | (0.064) | (0.064) | (0.064) | (0.065) |
| Student organization | 0.363\* | 0.357\* | 0.351\* | 0.356\* |
|  | (0.151) | (0.152) | (0.152) | (0.155) |
| Professional organization | -0.225 | -0.236 | -0.221 | -0.222 |
|  | (0.261) | (0.261) | (0.260) | (0.267) |
| Labor organization | 0.206 | 0.217 | 0.207 | 0.198 |
|  | (0.132) | (0.132) | (0.132) | (0.136) |
| Capital |  | 0.591\* |  | 0.568\* |
|  |  | (0.247) |  | (0.256) |
| Repression |  | 0.177 | 0.189 | 0.217 |
|  |  | (0.153) | (0.154) | (0.155) |
| Urban |  |  | 0.100 |  |
|  |  |  | (0.174) |  |
| V-Dem free expression |  |  |  | -0.931 |
|  |  |  |  | (1.059) |
| V-Dem free association |  |  |  | 0.278 |
|  |  |  |  | (1.069) |
| AIC | 3041.300 | 3038.091 | 3043.498 | 2942.160 |
| BIC | 3100.081 | 3109.935 | 3115.343 | 3026.648 |
| Log Likelihood | -1511.650 | -1508.045 | -1510.749 | -1458.080 |
| Num. obs. | 5071 | 5071 | 5071 | 4910 |
| Num. groups: admin2 | 856 | 856 | 856 | 838 |
| Num. groups: admin1 | 406 | 406 | 406 | 399 |
| Num. groups: country | 43 | 43 | 43 | 43 |
| Var: admin2 (Intercept) | 0.717 | 0.717 | 0.709 | 0.760 |
| Var: admin1 (Intercept) | 0.000 | 0.000 | 0.000 | 0.000 |
| Var: country (Intercept) | 0.243 | 0.184 | 0.238 | 0.204 |
| \*\*\*p < 0.001; \*\*p < 0.01; \*p < 0.05 | | | | |

## Hand-coding the size of participation

One core challenge to our modeling approach is that we use a proxy in order to estimate the possible size of the protest. After all, many single-day protests occur in highly populated areas, and it is possible that the size of a given protest is not correlated with the population of its environs. This is a strong and legitimate critique of our modeling choice. To address this, we have hand-coded the size of each protest found within the dataset using the comments and notes found within the dataset itself. Doing so yielded a total of 5,158 observations in which we felt confident in assigned an ordinal value to approximate the size of the protest.[[1]](#footnote-1) The results, which are found in the appendix, do not support the hypothesis that the size of individual protest events matters for extending – or contracting – the duration of an event. Across each of the models, the protest size variable is indistinguishable statistically from zero.

The advantage of this approach is that it provides granular, day-to-day variation in reporting the size of each protest. We firmly believe that such a granular approach is the future of social movements research. The disadvantage, however, is that only a small number of the total dataset can be coded with confidence. As a result, the promise of day-to-day variation is stymied by the reality that few multiday protest events actually register any change in size within the variable. Moreover, the dataset is cut down by nearly 80 percent. An examination of the parameters of the data with a non-NA values for protest size indicates that the presence of a protest size value is likely systematic. As such, we are hesitant to draw too much from these results, given the rather extreme censoring of the data.

Table - Logistic regression with expanded temporal window

|  | **Model 1** | **Model 2** | **Model 3** | **Model 4** | **Model 5** |
| --- | --- | --- | --- | --- | --- |
| Intercept | -4.213\*\*\* | -3.590\*\*\* | -4.026\*\*\* | -3.208\*\*\* | -2.451\*\*\* |
|  | (0.330) | (0.392) | (0.328) | (0.423) | (0.215) |
| Log of local population | 0.131\*\*\* | 0.074\* | 0.113\*\*\* | 0.074\* |  |
|  | (0.025) | (0.033) | (0.025) | (0.033) |  |
| Student organization | 0.300\*\*\* | 0.304\*\*\* | 0.294\*\*\* | 0.309\*\*\* | 0.283\*\*\* |
|  | (0.046) | (0.046) | (0.046) | (0.047) | (0.048) |
| Professional organization | 0.099 | 0.093 | 0.092 | 0.075 | 0.040 |
|  | (0.075) | (0.075) | (0.075) | (0.076) | (0.078) |
| Labor organization | 0.112\* | 0.107\* | 0.099\* | 0.107\* | 0.089 |
|  | (0.044) | (0.044) | (0.044) | (0.045) | (0.046) |
| Capital |  | 0.310\*\* |  | 0.306\*\* | 0.489\*\*\* |
|  |  | (0.115) |  | (0.116) | (0.087) |
| Repression |  | -0.152\*\*\* | -0.154\*\*\* | -0.153\*\*\* | -0.154\*\*\* |
|  |  | (0.045) | (0.045) | (0.046) | (0.045) |
| Urban |  |  | 0.206\*\*\* |  |  |
|  |  |  | (0.062) |  |  |
| V-Dem free expression |  |  |  | -0.672 | -0.652 |
|  |  |  |  | (0.382) | (0.379) |
| V-Dem free association |  |  |  | 0.107 | 0.092 |
|  |  |  |  | (0.415) | (0.412) |
| Number of actors |  |  |  |  | 0.085\* |
|  |  |  |  |  | (0.036) |
| AIC | 27258.277 | 27244.153 | 27240.166 | 26578.681 | 27006.877 |
| BIC | 27324.774 | 27327.274 | 27323.287 | 26678.108 | 27106.540 |
| Log Likelihood | -13621.138 | -13612.076 | -13610.083 | -13277.340 | -13491.439 |
| Num. obs. | 30095 | 30095 | 30095 | 29308 | 29888 |
| Num. groups: admin2 | 1978 | 1978 | 1978 | 1955 | 2044 |
| Num. groups: admin1 | 577 | 577 | 577 | 574 | 578 |
| Num. groups: country | 43 | 43 | 43 | 43 | 43 |
| Var: admin2 (Intercept) | 0.496 | 0.498 | 0.474 | 0.493 | 0.515 |
| Var: admin1 (Intercept) | 0.219 | 0.202 | 0.224 | 0.198 | 0.189 |
| Var: country (Intercept) | 0.276 | 0.267 | 0.266 | 0.294 | 0.281 |
| \*\*\*p < 0.001; \*\*p < 0.01; \*p < 0.05 | | | | | |

Table - Neg. Binomial with expanded temporal window

|  | **Model 1** |
| --- | --- |
| Intercept | -0.10 |
|  | (0.06) |
| Pre-event momentum | 0.01\*\*\* |
|  | (0.00) |
| Capital | 0.05\*\* |
|  | (0.02) |
| Student organization | 0.13\*\*\* |
|  | (0.02) |
| Professional organization | 0.07\* |
|  | (0.03) |
| Labor organization | 0.10\*\*\* |
|  | (0.02) |
| Repression day 1 | -0.04\* |
|  | (0.02) |
| Repression day 2 | 1.06\*\*\* |
|  | (0.06) |
| V-Dem free expression | -0.05 |
|  | (0.07) |
| V-Dem free association | 0.10 |
|  | (0.07) |
| Lag duration | 0.06\*\*\* |
|  | (0.01) |
| Log of max population | 0.00 |
|  | (0.01) |
| AIC | 52602.37 |
| BIC | 52707.41 |
| Log Likelihood | -26288.19 |
| Deviance | 3119.09 |
| Num. obs. | 23862 |
| \*\*\*p < 0.001; \*\*p < 0.01; \*p < 0.05 | |

Figure - Predicted duration based on lag of previous duration

Chart, histogram

Description automatically generated

Figure - Who gets repressed?

Chart, bar chart

Description automatically generated

# Examining variation within durations

Below, in Table 10, we examine the descriptive statistics associated with momentum-based factors such as protest size, participants, location, and experiences of repression. We break these out into four categories: single-day events, two-day events, three-day events, and events lasting four days or longer.

Table - Descriptive statistics by duration

| **Duration** | **Obs.** | **Size** | | **Participants** | | | **Location** | | **Repressed** | **Population** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Approx. size** | **Actors** | **Stud.** | **Prof.** | **Labor** | **Urban** | **Capital** |
| 1 day | 24,556 | 2.94 | 1.10 | 10.05% | 4.40% | 13.60% | 59.10% | 44.74% | 16.48% | 13.33401 |
| 2 days | 1,856 | 3.15 | 1.25 | 20.47% | 7.81% | 24.25% | 74.89% | 60.72% | 23.49% | 13.83138 |
| 3 days | 428 | 2.92 | 1.26 | 28.27% | 6.78% | 27.57% | 80.37% | 66.82% | 28.27% | 14.01767 |
| 4+ days | 225 | 3.11 | 1.29 | 26.67% | 7.56% | 28.44% | 74.67% | 66.22% | 29.33% | 13.82823 |

We observe that the main “cutpoint” does seem to be between single-day events and multi-day events. While the observed approximate size does not seem to vary much while moving across categories, we do observe that the number of observable actors jumps from 1.1 to 1.25 when moving from single-day events to multiday events, rising slowly monotonically thereafter. A similar observation can be made about the overall presence of each of the types of participants that we track, as well as presence in urban and capital regions.

# Is repression endogenous with duration?

Table - Negative binomial regression with lag of repression

|  | **Day-to-day** | **Total duration** |
| --- | --- | --- |
| Intercept | −2.366\*\*\* | −0.032 |
|  | (0.254) | (0.033) |
| Identifiable actors | 0.009 | 0.030\* |
|  | (0.043) | (0.013) |
| Students | 0.378\*\*\* | 0.129\*\*\* |
|  | (0.056) | (0.018) |
| Professional | 0.111 | 0.063\* |
|  | (0.095) | (0.028) |
| Labor unions | 0.164\*\* | 0.095\*\*\* |
|  | (0.055) | (0.017) |
| Repression | −0.190\*\*\* | 0.082\*\*\* |
|  | (0.057) | (0.016) |
| V-Dem Free Expression | −0.776+ | −0.019 |
|  | (0.466) | (0.068) |
| V-Dem Free Association | −0.134 | 0.062 |
|  | (0.515) | (0.073) |
| Lag of repression | −0.025 | −0.018 |
|  | (0.055) | (0.017) |
| 30-day pre-event momentum |  | 0.014\*\*\* |
|  |  | (0.002) |
| Num.Obs. | 27818 | 24277 |
| AIC | 19411.3 | 53787.2 |

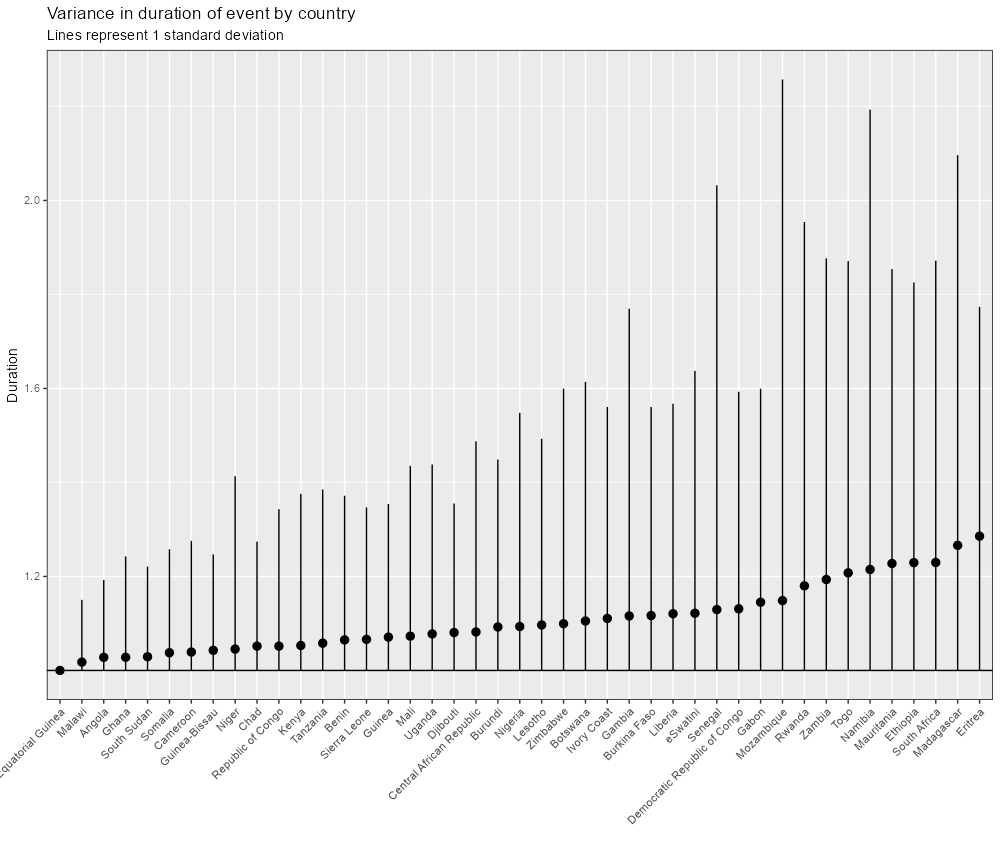
# Pooled models

Table - Pooled Day-to-day models

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) |
| (Intercept) | −4.713\*\*\* | −3.461\*\*\* | −4.282\*\*\* | −3.462\*\*\* | −2.710\*\*\* |
|  | (0.175) | (0.198) | (0.185) | (0.200) | (0.088) |
| lmaxpop | 0.192\*\*\* | 0.079\*\*\* | 0.148\*\*\* | 0.072\*\*\* |  |
|  | (0.013) | (0.015) | (0.014) | (0.017) |  |
| actor1\_student | 0.464\*\*\* | 0.489\*\*\* | 0.444\*\*\* | 0.487\*\*\* | 0.463\*\*\* |
|  | (0.051) | (0.051) | (0.051) | (0.052) | (0.053) |
| actor1\_professional | 0.030 | 0.012 | 0.016 | 0.033 | −0.028 |
|  | (0.087) | (0.087) | (0.087) | (0.088) | (0.090) |
| actor1\_labor | 0.209\*\*\* | 0.216\*\*\* | 0.173\*\*\* | 0.198\*\*\* | 0.180\*\*\* |
|  | (0.050) | (0.050) | (0.050) | (0.051) | (0.052) |
| capital |  | 0.555\*\*\* |  | 0.553\*\*\* | 0.666\*\*\* |
|  |  | (0.046) |  | (0.048) | (0.037) |
| repression |  | −0.238\*\*\* | −0.197\*\*\* | −0.201\*\*\* | −0.199\*\*\* |
|  |  | (0.052) | (0.052) | (0.054) | (0.053) |
| urban |  |  | 0.292\*\*\* |  |  |
|  |  |  | (0.045) |  |  |
| vdem\_free\_express |  |  |  | −0.902\*\*\* | −0.830\*\*\* |
|  |  |  |  | (0.207) | (0.203) |
| vdem\_free\_assoc |  |  |  | 1.072\*\*\* | 1.087\*\*\* |
|  |  |  |  | (0.216) | (0.213) |
| num\_actors |  |  |  |  | 0.101\* |
|  |  |  |  |  | (0.040) |
| Num.Obs. | 30095 | 30095 | 30095 | 29308 | 29888 |
| AIC | 21528.8 | 21368.4 | 21475.0 | 20829.6 | 21275.4 |
| BIC | 21570.4 | 21426.6 | 21533.2 | 20904.1 | 21350.1 |
| Log.Lik. | −10759.399 | −10677.190 | −10730.523 | −10405.781 | −10628.700 |
| F | 82.259 | 81.946 | 63.225 | 62.479 | 60.253 |
| RMSE | 0.32 | 0.32 | 0.32 | 0.32 | 0.32 |
| + p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001 |  |  |  |  |  |

Table - Pooled Duration Models

|  |  |  |  |
| --- | --- | --- | --- |
|  | (1) | (2) | (3) |
| (Intercept) | −0.021 | −0.039 | −0.054 |
|  | (0.056) | (0.028) | (0.092) |
| momentum\_30day | 0.013\*\*\* | 0.013\*\*\* | 0.010\* |
|  | (0.001) | (0.001) | (0.005) |
| capital | 0.051\*\*\* | 0.054\*\*\* | 0.002 |
|  | (0.014) | (0.012) | (0.032) |
| students | 0.125\*\*\* | 0.116\*\*\* | 0.050 |
|  | (0.017) | (0.018) | (0.053) |
| professional | 0.058\* | 0.047+ | 0.023 |
|  | (0.026) | (0.027) | (0.076) |
| labor | 0.097\*\*\* | 0.086\*\*\* | 0.023 |
|  | (0.016) | (0.017) | (0.042) |
| rep\_d1 | −0.043\*\* | −0.038\* | −0.003 |
|  | (0.016) | (0.016) | (0.047) |
| rep\_d2 | 1.067\*\*\* | 1.069\*\*\* | 1.188\*\*\* |
|  | (0.062) | (0.062) | (0.320) |
| lmaxpop | 0.004 |  |  |
|  | (0.004) |  |  |
| vdem\_free\_express |  | −0.038 | −0.051 |
|  |  | (0.065) | (0.177) |
| vdem\_free\_assoc |  | 0.088 | 0.060 |
|  |  | (0.068) | (0.176) |
| num\_actors |  | 0.031\* |  |
|  |  | (0.013) |  |
| protest\_size |  |  | 0.015 |
|  |  |  | (0.019) |
| Num.Obs. | 26526 | 26347 | 4250 |
| AIC | 58394.6 | 58048.9 | 8762.7 |
| BIC | 58476.5 | 58147.1 | 8838.9 |
| Log.Lik. | −29187.318 | −29012.457 | −4369.328 |
| F | 72.607 | 59.725 | 2.075 |
| RMSE | 0.50 | 0.50 | 0.24 |
| + p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001 |  |  |  |



# Evaluating pooled versus hierarchical modeling

## Overview

In order to determine whether a pooled or multilevel model is a best fit for the data, I use the lrtest to conduct likelihood ratio test, which compares the fit of two nested models. Unfortunately, the package does not support the two negative binomial models, and thus I re-run them as a pooled and multilevel linear regression. There are obvious disadvantages to this in terms of interpretation, but for the purposes of examining model fit, this is appropriate.

## Hypotheses

The null hypothesis is that the smaller, simpler model (the pooled model) fits the data as well as the larger, more complex model (the multilevel model). The alternative hypothesis is that the larger model fits the data better than the smaller model.

## Results

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Deg. Freedom** | **Log Likelihood** | **Deg. Freedom diff.** | **Chi Square** | **P-value** |
| Pooled | 10 | -19077 |  |  |  |
| Multilevel | 11 | -19059 | 1 | 36.395 | 0.000 |

The likelihood ratio test statistic is the difference in the log-likelihoods of the two models, multiplied by -2. The test statistic is 36.395, with one degree of freedom, which results in a p-value of 1.612e-09, a value much smaller than conventional levels of significance (p < 0.05). Therefore, I reject the null hypothesis and conclude that the larger model (multilevel) is a better fit for the data than the smaller model (pooled). The addition of the random effect for country significantly improved the fit of the model.

# Spatial autocorrelation

Table – Including spatial autocorrelation

|  | **Model 1** | **Model 2** | **Model 3** | **Model 4** |
| --- | --- | --- | --- | --- |
| Intercept | 0.060\*\*\* | 0.031\*\* | 0.032\*\* | 0.014 |
|  | (0.012) | (0.011) | (0.012) | (0.013) |
| Spatial autocorrelate | 0.054\*\*\* | 0.051\*\*\* | 0.051\*\*\* | 0.041\*\*\* |
|  | (0.006) | (0.006) | (0.006) | (0.007) |
| Students |  | 0.128\*\*\* | 0.128\*\*\* | 0.128\*\*\* |
|  |  | (0.017) | (0.017) | (0.017) |
| Labor |  | 0.097\*\*\* | 0.096\*\*\* | 0.096\*\*\* |
|  |  | (0.015) | (0.016) | (0.016) |
| Professional |  | 0.067\*\* | 0.066\*\* | 0.065\*\* |
|  |  | (0.025) | (0.025) | (0.025) |
| Repressed on day 1 |  |  | −0.009 | −0.012 |
|  |  |  | (0.016) | (0.016) |
| Capital |  |  |  | 0.052\*\*\* |
|  |  |  |  | (0.012) |
| SD (Intercept country) | 0.041 | 0.035 | 0.035 | 0.036 |
| Num.Obs. | 27065 | 27065 | 27065 | 27065 |
| R2 Marg. | 0.004 | 0.009 | 0.009 | 0.009 |
| R2 Cond. | 0.007 | 0.010 | 0.011 | 0.011 |
| AIC | 59936.9 | 59854.5 | 59856.1 | 59841.1 |
| BIC | 59969.7 | 59911.9 | 59921.8 | 59915.0 |
| ICC | 0.0 | 0.0 | 0.0 | 0.0 |
| RMSE | 0.52 | 0.51 | 0.51 | 0.51 |
| + p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001 |  |  |  |  |

**Note:** multilevel negative binomial regression structured with random intercepts at the country level. Administrative districts not included because of singular binding.

Table - Effect of urban/rural on protest size

|  | **Number of actors** | | **Size of protest** | |
| --- | --- | --- | --- | --- |
|  | **Rural** | **Urban** | **Rural** | **Urban** |
| (Intercept) | −0.033 | −0.270\*\*\* | −0.032 | −0.111 |
|  | (0.068) | (0.074) | (0.152) | (0.158) |
| Pre-event momentum | 0.005\* | 0.015\*\*\* | 0.002 | 0.008 |
|  | (0.003) | (0.002) | (0.017) | (0.005) |
| Number of actors | 0.020 | 0.025+ |  |  |
|  | (0.029) | (0.014) |  |  |
| Students | 0.056 | 0.138\*\*\* | 0.029 | 0.052 |
|  | (0.037) | (0.021) | (0.113) | (0.061) |
| Professional | 0.012 | 0.070\* | −0.018 | 0.040 |
|  | (0.055) | (0.031) | (0.167) | (0.086) |
| Labor | 0.020 | 0.107\*\*\* | 0.005 | 0.033 |
|  | (0.034) | (0.019) | (0.084) | (0.048) |
| Repressed day 1 | −0.025 | −0.031 | −0.007 | −0.001 |
|  | (0.028) | (0.020) | (0.085) | (0.056) |
| Repressed day 2 | 1.047\*\*\* | 1.062\*\*\* | 1.387\* | 1.070\* |
|  | (0.125) | (0.067) | (0.555) | (0.416) |
| Lag of duration | 0.038\* | 0.064\*\*\* | 0.009 | 0.070\* |
|  | (0.019) | (0.012) | (0.060) | (0.033) |
| lmaxpop | 0.002 | 0.016\*\* | 0.000 | −0.001 |
|  | (0.005) | (0.005) | (0.012) | (0.012) |
| protest\_size |  |  | 0.012 | 0.017 |
|  |  |  | (0.037) | (0.022) |
| SD (Intercept admin1) | 0.000 | 0.000 | 0.000 | 0.000 |
| SD (Intercept country) | 0.000 | 0.011 | 0.000 | 0.000 |
| Num.Obs. | 8558 | 15984 | 1207 | 2836 |
| AIC | 18237.6 | 35842.1 | 2479.5 | 5893.8 |
| BIC | 18329.3 | 35941.9 | 2545.8 | 5971.2 |
| ICC |  | 0.0 |  |  |
| RMSE | 0.40 | 0.53 | 0.19 | 0.26 |

# Alternative datasets

## SCAD Dataset

There are several important ways in which the SCAD data differ from the ACLED data. First, the temporal scope is different. Whereas the ACLED data include incidents from 1997 to 2021, the SCAD data include incidents from 1990 to 2017. Second, in terms of size, the datasets are quite different. The SCAD data include 4,867 single- and multi-day protests while the ACLED data include 27,065 single- and multi-day events. This limited size has been the source of recent criticism of the SCAD dataset (Herkenrath and Knoll 2011; Demarest and Langer 2018).

Almost all the variables tested in Approach 2 can be replicated in the SCAD data, save the intra-event repression variables. To address this, we include a simple variable that measures whether an event experienced repression at all. The drawback to this variable is obvious: within multiday events, it is impossible to know when the repression occurred. The other variables are replicated as best as possible, including whether an event occurred in the capital as well as the participation of student, professional, and labor groups.

Using the SCAD data does provide the opportunity to measure size, albeit imperfectly. The SCAD data include a variable that measures the estimated “total number of participants” in an event on a 7-point scale ranging from “less than 10” to “more than 1,000,000.” This variable provides three challenges. First, the interval does not allow for enough granularity to differentiate between an event with 10,001 participants and 50,000 participants. Second, this variable takes NA values in nearly 40 percent of observations within the SCAD data. While there is not necessarily any reason, *prima facie*, to believe that NA values are systemically included in the data, the process of collecting the data may favor mid-size and large events over small events. If that is the case, then event duration will be spuriously correlated with mid-size and large events. Third, because the SCAD data cannot provide information about intra-event developments, it is impossible to know when within the event the size was recorded. This is a potential source of endogeneity, as size may be a function of duration rather than the other way around.

We turn again to a multilevel negative binomial regression model structured at the first administrative district and country level to test the SCAD data. The results are visualized below in **Error! Reference source not found.** and the coefficient table can be found in the supplementary materials. H1 finds support, and the variable for size is positively and statistically significantly correlated with overall event duration. H2 also finds support, as the variable indicating that an event occurred in a capital is also positive and statistically significant. H3 finds partial support: while student and professional organization participation both positively and significantly correlate with extended duration, the variable for labor is indistinguishable from zero. The repression variable is positively and significantly correlated with duration, which runs counter to H4 but lends some support to the interesting finding regarding repression on the second day of an event.

Chart

Description automatically generated

# Addressing selection issues and endogeneity

## Selecting on repression

In the first selection model, we examine whether selecting on repression affects the outcomes in our main results. We develop a logistic selection model that regresses whether an event is repressed (no = 0, yes = 1) and then run a negative binomial model that includes our main factors (event size, participants, location). This model can be found below in Table 16.

Table - 2 Stage Selection Model on Repression

|  |  |  |  |
| --- | --- | --- | --- |
|  | Stage 1  Logistic | Stage 2  Neg. Bin | Stage 3  Neg. Bin |
| Intercept | −1.838\*\*\* | −0.069\*\* | 0.969\*\*\* |
|  | (0.063) | (0.023) | (0.021) |
| Lag of repression | 1.117\*\*\* |  |  |
|  | (0.038) |  |  |
| Lag of duration | −0.105\*\* | 0.070\*\*\* | 0.080\*\*\* |
|  | (0.034) | (0.009) | (0.009) |
| Pre-event momentum | 0.026\*\*\* |  |  |
|  | (0.004) |  |  |
| Capital | 0.204\*\*\* | 0.071\*\*\* | 0.060\*\*\* |
|  | (0.035) | (0.012) | (0.010) |
| Number of actors | −0.019 | 0.044\*\*\* | −0.271\*\*\* |
|  | (0.040) | (0.013) | (0.041) |
| Issue: Hum. Rights / Democracy | −0.306\* |  |  |
|  | (0.133) |  |  |
| Issue: Economy | −0.437\*\*\* |  |  |
|  | (0.077) |  |  |
| Students |  | 0.138\*\*\* | 1.458\*\*\* |
|  |  | (0.018) | (0.244) |
| Labor |  | 0.102\*\*\* | 1.139\*\*\* |
|  |  | (0.017) | (0.106) |
| Repressed |  | 0.075\*\*\* | 0.082 |
|  |  | (0.016) | (0.068) |
| Propensity score |  | −0.074 |  |
|  |  | (0.079) |  |
| Num.Obs. | 24969 | 24969 | 24969 |
| AIC | 22027.3 | 55314.6 | 54325.5 |
| BIC | 22092.3 | 55387.8 | 54390.6 |

**Note:** correlation between repressed and propensity score = 0.21, p < 0.01

We find that the results are concordant with our primary results. The presence of students and labor unions remain positively and significantly correlated with extended event duration. Protesting in the capital also correlates with extended duration. Interestingly, the number of actors correlates inversely with overall duration, suggesting that further study should be done. The propensity score, which draws the likelihood of treatment from the first stage, is not correlated with extended duration in the Stage 2 model. It is likely that repression does not attain significance because the selection model already accounts for whether an event is or is not repressed.

## Selecting on student participants

Secondly, it is likely important to evaluate whether incorporating the factors that lead students to participate in a protest also affect the effect of other coefficients. We develop a logistic selection model that regresses whether students are present (no = 0, yes = 1) on covariates that we believe are important, including a lag of repression, lag of duration, pre-event momentum, whether an event occurs in an urban area, and several key issues (economy, selections, human rights).

Table - 2 Stage Selection Model on student participation

|  |  |  |  |
| --- | --- | --- | --- |
|  | Stage 1  Logit | Stage 2  Neg. Bin | Stage 3  Neg. Bin |
| Intercept | −2.450\*\*\* | −0.070\*\* | −0.842\*\*\* |
|  | (0.055) | (0.025) | (0.164) |
| Lag of repression | 0.128\* |  |  |
|  | (0.051) |  |  |
| Lag of duration | 0.049 | 0.074\*\*\* | 0.061\*\*\* |
|  | (0.034) | (0.010) | (0.011) |
| Pre-event momentum | 0.028\*\*\* |  |  |
|  | (0.005) |  |  |
| Urban setting | 0.463\*\*\* | 0.061\*\*\* | −0.047\* |
|  | (0.045) | (0.015) | (0.019) |
| Issue: economy | −0.789\*\*\* |  |  |
|  | (0.101) |  |  |
| Issue: election | −1.283\*\*\* |  |  |
|  | (0.176) |  |  |
| Issue: hum. Rights/democracy | −1.549\*\*\* |  |  |
|  | (0.272) |  |  |
| Number of actors |  | 0.042\*\*\* | 1.651\*\*\* |
|  |  | (0.013) | (0.164) |
| Students |  | 0.134\*\*\* | 0.345 |
|  |  | (0.018) | (0.222) |
| Labor |  | 0.097\*\*\* | 0.210 |
|  |  | (0.017) | (0.139) |
| Repressed |  | 0.077\*\*\* | 0.115 |
|  |  | (0.015) | (0.080) |
| Propensity score |  | −0.157 |  |
|  |  | (0.198) |  |
| Num.Obs. | 24969 | 24969 | 24969 |
| AIC | 17367.2 | 55329.1 | 64943.3 |
| BIC | 17432.2 | 55402.2 | 65008.3 |

**Note:** correlation between propensity and students is 0.12, p < 0.01

We find that, when incorporating the selection factors that lead students to participate in protests, the number of actors attains a large, positive coefficient that is statistically significant. Students are significantly more likely to protest in urban settings, but once this propensity is accounted for, urban settings themselves correlate negatively with overall duration. Unsurprisingly, students are more likely to protest when a previous event has been repressed. Quite surprisingly, students are less likely to protest when major issues with the economy, elections, or human rights/democracy are at play. When accounting for the factors that lead students to protest, labor unions no longer correlate with longer duration events.

## Selection on labor union participation

Thirdly, we examine the factors that lead labor unions to engage in protest. We regress labor union participation (no = 0, yes = 1) on a lag of repression, pre-event momentum, presence in the capital, economic issues, election issues, and human rights issues). Unsurprisingly, economic issues correlate strongly and significantly with the likelihood that a labor union will participate in a protest.

Table - 2 Stage Selection Model on labor participation

|  |  |  |  |
| --- | --- | --- | --- |
|  | Stage 1  Logit | Stage 2  Neg. Bin | Stage 3  Neg. Bin |
| Intercept | −1.899\*\*\* | −0.119\* | −0.490\* |
|  | (0.029) | (0.060) | (0.197) |
| Lag of repression | −0.276\*\*\* |  |  |
|  | (0.051) |  |  |
| Pre-event momentum | 0.024\*\*\* | 0.012\*\*\* |  |
|  | (0.004) | (0.002) |  |
| Capital | 0.049 | 0.046\*\*\* | 0.056\*\*\* |
|  | (0.037) | (0.012) | (0.016) |
| Issue: economy | 1.225\*\*\* | −0.058 | −0.266+ |
|  | (0.053) | (0.104) | (0.139) |
| Issue: election | −0.221\* | 0.062+ |  |
|  | (0.105) | (0.033) |  |
| Issue: hum. Rights / democracy | −0.059 | 0.052 |  |
|  | (0.131) | (0.041) |  |
| Number of actors |  | 0.051\*\*\* | 1.193\*\*\* |
|  |  | (0.012) | (0.250) |
| Labor |  | 0.072\*\*\* | 1.507\* |
|  |  | (0.017) | (0.718) |
| Repressed |  | 0.074\*\*\* | 0.320\* |
|  |  | (0.016) | (0.125) |
| Propensity score |  | 0.685 |  |
|  |  | (0.464) |  |
| Num.Obs. | 24969 | 24969 | 24969 |
| AIC | 20683.4 | 55316.7 | 67832.9 |
| BIC | 20740.3 | 55406.1 | 67889.7 |

**Note:** correlation between propensity and students is 0.17, p < 0.01

We find that when selecting for labor participation, protests in the capital are positively and significantly correlated with longer durations. The number of actors is also positively and significantly correlated with extended duration. This makes sense, as labor unions may be more likely to band together to achieve shared goals. In fact, a T-Test measuring the average number of actors present at protests demonstrates that when labor unions are present, more observable actors are present:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | No labor | Labor | Lower | Upper | P |
| Num. Actors | 1.07 | 1.36 | -0.32 | -0.27 | 0.00 |

Having examined several potential selection issues, we now turn to the use of a statistical matching process. To do so, we generate pseudo “control” and “treatment” groups by matching similar protest events along our core parameters as well as additional parameters that create a set of “matched” events. In addition to matching along our core covariates, we also ensure that events are matched to the country, admin1, and admin2 levels. We also ensure that protests occur in the same year and are about similar issues (economics, elections, human rights and democracy). Doing this allows us to run bivariate regression with our “treatment” variable and to estimate its quasi-causal effect on the dependent variable.

## Coarsened Exact Matching – Number of actors

We turn to the results of a coarsened exact match in which we establish the number of actors – or, specifically, whether more than one actor is present – as the core quasi-causal variable. This variable is called “Binary Actors” and takes a 0 if only 1 actor is present and a 1 if more than 1 actor is present.

Table - Matching data on number of actors

|  |  |  |
| --- | --- | --- |
|  | Untreated  (1 actor present) | Treated  (2+ actors present) |
| All | 22596 | 2373 |
| Matched | 21043 | 2313 |
| Unmatched | 1553 | 60 |

Table - Causal estimate for number of actors

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | T Value | P-Value |
| Intercept | 1.197909\*\*\* | 0.004169 | 287.3339 | 0.00 |
| Binary Actors | 0.079220\*\*\* | 0.013248 | 5.9798 | 0.00 |

## Coarsened Exact Matching – Students Present

Table - Matching data on number of actors

|  |  |  |
| --- | --- | --- |
|  | Untreated  (no students) | Treated  (students) |
| All | 22123 | 2846 |
| Matched | 20122 | 2822 |
| Unmatched | 2001 | 24 |

Table - Causal estimate for number of actors

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | T Value | P-Value |
| Intercept | 1.1212614\*\*\* | 0.0035749 | 313.649 | 0.00 |
| Students | 0.1654147 | 0.0101934 | 16.228 | 0.00 |

## Coarsened Exact Matching – Labor Present

Table - Matching data on number of actors

|  |  |  |
| --- | --- | --- |
|  | Untreated  (no labor) | Treated  (labor) |
| All | 21188 | 3781 |
| Matched | 19904 | 3731 |
| Unmatched | 1284 | 50 |

Table - Causal estimate for number of actors

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | T Value | P-Value |
| Intercept | 1.1229912\*\*\* | 0.0037084 | 302.827 | 0.00 |
| Labor | 0.1144789\*\*\* | 0.0093336 | 12.265 | 0.00 |

## Coarsened Exact Matching – In the capital

Table - Matching data on number of actors

|  |  |  |
| --- | --- | --- |
|  | Untreated  (not in capital) | Treated  (in capital) |
| All | 12897 | 12072 |
| Matched | 12841 | 11792 |
| Unmatched | 56 | 280 |

Table - Causal estimate for number of actors

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | T Value | P-Value |
| Intercept | 1.0955395\*\*\* | 0.0046095 | 237.667 | 0.00 |
| In capital | 0.0861938\*\*\* | 0.0066623 | 12.938 | 0.00 |

## Coarsened Exact Matching – Repressed

Table - Matching data on number of actors

|  |  |  |
| --- | --- | --- |
|  | Untreated  (not repressed) | Treated  (repressed) |
| All | 20658 | 4311 |
| Matched | 19649 | 4244 |
| Unmatched | 1009 | 67 |

Table - Causal estimate for number of actors

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | T Value | P-Value |
| Intercept | 1.1191809\*\*\* | 0.0035948 | 311.33 | 0.00 |
| Repressed | 0.0898200\*\*\* | 0.0085296 | 10.53 | 0.00 |

# Including riots in the analysis

Table - Next day (including riots)

|  | (1) | (2) | (3) | (4) | (5) |
| --- | --- | --- | --- | --- | --- |
| Intercept | −4.013\*\*\* | −3.574\*\*\* | −3.899\*\*\* | −3.453\*\*\* | −2.760\*\*\* |
|  | (0.264) | (0.314) | (0.265) | (0.349) | (0.187) |
| Log of local population | 0.102\*\*\* | 0.062\* | 0.091\*\*\* | 0.062\* |  |
|  | (0.020) | (0.026) | (0.020) | (0.026) |  |
| Number of actors |  |  |  |  | 0.026 |
|  |  |  |  |  | (0.036) |
| Actors: students | 0.262\*\*\* | 0.266\*\*\* | 0.255\*\*\* | 0.272\*\*\* | 0.261\*\*\* |
|  | (0.039) | (0.039) | (0.039) | (0.040) | (0.040) |
| Actors: professionals | 0.052 | 0.052 | 0.051 | 0.056 | 0.038 |
|  | (0.079) | (0.079) | (0.079) | (0.080) | (0.082) |
| Actors: labor union | 0.069+ | 0.068 | 0.058 | 0.060 | 0.065 |
|  | (0.042) | (0.042) | (0.042) | (0.043) | (0.044) |
| Capital |  | 0.219\* |  | 0.229\* | 0.371\*\*\* |
|  |  | (0.089) |  | (0.091) | (0.069) |
| Repression / Day 1 Rep. |  | −0.186\*\*\* | −0.189\*\*\* | −0.185\*\*\* | −0.187\*\*\* |
|  |  | (0.048) | (0.048) | (0.049) | (0.048) |
| Urban |  |  | 0.139\*\* |  |  |
|  |  |  | (0.045) |  |  |
| V-Dem free expression |  |  |  | −0.146 | −0.163 |
|  |  |  |  | (0.325) | (0.332) |
| V-Dem free association |  |  |  | −0.077 | −0.079 |
|  |  |  |  | (0.342) | (0.350) |
| SD (Intercept admin2) | 0.589 | 0.596 | 0.583 | 0.608 | 0.622 |
| SD (Intercept admin1) | 0.401 | 0.388 | 0.399 | 0.393 | 0.399 |
| SD (Intercept country) | 0.464 | 0.460 | 0.461 | 0.479 | 0.453 |
| Num.Obs. | 50461 | 50461 | 50461 | 49107 | 50045 |
| R2 Marg. | 0.009 | 0.011 | 0.012 | 0.011 | 0.011 |
| R2 Cond. | 0.187 | 0.188 | 0.187 | 0.196 | 0.195 |
| AIC | 39023.1 | 39006.4 | 39003.1 | 38015.9 | 38604.7 |
| BIC | 39093.8 | 39094.7 | 39091.4 | 38121.5 | 38710.6 |
| ICC | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |
| RMSE | 0.34 | 0.34 | 0.34 | 0.34 | 0.34 |

Table - Duration (including riots)

|  | (1) | (2) | (3) | (4) |
| --- | --- | --- | --- | --- |
| Intercept | −0.078 | −0.131 | −0.120 | −0.082 |
|  | (0.060) | (0.086) | (0.084) | (0.060) |
| Pre-event momentum | 0.012\*\*\* | 0.007 | 0.008 | 0.012\*\*\* |
|  | (0.002) | (0.005) | (0.005) | (0.002) |
| Capital | 0.049\*\* | 0.001 | 0.000 | 0.048\*\* |
|  | (0.016) | (0.032) | (0.033) | (0.015) |
| Actors: students | 0.140\*\*\* | 0.042 | 0.045 | 0.147\*\*\* |
|  | (0.018) | (0.055) | (0.055) | (0.018) |
| Actors: professionals | 0.068\* | 0.032 | 0.038 | 0.070\*\* |
|  | (0.027) | (0.079) | (0.078) | (0.027) |
| Actors: labor union | 0.100\*\*\* | 0.016 | 0.021 | 0.089\*\*\* |
|  | (0.016) | (0.044) | (0.043) | (0.017) |
| Repression / Day 1 Rep | −0.018 | 0.013 | 0.013 | −0.015 |
|  | (0.017) | (0.048) | (0.048) | (0.017) |
| Lag of duration | 0.058\*\*\* | 0.053+ | 0.054+ | 0.058\*\*\* |
|  | (0.010) | (0.029) | (0.029) | (0.010) |
| Log of local population | 0.003 |  |  | 0.002 |
|  | (0.005) |  |  | (0.005) |
| V-Dem free expression |  | −0.043 | −0.047 |  |
|  |  | (0.102) | (0.107) |  |
| V-Dem free association |  | 0.065 | 0.070 |  |
|  |  | (0.107) | (0.102) |  |
| Protest size |  | 0.016 | 0.016 |  |
|  |  | (0.019) | (0.019) |  |
| Number of actors |  | 0.012 |  |  |
|  |  | (0.031) |  |  |
| Issue: economy |  |  |  | 0.087\*\*\* |
|  |  |  |  | (0.022) |
| Issue: service delivery |  |  |  | 0.076\*\* |
|  |  |  |  | (0.027) |
| Issue: election |  |  |  | 0.078\* |
|  |  |  |  | (0.031) |
| Issue: corruption |  |  |  | 0.110\* |
|  |  |  |  | (0.044) |
| Issue: Pandemic |  |  |  | 0.005 |
|  |  |  |  | (0.040) |
| Num.Obs. | 24542 | 3977 | 3977 | 24542 |
| R2 Marg. | 0.013 | 0.003 | 0.003 | 0.015 |
| AIC | 54279.8 | 8226.3 | 8224.5 | 54256.0 |
| BIC | 54377.1 | 8320.7 | 8312.5 | 54393.9 |
| RMSE | 0.51 | 0.25 | 0.25 | 0.50 |

# Holding the day of the event constant

Does it matter what day of an event it is in terms of understanding the likelihood of observing another one? To address this, we examine the base likelihood that a protest will continue on to the next day based on the day of the protest. We chart these below in Table 31, along with the number of observations associate with each day.

Table - Likelihood of observing another day

| Day of event | Likelihood of next day | Observations |
| --- | --- | --- |
| 1 | 9.27% | 27,065 |
| 2 | 26.03% | 2,509 |
| 3 | 34.46% | 653 |
| 4 | 46.67% | 225 |
| 5 | 51.43% | 105 |
| 6 | 62.96% | 54 |
| 7 | 44.12% | 34 |
| 8 | 66.67% | 15 |
| 9 | 80.00% | 10 |
| 10 | 75.00% | 8 |
| 11 | 33.33% | 6 |
| 12 | 100.00% | 2 |
| 13 | 50.00% | 2 |
| 14 | 100.00% | 1 |
| 15 | 0.00% | 1 |

What one notes is that the likelihood of observing a next day drastically increases after the first day has completed. Most events are single-day events, and the base likelihood that an event will stretch into the next day sits at 9.27 percent. However, if an event makes it to the second day, then the likelihood that it will see a third day nearly triples to 26 percent. This trend continues, though as events stretch past a week, the sample becomes small enough that drawing either representative statistics or inference becomes a risky proposition.

We then re-ran our “next-day” models and included a variable for the day of the protest as a covariate (see Table 32). We find a substantively large and statistically significant correlation. We then plot out the predictions in Figure 3.

Because we have (quite substantially) overrun the word limit, we relegate these findings to the appendix with a quick footnote that mentions this very problem. Thank you for the suggestion!

Table - Likelihood of next day, holding day of event constant

|  | (1) | (2) | (3) |
| --- | --- | --- | --- |
| Intercept | −4.109\*\*\* | −3.338\*\*\* | −3.423\*\*\* |
|  | (0.447) | (0.232) | (0.464) |
| Log of local population | 0.069\* |  |  |
|  | (0.034) |  |  |
| Number of actors |  | 0.019 |  |
|  |  | (0.043) |  |
| Protest size |  |  | −0.075 |
|  |  |  | (0.065) |
| Actors: students | 0.262\*\*\* | 0.245\*\*\* | 0.320+ |
|  | (0.066) | (0.067) | (0.188) |
| Actors: professionals | 0.118 | 0.107 | −0.227 |
|  | (0.094) | (0.095) | (0.271) |
| Actors: labor union | 0.145\*\* | 0.147\*\* | 0.205 |
|  | (0.055) | (0.056) | (0.139) |
| Capital | 0.312\*\* | 0.468\*\*\* | 0.636\*\*\* |
|  | (0.117) | (0.086) | (0.176) |
| Repression / Day 1 Rep. | −0.172\*\* | −0.173\*\* | 0.221 |
|  | (0.057) | (0.056) | (0.155) |
| V-Dem free expression | −0.390 | −0.384 | −0.686 |
|  | (0.456) | (0.445) | (1.014) |
| V-Dem free association | −0.103 | −0.096 | 0.303 |
|  | (0.492) | (0.483) | (1.033) |
| Issue: economy | 0.168\* | 0.163\* | 0.321+ |
|  | (0.069) | (0.069) | (0.167) |
| Issue: hum. rights/dem | −0.080 | −0.080 | 0.318+ |
|  | (0.075) | (0.075) | (0.172) |
| Issue: education | 0.195\* | 0.199\*\* | 0.171 |
|  | (0.077) | (0.077) | (0.218) |
| Issue: service delivery | −0.082 | −0.080 | −0.337 |
|  | (0.087) | (0.086) | (0.241) |
| Issue: corruption | 0.018 | 0.011 | 0.205 |
|  | (0.144) | (0.144) | (0.292) |
| Day within the event | 0.474\*\*\* | 0.473\*\*\* | 0.394\*\*\* |
|  | (0.023) | (0.023) | (0.074) |
| Num.Obs. | 29308 | 29888 | 4991 |
| AIC | 19228.5 | 19577.7 | 2954.0 |
| BIC | 19377.6 | 19727.2 | 3071.3 |

Figure - Predicted "next day" based on current day of event

A graph with a line

Description automatically generated

# Additional qualitative evidence

## On the South African student protests

An example of such mechanisms emerges from student-led protests in South Africa. October 2015 saw the birth of the #FeesMustFall protest movement, whereby students took to the streets to voice their frustration against a proposed increase in university tuition. Starting at the University of the Witwatersrand, Johannesburg, the revolt then spread nationwide, incorporating at least ten other higher education institutions (Booysen 2016; Hewlett et al. 2016; Mavunga 2019). The movement stemmed from students themselves, many of whom had outstanding grievances with higher educational institutions, with the #RhodesMustFall protests causing uproar amongst this population earlier in the year (Eve Fairbanks 2015). #FeesMustFall saw thousands of students mobilize, a notable figure given the reported increase in South Africa’s campus-based activism and a further signal of student frustrations. Before this, hundreds of students had stormed Cape Town’s parliamentary gates. Despite initial reports of violence, the protests continued in duration, resulting in the closure of 26 universities by students. In response, the Zuma-led African National Congress intervened, conceding to the movement's demands (Cini 2019).

1. The use of ordinal values rather than exact point estimates is well-established in datasets such as NAVCO (citation) and SCAD (citation). [↑](#footnote-ref-1)