Online Appendix for *Chaos on Campus: Universities and Mass Political Protest*

January 13, 2020

1 Overview

This appendix performs several additional tests that elaborate on, and assess the sensitivity of, our main result. First, in section 2 we conduct country-year analyses probing whether the effects of universities identified at the grid-cell level aggregate to the level of countries, supplementing the analyses without country-fixed effects (that also allow for country-level variation) in table 1 in the main text. This speaks to the concern that universities do not generate sui generis protests, but simply attract already existing protests happening elsewhere. In section 2.1 we interact our university measure with economic crises, to explore heterogeneous effects of universities in situations with and without economic crises. Section 3 estimates models where we replace our benchmark university indicator with a measure of “university stock” to capture more long-run effects of universities on protest. In section 4 we re-investigate our main result while controlling for education levels, since a number of our mechanisms concern other pathways than that between universities and the education level of citizens.

Section 5 presents and discusses the instrumental variable analysis. In section 6, we provide tests to probe whether our results are due to reporting bias – which would occur if university locations are more likely to generate news reports of protest, conditional on there being an actual protest. In section 7, we gauge how sensitive our results are to using different estimators than the ones used in the baseline models. After this, section 8 looks at whether results differ when we dis-aggregate our data on universities into universities that are privately owned and universities that are public. Section 9 conducts a placebo analysis. Section 10 tests whether the identified university coefficient is stable across the two world regions in our sample (Africa and Central America/the Caribbean). In section 11 we conduct the semester analysis, and describe the procedure for coding the academic semester (and sources of missingness). Section 12 presents some descriptives for the main variables used in the core
empirical analyses, while section (section 13) describes the variables and data used in the main analysis. Finally, section 13.4 provides an overview of universities by country (and the percentage of universities established after the start of our time-series for each country).

2 Country-year analyses

Models 1.7 and 1.8 (which do not include cell-fixed effects) and allow for between-country comparisons address the question of whether universities simply attract protests rather than creating new protests. However, a more clean test of this question is to estimate models at the country level. If universities do nothing more than change the geographic location of protest, without affecting the aggregate number of protests, we would observe a positive relationship when investigating grid-cells, but no relationship when investigating country-years as our unit of analysis. This is because geographic “re-shuffling” of protest within a country will not change the aggregate level of protests in a country. Table 1 addresses this issue. It investigates the relationship between (the number of) universities and protest incidence at the country-year level, estimating OLS models (of the number of protests). Some models include covariates for GDP, population, and urbanization, which should could affect both universities and protest. We include country- and year-fixed effects, plus a lagged dependent variable, in different models. Since the covariates are potentially post-treatment, they are not included in all models. Since variation is at the country-year level, it is not possible to include country-year interactions (as we do in the grid-cell models). These country-level results are shown in table 1. These estimations yield similar results as the grid-cell analyses: Increases in the number of universities yield increases in protest, also at the country level. This suggests that universities affect protest in other ways than simply attracting protest activity that would otherwise have occurred elsewhere.
Table 1: Country-year models

|                | OLS  | OLS  | OLS  | OLS  | OLS  | OLS  |
|----------------|------|------|------|------|------|------|
|                | (A1) | (A2) | (A3) | (A4) | (A5) | (A6) |
| L(universities) | 0.329*** | 0.343* | 0.334* | 0.186* | 0.472*** | 0.468*** |
|                | (3.18) | (2.18) | (2.16) | (2.05) | (4.00) | (4.02) |
| Protest\(_{t-1}\) | 0.036* | 0.015* | 0.076*** | 0.025* |
|                | (2.15) | (2.53) | (4.89) | (2.65) |
| GDP p.c.       | -0.348 | -0.157 | -0.172 |
|                | (-1.91) | (-0.41) | (-0.48) |
| Population     | 0.493** | -0.608 | -0.556 |
|                | (3.45) | (-0.37) | (-0.35) |
| Urbanization   | 0.142 | -15.934*** | -15.267*** |
|                | (0.14) | (-3.80) | (-3.86) |

|                | Country-FE | Year-FE |
|----------------|-------------|---------|
|                | ✓           | ✓       |
| N              | 1440        | 1440    |
| R\(^2\)       | 0.195       | 0.108   |
| Countries      | 61          | 58      |
| Years          | 23          | 23      |

Notes: Standard errors clustered on countries. Intercept omitted from table. T-values (OLS) in parentheses. The independent variable is log number of universities in a country, GDP p.c., urbanization, and population taken from V-DEM dataset.
2.1 Interaction with economic crises

One interesting question is whether there is an interaction between universities and economic crises on protest frequency. This is suggested by Campante and Chor (2012), who find that the highly educated are more likely to protest when economic circumstances are dire, because people with education will find it harder to translate education into a job, causing frustrations. While they find evidence for this at the individual level, it is not given that the general equilibrium effect is positive at the macro-level of countries undergoing economic crises. Countervailing effects could operate as well, for example if university education can serve as an alternative to being on the job market in economic downturns. To gauge this, we estimate an interaction term between our university variable and economic crises, with protest as the outcome. Since crises are conventionally measured at the country-level, we implement this in a country-year framework. Since there is no consensus regarding what constitutes an economic crisis, we opt for the arguably least controversial measure, using an indicator that equals 1 if there is negative growth in a given year. We interact this measure with the log of universities variable. These models include the same covariates as the country year models in table 1 above. We estimate models both with and without covariates and fixed effects. The results can be found in table 2 below. Here, we find no evidence of a significant relationship between crises, universities and protests.

Table 2: Country-year models with economic-crisis interactions

|                  | OLS  | OLS  | OLS  | OLS  | OLS  | OLS  |
|------------------|------|------|------|------|------|------|
|                  | (A7) | (A8) | (A9) | (A10)| (A11)| (A12)|
| L(universities)  | 0.388** | 0.319* | 0.321* | 0.183* | 0.476*** | 0.474*** |
|                  | (3.28) | (2.15) | (2.20) | (2.06) | (4.25) | (4.31) |
| Econ. crisis     | 0.762** | 0.133 | 0.102 | 0.282 | 0.085 | 0.067 |
|                  | (2.66) | (0.62) | (0.49) | (1.30) | (0.39) | (0.31) |
| Econ. crisis × L(universities) | -0.057 | 0.049 | 0.027 | 0.018 | -0.012 | -0.017 |
|                  | (-0.58) | (0.78) | (0.45) | (0.26) | (-0.23) | (-0.31) |
| Protest\_{t-1}  | 0.035* | 0.014* | 0.074*** | 0.025* | 0.035 | 0.025 |
|                  | (2.15) | (2.53) | (4.79) | (2.64) | (2.61) | (2.61) |
| GDP p.c.         | -0.299 | -0.146 | -0.169 | -0.146 | -0.169 | -0.169 |
|                  | (-1.57) | (-0.38) | (-0.47) | (-0.38) | (-0.47) | (-0.47) |
| Population       | 0.506*** | -0.576 | -0.533 | -0.576 | -0.533 | -0.533 |
|                  | (3.57) | (-0.35) | (-0.33) | (-0.35) | (-0.33) | (-0.33) |
| Urbanization     | 0.104 | -15.996*** | -15.326*** | -15.996*** | -15.326*** | -15.326*** |
|                  | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Country-FE       | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year-FE          | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| N                | 1440 | 1440 | 1440 | 1215 | 1215 | 1215 |
| R²               | 0.206 | 0.109 | 0.121 | 0.315 | 0.129 | 0.135 |
| Countries        | 61 | 58 | 58 | 61 | 58 | 58 |
| Years            | 23 | 23 | 23 | 23 | 23 | 23 |

Notes: Standard errors clustered on countries. Intercept omitted from table. T-values (OLS) in parentheses. The independent variable is log number of universities in a country, covariates are GDP p.c., population, and urbanization, taken from V-DEM dataset.
To gauge the impact of having a long history of universities in a location (on protest), we create a university-stock variable, and estimate the relationship between this and protest. This primarily relies on between-cell variation since a large component of the stock variable is time-invariant and the idea behind a stock variable is that there are long-term effects of universities on protest. We therefore estimate this in a cross-section. The stock variable is created by multiplying the number of universities in a grid-cell with the number of years these universities have been present in the cell. We calculate this variable in a cross-sectional dataset (since this is the dimension of variation employed) and use it to estimate the outcome; the absolute number of protest events in the time-series we have data for in SCAD, 1990-2016. We operationalize our covariates by taking their means over the period. Furthermore, we separate between pre-treatment covariates (that are exogenous to universities) and economic and demographic covariates (income, urbanization etc.) that also have some post-treatment component (that are partly affected by universities). The former category includes share of mountainous terrain, distance to capital and border, forest cover, shrubland, barren land, and agricultural suitability, while the latter category includes gross-cell product (p.c.), population, urban density, night lights and a spatial lag of the outcome.

We estimate the association between log number of protest events in the period and university-stock in different models in table 3. We proceed with a parsimonious model only including fixed effects (column 1), adding a spatial lag of the outcome (column 2), with pre-treatment geographic covariates (column 3) and with covariates (economic and demographic) that have some post-treatment component (column 4). University-stock is strongly related to higher protest frequencies, across all four specifications.

Table 3: Estimating the impact of university-stock

| Outcome: university-stock | (protest events) |
|---------------------------|------------------|
| L(university stock) | 0.457*** 0.454*** 0.376*** 0.302*** |
| | (11.45) (11.40) (10.00) (8.00) |
| Geographic covariates | ✓ ✓ ✓ ✓ |
| Economic and demographic covariates | ✓ ✓ ✓ ✓ |
| LDV | ✓ ✓ ✓ ✓ |
| N | 12086 12086 11309 11054 |
| R² | 0.188 0.190 0.216 0.244 |
| Country-FE | ✓ ✓ ✓ ✓ |

Notes: Unit of analysis: Grid-cells cross section 1990-2016. Standard errors clustered on grid-cells. Intercept omitted from table. T-values (OLS) in parentheses. The independent variable is the log of university stock.
4 Controlling for local-level education

We investigate whether universities increase protests through increasing local education levels. Using local-level data from the DHS surveys, Dahlum and Wig (2019) show that education increases local-level protests in Africa. Since we conjecture that universities cause protests also when holding local education levels constant, partly by improving coordination, but also through affecting the attitudes of not-yet graduated students, we expect a link between universities and protest also when controlling for local education levels.

Table 4 re-estimates our baseline models while controlling for the local (average) level of education, drawing on the DHS surveys as they are used and operationalized in grid-cell format in the paper by Dahlum and Wig (2019). We use an indicator measuring the average highest level of education received, ranging from no formal education to completed tertiary education. Controlling for education reduces the number of observations in the sample quite dramatically, since only a few grid-cells and years have DHS coverage. The sample is reduced to around 6000 observations in most of the models. Education is also a potentially post-treatment control, making us reluctant to use this in our benchmark model. Using this control, we estimate models with lagged dependent variables, country- and year-fixed effects, as well as with country-year interactions and grid-cell fixed effects (in different combinations), in similar fashion as in the benchmark table. We estimate models with and without our benchmark covariates. Models controlling for education are shown in Table 4. Even with this greatly reduced sample, and this added control, the university coefficient is in the expected direction and is precisely estimated in all models.

The education coefficient is positive in some of the models, in line with the findings in Dahlum and Wig (2019), but much weaker (and insignificant in most models) than the university coefficient. This indicates that universities work their effect through other pathways than just through their impact on the education level of populations. We discuss several such potential pathways in the main text.
Table 4: OLS models of protest events, controlling for local-level education from DHS

|                  | (A13) OLS | (A14) OLS | (A15) OLS | (A16) OLS | (A17) OLS |
|------------------|-----------|-----------|-----------|-----------|-----------|
| **L(universities)** | 0.150***  | 0.148***  | 0.082***  | 0.079***  | 0.065**   |
|                  | (7.51)    | (7.36)    | (4.26)    | (4.01)    | (2.95)    |
| Education        | 0.073*    | 0.077*    | 0.010     | 0.011     | 0.011     |
|                  | (2.47)    | (2.51)    | (0.30)    | (0.33)    | (0.18)    |
| LDV              | ✓         | ✓         | ✓         | ✓         | ✓         |
| Country-FE       | ✓         | ✓         | ✓         | ✓         | ✓         |
| Year-FE          | ✓         | ✓         | ✓         | ✓         | ✓         |
| Country-year interaction | ✓       | ✓         | ✓         | ✓         | ✓         |
| Cell-FE          | ✓         | ✓         | ✓         | ✓         | ✓         |
| Static covariates | ✓        | ✓         | ✓         | ✓         | ✓         |
| Time varying covariates | ✓       | ✓         | ✓         | ✓         | ✓         |
| **N**            | 6699      | 6699      | 6151      | 6151      | 6151      |
| **r2**           | 0.319     | 0.347     | 0.353     | 0.359     | 0.036     |

**Notes:** Standard errors clustered on grid-cells. Intercept omitted from table. T-values (OLS) in parentheses. The independent variable is log number of universities in a grid-cell. The education variable is taken from Dahlum and Wig (2019) and measures the average level of education in the grid-cell.
5 Instrumental variable models

As a complement to our analysis leveraging the within-unit variation in university establishment, we estimate 2SLS models endogenizing university placement by leveraging the location of colonial-era Christian mission stations in Africa. These missions often set up educational facilities that seeded the later evolution of universities (Nunn, 2014). Mission locations have recently been used as instrument for local levels of education in the post-colonial era (Acemoglu, Gallego and Robinson, 2014). Two steps are crucial for this strategy: First, we must control for plausible determinants of colonial-era missions, and try to block potential violations of the exclusion restriction using contemporary factors. To address the first issue, we control for several variables taken from the empirical literature on local-level patterns of colonial settlement/missionary activity.

The best treatment of this question is Jedwab, zu Selhausen and Moradi (2019). They discuss the factors that lead colonizers (and mission societies) to develop missions, and investigate the predictors of mission placement using detailed data for Ghana. They identify geography, political conditions, transportation, population, and economic activities as factors that lead to the development of colonial-era missions (with a focus on Africa). We use their framework and include covariates falling into these general categories. We include a range of geographic factors, density of jungle/forest, irrigation potential, ruggedness, desert, shrubland, latitude and longitude. We control for several additional factors that have been claimed to affect colonial settlement (and, by extension, missions). Acemoglu, Johnson and Robinson (2001) argue that environments with high prevalence of diseases like malaria made colonists averse to creating settlements. Such environments should decrease the likelihood of missions too. To get at this, we include an indicator of underlying disease environment, using the mapping of the Malaria Stability Index (Kiszewski, Mellinger, Spielman, Malaney, Sachs and Sachs, 2004) from Nunn and Wantchekon (2011).

Another factor that affected colonial settlement was the extent of pre-colonial state centralization (see Hariri, 2012). To capture this, we control for the presence and strength of pre-colonial kingdoms, relying on the mapped version of the Ethnographic Atlas (Murdock, 1959). We also control for observed proxies for colonial settlement, including the presence of colonial-era railroads, as well as the initial explorer routes taken by colonizers, both taken from Nunn and Wantchekon (2011). We also include distance to the nearest “natural harbor”, using the World Port Index, to account for the possible correlation between places where colonizers would make land-fall, mission stations, and contemporary political outcomes.\(^1\) The key assumption when estimating these IV models, is that mission placement is exogenous to contemporary protest when conditioning on these correlates of colonial-era mission placement.

\(^1\)The world port index is a global list of seaports compiled by the National Geospatial Intelligence Agency, and can be found here: https://msi.nga.mil/NGAPortal/MSI.portal?n_fpb=true&ageLabel=msiportal&page=2&pubCode=0015
protest through other channels than present-day universities. We include a number of covariates to account for potential channels that violate the exclusion restriction. To block the pathway from colonial-era missions to protest through economic development and urbanization, we condition on urbanization, night-light density, and local estimates of contemporary wealth. Furthermore, since colonial-era missions should affect the current religious make-up of a region, we control for the share of Christians in the grid-cell, using the World Religions Database (local estimates in GIS format) (Johnson and Grim, 2013).\footnote{The shape file is available at http://worldmap.harvard.edu/data/geonode:wrdr,province,religion,q0.} Finally, to isolate the effect of universities from the effect of general levels of education, which will likely increase as a result of colonial-era missions (Nunn, 2014), and also should affect local protest, we control for local-level average education, relying on geolocated Demographic and Health Survey data from Dahlum and Wig (2019).

While these IV models provide interesting information, they rest on the uncertain exclusion restriction discussed above. We therefore do not claim that these IV models provide very strong evidence for a causal relationship and place more credence in our difference-in-difference/fixed effects estimation strategy. But, they add some additional credence to the causal story for those who put more stock in this exclusion restriction than the assumptions underpinning the diff-in-diff models we rely on in the main paper. The fact that we find similar results in two analyses that rely on orthogonal sources of variation in the same independent variable is, in our view, a strength. We therefore do not let our results hinge on the IV analyses, but provide them here as additional information for interested readers.

The IV models are shown in table 5. Since the mission stations are time-invariant, we do not include cell-fixed effects. The first column shows the first stage where the university measure is regressed on the missions instrument and covariates. This shows that missions are a strong predictor of universities in grid-cells (T-value$= 6.73$). The F-test of instrument strength shows a strong first-stage, with F-values no lower than 42.68. The second column estimates a 2SLS model with no covariates (except country-and year fixed effects), while the next column adds covariates. This expectedly reduces the coefficient greatly, but we still find a positive coefficient for universities. Finally, we regress the outcome on our university indicator in an OLS model where all the controls (including all the additional controls for the 2SLS) are added. The university coefficient is positive also in this model and similar to the ones obtained in our benchmark analysis. The local average treatment effect (LATE) estimated in the 2SLS models is the effect of universities on protest for areas where universities were affected by the establishment of christian missions in the colonial period. We should not immediately generalize this treatment effect to “all” university cells (in our sample), since these mission-induced university locations might be different from university-locations that never had a colonial mission. In particular, these places will often have much older universities, and thus a stronger university culture, which may again be linked to a stronger
strain of political activism and radicalism than locations with newer universities. That we find a stronger coefficient for universities in the 2SLS models than in most of our benchmark models might reflect this.

Table 5: First-stage, 2SLS models where universities are instrumented for by colonial missions, and model with all 2SLS covariates

| OLS (First stage) | 2SLS | 2SLS | OLS |
|-------------------|------|------|-----|
| (A18)             | (A19)| (A20)| (A21)|
| Outcome           | Universities | Protest | Protest | Protest |
| L(missions)       | 0.058***   | (6.73)  | 0.375*** | 0.168*** | 0.096*** |
| L(universities)    |               | (11.29) | (4.68)   | (6.96)   |         |
| Benchmark controls | ✓      | ✓     | ✓      | ✓       |
| Additional IV-controls | ✓  | ✓     | ✓      | ✓       |
| LDV               | ✓      | ✓     | ✓      | ✓       |
| Country-FE        | ✓      | ✓     | ✓      | ✓       |
| Year-FE           | ✓      | ✓     | ✓      | ✓       |
| First-stage F value | 160*** | 42.68*** |         |         |
| N                 | 183,406 | 233,750 | 183,406 | 183,406 |
| R²                | 0.279  | 0.084 | 0.253  | 0.277   |

Notes: Standard errors clustered on grid-cells. Intercept omitted from table. T-values in parentheses.
6 Sensitivity to reporting-bias

One of the most plausible threats to making causal claims regarding the universities-protest relationship is the potential for reporting-bias in the protest-event data. Recent studies suggest that data based on media reports (of which the SCAD data is an instance), is subject to reporting biases that correlate with right-hand side variables (Weidmann, 2016). For example, news reporting of conflict might be more frequent in urban, wealthy and more populated areas where there is infrastructure that facilitates the operation of journalists, and events that attract media attention. This is a particular problem for universities: Universities stimulate a demand for- and supply of local media content.

To assess this potential reporting-bias, we follow the prescription of Weidmann (2016). He proposes an assessment of the degree of reporting bias, by investigating whether estimated coefficients vary by the size of protest-incidents (measured in terms of the number of reported participants). Larger events (with large numbers of participants) are more likely to be reported in the media, and will thus have a higher baseline probability of being reported, as a function of reporting bias. Hence, if estimates attenuate for high-participation events, this indicates that the initial results suffer from reporting-bias. We gauge this in figure 1. It shows that coefficients attenuate slightly when moving from the events with lowest participation to those at the levels just above. However, coefficients stabilize already when considering events with 101-1000 participants, and most of the confidence intervals overlap. This suggests that our results are quite robust to reporting-bias. This could partly result from the fact that we already control for a number of confounders that should capture reporting bias in university locations, such as urbanization, population and local wealth-estimates, meaning that the coefficient for universities net of these covariates is more or less purged of such bias.
Figure 1: Testing for reporting bias

Notes: Coefficient estimates for the L(universities) coefficients, in a model based on model 1.6 in table 1 in the main text. 90% (light blue) and 95% (dark blue) confidence intervals around the point estimate (red).
7 Different functional forms

Table 6 probes the sensitivity of the main result to the choice of estimator in the main analysis (which uses OLS). It estimates a negative binomial model, a poisson model, a linear probability model, a probit model, and a logit model (the latter three use a binary outcome). The algorithms for the GLM models would not converge for models using country-year interactions and grid-cell fixed effects, which is not unsurprising given the sensitivity and computational difficulty of these models. We do therefore not adopt this most stringent specification when testing whether our result hinges on functional form assumptions. However, the main result remains very similar to the OLS result, across these different estimator choices, and give us no reason to suspect that our findings depend on functional form assumptions.

![Table 6: Assessing sensitivity to functional form assumptions](image)

| Outcome: (protest events) | Negative binomial | Poisson | LPM | Probit | Logit |
|---------------------------|-------------------|--------|-----|--------|-------|
|                           | (A22)             | (A23)  | (A24) | (A25)  | (A26) |
| L(universities)           | 0.364***          |        | 0.006*** | 0.140*** | 0.299*** |
|                           | (14.29)           |        | (10.57) | (15.74) | (13.6) |
| Country-FE                | ✓                 |        | ✓    | ✓      | ✓     |
| Year-FE                   | ✓                 | ✓      | ✓    | ✓      | ✓     |
| N                         | 347,752           | 347,752 |     | 347,752 | 347,752 |
| Log likelihood            | -1.63e+04         | -1.63e+04 | -1.10e+04 | -110    |
| R²                        |                   |        | 0.219 |

Notes: Standard errors clustered on grid-cells. Intercept omitted from table. Z-scores (GLM models) and T-values (LPM model) in parentheses. The independent variable is log number of universities in a grid-cell.
8 Additional conditional effects: Public vs. private education

It could be argued that some universities are more likely than others to generate protest. For instance, variations in university ownership structure could influence to what extent students gain democratic preferences and/or grievances against current regimes. To explore this, we investigate whether the effect of universities depend on whether they are privately or publicly owned, drawing on information from our universities-dataset. It could be argued that publicly owned universities, at least in dictatorships, can be used to indoctrinate and will therefore reduce political opposition.

This section investigates whether there is a difference between public and private education in terms of the effect of universities on protest, and whether this differs across democracies and dictatorships. It uses the information in the university-dataset on whether a university is public or privately owned, and re-estimates the baseline model with the university terms disaggregated: Including one term for public education, and one for private education. We run these analyses for both democracies and dictatorships, using the Regimes of the World data (described in the text). Here, we classify “liberal democracies” and “illiberal democracies” as Democracies, and “closed autocracies” and “electoral autocracies” as dictatorships. Figure 7 displays these results. Both public and private universities seem to matter for protest, while private universities seem to have somewhat stronger effects overall. Their coefficients are, however, not significant in the most demanding models, with country-year interactions and cell-fixed effects. This is probably due to much more limited variation when introducing public and private universities as different terms. Comparing the models show that both the public universities have stronger relationships to protest in dictatorships than in democracies, and that public universities seem to have no effect in democracies. It should be noted that there are many more autocracies than democracies in the sample. That private universities have stronger effects in dictatorships than in democracies is in line with the notion that the regime has more control over public universities. We leave further important questions about this heterogeneity to be explored in future research.
| Regimes       | OLS All (A27) | OLS All (A28) | OLS Dictatorships (A29) | OLS Dictatorships (A30) | OLS Democracies (A31) | OLS Democracies (A32) |
|--------------|---------------|---------------|------------------------|------------------------|-----------------------|-----------------------|
| L(private universities) | 0.072*** (5.65) 0.016 (1.10) | 0.095*** (5.57) (0.05) | 0.001 (3.26) (1.43) | 0.057*** (3.26) (1.43) | 0.043 (1.48) (0.93) |
| L(public universities) | 0.039*** (7.55) 0.011 (1.57) | 0.059*** (8.49) (1.43) | 0.015 (1.48) (1.48) | 0.009 (1.48) (1.48) | 0.009 (1.48) (1.48) |
| LDV          | ✓             | ✓             | ✓                      | ✓                      | ✓                     | ✓                     |
| Country-year interaction | ✓             | ✓             | ✓                      | ✓                      | ✓                     | ✓                     |
| Cell-fixed effects | ✓             | ✓             | ✓                      | ✓                      | ✓                     | ✓                     |
| N            | 244,519       | 244,519       | 197,809                | 197,849                | 70,407                | 70,407                |
| \(R^2\)     | 0.284         | 0.051         | 0.292                  | 0.052                  | 0.279                 | 0.053                 |

Notes: Standard errors clustered on grid-cells. Intercept omitted from table. T-values (OLS) in parentheses. A country is deemed “democratic” is classified as either a “liberal democracy” or an “illiberal democracy”, based on the Regimes of the World dataset.
9 Random placebo in “ever-treated” cells

As a placebo test, we construct a random university variable in the cells that are (in reality) treated at some point. If a cell for example establishes a (new) university in 2011, it gets the score 1 on this “ever treated” variable for the whole period, and all cells that are never treated get a 0 on this variable. We then construct a random variable within treated cells with a distribution with 10% positive values (scores of 1). This simulates “university establishment” in these cells that in the real world at some point gain one or more universities. If this random variable within these cells has a significant relationship to our outcome, this would be a red flag that our results could be picking up spurious processes in these cells. Table 8 below re-estimates models 1.1-1.6 in the main table with universities replaced by “placebo universities” (in treated cells). The table shows no indication of a significant relationship between the placebo and the outcome.

Table 8: Replicating the benchmark table with “placebo universities” in cells that are (actually) treated

| OLS  | OLS  | OLS  | OLS  | OLS  | OLS  |
|------|------|------|------|------|------|
| (A33) | (A34) | (A35) | (A36) | (A37) | (A38) |
| Outcome: | L(protest events) |
| Placebo universities | 0.007 | 0.009 | 0.004 | 0.008 | 0.003 | 0.003 |
| (0.52) | (0.52) | (0.29) | (0.44) | (0.21) | (0.18) |
| Fixed effects: |
| Country X year | ✓ | ✓ | ✓ | ✓ | ✓ |
| LDV | ✓ | ✓ | ✓ | ✓ | ✓ |
| Grid-cell | ✓ | ✓ | ✓ | ✓ | ✓ |
| Covariates: |
| Time-varying | ✓ | ✓ | ✓ |
| Clustering: | Cell | Cell | Cell | Cell | Cell |
| N | 272,067 | 272,067 | 272,067 | 272,067 | 272,067 |

Notes: Standard errors clustered on grid-cells. Intercept omitted from table. T-values (OLS) in parentheses. The “placebo universities” are created by constructing a random binary variable within “ever treated” cells.
10 Additional conditional effects: Regional variation

This section re-estimates our baseline models to see if there is a difference between the Central-America and Caribbean sample, and the (significantly larger) Africa sample. This is done in the table 9 below, where we split the sample by region. The central America and Caribbean sample includes much fewer grid-cells than the African region (approx. 21000 vs. approx. 230000). The university coefficient is nevertheless somewhat stable across these two samples. However, it must be noted that the university coefficient does not attain statistical significant (and can’t be distinguished from zero) in the Central America/Caribbean sample when we include cell-fixed effects.

Table 9: Effect of universities across regions

| Outcome:          | OLS (Central america, Caribbean) | OLS Africa |
|-------------------|---------------------------------|------------|
| L(universities)   | 0.074*** (8.40)                 | -0.002 (-0.19) |
| LDV               | ✓                               | ✓          |
| Country-year interaction | ✓                           | ✓          |
| Cell-fixed effects | ✓                             | ✓          |
| N                 | 21,742                          | 222,777    |

Notes: Standard errors clustered on grid-cells. Intercept omitted from table. T-values (OLS) in parentheses.
11 Semester analysis

Figure 2: Differences in protest in academic semester vs. non-semester months, in university-grid cells and cells without universities

Notes: Coefficient from OLS model where protests are regressed on semester months in university and non-university locations. The outcome is the log number of protests in a grid-cell, errors clustered on months and grid-cells.

Figure 2 shows the semester coefficients in university and non-university cells (semester=1), from a model with cell-, year and country-month linear terms. This shows that protest activity is higher during the academic semester in university-cells but not in cells without universities. While it is hard to rule out that other things correlate with semester months in different countries (we would ideally like to have country-month fixed effects, but the data does not support this), this pattern is consistent with the claim that universities affect protest. While far from a smoking gun, we take this as an additional piece of evidence, strengthening the causal interpretation of the university-protest link.

Academic semester months are coded by using two proxies: The academic semester as listed by the department of education (some times on their home page in the given country), or by using the semester months from the biggest university (in the capital) in the country as a proxy for other universities. Using this procedure, we are able to find data for 37 cases (out of 62). These coding rules assume that there are not (major) differences within countries when it comes to the academic calendar, an assumption that is supported by the fact that the academic semester structure is often publicly announced (yearly) by departments of education and even covered in the local news. Semester months are understood as months with ongoing classes during the majority of days.

These countries for which we have data on academic semesters are: Algeria, Angola, Benin, Botswana, Burundi, Cameroon, Cote d’Ivorie Egypt, Eritrea, Ethiopia, Gambia, Ghana, Guinea, Kenya, Lesotho, Liberia, Malawi, Mauritius, Morocco, Mozambique, Namibia, Nigeria, Rwanda, Sierra leone, Somalia, South Africa, South Sudan, Cuba, Dominican Republic, El Salvador, Guatemala, Haiti, Honduras,
Jamaica, Mexico, and Nicaragua.
12 Descriptive statistics

Table 10: Summary statistics for main covariates (Baseline model and 2SLS models)

| Variable                          | Mean   | Std. Dev. | N    |
|-----------------------------------|--------|-----------|------|
| L(universities)                   | -4.474 | 0.799     | 24375|
| Protest                           | 0.027  | 0.542     | 24375|
| Grid-cell product ppp             | 0.159  | 0.72      | 230319|
| Population                        | 72206.58 | 231951.161 | 24375|
| Forest (globcover)                | 24.41  | 30.994    | 24375|
| Shrubland (Globcover)             | 7.392  | 13.79     | 24375|
| Night lights (satellite)          | 0.035  | 0.048     | 24375|
| Urbanization                      | 0.105  | 0.562     | 236417|
| Excluded groups (EPR)             | 0.361  | 0.604     | 24375|
| Capital distance                  | 6.23   | 0.788     | 24375|
| Distance to border                | 4.382  | 1.271     | 24375|
| Mountains                         | 0.141  | 0.261     | 240787|
| Temperature                       | 24.541 | 3.95      | 240648|
| Latitude                          | 18.42  | 15.771    | 24375|
| Longitude                         | 6.258  | 17.699    | 24375|
| L(mission stations)               | -6.334 | 1.956     | 24375|
| Colonial railroad                 | 0.053  | 0.223     | 244214|
| Christian population              | 0.424  | 0.494     | 244145|
| Colonial explorer route           | 0.163  | 0.37      | 244214|
| Malaria                           | 0.743  | 0.324     | 209415|
| Pre-colonial state centralization | 2.232  | 0.927     | 209415|
13 Variables and data sources

13.1 The university data

The university data is taken from the website www.4ICU.com. This website lists universities around the world, whether they are public or private, religious and other characteristics. There are around 12500 universities in the database. Figure 3 shows the increase in universities in our time-period, in the two world regions where we have SCAD data. It shows the steep increase in universities in these regions. The geocoding of these universities is done by taking their street address and city, and using the ggcode package in R. This matches addresses to google-map coordinates. It has a very strong hit-rate for our university sample, managing to geocode 12326 out of 12482 universities (close to 99%). Our main variable used in the main analysis is the (log) number of universities (or the presence of a university) in a grid-cell, as listed in the 4ICU database. We have fairly scant information on the size of these universities (number of enrolled students), and the degrees offered. Hence, we operate with a fairly wide conception of a “university”, allowing for a large number of universities.

Figure 3: Universities in our sample

13.2 The protest data

The Social Conflict in Africa Database (now expanded to more countries, and labeled the Social Conflict Analysis Database), includes information on protests, riots, strikes and other “social disturbances” in Africa since 1991. It uses the search engine LexisNexis to identify events, and the coding procedure is
described in the codebook (Salehyan and Hendrix, 2016). It covers all countries included in our dataset, but not all countries experience a protest event in a given year. To operationalize our “protest” variable, we combine the events: demonstrations and riots, both “organized” and “unorganized”. In the codebook (Salehyan and Hendrix, 2016, 4), these are described as follows:

- **Organized Demonstration.** Distinct, continuous, and largely peaceful action directed toward members of a distinct “other” group or government authorities. In this event, clear leadership or organization(s) can be identified.

- **Spontaneous Demonstration.** Distinct, continuous, and largely peaceful action directed toward members of a distinct “other” group or government authorities. In this event, clear leadership or organization cannot be identified.

- **Organized Violent Riot.** Distinct, continuous and violent action directed toward members of a distinct “other” group or government authorities. The participants intend to cause physical injury and/or property damage. In this event, clear leadership or organization(s) can be identified.

- **Spontaneous Violent Riot.** Distinct, continuous and violent action directed toward members of a distinct “other” group or government authorities. The participants intend to cause physical injury and/or property damage. In this event, clear leadership or organization(s) cannot be identified.

13.3 Covariates

Most of our baseline covariates come from the PRIO-GRID database Tollefsen, Strand and Buhaug (2012), version 2.0, and are described at [http://grid.prio.org/](http://grid.prio.org/). These include Grid-cell product ppp (G-Econ dataset), Cell population (World population database), Forest (globcover), Shrubland (Globcover), Night lights, Urbanization, Capital distance, Distance to border, Mountains, Temperature, Latitude, Longitude, and the presence of an excluded ethnic group. All sources and other details for these variables can be found.

The additional covariates that are added to the 2SLS models, are taken from different sources. Colonial explorer routes, colonial railroads, malaria and pre-colonial state centralization are taken from replication data from Nunn (2014), while mission stations are taken from replication data for Nunn and Wantchekon (2011). The democracy, corruption, GDP and population data comes from the extended V-DEM dataset (Coppedge, Gerring, Lindberg, Skaaning, Teorell, Altman, Andersson, Bernhard, Fish, Glynn et al., 2017), version 8.
### 13.4 Universities by country and change over time

Table 11: Universities in different countries

| Country       | Universities first year (1990) | Universities last year (2014) | change | % Change |
|---------------|--------------------------------|-------------------------------|--------|----------|
| Senegal       | 2                              | 2                             |        | 100      |
| Guinea        | 1                              | 1                             |        | 100      |
| Somalia       | 1                              | 1                             |        | 100      |
| Burkina Faso  | 1                              | 1                             |        | 100      |
| Malawi        | 2                              | 2                             |        | 100      |
| Mali          | 1                              | 1                             |        | 100      |
| Nauru         | 3                              | 3                             |        | 100      |
| Togo          | 4                              | 4                             |        | 100      |
| Iran          | 5                              | 5                             |        | 100      |
| Tunisia       | 6                              | 6                             |        | 100      |
| Uganda        | 6                              | 6                             |        | 100      |
| Botswana      | 1                              | 1                             |        | 100      |
| Zambia        | 2                              | 2                             |        | 100      |
| Burundi       | 3                              | 3                             |        | 100      |
| Senegal       | 2                              | 2                             |        | 100      |
| Nicaragua     | 1                              | 1                             |        | 100      |
| Ethiopia      | 8                              | 8                             |        | 100      |
| Zimbabwe      | 15                             | 15                            |        | 100      |
| Nigeria       | 32                             | 32                            |        | 100      |
| Panama        | 7                              | 7                             |        | 100      |
| Kenya         | 15                             | 15                            |        | 100      |
| Colombia      | 5                              | 5                             |        | 100      |
| Guatemala     | 5                              | 5                             |        | 100      |
| Egypt         | 14                             | 14                            |        | 100      |
| Chad          | 1                              | 1                             |        | 100      |
| Namibia       | 1                              | 1                             |        | 100      |
| Ghana         | 24                             | 24                            |        | 100      |
| Sudan         | 13                             | 13                            |        | 100      |
| Congo - Kinshasa | 8                 | 8                             |        | 100      |
| Mexico        | 235                            | 235                           |        | 100      |
| Côte d'Ivoire | 5                              | 5                             |        | 100      |
| Cameroon      | 8                              | 8                             |        | 100      |
| Eritrea       | 1                              | 1                             |        | 100      |
| Honduras      | 6                              | 6                             |        | 100      |
| Liberia       | 2                              | 2                             |        | 100      |
| Togo          | 1                              | 1                             |        | 100      |
| Israel        | 27                             | 27                            |        | 100      |
| Togo          | 3                              | 3                             |        | 100      |
| Libya         | 8                              | 8                             |        | 100      |
| Dominica      | 26                             | 26                            |        | 100      |
| Algeria       | 42                             | 42                            |        | 100      |
| Mauritania    | 2                              | 2                             |        | 100      |
| Morocco       | 19                             | 19                            |        | 100      |
| Haiti         | 13                             | 13                            |        | 100      |
| South Africa  | 36                             | 36                            |        | 100      |
| Congo - Brazzaville | 7   | 7                             |        | 100      |
| Mauritania    | 3                              | 3                             |        | 100      |
| Sierra Leone  | 3                              | 3                             |        | 100      |
| El Salvador   | 21                             | 21                            |        | 100      |
| Jamaica       | 5                              | 5                             |        | 100      |
| Cuba          | 38                             | 38                            |        | 100      |
| Lebanon       | 2                              | 2                             |        | 100      |
| Madagascar    | 6                              | 6                             |        | 100      |
| Niger         | 1                              | 1                             |        | 100      |
| South Sudan   | 10                             | 10                            |        | 100      |
| Swaziland     | 1                              | 1                             |        | 100      |
| Central African Republic | 0 | 0 | 0 | 0 |
| Guinea-Bissau | 0                              | 0                             |        | 100      |

*= Start year for all countries is in 1990, with two exceptions: Eritrea (1993) and South Sudan (2011)
References

Acemoglu, Daron, Francisco A Gallego and James A Robinson. 2014. “Institutions, human capital, and development.” Ann. Rev. Econ. 6(1):875–912.

Acemoglu, Daron, Simon Johnson and James A. Robinson. 2001. “The Colonial Origins of Comparative Development: An Empirical Investigation.” American Economic Review 91(5):1369–1401.

Campante, Filipe R and Davin Chor. 2012. “Why was the Arab world poised for revolution? Schooling, economic opportunities, and the Arab Spring.” The Journal of Economic Perspectives 26(2):167–187.

Coppede, Michael, John Gerring, Staffan II Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, Frida Andersson, Michael Bernhard, M Steven Fish, Adam Glynn et al. 2017. “V-Dem Codebook v7.”.

Dahlum, Sirianne and Tore Wig. 2019. “Educating Demonstrators: Education and Mass Protest in Africa.” Journal of Conflict Resolution 63(1):3–30.

Hariri, Jacob Gerner. 2012. “The Autocratic Legacy of Early Statehood.” American Political Science Review 106(03):471–494.

Jedwab, Remi, Felix Meier zu Selhausen and Alexander Moradi. 2019. “The Economics of Missionary Expansion: Evidence from Africa and Implications for Development.”.

Johnson, Todd M and Brian J Grim. 2013. The world’s religions in figures: An introduction to international religious demography. John Wiley & Sons.

Kiszewski, Anthony, Andrew Mellinger, Andrew Spielman, Pia Malaney, Sonia Ehrlich Sachs and Jeffrey Sachs. 2004. “A global index representing the stability of malaria transmission.” The American journal of tropical medicine and hygiene 70(5):486–498.

Murdock, George Peter. 1959. Africa: Its peoples and their culture history. New York: McGraw-Hill.

Nunn, Nathan. 2014. Gender and Missionary Influence in Colonial Africa. In Africa’s Development in Historical Perspective, ed. Emmanuel Akyeampong, Robert Bates, Nathan Nunn and James A. Robinson. New York: Cambridge U. Press pp. 489–512.

Nunn, Nathan and Leonard Wantchekon. 2011. “The Slave Trade and the Origins of Mistrust in Africa.” American Economic Review 101(7):3221–52.

Salehyan, Idean and Cullen Hendrix. 2016. “Social Conflict Analysis Database (SCAD) Version 3.2.” Codebook .

24
Tollefsen, Andreas Forø, Havard Strand and Halvard Buaug. 2012. “PRIO-GRID: A unified spatial data structure.” Journal of Peace Research 49(2):363–374.

Weidmann, Nils. 2016. “A closer look at reporting bias in conflict event data.” American Journal of Political Science 60(1):206–218.