**ONLINE APPENDIX**

This online appendix for “Revolutionary Homophobia: Explaining State Repression of Sexual Minorities” has two sections. First, an extended discussion of the data collection and coding process is outlined. Second, the various robustness checks discussed in the main text are presented.

**Section I: Data Collection and Coding for Dependent Variable**

As mentioned in the mentioned in the article, the primary dependent variable is a binary indicator for whether a state initiates anti-LGBT repression in a given year. Several different types of sources were used in the data collection process. First, and most instrumental, was the large and growing body of secondary sources research on LGBT history, rights, and community. Over the past several decades, scholars across various academic fields have taken interest in these topics across space and time. Many of their works are detailed histories and narratives that leverage in-depth, case-specific knowledge, and are based on extensive primary source information such as archival documents, elite interviews, and ethnography. These resources are necessary because it is impossible for any single scholar, or even group of scholars working in concert, to amass a similar wealth of primary source information. Additionally, the many secondary studies based on historical documentation helps alleviate the problem that interest in LGBT issues has increased over time. Although it cannot overcome the limitation that journalistic accounts are likely sparse in historical cases, relying on expert histories should decrease the number of false negatives in coding anti-LGBT repression based on contemporary interest.

 By collating this information in one place, this article was able to create a dataset on anti-LGBT repression. Using the Cuban case study in the manuscript as a running example, we can demonstrate the utility of such secondary sources. Allen Young’s (1981) book (all sources cited in the main document) is based on his personal experiences in Cuba during the period in question, interviews with elite Cuban officials such as Castro, and other primary sources on the topic, such as periodicals and other interviews. Ocasio’s (2002) article is based on in-depth research and reading into the works and lives of intellectuals who were subjects of Cuba’s repression against sexual minorities. Arguelles and Rich’s (1991) book chapter is a historical study based on primary source Cuban documents and personal interviews. Although the Cuban case is particularly well-documented, similar sources are available for most countries and regions of the world.

 Relatedly, several encyclopedia of LGBT rights have remarkable coverage that proved useful. *The Greenwood Encyclopedia of LGBT Issues Worldwide*, for instance, contains expert case histories of LGBT rights and issues for every country in the world. Resources such as these protect against the potential bias that information was sought disproportionately for countries likely to turn up positive results or to support the theory; sources, such as those discussed above regarding Cuba, are likely more abundant for cases where homophobic repression occurred, and relying on sources with global coverage of LGBT issues in every country can help ensure that overreliance on countries that are especially well-covered in the secondary literature is not biased coding decisions. Since these encyclopedias provided limited information compared with country-level secondary sources, a process was carried out whereby additional secondary sources were sought for each country, and where no such sources were found the lowest regional sphere for which information was available was sought (e.g. if there are no sources on LGBT rights in Tunisia, a source on North Africa would be consulted).

 Primary source documents were sought for cases in which no secondary source narratives had already integrated them into an LGBT rights history or case study. Academic search engines, such as Factiva and LexisNexis, were used to find periodicals on LGBT rights in individual states. Various combinations of relevant terms were searched through these databases; for example, [country name OR state leader] AND [LGBT OR LGBT rights OR homosexual OR gay OR sexual minority]. Every country was the subject of a search for evidence of LGBT human rights abuses. Other sources include country-level information from NGOs—Amnesty International, Human Rights Watch, and the International Lesbian, Gay, Bisexual, Trans, and Intersex Association (ILGA)—were consulted. Many times, anti-LGBT abuses would only come to light following post-conflict reconciliation efforts, and therefore various Truth and Reconciliation commissions were searched for evidence of homophobic repression. Although most of these documents are made available in English, some Spanish and Portuguese commissions were used when only they were available (e.g. for the Brazilian Truth and Reconciliation commission report).

 What sort of descriptions indicate that anti-LGBT repression has been initiated? As discussed in the article, the indicators used are similar to those used for the Physical Integrity Rights index; this includes evidence of anti-LGBT execution, torture, extrajudicial killings, imprisonment, disappearance, and displacement. An illustration is again provided with the Cuban case study. Multiple sources are in agreement that Cuba began arbitrarily imprisoning LGBT people in 1961 following a series of raids and without due process (Ocasio 2002; Lumsden 1996). Because they were conducted through raids, and also explicitly mentioned targeting individuals for their homosexuality, it is clear that sexual minorities were targeted as a group, rather than merely some homosexuals having been incidentally targeted during a broader campaign of repression. Such situations where LGBT people become suddenly targeted as a group are representative of the positive cases in the dataset. For this reason, a binary indicator for anti-LGBT repression seems an appropriate coding choice. In several cases, the exact year in which homophobic repression was initiated remains uncertain. Despite clear evidence that homophobic repression occurred during the Cultural Revolution, for instance, reference to a specific date is lacking. However, this is not problematic for the empirical analysis since results are not sensitive to using alternative dates. Since China is coded as a revolutionary government throughout the entire Cultural Revolution, using any year during this time-period leaves unaltered; this remains true for the other cases where onset time is ambiguous (e.g. Cambodia under the Khmer Rouge).

**Section II: Robustness Checks**

First, the results are replicated using an alternative dependent variable. Instead of repression onset, this variable is a binary indicator for whether sexual minorities at all were repressed in a state-year. Coding for the timing of homophobic repression is gathered from the same primary and secondary sources mentioned in the main text for the onset variable. Figures A1 through A4 present replications of the three figures from the paper using this variable. As mentioned, the coefficient estimates for revolutionary government are substantially larger when using this variable, and therefore lend strong support for the hypothesis that these states are more likely to repress sexual minorities.



**Figure A1:** *Replication of Figure 2 using Alternative Dependent Variable*



**Figure A2:** *Replication of Figure 3 using Alternative Dependent Variable*



**Figure A3:** *Replication of Figure 4 using Alternative Dependent Variable*



**Figure A4:** *Replication of Figure 5 Using Alternative Dependent Variable*

Next, additional control variables are added to militate against omitted variable bias. Perhaps the most important excluded variable from the main analysis is *irregular transition*. If it is irregular transitions in general that are more likely to produce homophobic repression, then it is possible that the relationship between revolutionary governments and repression is spurious. However, that is not what is found in Figure A5. Including a variable for irregular transition, taken from Colgan’s (2012) dataset on revolutionary governments, does not impact the substantive or statistical significance of revolutionary government. Moreover, the coefficient for irregular transition is not itself statistically significant. We can therefore rule out the possibility that the results are driven solely by irregular transition. Another factor to control for is regime type. While the main text contrasts personalist and non-personalist revolutionary governments, the rest of the authoritarian regimes from the Geddes’ typology are excluded. In Figure A6, the four regime types—personalist, military, party, and monarchy—are included with democracy serving as the reference category. None of the four regimes are statistically significant and revolutionary government remains unchanged. It does not appear, in other words, that other authoritarian institutions are responsible for homophobic repression.



**Figure A5:** *Controlling for Irregular Transitions*



**Figure A6:** *Controlling for Other Authoritarian Regime Types*

Since one concern mentioned in the text is that results using the onset dependent variable could be biased due to limited positive values on the dependent variable, the analysis is also replicated using two additional estimation strategies. First, Firth’s (1993) penalized likelihood logit model is used in figures A6 – A10. Second, jackknife clustered standard errors are used to iteratively estimate the models while dropping one country-level cluster each time to ensure that outliers are not driving the results.

****

**Figure A7:** *Replication of Figure 2 Using Penalized Likelihood Logit*

****

**Figure A8:** *Replication of Figure 3 Using Penalized Likelihood Logit*

****

**Figure A9:** *Replication of Figure 4 Using Penalized Likelihood Logit*

****

**Figure A10:** *Replication of Figure 5 Using Penalized Likelihood Logit*

****

**Figure A11:** *Replication of Figure 2 Using Jackknife Clustered Standard Errors*

****

**Figure A12:** *Replication of Figure 3 Using Jackknife Clustered Standard Errors*

****

**Figure A13:** *Replication of Figure 4 Using Jackknife Clustered Standard Errors*



**Figure A14:** *Replication of Figure 5 using Jackknife Clustered Standard Errors*

Since over 1,000 observations are omitted from the full models due to missing values, the results are replicated using multiple imputation to fill in this missing information. Stata’s mi command is a straightforward way to perform multiple imputation. Several covariates not estimated in the full model are used for imputation because there is no penalty in multiple imputation for using as many variables as possible to predict missing values. Ten imputations are used for these figures, although the results are the same when using a higher number (25, 50, and 100).



**Figure A15:** *Replicating Figure 2 Using Multiple Imputation*



**Figure A16:** *Replicating Figure 3 Using Multiple Imputation*



**Figure A17:** *Replicating Figure 4 Using Multiple Imputation*



**Figure A18:** *Replication Figure 5 Using Multiple Imputation*