

## Chapter 3: Quantitative Analysis

The literature review in Chapter 2 highlighted existing limitations in the quantitative literature on the impact of technology on democratization, dictatorship and activism. The qualitative literature proved more ambiguous, riddled with selection bias. “And while debate continues, there is no doubt that rigorous and data-driven analysis of this relationship will benefit scholars and policy makers alike. Indeed, the majority of earlier studies of the effects of the Internet on democracy are case studies and/or largely theoretical analyses. Few previous studies approach the issue of Internet and democracy with data-driven analysis” (Best and Wade, 2009).

The purpose of Chapter 3 is thus to carry out a large-N quantitative study that overcomes some of the shortcomings of earlier studies. More specifically, this study will use negative binomial regression analysis to determine whether the diffusion of information and communication technology (ICT) is a statistically significant predictor of anti-government protests. If the results of the analysis reveal that an increase in access to ICTs is a significant predictor of protest events in countries under repressive rule, then this may imply that technology is indeed an important explanatory variable. An authoritarian state that is unable to prevent anti-government protests as a result of greater access to ICTs faces a clear challenge to its power.

This chapter is structured as follows: the first section develops the econometric model based on the findings from the literature review. The second defines the variables and explains the case selection criteria. Section 3 provides descriptive statistics, correlation analysis and identifies the appropriate regression technique for the analysis. Section 4 lists the results of the negative binomial regression analysis. The fifth and final section of this chapter reviews the results and lays out the rationale for the comparative qualitative case studies in Chapter 4.

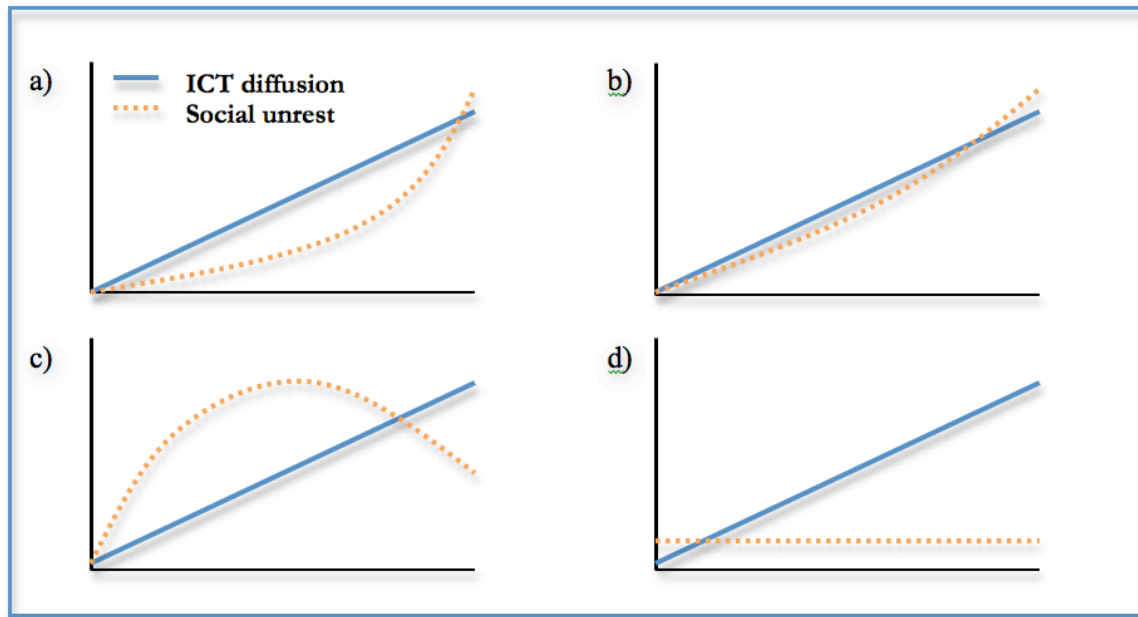
### 3.1: The Model

In Chapter 2, the meta-analysis of the quantitative literature provides decidedly mixed results, with some studies identifying statistically significant relationships between technology and democracy, and others not. The most important limitation of the data-driven studies reviewed are that: (1) the data analyzed typically goes through 2003, well before the Web 2.0 revolution; (2) the analysis tends to focus on the impact of the Internet or mobile phones, but not both; (3) the studies tend to aggregate data on democratic and authoritarian states, thus running the risk of not capturing more subtle effects regarding the impact of information and communication technologies (ICTs) on repressive regimes. Associated with this limitation is the use of aggregate measures for democracy that renders the conclusions derived from quantitative studies challenging to unpack. Indeed, “large-scale, quantitative, and cross-sectional studies must often collapse fundamentally different political systems—autocracies, democracies, emerging democracies, and crisis states—into a few categories or narrow indices” (Howard 2011).

The first limitation is perhaps the most serious. Major social media platforms are still very new even if they are rapidly growing in use and membership. Twitter was first launched in 2006, but is only now beginning to be employed by civil society groups in repressive environments. Collectively, Facebook, YouTube, and Flickr are only about five years old. The study by Howard (2011) is the only one that uses data through 2008. The second limitation is problematic because Internet access in many repressive regimes is considerably more limited than access to mobile phones. That said, the information revolution is increasingly conceived as an ecosystem of more integrated connection technologies (Schmidt and Cohen 2010). Studying the nodes of this ecosystem in isolation is thus an important constraint on the analysis. The third limitation is no less problematic with all quantitative studies in the literature combining both democratic and non-democratic countries in their analysis. This runs the risk of having important underlying effects cancel each other out. This is perhaps the only problem with Howard’s (2011) study, which is otherwise particularly robust and overcomes the other limitations identified in the literature review.

To address the first limitation, this quantitative study will draw on data from 1990 through to 2007. The study will draw on both Internet and mobile phone data to overcome the second limitation found in previous quantitative studies. In response to the third limitation, the study will focus exclusively on authoritarian states. In addition, the study will not use an aggregate measure of democracy as the dependent variable but will instead use protest events to better understand the linkages between access to ICTs (Internet and mobile phones) and such events. (As described in Section 3 below, the protest data used for this regression analysis is proprietary data and limited to 2007). The use of protest events as the dependent variable also makes the subsequent qualitative case study analysis less abstract.

The purpose of this regression analysis is thus to determine whether the diffusion of ICTs is a statistically significant predictor of protest events and if so, whether that relationship is positive or negative. The dependent variable protests is the number of protests per country-year. The ICT variables used in the model are: Internet users, mobile phone subscribers and number of telephone landlines per country-year. The control variables, identified in the literature review in Chapter 2, are percentage change in GDP, unemployment rate, the degree of autocracy per country-year, internal war and elections. Figure 1 below summarizes the four plausible outcomes of the regression analysis.



**Figure 1:** Plausible time series of ICT diffusion versus anti-government protests.

Graph (a) suggests a statistically significant relationship between the use of ICTs and the variation in incidents when controlling for other effects. This may imply that the impact of the information revolution empowers civil resistance movements at the expense of tech-savvy regimes. Graph (b) suggests a marginal increase in the frequency of incidents in response to the diffusion of ICTs, i.e., the relationship is not statistically significant. The graph depicted in (c) suggests a statistically significant—albeit inverse—relationship between ICTs and incidents. This may imply that regimes have been effective repressing anti-government protests. Graph (d) may also suggest that governments have been effective at maintaining their information blockade. The econometric model for this study is specified in Figure 2 below.

### Econometric Model

Using negative binomial regression analysis, this large-N quantitative study will test whether the diffusion of information communication technology is a statistically significant predictor of protests ( $R$ ) in 38 countries between 1990-2007 by estimating the effect of technology ( $T$ ), political ( $P$ ), economic vectors/variables on the incidence of protest:

$$(1) R = f(Tech + Pol + Econ) + \varepsilon$$

Or more formally:

$$(2) P_{c,t} = \beta_0 + \beta_1 T_{c,t} + \beta_2 P_{c,t} + \beta_3 E_{c,t} + \varepsilon_{i,t}$$

Where:

$P_{c,t}$  = the number of protest/riot events in a country  $c$ , in a given year,  $t$ .

$T_{c,t}$  = a vector of information communication technologies relevant for capturing the diffusion of the information revolution in country  $c$ , in a given year,  $t$ .

$P_{c,t}$  = a vector of political variables to control for influences on social unrest.

$E_{c,t}$  = a vector of economic variables to control for influences on social unrest.

$\varepsilon_{i,t}$  = error term, distribution assumed to be normal.

**Figure 2:** The econometric model for the quantitative analysis section of the dissertation.

Negative binomial regression analysis is used to carry out the statistical analysis. This type of regression analysis was used because the dependent variable, protest events, is highly skewed which means that regular object least squared regression analysis is not possible. The quantitative study by Miard (2009), which tested the impact of mobile phones on political protests, used the same approach for the same reason. The rationale for using this type of regression analysis is further articulated in Section 3.

### 3:2: The Data

This section reviews the datasets for the 10 variables used in this study. The dependent variable is the total number of protests per country-year. The three independent predictor variables are (1) number of Internet users, (2) number of mobile phone subscribers and (3) number of landlines. The seven control variables are (1) Autocracy, (2) Unemployment, (3) GNI, (4) GDP, (5) Population, (6) Internal war, and (7) No election. Data for 38 countries were included in the study: Algeria, Armenia, Azerbaijan, Bahrain, Belarus, Burkina Faso, Burma, China, Cote d'Ivoire, Cuba, DRC, Egypt, Gabon, Guinea, India, Iran, Iraq, Jordan, Kazakhstan, Kenya, Malaysia, Morocco, Pakistan, Philippines, Russia, Saudi Arabia, Singapore, Sudan, Syria, Tajikistan, Thailand, Tunisia, Turkey, Ukraine, United Arab Emirates, Uzbekistan, Venezuela and Zimbabwe. The country selection criteria is explained later in this section.

#### ***Dependent Variable Data***

While many previous empirical studies have been limited to examining “protest potential,” this represents an inadequate indicator of actual protest behavior (Norris 2006). Since surveys are usually more appropriate for capturing attitudes rather than actual behavior, this study must focus on datasets that document past incidents of protests instead of surveys. There are a number of datasets available that focus on (or include) protest event-data. The most commonly used include the *World Handbook of Political and Social Indicators* (Taylor and Jodice 1983), the *Prodat Project* (Kriesi *et al.* 1995), the *Cross-National Time Series* (Banks 2001) and the *Protest and Coercion Data* (multiple contributors, 2008). Table 2 below compares these datasets.

Dataset	Time Period	Countries
1948-1992	156	All members of the United Nations in 1975
Prodat Project	1975-1989	France, Germany, the Netherlands, Switzerland
Cross-National Series	1815-1999	167+
Protest/Coercion Data	Multiple	27 European states (1980-1995), Burma (1988), S. Korea (1990-1991), Colombia (1988-1997), El Salvador (1979-1991), Peru (1980-1995).

**Table 2:** Available datasets that document past instances of protest events.

As Nam (2006) rightly notes, a serious debate on the quality of protest datasets has been lacking in the literature. The datasets listed above have practical and theoretical limitations. First of all, the categories of events coded in these traditional datasets “are often too general, arbitrary and inaccurate” (Nam 2006). The second problem is that the protest data is only available in annually aggregated numbers, thus “eliminating the possibility of quarterly or monthly, let alone weekly or daily analyses (Nam 2006).

Indeed, “the annual measures omit the substantive ebb and flow of daily and weekly interactions between citizens and authorities (Dahlerus 2006). Furthermore, for the purpose of this study, “if communication and information technologies are a part of political protests, then we must get more detailed information on the timing of the protests to see if they occur in clusters, which we would expect to happen” (Eyck 2001). Third, and perhaps most important, these traditional datasets have typically relied on a limited number of sources. When protests are sampled from just one or two sources, especially non-local sources, the data is very likely to be incomplete and biased (Dahlerus 2006; Nam 2006).

These traditional datasets generally draw on the “hand-coding” of event-data based on national and international news reports. Strictly speaking, then, event data are no more

than simple observations of media reporting patterns as opposed to a valid sample of protest and repressive events (Danzger 1978; Franzosi 1987; Mueller 1999; Hocke 1998). To complicate matters, “news-sources contain internal biases that influence how protest and repressive events are reported, who is involved, what happened, and why” (Dahlerus 2006; see also Davenport 2001). To be sure, “coverage in Western European and international media sources are more likely to pay attention to larger protests that involve property destruction or other more visible qualities” since these obviously make for good headlines which increases profits (Dahlerus 2006; see also Clutterbuck 1980; Franzosi 1987; Herman and Chomsky 1988; Turow 1997; Wolsfeld 1997).

Furthermore, international media coverage of “protest is skewed toward political contention in established democracies, while protest and in particular repression in authoritarian states receive less attention outside of large scale dramatic events” (Dahlerus 2006). Local media sources also come with a host of issues. Press freedom is endogenous to democratization and consolidation, which means that in the context of repressive or closed regimes, getting accurate reports on anti-government protests from national sources may be difficult to come by. Indeed, national media outlets in nondemocratic countries are almost always state-run and thus virtually guaranteed to omit any indication of resistance against the ruling regime.

While using newspapers for events-based information collection is far from ideal, one must recognize that no other available source could realistically fulfill the stringent requirement of objectivity and completeness. To be sure, it is very well known biases exist in official state reports especially in authoritarian systems that seek to restrict the outflow of information (Dahlerus 2006). While police reports will underreport the number of people participating in anti-government protests, social movements tend to inflate crowd magnitude (Beissinger 1998; McPhail *et al.* 1998).

Moreover, despite the implicit biases that exist in news-generated events data, “scholars who rely on newspapers for information about contentious [sic] suggest that



newspapers are overall accurate in the information they do report” (Dahlerus 2006; see also Franzosi 1987, 7; Wolsfeld 1997; Beissinger 1998; Olzak and Olivier 1998; Ekiert and Kubik 1999). To this end, while all events are not captured by the news media, one can still “rely on an overall accuracy of information contained in reports of the essential ‘who,’ ‘what,’ ‘where,’ and ‘when’ issues of contentious events” (Dahlerus 2006). In short, newspapers may provide biased coverage, but they seldom print false information (Franzosi 1987; Dahlerus).

Since this study focuses on the likely impact of ICTs on the frequency of protest events, the ideal dataset would need to cover the time period prior to the information revolution through beyond 2003, which is where most studies stop. In the context of nondemocratic (generally developing) countries, the ICT data (described in more detail below) suggests that the information revolution begins to take around the year 2000. None of the traditional datasets listed above include protest events-data beyond 1999.

Since there is rich evidence that large-scale protests are valuable in cultivating deliberative and sustainable democracies (Ekiert and Kubik 1999; Ackerman and DuVall 2000; Inglehart and Catterberg 2002; Norris 2006; Stephan and Chenoweth 2008), this study is not particularly interested in small-scale protests that may not be reported by mainstream media. Moreover, sizeable events will require more extensive use of ICTs to mobilize, organize and coordinate. Using this reasoning, then, the constraints of international news media as a source for protest event-data are not particularly problematic. On the contrary, relying on international news serves to filter out small-scale, politically insignificant protest events.

This study therefore draws on the conflict event-dataset developed by the private company Virtual Research Associates, Inc. (VRA). VRA uses a patented natural language parsing algorithm to parse Reuters newswires in near real-time for over 60 countries. The algorithm codes events into a 157-indicator framework called the Integrated Data for Events Analysis (IDEA) framework. For each event, the parser codes

the following parameters, “who (source), did what (event), to whom (target), where (country) and when (day)?” (Bond *et al.* 1997; Bond *et al.* 2001; Bond *et al.* 2003; King and Lowe 2003). To this end, the dataset specifies whether the state or society is the initiator of the event.

Unlike traditional datasets on protest events, the VRA data uses automated coding. An important question then is how hand- and automated coding compares in terms of reliability. Harvard University Professor Gary King and Dr. Will Lowe carried out an independent and comprehensive evaluation of the VRA dataset in 2003. They conclude their in-depth study with some optimism:

In our view, the results in this article are sufficient to warrant a serious reconsideration of the apparent bias against using events data, and especially automatically created events data, in the study of international relations. If events data are to be used at all, there would now seem to be little contest between the machine and human coding methods. With one exception, performance is virtually identical, and that exception (the higher propensity of the machine to find “events” when none exist in news reports) is strongly counterbalanced by both the fact that these false events are not correlated with the degree of conflict of the event category, and by the overwhelming strength of the machine: the ability to code huge numbers of events extremely quickly and inexpensively.

Although the machine performed approximately equally to our trained human coders in this study, the machine would be far better over the long run. Hiring people of the quality we were able to recruit to code many more events than we asked of them is probably infeasible, and doing so for the many years it would take to do this right would undoubtedly reduce performance to levels significantly below that of the machine. Longer-term coding by human coders would result in lower performance, either because

we would have to resort to using less-qualified coders or because their attention to the extremely tedious and boring task would wane over time.

Unlike the more common datasets describe above, the VRA parser picks up distinct events as they occur throughout a given day or week. For example, as a demonstration unfolds, the natural language parser will not code one event only but rather a series of events just as long as the underlying news reports are different. Only if two reports are exactly alike does the parser only code the event as one event. To this end, both the frequency and duration of protests can be extracted from the VRA database. Furthermore, because the data is generated using newswires, any period of specific interest in the time series data can be interpreted by reading the associated news reports themselves. These can easily be found using Lexus Nexus and Factiva, as well as Google News archives. In sum, the VRA data is not aggregated by year. Furthermore, the dataset includes the following six directly relevant event types:

- **Protest altruism:** protest demonstrations that place the source (protestor) at risk for the sake of unity with the target.
- **Protest defacement:** performance protests, graffiti and desecration of symbols.
- **Protest obstruction:** sit-ins and other non-military occupation protests.
- **Protest procession:** picketing and parading protests.
- **Protest demonstration:** all protest demonstrations not otherwise specified.
- **Riot:** civil or political unrest explicitly characterized as riots.

These six event types are aggregated in the regression analysis to serve as a proxy for protest incidents. Finally, the VRA dataset covers over 60 countries between 1990 and

2007, i.e., from pre-information revolution and through to the current Web 2.0 and mobile technologies revolution. Furthermore, the dataset is not limited to Western democracies, which is another distinct advantage over traditional datasets.

### ***Independent Variable Data***

The data on ICT diffusion is drawn from the 2008 edition of The World Telecommunication ICT Indicators Database produced by the International Telecommunications Union (ITU). This database contains annual time series data from 1975 to 2007 for around 100 communications statistics including the number of mobile phone users and Internet users. Data on Internet users, mobile phone subscribers and total number of telephone lines per country year is taken from this dataset for the regression analysis.

All the quantitative studies reviewed in Chapter 2 make use of this dataset. As Howard (2011) rightly notes, however, this is problematic given “how many scholars rely on a few data sources, chiefly the International Telecommunications Union, the World Bank, and the World Resources Institute. Indeed, these organizations often just duplicate each other's poor quality data. Many researchers rely heavily on this data for their comparative or single-country case studies, rather than collecting original observations or combining data in interesting ways. The same data tables appear over and over again.” Unfortunately, collecting original data for 30+ authoritarian states is beyond the scope of this dissertation research. The regression analysis is therefore limited to using the ITU data.

### ***Control Variable Data***

The statistical analysis includes seven control variables. Based on a thorough literature review, Autocracy, Unemployment, Gross National Income (GNI), Gross Domestic Product (GDP), Population, Internal war and No election were identified as statistically or potentially significant predictors of protest events. Increasing repression can backfire

and trigger protests. Increasing unemployment may also serve to foment anti-government sentiments. “In regimes where political parties cannot work to improve the quality of life for their membership, such disparities can be the motivation for social rebellion. Poverty is most likely to motivate people to rebel when there is also a perception of relative deprivation, with an enormous gulf between rich and poor that seems insurmountable to the poor” (Howard 2011). GNI and GDP also need to be held constant since they may be correlated with increasing access to ICTs. In addition, “education and economic wealth—especially inequities in the distribution of wealth—can play significant causal roles in social unrest” (Howard 2011). Large country populations may also increase the probability of unrest as regimes are not able to repress effectively beyond their own capitals. “Large countries often also have a ‘youth bulge,’ with significant numbers of people under the age of 15 who must be accommodated in schools and the labor force. If they are not accommodated, cohorts of disaffected youth can cause trouble for a regime. This variable was tested, however, and consistently dropped out of the reduced causal sets as a factor that neither contributed to nor detracted from membership in the set of countries that democratized” (Howard 2011). Finally, unrest may be associated with internal wars and presidential or parliamentary elections. The regression analysis will thus employ these variables as control variables.

The autocracy measure used in the subsequent regression analysis is taken from the Polity2 indicator in Polity IV dataset. The Polity data series is a widely used data series in political science research. The latest version, Polity IV contains coded annual information on regime authority characteristics and transitions for all independent states (with greater than 500,000 total population) in the global state system and covers the years 1800-2006. A detailed analysis by Munck and Verkuilen (2002) highlights some of the strengths and weaknesses of existing datasets on democracy and authoritarianism. While the data by Freedom House is widely used, the data conflates multiple problems of measurement and uses an inappropriate aggregation procedure. Worse still, Freedom House has refused to make their disaggregated data public. While certainly not perfect, the Polity IV data is still considered of higher quality than Freedom House (Munck and Verkuilen 2002).

The Polity2 indicator is computed by subtracting PolityIV’s Autocracy score from the Democracy score, with the resulting scale ranging from +10 (strongly democratic) to – 10 (strongly autocratic). This indicator provides a convenient avenue for examining general regime effects in analyses but researchers should note that the middle of the implied [Polity2] “spectrum” is somewhat muddled in terms of the original theory, masking various combinations of [Democracy] and [Autocracy] scores with the same [Polity2] score” (Gurr and Jagers 2010).

Polity IV’s Autocracy score is defined “operationally in terms of the presence of a distinctive set of political characteristics” (Gurr and Jagers 2010). The autocracy score is an additive 11-point scale and is derived from these other Polity IV indicators: competitiveness of political participation, the regulation of participation, the openness and competitiveness of executive recruitment and constraints on the chief executive. The additive weights used for coding purposes are listed in Table 3 below.

<i>Competitiveness of Executive Recruitment (XRCOMP):</i>  (1) Selection +2  <i>Openness of Executive Recruitment (XROPEN):</i> only if XRCOMP is coded Selection (1)  (1) Closed +1 (2) Dual/designation +1  <i>Constraints on Chief Executive (XCONST):</i>  (1) Unlimited authority +3 (2) Intermediate category +2 (3) Slight to moderate limitations +1  <i>Regulation of participation (PARREG):</i>  (4) Restricted +2 (5) Sectarian +1  <i>Competitiveness of Participation (PARCOMP):</i>  (1) Repressed +2 (2) Suppressed +1
--

**Table 3:** Authority coding and weight scale for Autocracy

Polity IV's Democracy score is also an additive 11-point scale. The operational indicator of this score is derived from codings of competitiveness of political participation, the openness and competitiveness of executive recruitment and constraints on the chief executive. The weight scale in Table 4 below is used to code this Democracy score.

<i>Competitiveness of Executive Recruitment (XRCOMP):</i>  (1) Election +2 (2) Transitional +1  <i>Openness of Executive Recruitment (XROPEN):</i> only if XRCOMP is Election (3) or Transitional (2)  (1) Dual/election +1 (2) Election +1  <i>Constraint on Chief Executive (XCONST):</i>  (1) Executive parity or subordination +4 (2) Intermediate category +3 (3) Substantial limitations +2 (4) Intermediate category +1  <i>Competitiveness of Political Participation (PARCOMP):</i>  (1) Competitive +3 (2) Transitional +2 (3) Factional +1
---

**Table 4:** Authority coding and weight scale for Democracy

Unemployment, Gross National Income (GNI) and Gross Domestic Product (GDP) are standard variables that are available from the International Labor Organization’s (ILO) labor statistics and the World Bank’s data development indicators. These variables and datasets are regularly used in numerous quantitative political science studies.

The two variables on “Internal war” and “No election” are dichotomous variables. Internal war is coded as 1 if an internal war takes place in a given country-year. No election is coded 1 when there are no major elections in a given country year. The data for these variables is drawn from the Polity IV dataset and Political Risk Services (PRS).



### **Case Selection**

A total of 38 countries were used in the subsequent regression analysis were primarily selected on the basis of the Polity IV and protest datasets. Any country that met the following three criteria were included in the study:

1. *A Polity2 score between -5 and -10 for at least one year between 1990 and 2007.*

**Or**

2. *Polity2 scores above 0 for all years between 1990 and 2007 but known to be a repressive regime and the frequently the subject of case study analyses in the literature.*

**And**

3. *Protest data and control variable data available.*

The above selection criteria were used for several reasons. First, the purpose of this study is to assess the impact of the information revolution on strong authoritarian states, not weak states or democratic ones. Second, a small number of countries in the Polity IV dataset have Polity2 scores greater than zero but are nevertheless clearly repressive, such as Russia. In addition, these countries figure as principle case studies in the digital activism literature. Third, the dependent variable for this study is the number of protests per country-year. Given that this quantitative study draws on the most comprehensive protest data available, if the majority of that data is missing for a particular country between 1990 and 2007, it simply cannot be included in the regression analysis. In addition, time series data on unemployment in repressive regimes stretching back to 1990 is often not available.

A total of 38 countries were included in this study based on the criteria just explained. These are: Algeria, Armenia, Azerbaijan, Bahrain, Belarus, Burkina Faso, Burma, China,

Cote d’Ivoire, Cuba, DRC, Egypt, Gabon, Guinea, India, Iran, Iraq, Jordan, Kazakhstan, Kenya, Malaysia, Morocco, Pakistan, Philippines, Russia, Saudi Arabia, Singapore, Sudan, Syria, Tajikistan, Thailand, Tunisia, Turkey, Ukraine, United Arab Emirates, Uzbekistan, Venezuela and Zimbabwe. Of these, 8 were included based on criterion number 2: Malaysia, Philippines, Russia, Singapore, Thailand, Turkey, Ukraine and Venezuela. The following 14 countries from the PolityIV dataset had to be excluded due to criterion number 3: Bhutan, Equatorial Guinea, Gambia, Kuwait, Laos, Libya, Mauritania, North Korea, Oman, Qatar, Swaziland, Togo, Turkmenistan and Vietnam.

### 3.3: Descriptive Statistics, Correlation Analysis and Regression Model

SPSS v15.0 and STATA v10.0 were used for all descriptive and inferential analyses.<sup>1</sup>

Descriptive data with measures of central tendency were defined for the variables of the dataset. **Table 1** (see Appendix) shows descriptive data as relates to the overall sample ( $N = 684$ ) and for each of the 38 country sub-groups ( $n = 18$ ).

For all 38 countries over the 18-year period ( $N = 684$ ):

- **Protests** ranged from 0 -116 per year, with a mean of 10.46 per year ( $SD = 15.00$ ).
- **Internet users** ( $M = 4.26$ ,  $SD = 9.63$ ) ranged from 0 to 62%.
- **Mobile subscribers per 100 population** ( $M = 12.72$ ,  $SD = 25.12$ ) ranged from 0 to 173, indication that for some countries the number of mobile phones exceeds one per person.
- **Fixed phone lines per 100 persons** ( $M = 9.66$ ,  $SD = 10.18$ ) ranged in number from 0.12 – 48.44.
- **Autocracy scale values overall** ( $M = -1.95$ ,  $SD = 6.02$ ) ranged from scores of -10 to scores of 10.
- **The percentage change in unemployment rate** for the sample ( $M = .09$ ,  $SD = 0.84$ ) ranged from -1% to 14.08%.

---

<sup>1</sup> Acknowledgements: Ben Mazzotta and Elaine Bellucci for their guidance on the regression analysis; Ginn Library Research Staff and Christine Martin for their assistance in data development.

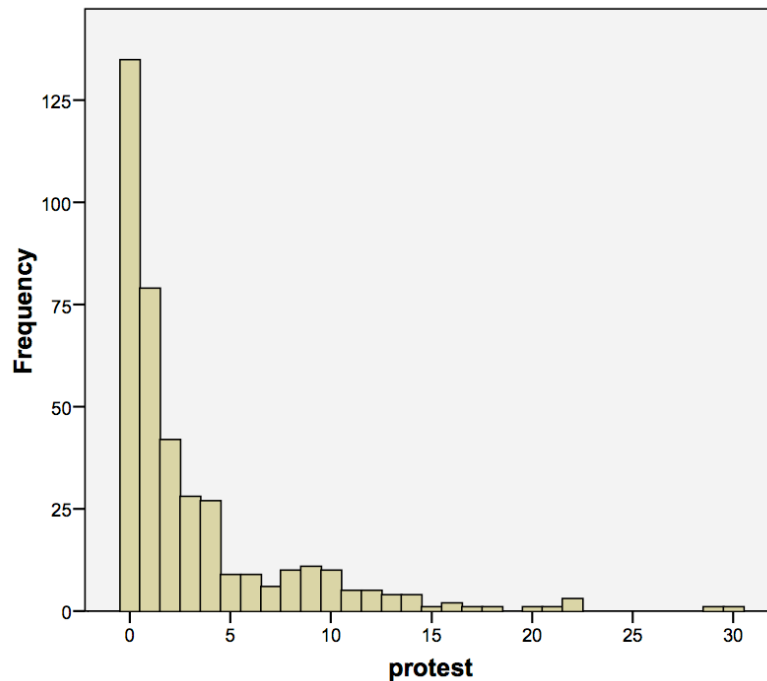
- **Gross domestic product values overall** (GDP,  $M = 3.87$ ,  $SD = 7.02$ ) ranged from -41.3 to 52.3.
- **Gross national income per capita** (GNI,  $M = 3173.98$ ,  $SD = 5440.55$ ) ranged from 128 to 41,031.
- **Population figures** ( $M = 942,778.44$ ,  $SD = 3208592$ ) for the sample ranged from 493 to 19,268,303.

The number of instances of internal wars for the 38 countries according to each of the 18 years included in the study was 157 (23% of all 684 records included in the study). The number of instances of no elections for the 38 countries according to each of the 18 years included in the study was 529 (77.3% of all 684 records).

Table 2 (see Appendix) lists the 5 highest and 5 lowest scoring countries according to mean values on each of the nine continuous variables used in this study.

### ***Assumptions for Inferential Analysis***

The mean values were lower than the standard deviations for all nine continuous variables studied as relates to the entire sample. Additionally, the median values for all one variable (% GDP change) were smaller than the mean values. This information indicates that the variables were not normally distributed. Of interest in this study was the variable *Protests*, which was used as a dependent variable in regression analysis. A histogram of the *Protests* variable indicated a long positive skew, with most values close to the lower end of the scale, but with a number of values dispersed at the high end of the range.



The data was count data, and the distribution was Poisson. Transformations of the Protests variable were performed with logarithms, square roots, and inverse-logarithms, but the distribution did not transform to normal. It was determined that ordinary least squares (OLS) type of regression would not be used for the dataset. Also of concern was the mean to variance numbers for the dependent variable of Protests. The variance ( $s^2 = 225.01$ ) was almost 22 times larger than the mean ( $M = 10.46$ ) indicating over-dispersion of the data points for the Protest variable. It was determined that a negative binomial regression would be used to evaluate the data, as the negative binomial distribution allows for a variance with a larger value than the mean. STATA command for a negative binomial regression model with panel data “`xtnbreg`” was used to fit a negative binomial model to the dataset, with the autoregressive (AR1) correlation option chosen due to the repeated measures nature of the data.

Correlation analysis was also performed on the dataset. Normality was not assumed for any of the variables in the data. Also, linearity was not assumed. Spearman’s rho correlations are based on rank order of data values and not on a normal distribution,

and they can be used with count and ordinal data. Therefore this nonparametric alternative was chosen for use in this study in order to maintain a conservative approach to assessing correlation significance.

### **Correlation Analysis**

Spearman’s rho correlations were performed on the study variables of (a) number of protests (Protests), (b) Autocracy, (c) percentage change in unemployment (Unemployment), (d) GNI, (e) GDP, (f) Population, (g) Internal war, (h) No election, (i) Internet use, (j) phones per 100 population (Phones), and (k) mobile phones per 100 population (Mobile phones). **Table 3** below lists the bi-variate correlations between the variables.

- **The variable of Protests** was significantly directly correlated with Autocracy ( $\rho = .282, p < .01$ ), indicating that when the number of protests increases or decreases, the value of Autocracy moves similarly. The Protests variable was also significantly directly correlated with GNI ( $\rho = .104, p < .01$ ), Population ( $\rho = .361, p < .01$ ), Internal war ( $\rho = .303, p < .01$ ), Internet use ( $\rho = .098, p < .05$ ), and Phones ( $\rho = .127, p < .01$ ).
- **Autocracy** was significantly negatively correlated with Unemployment ( $\rho = -.084, p < .05$ ), and No election ( $\rho = -.144, p < .01$ ). Indicating that higher numbers of values on the Autocracy variable are associated with lower values of the Unemployment and no election variables, and vice versa. Autocracy was significantly directly correlated with the variables of Population ( $\rho = .240, p < .01$ ), Internal war ( $\rho = .133, p < .01$ ), Internet use ( $\rho = .189, p < .01$ ), Phones ( $\rho = .137, p < .01$ ), and Mobile phones ( $\rho = .208, p < .01$ ).

- **The Unemployment variable** indicated significant negative correlations with GNI ( $\rho = -.112, p < .01$ ), GDP ( $\rho = -.303, p < .01$ ), Internet use ( $\rho = -.185, p < .01$ ), and Mobile phones ( $\rho = -.282, p < .01$ ).
- **GNI** was significantly directly correlated with GDP ( $\rho = .122, p < .01$ ), Internet use ( $\rho = .434, p < .01$ ), Phone use ( $\rho = .686, p < .01$ ), and Mobile phones ( $\rho = .541, p < .01$ ). A significant negative correlation was found between GNI and Population ( $\rho = -.216, p < .01$ ) and Internal war ( $\rho = -.107, p < .01$ ).
- **GDP** was significantly directly correlated with Internet use ( $\rho = .263, p < .01$ ), Phones ( $\rho = .161, p < .01$ ), and Mobile phones ( $\rho = .321, p < .01$ ).
- **The variable for Population** had a significant direct relationship with Internal war ( $\rho = .330, p < .01$ ) and a significant negative relationship with Phones ( $\rho = -.365, p < .01$ ).
- **Internal war** indicated a significant negative relationship with regular Phone use ( $\rho = -.168, p < .01$ ).

Strong direct significant relationships were found between the variables of Internet use and Mobile phones ( $\rho = .883, p < .01$ ), and Phones and Mobile phones ( $\rho = .503, p < .01$ ).

**Table 3**

*Spearman’s Rho Correlation Coefficients of Inferential Study Variables (N = 684)*

Variable	1	2	3	4	5	6	7	8	9	10
1. Protest										
2. Autocracy	.282**									
3. Unemployment rate (% change)	.035	-.084*								
4. GNI	.104**	0.63	-.112**							
5. GDP	.017	.001	-.303**	.122**						
6. Population	.361**	.240**	-.017	-.216**	-.027					
7. Internal War	.303**	.133**	-.009	-.107**	.003	.330**				
8. No elections	-.009	-.144**	-.001	-.007	.038	-.036	-.053			
9. Internet use	.098*	.189**	-.185**	.434**	.263**	-.059	-.054	-.057		
10. Phone use	.127**	.137**	-.027	.686**	.161**	-.365**	-.168**	-.035	.498**	
11. Mobile phone use	.060	.208**	-.191**	.541**	.321**	-.028	-.065	-.051	.883**	.503**

*Note.* GNI = Gross National Income; GDP = Gross Domestic Product

\*  $p < .05$

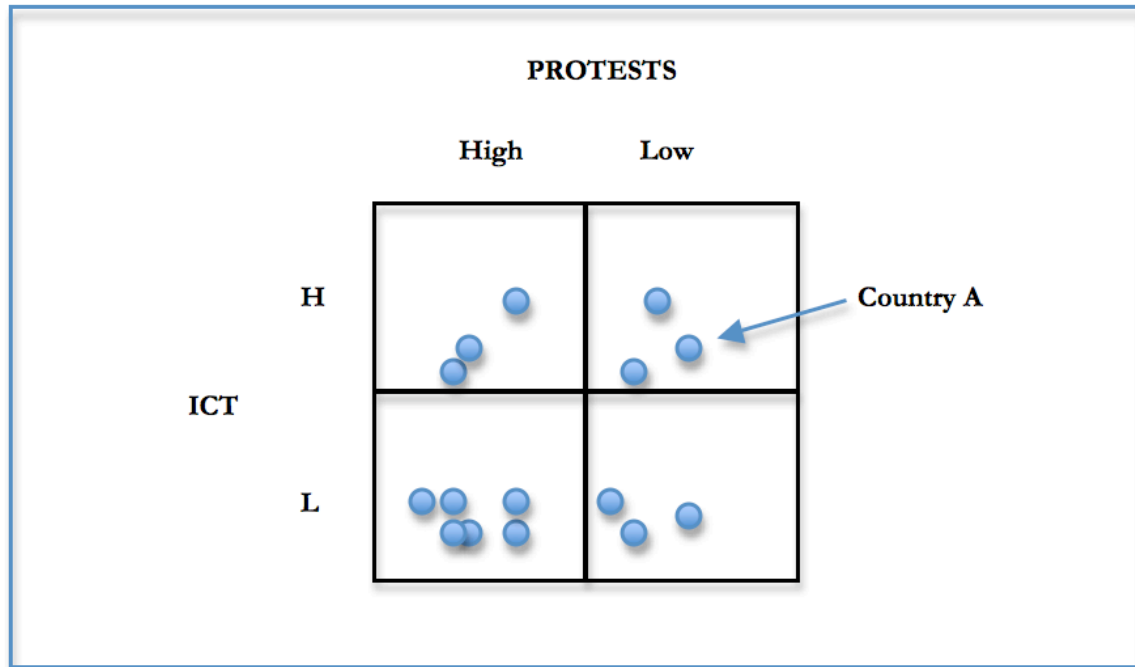
\*\*  $p < .01$

### 3.4: Negative Binomial Regression Analysis

A negative binomial regression analysis was performed on the dependent variable of Protests with three independent predictor variables of: (1) Internet use, (2) Mobile phones, and (3) Phones, and seven control variables of (1) Autocracy, (2) Unemployment, (3) GNI, (4) GDP, (5) Population, (6) Internal war, and (7) No election. STATA v10.0 was used for inferential analysis with the code “xtnbreg” for panel data with autoregressive correlation (AR1). The negative binomial regression was chosen over a Poisson model because the data for the Protest variable were over-dispersed ( $M = 10.46$ ,  $Variance = 225.01$ ) and the variable was also zero-inflated, containing 116 (16.5%) records with the number of protests counted as zero.

**Five separate regressions** were performed using STATA command “xtnbreg” for a negative binomial regression with panel data. The first regression included all 38 countries ( $n = 675$ ). The dataset was then divided into two groups according to number of protests for countries groups as (a) low protest, and (b) high protest. The groups were formed using a median split on the protest variable, with records containing protests of less than six being considered low (21 countries,  $n = 370$ ), and protests as greater than or equal to six as high protest (17 countries,  $n = 305$ ). The split was performed on the median in an attempt to achieve a number as equal as possible between the two country groups while still making use of a parameter of central tendency. The mean was not chosen for the split because the data distribution was not normal, and the mean was inflated due to the over-dispersion of the data. After two regressions were performed on the protest sub-groups, the aggregated data was then divided into two groups using a median split on the mobile phone variable. The two groupings were categorized as (a) mobile use low (25 countries,  $n = 441$ ), and (b) mobile use high (13 countries,  $n = 234$ ). Again the median of the mobile phone variable was used to define the split, with countries at less than or equal to 1 percent mobile use assigned to the low group, and countries with greater than 1 percent mobile phone use assigned to the high group.





The statistical hypotheses for all of the five regression analyses were as follows:

**Null Hypothesis:** None of the predictors of (a) Internet use, (b) mobile phone use, (c) phone use, are statistically significant predictors of the number or protests, when controlling for additional independent variables of (1) Autocracy, (2) Unemployment, (3) GNI, (4) GDP, (5) Population, (6) Internal war, and (7) No election.

**Alternative Hypothesis:** At least one of the predictors of (a) Internet use, (b) mobile phone use, (c) phone use, are statistically significant predictors of the number or protests, when controlling for additional independent variables of (1) Autocracy, (2) Unemployment, (3) GNI, (4) GDP, (5) Population, (6) Internal war, and (7) No election.

**Regression 1, All Data (n = 675)**

The model for all data was significant (Wald  $\chi^2 = 27.57$ ,  $p = .002$ ), indicating that the predictor model using the dataset was improved over a model in which all predictors were set to zero. Mobile phone use was a significant predictor ( $z = -4.12$ ,  $p < .0005$ ). Autocracy was a significant covariate ( $z = 2.25$ ,  $p = .024$ ). Incidence rate ratios (IRR) were computed for the two significant variables. The IRR for mobile phone use (.990) indicates that if all other predictor variables are held constant, then a one point increase on the percentage of a country’s mobile phone use would decrease the number of protests by a factor of .990. The IRR for autocracy (1.01) indicates that if all other predictor variables are held constant, then a one point increase in the autocracy score would increase the number of protests by a factor of 1.01.

*Table 4: Negative Binomial Regression Coefficients for Predictors on Criterion of Number of Protests for Entire Sample (n = 675)*

Variable	<i>B</i>	<i>SE B</i>	<i>z</i>	<i>p</i>
Internet use	-0.008	0.005	1.42	.155
Phone use	0.010	0.007	1.54	.125
Mobile phone use	-0.010	0.002	-4.12	<.0005
Autocracy	0.019	0.008	2.25	.024
Unemployment rate (% change)	0.005	0.037	0.14	.890
GNI	<0.0005	<0.0005	1.92	.054
GDP	-0.0003	0.005	-0.05	.958
Population	<0.0005	<0.0005	0.53	.599
Internal war	0.079	<0.0005	0.83	.407
No elections	-0.047	0.104	-0.65	.517
Wald $\chi^2 = 27.57$ $p = .0021$				

*Note.* GNI = Gross National Income; GDP = Gross Domestic Product.

***Regression 2, Low Protest Group (n = 370)***

The model for the low protest group was not significant (Wald  $\chi^2 = 15.86$ ,  $p = .10$ ), indicating that the predictor model using the dataset was *not* an improvement over a model in which all predictors were set to zero. Since the overall model was not statistically significant, the model was not investigated further for significant predictors variables.

**Regression 3, High Protest Group (n = 305)**

The model for the high protest group was significant (Wald  $\chi^2 = 20.00$ ,  $p = .03$ ), indicating that the predictor model using the dataset was improved over a model in which all predictors were set to zero. Mobile phone use indicated significance ( $z = -2.99$ ,  $p = .003$ ), with incidence rate ratios (IRR) of .992 indicating that if all other predictor variables are held constant, then a one point increase on the percentage of a country's mobile phone use would decrease the number of protests by a factor of .992.

*Table 5: Negative Binomial Regression Coefficients for Predictors on Criterion of Number of Protests for High Protest Sub-Group (n = 305)*

Variable	<i>B</i>	<i>SE B</i>	<i>z</i>	<i>p</i>
Internet use	0.004	0.006	0.73	.463
Phone use	0.004	0.008	0.57	.566
Mobile phone use	-0.008	0.003	-2.99	.003
Autocracy	0.013	0.009	1.47	.142
Unemployment rate (% change)	0.050	0.040	1.29	.198
GNI	<0.0005	<0.0005	0.02	.987
GDP	-0.006	0.006	0.93	.355
Population	<-0.0005	<0.0005	0.17	.862
Internal war	0.125	0.113	1.26	.209
No elections	-0.148	0.072	-1.78	.075
Wald $\chi^2 = 20.00$ $p = .029$				

*Note.* GNI = Gross National Income; GDP = Gross Domestic Product.

**Regression 4, Low Mobile Use Group (n = 441)**

The model for the low mobile use group was significant (Wald  $\chi^2 = 32.28$ ,  $p = .0004$ ), indicating that the predictor model using the dataset was improved over a model in which all predictors were set to zero. Significant predictors included phone use ( $z = 2.35$ ,  $p = .019$ ), mobile phone use ( $z = -3.05$ ,  $p = .002$ ), and no elections ( $z = -2.01$ ,  $p = .045$ ). Incidence rate ratios (IRR) were computed for the three significant predictors. The IRR for phone use (1.03) indicates that if all other predictor variables are held constant, then a one point increase on the phone use rate score would increase the number of protests by a factor of 1.03. The IRR for mobile phone use (.988) indicates that, given the other predictors are held constant; a percentage point increase in the mobile phone use rate would decrease the number of protests by a factor of .988. The IRR for no elections (.841) indicates that, given the other predictors are held constant; a year in which a country did not hold elections would result in a decrease the number of protests by a factor of .841.

*Table 6: Negative Binomial Regression Coefficients for Predictors on Criterion of Number of Protests for Low Mobile Use Sub-Group (n = 441)*

Variable	<i>B</i>	<i>SE B</i>	<i>z</i>	<i>p</i>
Internet use	-0.018	0.011	-1.57	.116
Phone use	0.025	0.011	2.35	.019
Mobile phone use	-0.012	0.004	-3.05	.002
Autocracy	0.007	0.011	0.67	.505
Unemployment rate (% change)	-0.002	0.040	-0.05	.961
GNI	0.001	<0.0005	1.88	.060
GDP	0.007	0.006	1.25	.212
Population	<0.0005	<0.0005	-1.04	.300
Internal war	-0.060	0.105	-0.54	.591
No elections	-0.173	0.073	-2.01	.045
Wald $\chi^2 = 32.28$ $p < .0005$				

*Note.* GNI = Gross National Income; GDP = Gross Domestic Product.

**Regression 5, High Mobile Use Group (n = 234)**

The model for the high mobile use group was significant (Wald  $\chi^2 = 21.88$ ,  $p = .016$ ), indicating that the predictor model using the dataset was improved over a model in which all predictors were set to zero. Significant predictors included Internet use ( $z = 2.39$ ,  $p = .017$ ), mobile phone use ( $z = -2.92$ ,  $p = .004$ ), and autocracy ( $z = 2.04$ ,  $p = .041$ ). Incidence rate ratios (IRR) were computed for the three significant predictors. The IRR for Internet use (1.019) indicates that if all other predictor variables are held constant, then a one point increase on the percentage of a country’s Internet use would increase the number of protests by a factor of 1.019. The IRR for mobile phone use (.988) indicates that if all other predictor variables are held constant, then a one point increase in the mobile phone use rate would decrease the number of protests by a factor of .988. The IRR for autocracy (1.034) indicates that, given the other predictors are held constant; a percentage point increase in a country’s autocracy score would increase the number of protests by a factor of 1.034.

*Table 7: Negative Binomial Regression Coefficients for Predictors on Criterion of Number of Protests for High Mobile Use Sub-Group (n = 234)*

Variable	<i>B</i>	<i>SE B</i>	<i>z</i>	<i>p</i>
Internet use	0.019	0.008	2.39	.017
Phone use	0.020	0.011	1.81	.070
Mobile phone use	-0.002	0.004	-2.92	.004
Autocracy	0.033	0.017	2.04	.041
Unemployment rate (% change)	0.072	0.124	0.62	.535
GNI	<0.0005	<0.0005	0.96	.339
GDP	-0.011	0.011	-0.99	.325
Population	<0.0005	<0.0005	0.60	.546
Internal war	0.253	0.246	1.32	.187
No elections	0.103	0.138	0.83	.408
Wald $\chi^2 = 21.88$ $p = 0.016$				

*Note.* GNI = Gross National Income; GDP = Gross Domestic Product.

### 3.5: Discussion and Interpretation

The purpose of this chapter was to use econometric analysis to test whether access to information and communication technologies (ICTs) is a statistically significant predictor of the protest events in non-permissive environments using data from 1990 to 2007. A total of 38 countries were selected for the negative binomial regression analysis. Regressions were run on five different country clusters. The first cluster included all 38 countries. The remaining four clusters were divided by high and low levels of ICT access and protest levels.

All five negative binomial regression models on the entire 18-year time panel for the study data were significant. Of note, however, is the non-significance of the Internet variable in all models analyzed. Mobile phones were only significant in the regression models for the “Low Protest” and “High Mobile Phone Use” clusters. However, the relationship was negative in the former case and positive in the latter. In other words, an increase in mobile phone users in countries with low protest counts decreases the number of protests. The overall correlation analysis showed a strong significant negative relationship between mobile phone use and unemployment. This may imply that an increase in mobile phone users in low protest countries leads to a decrease in unemployment and hence fewer protests. Another explanation might be that regimes in the “low protest cluster” are tech-savvy and able to prevent mass social unrest. The regression analysis on the “high mobile phone use” cluster revealed a significant positive relationship between mobile phone users and protests. This may imply that social unrest is facilitated by the use of mobile communication in countries with widespread access to mobile phones, keeping other factors constant.

These conclusions from the regression analysis require some important qualifications. First, as discussed in the data section, the protest data may suffer from media bias. Second, the protest data does not provide any information on the actual magnitude of the protests. Third, economic data on countries under repressive rule need to be treated with suspicion since some of this data is self-reported. For example, authoritarian regimes are unlikely to report the true magnitude of unemployment in their country.

ICT data is also self-reported. Fourth, the data is aggregated to the country-year level, which means potentially important sub-national and sub-annual variations are lost. Fifth and finally, the regression results may be capturing other dynamics that are not immediately apparent given the limits of quantitative analysis. Taken together, these qualifiers may explain why Howard (2011) argues that, “a large-N quantitative approach, with its assumption of well-defined categories and populations and quest for the net effects of independent variables in linear models, is the least appropriate template for this research.” Moreover, “many of the traditional statistical techniques do not lead to conclusions about causal connections. Instead, they lead to models of ‘explained variation,’ a different thing” (Howard 2011).

Qualitative comparative analysis is therefore needed to test and potentially validate the results derived from this quantitative study. Indeed, “perhaps the best reason to proceed in a qualitative and comparative way is that the categories of ‘democracy’ and ‘technology diffusion’ are themselves aggregates and proxies for other measurable phenomena” (Howard 2011). Unpacking and then tracing the underlying causal connections between ICT use and protests requires qualitative methodologies such process-tracing and semi-structured interviews. The conceptual framework developed in Chapter 2 serves as an ideal framework to inform both the process-tracing and interviews. The next chapter will therefore introduce two qualitative case studies to critically assess the impact of ICTs on state-society relations in countries under repressive rule.