Explaining Cross-National Variation in Voter Turnout:

Aggregate and Temporal Patterns

Arild Een

Master Thesis
Department of Comparative Politics
University of Bergen
September 2007

Abstract

Over the last 25 years, cross-national variation in voter turnout has received increased attention from social science researchers. The dominant view in the field is that existing research on voter turnout has established some robust patterns and we know relatively well why voter turnout is higher in some countries than in others. Key variables for explaining cross-national variation in voter turnout are compulsory voting, electoral system, level of economic development, unicameralism, size of country, and literacy rate. This thesis formulates hypotheses concerning the causal effects of these variables alongside additional theoretically important variables, estimates their causal significance and checks for the robustness of their effects. By conducting a comprehensive research strategy involving both general and time-specific cross-sectional analyses, this thesis tests a series of research hypotheses on a data set that spans 90 countries across a long time period (1950-2000). The results indicate that the majority of the determinants behind cross-national variation in voter turnout are time-specific. This means that we cannot (as previous research has done) discuss the determinants behind cross-national variation in voter turnout without taking the time dimension into account. Only very few determinants, like compulsory voting, economic development, flow of information and equality in the distribution of income can be said cause variation in turnout across countries irrespective of time.

Acknowledgements

I leave this thesis indebted to my supervisor, Michael Alvarez, whose firm assistance has been invaluable. Thank you for your flexibility in the face of my ad hoc visits and irregular email correspondence, for helpful and constructive comments, for introducing me to the Nlogit software package, and for your linguistic advice.

I would also like to thank the Department of Comparative Politics at the University of Bergen as well as Pål Hermod Bakka at the Social Science Faculty Library. Thank you for your kind assistance upon my request for books, articles and datasets.

And last, but not least, thank you for all your encouragement, Annelise. Your support really means a lot to me.

Errors and shortcomings are my responsibility alone.

Bergen, September 2007
Arild Een

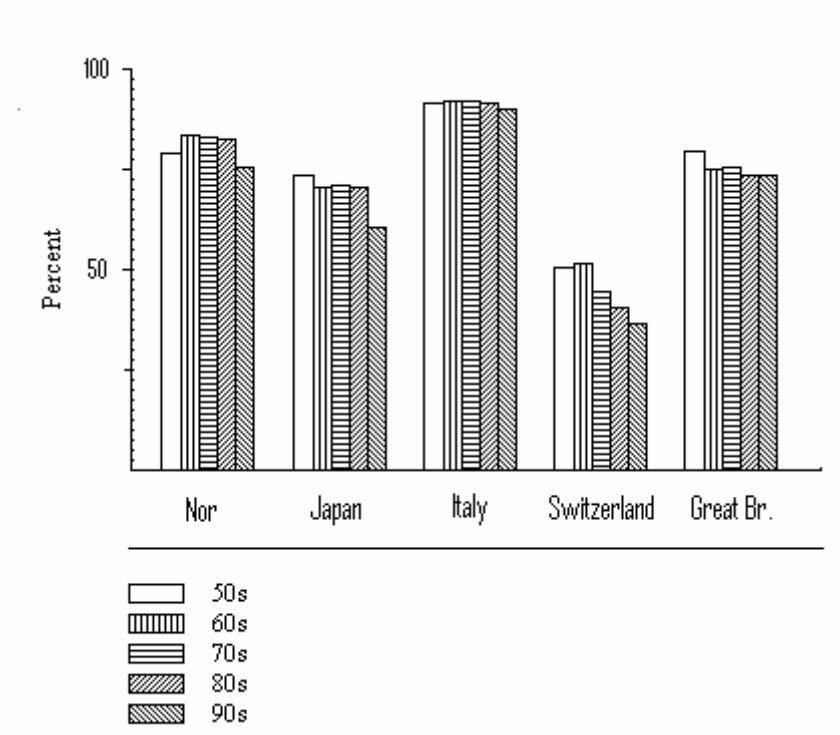
List of Contents

List of Figures and Tables

I. Introduction

A central democratic institution is the *election*. By showing up at the polls and casting their vote, the electorate decides which party or coalition of parties is to rule them for a fixed period of time. The number of people casting their vote at an election constitutes the *electoral turnout* of the election.[1] Even though voting is the form of electoral participation in which most people engage, we find that there is considerable variation across the democracies of the world in voter turnout rates. For example, average voter turnout in the 1970s was 44% in Switzerland, 75% in Great Britain, 80% in Norway, 72% in Japan, and 94% in Italy. As Figure 1 below shows, while the degree of variations *within* countries is not particularly striking, the degree of variation *across* countries is quite significant.

Figure 1: *Mean Vote/VAP, 1950s-90s*[2]



Note: Mean Vote/VAP is measured as the number of valid votes as a proportion of the Voting Age Population in parliamentary elections. VAP, Voting Age Population. Source: International IDEA database, available online at www.idea.int.

[1] Voter turnout has commonly been measured in two different ways: turnout as a proportion of the registered electorate or turnout as the proportion of the voting age population (VAP) that cast a vote (discussed below).
[2] Votes cast as a proportion of the voting age population, 1950s, 60s, 70s, 80s and 90s.

How can these cross-national differences in turnout be explained? More specifically: Why does Italy enjoy a higher turnout than Norway? Why do Norwegians vote at a higher rate than the Japanese? Furthermore, do factors that explain cross-national variation in turnout in the 1960s, also explain variations in the 1990s? In other words, are the factors behind cross-national variation in voter turnout consistent across time? Questions like this motivate this thesis.

1.1 Where we are

The dominant view in the literature is that the existing research on voter turnout has established some robust patterns and we know relatively well why voter turnout is higher in some countries than in others. Some argue that differences in voter turnout are inevitable because countries differ in terms of political culture – that is, in terms of their citizens' "subjective orientation to politics" (Pye and Verba 1965:513). In other words, we are dealing with participatory cultures, and some countries are said to have more of it than others (Almond and Verba, 1963). Cultures that foster such participatory values enhance voting turnout. An alternative to the cultural explanation of voting turnout centers on institutional factors. As Jackman (1987) argues, it makes good intuitive sense that voter turnout should respond to institutional patterns. Voting is everywhere systematically governed by laws and institutional arrangements that vary markedly from nation to nation. Jackman lists a number of institutional characteristics that affect voter turnout, including compulsory voting, electoral disproportionality (in turning votes into seats), unicameralism and the party system (number of parties). His findings are supported by subsequent studies (Blais and Carty 1990; Franklin 1996; Blais and Dobrzynska 1998; Norris 2004) which also add other characteristics to the "institutional" explanation: electoral frequency, electoral system, electoral decisiveness and district size.

Verba, Nie, and Kim (1978), in their study of different forms of political participation, propose a distinction between two types of forces that affect political activity. On the one hand, individuals bring attitudes and characteristics to the political arena, and, on the other hand, this 'participation' is facilitated or hindered by the institutional context within which an individual acts. As for the first dimension, participation is, in general, facilitated by greater socio-economic resources and by general levels of political awareness and self-confidence. This is supported by a number of studies (Powell 1982, 1986; Blais and Dobrzynska 1998;

Norris 2004) which present a number of socio-economic variables that affect turnout: level of economic and societal development, literacy rate, the size of the country and party group linkages. The second dimension affects the first, as legal rules, social and political structures, and configurations of partisanship all present the individual with conditions that shape his or her choices. Furthermore, these conditions are relatively difficult for the individual to change. Hence, voter turnout is affected by what the individual brings to the participatory scene (socio-economic and cultural "baggage"), a *scene* which itself play a role in shaping participation through rules, structures and institutional characteristics. However, when cross-national variation in voter turnout is discussed there is one distinction that is important to keep in mind. Voter turnout at the aggregate level is a feature of an electorate, not a voter. Even though an electorate is an aggregate of voters, the process of aggregation is not simply one of adding up relevant features of the individuals who form part of it. An electorate is not a voter writ large, any more than an economy is a consumer writ large (Franklin 2004).[3] This paper, like other studies dealing with cross-national variation in voter turnout, concerns itself with turnout at the *aggregate* level, meaning that it does not ask questions like "Why do people vote (in the first place)?". The question is rather: "Why do Italians vote at a higher rate than Norwegians?" and "Why do Norwegians vote at a higher rate than the Japanese?", and so on.

1.2 Purpose of paper

Research has dealt specifically with these kinds of questions for over 25 years. Do we need more research on the topic? The answer is yes and the reasons are threefold:

- Too few cases. Research dealing with cross-national variation in voter turnout has not exploited the richness of the data that exist. Instead they too often rely on a limited number of cases (19-32) upon which the possibility of making generalizations is limited. For example, one of the most famous studies on cross-national variation in voter turnout, Jackman's (1987) *Political institutions and voter turnout in industrial democracies*, rely upon 19 countries from which he generalizes across the entire universe of democracies.
- Too short time period. Research typically examines short time periods upon which the possibility of making generalizations across time is limited. For example, Norris

[3] In his famous studies of suicide rates, Durkheim confronted the same problem of distinguishing between the causes of suicides and the prevalence of high suicide rates.

(2004), in her *Electoral Engineering*, generalizes upon findings conducted across only six years (1996-2002).

- Contradictory findings. A number of variables are found to have different effects and sometimes even different directions of correlation across the studies.

I am not convinced by previous research on cross-national variation in voter turnout. This paper has come forth as a result of this and I will make a unique contribution to our knowledge of cross-national variation in voter turnout, by addressing the issues above. In my study I will:

- Include a great number of cases. By including every democratic country in the study, the results I obtain are more robust. Based on ACLP's[4] dichotomous approach (discussed below), I have come up with a list of 90 countries to include in my analyses. Only by including as many democratic regimes as possible in the analyses, can the results be generalized upon, irrespective of the space dimension.
- Expand the time period under review. By examining a long time period (1950-2000), conducting analyses across each decade and looking for time-specific variance, the results I obtain are more robust. Only by conducting across-time sensitivity regression analyses can the results be generalized with confidence, irrespective of the time dimension.
- Once and for all try to determine the direction and effect of each variable's correlation with voter turnout. This is achievable based on the above, and the inclusion of a wide range of control variables, some old and "established", other new and never before included in research on cross-national variation in voter turnout. Only by including a great variety of explanatory variables will the true effect of each respective variable upon turnout stand out.

The discussion of voter turnout is often followed by a discussion of the health of the democratic institutions. Low electoral participation is often considered to be bad for democracy, whether inherently or because it calls legitimacy into question by suggesting a lack of representation of certain groups and in-egalitarian policies (Franklin 2004). Above all, as Franklin (2004) notes, "low turnout seems to be seen by commentators as calling into

[4] ACLP is an acronym for a comprehensive dataset, compiled by Alvarez, Cheibub, Limongi and Przeworski in the preparations for *Democracy and Development* (2000).

question the civic mindedness of a country's citizens and their commitment to democratic norms and duties" (2). However, differences in participatory norms are not, in turn, systematically linked to turnout rates. For example, as Crewe (1981) noted over 25 years ago,

> The best known (and still most useful) comparative study of subjective orientations, *The Civic Culture*, found that interest in politics, attention to political affairs in the media, feelings of civic duty and individual political efficacy, and trust in political as opposed to other solutions to individual and communal problems, were consistently higher in the United States, followed by Britain, then Germany, and finally Italy – exactly the *reverse* of their rank order for postwar turnout! (239: emphasis added).

Whether or not low turnout is bad for democracy and whether or not countries that experience higher turnout are necessarily better democracies than those experiencing lower turnout is not the concern of this paper. For a discussion on the topic, see Franklin (2004) and Norris (2004).

1.3 Structure of paper

As noted above, the dominant view in the literature is that we know relatively well why turnout is higher in some countries than others. I start out, in *Chapter 2*, with a presentation of the findings in the literature, focusing on six studies which represent important contributions to our understanding of cross-national variation in voter turnout. Each study is discussed both in terms of their sample selection and findings. This approach will allow me to build upon the existing literature and hence make it easier for me to make a contribution to this field of research. In *Chapter 3* I discuss my basis for sample selection and identity the variables I will include in my analyses. All of the variables from the previous chapter (alongside some new variables) are included, and each variable is discussed in terms of measurement and data source. The chapter ends with a discussion of the validity and reliability of my measurements. In *Chapter 4* I present my methodological approach by briefly discussing the quantitative method, the regression analysis techniques at hand, namely the OLS-method, alongside my research strategy. In *Chapter 5* the strategy is played out as I conduct the analyses for the general model, which pertains to explain cross-national variation in voter turnout for the entire time period under review. My analyses are time-specific in *Chapter 6*, meaning that the

analyses are conducted on the basis of shorter time periods (decades). The final chapter summarizes the findings and provides suggestions for future research.

II. Previous research

Comparative research has long sought to understand the reason for voting participation and the explanations for cross-national differences.[5] I will now present some selected works in the literature on cross-national research in order to establish a basis for my own analysis. Many studies have been conducted to try to explain the differences in voter turnout one observe among democratic regimes. I will present some of these studies and have decided to focus my attention on the following seven publications: Powell (1982; 1986), Jackman (1987), Blais and Carty (1990), Franklin (1996), Blais and Dobrzynska (1998), and Norris (2004). These are all regarded as important contributions to our understanding of cross-national differences in voter turnout (Blais 2006). By reviewing these studies and identifying their results, we get a good picture of what we know today about the determinants of cross-national variation in voter turnout. Furthermore, by identifying significant variables from the literature, I will establish a theoretical platform from which I can base my own analysis. This way it is possible to build upon the existing works of other scholars and, in addition, make my own contribution to this field of research.

The first necessary step in any study of voter turnout under democracies is to identify the population of existing democratic regimes. Before we turn to a discussion of the findings, a brief note on each study and their approach to this matter is in place.

2.1 Scope of samples in existing studies[6]

By defining five criteria that need to be satisfied in order for a country to be classified as democratic, and by reviewing the work of other scholars, Powell, in his award-winning book *Contemporary Democracies* (1982), came up with a list of 29 countries to include in his analysis. Powell's American Political Science Review article, *American voter turnout in comparative perspective* (1986), examined 20 countries in the 1970s, including Western European countries, Israel, Canada, the Unites States, Australia, New Zealand and Japan. This article's main emphasis was American voter turnout in comparative perspective, so the low number of cases is justifiable. In *Political institutions and voter turnout in industrial*

[5] Powell 1982, 1986; Jackman 1987; Blais and Carty 1990; Jackman and Miller 1995, Franklin 1996; Blais and Dobrzynska 1998; Norris 2004.
[6] A complete list of countries and years for each study can be found in Appendix D.

democracies (1987), Jackman chose to confine his analysis to 19 industrial democracies in the 1970s. Jackman's list of countries is the same as Powell's (1986), except that Jackman excludes Spain in view of the fact that "Franco did not die until late 1975, and the subsequent transition to democracy was neither smooth nor immediate" (Jackman 1987: 409).

Blais and Carty, in their European Journal of Political Research article *Does proportional representation foster turnout?* (1990), rely on Mackie and Rose's *International Almanac of Electoral History* (1982) as updated by the *European Journal of Political Research* to 1985. Of the 24 countries included there, they exclude Greece, Portugal and Spain because of their interrupted electoral history along with the United States, and end up with 20 countries to include in their analysis. Franklin's study, *Electoral participation* (1996), includes 29 countries in his analysis, mostly European countries along with Japan, India, the United States, Brazil, Australia, New Zealand Venezuela, and Costa Rica. He relies on Mackie and Rose's *International Almanac of Electoral History* (1991) and Katz (1996).

Blais and Dobrzynska, in *Turnout in electoral democracies* (1998) argue in favour of the importance of including *all* democratic elections: "if we wish to arrive at comprehensive understanding of the sources of cross-national variations in turnout, we should look at as many cases as possible and exploit the richness of data provided by the process of democratization" (240). Using data from Freedom House to classify regimes, they are able to study turnout in 324 democratic elections held in 91 different countries between 1972 and 1995. Since Blais and Dobrzynska relies on the judgement of one standard source as to whether a country is democratic or not and include all democratic elections, their study constitutes a major improvement over previous studies.

In *Electoral engineering* (2004), Norris relies on multiple sources of data. The most important one concerns survey research drawn from the Comparative Study of Electoral Systems (CSES) and includes Module I, which allows her to compare surveys of a representative cross-section of the electorate in 37 legislative and presidential elections in 32 countries.

Notwithstanding the impressiveness of the studies considered here, a challenge for new research would be to expand the range of cases even further, incorporating a wider range of regions, states, and years. I will address this issue in chapter 3, where I provide a rigorous justification for the inclusion of more cases. Let us now examine the findings that the studies

above have come up with. I have organised the findings along the lines of Powell (1982) and Blais and Dobrzynska (1998) who both distinguish different kinds of factors that influence political performance in general and voting participation in particular: institutional determinants, the socio-economic environment and the party system.

2.2 Institutional variables

Institutional variables are important determinants of cross-national variation in voter turnout because, as Jackman (1987) argue, voting is everywhere systematically governed by laws and institutional arrangements. Legal rules, social and political structures, and configurations of partisanship varies cross-nationally and all present the individual with conditions that shape his or her choices. The following section presents the institutional variables from the literature.

Compulsory voting

The use of compulsory voting laws can be expected to have an impact on turnout. As Tingsten (1975) demonstrated quite clearly over seventy years ago, the imposition of relatively small fines or other penalties can have a major impact on voting turnout. Jackman (1987) estimates that compulsory voting increases turnout by about 13 percentage points. Blais and Carty (1990) find that compulsory voting increases turnout by "twelve percentage points" (176). Franklin (1996) estimates that countries that make voting obligatory experience an increase of seven percentage points on voter turnout (226). Blais and Dobrzynska's (1998) findings support the trend: "compulsory voting boosts turnout by 11 points" (246) and so does Norris's (2004): "a turnout increase of almost 8 percentage point is expected" (169).[7] Although these findings are hardly surprising, it is important to take account of this factor before we can assess the impact of other institutional and social conditions. "Compulsory voting increases turnout" can be construed as a well-established proposition (Blais 2006: 113).

Nationally competitive districts and PR

It has frequently been suggested that voting participation is or would be enhanced by the introduction of proportional representation and multimember legislative districts, as opposed to single member district plurality or majority representation. The obvious argument in favour of this suggestion is that with single member districts, "it is likely that some districts will be

[7] This applies only to older democracies, see Norris 2004: 169.

non-competitive, giving citizens less incentive to vote; parties, less incentive to campaign there" (Rose 1980: 12). According to incentive-based explanations, the electoral formula can be expected to affect voter turnout, as citizens are faced with differential rewards under alternative electoral arrangements. PR-systems should generate higher turnout than majoritarian systems as supporters of minor and fringe parties in the latter system "may believe that casting their votes will make no difference to who wins in their constituency, still less to the overall composition of government and the policy agenda" (Norris 2004: 162). In constituencies where the incumbent party is unlikely to lose, the wasted vote argument travels well. In contrast,

> PR elections with low thresholds and large district magnitudes … increase the opportunities for minor parties with dispersed support to enter parliament with only a modest share of the vote, and, therefore this could increase the incentives for their supporters to participate (Norris 2004: 162).

Jackman's (1987) four-category ordinal variable takes into account the electoral formula and the size of the districts. The same variable, or dummy variables that distinguish electoral formulas, or a summary disproportionality index, has been utilized in further research. Blais (2006) writes that the studies that have been confined to advanced democracies (Jackman and Miller 1995, Radcliff and Davis 2000) as well as one study of turnout in post-communist countries (Kostadinova 2003) "have confirmed that turnout is higher in proportional representation (PR) and/or larger districts, whereas research dealing with Latin America reports no association (Perez-Linan 2001, Fornos et al. 2004), and an analysis that incorporates both established and non-established democracies concludes that the electoral system has a weak effect" (113-14).

These findings provide a basis for interpretation. Is it the case that PR increases turnout except perhaps in Latin-America, or is it the case that once one move outside of Europe there is no generalized correlation between the electoral system and turnout? Blais (2006) leans towards the second, more sceptical interpretation. Even though Jackman's finding on this variable supports Powell (1986), Blais and Dobrzynska (1998) question these findings on the grounds of variable specifications. They argue that since both Powell and Jackman rely on a variable they call "nationally competitive districts", which encapsulates the combined effect of the electoral formula and of district magnitude; the result is that it is not clear what it measures

exactly (245). Blais and Dobrzynska thus claim that it is appropriate to re-examine, with a larger sample if elections and a more systematic set of control variables, the impact of electoral systems on turnout. They do this by creating dummy variables for *PR, plurality, majority, and mixed systems* (245). Their final model establishes that, all else held equal, "turnout is three points higher in PR systems" (248). The difference is small and displays a relatively low significance (t-statistic of 1.6). Blais and Carty (1990) find that "turnout is seven percentage points lower in a plurality system, and five percentage points lower in a majority system, compared with PR" (174-5). Norris (2004) found that "legislative elections held during the 1990s in the CSES countries under comparison generated 75% turnout (Vote/VAP) under PR systems, 10% higher than under those elections contested in majoritarian systems, and a similar pattern was confirmed in a broader comparison of all 164 nations holding competitive elections worldwide during the 1990s" (161). Because, as Norris points out, the type of electoral system is a categorical rather than a continuous variable, she provides further details about the impact of different electoral systems on worldwide levels of turnout in the 1990s. The results, without any controls, confirm that average turnout was highest among nations using PR, namely, party lists and the STV electoral systems. Norris's results indicate that the basic type of electoral systems indeed, shapes the incentives to participate, with the key distinction being that between PR systems and all others (161-2).

Electoral Disproportionality

Most electoral systems produce a degree of *disproportionality* in favour of the largest party, but some systems generate a good deal more than others (Jackman 1987: 407). The greater the disproportionality, the more likely are the votes of minor-party supporters to be wasted. Jackman's findings, which are statistically significant, support this but the effect is small. Franklin (1996) finds that proportionality is worth "about six tenths of a percentage point in turnout for every percent by which the distribution of seats in the legislature approaches proportionality with the distribution of votes" (226). Countries vary in terms of the proportionality of their electoral systems from a low of 79 in Britain to a high of 99 in Germany (numbers from Franklin's analysis). That is a 20-point difference which (multiplied by 0.6) translates into a difference of 12 percent in turnout.

The inclusion of electoral disproportionality as a control variable by Franklin (1996) was only made possible by his exclusion of electoral formula (electoral system) as a control variable. As the two variables correlate too highly they can not be included in the same model.

Jackman (1987) includes a variable called nationally competitive districts that takes into account the electoral formula and the size of the districts. (I am surprised that correlation with electoral disproportionality did not rule out this variable.)[8]

Unicameralism

On unicameralism we follow the argument of Jackman (1987):

> ….unicameralism is important in producing decisive governments. Where there is no second house (as in New Zealand), governments based on the first house do not have to compete and compromise with another legislative chamber. In contrast, where there is strong bicameralism (as in Switzerland), legislation can only be produced by compromise between members of the two houses. This means that elections for the lower house play a less decisive role in the production of legislation where bicameralism is strong. Unicameralism should therefore foster turnout (408).

The more powerful the body that is being elected, the stronger the incentive to vote. Jackman measures unicameralism by using the criteria proposed by Lijphart (1984), which operate with four scores, depending on "how unicameral" the system is. The most unicameral system is assigned the score of four and the most bicameral system is assigned the score of one (see Lijphart 1984, 213, for the scoring system). Jackman estimates that each 'score' is worth an almost 2 percentage point increase in turnout, meaning that a country with a strong unicameral legislature will enjoy an increased turnout at elections of about eight points compared with an election for a legislature that is strongly bicameral. The findings about the impact of unicameralism on turnout are mixed. While Jackman (1987), Jackman and Miller (1995), and Fornos et al. (2004) report positive findings, Blais and Carty (1990), Black (1991), Radcliff and Davis (2000), and Perez-Linan (2001) indicate no effect (Blais 2006: 114).

Decisiveness of elections

Blais and Dobrzynska (1998) do not include unicameralism as a control variable. Instead they include an institutional characteristic that pertains to the decisiveness of elections. As they look at turnout for elections of the national lower house, they argue that the more powerful the

[8] Jackman was aware of the possible problem

national lower house, the more decisive the election, and thus, the higher the expected turnout. In many respect, their variable measures the same phenomenon as the variable that takes into account unicameralism. We follow their argument, which:

> . . . predict[s] that turnout will be lower if there is an elected upper house or president or if the country is a federation. This prediction holds only if subnational, presidential or upper house elections are *not* held at the same time. The presence of an elected upper house, for instance, matters only if the lower and upper house elections are not held at the same time: in such a context, the lower house election can be construed as being less decisive, and turnout could be lower. If the two elections are held at the same time, the situation is equivalent to there being one house (245-46: italics in original).

By constructing dummy variables, they take into account the presence and timing of subnational, upper house, and/or presidential elections. The prediction is that turnout is higher when the election is more decisive. The results, which are consistent with Jackman's (1987) findings, indicate that turnout is affected by the decisiveness of elections: all else held equal, "turnout is reduced by 6 points when lower house elections are least decisive" (Blais and Dobrzynska 1998: 246).

Frequency of elections

Norris (2004) includes a variable that measures the frequency of elections, which proves to be strong and statistically significant in a negative direction, arguing that when the frequency of elections goes up, the voter is facing increasing costs and possible voter fatigue, which in turn affects voter turnout.

District size

Norris (2004) also includes a variable that take into account the number of electors per member of parliament. She argues that the higher the number of electors, the lower the turnout because it is more difficult to mobilize many voters than few voters for the representative. The possibility for increased information, familiarity and contact between voters and their representatives is lower if the number of electors per representative is high. Her finding, which supports her assumption, is not statistically significant.

2.3 Socio-economic variables

Socio-economic variables are important determinants of cross-national variation in voter turnout because political participation is facilitated by greater socio-economic resources (Norris 2004). The following section presents the socio-economic findings in the literature.

Level of development – economic (GNP per capita) and societal (HDI)

A large body of writing on theories of cross-national participation is the literature on social modernization and political mobilization. In *Electoral Participation* (1980), Powell argues that "economic development" has "important consequences for mass political activity, as the achievement of higher levels of economic development is associated with major transformations of the social and economic structure of society" (21). Powell (1982) posited that economic development fosters turnout and Blais and Dobrzynska (1998) confirm this. They measure economic development as GNP per capita. As the relationship is logarithmic, it implies that the main difference is at the low end of economic development: "everything else being equal, turnout increases by 13 percentage points when GNP per capita moves from 163 American dollars (Malawi in 1994), the lowest in the sample, to the average (7,614) but only by 5 points when it moves from the average to the highest, 30,433 (Luxembourg in 1994)" (244-43).[9] Jackman (1987) leaves out socio-economic variables altogether, perhaps because he deals with a small number of cases among which there is little variance in the level of socio-economic development. Norris (2004) finds turnout to be slightly higher in more developed societies, as gauged by the United Nations Development Programme (UNDP) Human Development Index. The societal changes associated with the modernization process do strengthen electoral participation, according to her study.

Literacy Rate

Even though Verba, Schlozman and Brady (1995) show that voting is the least demanding form of political activity and the one that is least dependent on the possession of civic skills,

[9] After taking the natural logarithm of the independent variable, the estimated regression coefficient is interpreted in the following manner. The coefficient shows the absolute change in the dependent variable that results from a given proportional change in the independent variable. Since the function is nonlinear, the absolute impact upon the dependent variable will vary, depending upon whether we are at lower or higher values of the independent variable. Therefore, one can compare the change in the dependent variable that results from moving from the minimum value of the independent variable to its average, as compared to the impact that results from moving from the average value of the independent variable to its maximum (the actual numbers for each variable – minimum value, average, maximum value – can be read directly from the descriptive statistics tables, presented in several of the Appendixes).

they note that those with little linguistic skills are less likely to vote. A minimum degree of literacy is almost a prerequisite to good turnout. Blais and Dobrzynska (1998) support this: all else held equal, "turnout increases by a hefty 16 points when the *il*literacy rate moves from the highest level (85%) to the average (12%) but by only one point when it moves from the average to the lowest level (2.5%)" (244, emphasis added).

Size of country - population

The last socio-economic variable to have an impact on voter turnout in Blais and Dobrzynska's study is the size of the country. As this relationship is logarithmic, the important difference is between smaller countries and all other countries. All else held equal, "turnout is 7 points higher in a country of 100,000 people than in one of 26 million, which is the average in the sample; the difference between turnout in a country of 26 million and one of 100 million is only 2 points" (Blais and Dobrzynska 1998: 244). This confirms the view that smaller countries are able to arouse a greater sense of community, which itself fosters a higher turnout. Blais and Carty's (1990) data confirm that electoral participation tends to be somewhat higher in smaller polities: "the regression indicating that we should expect, everything else being equal, electoral turnout to be seven points higher in Luxembourg than in Britain" (176).

2.4 Party System

In legislative elections, voters are offered to make a choice among parties and candidates. Turnout depend on the choice offered to the electorate and that choice is very much structured by the party system.[10] I will now present the findings from the literature on this bloc.

Number of political parties

The *multipartyism*-variable follows the arguments of Downs (1957). Downs argues that voters in a multiparty system face a fundamental problem: they do not directly select the government that will govern them, as the government most likely will be selected by the parties in the legislature. The paradox for multiparty systems becomes apparent:

[10] Some would argue in favour of placing the party system variables among the institutional variables. This could be justifiable (I do it below in my own research) but some (Blais and Dobrzynska 1998) argue that, for example, the number of parties are consequences of the institutional setting (electoral system), rather than an institutional characteristic per se.

> The type of political system which seems to offer the voter a more definite choice among policies in fact offers him a less definite one. This system may even make it impossible for him to choose a government at all. Instead, it may force him to shift this responsibility onto a legislature over which he has very little control between elections (Downs 1957: 156).

This implies that elections play a less decisive role on government formation within multiparty systems. Multipartyism should therefore depress turnout, and does, by three percentage points, according to Jackman's (1987) study.

Blais and Dobrzynska (1998) also include a multiparty-variable as they predict that *the greater the number of parties, the more choice electors are offered and higher the turnout*. Still, as they acknowledge, the door swings both ways on this matter. The greater the number of parties, the more difficult it can be for electors to make up their minds. Furthermore, the greater the number of parties, the less likely it is that there will be a one-party government. Thus, they formulate a contrary hypothesis: *the greater the number of parties, the smaller the probability of a one-party majority government and the lower the turnout*. Blais and Dobrzynska's findings confirm Jackman's (1987) finding that turnout tends to be reduced when the number of parties increases. The relationship is logarithmic: "turnout declines by 4 points when the number of parties moves from 2 to 6, but by only 2 points from 6 parties to 10 and from 10 to 15" (Blais and Dobrzynska 1998: 249). Almost all empirical research has found a negative correlation between the number of parties and turnout (Jackman 1987, Blais and Carty 1990, Jackman and Miller 1995, Blais and Dobrzynska 1998, Radcliff and Davis 2000, Kostadinova 2003). The only exceptions are Norris (2004), who finds the relationship to be positive and small but statistically significant, and studies of turnout in Latin America, where there appears to be no relationship at all (Perez-Linan 2001; Fornos et al. 2004).

Party-group linkages

The party-group linkages enhance turnout because "partisan choice should seem simpler to the less involved; cues from the personal environment of the individual (friends, family and co-workers) should be more consistent; party organizers can more easily identify their potential supporters in making appeals and in helping voters to the polls on election day" (Powell 1986: 22). Voting is simpler when and where groups (e.g., unions, churches, professional associations) are clearly associated with specific parties (Blais 2006: 1). Powell's

finding on this variable were not replicated by Jackman (1987), and subsequent studies have left out this variable, which is the approach that I too will follow.

2.5 Summary of findings

The survey just discussed is summarized in the table presented below. The explanatory variables are ordered just as they were discussed above, divided into three "blocs" of variables. For each of the six studies surveyed (Powell 1982 is left out) and each of the independent variables, I indicate the direction of causality (+ or -) as well as whether the variable was found to be statistically significant at the 95% confidence level (indicated by a *).

Table 1: *Literature review - variables, direction and significance*

	Study					
	Powell (1986)	Jackman (1987)	Blais & Carty (1990)	Franklin (1996)	Blais & Dobrzynska (1998)	Norris (2004)
Institutional Variables						
Compulsory voting		+*	+*	+*	+*	+*
PR			+*		+*	+*
Nat. com. districts	+	+*				
Electoral decisiveness					+*	
Electoral disproport.		–*		–*	–*	
Unicameralism		+*				
Electoral frequency						–*
District size						+
Socio-economic Variables						
GNP per capita					+*	
HDI						+*
Literacy					+*	
Size			–*		–*	
Party System Variables						
Multipartyism		–*	–*		–*	+*
Party group linkages	+*					

*=sig at 0.05

In Table 1 we see that the variables behind cross-national variation in voter turnout are stable in their direction of correlation. The effect of only one variable, multipartyism, varies across the studies, with Norris (2004) postulating a *positive* correlation between number of parties

and voter turnout. I will include the variables from Table 1 in my own research and check if their significance and direction of correlation changes when included in regression analyses conducted across more cases and over a longer time period.

I turn now to Chapter 3, which discusses my sample selection, presents my dependent variable, electoral turnout, and identifies my independent variables.

III. Sample selection and research hypotheses

In this chapter I will discuss my basis for sample selection as well as identity the variables and hypotheses for my analyses. Building upon the discussion from the previous chapter, I will argue in favour of including as many cases as possible in the analyses. Furthermore, I will include the variables identified in the previous chapter and in addition list the variables created specifically for this paper. I start out with a discussion of my sample selection.

3.1 Sample selection in this thesis

Recall from the discussion in the previous chapter that the studies concerned with cross-national variation in voter turnout had different approaches to the sample selection procedure. I argue for including as many cases as possible when studying cross-national variation in voter turnout. The rationale is clear: if we wish to arrive at a comprehensive understanding of the sources of cross-national variations in turnout, we should look at as many cases as possible (see King, Keohane and Verba, 1994). Furthermore, the inclusion of elections in the study should rely on one standard source as to whether or not a country is democratic. This will give the analyses conducted more credibility and provide opportunity for replication.

3.1.1 Limitations of previous studies

Powell's basis for sample selection in *Contemporary democracies* (1982) has been criticized, as he "relies on the work of others who did not exactly have the same criteria and because there is disagreement among authors with respect to a number of cases, the inclusion and exclusion of which appear somewhat arbitrary" (Blais and Dobrzynska 1998: 239). Why Powell relies on multiple sources in his sample selection is not clear. In any case, Powell's approach is not the way to go.

Jackman (1987) confines his analysis to industrial democracies only. Why he excluded non-industrial democracies is not clear. One advantage in comparing countries as similar as possible in terms of socio-economic characteristics is that one thereby controls for these variables and can more easily isolate the impact of political institutions. One shortcoming, on the other hand, is that the number of cases is small, and that the results may not be generalizeable to the universe of democracies. Moreover, this approach does not let us measure the possible impact upon turnout of interesting explanatory variables like economic

and "social" development. I argue in favour of including as many cases as possible, in order to measure the impact of as many variables as possible.

Blais and Dobrzynska (1998) use data from Freedom House and their analysis is confined to post-1972 (the year first year of data available). By using a different measure of democracy, one can maximize the time aspect and thus include more cases in order to effectively test the robustness of the findings observed among established democracies. I turn now to a discussion of my basis for sample selection.

3.1.2 My approach

As I study electoral turnout in *democracies* it is necessary, in order to establish a basis for sample selection, to decide upon 'what' democracy is. That is to say, elections takes place in many regimes, I am interested in the elections taking place in democratic regimes. I have chosen the employ the regime classification data presented in Przeworski et al. (2000) (called ACLP for short). Their approach is dichotomous, as they classify every regime in the world between 1950-1990[11] as either a democracy or a dictatorship. They focus on contestation, and argue that only regimes, in which those who govern are selected through contested elections, that is, the chief executive office and the seats in the effective legislative assembly, can be classified as democracies. Contestation takes place when there exists an opposition which can assume office as a result of elections. They argue that Przeworski's dictum, "democracy is a system in which parties lose elections" (Alvarez et al. 2000: 16) is taken literally and alternation in office constitutes prima facie evidence of contestation they argue. Moreover, contestation entails three features: ex-ante uncertainty, ex-post irreversibility, and repeatability (on an elaboration of these matters see Alvarez et al. 2000: chapter one). My basis for sample selection, drawn from ACLP is clear: "Democracy is a system in which incumbents lose elections and leave office when the rules so dictate" (2000: 54).

Munck and Verkuilen (2002), in their review of nine democracy indices on the grounds of conceptualization, measurement, and aggregation, highlight ACLP as "particularly insightful concerning the selection of indicators and especially clear and detailed concerning coding rules" (27). The purpose of Munck and Verkuilen's paper is to provide a systematic assessment of the large-N data sets on democracy that are most frequently used in current

[11] It has been updated to 2002 which is essential for my purpose.

statistical research. Even though they conclude that "no single index offers a satisfactory response to all three challenges of conceptualization, measurement and aggregation" (28), the ACLP is regarded as one of the stronger indices in Munck and Verkuilen's review. As I am not in need of a continuous variable to include in my model, only one that makes my sample selection easy, the ACLP's dichotomous approach suits my purpose perfectly.

Furthermore, my approach represents a major improvement over previous studies as I, among very few, rely only on the judgment of one standard source as to whether a country is democratic or not. Blais and Dobrzynska (1998) also rely on one source, Freedom House, and their analysis is thus confined to post-1972 (the year from which Freedom House provide data). Relying on surveyed-based indices of political rights and civil liberties ranging from 1 to 7, Freedom House averages each pair to determine an overall status of "Free" (1.0-2.5), "Partly Free" (3.0-5.0), or "Not Free" (5.5-7.0). The Freedom House index has been harshly criticized, not least for its maximalist definition of democracy. The analytical usefulness of its index is severely restricted due to the inclusion of attributes such as "socio-economic rights", "freedom from gross socio-economic inequalities," "property rights, " and "freedom from war", which are more fruitfully seen as attributes of some other concept (Munck and Verkuilen 2002: 9-11). Furthermore, the Freedom House index includes so many components under its two attributes "political rights" and "civil rights" (9 and 13, respectively) and does so with such little thought about the relationship among components and attributes that it is hardly surprising that a large number of distinct or at best vaguely related aspects of democracy are lumped together (Munck and Verkuilen 2002: 14).

By employing only one democracy measure, my model constitutes a major improvement over previous studies. Based on the ACLP indices I have come up with a list of 90 democracies to include in my analysis.[12]

3.2 Variables and hypotheses

In the previous chapter I discussed the findings in the literature and identified important explanatory variables behind cross-national variation in voter turnout. All of the variables identified in that chapter will be included in my analyses alongside some new variables,

[12] See Appendix D for overview of countries and elections included in my analysis.

created specifically for this paper. I will turn to an identification of the variables I intend to include in this analysis, and I start out with the dependent variable, electoral turnout.

3.2.1 Dependent variable – electoral turnout

Electoral turnout has commonly been measured in two different ways. Turnout as a proportion of the registered electorate can be calculated as the number of votes divided into the number of citizens who are legally registered to vote. This measurement has been utilized by many scholars, including Blais and Dobrzynska (1998), Crewe (1981), Blais and Carty (1990), Black (1990), and Franklin (1996). However, as Norris (2001) argues, this method of measuring voter turnout can be misleading in situations with a restricted franchise, for example if only men or whites (e.g. apartheid South Africa) are eligible to vote, since in these countries official estimates of turnout can be relatively high even if the voices of all women or ethnic majorities are excluded. For these reasons, it is more satisfactory to compare turnout as a proportion of the voting-age population (VAP), representing the number of valid votes divided into the size of the population over the minimal legal voting age, whether enfranchised and registered or not. This way of measuring voter turnout has been utilized by Powell (1982, 1986) and Jackman (1987) amongst others. However, the measure has one major shortcoming, namely the difficulties in estimating the size of the eligible population. The eligible population is assumed to be the voting age population at the time of the election. As Powell (1986: 40) acknowledges:

> in most democracies voting eligibility is limited to citizens. Population figures…include noncitizens resident for a year or more. Countries vary substantially in the percentage of such aliens … we do not have good data on percentage of residents…who are aliens of voting age, and cannot systematically adjust our turnout data to remove them.

Furthermore, Black (1991), who reports such estimations for a small number of countries, indicates that "the entire exercise required drawing some overly simplified assumptions" (Blais and Dobrzynska 1998: 241). In the end, although this measurement has its shortcomings, it has become the standard measure adopted in cross-national research, as it provides a more consistent yardstick for comparing countries than the alternative.

In my thesis the dependent variable is voter turnout calculated as the percentage of the eligible population that cast a vote. As discussed above, this measure has its shortcomings but I find

them less worrying than the shortcomings associated with the alternative. All data on this variable are collected from The International Institute for Democracy and Electoral Assistance, IDEA, available online at www.idea.int.[13] IDEA provides the most comprehensive record of turnout in national elections that is currently available. They cover 1620 national-level elections held from 1945 to 2000 in all 193 independent nation-states, including 1218 parliamentary and 402 presidential elections.

3.2.2 Which elections?

Reif and Schmitt (1980) put forward a model explaining the difference in election participation and election profits of individual political parties on regional, national and European levels. They argue that national parliament elections are more salient, for both political parties and the public, and an increased political participation and turnout can thus be expected. Reif and Schmitt coined the phrase 'first-order' elections to differentiate such elections from other, less important, elections. Even though Reiff and Schmitt's work were on elections at the European level, their differentiation of elections based on the relative importance of the election is quite useful (Van der Eijk and Franklin: 1996). As I study voter turnout cross-nationally, first-order elections will constitute the basis for my dependent variable and I will hence study parliamentary elections for the lower house of the parliament.

3.2.2 Independent variables and hypotheses

Above I reviewed the literature and reported the most important variables that explain cross-national variation in voter turnout. I am interested in checking the significance and robustness of those variables in a specification that contains a wider range of control variables. The analyses in this thesis will thus constitute a robustness test of these former findings. As indicated, I also wish to include some new variables to contribute to our understanding of the cross-national variation in voter turnout. I proceed now to an identification of the independent variables I will include in my analysis. A brief discussion of measurement will follow each variable alongside data source and proposed hypothesis. Some of the independent variables are transformed, taking the natural logarithm of each respective variable. The transformation serves two purposes; first, the difference between high and low values on the variable are reduced as the variable becomes more symmetric, which in turn reduces the possibility of heteroscedasticity; second, I am interested in finding the absolute change in Y for a percent

[13] Given the lack of data, I was compelled to exclude four elections from my study; Switzerland 1971; Jamaica 1983; St. Kitts and Nevis 1993; and Kiribati 1998.

change in X.[14] The latter point refers to how the variables are measured; it makes more sense to interpret them in the terms of percentage changes rather than in unit changes.

I will, inspired by Powell (1982) and Blais and Dobrzynska (1998), organize my variables in different categories as I believe it is a fruitful and clear approach. I will operate with a set of five categories of variables: institutional; socio-economic; information circulation; activism; political legacy. I begin with an identification of the institutional variables.

3.2.2.1. Institutional variables[15]

(1) *Compulsory voting - COMP*

An inclusion of a compulsory voting variable is self-evident in any study on cross-national variation in voter turnout. I expect a higher turnout where compulsory voting laws are in practice because the electors are facing varying degrees of punishment if they fail to vote under such systems. The measurement however, is not as straightforward as one would presume. As a number of countries impose compulsory voting laws on its citizens, the same countries vary in how strict these laws are enforced.[16] Furthermore, there is also the issue of consequences facing the voters if they do not participate. These vary substantially across countries, from small fines in Brazil to denial of certain goods and services provided by public offices in Peru and Greece. In other words, it is a matter of consequences and chances of these consequences actually happening. I have, based on the guidelines of IDEA and by consulting Blais et al. (2003) come up with a list of 14 countries to include in my analysis (available in Appendix A).

H1: The presence of compulsory voting increases turnout

(2) *Electoral frequency - FREQ*

The frequency of elections is an important determinants for voter participation as it increases the costs facing electors and may produce voter fatigue. By calculating the years since the last national-level parliamentary or presidential election it is possible to measure the frequency of

[14] Gujarati (2003) refer to such a models as **lin-log models**, see p. 181

[15] I include a variable called NUMPART (number of parties – see below) in this bloc acknowledging that it is not really an institutional factor, rather, a consequence of the electoral system which itself have an impact on turnout.

[16] Brazil's voting laws are non compulsory for citizens aged 16-17 and above 70. In Chile, enrolment is voluntary.

elections quite effectively. However, it is important to note that this estimate is likely to represent a conservative estimate as it does not count many other types of contests held during the time period under examination, including national and local referenda and initiatives, primaries, or European, regional/state, and local contest. It is, however, the most consistent and reliable cross-national indicator that is available. Data on this variable are from IDEA.

H2: The higher the frequency of elections, the lower the turnout

(3) Closed or Open Party List - CLOSED

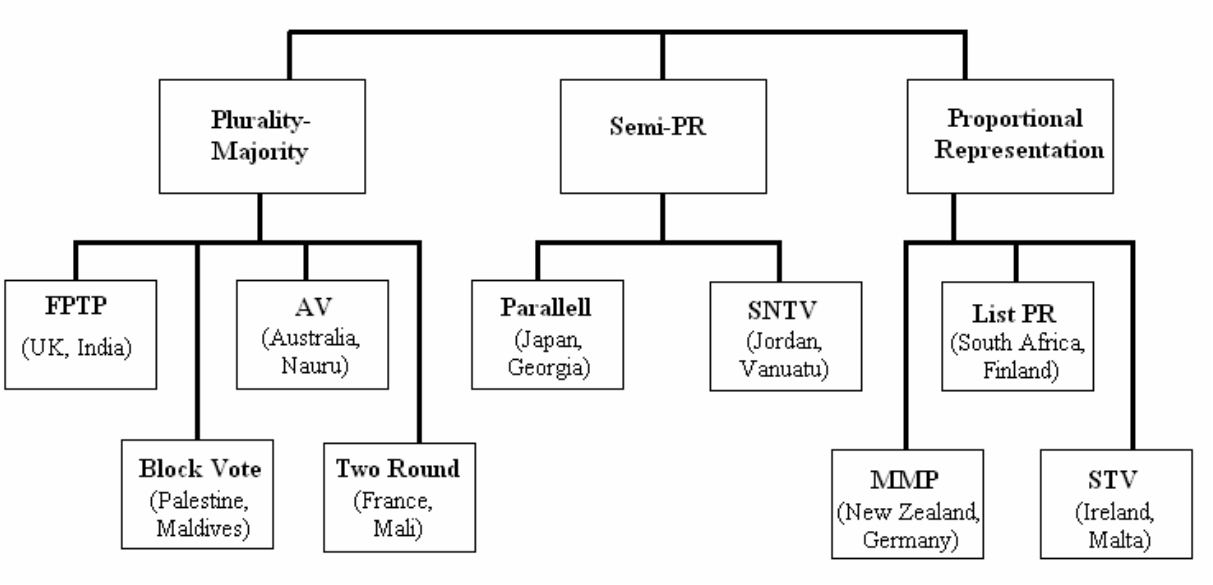
If a party list is closed it means that the voter can not express preferences for candidates within the list. As this limits the influence on the order in which a party's candidates are elected and hence gives the voter less of a choice as of preferences this may give the voter less incentives to show up at the polls. Hence, a depressed turnout is expected where closed lists are in practise. I will include a dummy variable that assumes the value (1) where closed lists are in use and the value (0) where this is not the case. The data for this variable were collected from the Database on Political Institutions (DPI) also known as the Beck database on political institutions.

H3: The presence of closed lists depresses turnout

(4) *Electoral system – PLUMAJ and PR*

I will measure the electoral system in a different way than has been done in the literature up till this point. As the authors of *The International IDEA Handbook of Electoral System Design* (1997) note, there are countless electoral system variations, but essentially they can be split into nine main systems which fall into three broad categories (see figure below). Usually, electoral systems are grouped on the basis of how they translate the electoral vote won into parliamentary seats won. In other words, it is a matter of how proportional the system is. When the proportionality principle, along with some other considerations such as how many members are elected from each district and how many votes the voter has, is taken into account the three main families, along with their members and some examples, are illustrated in Figure 2 below:

Figure 2: *Electoral system families*



Source: *The International IDEA Handbook of Electoral System Design* (1997)

I will, based on this framework, classify my cases along the "family lines" proposed, that is, either as a Plurality-majority-system, as a Semi-PR-system or as a PR-system. I can easily identity the different electoral systems I am interested in (Plurality-majority and PR) by the inclusion of two dummy variables, the one called PLUMAJ, indicating that the electoral system belongs to the Plurality-majority-family, and one called PR, indicating that the electoral system belongs to the Proportional Representation family above. We know that the electoral formula can be expected to affect voter turnout, as citizens are faced with differential rewards under alternative electoral arrangements. I expect a lower turnout in PLUMAJ systems because they are less proportional in their translation of votes into seats, meaning that electors of smaller parties are given less incentives to show up at the polls. I expect an increased turnout in PR-systems because they increase the opportunities for minor parties with dispersed support to enter parliament with only a modest share of the vote, and, therefore this increases the incentives for their supporters to participate.

H4: The presence of Plurality-majority-systems depresses turnout

H5: The presence of PR-systems increases turnout

(5) Unicameralism - UNICAM

Unicameralism means that the parliament consists on one house only. One would therefore expect an increased turnout where unicameralism is present, as it does not have to compete

for power against a possible upper house. However, recall that the findings on this variable were mixed. Whereas some authors report this variable as having a positive effect upon turnout (Jackman 1987), others reported no such effect (Blais and Carty 1990). I will include unicameralism as a control variable in my analysis by creating a dummy variable that is indicating the presence (1) or the absence (0) of unicameralism. My numbers are from the Inter-Parliamentary Union.

H6: The presence of unicameralism increases turnout

(6) *Legislative effectiveness - LEGEFF*

I will include a variable that takes into account the legislature's relative power vis-a-vis the executive power. The Banks Dataset[17] provides data on the relative power of the legislature and assigns each case a score according to this rank:

(0) None. No legislature exists
(1) Ineffective. There are three possible bases for this coding: first, legislative activity may be essentially of a "rubber stamp" character; second, domestic turmoil may make the implementation of legislation impossible; third, the effective executive may prevent the legislature from meeting, or otherwise substantially impede the exercise of its functions.
(2) Partially Effective. A situation in which the effective executives power substantially outweighs, but does not completely dominate that of the legislature.
(3) Effective. The possession of significant governmental autonomy by the legislature, including, typically, substantial authority in regard to taxation and disbursement, and the power to override executive vetoes of legislation.

All of the cases included in this paper can be classified as having a legislature that is either partially effective (2) or effective (3), according to Bank's rank. I can very easily measure the effectiveness of the legislature by the inclusion of a dummy variable. I will include a dummy variable that assumes the value (1) where the legislature is effective and the value (0) where the legislature is partially effective. There are fewer incentives to vote if one knows that the legislature has lesser power when it comes to important subject as, for example, taxation and

[17] Cross-national time-series data archive: Arthur S. Banks: see below for description. I will refer to this dataset as *the Banks Dataset* throughout the paper.

disbursement. The stronger the legislature, the more incentives to vote in parliamentary elections and hence, a higher turnout is expected. The data for this variable are collected from the Banks Dataset.

H7: The more powerful the legislature relative to the executive, the higher the turnout

(7) *Number of effective political parties - NUMPART*

As the discussion above revealed, the findings in the literature are inconsistent regarding this variable. Most authors (Jackman 1987; Blais and Cart 1990; Blais and Dobrzynska 1998) report a negative correlation between the number of parties and turnout whereas others report a positive one (Norris 2004). Again, others report no relationship at all (Perez-Linan 2001; Fornos et al. 2004). Testing the variable in a larger dataset would therefore prove useful in order to establish a final argumentation regarding this variable's effect upon voter turnout. All my numbers for this variable are from ACLP, hence, so is my measurement: number of effective political parties in the legislature measured by utilizing this formula: 1/ (1-F), where F=Party Fractionalization Index as presented in Banks (see Appendix B). Based on the 'inconsistent' direction of this variable's correlation with turnout, I am compelled to create two hypotheses.

H8a: The higher the number of parties, the higher the turnout

H8b: The lower the number of parties, the higher the turnout

(8) *District size – DISTRICT (log)*

The population size of the electoral district can be expected to have an impact on voter turnout. The linkages between voters and their representatives are affected by the number of electors per member of parliament. If a district is small, the possibility for increased information, familiarity and contact between voters and their representatives is present. As Norris (2004) notes: "…..the smaller the number of electors for members of parliament, the greater the potential for constituency service and for elected representatives to maintain communications with local constituents, and, therefore, the higher the incentive to turnout, based on any "personal" vote" (163). I therefore expect a higher turnout in countries with smaller districts. I will measure district size by dividing the number of seats in the lower

house of the parliament into the total population in each country.[18] The variable is expressed through its natural logarithm.

H9: The larger the district size, the lower the turnout

3.2.2.2 Socio-economic variables

Any study on cross-nationally variation in voter turnout should take into account socio-economic variables and the following section identifies which variables I will include in my analyses.

(9) *Economic development – GDP per capita (log)*

Economic development correlates positively with voter turnout (Powell 1980; Blais and Dobrzynska 1998). Economic development in this thesis is indicated by the gross domestic product (GDP) per capita, as defined in *The Penn World Table*. There are three real-GDP-per-capita measures in the data: RGDP, CGDP, and RGDPCH. RGDP is real GDP per capita, based on 1985 price levels. According to Feng (2003) RGDP is suitable for studies that involve relatively "short" time series close to 1985. CGDP is current-year real GDP per capita and is ideal for cross-country, single-year analysis. RGDPCH is real GDP per capita that uses a price chain index with the base year changed from year to year. Of the three, this thesis focuses on RGDPCH, which is adjusted both annually to capture price changes and cross-sectionally to reflect purchasing-power parity. By design, "it is the best indicator of long-run economic growth" (Feng 2003: 4). All my numbers are taken from Penn World Table 5.6 and 6.2 and the variable, which is expressed through its natural logarithm, will simply be called GDP.

H10: The higher the level of economic development, the higher the turnout

(10) *Level of income inequality - GINI*

The level of income inequality can be expressed in a numerous of ways. The most comprehensive summary measure, used to compare changes in one society over time or, more importantly in my case, two or more at a given moment, is known as the GINI Index (or coefficient or ration). The GINI is a number between 0 and 1, with 0 representing a society

[18] This measurement does not technically measure the number of *voters* per member of parliament but as the relationship between population, voters and countries are relatively stable I feel confident that the validity of the measurement is preserved. Norris (2004) utilizes the same procedure (163).

where all members have the same income, and 1 representing perfect inequality, where one member takes all the income. The data for this variable were collected from Human Development Report 2003 and *World Development Indicators* provided by the World Bank. Because data were sparse on this variable my analysis on this variable's impact on turnout is confined to the 1990s model only. We know from Downs (1957) that voting is associated with costs (time, money, energy etc) for the elector. In a society were not everybody is secured at least a minimum of resources (income), less people can "afford" to vote. I therefore expect a decreased turnout in countries that enjoy a more unequal distribution of income.

H11: The more unequal the distribution of income in society, the lower the turnout

(11) *Literacy rate - LITERACY*

Literacy is the ability to read and write and involves "a continuum of learning to enable an individual to achieve his or her goals, to develop his or her knowledge and potential, and to participate fully in the wider society" (UNESCO). Participation in the wider society would also implicate voting at elections and taking part in other political activity. The data for this variable was collected from two sources; the Banks Dataset and UNESCO's Institute for Statistics.[19] The numbers reflect the literacy rate of the population aged 15 years and over, in percentage points. Voting is the least demanding form of political activity and the one that is least dependent on the possession of civic skills. Those with little linguistic skills are however less likely to vote as studies show that a minimum degree of literacy is almost a prerequisite to good turnout (Blais and Dobrzynska 1998). I therefore expect a higher turnout in countries that enjoy a higher literacy rate.

H12: The more literate the population, the higher the turnout

(12) *Human Development Index - HDI*

HDI measures the level of well-being in a country at a given time, taking into account four dimensions; literacy, life expectancy, standard of living and education. The index very effectively measures the level of development a country enjoys by ranking each country on a scale that goes from 0 (least developed) to 1 (most developed). A higher turnout is expected in

[19] As data was difficult to collect for this variable I had to interpolate the data for some missing cases. The logic was this: If, say, Spain had a literacy score of 70 in 1970 and 80 in 1980, then it would receive the score of 70 from 1966 through 1975 and the score of 80 from 1976 through 1985.

more developed countries. The data for this variable were collected from Human Development Report 2006.[20] Due to lack of data on this variable my analysis on this variable's impact on turnout is confined to the 1980s- and 1990s-models only. Norris (2004) argues that the societal changes associated with the modernization process strengthen electoral participation (174). I therefore expect a higher turnout in more developed countries.

H13: The higher the level of development, the higher the turnout

(13) *Population size – POP (log)*

The size of the population correlates negatively with voter turnout (Blais and Carty 1990; Blais and Dobrzynska 1998) as smaller countries are able to arouse a greater sense of community which itself fosters a higher turnout. I therefore expect turnout to be lower in countries that enjoy a large population versus countries that enjoy a small population. I will include a population size variable and call it POP. The variable is expressed through its natural logarithm and my numbers are from the Banks Dataset.

H14: The larger the size of the population, the lesser the turnout

(14) *Population density – POPDEN (log)*

It makes intuitive sense to believe that the more densely populated people are, the easier they are to mobilize, that is, they are easier to find and it takes less resources to inform them. I therefore expect an increased turnout the more densely populated a country is. Population density can easily be measured by dividing the population of a country by the size of the country. As I utilize the same measurement as the Banks Dataset, I operate with square miles as indicator of country size. All my numbers are from the Banks Dataset and the variable is expressed through its natural logarithm.

H14: The more densely populated the population, the higher the turnout

(15) *Catholic - CAT*

I want to see if turnout can be predicted along religious lines. What know that some religions work better along the democratic lines than others[21] and that democratic regimes are more

[20] Same logic about interpolation applies as to the LITERACY variable.
[21] See Haynes (2001) for a discussion on this issue

frequent in countries that can be classified as belonging to the Christian cultural heritage. My dataset consists of 90 countries of which about 95% belong to this heritage. Of these countries, the majority are either of a catholic of protestant orientation. To identify the catholic orientation, I simply employ a dummy variable that is assigned the value (1) in cases where the catholic population exceeds 50% of the total population and (0) otherwise. All my numbers for this variable are from ACLP. Since I have no clear cut expectation regarding the direction of this variable's correlation, I am compelled to create two competing hypotheses.

H15a: An increased turnout can be expected in countries that enjoy a catholic majority
H15b: A decreased turnout can be expected in countries that enjoy a catholic majority

3.2.2.3 Information circulation variables

(16) *Information circulation – RADIOS, TVS, NEWSPAPER and INFO (all logged)*

I will include a variable that takes into account the flow of information in society. It makes good intuitive sense to believe that the higher the flow of information, the greater the turnout at elections. It is difficult to measure this flow precisely, but by taking into account the means by which this information flows, that is, the media, it is possible to get an impression of the phenomenon. I will therefore include four media variables: RADIOS, TVS, NEWSPAPER and INFO. I will measure the three former variables as the Banks Dataset does: radios, TVs and newspaper circulation per capita. The INFO variable is a compilation of the three different variables, meaning that the value of the three are summed and then divided by three. All my numbers are from the Banks Dataset and the variables are expressed through their respective natural logarithm. I expect an increased level of turnout in countries that enjoy an increased level of either of these media variables. As no study previously has included variables that take into account the flow of information in society it will be very interesting to check its explanatory power on cross-national differences in voter turnout.

H16: The higher the flow of information, the higher the turnout

3.2.2.4 Activism variables

I want to include variables that measure the level of political mobilization and activism in society. As a result, I have created four variables that measure the phenomenon in different ways.

(17) *Number of peaceful demonstrations - DEMO*

The first 'activism' variable deals with peaceful demonstrations. Banks Dataset provides data on the number of peaceful demonstrations in a country in a given year. As these demonstrations are peaceful, it makes intuitive sense to believe that the higher the number of peaceful demonstrations, the more political awareness and mobilisation the country's citizens enjoy. It is therefore possible to expect a higher turnout where these demonstrations are more frequent. However, an increased number of demonstrations could also mean that a country's citizens are frustrated and feel that the best way to influence governmental policies are not going through the polls. I will therefore be aware of this variable's possibility to swing both ways. The variable is measured as in the Banks Dataset: any peaceful public gathering of at least 100 people for the primary purpose of displaying or voicing their opposition to government policies or authority, excluding demonstrations of a distinctly anti-foreign nature.

H17a: The higher the number of demonstrations, the higher the turnout

H17b: The higher the number of demonstrations, the lower the turnout

(18) *Number of riots - RIOTS*

The second 'activism' variable, RIOTS, is included to measure the level of *conflict*. The reason is this: if the level of conflict is high, it could be because going to the polls is not regarded to be the most "meaningful" way to express one's political beliefs. I hence expect a decreased turnout where the number of riots is higher. All my numbers are from the Banks Dataset and indicate the number of riots in a country in a given year. The variable is measured as in the Banks Dataset: any violent demonstration or clash of more than 100 citizens involving the use of physical force.

H18: The higher the number of riots, the lower the turnout

(19) *Governmental crisis - GOVCRIS*

The third variable in this block, GOVCRIS, deals with the state of the government/regime. The more unstable the regime, a decreased turnout is expected on my part. Again, I expect people to discredit the polls as a way of influencing policy formation because of the unstable political climate. My measurement follows Banks: Any rapidly developing situation that threatens to bring the downfall of the present regime - excluding situations of revolt aimed at such

overthrow. My numbers are also from the Banks Dataset and indicate the number of governmental crises in a country in a given year.

H19: The higher the number of governmental crises, the lower the turnout

(20) *Weighted Conflict Index – WCI (log)*

The last variable in the activism-block, WCI, is included for the same reasons as the RIOTS- and the GOVCRIS-variable, differing only in the complexity of the measurement.[22]

In and of themselves, the different factors constituting the WCI could have an effect upon turnout, in a negative direction. I intend, as shown above, to include three of these factors, namely DEMO, RIOTS and GOVCRIS. I expect that the higher the score on the weighted conflict index, the lower the turnout. Again, my numbers are from the Banks Dataset and indicate the score a country is assigned in a given year, expressed through its natural logarithm.

H20: The higher the level of conflict, the lower the turnout

3.2.2.5 Political legacy

(21) *Eastern Europe - EASTERN*

I have also included a dummy variable that indicates whether or not the country under analysis is a former Soviet Union satellite. I have called this variable EASTERN, as most of the former Soviet satellites identified in my dataset are found in Eastern Europe. I wish to see if the historical legacy of former authoritarian rule has had an impact on voter turnout. We know from the landmark book by Almond and Verba, *The Civic Culture* (1963) that people, given the historical political legacy of their country, vary in terms of their political participation. Almond and Verba distinguish between *participants, subjects* and *parochials*. My interest lies in the subjects and parochials, the former identified as a person having experienced a regime of an authoritarian character, and the latter being identified by a person having experienced one-party rule for a long time. As a result, they are both characterised by a low level of political participation because they could see politics as a dangerous thing to get involved in. Based on this, I expect that the countries associated by former Soviet rule, are characterized

[22] The weighted conflict index is calculated (by Banks) in the following manner: Multiply the value of the number of Assassinations by 24, General Strikes by 43, Guerrilla Warfare by 46, Government Crises by 48, Purges by 86, Riots by 102, Revolutions by 148, Anti-Government Demonstrations by 200. Sum the 8 weighted values and divide by 9. See Appendix B for a description of the different factors.

by a lower turnout as they have both experienced long term relationship with one-party rule and regimes of an authoritarian character.

H21: *Soviet-type authoritarian political legacy decreases turnout*

3.2.2.5 Summary of variables

All the variables discussed above, have been included in the table below. The table provides an overview of the different variables and their expected correlation with turnout.

Table 2: *Hypothesized effect of each independent variable on turnout*

Variable	Hyp. relation	Variable	Hyp. relation
Institutional		**Information circulation**	
COMP	+	RADIOS	+
FREQ	–	TVS	+
CLOSED	–	NEWSPAPER	+
PLUMAJ	–	INFO	+
PR	+		
UNICAM	+	**Activism**	
LEGEFF	+	DEMO	+/–
NUMPART	+/–	RIOTS	+/–
DISTRICT	–	GOVCRIS	+/–
		WCI	+/–
Socio-economic			
GDP	+	**Political legacy**	
POP	–	EASTERN	–
POPDENS	+		
GINI	+		
LITERACY	+		
HDI	+		
CAT	+/–		

3.3 Quality of measurement

The major criteria to judge the quality of measurements are *validity* and *reliability*. The validity of measurement is defined as the degree to which one actually measures whatever concept the measurement procedure purports to measure. It refers to the closeness of the correspondence between the measurements and the concept being measured (Pennings et al. 2006). Regarding the validity of the measurements above, I feel confident that the variables that I have defined are measuring the phenomenon/concept that I am interested in analysing. Keep in mind that the majority of the variables included in this paper are replicates of

variables included in previous research (see chapter 2). I am, however, aware of the possibility that perhaps some of my measurements represent a conservative estimate (FREQ) or the opposite (DISTRICT). The validity however, is preserved.

As for reliability, measurements are reliable to the extent that measurements with respect to the same units deliver consistent results. The measurements are reliable if they are trustworthy, that is, they measure what they are supposed to measure. I will address the issue of reliability by briefly describing the data sources used in the collection of the data for this paper. All the data sources below are well known and regarded as highly reliable by social scientists studying voting behaviour in general and voter turnout in special.[23]

I turn now to a presentation of my methodological approach and research strategy.

[23] The ACLP dataset is a culmination of ambitious efforts to collect a vide variety of variables put forward in the landmark book *Democracy and Development: Political Institutions and Well-Being in the World 1950-1990*. The dataset has been updated to 2002, making it suitable for my purposes. The unit of analysis in the ACLP dataset is a given country in a given year, and the over 100 variables included represents one of the most comprehensive sources of post-World War II indicators for "large-N" cross-sectional time-series studies. The Banks Dataset, also known as The Cross-National Time-Series Data Archive, abbreviated CNTS, provide a wide range of variables and excellent coverage. It was a product of the State University of New York, launched in the fall of 1968 by Arthur S. Banks (hence the name). The numbers I use from the Banks dataset come from the 2004 version. IDEA provides the most comprehensive record of turnout in national elections that is currently available. They cover 1620 national-level elections held from 1945 to 2000 in all 193 independent nation-states, including 1218 parliamentary and 402 presidential elections. The Inter-Parliamentary union is an international organization of parliaments of sovereign states. They provide comprehensive data on political variables through their PARLINE database (a derivative of Parliaments on-line). UNESCO is a specialized agency of the United Nations, established in 1945. The UNESCO Institute for Statistics regularly updates and disseminates literacy estimates and projections for more than 100 countries. My numbers are from 2000. The Human Development Report (HDR) is an annual milestone publication from the United Nation Development Programme. HDR is an independent report and is commissioned by the United Nations Development Programme (UNDP). It is the product of a selected team of leading scholars, development practitioners and members of the Human Development Report Office of UNDP. In this paper I used the HDR 2003 and 2006. The World Bank also provides annual publications on development, called World Development Indicators. As of 2006, it covers more than 150 economies over 900 indicators. Regarding political variables, the World Bank provides a large cross-country database on political institutions: the Database on Political Institutions (DPI). As one of the creators of this dataset is named Thorstein Beck, the dataset is also known as Beck database. Finally, the Penn World Table provides purchasing power parity and national income accounts converted to international prices for 188 countries for some or all of the years 1950-2004. In this paper I have utilized the Penn World Table 5.6 and 6.2.

IV. Methodological approach and research strategy

In this chapter I will outline my methodological approach and research strategy. By utilizing the quantitative method, specifically the regression analysis of cross-sectional data, I will test my research hypotheses. I begin with a presentation of my methodological approach.

4.1 Methodological approach

4.1.1 Quantitative method

In social science, a variety of approaches can be taken to test hypotheses. These approaches have traditionally been divided into two research methods, the quantitative (variable-oriented) and the qualitative (case-oriented), where the former generally refers to studies of a large number of units and the latter referring to detailed analysis of one or a few cases. The different methods are used to answer different questions and vary accordingly in their ability to produce findings upon which one can make generalizations: the quantitative method gives broader and more general answers based on a larger number of cases, suitable for generalizations, while the qualitative method focuses its attention on a narrower scale, answering questions on a small number of cases (Ragin 1987; Ragin and Zaret 1983; Ringdal 2001; Skog 2004).

The goal of most comparative social science is to produce explanations of macro-social phenomenon that are general but also show an appreciation of complexity (Ragin 1987). As we recognize that a good social scientific explanation is relevant to a variety of cases, and, at the same time, take into account the complexity of social phenomenon we realize that a general explanation is a partial explanation at best. Generality and complexity compete, and an appreciation of complexity sacrifices generality; an emphasis on generality encourages a neglect of complexity. As Ragin put it: “It is difficult to have both” (1987: 54). In the study of cross-national variation of voter turnout, generality is given precedence over complexity because it is a variable-oriented study. By specifying the hypotheses to be tested and then delineating the widest possible population of relevant observations, it is possible to generalize upon findings. The wider this population the better, because a wide population provides a basis for a more exacting test and, in addition, gives the investigator the opportunity to demonstrate the breadth of an argument (Ragin 1987: 55). In my study I include, based on the indices proposed by ACLP, 90 countries. Recall that the studies I reviewed in Chapter 2

(studies which represent the tradition found within this field of research) operate with a number of cases ranging from 19 to 32.[24] I find it problematic that the same studies, given their low number of cases, generalize upon their findings. As my study includes 90 countries (96 cases), it represents a major improvement over previous research as it allows me to generalize more confidently, based upon my findings.

Since the emphasis of the quantitative strategy is on general features of social structure (variables), it has allowed importations of quantitative methods, especially multiple regression techniques, from mainstream social science. On the basis of relatively small data sets, these techniques allow investigators to make broad statements about cross-societal patterns, such as voter turnout variations. Thus, in the study of cross-national variation in voter turnout, quantitative methods are utilized, and the dominant approach is to use a multiple regression technique, namely *Cross Sectional Analysis* (CS).

I proceed next to a brief description of the technique at hand.

4.1.2 Multiple regression analysis

In a multiple regression model, more than one independent variable is assumed to have an impact on the dependent variable. By incorporating more than one independent variable into the equation, we accomplish two things. First, since few phenomenona are products of a single cause, a multivariate model offers a fuller explanation of the dependent variable. Second, the effect of a particular independent variable is made more certain, for the possibility of distorting influences from the other independent variables is removed (Lewis-Beck 1980: 47). The value for case *i* on the dependent variable, Y_i, is assumed to be a linear combination of the values of case *i* on the independent variables, X_{1i}, X_{2i}, … , X_{ki}, except for a *residual*, ε_i, that is not accounted for by the independent variables (Pennings et al 2006: 153). The residual for case *i* is the difference between the value of the dependent variable, Y_i, and the predicted value, $\hat{Y}_i$:

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \Lambda + b_k X_{ki} + \varepsilon_i$$

[24] Blais and Dobrzynska (1998) being the exception. See Appendix D for an overview.

In the case of *k* independent variables, X_1, X_2, …, X_k the predicted value of case *i*, $\hat{Y}_i$, is a linear function of X_{1i}, X_{2i}, … , X_{ki} multiplied by their respective regression slope coefficients b_1, b_2, … , b_k, and the regression constant b_0

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \Lambda + b_k X_{ki}$$

The *Ordinary Least Squares* (OLS) method to estimate the regression coefficients b_0, b_1, b_2, … , b_k prescribes that the sum of all squared residuals, ε_i^2, also denoted as SSR, should be minimized:

$$\text{Minimize SSR, where SSR} = \sum_i (Y_i - Y_i)^2 \qquad \text{(Pennings et al 2006).}$$

In other words: the central goal in the OLS-method is to find the regression line that best fits the sample data. The OLS-method, which I will use in this paper, is a widely used technique within linear regression analysis and has become, due to the possession of some attractive statistical properties, one of the most powerful and popular methods of regression analysis. It rests, however, on certain assumptions (see Appendix C). Violating one of more of the assumptions can seriously harm the accuracy of the OLS-estimates. Some of these assumptions, like the absence of perfect multicollinearity and the absence of heteroscedasticity, are more important than others. In the discussion of this paper's analyses, in chapter 5 and 6, I will specifically discuss these assumptions and explain how they are dealt with in a manner which minimizes the bias and maximizes the efficiency of the estimated regression coefficients.

4.1.3 The 'Krieckhaus approach'

As indicated above, the dominant approach in cross-national voter turnout research has been a Cross Sectional Analysis of variation. This approach is appropriate for sorting out the effects of variables that tend to be stable over time, such as the socio-economic environment, the electoral system, compulsory voting and other institutional factors (Blais 2006:121). Two recent articles (Krieckhaus 2004, 2005) which survey and analyze the recent literature on the effects of political regimes upon economic growth prove to be very useful for the purposes of my interest in cross-sectional variation in voter turnout. Krieckhaus shows how the effects of regimes upon growth vary across time. Therefore, the general cross-sectional findings may in fact vary and not be constant across time. Following the insights of Krieckhaus, I propose to

make an additional contribution to our knowledge of the determinants of voter turnout in the following manner. In order to be sensitive to variation of variable effects across time, it is necessary to conduct a series of cross-sectional analyses. If I want the robustness of the results across time to come forward I will have to examine shorter time periods. To allow comparison across periods, I intentionally examine each cross-section separately rather than employ a pooled analysis. By averaging the data on each variable for, say, every decade, it is possible to employ cross-sectional analysis quite effectively. This approach will allow me to compare the decade-specific results to each other as well as to the results based upon the entire time period.

4.2 Research strategy

4.2.1 Research strategy outlined

As part of the preliminary analysis for this paper's research, I will start out by focusing only on those explanatory variables that have been studied in previous research, as identified in my literature review above. It will be interesting to test the robustness of these variables from the literature when included in regression models based on a dataset that covers many cases across a long time period. For this purpose I will utilize a dataset that I have compiled, called "Grand Dataset" (see below). Recall that a number of the findings reported in the literature review chapter have come about as a result of regression analyses conducted across relatively few cases (19-32) over a relatively short time period (10-20 years).[25] Only when included in regression models conducted on the basis of a dataset that incorporates as many democratic elections as possible can the true explanatory power of the variables come forward.

The next step in my research strategy will be to go on to build a "grand model" that explains cross-national variation in voter turnout in general. The regression analyses for this model will be conducted on the basis of the same dataset as the literature model above, namely the Grand Dataset. In addition to variables identified in the literature review chapter, I intend to include some new variables which have never before been included in research on cross-national variation in turnout. The new variables that I have prepared for this thesis were discussed above and include: LEGEFF, POPDENS, GINI, CAT, EASTERN, RADIOS, TVS, NEWSPAPER, INFO, DEMOS, RIOTS, GOVCRIS and WCI. I will begin the analysis by running

[25] See Appendix D for an overview of samples and time period of the different studies.

simple bivariate regression models for each independent variable and present the variables that have effects that are the most statistically significant. I can thus generate some preliminary assessments of the variables' effects and statistical significance.

Based on these preliminary results, I will go on to build a general model. I will do so by identifying a core of central variables to which I will add one or two variables at a time. As the model continues to expand, I will continue to add additional variables to the model specification and test their significance in the larger multivariate model. This process will continue until the final model is complete. By running regression models on the Grand Dataset I will determine the "average" effects of each independent variable across the entire population of countries for which data exist. (These "aggregate" or "overall" effects will later be compared to the decade-specific results.)

Thirdly, I will check for time-specific variance in the explanatory variables on my dependent variable, by running separate cross-sections for each decade. By averaging every variable for each decade, and running all countries, it is possible to see how the effect of each variable varies across time. The time-specific analyses will be called Decade Models, and conducted on the basis of a dataset I have compiled specifically for this purpose. The dataset, which is discussed below, will be called Decade Dataset. I will first run bivariate regression models on each variable, across each decade to see if any patterns emerge. The bivariate regressions will serve the purpose as preliminary assessment for the variables, and those which turn out to be significant will be included in multivariate regression models across each decade. Only by including the variables in multivariate regression models can the effect of each individual variable be determined.[26]

Only by checking the robustness of the explanatory variables across time can we be certain of their importance for explaining cross-national variation in voter turnout. By conducting a comprehensive research strategy like this, it is possible to test our hypotheses more thoroughly, and make a unique contribution to our knowledge of why countries differ in voter turnout.

[26] In my analysis of the 1990s decade I will extend my analysis to building a final, "best" model. This will allow me to fully exploit the potential for studying the impact of additional independent variables for which data exist only for the 1990s.

4.2.2 Datasets

In the preparations for this thesis I was compelled to create two datasets, one dealing with the grand cross-sectional analysis, called 'Grand Dataset', and one dealing with the time-specific analysis, called 'Decade Dataset'. There are some points worth mentioning regarding the creation of the datasets, and I begin with the Grand Dataset.

In order for the analyses to be sensitive to institutional changes in the cases under review, some of the cases had to be divided into sub-cases. In other words, some countries experienced fundamental institutional changes over the time period covered and I had to account for that in the creation of the dataset. For example, Sweden changed from a bicameral to a unicameral legislature in 1970 meaning that for Sweden I had to create two sub-cases, Sweden I and Sweden II, the former referring to Sweden when it had bicameralism (1950-1970) and the latter to when it had unicameralism (1970-2000). New Zealand is another example, as it changed its electoral system in 1993. I created two sub-cases, New Zealand I, referring to the Plurality/majority period (1950-1993), and New Zealand II, referring to the Proportional Representation period (1993-2000). Four countries also changed their mandatory voting laws during the time period under analysis. These include: Chile (changed from non-compulsory to compulsory in 1962), Uruguay (changed from non-compulsory to compulsory in 1970), Venezuela (changed from compulsory to non-compulsory in 1993), and the Netherlands (changed from compulsory to non-compulsory in 1970). This implied that sub-cases had to be created taking into account the presence or absence of mandatory voting laws for the different countries. By creating the sub-cases for Sweden, New Zealand, Chile, Uruguay, Venezuela and the Netherlands, my cross-sectional analysis is sensitive to the institutional changes, and hence, the results are more robust. This also implies that the number of cases under comparison is 96, even though only 90 countries are included in the analysis.

As noted, I have also created a dataset for the time-specific regression analyses called Decade Dataset. By averaging each variable for each decade across each country, I can run regressions on my data that are time sensitive. The dataset is sensitive to institutional changes, as the dataset above, because of the timing of the institutional changes across my countries. Sweden, Uruguay and the Netherlands all experienced institutional changes in 1970s (see above). Hence Sweden was bicameral in the decades of 1950 and 1960, and unicameral in the following three decades; the same logic applies to Uruguay and the Netherlands regarding

their compulsory voting laws. Chile changed from non-compulsory to compulsory in 1962. I therefore classified Chile as non-compulsory in the 1950s decade and compulsory from the 1960s decade and onwards. Finally, New Zealand and Venezuela experienced institutional changes in 1993, as New Zealand changed their electoral system and Venezuela abolished compulsory voting. The same logic as above applies to these two cases making their 1990s decades differ from the other decades on the respective institutions.

4.2.3 Software

As for statistical software, I will use Limdep version 8.0, which is an integrated program for estimation and analysis of linear and nonlinear models, with cross-section, time series and panel data. It has interesting properties, like for instance the possibility of running the regression models implementing White's heteroscedasticity-consistent variances and standard errors. Because the hypothesized effects of a number of the variables included in this paper are uncertain regarding the direction of their correlation, all of my regression models are two-tailed. Furthermore, unless otherwise explicitly mentioned, whenever I refer to a variable being significant or not, I mean at the 0,05 level (significant at a 95% confidence interval).

I turn now to the analyses, beginning with the building of a general model which pertains to explain why countries vary in voter turnout. By utilizing the Grand dataset I will engage in an extensive model building procedure which will result in a model that best explain cross-national variation in voter turnout between 1950 and 2000.

V. Aggregate empirical patterns

The purpose of this paper is to explain cross-national variation in voter turnout. As we saw from the literature review chapter above, the different variables were associated with different effects and significance across the different studies. We do know that some variables matter more than others, like for instance compulsory voting and economic development. The problem is, however, that the different studies associated with the research question differ in their sample selection, number of cases, time period included and level of sophistication. This thesis addresses those shortcomings by expanding the number of cases and extending the time period covered. By creating a dataset that covers more countries over more years than any previous study, and by including variables that have not yet been included in previous models, I hope to make a unique contribution to our knowledge of why we see cross country differences in voter turnout.

I will start out, as elaborated above, by running multiple regression models on the Grand Dataset. The purpose of these regression models will be to detect any general patterns in the determinants of cross-national variation in voter turnout between 1950 and 2000. Before I turn to the general model building procedure using both variables from the literature review chapter and variables that are unique to this study, I will run some preliminary analyses, including the 'Literature model' and bivariate regression models. But first, we will have a look at the descriptive statistics, the presence of outliers and the issue of collinearity.

5.1 Descriptive statistics

The operationalization of the dependent variable in this study, electoral turnout, is calculated as the percentage of the voting age population (VAP) that cast a vote. As we see from Table 3 below, our dependent variable, TURNOUT, goes from a minimum value of about 21 to a maximum value of about 94 percent. The average turnout for the sample is 67 percent. The minimum value of 21 percent turnout reflects the case of Mali. Given the low level of turnout, Mali will be regarded as an outlier in this dataset (discussed below). Seven of my variables are dummy variables and another eight are transformed by taking their natural logarithm. The descriptive statistics for the variables from the Grand Dataset are presented in Table 3.

Table 3: *Descriptive statistics – all variables in Grand Dataset*

Variables	Mean	St.dev.	Min.	Max.	Cases
TURNOUT	67.9292970	15.3504818	21.6100000	93.9500000	96
COMP	.145833333	.354791760	.000000000	1.00000000	96
FREQ	3.99471242	1.62862696	1.66666667	14.0000000	96
CLOSED	.611940299	.490986141	.000000000	1.00000000	67
PLUMAJ	.312500000	.465945558	.000000000	1.00000000	96
PR	.541666667	.500876425	.000000000	1.00000000	96
UNICAM	.541666667	.500876425	.000000000	1.00000000	96
LEGEFF	.552083333	.499890339	.000000000	1.00000000	96
NUMPART	3.69659473	2.70693783	1.00000000	19.7286350	94
DISTRICT*	86.2034318	159.485550	2.21500000	1298.00000	91
DISTRICT(log)	3.68234035	1.25581024	.795252403	7.16857990	91
GDP	8350.72844	5359.14003	793.555950	21553.6682	94
GDP(log)	8.76951167	.800674764	6.67652405	9.97830130	94
POP*	27127.4734	77693.7883	30.5000000	693331.292	95
POP(log)	8.44114723	2.18725954	3.41772668	13.4492632	95
POPDEN	3278.57398	4391.54879	44.8000000	28647.0000	95
POPDEN(log)	7.27378215	1.46279532	3.80220814	10.2628040	95
LITERACY	85.4991207	19.0274550	20.4000000	99.8000000	83
CAT	.427083333	.497251216	.000000000	1.00000000	96
RADIOS**	4401.21395	2752.91660	384.000000	15823.7200	94
RADIOS(log)	8.13766028	.797691637	5.95064255	9.66926536	94
TVS***	21611.6449	15471.0757	12.6666667	54509.0000	94
TVS(log)	9.36736498	1.64894513	2.53897387	10.9061211	94
NEWSPAPER**	1705.12466	1431.44510	10.0000000	5541.50000	83
NEWSPAPER(log)	6.92497463	1.25377698	2.30258509	8.62002050	83
INFO	3188.45662	2198.21737	96.8333333	8825.41667	95
INFO(log)	7.66566695	1.10463514	4.57299129	9.08539110	95
DEMO	.604192145	1.08300417	.000000000	5.92000000	95
RIOTS	.537093889	1.09129802	.000000000	7.08333333	95
GOVCRIS	.278455573	.398898099	.000000000	2.08333333	95
WCI	2903.63708	4379.28435	.000000000	20741.0000	95
WCI(log)	5.61658645	3.49693948	.000000000	9.93986770	95
EASTERN	.145833333	.354791760	.000000000	1.00000000	96

* scaling 1000
** scaling 0.0001
*** scaling 0.00001

5.1.1 Outliers

An *outlier* is a single case which affects the slope of the complete regression line disproportionately (Pennings et al 2006). By inspecting the partial regression plots I detected, beside Mali at 21 percentage points turnout, another two outliers, namely, Chile I (Chile pre-compulsory 1950-62) at 25 percent turnout and Guatemala at 27 percent turnout. There are different ways to deal with outliers. Some argue in favour of dropping the cases at hand

completely from the model. A technical alternative to dropping the outliers completely, which is the approach I will follow, is to create a dummy variable for each outlier, coded 1 for the outlier case, 0 for all other cases. The slope coefficient for such a dummy variable indicates the difference between the value of the outlier on the dependent variable and the prediction for the outlier produced by the regression equation based on the remaining cases (Pennings et al. 2006). It indicates precisely how different the outlier is from the remaining cases, but it can not tell the story as to why this is so. (Dummy variables indicating different outliers will be included in all of the regression models presented in this thesis.)

5.1.2 Collinearity

An important assumption in multivariate regression analysis is that the effect of each of the independent variables, x, on the dependent variable, y, should be independent of the effects of the other independent variables. High correlation between two or more of the independent variables indicates that there is a problem of collinearity in the data which makes the estimation of each independent variable's effect upon the dependent variable difficult. By conducting a correlation analysis, resulting in a correlation matrix (available in Appendix E), the degree of association between the variables is measured. Several of the pair-wise correlations are quite high, suggesting that there may be a collinearity problem. As a rule of thumb, if two variables show a correlation coefficient of 0,7 or higher they should not be included in the same model. However, even if no high values are found, one can not yet conclude that multicollinearity is not a problem in one's data. When specifying my final model for explaining cross-national variation in voter turnout below, more sophisticated methods of checking for multicollinearity (and heteroscedasticity) will be conducted. For now, it is sufficient to look at the correlation matrix, as a guide to which variables *not* to include in the same regression model. A close inspection of the correlation matrix reveals that a number of variables should not be included in the same model. They are presented in Table 4.

Table 4: *Collinear variables in Grand Dataset*

NEWSPAPER-INFO	(r = .85)	INFO-RADIO	(r = .83)	GDP-TV	(r = .72)
NEWSPAPER-RADIO	(r = .77)	INFO-TV	(r = .88)	GDP-LITERACY	(r = .75)
NEWSPAPER-TV	(r = .75)	INFO-LITERACY	(r = .85)	POP-DISTRICT	(r = .92)
NEWSPAPER-LITERACY	(r = .84)	INFO-GDP	(r = .82)		
NEWSPAPER-GDP	(r = .82)	RADIO-LITERACY	(r = .83)		

Based on this, I know which independent variables not to include in the same regression model. As a consequence I am compelled, when conducting my analyses below, to run separate analyses for the variables that are correlated.

5.2 Preliminary assessment

First of all, I will test the findings presented in chapter 2, the literature review chapter. They will be included in regression analyses, called "literature models," conducted on the basis of my Grand Dataset. As this dataset consists of a large number of cases, over a long time period, it will be interesting to see whether or not the different findings from Chapter 2 turn out to be robust.

5.2.1 'Literature model'

These analyses will provide some preliminary assessments of how significant the variables from the literature review chapter are. I am however, aware of the likelihood that any of these variables can change effect and significance when I include them in a model containing a wider range of control variables.

Since the different studies in the review chapter deal with different approaches, alongside the likelihood of multicollinearity, the different variables from the literature have never before been included in the same model. Recall from above, that some of the independent variables correlate too much to be included in the same regression model, including GDP-LITERACY ($r = 0,75$) and POP-DISTRICT ($r = 0,92$). I am therefore compelled to run a number of regression models to see whether or not the variables from the literature are significant in a dataset that contains more cases over more years. Table 5 below presents the hypothesized direction of effect of each variable as well as the results of my estimations, displaying the "best showing" of each of the respective "literature" variables. It is important to note that the variables showed consistency regarding direction, effect and significance across the different models. Variables that were significant have t-statistics presented in bold.

Table 5: *Variables from the literature, best showing*[27]

Variables	Hyp. effect	b	T-statistic
COMP	+	10.1	**2.8**
PR	+	2.3	0.9
UNICAM	+/none	1.8	0.7
FREQ	–	–1.6	**–2.1**
NUMPART	–/+	–0.3	–0.7
DISTRICT(log)	–	–3.0	**–2.3**
GDP(log)	+	6.0	**2.8**
LITERACY	+	0.2	1.4
POP(log)	–	–1.2	–1.5

5.2.1.1 Brief discussion of findings

All of the variables, except FREQ, show expected direction in its correlation with electoral turnout. Only four of the ten variables included in the regressions were statistically significant at a 95% confidence interval level.[28]

Not surprisingly, COMP had a positive and significant effect with an expected increase, all else held equal, of about 10 percentage points on voter turnout, when present. Recall from the literature review chapter that an increase ranging from 7 percentage points (Franklin 1996) to 13 points (Jackman 1987) are expected when compulsory voting is present so my finding is consistent with the findings in the literature. My finding on this variable is also highly significant with a t-statistic score of almost 3. As with PR, we see that an election taking place under PR-systems are expected to enjoy, all else held equal, an increased turnout of 2,3 percentage points. This finding is consistent with Blais and Dobrzynska's (1998) finding but the effect is not statistically significant in my regression. Unicameralism also have a positive effect upon turnout according to the literature (Jackman 1987), even though some report no correlation at all. My finding is consistent with Jackman (1987); all else held equal, an increase of 2 percentage points is expected with the presence of unicameralism. The effect however, is not statistically significant. The frequency of elections is expected to influence turnout in a negative direction (Norris 2004). However, my finding on this variable does not

[27] In every regression, MALI, GUATEMALA and CHILE I were included as dummy variables.

[28] All the regressions in this paper are two-tailed

support this assumption. Recall that I measure FREQ as the number of years since the last national election. Hence, if the impact of this variable is to affect turnout in a negative direction, the coefficient ought to have a positive sign. To the contrary, it has a negative sign meaning that there will be an increasingly negative effect on turnout the more years since the last election. This is quite interesting. What my numbers indicate is that each year since the last election is, all else held equal, worth 1,6 percentage points of decreased turnout. The finding is highly significant, displaying a t-statistic score of 2,1.

The findings in the literature regarding the NUMPART variable were mixed. My finding supports Jackman (1987) and Blais and Dobrzynska (1998) which report a negative correlation between the number of parties and turnout. The effect, which is not statistically significant, is worth about 1 percentage point decrease in turnout for every 3 new parties present in the legislature, given that all else is held equal. DISTRICT also affects turnout negatively, as expected from Norris (2004). The relationship is logarithmic, implying that the main difference is at the lower end of the district size: all else held equal, turnout decreases by 10,8 percentage points when the district size goes from 2221 voters per member of parliament (Andorra), the lowest in the sample, to 86 200 voters per member of parliament (about the size of Italy), which represents the average. The decrease in turnout is at 2,85 percentage points, all else held equal, when district size goes from the average to the highest, 1,3 million voters per member of parliament (India). The finding is statistically significant with a t-statistic score of 2,3.

The level of economic development, measured by the GDP variable, correlates positively with electoral turnout. This relationship is also logarithmic which again implies that the main difference is at the low end of economic development: everything else being equal, turnout increases by 1,9 points when GDP per capita moves from 793 American dollars (Central African Republic), the lowest in the sample, to the average, 8350 American dollars (about the GDP level of Slovakia), but only 0,8 points when it moves from the average to the highest, 21554 (Luxembourg). The effect, which is statistically significant, is very modest compared to the findings reported in the literature chapter above. According to Blais and Dobrzynska (1998), the effect is over 10 percentage point on turnout when one moves from the lowest level of economic development to the average level in their sample. LITERACY also correlates positively with turnout, as expected from the literature. For every extra percentage points in

increased level of literacy we see an increased level of turnout, by 0,2 percentage points. Again, we note that the effect, which is not statistically significant, is smaller than the effect one would expect from the literature findings.

The last variable to be included from the literature chapter is the size of the population, measured through the POP variable. It correlates, as expected, negatively with turnout. As with DISTRICT and GDP, the relationship is logarithmic. All else held equal, turnout decreases by 1,8 points when the population size moves from 30 500 (Lichtenstein), the lowest in the sample, to the average (27 127 000), but only 0,7 points when moving from the average to the highest, 693 331 000 (India). Yet again, the effect is more moderate than one would expect from the literature review. Blais and Dobrzynska (1998) report a finding of 7 points on turnout when moving from the low to the average in population size in their sample.

The robustness of the findings surveyed in Chapter 2 is strengthened as a result of the literature model-regressions. By running multiple regressions on my dataset I was able to test the variables and can determine their relative importance in explaining the cross-national variation in voter turnout. Every variable, except FREQ, showed expected direction in their correlation with voter turnout. The downside is of course the lack of significance that most of the variables show. I would have expected that more than four variables (COMP, FREQ, DISTRICT and GDP) were significant at a 95 percent confidence interval, the level at which my own analysis below will be conducted. Furthermore, I am struck by the modest effect each variable has upon turnout, given the positive impression they gave when they were discussed above. I will leave the discussion at that, turning next to a more thorough testing of the different variables. If I wish to test the different variables' explanatory power on voter turnout differences, I will have to build a new model, from scratch. By including the variables from the literature, alongside the new variables I have created in the preparations for these analyses, I can thoroughly analyze and test which variables have an impact on cross-national variation in voter turnout. It is to this model building I now shall turn.

5.3 Cross-national variation in voter turnout – building a model

Based on the above I now have a good platform from which I can build a model that best explains cross-national variation in voter turnout between 1950 and 2000. I will begin by

running simple bivariate regression models between my dependent variable and the independent variables.

5.3.1 A second preliminary assessment: Bivariate regressions

A feeling for the data precedes any serious data analysis (Pennings et al 2006). I will therefore run simple bivariate regressions between the dependent variable and the different independent variables. The bivariate regressions will suggest which Xs are correlated with Y and if this correlation is in the expected direction. As I have 22 independent variables, amounting to 22 bivariate regressions, I will only present the variables that have the highest significance in their correlation with turnout. A complete list of every variable's bivariate regression is available in Appendix G. It is important to keep in mind that the effects of the different variables are likely to change when the variables are included in a multivariate regression model. The variables that display significant correlation with voter turnout have t-statistics presented in bold.

Table 6: *Bivariate regressions – most significant variables*[29]

Variable	Coefficient	T-statistic
Institutional		
COMP	9.1	**2.7**
PR	5.4	**2.0**
LEGEFF	7.4	**2.7**
DISTRICT(log)	-2.2	**-2.2**
Socio-economic		
GDP(log)	6.7	**3.8**
LITERACY	0.2	**2.6**
Information circulation		
RADIOS(log)	4.0	**2.2**
TVS(log)	1.1	1.5
NEWSPAPER(log)	0.8	1.4
INFO (log)	3.0	**2.1**

Table 6 lists the most significant variables, and shows that the variables possess the expected direction in its correlation with the dependent variable. Only eight out of 22 variables I ran bivariate regressions on are significant at a 95 % confidence interval. Of the eight variables that turned out to be significant, four variables stand out. Two of these are institutional, COMP

[29] In every regression, MALI, GUATEMALA and CHILE I were included as dummy variables.

and LEGEFF and two are socio-economic, namely GDP and LITERACY. The latter two variables correlate, as noted above, too much (r = 0,75) to be included in the same model. Each will have to be tested thoroughly in multivariate regressions to check whether or not they should be included in the final model. All of the INFORMATION-variables also did well in the bivariate regressions, with RADIOS and INFO standing out.

5.3.2 Building a model

Given the results in the preliminary assessments alongside theoretical considerations discussed in the literature review chapter, I am compelled to include COMP (the presence of compulsory voting laws), in my final model. I will therefore run two-variable regressions with COMP, adding one independent variable at a time, starting with the most significant variables from the bivariate regressions, listed above. The variables that stay significant, when run along COMP in the two-variable regressions, will be added to the model and taken to the 'second round'. Again, variables will be tested against this 'new' final model, and the one's that matter, will be added. By conducting a comprehensive statistical strategy like this, I will end up with the final model that explains cross-national variation in voter turnout the best.

The two-variable regressions, including COMP and the other independent variables added one at a time, are presented below, in Table 7. The variables displaying significant correlation with turnout have t-statistics presented in bold.

Table 7: *Two-variable regressions – COMP plus one ind. variable added at a time*[30]

Variable	T-statistic	Variable	T-statistic
FREQ	−1.6	RADIOS(log)	**2.1**
CLOSED	0.9	TVS(log)	1.8
PLUMAJ	−0.4	NEWSPAPER(log)	**2.9**
PR	1.5	INFO(log)	**2.1**
UNICAM	0.1		
LEGEFF	**2.6**	DEMO	−0.5
NUMPART	0.03	RIOTS	−0.9
DISTRICT(log)	**−3.5**	GOVCRIS	−0.2
		WCI(log)	−0.2
GDP(log)	**3.6**		
POP(log)	−1.5	EASTERN	−0.04
POPDENS(log)	1.5		
LITERACY	**2.4**		
CAT	−1.3		

In the two-variable regression models, seven variables (specified alongside COMP) are significant, among which DISTRICT and GDP stand out. We know that GDP correlates too much with the other significant variables from the two-variable regressions (LITERACY, RADIOS, NEWSPAPER and INFO) to ever be included in the same model. DISTRICT on the other hand, only correlates with population size (POP) and since this variable does not seem to possess significant correlation with voter turnout, the exclusion of POP from future regression models is justifiable. I will, based on the above, include DISTRICT in the two-variable regression model, proceeding next to expand this to include three independent variables. The other independent variables will be run against this model, one at a time, to check each independent variables significance in explaining cross-national variation in voter turnout.

The multivariate regressions, including COMP, DISTRICT and the other independent variables added one at time, are presented below, in Table 8. The variables displaying significant correlation with voter turnout have t-statistics presented in bold.

[30] In every regression, MALI, GUATEMALA and CHILE I were included as dummy variables.

Table 8: *Multivariate regressions – COMP, DISTRICT plus one ind. variable added at a time*[31]

Variable	T-statistic	Variable	T-statistic
FREQ	**-2.5**	RADIOS(log)	1.5
CLOSED	1.8	TVS(log)	**2.0**
PLUMAJ	-0.9	NEWSPAPER(log)	**3.0**
PR	1.3	INFO(log)	1.8
UNICAM	-0.5		
LEGEFF	**2.8**	DEMO	1.2
NUMPART	0.8	RIOTS	1.4
		GOVCRIS	0.9
GDP(log)	**3.2**	WCI(log)	1.3
POPDENS(log)	**2.0**		
LITERACY	1.6	EASTERN	-0.4
CAT	-1.6		

Going from bivariate to multivariate regression models resulted in five of the variables changing direction in their correlation with voter turnout. UNICAM went from a positive to a negative correlation while all of the Activism-variables changed correlation in the opposite direction, from negative to positive. As none of those variables are even close to possessing any significance in their correlation with turnout I will not address the issue of direction change. A couple of variables did, however, change the significance of their correlation with voter turnout. RADIOS lost its significance while population density (POPDENS) appears to possess significance in its correlation with voter turnout.

As for the three-variable regression models above, we see that GDP again stands out in its significant correlation with voter turnout in the multivariate regression model. NEWSPAPER also displays a highly significant correlation. Including both GDP and NEWSPAPER in the same model is impossible due to multicollinearity. Hence, I will have to make a choice regarding which of the two variables I want to include in my final model. Including GDP has fundamental implications for which other variables to include in the model (see above).

Given the specification constraints I face, I choose to include NEWSPAPER instead of GDP, as the former captures a specific causal phenomenon, while the latter is a broader, less specific gauge of many other phenomena related to income. Due to collinearity issues, including

[31] In every regression, MALI, GUATEMALA and CHILE I were included as dummy variables.

NEWSPAPER rules out both LITERACY (r = 0,84) and GDP (r = 0,82) from the next regression models. Other variables that correlate significantly with turnout include, LEGEFF, POPDENS and TVS. One of these variables will be included in the model for the next regressions, alongside NEWSPAPER. As none of the socio-economic variables have yet been specified in my models, I will include POPDENS, even though LEGEFF displays a higher significance in its correlation with turnout.

Based on the above, I will include NEWSPAPER and POPDENS in an expanded regression model and run it against the other variables one at a time. The results are presented below, in Table 9. The variables displaying significant correlation with voter turnout have t-statistics presented in bold.

Table 9: *Core of final model – COMP, DISTRICT, NEWSPAPER and POPDENSITY plus one ind. variable added at a time*[32]

Variable	T-statistic	Variable	T-statistic
FREQ	−1.6	DEMO	0.8
CLOSED	1.5	RIOTS	1.2
PLUMAJ	0.1	GOVCRIS	1.2
PR	0.2	WCI(log)	1.4
UNICAM	0.6		
LEGEFF	1.6	EASTERN	−0.5
NUMPART	1.1		
CAT	**−2.1**		

The inclusion of more variables into the regression model has again led some variables to change the direction of their correlation with voter turnout. UNICAM is again correlating positively with turnout but as for the other regression model, the correlation does not display significance. Recall from the literature review chapter above, that there were mixed findings regarding the correlation of unicameralism. As Jackman (1987), Jackman and Miller (1995), and Fornos et al. (2004) report a positive effect, Blais and Carty (1990), Black (1991), Radcliff and Davis (2000), and Perez-Linan (2001) indicate no effect. Based on the regression models so far, with unicameralism going back and from in the direction of its correlation, I

[32] In every regression, MALI, GUATEMALA and CHILE I were included as dummy variables.

support the conclusion of the latter studies, namely the one indicating that unicameralism has no effect upon turnout.

Only CAT, the dummy variable indicating a catholic majority, is significant. CLOSED and LEGEFF have also displayed a relatively high significance through the model building procedure, with LEGEFF standing out. Based on this have I decided to include both CAT and LEGEFF in the multivariate model, and see if their significance is robust. If their significance is robust when included in the same model, they will be added to the final model.

The significance of CAT is robust, when included in the multivariate regression model, displaying a t-statistic score of over two. I will thus add CAT to my final model. I subsequently added LEGEFF and NUMPART to this expanded five-variable model (COMP, DISTRICT, POPDENS, CAT and NEWSPAPER) and they performed well, displaying t-statistics of between 1,3 and 1,5. As a result of this, they will both be added to the final model, which now is complete. The final model for explaining cross-national variation in voter turnout between 1950 and 2000 is presented in Table 10.

Table 10: *Determinants of cross-national variation in voter turnout – 1950-2000*

```
+-----------------------------------------------------+
| Ordinary    least squares regression                |
| Model was estimated May 07, 2007 at 05:04:16PM      |
| LHS=TURNOUTC Mean                 =   68.05178      |
|              Standard deviation   =   16.06015      |
| WTS=none     Number of observs.   =         80      |
| Model size   Parameters           =         11      |
|              Degrees of freedom   =         69      |
| Residuals    Sum of squares       =   9195.295      |
|              Standard error of e  =   11.54405      |
| Fit          R-squared            =   .5487270      |
|              Adjusted R-squared   =   .4833251      |
| Model test   F[ 10,    69] (prob) =   8.39 (.0000)  |
| Autocorrel   Durbin-Watson Stat.  =  1.7863293      |
|              Rho = cor[e,e(-1)]   =   .1068353      |
| White heteroscedasticity robust covariance matrix   |
| Br./Pagan LM Chi-sq [ 10]  (prob) =   3.35 (.9721)  |
+-----------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |t-ratio |P[|T|>t] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant       53.8182012     8.70476631     6.183   .0000
 COMP           13.1800157     3.59715654     3.664   .0005     .17500000
 DISTRICT(log) -4.05827247     1.13410168    -3.578   .0006    3.84761389
 LEGEFF         5.04759720     2.71193734     1.861   .0670     .55000000
 NUMPART         .57270107      .33621105     1.703   .0930    3.87067172
 POPDENS(log)   1.61481780      .71058105     2.273   .0262    7.14785371
 CAT           -6.09368434     2.66228897    -2.289   .0252     .45000000
 NEWSPAPER(log) 2.16804345     1.12758584     1.923   .0586    6.90068298
 CHILE         -39.4494735     3.08059846   -12.806   .0000     .01250000
 GUATE         -28.0832260     2.93459618    -9.570   .0000     .01250000
 MALI          -29.4654792     5.95151316    -4.951   .0000     .01250000
```

5.3.2.1 Interpretations of findings

Four institutional, two socio-economic and one information variable, alongside three country dummies, constitute the model that best explains cross-national variation in voter turnout between 1950 and 2000. Not surprisingly, compulsory voting laws are significant for explaining voter turnout, yielding an effect of over 13 percentage points, all else held equal. The effect is highly significant, with a t-statistic of 3,6. My finding supports the study of Jackman (1987), which also estimates an increase of 13 points in turnout due to compulsory voting laws. Why does COMP foster turnout? An obvious reason for this is that under such systems, the electors are likely to be punished if they do not vote at elections. I say likely, because the punishment for not complying with these laws varies between the different countries. Some countries have very strict implementation of the laws, ranging from fines to

imprisonment,[33] while in other countries the implementation is enforced more loosely. Also, it has often been overlooked that fines and sanctions are just one aspect of compulsory voting: systems that employ it usually reciprocate by reducing the cost of turnout for its citizens, via weekend voting, simple registration procedures, and the creation of a centralized, professional bureaucracy concerned with all aspects of election administration. In short, if a country imposes compulsory voting on its citizens, it follows through in every aspect, not just the one dealing with fines and sanctions.

The next institutional variable to have an impact on voter turnout is the district size measured by the DISTRICT-variable. The relationship is logarithmic, implying that the main difference is at the low end of district size: all else held equal, turnout decreases by over 14 percentage points when the district size goes from 2221 voters per member of parliament (Andorra), the lowest in the sample, to 86200 voters per member of parliament (about the size of Italy), which represents the average. The decrease in turnout is only 3,8 percentage points, all else held equal, when the district size goes from the average to the highest, 1,3 million voters per member of parliament (India). The effect, which is formidable, is highly significant with a t-statistic score of 3,5. My finding supports the study of Norris (2004), even though her DISTRICT-variable did not turn out significant. Why would the size of the district affect turnout? As noted in the measurement chapter, the smaller the number of electors for member of parliament, the greater the potential for constituency service and for elected representatives to maintain communications with local constituents. If the voter "knows" the representative, or party in PR-system, the higher the incentive to turnout based on any "personal" vote.

The relative power of the legislature vis-a-vis the executive power, measured by LEGEFF, also has an impact on turnout. An election for a legislature that possesses significant governmental autonomy, including, typically, substantial authority in regard to taxation and disbursement, and the power to override executive vetoes of legislation can expect a turnout that is five percentage points higher, all else held equal, than an election taking place in a situation in which the effective executives power substantially outweighs, but does not completely dominate that of the legislature. The effect is highly significant, yielding a t-statistic score of almost 1,9. The more powerful the body that is being elected, especially on important issues

[33] In cases where the non-voter does not pay the fines after being reminded or after refusing several times, the courts may impose a prison sentence. This is usually classified as imprisonment for failure to pay the fine, not for failure to vote. Examples include Australia and Cyprus.

like taxation and disbursement, the higher the turnout.

The last institutional variable to be included in my final model is effective number of parties, NUMPART. It correlates positively with turnout, yielding an effect of about, all else held equal, one point for every two parties added to the legislature. My finding supports Norris (2004) but contradicts almost all other empirical research on this variable, which posts a negative correlation between the number of parties and turnout (Jackman 1987, Blais and Carty 1990, Jackman and Miller 1995, Blais and Dobrzynska 1998, Radcliff and Davis 2000, Kostadinova 2003). That is quite interesting. It appears to be the case, that the higher the number of parties, the richer the choice offered the electorate which in turn leads to a higher turnout. The notion that the greater the number of parties, the more difficult it can be for voters to make up their minds (Blais and Dobrzynska) and the greater the number of parties, the less likely it is that there will be a one-party government which in turn will depress turnout (Jackman 1987 and Blais and Carty 1990) seems to be farfetched. In this regard it is important to note that NUMPART has displayed a positive correlation with voter turnout in *every* regression model it has been included in, except for the 'literature model' above. The effect of NUMPART on my data also displays a relatively high significance with t-statistic score of over 1,7.

Population density correlates positively with voter turnout because the more densely populated the voters are, the easier they are to mobilize for the different political parties. The relationship is logarithmic, implying that the main difference is at the low end of population density: all else held equal, turnout increases by about 1,5 points when density goes from 45 inhabitants per square mile (Australia), the lowest in the sample to the average (3278 inhabitants per square mile) but only 0,6 points when going from the average to the most densely populated country, at 28 647 inhabitants per square mile (Malta). Population density displays a fairly modest effect on voter turnout. One of the reasons for this could perhaps be due to cases like Australia, which enjoy a highly urbanized population (91% as of 2004)[34] but still come out as a low-density country due to the size of the country relative to size of population. My point is: Australia *is* densely populated where people actually live, along the coast to the east, southeast and southwest, it is just that when one take the entire country into consideration, it is not densely populated at all. The effect of POPDENS, however moderate, is highly significant, with a t-statistic score of almost 2,3.

[34] World in Figures (2004) *The Economist.*

CAT, indicating a catholic population of over 50%, correlates negatively with voter turnout. I included this variable because I wanted to see if turnout could be predicted along religious lines, and it appears it can. The effect upon turnout, by a predominantly catholic electorate, is quite astonishing. An election taking place in a 'catholic country', all else held equal, can expect a decreased turnout of about six percentage points. The effect is also highly statistically significant, with a t-statistic score of over 2,2. Why catholics vote at a lesser rate than others is not clear. Maybe catholics are more difficult to mobilize for the democratic 'institution' because of their commitments to other 'institutions'. A more plausible explanation for the lower turnout observed in catholic countries could be due to a spurious correlation. The 'catholic' democracies in my sample, with the majority coming from Latin America, have an authoritarian past. Turnout could be depressed in these countries for the same reasons as why it is depressed in the former Soviet satellites, measured through my EASTERN-variable. People that have experienced authoritarian rule and one-party government are characterised by a low level of political participation (Almond and Verba: 1963). Practically all of my 'catholic' democracies have such a political legacy which in turn could explain the depressed turnout we observe in such countries.

The last variable to be included in my final model deals with the flow of information in society, namely NEWSPAPER. The correlation is positive and logarithmic: all else held equal, turnout increases by 6,6 points when the newspaper circulation per capita goes from 0,001 newspaper per inhabitant (Central African Republic), the lowest in the sample, to the average (0,17 newspaper per inhabitant) but only 0,5 points going from the average to the highest (0,5 newspaper per inhabitant), the level of which Czechoslovakia was at between 1990 and 1993. As we see, the effect at the low end of newspaper circulation increase is quite formidable. The flow of information, measured through the distribution of newspapers, is important for the political mobilisation of the electorate, as information and political propaganda is more easily distributed by the political parties. NEWSPAPER is also statistically significant with a t-statistic score of over 1,9.

As for the dummy variables included to control for the three outliers, the effect is striking: all else held equal, an election taking place in Chile between 1950 and 1962 can expect a decrease in turnout of almost 40 percentage points. For Guatemala and Mali the numbers are 28 and 29 respectively.

5.3.2.2 Other interpretations

When building my final model for explaining cross-national variation in voter turnout I had to exclude a number of variables that did not 'fit' the model. Some of these variables were thought to be important and their exclusion merits attention. I start out with the two dummy variables dealing with electoral system, PR and PLUMAJ, which based on the literature review chapter above, were expected to yield high significance in their correlation with turnout. Both PR and PLUMAJ show the expected directions in their correlation with voter turnout. An election taking place under PR-systems, including Mixed Member Proportional, Single Transferable Vote and List PR, can expect an increase of about three percentage points, all else held equal. The finding, however, is not statistically significant. Elections taking place under Majority-Plurality-systems, including First Past The Post, Block Vote, Alternative Vote[35] and Two Round System, can expect a decrease of about 2,1 percentage points, all else held equal. As with PR, PLUMAJ is not statistically significant either. Why does not electoral system seem to matter? An answer to this question could lie in my basis for sample selection. Recall from the literature review chapter, that the studies that reported a positive and significant correlation between electoral system and turnout based their studies on established democracies.[36] Research dealing with Latin America reports no association between electoral system and turnout, and an analysis that incorporates both established and non-established democracies (Blais and Dobrzynska 1998) report a weak effect which yields a relatively low significance (t = 1,6). Hence, the reason why my electoral system variables are not significant must be found in my sample selection criteria (ACLP), which allows me to include 90 countries (96 cases) of which many are not established and many are found in Latin America.

The frequency of elections can not explain cross-national variation in electoral turnout, according to my regression models. The reason for this could be the conservative estimate I employ when measuring FREQ. As I calculate the electoral frequency by taking into account national level elections (parliamentary and presidential), my measurement does not count many other types of contests held during the time period under examination, including national and local referenda and initiatives, primaries, or European, regional/state, and local contests which in and of themselves could result in voter fatigue. We know that among western democracies, the cases of Switzerland and the United States are commonly cited as

[35] Also known as Instant runoff voting (IRV).
[36] Sample selection of each study is available in Appendix D.

exemplifying nations with frequent elections for office at multiple levels as well as with widespread use of referenda and initiatives, and, at the same time are characterized by exceptionally low voter participation.

The inclusion of NEWSPAPER had collinearity implications, which forced me to drop GDP from my final regression models. By running an additional version of the final model, replacing NEWSPAPER with GDP, I can check GDP's impact (logarithmic) on voter turnout: all else held equal, turnout increases by 1,6 points when GDP per capita goes from 793 American dollars (Central African Republic), the lowest in the sample, to the average (8350) but only 0,6 points when it goes from the average to 21 554 American dollars (Luxembourg). The effect is statistically significant, yielding a t-statistic score of almost two.

As for the variables dealing with political activism, they all correlated positively with turnout in the final multivariate regression models. Their correlations, however, were not statistically significant. My political legacy variable (EASTERN), correlated, as hypothesized, negatively with turnout but the correlation did not prove significant in any of my multivariate regression models.[37]

5.3.3 Methodological considerations

As noted in the methodology chapter, regression analysis rests on certain assumptions. Two of these assumptions, the absence of heteroscedasticity and the absence of perfect multicollinearity, are of particular importance, and I will address them here. The other assumptions are dealt with in Appendix C.

5.3.3.1 Heteroscedasticity

Violating the assumption of the homoscedasticity is serious. Homoscedasticity and heteroscedasticity are antonyms that refer to the correspondence of the *spread* of residuals with the independent variable. The residuals are *homescedastic* if they have a constant variance and *heteroscedastic* if their variance is variable. Heteroscedasticity impairs the efficiency of OLS estimators of the regression coefficients, but these coefficients are still

[37] Out of curiosity I ran F-tests to decide which of the two categories, institutions or socio economic variables, was the better predictor of cross-national variation in voter turnout. Institutional variables came out as the better predictor. Multicollinearity issues as well as limitations in scope and time of this thesis forced me to drop the issue from further inquiries.

unbiased (Pennings et al. 2006: 161). By saving the unstandardized residuals for the final model, and plotting them against each individual explanatory variable I visually checked the pattern to see whether or not heteroscedasticity could pose a problem. I found none. Nonetheless, I run my models implementing *White's Heteroscedasticity-Consistent Variances and Standard Errors*, also known as *robust standard errors***.** Since White's heteroscedasticity-consistent estimators of the variances are available in established regression packages, it is recommended that the reader report them. As noted in Gujarati (2003):

> Generally speaking, it is probably a good idea to use the WHITE option routinely, perhaps comparing the output with regular OLS output as a check to see whether heteroscedasticity is a serious problem in a particular set of data (418).

I ran my final model both with and without the White estimator and by comparing the output of the two models I can safely conclude that heteroscedasticity poses no problem in my model.

5.3.3.2 Multicollinearity

One of the assumptions of the ordinary least squares estimation method is the absence of perfect multicollinearity, which is the phenomenon of *highly* correlated independent variables. If ones models suffer from perfect multicollinearity, it is impossible to arrive at a unique solution for the least squares parameter estimates. Even though X is actually associated with Y in the population, our estimated regression coefficient may be so unstable that it fails to achieve statistical significance. Our parameter estimates also become unreliable, as the magnitude of the partial slope estimate varies considerably from sample to sample. For a diagnosis of multicollinearity we must look directly at the correlations between the independent variables. The correlation matrix provided useful preliminary checks for multicollinearity as I could examine the bivariate correlations. While suggestive, this approach is not satisfactory, for it fails to take into account the relationship of one independent variable with all the other independent variables. It is possible that one of the independent variables is a perfect linear combination of the other variables, even though the bivariate correlations were low. Hence, we need a better measure of multicollinearity. A widespread measure is referred to as the *tolerance*. The tolerance is defined as one minus the explained variance in one independent variable *j* due to the other independent variables in the

regression equation ($1-R_j^2$). In other words, the procedure is to regress each respective independent variable on the full set of other independent variables. In research based on data from official statistical agencies, as my data are, tolerance between 0,01 and 0,25 may still be acceptable, since these data are not prone to measurement errors (Pennings et al 2006). In general we can say that the closer the tolerance is to one, the greater the evidence that *Xj* (regressor under inspection) is not collinear with the other regressors. By running the tolerance test on my model the tolerance scores presented in Table 11 were obtained:

Table 11: *Tolerance test – general model*

Variable	Tolerance
COMP	0.78
LEGEFF	0.60
DISTRICT	0.99
NUMPART	0.88
POPDENS	0.99
CAT	0.80
NEWSPAPER	0.54
MALI	0.88
GUATEMALA	0.99
CHILE I	0.99

I can safely conclude, on the basis of the tolerance tests conducted, that multicollinearity poses no problem in my model.

5.3.3.3 How good is my model?

How good is my final model for explaining cross-national variation in voter turnout? The *multiple coefficient of determination*, R^2, could tell us this. An important property of R^2 is, however, that it is a nondecreasing function of the number of explanatory variables; as the number of variables goes up, R^2 almost invariably increases and never decreases (Gujarati 2003: 217). I therefore rely on the adjusted R^2, denoted by $\bar{R}^2$. The term *adjusted* means adjusted for the df (degrees of freedom) associated with the sums of squares entering into the regression. The $\bar{R}^2$ does not give an overly optimistic picture of the fit of the regression, particularly when the number of explanatory variables is not very small compared with the number of observations. My model explains close to 50% of the variation in voter turnout observed cross-nationally. Given the wide range of explanatory variables I have included in this thesis, the explained variance may appear to be low. Remember, however, that in a

regression analysis, our objective is not to get an as high an $\bar{R}^2$ as possible; rather, it should be to obtain dependable estimates of the true population regression coefficients and draw statistical interferences about them. As Gujarati (2003) notes: "the researcher should be more concerned about the logical and theoretical relevance of the explanatory variables to the dependent variable and their statistical significance" (222). If a model's $\bar{R}^2$ is low, it does not mean the model is necessarily bad.[38] Based on the logical and theoretical relevance of the explanatory variables included in my final model, I conclude that my model is quite solid and insightful for explaining cross-national variation in voter turnout between 1950 and 2000.

I turn now to a time-specific analysis of cross-national variation in voter turnout. By examining shorter time periods (decades) and comparing the findings across time, I can check whether or not the determinants behind cross-national variation in voter turnout varies or stay the same across time.

[38] See Gujarati (2003: chapter 7) for a discussion on the topic and Lieberson (1985) for a critique of methods that use proportion of explained variation as a fundamental criterion for evaluating models.

VI. Time-specific empirical patterns

In the previous chapter, the regression analyses were conducted on the basis of a dataset that covered the entire time period, from 1950-2000, and provided a general model for explaining cross-national differences in voter turnout, including the following variables: COMP, DISTRICT, LEGEFF, NUMPART, POPDEN, CAT, and NEWSPAPER.

In this chapter I am interested in checking for time-specific variance in the explanatory variables. Are the variables from the general model robust in their effects across shorter time periods? Did the same variables matter in the 1950s as in the 1990s? Are some variables significant across each decade? Questions like this motivate the following analyses. In this chapter I will therefore utilize the Decade Dataset I created in the preparations for this thesis. It allows me to run regressions across the independent variables for each decade. By running regressions that are sensitive to time, I am able to identify time-specific patterns in the explanatory variables behind the cross-national variation in voter turnout.

My research strategy in this chapter will differ from the one utilized in chapter 5, as I will not, except for the 1990s, build a final model for each decade. I will start out, as I did for the general model, by running simple bivariate regressions for each decade and present their findings. After a brief discussion of the bivariate findings I will go on to include the significant variables from the bivariate regressions in multivariate regression models. The multivariate regression models will constitute the basis from which the discussion of the different variables can be conducted. As many of the variables correlate too much to ever be included in the same model I will be compelled to run a great number of multivariate regression models. (Every model that includes a finding presented in this paper was checked for the presence of heteroscedasticity[39] and multicollinearity.)

6.1 Preliminary assessment

As for the general model above, I wish to run some simple bivariate regression models to get a feel for the data. The bivariate regressions will suggest which Xs are correlated with Y across each decade and whether this correlation is in the expected direction.

[39] As with the analyses above, the White estimator was included in every model.

Before we turn to the bivariate regression analyses a note on outliers is in place.

6.1.1 Outliers

The descriptive statistics tables for each decade are available in Appendix F. As for the general model above, I visually checked the partial regression plots on the dependent variable to see if there were any outliers across the decades. A number of cases were identified as outliers and will accordingly be included in every regression model as dummy variables so that they are controlled for. The outliers I identified are:

1950s: *Brazil, Chile and Ecuador* **1960s**: *Guatemala and Panama* **1970s**: *Guatemala*
1980s: no outlier identified **1990s**: *Mali*

6.1.2 Bivariate regressions

Table 12 presents the t-statistic score for each independent variable when included in bivariate regression models across each decade. Also included, for the purpose of comparison, is the score for each variable in the bivariate regressions conducted when the general model was built above (indicated by GM). The variables that display significance in their correlation have t-statistics presented in bold.

Table 12: *Bivariate regressions – t-statistic score for each variable*

Variable	*Decade*					
	50s	**60s**	**70s**	**80s**	**90s**	**GM**
COMP	0.7	0.6	**3.6**	**2.4**	**3.9**	**2.7**
CLOSED	**2.2**	1.1	1.0	0.9	−0.2	1.0
FREQ	**−2.1**	0.8	−0.3	−0.1	−0.7	−1.3
PLUMAJ	−0.4	−0.8	**−2.3**	−1.4	−1.1	−1.0
PR	1.1	1.9	**3.0**	1.6	**2.1**	**2.0**
UNICAM	−0.1	1.1	1.3	**2.4**	−0.06	−0.3
DISTRICT(log)	−0.3	**−2.7**	−1.6	−1.9	**−3.4**	**−2.2**
LEGEFF	**2.0**	1.8	1.5	**2.8**	**2.2**	**2.7**
NUMPART	0.7	−0.7	1.3	0.5	0.2	−0.1
GDP(log)	**2.1**	**2.9**	**2.1**	**2.4**	**2.8**	**3.8**
LITERACY	**2.7**	**2.8**	**2.4**	**2.4**	**2.3**	**2.6**
CAT	−0.5	−1.0	0.8	−0.06	0.4	−0.3
POP(log)	−0.1	−1.2	−0.3	−1.2	−1.6	−1.0
POPDEN(log)	0.1	0.6	−0.01	0.02	0.8	1.0
HDI	–	–	**–**	**2.6**	**3.1**	–
GINI	–	–	–	–	**−2.2**	–
RADIOS(log)	**2.1**	**2.7**	1.7	1.8	1.6	**2.2**
NEWSPAPER(log)	**2.8**	**3.0**	**2.6**	**2.5**	**2.2**	1.4
TVS(log)	–	–	**2.6**	**2.3**	1.5	1.5
INFO(log)	**2.3**	**2.8**	**2.8**	**2.4**	1.9	**2.1**
DEMO	−0.1	**−3.0**	−0.8	−0.2	−0.3	−0.6
RIOTS	0.1	**−4.5**	−0.1	−0.8	−0.03	−1.0
GOVCRIS	1.4	−0.06	0.5	1.6	−0.1	0.1
WCI(log)	0.3	−1.2	1.3	−1.0	−0.6	0.2

Note: Outliers for each decade was included as dummy variable in each bivariate regression. Outliers include the following countries for each decade: 50s: Brazil, Chile and Ecuador; 60s: Guatemala and Panama; 70s: Guatemala; 90s: Mali.

6.1.3 Brief discussion of bivariate results across decades

Before I turn to a more detailed interpretation of the findings for each decade, which also include multivariate regression models, a general discussion of the patterns (from Table 12) across the decades are in place.

None of the institutional variables are significant across the entire time period. Compulsory voting laws (COMP) are not significant in the 1950s and 1960s but from 1970 and onwards the variable displays a strong significance in its correlation with voter turnout. As for the electoral

systems, both PLUMAJ and PR show the expected direction in their correlation with voter turnout across every decade. The effect, however, is only significant for the 1970s for PLUMAJ and for the 1970s and 1990s for PR. DISTRICT is, from the 1960s and onwards, highly significant and the effect, which is negative, is of expected direction across the decades. LEGEFF shows strong significance in its positive correlation with voter turnouts, but the t-statistic score is greater than two only for the 1950s and 1980s.

Two socio-economic variables are significant across all five decades, namely GDP and LITERACY. HDI and GINI are also significant but numbers on these variables are limited, the former only covering the last two decades, the latter only the last decade. The circulation of newspaper's (NEWSPAPER) correlation with turnout, which is positive, is also strongly significant across all decades. The other information circulation variables (RADIOS, TVS and INFO) are also highly significant across all decades but NEWSPAPER stands out. Interestingly enough, RADIOS seem to play a more dominant role in the 1950s and 60s, when the competition from television was not present. In the 1970s and 80s we observe that TVS is significant, RADIOS not, but TVS's significance eventually fades in the 1990s. As of variables dealing with mobilization and activism, we see that DEMO and RIOTS are the only variables that display significance and only for the 1960s. That is interesting and I will address those particular findings when discussing the 1960s model.

6.2 Decade models

The variables were also included in multivariate regression models across each decade. I will now present the findings and I begin with the first decade in the analysis, namely the 1950s.

6.2.1 The 1950s[40]

A number of variables are significant in the bivariate regressions for the 1950s. As for institutional variables they are: CLOSED (closed or open party list), FREQ (frequency of elections) and LEGEFF (legislative effectiveness). What is striking about the CLOSED variable, given that it is significant, is the direction of its correlation with voter turnout. Where closed lists are in practise, voters can not express preferences for candidates within the list. As this limits the influence on the order in which a party's candidates are elected and hence gives the

[40] Brazil, Chile and Ecuador were, due to relative low turnout, identified as outliers and included in every regression model as dummy variables. N = 30.

voter less of a choice as of preferences I expected a lower turnout where closed lists are in practise. The opposite is in fact the case. When included in multivariate regression models, CLOSED is still significant and worth about eight percentage points on voter turnout. Why CLOSED is positively correlated with voter turnout is not easy to explain. It could perhaps be the case that the ballot forms used in closed list-systems are easier to interpret and use for the voter which in turn will lead to higher turnout. A more plausible explanation is perhaps that there are factors associated with closed list-system which increase turnout and that my measurement is not finely grained enough to capture what I am interested in measuring, namely the lack of preference expression for the voter on the ballot form. Since CLOSED is only significant for the 1950s, its significance could also be due to the nature of the particular data for that decade.

FREQ is also significant in the bivariate regression and this variable too has an unexpected direction in its correlation indicating an increased turnout as frequency goes up. I expected the opposite effect of frequency, associating it with increased costs for the elector which in turn could lead to voter fatigue. I measure frequency as number of years since last national level election. It is important to note that my estimate is likely to represent a conservative estimate, as it does not count many other types of contest held during the time period under examination, including national and local referenda and initiatives, primaries, or European, regional/state, and local contest. If the frequency of election at national level is not positively correlated with elections at lower level, referenda and such, I have no reason to believe that FREQ should increase turnout, but it does, according to the multivariate regression model for the 1950s: all else held equal, turnout decreases by three points for every extra year since last national election. As with CLOSED, FREQ is only significant for the 1950s, making me wonder if it is the particular data at hand for that decade that produce such interesting results.

The relative power of the legislature vis-a-vis the executive power (LEGEFF) is also significant for the 1950s. It loses, however, its significance when included in multivariate regression models. All the other institutional variables show, when included in bivariate regressions, expected direction in their correlation with voter turnout, but none are significant.

Not even compulsory voting, COMP, is significant in the 1950s. It may come from the fact that only eight cases had mandatory voting laws in the 1950s and two of these, Brazil and Costa Rica, had poor turnout at 31,8% and 49,8% respectively.

As for socio-economic variables, two are highly significant, namely GDP and LITERACY. GDP is also highly significant when included in multivariate regression models with a coefficient of about six. The relationship is logarithmic, implying that the main difference is at the low end of economic development: everything else being equal, turnout increases by 2,5 points when GDP per capita moves from 806 American dollars (India), the lowest in the sample, to the average, 6609 American dollars, but only by 1,4 points when it moves from the average to the highest, 13551 American dollars (Switzerland). LITERACY is also highly significant in multivariate regression models, with an effect of about 0,3 percentage points increase in turnout for every point increase in literacy. None of the other socio-economic variables are significant for 1950s bivariate and multivariate regression models.

All of the variables dealing with the flow of information are significant, including RADIOS, NEWSPAPER and INFO. When included in different multivariate regression models, it is the circulation of newspaper per capita (NEWSPAPER) that yields the highest significance. The correlation is positive and has the expected effect, which is logarithmic: all else held equal, 4,5 points increase in turnout when the circulation of newspaper increases from the level of India (0,008 papers per inhabitants), the lowest in the sample, to the average (0,25 papers per inhabitants) but only one point when it moves from the average to the highest level, which is the level of United Kingdom (0,5 papers per inhabitants).

None of variables dealing with mobilization and activism were significant for the 1950s regression models.

We thus observe the following overall tendency: socio-economic variables are important for explaining cross-national variation in voter turnout for the 1950s. The level of economic development and literacy are the two most significant variables. Important are also the variables dealing with the flow of information in society, with the distribution of newspapers standing out. Institutional characteristics do not seem to matter much, not even compulsory voting. None of the variables in the activism bloc matter either for explaining cross-national variation in voter turnout for the 1950s.

6.2.2 The 1960s[41]

The bivariate regressions for the 1960s also produced some interesting results. None of the institutional variables, except the district size (DISTRICT), were significant even though PR came close. DISTRICT correlates negatively with voter turnout, as expected, but the significance from the bivariate regression is not robust in the multivariate regression models. PR, when included in multivariate regression models, yields an effect of about four points on voter turnout, but the effect is not statistically significant. COMP was not significant for the 1960s either. Ten countries had mandatory voting laws in the 1960s, including two low-turnout countries, Brazil at 36,9% and Chile at 42,6% and one with relatively low turnout given the laws at hand, Luxembourg at 70%, which could perhaps explain why COMP did not turn out as significant.

As for the 1950s, GDP and LITERACY are statistically significant for the 1960s when included both in the bivariate and multivariate regression models. The latter analyses tell us that the effect of GDP is slightly stronger in the 1960s than it was in the previous decade: all else held equal, turnout increases by 2,6 points when GDP per capita moves from 959 American dollars (Sri Lanka), the lowest in the sample, to the average 7898 American dollars and by 1,2 points when it moves from the average to the highest, 17807 American dollars, which is the level of Switzerland. The effect of literacy is also a bit stronger in the 1960s compared to the 1950s: all else held equal, turnout increases by 0,5 points when literacy increases by one point.

All of the variables dealing with the flow of information are significant for the 1960s, including RADIOS, NEWSPAPER and INFO. When included in different multivariate regression models, RADIO and NEWSPAPER stand out as the most significant, with RADIO yielding the largest effect: all else held equal, turnout increases by 4,8 points when the number of radios goes from the level of India (at 0,001 radios per inhabitant), the lowest in the sample, to the average (0,3 radios per inhabitants) and 1,6 points when it moves from the average to the highest, 1,1 radios per inhabitants, which is the level of the US.

What is particularly interesting regarding the 1960s regression models is the significance displayed by two of the Activism-variables, namely DEMO (number of peaceful demonstrations a year) and RIOTS (number of riots a year). They are both highly significant in

[41] Guatemala and Panama were, due to relative low turnout, identified as outliers and included in every regression model as dummy variables. N=39.

the bivariate regressions, only DEMO for the multivariate regression models. Recall from the measurement chapter that DEMO was included to measure the level of peaceful political mobilisation. As I did not have a clear cut expectation regarding the direction of the correlation between DEMO and turnout, I hypothesised both a positive and negative correlation between turnout and the number of peaceful demonstrations. For the 1960s this correlation is negative, yielding an effect of about 2,5 points decrease in turnout for every peaceful demonstration, all else held equal. The reason for this could be that people see demonstrations as a more powerful tool in influencing governmental policies than that of going through the polls. When the voter does not get what he or her wants by going through the polls, he utilizes different means to achieve his or her political ends.

We thus observe the following tendencies: socio-economic variables were important in the 1960s, just as they were in the previous decade, and again is it the level of economic development and the level of literacy that stand out. The effect they display is also slightly higher in the 1960s compared to the 1950s. The information-variables are also important for the 1960s, only this time it is the number of radios that stand out. As for the activism-variables, we see that the number of demonstrations play a significant role in explaining cross-national variation in voter turnout. Again, institutional variables seem to play no role in determining turnout.

6.2.3 The 1970s[42]

A number of institutional variables are significant in the bivariate regressions for the 1970s, including COMP, PR, and PLUMAJ. When included in multivariate regression models, COMP is highly significant, yielding an effect of about nine points on voter turnout, all else held equal. As for electoral system variables, both PLUMAJ and PR are significant in the bivariate regression. In the multivariate regression models however, only PR's significance is robust, yielding an effect upon turnout of about 8,6 points, all else held equal. The relative importance of the legislature vis-a-vis the executive power (LEGEFF) yields an effect of, all else held equal, about ten points. That is quite a formidable effect. The finding is, however, not statistically significant with a t-stat of 1,6. The size of the district correlates, as expected, negatively with turnout, yielding an effect (logarithmic), of about four points when moving from the smallest district (Iceland) to the average district, but only two points going from the

[42] Guatemala was, due to relative low turnout, identified as an outlier and included in every regression model as a dummy variable. N=43.

average to the largest district (India). The effect, however, is not statistically significant with a t-stat of 1,5. UNICAM, meaning that the legislature consists of only one house, is significant (t = 2,5) when included in the multivariate regression models, yielding an effect of about five points on turnout, all else held equal.

GDP and LITERACY were again significant in the bivariate regressions but unlike the previous two decades, none were significant when included in multivariate regression models. None of the other three socio-economic variables were significant in the multivariate regression model either. That is quite interesting. Also taking into account that none of the variables dealing with neither the flow of information in society, nor any of the variables dealing with political activism were significant in any regression models for the 1970s makes the institutional variables stand out when examining the explanatory factors behind cross-national variation in voter turnout in the 1970s.

We thus observe the following tendency: the patterns from the two previous decades are reversed, with institutional variables replacing the socio-economic variables in explaining cross-national variation in voter turnout. The dominant role of institutions stands out when taking into account that none of the information- or activism-variables seems to matter either.

6.2.4 The 1980s[43]

Three institutional variables were significant in the bivariate regressions for the 1980s, namely COMP, UNICAM and LEGEFF. When included in multivariate regression models, their significance proved to be robust. Compulsory voting laws increase turnout by almost 9 points, all else held equal. That is the same effect as COMP displayed in the 1970s model. UNICAM is positively correlated with turnout and yields an effect of 8,1 points, all else held equal. This finding, which is highly significant (t = 2,9) is consistent with Jackman (1987). He estimates that the presence of unicameral legislatures in countries like Finland and Norway produces a turnout increase of about eight points over the rate that prevails with strong bicameralism (414-415). LEGEFF also yields an effect of about eight points, all else held equal. The effect, which is slightly more moderate than in the 1970s, is statistically significant in the 1980s, with a t-statistic score of over two. When included in the multivariate regression models, a fourth institutional variable proved significant for the first time, namely DISTRICT, yielding a t-statistic score of 2,3. The negative effect of district size is logarithmic with the main

[43] No outlier identified (N=60).

difference at the low end of district size: all else held equal, turnout decreases by 5,4 percentage points when the size of the district goes from 3575 voter per member of parliament (Dominican Republic), the smallest district size in the 1980s, to the average (100 544 voters per member of parliament), but only 2,7 percentage points going from the average to the largest district size (1 383 413 voters per member of parliament), the level at which India were in the 1980s.

GDP, LITERACY and HDI (introduced for the first time) were all significant in the bivariate regression models for the 1980s. However, when included in multivariate regression models, their significance did not prove to be robust. This makes the institutional variables stand out as they did in the 1970s model. This impression is strengthened when looking at the other variables. None of the variables dealing with the flow of information prove to be significant when included in multivariate regression models. The only non-institutional variable that proves to be significant is the GOVCRIS-variable (t = 2,1) indicating that turnout will increase with seven percentage points for every rapidly developing situation that threatens to bring the downfall of the present regime (excluding situations of revolt aimed at such overthrow).

We observe the following tendency: as was the case for the 1970s, institutional variables mattered the most for explaining cross-national variation in voter turnout during the 1980s. The socio-economic variables were not important for the 1980s, continuing their poor showing from the previous decade. The information-variables alongside the activism-variables, except GOVCRIS, did not play a role in determining the cross-national variation in voter turnout.

6.2.5 The 1990s[44]

As the 1990s is the decade from which I have the most data and variables, and I will thus systematically build a model that explains cross-national variation in voter turnout for that decade.

Four institutional variables were significant for the 1990s bivariate regression models, namely COMP, PR, DISTRICT and LEGEFF. As correlation between these variables poses no problem, my intention was to let them constitute the core of my explanatory model for the 1990s. When

[44] Mali was, due to relative low turnout, identified as an outlier and included in every regression model as a dummy variable. N=89.

run against the other significant variables however, PR and LEGEFF completely lost significance, persuading me drop them from the model altogether. Four socio-economic variables are significant in the bivariate regressions, including GDP, LITERACY, HDI and GINI. When added to the institutional core above, HDI displays the highest significance of the four and will be added, alongside POPDENS to the model. I will also include NUMPART and CAT, and thus the final model for the 1990s includes seven variables. None of the information-variables seem to play a role in determining voter turnout in the 1990s, as they all displayed a poor significance when included in the final model.[45] When the activism-variables were added to the final model one at a time, DEMO stood out, displaying a high significance in its correlation with voter turnout (t = 2,3). I will hence include it in my final model which now is complete. It is presented in Table 13.

Table 13: *Determinants of cross-national variation in voter turnout – 1990s*

```
+----------------------------------------------------+
| Ordinary    least squares regression               |
| Model was estimated May 23, 2007 at 01:09:49PM     |
| LHS=TURNOUTC Mean                 =   67.47602     |
|              Standard deviation   =   16.14341     |
| WTS=none     Number of observs.   =         71     |
| Model size   Parameters           =          9     |
|              Degrees of freedom   =         62     |
| Residuals    Sum of squares       =   10728.76     |
|              Standard error of e  =   13.15464     |
| Fit          R-squared            =   .4118869     |
|              Adjusted R-squared   =   .3360013     |
| Model test   F[  8,    62] (prob) =   5.43 (.0000) |
| Autocorrel   Durbin-Watson Stat.  =  2.1777595     |
|              Rho = cor[e,e(-1)]   =  -.0888798     |
| White heteroscedasticity robust covariance matrix  |
| Br./Pagan LM Chi-sq [  8]  (prob) =   6.24 (.6208) |
+----------------------------------------------------+
+---------+-------------+----------------+--------+---------+----------+
|Variable | Coefficient | Standard Error |t-ratio |P[|T|>t] | Mean of X|
+---------+-------------+----------------+--------+---------+----------+
 Constant      49.1373041     10.4841131     4.687   .0000
 COMP          16.1611832     5.06650377     3.190   .0022     .15492958
 DISTRICT(log)-5.31071329     1.71588105    -3.095   .0030    4.01006716
 NUMPART        .55971374      .34603385     1.618   .1108    4.24886286
 HDI           27.4568878     11.3071729     2.428   .0181     .77679577
 POPDENS(log)  1.91709987      .97164456     1.973   .0530    7.31772622
 CAT          -4.31066011     3.24675894    -1.328   .1892     .46478873
 DEMO          2.88638587     1.23069436     2.345   .0222     .60328638
 MALI         -24.2774262     6.72352385    -3.611   .0006     .01408451
```

[45] Correlation between HDI and the information-variables are: NEWSPAPER (r = 0,6), RADIO (r = 0,5), TV (r = 0,7) and INFO (r = 0,7).

6.2.6 Interpretation of findings

Three institutional, three socio-economic and one activism variable, alongside one country dummy variable, constitute the final model for explaining cross-national variation in voter turnout in the 1990s. Compulsory voting laws are (as in the general model above) strongly correlated with voter turnout, displaying an even stronger effect than for the general model: all else held equal, compulsory voting increases turnout by over 16 percentage points. The effect is highly significant, displaying a t-statistic score of 3,2. The reasons why compulsory voting has a substantial impact on voter turnout are straightforward: in countries were compulsory voting is in use, the electors are likely to be punished if they do not vote at elections. Furthermore, countries that employ compulsory voting laws are also engaged in reducing the cost of turnout for its citizens, via weekend voting, simple registration procedures, and the creation of a centralized, professional bureaucracy concerned with all aspects of election administration.

The next institutional variable to have an impact on voter turnout in the 1990s is the district size measured by the DISTRICT-variable. The relationship is logarithmic, implying that the main difference is at the low end of district size: all else held, turnout decreases by about 24 percentage points when the district size goes from 2215 voters per member of parliament (Andorra), the lowest in the sample, to 100 653 voters per member of parliament, the average. The decrease in turnout is only 5,3 percentage points, all else held equal, when the district size goes from the average to the highest, 1,7 million voters per member of parliament (India). The effect of the district size is quite formidable and also highly significant (t = -3). Recall the effect of DISTRICT in the general model above: 14 points when going from minimum to average and almost four points when going from average to maximum. The district size is an important explanatory factor behind voter turnout because the linkages between voters and their representatives are affected by the number of electors per member of parliament. If a district is small, the possibility for increased information, familiarity and contact between voters and their representatives is present which in turn will increase turnout.

The last institutional variable to have an impact in the 1990s is the number of parties, measured through the NUMPART-variable. The positive effect, all else held equal, of one point for every two parties is the same as for the general model above. The correlation however is

not as significant as it was above, displaying a t-statistic score of 1,6. What is interesting about the finding is that NUMPART once again, as in the general model above, displays a positive correlation with voter turnout. My finding thereby contradicts the findings of Jackman (1987), Blais and Carty (1990) and Blais and Dobrzynska (1998) among others which all report a negative correlation between the number of parties and turnout. In this paper NUMPART has displayed a positive correlation in every multivariate regression model it has been included in making its positive correlation with turnout robust. It appears to be the case that the higher the number of parties, the richer the choice offered the electorate which in turn leads to a higher turnout. As Norris (2004) notes: " … that wider electoral choices across the ideological spectrum mean that all sectors of public opinion and all social groups are more likely to find a party to represent their views, preference and interests" (166). My findings on the NUMPART-variable throughout this paper support the argument of Norris.

The level of well-being in a country, calculated on the basis of the literacy rate, life expectancy, standard of living and education have a positive impact on voter turnout. It is measured through my HDI-variable and increase turnout, all else held equal, by 2,7 percentage points for every 0,1 point increase on the HDI (which runs from 0,3 to 0,95). The finding is highly statistically significant, displaying a t-statistic score of over 2,4. My finding supports that of Norris (2004). Why would a higher level of development, measured through the HDI, result in a higher turnout? Recall that HDI compromises four dimensions: literacy, life expectancy, standard of living and education. The literacy dimension of HDI's correlation with turnout follows the same argument as the LITERACY variable created for this paper. Those with little linguistic skills are less likely to vote and studies (Blais and Dobrzynska: 1998) show that a minimum degree of literacy is almost a prerequisite to good turnout. The life expectancy dimension deals with the well being of the population. Alongside the standard of living dimension it effectively measures the quality and quantity of goods and services available to people also taking the distribution of such goods and services into account. People have basic needs and these needs must be met before the majority can engage in political participation. The last dimension is education and it follows the same logic as the literacy dimension. Education enhances civic skills, which in turn enhance turnout.

Population density correlates positively with voter turnout in the 1990s, as it did in the general model above. The relationship is logarithmic, implying that the main difference is at the low end of population density: all else held equal, turnout increases by about 1,6 points when

density goes from 60 inhabitants per square mile (Australia), the lowest in the sample to the average (3687 inhabitants per square mile) but only 0,7 points when going from the average to the most densely populated country, at 31 010 inhabitants per square mile (Malta). The effect, which is as moderate for the 1990s model as it was for the general model, is statistically significant, displaying a t-statistic score of almost two. CAT, indicating a catholic population of over 50%, correlates negatively with voter turnout for the 1990s as it did in the general model above. The effect, which does not display a high significance (t = -1,3), is worth about 4,3 percentage points of depressed turnout. The effect is smaller in the 1990s than for the entire time period covered by the general model above.

The last variable to be included in the 1990s model was an activism-variable, namely DEMO. In the multivariate regression models for the 1960s, DEMO displayed a negative correlation with voter turnout. For the 1990s however, the correlation is a positive one, displaying an effect of almost three percentage points on voter turnout for every demonstration, all else held equal. Whereas people in the 1960s saw demonstrations as an alternative channel for expressing political views, people in the 1990s saw demonstrations as a supplement to that of showing up at the polls. The finding on this variable is highly statistically significant, displaying a t-statistic score of over 2,3.

As for the dummy variable indicating the outlier case of Mali, the effect is formidable: all else held equal, an election taking place in Mali in the 1990s can expect a decreased turnout of over 24 percentage points.

6.2.7 Other interpretations

The inclusion of HDI in my model had collinearity implications for two other variables, GINI and GDP, which in turn had to be dropped from the final model. By substituting HDI in the final model with GDP and GINI respectively, the following effects were obtained.

All else held equal, turnout increases by 1,2 percentage points when GDP per capita goes from 793 American dollars (Central African Republic), the lowest in the sample, to the average (10 839) but only 0,6 points when it goes from the average to 40 583 American dollars (Luxembourg). The effect is statistically significant, yielding a t-statistic score of almost two. Interestingly enough, the effect of economic development in determining cross-

national variation in voter turnout for the 1990s is identical both in terms of effect and significance as the effect of economic development in the general model above.

The level of income inequality, measured through GINI, displays a highly significant correlation with voter turnout (t = -3,5). The effect is, all else held equal, 0,5 percentage points of decreased turnout for every 0,1 point increase on the GINI-index. In other words: the more unequal the distribution of income in a society is, the lower the turnout. My finding is consistent with my expectations. If people's basic needs are not met, the majority can not "afford" or have not the means to engage in political participation.

6.2.8 Methodological considerations

All the regression models run in this paper have been done so implementing White's Heteroscedasticity-Consistent Variances and Standard Errors. By comparing the output of my final model for the 1990s run both with and without the White's estimator and comparing the output of the two, I can safely conclude that heteroscedasticity does not pose a problem in my model.

As of multicollinearity, I ran a tolerance test on my model, as I did for the general model above, and the results are presented in Table 14:

Table 14: *Tolerance test – 1990s model*

Variable	Tolerance
COMP	0.79
DISTRICT	0.71
NUMPART	0.99
HDI	0.77
POPDENS	0.90
CAT	0.82
DEMO	0.77
MALI	0.80

All of the variables in my model appear to possess a low level of collinearity with the other regressors. I can firmly conclude that multicollinearity poses no problem in my model. As of other assumptions, see Appendix C.

6.3 Cross-time variance in the determinants of voter turnout

The regression analyses conducted across the decades from 1950 to 2000 revealed that some variables are time-specific whereas other variables are more consistent across time. By consistent I mean both in the direction and significance of their correlation with voter turnout. I will begin by identifying the most consistent variables, as the identification of such variables constitutes the primary purpose of these analyses. When discussing the cross-time-specific determinants of voter turnout it can be fruitful to keep in mind the findings from the general model above. Recall that the variables behind cross-national variation in voter turnout between 1950 and 2000 are the following: COMP, LEGEFF, NUMPART, DISTRICT, POPDENS, CAT and NEWS. Are these variables consistent in their correlation with voter turnout across shorter time periods as well? Some of them are more consistent than others and I turn now for an identification of the most consistent variables. Table 15 below provides an overview of the different variable's direction and significance (* sig. at 0.10; ** sig. at 0.05; *** sig. at 0.01).

Table 15: *Direction and significance of correlation – multivariate regression models*

	Decade					
Variable	**50s**	**60s**	**70s**	**80s**	**90s**	**GM**
COMP	+	+	+**	+**	+***	+***
CLOSED	+**	+	+	+	−	+
FREQ	−***	+	−	−	−	+
PLUMAJ	−	−	−	−	−	−
PR	+	+	+**	+	+	+
UNICAM	+	+	+***	+***	+	+
DISTRICT	−	−	−	−**	−***	−***
LEGEFF	+	+	+	+**	+	+*
NUMPART	+	−	+	+	+	+*
GDP	+***	+**	+	+*	+**	+**
LITERACY	+**	+***	+	+	+	+
CAT	−	−	+	−	−	−**
POP	−	−	−	−	−	−
POPDEN	+	+	−	+	+**	+**
HDI				+	+***	
GINI					−***	
RADIOS	+*	+***	+	+	+	+
NEWSPAPER	+**	+**	+	+	+	+*
TVS			+	+	+	+*
INFO	+ *	+**	+	+*	+	+
DEMO	−	−**	−	−	−	+
RIOTS	+	+	−	−	−	+
GOVCRIS	+*	−	+	+**	−	+
WCI	+	−	+	−	+	+

6.3.1 Consistent variables

Compulsory voting laws increase turnout. COMP has correlated positively with turnout in every regression model it has been included in so its direction of correlation is undisputed. The significance of COMP's correlation has also been undisputed, except for the 1950s and 1960s regression models. As discussed above, the lack of significance displayed by COMP for those decades probably comes from the cases that both enjoyed compulsory voting laws and low turnout. Based on the regression analyses conducted in this paper, alongside evidence on the variable presented in the literature review chapter, compulsory voting stands out as an important explanatory factor behind cross-national variation in voter turnout.

The flow of information in society, measured through the circulation of newspaper (NEWSPAPER), number of radios per capita (RADIOS), number of television sets per capita (TVS) or as a compilation of the three (INFO) are also correlating positively with voter turnout. As of significance, NEWSPAPER and INFO stand out. NEWSPAPER displays a high significance in its correlation with voter turnout in every regression model it has been included in, except for the 1970s, 1980s and 1990s models. INFO does even better, displaying significance in all regression models, except in the 1970s and general model.[46] The relatively consistent significance displayed by NEWSPAPER and INFO emphasizes their importance when cross-national variation in voter turnout is being studied. The flow of information does increase turnout.

Three more variables stand out as of stability in the direction and significance of their correlations with voter turnout, namely GDP, DISTRICT and GINI. The level of economic development correlates positively with voter turnout, both when included in the general model above, as well as when included in the time-specific multivariate regression models. The correlation is also significant across each and every regression model, except for the 1970s and 1980s model. Blais and Dobrzynska (1998) also reported a positive and significant correlation between economic development (GNP per capita) and voter turnout, making it clear that economic development increases turnout.

[46] It was not included in the final 1990s model due to collinearity issues with HDI. When HDI was substituted with INFO, the latter came out as significant.

The district size is also stable in the direction of its correlation with turnout, displaying a negative correlation throughout this paper. The significance however, has been limited to the general model plus the 1980s and 1990s model. Norris (2004) was right when she found district size to correlate positively with turnout. Her finding, which was in fact not significant even though she list it as being significant,[47] came about partly as a result of the time period she covered in her study, namely the 1990s. Based on the regression analyses conducted in this paper, district size does stand out as an important explanatory factor behind cross-national variation in voter turnout.

The level of income inequality (GINI) has, due to the lack of data, only been included in one regression model, namely the 1990s model. The correlation, which was negative, was highly significant. Further research is needed before we can conclude anything final on this variable's explanatory power on cross-national variation in voter turnout but for now we can say that the level of income inequality does seem to decrease turnout.

I turn now to an identification of the inconsistent variables.

6.3.2 Inconsistent variables

The majority of the variables in this paper, including four from the general model (LEGEFF, NUMPART, POPDENS and CAT), are consistent in the direction of their correlation but seem to display a *time-sensitive* significance in their correlation with voter turnout. Their correlation's inconsistence as of significance makes them unreliable as determinants of cross-national variation in voter turnout.

- LEGEFF. Legislative effectiveness correlates positively with turnout throughout this paper but the correlation is only significant in the 1980s model.

- NUMPART. Multipartyism correlates positively with voter turnout (except in the 1960s model) but not once in my study is the correlation significant. Its best showing is in the general model (t = 1,7) and in the 1990s model (t = 1,6). My findings on NUMPART support, as noted above, the finding of Norris (2004). She is one, of only a few, to have found a positive correlation between voter turnout and number of parties. Her finding is also

[47] With a coefficient of -0,001 and standard error of 0,035, her district size variable is far from significant (p. 158).

statistically significant. Recall that she studied cross-national variation in the 1990s, the decade from which my finding on the variable came close to significance.

- POPDENS. Population density has a similar pattern: positive correlation throughout the paper; significant, however, only for the general model and in the 1990s model.

- CAT. The variable taking into account catholic majorities is also stable in its correlation with voter turnout, displaying a negative correlation in every regression model it has been included in. The significance of its correlation however, is limited to the general model.

- LITERACY. Literacy has displayed a positive correlation with voter turnout throughout this paper. Its significance however, has been limited to the 1950s and 1960s models.

- HDI. The level of well being has only been included in two models, the 1980s and 1990s model. The correlation with turnout was positive in both decades but only significant in the 1990s model. Hence, Norris (2004) was right about HDI's impact on turnout, but her finding came out significant only because she studied turnout in the 1990s; HDI's impact on turnout is time-specific.

- PR and PLUMAJ. From the literature review chapter, we know that elections taking place in PR-system are associated with higher turnout and elections taking place in PLUMAJ-systems are associated with lower turnout. Both my PR and PLUMAJ variables correlate according to theory, positively for PR and negatively for PLUMAJ, but only PR displays a significant correlation, and only for the 1970s model. Hence, Jackman (1987) was right about the impact that electoral disproportionality (the main criteria behind my PR-variable) has on voter turnout, but only in his 1970s model. I have also speculated on the possibility that my electoral system variables are suffering as a result of my sample selection criteria (ACLP), which allows me to incorporate both established and non-established democracies. From the literature review chapter we know that PR's impact on turnout in established democracies is undeniable. If one takes non-established democracies into account as well, my regression analyses show that PR's impact on voter turnout is time-specific.

- UNICAM. If the legislative body consists of only one house, this will increase voter turnout according to Jackman (1987). My regression analyses support this assumption but the correlation is only significant for the 1970s and the 1980s models.

6.3.3 What are the implications of these findings?

The most interesting result that has emerged from the time-sensitive regression analyses conducted above is the fact that a number of the variables that were thought to be important determinants of cross-national variation in voter turnout *generally,* are in fact time-specific. These variables, which include LEGEFF, NUMPART, POPDENS, CAT, LITERACY, HDI, PR/PLUMAJ and UNICAM, are still important determinants of voter turnout, albeit their impact is time-specific. This means that one has to take the time dimension into account whenever discussing the variable's impact on turnout.

As much as the comprehensive research strategy implemented in this paper has "discredited" a number of variables, it has also reinforced other variables. A number of variables, including COMP, DISTRICT, GDP, NEWSPAPER, INFO and GINI, have proved to be relatively robust and consistent in their effects across time. Based on my study, we know that turnout is likely to be highest in rich countries which implement compulsory voting, where the distribution of income is more equal, the district size small, and the flow of information is at a relatively high level. Furthermore, we know these variables matter, irrespective of time.

VII. Conclusion

This thesis was motivated by the desire to understand and explain variation in voter turnout in democratic regimes cross-nationally. I set out to answer a question that has been receiving increasing interest over the last quarter of a century, namely: Why do countries vary in terms of voter turnout? By utilizing cross-sectional data and the OLS-method, this thesis has provided some interesting answers. Some of them are in line with previous research; others seem to indicate that previous research has been a bit hasty in its conclusions, especially with respect to generalizations across time and space.

This concluding chapter has two goals. First, it provides an overview of the findings the thesis has produced. Second, it considers how the thesis can serve as a basis for future research on cross national variation in voter turnout.

7.1 Cross-national variation in voter turnout

Following *Chapter 1*, where I presented the research question and my overall research strategy, *Chapter 2* outlined previous research on the topic by focusing on seven studies that represent major contributions to this field of research, presenting the key explanatory variables, and organizing them in blocs. *Chapter 3* proceeded to include the findings from the literature alongside new variables created specifically for this thesis, and derived hypotheses about each variable's anticipated association with voter turnout. *Chapter 4* outlined my methodological approach and research strategy.

By running extensive preliminary analyses alongside a comprehensive step-by-step model-building procedure I proposed, in *Chapter 5*, a model that explains cross-national variation in voter turnout for the time period between 1950 and 2000. A country that is heavily populated, enjoys compulsory voting, multipartyism and relatively small district magnitude, is non-catholic, and where the flow of information is at a relatively high level can expect an increased turnout at elections.

The time-sensitive models however, conducted on the basis of my Decade Dataset and presented in *Chapter 6*, show that the majority of determinants behind cross-national variation in voter turnout are indeed time-specific. By conducting regression analyses across shorter

time periods (decades), I was able to show that even though some variables (such as compulsory voting and level of economic development) are more robust across time than others, the majority are dependent upon time. This is quite interesting and has implications for how we think regarding determinants of cross-national variation in voter turnout. The risk of generating unreliable and non-robust findings is very high when we simply study (as the majority of research on variation in voter turnout has done) turnout variation across a limited number of countries based on shorter time periods, and then proceed to use these findings to generalize across time and space. Only by conducting comprehensive research strategies, like the time-sensitive regression analyses utilized in this thesis, can we assess each explanatory variable's impact on the variation in turnout that we observe across regimes.

7.2 Contributions of this thesis

By conducting a comprehensive step-by-step model building for the general model, and then estimating and comparing a variety of time-specific models, this paper has shown that the majority of determinants behind cross-national variation in voter turnout really are time-specific. What are the implications of my findings?

For one, these findings mean that Jackman (1987) was "right" in his findings. He was right because he studied turnout in the 1970s; but it also means that Jackman was right *only* for the 1970s model. Variables that mattered in the 1970s do not necessarily matter in the 1990s. A case in point is unicameralism. According to Jackman, unicameralism has a strong and significant impact on turnout. My finding supports this, but *our* common finding regarding the effects of unicameralism cannot be generalized: the effect of unicameralism is *time-specific*.

Furthermore, Norris (2004) was "right" about how societal development, measured by the HDI-variable, has a positive impact on turnout. She was right, because she studied variation in turnout in the 1990s. But, again, HDI's impact on turnout is time-specific and, according to my study, limited to the 1990s.

We can not discuss what determines cross-national variation in voter turnout without taking the time dimension into account. My extensive analyses have shown that the majority of determinants behind cross-national variation in voter turnout are in fact time-specific. Only a

very few determinants, like compulsory voting, economic development, flow of information and equality in the distribution of income can be said to systematically influence variation in turnout across countries, irrespective of time.

7.3 Suggestions for future research

Whereas some steps have been taken to answer the research question(s) put forth in this thesis, there are related questions that this thesis does not address as well as questions raised as a result of the conclusions reached by this thesis. Although this thesis has adopted a rather involved research methodology alongside a comprehensive model-building procedure in order to answer the research question(s), the time and scope of the thesis rendered necessary the omission of a more detailed approach to the topic.

For example, research on cross-national variation in voter turnout in Latin America indicates that other variables explain the variation in voter turnout for that region, as compared to Europe (Perez-Linan 2001 and Fornos et al. 2004). Does this hold across time as well? What about other regions? By constructing datasets that take into account the different regions one can easily check for region-specific variance behind voter turnout, also across time.

It is also important that we move beyond established democracies ("Western world") and check whether the patterns we observe among them hold in new democracies (Eastern Europe, Latin America, Africa, and Asia). My study indicates that, for instance, the electoral system has an impact on cross-national variation in voter turnout *only* when established democracies are taken into account. If some variables, like electoral system, appear to have an impact only in some subset of countries, we have to develop a more complex theory which explains when and where they matter more or less.

Furthermore, the conclusion by this thesis, that the majority of determinants behind cross-national variation in voter turnout are time-specific, raises an interesting follow up question, namely: *Why* are they time-specific?

As noted above, institutions seem to matter more than socio-economic variables for explaining cross-national variation in voter turnout. Jackman (1987) noted the same 20 years ago, but further research is needed in support of these assumptions. If institutions matter more

than socio-economic variables, we need to check whether or not this is a time-specific and/or a region-specific phenomenon and develop a theory accordingly.

Finally, with the advent of datasets such as the Comparative Study of Electoral Systems (CSES), utilized by Norris (2004), it is possible to examine the conditional impact of institutions on different types of voters. Multilevel analyses which link institutional variables with individual voter characteristics are needed because, as a result of an important habit component in voting, contextual factors have a greater impact on new cohorts. This opens up a fascinating avenue of research.

This thesis has, through its comprehensive research strategy, employed a new approach to contribute to our knowledge of why countries vary in electoral turnout. The time-sensitive regression models can be used as a basis for future research perhaps combined with a region-specific approach and one that employs a multi-level analysis. My basis for sample selection, which allowed me to compare 90 countries across 50 years, is also a fruitful approach to follow. Only through the inclusion of as many cases as possible (both established and nonestablished democracies), and by running regression models that are sensitive to possible coefficient variation over time, can any general findings be generalized across time and space. With the abovementioned suggestions for future research, the approach and structure of this thesis can be used as a baseline model for a more comprehensive exploration of the determinants behind cross-national variation in voter turnout. The search for new ways to approach the study of cross-national variation in voter turnout does not end here.

VIII. Bibliography

8.1 Books and articles

Almond, Gabriel, and Sidney Verba. 1963. *The Civic Culture: Political Attitudes and Democracy in Five Nations*. Princeton, Princeton University Press.

Black, J. 1991. Reforming the context of the voting process in Canada: Lessons from other democracies. In H. Bakvis (ed.), *Voter turnout in Canada*. Toronto: Dundurn Press.

Blais A. 2000. *To Vote or Not to Vote? The Merits and Limits of Rational Choice*. Pittsburgh: University Pittsburgh Press.

—–——. 2006. What effects voter turnout? *Annual Review Political Science 9*.

Blais A. Gidengil E, Nevitte N, Nadeau R. 2004. Where does turnout decline come from*?* *European Journal of Political Research* 43.

Blais A, Aarts K. 2005. *Electoral system and turnout*. Presented at Int. Expert Meet. "Changing the Electoral System: The Case of Netherlands", Amsterdam, Sept. 14-15.

Blais, Andre and Ken Carty. 1990. Does proportional representation foster voter turnout? *European Journal of Political Research* vol. 18: pp. 167-81.

Blais, Andre and Agnieszka Dobrzynska. 1998. Turnout in electoral democracies, *European Journal of Political Research* vol 33: 239–61.

Crewe I. 1981. Electoral participation. In D. Butler, H.R. Penniman & A. Renney (eds.), *Democracy at the polls: A comparative study of competitive national elections.* Washington DC: American Enterprise Institute.

Dahl, R., 1956. *A Preface to Democratic Theory*, Chicago: Univ. Chicago Press.

Downs, A. 1957. *An Economic Theory of Democracy*. New York: Harper.

Economist, The. *Pocket World in Figures* - 2004 edition

Fornos CA, Power TJ, Garand JC. 2004. *Explaining voter turnout in Latin America, 1980 to 2000*. Comp. Polit. Stud. 37(8):909-40.

Franklin Mark N. 1996. Electoral participation. In *Comparing Democracies: Elections and Voting in Global Perspective*, ed. L Leduc, RG Niemi, P Norris, pp. 216-35. Beverly Hills, CA: Sage.

—–——. 2004. *Voter Turnout and the Dynamics of Electoral Competition in Established Democracies Since 1945*. Cambridge. UK: Cambridge University Press.

Gujarati, Damodar N. 2003. *Basic Econometrics*, forth edition. New York. McGraw-Hill Higher Education.

Haynes, Jeff. 2001. *Democracy in the Developing World*. Oxford, UK. Blackwell Pub.

Jackman, Robert. 1987. *Political institutions and voter turnout in industrial democracies*, American Political Science Review, vol. 81, no. 2.

Jackman, Robert and Ross A. Miller. 1995. *Voter turnout in the industrial democracies during the 1980s*. Comp. Polit. Stud. 27(4):467-92.

Katz, Richard. 1996. Party organization and finance. In *Comparing Democracies: Elections and Voting in Global Perspective*, ed. L Leduc, R. G. Niemi and Pippa Norris. Beverly Hills, CA: Sage.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing social inquiry. Scientific Inference in Qualitative Research*. Princeton. Princeton University Press.

Kostadinova T. 2003. Voter turnout dynamics in post-Communist Europe. *European Journal of Political Research* 42(6):741-59.

Krieckhaus, J. 2004. *The Regime Debate Revisited: A Sensitivity Analysis of Democracy's Economic Effect*. B.J.Pol.S 34, 635-655.

—————. 2005. *Democracy and Economic Growth: How Regional Context Influences Regime Effects,* B.J.Pol.S 35.

Lewis-Beck, Michael S. 1980. *Applied regression: An introduction*. Sage University Paper Series on Quantiative Applications in the Social Sciences, 07-022. Newbury Park, CA: Sage.

Lieberson, Stanley. 1985. *Making it count: The Improvement of Social Research and Theory*. Berkeley: University of California Press.

Mackie, Thomas T. and Richard Rose. 1991. *International Almanac of Electoral History*. Third edition. Washington, DC: CQ Press.

Milbrath, L. W. 1965. *Political Participation*, Chicago, Rand McNally.

Munck, Gerardo L. and Jay Verkuilen, 2002. Conceptualizing and Measuring Democracy: Evaluating Alternative Indices, *Comparative Political Studies* 35,1 (February) 5-34.

Norris, Pippa. 2001. Democratic Phoenix: Political Activism Worldwide Chapter 3.

—————. 2004. *Electoral Engineering*, Cambridge University Press, New York.

Pateman, Carole. 1970. *Participation and democratic theory*, Cambridge University Press.

Pennings, Paul, Hans Keman and Jan Kleinnijenhuis. 2006. *Doing Research in Political Science*, second edition. London. Sage Publications.

Perez-Linan A. 2001. Neoinstitutional accounts of voter turnout: moving beyond industrial democracies. *Elect. Stud*. 20(2):281-97.

Poe, Steven, Neal Tate and Linda Camp Keith. 1999. Repression of Human Rights to Personal Integrity Revisited: A Global Cross-National Study Covering the Years 1976-93. *International Studies Quarterly* **43**: 291-313.

Powell, G. Bingham, Jr. 1980. Voting Turnout in Thirty Democracies: Partican, Legal, and Socio-Economic Influences. In R, Rose (ed.), *Electoral Participation*, SAGE publications, London.

―――.1982. *Contemporary Democracies: Participation, Stability, and Violence*. Cambridge, Harvard University Press.

―――. 1986. American voter turnout in comparative perspective, *American Political Science Review*, vol. 80, no. 1.

Przeworski A., S. Stokes and B. Manin. 1999. *Democracy, Accountability, and Representation*. Cambridge University Press.

Przeworski, Adam, Michael Alvarez, Jose Antonio Cheibub, and Fernando Limongi. 2000. *Democracy and development*, New York: Cambridge University Press.

Pye, Lucian W., and Sidney Verba 1965. *Political Culture and Political Development*. Princeton University Press, New Jersey.

Radcliff B, Davis P. 2000. Labour organization and electoral participation in industrial democracies. *American Political Science Review* 44.

Rae, Douglas. in "A Note on the Fractionalization of Some European Party Systems", *Comparative Political Studies*, 1 (October 1968), 413-418.

Ragin, Charles. 1987. *The Comparative Method. Moving Beyond Qualitative and Quantitative Strategies*. University of California Press. Berkeley Los Angeles.

Ragin, Charles & David Zaret. 1983. "Theory and Method in Comparative Research: Two Strategies". *Social Forces* 61(3): 731-754.

Reif, K. and H. Schmitt. 1980. Nine Second-order National Elections: A Conceptual Framework for the Analysis of the European Election Results. *European Journal of Political Research*. 8:3-44.

Reynolds, Andrew and Ben Reilly. 1997. *The international IDEA Handbook of Electoral System Design.* Stockholm.

Ringdal, Kristen. 2001. *Enhet og mangfold: Samfunnsvitenskapelig forskning og kvantitativ metode*. Bergen, Fagbokforlaget.

Skog, Ole Jørgen. 2004. *Å forklare sosiale fenomener: En regresjonsbasert tilnærming*. Oslo. Gyldendal Norsk Forlag.

Tingsten, H. 1975. *Political Behaviour: studies in election statistics*. New York: Arno Press.

Van der Eijk, Cees, and Mark Franklin. 1996. *Choosing Europe*. University of Michigan Press.

Verba, Sidney, Norman H. Nie, and Jao-on Kim. 1978. *Participation and Political Equality: A Seven-Nation Comparison*. New York: Cambridge University Press.

Verba, Sidney, Kay Lehman Schlozman and Henry E. Brady. 1995. *Voice and equality: Civic voluntarism in American Politics*. Cambridge, MA: Harvard University Press.

8.2 Data sources

IDEA database
Available online at www.idea.int (01.06.07)

Inter-Parliamentary Union
Available online at: www.ipu.org/english/home.htm (01.06.07)

Penn world tables
PWT 5.6 The Penn World Table (Mark 5.6)--denoted PWT 5.6-- is a revised and updated version of the proceeding (Mark 5) version that was described in "The Penn World Table (Mark5): An Expanded Set of International Comparisons, 1950-1988" by Robert Summers and Alan Heston, Quarterly Journal of Economics, May 1991.

PWT6.2 Alan Heston, Robert Summers and Bettina Aten, Penn World Table Version 6.2, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, September 2006.

Banks dataset
Cross-national time-series data archive : Arthur S. Banks [electronic resource]. - Binghamton, N.Y. : Databanks International. 2004

Beck's Database on Political Institutions
Thorsten Beck, George Clarke, Alberto Groff, Philip Keefer, and Patrick Walsh, 2001. "New tools in comparative political economy: The Database of Political Institutions." 15:1, 165-176 (September), World Bank Economic Review.

IX. Appendixes

9.1 Appendix A – Countries with compulsory voting laws

Table 16: *Countries with compulsory voting laws*

Code	Country	Year*
Latin America & the Caribbean		
02	Argentina	
07	Brazil	
08	Chile	From 1962 onwards
10	Costa Rica	
18	Honduras	
22	Peru	
28	Uruguay	From 1970 onwards
29	Venezuela	Until 1993
Europe		
33	Belgium	
42	Greece	
46	Italy	
50	Luxembourg	
53	Netherland	Until 1970
Other		
66	Australia	

* Entire sample period if empty

9.2 Appendix B – Variable coding

Numpart
Number of effective parties is based on the following:
Field S20F5 (code utilized in Banks dataset) is a party fractionalization index, based on the formula proposed by Douglas Rae in "A Note on the Fractionalization of Some European Party Systems", Comparative Political Studies, 1 (October 1968), 413-418. The index is constructed as follows:

$$F = 1 - \sum_{i=1}^{m} (ti)2$$

where ti = the proportion of members associated with the ith party in the lower house of the legislature.

WCI
The Weighted Conflict Index is based on the following factors (including code from Banks dataset):
S17F1 Assassinations. Assassinations. Any politically motivated murder or attempted murder of a high government official or politician.
S17F2 General Strikes. Any strike of 1,000 or more industrial or service workers that involves more than one employer and that is aimed at national government policies or authority.
S17F3 Guerrilla Warfare. Any armed activity, sabotage, or bombings carried on by independent bands of citizens or irregular forces and aimed at the overthrow of the present regime.
S17F4 Government Crises. Any rapidly developing situation that threatens to bring the downfall of the present regime - excluding situations of revolt aimed at such overthrow.
S17F5 Purges. Any systematic elimination by jailing or execution of political opposition within the ranks of the regime or the opposition.
S17F6 Riots. Any violent demonstration or clash of more than 100 citizens involving the use of physical force.
S17F7 Revolutions. Any illegal or forced change in the top governmental elite, any attempt at such a change, or any successful or unsuccessful armed rebellion whose aim is independence from the central government.
S18F1 Anti-Government Demonstrations. Any peaceful public gathering of at least 100 people for the primary purpose of displaying or voicing their opposition to government policies or authority, excluding demonstrations of a distinctly anti-foreign nature.

The Weighted Conflict Index is calculated in the following manner:
Multiply the value of the number of Assassinations by 24,
General Strikes by 43,
Guerrilla Warfare by 46,
Government Crises by 48,
Purges by 86,
Riots by 102,
Revolutions by 148,
Anti-Government Demonstrations by 200.

Sum the 8 weighted values and divide by 9. The result is the value (with decimal) stored as the Weighted Conflict Index.

9.3 Appendix C – Assumptions of the OLS

9.3.1 Assumptions of OLS estimation method[48]

1. No specification error
    a. The relationship between X_i and Y_i is linear.
    b. No relevant independent variables have been excluded.
    c. No irrelevant independent variables have been included.
2. No measurement error
    a. The variables X and Y are accurately measured.
3. The following assumptions concern the error term, ε_i:
    a. Zero mean: E(ε_i) = 0
        i. For each observation, the expected value of the error term is zero.
    b. Homoscedasticity
        i. The variance of the error term is constant for all values of X_i.
    c. No autocorrelation
        i. The error terms are uncorrelated.
    d. The independent variable is uncorrelated with the error term.
    e. Normality.
        i. The error term, ε_i, is normally distributed.
4. No multicollinearity
    a. Absence of perfect multicollinearity.

9.3.2 About the assumptions

When assumption one through 3d are met in a bivariate regression, desirable estimators of the population parameters, α and β, will be obtained. Technically speaking, they will be the "best linear unbiased estimates", BLUE. However, for a multiple regression to produce BLUE, it must meet the bivariate regression assumptions, plus assumption four; absence of perfect multicollinearity. Let us examine the assumptions in more detail.

The absence of specification error is critical as it asserts that the theoretical model embodied in the equation is correct. The assumption of linearity asserts that the functional form of the relationship is actually a straight line. As with the exclusion and inclusion of independent variables, the logic is self-evident. If we wish to arrive at a more complete explanation of a give social phenomenon it is important to include the necessary explanatory variables, and exclude the irrelevant variables, in the equation. To evaluate the relevance of each variable, we perform tests for statistical significance.

[48] From Lewis-Beck (1980).

The second assumption deals with the issue of measurement errors. It goes without saying that for us to arrive at accurate estimates, our variables must be accurately measured. Validity and reliability of the measurement is essential in this regard. In my paper I have addressed the issue of reliability and validity in chapter 3.

The third set of assumptions involves the error term. A zero mean of the error term is not critical as it will only bias the intercept estimate. As this estimate is of secondary interest in social science research, and as long as the least square estimate of the slope is unchanged, we need not worry too much about this assumption (Lewis-Beck 1980: 28). The assumption of homoscedasticity (or absence of perfect heteroscedasticity) has been discussed above (both for the General Model and the 1990s model) and will hence not be addressed here. I will however remind the reader again, that I run my regression models implementing the White Estimator.

Autocorrelation may be defined as "correlation between members of series of observations ordered in time (as in time series data) or space (as in cross-sectional data)" (Gujarati 2003: 442). Autocorrelation, which appears more frequently with time-series variables than with cross-sectional variables, invalidates the significance tests, by indicating significance when in fact that's not the case. The least square estimates, however, are unbiased (Lewis-Beck 1980:28).

The next assumption, that the independent variable is uncorrelated with the error term, can be difficult to meet in non-experimental research. If our X, which we observe in society and can only account for, is related to the error term, than the least squares parameter estimates will be biased. As the error term really is a collection of excluded variables, the remedy would be an incorporation of missing explanatory variables into the model. I feel confident that my thesis includes the majority of the important explanatory variables. The last assumption regarding the error term is that it is normally distributed. We need only worry about the distribution of Y_i as it is the same as the distribution of ε_i (only their means are different) (Lewis-Beck 1980: 29). Figure 3 and 4 below present some evidence that the residuals in my two main models are normally distributed.

Figure 3: *Residuals – General model*

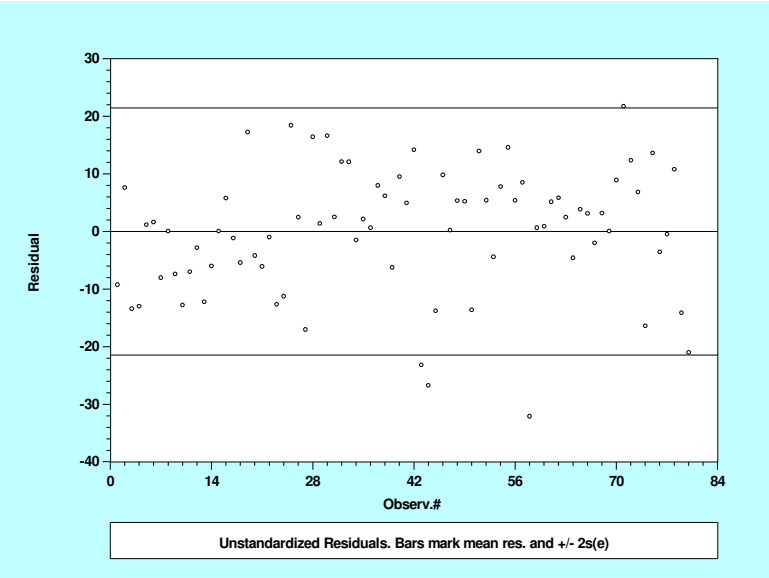
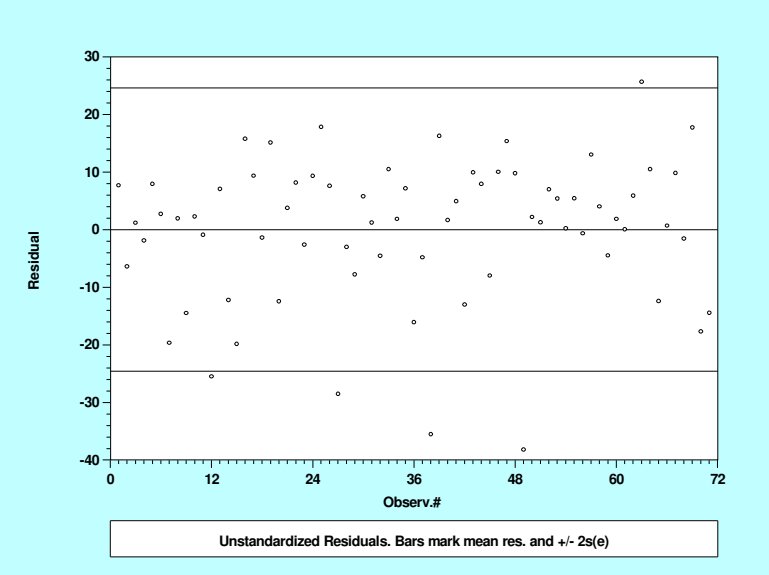


Figure 4: *Residuals – 1990s model*



The final assumption of the ordinary least squares estimation method is the absence of perfect multicollinearity. I will not discuss it here as I discussed it thoroughly in previous chapters.

9.4. Appendix D – Sample selection

9.4.1 Democratic elections and turnout rate – cases in my study

Table 17: *Democratic elections and turnout rate*

Country and election year	Turnout rate (%)	Country and election year	Turnout rate (%)	Country and election year	Turnout rate (%)
Latin America & the Caribbean					
Antigua and Barbuda 1980	60,6	Chile 1957	22,0	Ecuador 1979	42,5
Antigua and Barbuda 1984	49,7	Chile 1961	32,1	Ecuador 1984	56,1
Antigua and Barbuda 1989	72,5	Chile 1965	50,0	Ecuador 1986	64,9
Antigua and Barbuda 1994	85,6	Chile 1969	45,7	Ecuador 1988	67,0
Antigua and Barbuda 1999	72,3	Chile 1993	81,9	Ecuador 1990	64,7
Argentina 1983	77,5	Chile 1997	73,1	Ecuador 1994	66,3
Argentina 1985	77,8	Colombia 1958	65,0	Ecuador 1996	67,8
Argentina 1987	80,1	Colombia 1960	42,8	Ecuador 1998	48,5
Argentina 1989	82,1	Colombia 1962	49,8	El Sal.1985	48,3
Argentina 1991	89,4	Colombia 1964	35,7	El Sal.1988	43,5
Argentina 1993	78,1	Colombia 1966	38,5	El Sal.1991	44,0
Argentina 1995	79,8	Colombia 1968	30,71	El Sal.1994	51,5
Argentina 1998	78,1	Colombia 1970	46,0	El Sal.1997	82,2
Argentina 1999	79,4	Colombia 1974	48,4	Grenada 1984	68,8
Bahamas 1977	61,4	Colombia 1978	30,81	Grenada 1990	78,9
Bahamas 1982	64,2	Colombia 1982	38,7	Grenada 1995	84,0
Bahamas 1987	63,5	Colombia 1986	41,1	Grenada 1999	80,1
Bahamas 1992	68,5	Colombia 1990	40,0	Guatemala 1966	23,71
Bahamas 1997	67,9	Colombia 1991	25,61	Guatemala 1970	25,91
Barbados 1966	57,6	Colombia 1994	29,21	Guatemala 1974	25,61
Barbados 1971	70,0	Colombia 1998	40,5	Guatemala 1978	22,91
Barbados 1976	66,3	Costa Rica 1953	50,0	Guatemala 1990	41,0
Barbados 1981	74,7	Costa Rica 1958	49,6	Guatemala 1994	14,51
Barbados 1986	80,0	Costa Rica 1962	71,5	Guatemala 1995	33,4
Barbados 1991	67,4	Costa Rica 1966	69,8	Guatemala 1999	31,13
Barbados 1994	66,0	Costa Rica 1970	75,7	Guyana 1992	63,7
Barbados 1999	68,6	Costa Rica 1974	71,4	Guyana 1997	80,2
Belize 1984	65,9	Costa Rica 1978	75,2	Honduras 1985	77,8
Belize 1989	66,9	Costa Rica 1982	79,0	Honduras 1989	75,7
Belize 1993	68,7	Costa Rica 1986	74,7	Honduras 1993	63,5
Belize 1998	65,7	Costa Rica 1990	85,1	Honduras 1997	68,0
Bolivia 1985	65,2	Costa Rica 1994	84,2	Jamaica 1962	73,7
Bolivia 1989	51,0	Costa Rica 1998	73,7	Jamaica 1967	54,9
Bolivia 1993	50,0	Dominica 1980	80,9	Jamaica 1972	57,3
Bolivia 1997	64,5	Dominica 1985	74,7	Jamaica 1976	84,8
Brazil 1950	28,31	Dominica 1990	78,8	Jamaica 1980	74,7
Brazil 1954	31,51	Dominica 1995	82,4	Jamaica 1989	59,0
Brazil 1958	35,7	Dom. Rep. 1966	80,8	Jamaica 1993	44,7
Brazil 1962	36,9	Dom. Rep. 1970	66,3	Jamaica 1997	48,8
Brazil 1982	63,7	Dom. Rep. 1986	60,7	Nicaragua 1984	74,1
Brazil 1986	70,4	Dom. Rep. 1990	45,6	Nicaragua 1990	73,3
Brazil 1990	76,6	Dom. Rep. 1994	30,61	Nicaragua 1996	76,2
Brazil 1994	79,9	Dom. Rep. 1996	62,1	Panama 1960	23,13
Brazil 1998	81,0	Dom. Rep. 1998	45,6	Panama 1994	70,1
Chile 1953	22,91	Ecuador 1950	16,91	Panama 1999	76,1
		Ecuador 1952	19,21	Peru 1980	58,0
		Ecuador 1954	25,61	Peru 1985	64,8
		Ecuador 1956	30,71	St. Lucia 1979	68,0
		Ecuador 1958	24,71	St. Lucia 1982	65,8
		Ecuador 1962	31,11	St. Lucia 1987	67,5

Appendix D (continued)

Country and election year	Turnout rate (%)	Country and election year	Turnout rate (%)	Country and election year	Turnout rate (%)
Latin America & the Caribbean (continued)					
St. Lucia 1987	72,0	Venezuela 1988	72,7	Denmark 1950	76,6
St. Lucia 1992	76,8	Venezuela 1993	49,9	Denmark 1953	75,5
St. Lucia 1997	89,5	Venezuela 1998	42,7	Denmark 1953	78,9
St. Kitts &				Denmark 1957	82,1
Nevis 1984	71,7	*Europe*		Denmark 1960	85,9
St. Kitts &				Denmark 1964	84,8
Nevis 1989	66	Albania 1991	98,9	Denmark 1966	88,5
St. Kitts &		Albania 1992	89,2	Denmark 1968	88,0
Nevis 1995	69,6	Albania 1996	89,5	Denmark 1971	87,3
St. Vin. &		Albania 1997	63,4	Denmark 1973	87,4
the Gren. 1979	58,3	Andorra 1993	80,9	Denmark 1975	86,6
St. Vin. &		Andorra 1997	81,6	Denmark 1977	87,7
the Gren. 1984	80,5	Austria 1953	89,0	Denmark 1979	83,3
St. Vin. &		Austria 1956	89,3	Denmark 1981	86,3
the Gren. 1989	76,8	Austria 1959	89,7	Denmark 1984	86,0
St. Vin. &		Austria 1962	90,3	Denmark 1987	85,9
the Gren. 1994	73,2	Austria 1966	89,8	Denmark 1988	82,7
St. Vin. &		Austria 1970	89,5	Denmark 1990	80,4
the Gren. 1998	77,1	Austria 1971	88,2	Denmark 1994	81,7
Suriname 1991	67,8	Austria 1975	86,6	Denmark 1998	83,1
Suriname 1996	71,2	Austria 1979	86,8	Estonia 1992	40,9
Trinidad &		Austria 1983	86,9	Estonia 1995	48,8
Tob. 1966	67,9	Austria 1986	87,1	Estonia 1999	46,0
Trinidad &		Austria 1991	80,5	Finland 1951	74,0
Tob. 1971	25,61	Austria 1994	76,0	Finland 1954	80,3
Trinidad &		Austria 1995	78,6	Finland 1958	74,7
Tob. 1976	64,3	Austria 1999	72,6	Finland 1962	85,5
Trinidad &		Belgium 1950	86,3	Finland 1966	85,1
Tob. 1981	64,4	Belgium 1954	88,5	Finland 1970	83,6
Trinidad &		Belgium 1958	89,3	Finland 1972	84,5
Tob. 1986	80,4	Belgium 1961	88,0	Finland 1975	80,3
Trinidad &		Belgium 1965	88,0	Finland 1979	81,3
Tob. 1991	70,4	Belgium 1968	86,2	Finland 1983	81,1
Trinidad &		Belgium 1971	88,6	Finland 1987	77,2
Tob. 1995	67,3	Belgium 1974	85,7	Finland 1991	71,9
Uruguay 1950	51,4	Belgium 1977	89,9	Finland 1995	71,1
Uruguay 1954	50,9	Belgium 1978	87,8	Finland 1999	65,2
Uruguay 1958	54,3	Belgium 1981	94,3	France 1951	68,5
Uruguay 1962	60	Belgium 1985	86,3	France 1956	74,3
Uruguay 1966	66,9	Belgium 1987	86,5	France 1958	71,1
Uruguay 1971	88,2	Belgium 1991	85,1	France 1962	61,0
Uruguay 1989	96,9	Belgium 1995	83,2	France 1967	71,1
Uruguay 1994	96,1	Belgium 1999	83,2	France 1968	69,4
Uruguay 1999	94,6	Croatia 1992	74,9	France 1973	70,6
Venezuela 1963	78,3	Croatia 1995	72,2	France 1978	63,4
Venezuela 1968	84,6	Czechoslovakia 1990	93,1	France 1981	63,9
Venezuela 1973	81,1	Czechoslovakia 1992	83,8	France 1986	69,9
Venezuela 1978	74,5	Czech Rep. 1996	77,6	France 1988	58,1
Venezuela 1983	77,1	Czech Rep. 1998	76,7	France 1993	61,3

Appendix D (continued)

Country and election year	Turnout rate (%)	Country and election year	Turnout rate (%)	Country and election year	Turnout rate (%)
Europe (continued)					
France 1997	59,9	Ireland 1965	74,4	Netherland 1952	86,9
Germany 1953	80,6	Ireland 1969	77,2	Netherland 1956	88,1
Germany 1957	87,6	Ireland 1973	77,8	Netherland 1959	88,8
Germany 1961	86,9	Ireland 1977	85,2	Netherland 1963	87,9
Germany 1965	80,9	Ireland 1981	78,8	Netherland 1967	92,1
Germany 1969	79,9	Ireland 1982	76,3	Netherland 1971	77,8
Germany 1972	88,7	Ireland 1987	77,9	Netherland 1972	90,1
Germany 1976	83,8	Ireland 1989	71,2	Netherland 1977	86,3
Germany 1980	81,8	Ireland 1992	73,7	Netherland 1981	85,2
Germany 1983	81,0	Ireland 1997	66,7	Netherland 1982	80,3
Germany 1987	75,0	Italy 1953	91,5	Netherland 1986	84,2
Germany 1990	73,1	Italy 1958	93,9	Netherland 1989	78,0
Germany 1994	72,4	Italy 1963	95,3	Netherland 1994	75,2
Germany 1998	75,3	Italy 1968	93,2	Netherland 1998	70,1
Greece 1950	80,1	Italy 1972	94,7	Norway 1953	79,6
Greece 1951	80,2	Italy 1976	95,4	Norway 1957	76,9
Greece 1952	73,9	Italy 1979	92,7	Norway 1961	77,7
Greece 1956	67,9	Italy 1983	91,8	Norway 1965	85,0
Greece 1958	73,9	Italy 1987	94,4	Norway 1969	85,1
Greece 1961	86,3	Italy 1992	92,3	Norway 1973	80,1
Greece 1963	85,3	Italy 1994	90,8	Norway 1977	79,2
Greece 1964	83,6	Italy 1996	87,3	Norway 1981	82,3
Greece 1977	82,4	Latvia 1993	57,7	Norway 1985	83,6
Greece 1981	84,5	Latvia 1995	50,6	Norway 1989	81,5
Greece 1985	87,4	Latvia 1998	51,9	Norway 1993	74,5
Greece 1989	85,8	Lichtenstein 1993	55,6	Norway 1997	76,9
Greece 1989	87,5	Lichtenstein 1997	53,8	Poland 1989	44,4
Greece 1993	85,6	Lithuania 1993	70,2	Poland 1991	52,0
Greece 1996	83,9	Lithuania 1997	50,0	Poland 1993	48,8
Hungary 1994	69,4	Luxembourg 1951	37,3	Poland 1997	47,6
Hungary 1998	59,0	Luxembourg 1954	73,1	Portugal 1976	83,3
Iceland 1953	92,1	Luxembourg 1959	74,3	Portugal 1979	88,2
Iceland 1956	91,9	Luxembourg 1964	72,5	Portugal 1980	87,9
Iceland 1959	89,4	Luxembourg 1968	67,8	Portugal 1983	77,4
Iceland 1959	89,7	Luxembourg 1974	73,7	Portugal 1985	79,7
Iceland 1963	89,4	Luxembourg 1979	68,5	Portugal 1987	78,1
Iceland 1967	89,0	Luxembourg 1984	66,8	Portugal 1991	77,7
Iceland 1971	89,4	Luxembourg 1989	64,1	Portugal 1995	79,1
Iceland 1974	87,6	Luxembourg 1994	60,5	Portugal 1999	69,3
Iceland 1978	89,6	Luxembourg 1999	56,9	Romania 1992	76,2
Iceland 1979	90,6	Macedonia 1994	47,8	Romania 1996	78,2
Iceland 1983	88,2	Macedonia 1999	48,9	Russia 1993	47,2
Iceland 1987	92,3	Malta 1966	76,7	Russia 1995	62,8
Iceland 1991	88,7	Malta 1971	77,6	Russia 1999	59,9
Iceland 1995	87,8	Malta 1976	89,8	Slovakia 1994	75,9
Iceland 1999	86,2	Malta 1981	86,0	Slovakia 1998	78,9
Ireland 1951	74,4	Malta 1987	95,6	Slovenia 1992	85,5
Ireland 1954	75,3	Malta 1992	95,3	Slovenia 1996	75,7
Ireland 1957	70,4	Malta 1996	98,0	Spain 1977	79,4
Ireland 1961	71,0	Malta 1998	95,9	Spain 1979	72,3

Appendix D (continued)

Country and election year	Turnout rate (%)	Country and election year	Turnout rate (%)	Country and election year	Turnout rate (%)
Europe (continued)					
		Other			
Spain 1982	83,1				
Spain 1986	73,7	Australia 1951	84,9	India 1980	61,9
Spain 1989	71,3	Australia 1954	81,6	India 1984	64,5
Spain 1993	77,4	Australia 1955	78,1	India 1989	61,2
Spain 1996	80,6	Australia 1958	85,6	India 1991	57,2
Sweden 1952	77,3	Australia 1961	84,0	India 1996	61,1
Sweden 1956	77,3	Australia 1963	85,1	India 1998	67,1
Sweden 1958	75,6	Australia 1966	84,7	India 1999	65,5
Sweden 1960	82,8	Australia 1969	83,9	Israel 1951	72,8
Sweden 1964	80,8	Australia 1972	85,3	Israel 1955	82,2
Sweden 1968	86,6	Australia 1974	84,3	Israel 1959	81,8
Sweden 1970	87,3	Australia 1975	84,7	Israel 1961	80,4
Sweden 1973	85,8	Australia 1977	84,5	Israel 1965	82,4
Sweden 1976	88,5	Australia 1980	84,0	Israel 1969	83,0
Sweden 1979	86,9	Australia 1983	81,2	Israel 1973	81,8
Sweden 1982	88,6	Australia 1984	84,2	Israel 1977	80,4
Sweden 1985	86,2	Australia 1987	84,1	Israel 1981	79,9
Sweden 1988	82,7	Australia 1990	82,1	Israel 1984	80,4
Sweden 1991	82,8	Australia 1993	83,4	Israel 1988	82,4
Sweden 1994	83,6	Australia 1996	82,5	Israel 1992	81,7
Sweden 1998	77,7	Australia 1998	81,7	Israel 1996	84,7
Switzl. 1951	63,0	Bangladesh 1991	61,5	Israel 1999	84,4
Switzl. 1955	60,8	Bangladesh 1996	64,6	Japan 1952	77,4
Switzl. 1959	58,1	Benin 1991	46,5	Japan 1953	70,7
Switzl. 1963	53,5	Benin 1995	73,7	Japan 1955	73,6
Switzl. 1967	53,2	Benin 1999	65,9	Japan 1958	72,9
Switzl. 1975	43,6	Canada 1953	65,4	Japan 1960	71,4
Switzl. 1979	40,6	Canada 1957	67,9	Japan 1963	68,6
Switzl. 1983	40,8	Canada 1958	75,4	Japan 1967	73,9
Switzl. 1987	39,9	Canada 1962	73,3	Japan 1969	73,5
Switzl. 1991	39,7	Canada 1963	75,1	Japan 1972	73,9
Switzl. 1995	35,7	Canada 1965	70,7	Japan 1976	73,5
Switzl. 1999	34,9	Canada 1968	68,3	Japan 1979	68,2
Ukraine 1994	73,4	Canada 1972	71,3	Japan 1980	74,7
Ukraine 1999	68,1	Canada 1974	64,2	Japan 1983	67,6
U K 1950	81,6	Canada 1979	68,4	Japan 1986	71,5
U K 1951	81,4	Canada 1980	64,5	Japan 1990	74,9
U K 1955	75,7	Canada 1984	67,9	Japan 1993	66,3
U K 1959	77,5	Canada 1988	68,3	Japan 1995	44,9
U K 1964	75,1	Canada 1993	63,9	Japan 1996	59,8
U K 1966	73,8	Canada 1997	57,1	Kiribati 1982	74,7
U K 1970	71,2	Cape Verde 1991	71,8	Kiribati 1983	78,0
U K 1974	77,9	Cape Verde 1995	79,5	Kiribati 1991	62,0
U K 1974	72,5	Central Af. Rep. 1993	50,3	Korea Rep. 1988	73,8
U K 1979	75,1	Central Af. Rep. 1998	53,6	Korea Rep. 1992	74,6
U K 1983	71,7	India 1952	58,9	Korea Rep. 1996	65,3
U K 1987	75,2	India 1957	62,6	Mali 1992	21,91
U K 1992	75,4	India 1962	54,4	Mali 1997	21,31
U K 1997	69,4	India 1967	63,1	Mauritius 1976	84,4
		India 1971	57,2	Mauritius 1982	91,5
		India 1977	64,6	Mauritius 1983	77,0

Appendix D (continued)

Country and election year	Turnout rate (%)	Country and election year	Turnout rate (%)	Country and election year	Turnout rate (%)
Other (continued)					
Mauritius 1987	84,6	Pap. N. Gu. 1992	71,0	USA 1956	54,9
Mauritius 1991	82,4	Pap. N. Gu. 1997	98,8	USA 1968	44,5
Mauritius 1995	77,1	Philippines 1987	78,2	USA 1960	63,1
Micron. 1995	63,0	Philippines 1992	65,3	USA 1962	47,3
Micron. 1999	29,51	Philippines 1995	68,4	USA 1964	61,9
Nepal 1991	83,6	Philippines 1998	66,8	USA 1966	48,4
Nepal 1994	82,8	Sol. Island 1980	54,7	USA 1968	60,8
Nepal 1997	83,6	Sol. Island 1984	53,0	USA 1970	46,6
Nepal 1999	75,8	Sol. Island 1989	55,8	USA 1972	55,2
New Zea. 1951	95,1	Sol. Island 1993	60,6	USA 1974	38,2
New Zea. 1954	91,1	Sol. Island 1997	70,7	USA 1976	53,5
New Zea. 1957	85,5	Sri Lanka 1952	49,4	USA 1978	37,2
New Zea. 1960	85,6	Sri Lanka 1956	51,5	USA 1980	52,6
New Zea. 1963	83,3	Sri Lanka 1960	55,0	USA 1982	39,8
New Zea. 1966	79,3	Sri Lanka 1965	65,9	USA 1984	53,1
New Zea. 1969	85,6	Sri Lanka 1970	71,8	USA 1986	36,4
New Zea. 1972	85,3	Sri Lanka 1989	58,1	USA 1988	50,1
New Zea. 1975	81,7	Sri Lanka 1994	74,1	USA 1990	36,5
New Zea. 1978	82,3	Thailand 1992	58,4	USA 1992	55,1
New Zea. 1981	88,9	Thailand 1995	64,1	USA 1994	38,8
New Zea. 1984	87,4	Thailand 1996	65,0	USA 1996	49,1
New Zea. 1987	81,4	Turkey 1961	75,1	USA 1998	34,7
New Zea. 1990	78,6	Turkey 1969	59,3	Zambia 1991	34,2
New Zea. 1993	79,6	Turkey 1973	56,7	Zambia 1996	39,8
New Zea. 1996	83,0	Turkey 1977	66,8		
New Zea. 1999	76,1	Turkey 1983	75,5		
Pakistan 1988	42,9	Turkey 1987	91,8		
Pakistan 1990	43,4	Turkey 1991	79,8		
Pakistan 1993	37,6	Turkey 1995	79,1		
Pakistan 1997	31,51	Turkey 1999	80,4		
Pap. N. Gu. 1977	66,4	USA 1950	42,6		
Pap. N. Gu. 1982	75,7	USA 1952	59,7		
Pap. N. Gu. 1987	76,3	USA 1954	43,1		

Excluded elections due to lack of data:

Country	Year
Jamaica	1983
Kiribati	1998
St. Kitts & Nevis	1993
Switzerland	1971

9.4.2 Countries and years in the literature studies

Powell (1982): *Contemporary Democracies*.
29 countries: Australia, Austria, Belgium, Canada, Ceylon (Sri Lanka), Chile, Costa Rica, Denmark, Finland, France, Greece, India, Ireland, Israel, Italy, Jamaica, Japan, Netherlands, New Zealand, Norway, Phillipines, Sweden, Switzerland, Turkey, United Kingdom, United States, Uruguay, Venezuela, West Germany,
Time period: 1960s and 70s
Powell (1986): *American voter turnout in comparative perspective*.
20 countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, West Germany, Ireland, Israel, Italy, Japan, Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, United Kingdom, and the United States.
Time period: 1960s and 70s.
Jackman (1987): *Political institutions and voter turnout in industrial democracies*.
19 countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, West Germany, Ireland, Israel, Italy, Japan, Netherlands, New Zealand, Norway, Sweden, Switzerland, United Kingdom, and the United States.
Time period: 1960s and 70s.
Blais and Carty (1990): *Does proportional representation foster turnout?*.
20 countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Iceland, Ireland, Israel, Italy, Japan, Luxembourg, Netherlands, New Zealand, Norway, Sweden, Switzerland, and the United Kingdom.
Time period: Depends, - 1985
Franklin(1996): in *Comparing democracies*.
29 countries: Australia, Austria, Belgium, Brazil, Canada, Costa Rica, Denmark, Finland, France, Germany, Greece, Iceland, India, Ireland, Israel, Italy, Japan, Luxembourg, Malta, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, United States, and Venezuela.
Time period: Unknown
Blais and Dobrzynska (1998): *Turnout in electoral democracies.*
91 countries: Andorra, Antigua & Barbuda, Argentina, Australia, Austria, Bahamas, Bangladesh, Barbados, Belarus, Belgium, Belize, Benin, Botswana, Bolivia, Brazil, Bulgaria, Burkina Faso, Canada, Cape Verde, Chile, Colombia, Costa Rica, Cyprus, Czechoslovakia, Denmark, Dominica, Dominican Republic, El Salvador, Ecuador, Fiji, Finland, France, Gambia, Germany, Grenada, Guyana, Honduras, Hungary, Iceland, India, Israel, Kiribati, Korea, Rep., Lebanon, Lichtenstein, Lithuania, Luxembourg, Jamaica, Japan, Madagascar, Malawi, Mauritius, Mexico, Monaco, Namibia, Nauru, Nepal, New Zealand, Netherlands, Nigeria, Norway, Papua New Guinea, Peru, Philippines, Poland, Portugal, Romania, Russia, Samoa, San Marino, Sao Tome & Principe, Slovenia, Solomon Islands, South Africa, Spain, St. Kitts & Nevis, St. Lucia, St. Vincent & Grenadines, Surinam, Sweden, Switzerland, Taiwan, Thailand, Trinidad & Tobago, Turkey, Zambia, Ukraine, United States, Uruguay, Venezuela, Vanuatu.
Time period: 1972-1995
Norris (2004): *Electoral engineering*.
32 countries: Australia, Belarus, Belgium, Canada, Chile, Czech Republic, Denmark, Germany, Hungary, Korea, Rep., Iceland, Israel, Lithuania, Japan, Mexico, New Zealand, Netherlands, Norway, Peru, Poland, Portugal, Romania, Russia, Slovenia, Spain, Sweden, Switzerland, Taiwan, Thailand, Ukraine, and the United States.
Time period: 1996-2002

9.5 Appendix E – CORRELATION MATRIX – GENERAL MODEL

Table 18: *Correlation Matrix – General Model (LG indicating logged variables)*

	TURNOUT	COMP	EASTERN	PR	PLUMAJ	UNICAM	CAT	FREQ
TURNOUT	1.00000							
COMP	.24548	1.00000						
EASTERN	-.06293	-.22291	1.00000					
PR	.16424	.09261	-.13096	1.00000				
PLUMAJ	-.04032	-.01707	.04663	-.56729	1.00000			
UNICAM	.09166	-.10896	-.01546	-.02247	-.20363	1.00000		
CAT	-.19684	.31763	.07357	.25900	-.21592	-.07708	1.00000	
FREQ	.06436	.04925	-.16813	.10399	-.14363	.26538	.12694	1.00000

	TURNOUT	COMP	EASTERN	PR	PLUMAJ	UNICAM	CAT	FREQ
NUMPARTC	-.10031	-.12675	.36769	-.04808	-.13253	-.16840	-.05081	-.05782
LITERACY	.36807	.05150	.22266	.37360	-.48839	.01859	-.05452	-.13423
LEGEFF	.35933	.17124	-.18468	.41465	-.12521	-.12013	-.08958	-.13344
CLOSED	.08937	-.02131	-.05404	.08463	-.33371	-.01478	.09033	.01195
DEMO	.00469	-.09476	.15670	-.41387	.07587	-.09313	-.14596	-.08386
RIOTS	-.05829	-.07706	-.03591	-.38060	.31830	-.12892	-.20690	.03630
GOVCRIS	-.07472	.15170	.02303	-.22038	.06816	-.20390	-.03413	-.23006
GDPLG	.40463	.12082	-.04480	.37865	-.32982	-.02272	-.05004	-.12213

	NUMPARTC	LITERACY	LEGEFF	CLOSED	DEMO	RIOTS	GOVCRIS	GDPLG
NUMPARTC	1.00000							
LITERACY	.21980	1.00000						
LEGEFF	-.16262	.46904	1.00000					
CLOSED	.09471	.18325	-.07514	1.00000				
DEMO	-.09624	-.17783	-.14279	.16983	1.00000			
RIOTS	-.15189	-.40436	-.05009	.00592	.74171	1.00000		
GOVCRIS	.19528	-.14197	-.10817	.07817	.33683	.20814	1.00000	
GDPLG	.02159	.78010	.69455	.07054	-.15674	-.31538	-.03469	1.00000

	TURNOUT	COMP	EASTERN	PR	PLUMAJ	UNICAM	CAT	FREQ
POPLG	-.19296	-.00019	-.01028	-.31675	.28503	-.46106	-.18264	-.18422
WCILG	-.08809	.09340	-.26593	-.32438	.15209	-.02825	.11746	-.15994
POPDENLG	.09406	-.11723	.05289	-.08260	.00966	-.05184	.02512	.02212
DISTLG	-.34789	.04303	-.07005	-.41330	.38288	-.48729	-.10383	-.11302
RADIOLG	.38212	.05524	.07633	.25017	-.38275	.07386	-.18268	-.05241
TVLG	.40886	.06042	.20678	.09823	-.21834	.05238	-.02026	-.18329
NEWSLG	.37061	.03519	.05352	.34199	-.37722	.00300	-.22850	-.11107
INFOLG	.41069	.01008	.22117	.24426	-.38285	.06615	-.07105	-.17429

	NUMPARTC	LITERACY	LEGEFF	CLOSED	DEMO	RIOTS	GOVCRIS	GDPLG
POPLG	.23659	-.29849	-.14881	.10116	.47506	.50258	.50727	-.20993
WCILG	-.31937	-.41294	-.20824	.23856	.54040	.46067	.42137	-.25619
POPDENLG	-.00246	-.13449	-.01242	-.11253	.20453	.25779	.08440	-.02254
DISTLG	.15628	-.44982	-.24279	.09443	.42187	.50691	.44220	-.35933
RADIOLG	.14197	.81801	.38581	.13853	-.12439	-.30385	-.10345	.71579
TVLG	.09647	.50017	.30215	-.00833	.06603	-.14167	.10331	.62437
NEWSLG	.14025	.82760	.64649	.06224	-.17632	-.25954	-.18928	.85966
INFOLG	.17189	.79783	.44189	.09482	-.02983	-.25561	-.00438	.80615

	POPLG	WCILG	POPDENLG	DISTLG	RADIOLG	TVLG	NEWSLG	INFOLG
POPLG	1.00000							
WCILG	.49845	1.00000						
POPDENLG	.20458	.15356	1.00000					
DISTLG	.92474	.50615	.15329	1.00000				
RADIOLG	-.21161	-.31361	-.12280	-.34436	1.00000			
TVLG	.03544	-.00608	.21504	-.08967	.60301	1.00000		
NEWSLG	-.21582	-.38341	-.05153	-.38026	.75761	.54965	1.00000	
INFOLG	-.11851	-.21923	.05183	-.28187	.87247	.85714	.77676	1.00000

9.6 Appendix F – Descriptive statistics – Decade models

Table 19: *Descriptive statistics – 1950s*

All results based on nonmissing observations.

Variable	Mean	Std.Dev.	Minimum	Maximum	Cases
All observations in current sample					
TURNOUT	68.7911667	19.3380670	22.4550000	92.7000000	30
COMP	.266666667	.449776445	.000000000	1.00000000	30
PR	.666666667	.479463301	.000000000	1.00000000	30
PLUMAJ	.233333333	.430183067	.000000000	1.00000000	30
UNICAM	.366666667	.490132518	.000000000	1.00000000	30
CAT	.433333333	.504006933	.000000000	1.00000000	30
FREQ	3.48333333	1.25533345	1.75000000	7.66666667	30
NUMPART	3.29122024	1.29316801	1.74000000	7.62000000	28
LITERACY	87.6262222	18.2066917	22.2500000	98.8000000	30
LEGEFF	.933333333	.253708132	.000000000	1.00000000	30
CLOSED	.565217391	.506869802	.000000000	1.00000000	23
DEMO	.316555556	.485965916	.000000000	2.00000000	27
RIOTS	.632714693	.963904114	.000000000	3.66666670	27
GOVCRIS	.479761905	.633598988	.000000000	2.33333333	28
GDP	6609.82468	3410.35328	806.395000	13551.1000	30
GDP(log)	8.60372041	.723917853	6.69257370	9.51422300	30
POP*	34569.4867	77907.9123	164.412500	401539.000	30
POP(log)	9.02594965	1.75315001	5.10237851	12.9030599	30
WCI	2544.62037	2856.95102	.000000000	9325.00000	27
WCI(log)	5.93534729	3.37968454	.000000000	9.14045424	27
POPDENS	2158.71944	2368.76585	30.5000000	8363.00000	30
POPDEN(log)	6.77939540	1.62335762	3.41772668	9.03157249	30
DISTRICT*	81.1179278	146.618203	2.94000000	816.795000	30
DIST(log)	3.76851814	1.06795404	1.07840958	6.70538815	30
RADIOS**	2465.60690	1857.87064	24.5000000	7900.20000	29
RADIO(log)	7.35700944	1.24618651	3.19867312	8.97464335	29
NEWSPAPER**	2535.52667	1514.15244	86.0000000	5773.75000	30
NEWS(log)	7.53450427	.964458387	4.45434730	8.66107706	30
INFO	2459.47333	1450.53806	55.2500000	5667.50000	30
INFO(log)	7.48334094	1.03406215	4.01186834	8.64250338	30

Note:
* scaling 1000
** scaling 0.0001
*** scaling 0.00001

Table 20: *Descriptive statistics – 1960s*

All results based on nonmissing observations.

Variable	Mean	Std.Dev.	Minimum	Maximum	Cases
All observations in current sample					
TURNOUT	68.9887265	18.9770725	23.1300000	94.2000000	39
COMP	.256410256	.442359030	.000000000	1.00000000	39
PR	.615384615	.492864058	.000000000	1.00000000	39
PLUMAJ	.256410256	.442359030	.000000000	1.00000000	39
UNICAM	.410256410	.498310235	.000000000	1.00000000	39
CAT	.461538462	.505035374	.000000000	1.00000000	39
FREQ	3.66236752	1.39758304	1.75000000	9.50000000	39
NUMPART	3.03383846	1.17886922	1.46000000	6.87000000	38
LITERACY	85.3150427	18.7132352	29.6500000	99.1750000	39
LEGEFF	.871794872	.338688428	.000000000	1.00000000	39
CLOSED	.586206897	.501230014	.000000000	1.00000000	29
DEMO	1.03919444	3.07206610	.000000000	18.0000000	36
RIOTS	1.72803810	3.94595933	.000000000	18.0000000	35
GOVCRIS	.404228571	.708854548	.000000000	3.50000000	35
GDP	7898.62030	4627.00028	959.624950	17807.5450	39
GDP(log)	8.73820718	.777226176	6.86654253	9.78737752	39
POP*	32822.4414	83962.2206	191.540000	491451.100	39
POP(log)	8.85562266	1.80385357	5.25509666	13.1051177	39
WCI	5584.72143	12424.5295	.000000000	68185.0000	35
WCI(log)	5.92536087	3.72946231	.000000000	11.1299799	35
POPDENS	3247.35726	4821.50685	37.7500000	25983.0000	39
POPDEN(log)	7.10908530	1.60397574	3.63098548	10.1651978	39
DISTRICT*	89.2650812	163.227569	3.19000000	963.630000	39
DIST(log)	3.82337928	1.09846903	1.16002092	6.87070740	39
RADIOS**	3159.38889	2378.93819	110.500000	11806.2000	39
RADIO(log)	7.68564073	1.01322962	4.70501552	9.37638010	39
NEWSPAPER**	2210.74430	1544.83196	123.000000	4944.66667	38
NEWS(log)	7.34687253	.966882372	4.81218436	8.50606483	38
INFO	2692.61118	1841.18803	116.750000	7486.00000	38
INFO(log)	7.54618806	.985093436	4.76003490	8.92078989	38

Note:
* scaling 1000
** scaling 0.0001
*** scaling 0.00001

Table 21: *Descriptive statistics – 1970s*

All results based on nonmissing observations.

Variable	Mean	Std.Dev.	Minimum	Maximum	Cases
All observations in current sample					
TURNOUT	72.0263953	15.8955100	24.8100000	94.2666667	43
COMP	.209302326	.411625091	.000000000	1.00000000	43
PR	.558139535	.502485517	.000000000	1.00000000	43
PLUMAJ	.348837209	.482242822	.000000000	1.00000000	43
UNICAM	.465116279	.504684588	.000000000	1.00000000	43
CAT	.395348837	.494711791	.000000000	1.00000000	43
FREQ	3.86666658	1.14500732	1.00000000	6.50000000	41
NUMPART	3.00908140	1.16713626	1.35000000	5.55250000	43
LITERACY	86.7600833	17.2790977	34.4000000	99.4000000	40
LEGEFF	.837209302	.373543684	.000000000	1.00000000	43
CLOSED	.620689655	.493803974	.000000000	1.00000000	29
DEMO	.782170550	1.95418556	.000000000	10.5000000	43
RIOTS	.581007752	1.04091294	.000000000	5.00000000	43
GOVCRIS	.461627907	.638973785	.000000000	2.66666667	43
GDP	10835.4431	5884.17195	1219.33000	21887.2550	43
GDP(log)	9.07712893	.746989316	7.10605681	9.99365978	43
POP*	33597.8932	97436.0474	97.0100000	607396.600	43
POP(log)	8.64572445	2.01560979	4.57481407	13.3169372	43
WCI	3480.36822	6872.27972	.000000000	37693.5000	43
WCI(log)	5.22011393	3.70966605	.000000000	10.5372429	43
POPDENS	3820.80659	4926.05922	44.7500000	26844.0000	43
POPDEN(log)	7.39486190	1.55618499	3.80109144	10.1977976	43
DISTRICT*	89.8996783	183.170898	3.63750000	1139.83500	43
DIST(log)	3.71579028	1.20000986	1.29129663	7.03863879	43
RADIOS**	4540.82868	3328.91824	277.500000	17609.4000	43
RADIO(log)	8.05814018	1.00552070	5.62582093	9.77618813	43
TVS***	21060.7439	14245.6816	87.0000000	51160.8000	41
TV(log)	9.54519920	1.20432944	4.46590812	10.8427289	41
NEWSPAPER**	2206.41667	1596.26224	90.0000000	5442.50000	39
NEWS(log)	7.32659460	1.00498540	4.49980967	8.60199379	39
INFO	9217.20743	6329.86249	177.666667	23890.0667	43
INFO(log)	8.69867892	1.18296735	5.17990914	10.0812180	43

Note:
* scaling 1000
** scaling 0.0001
*** scaling 0.00001

Table 22: *Descriptive statistics – 1980s*

All results based on nonmissing observations.

Variable	Mean	Std.Dev.	Minimum	Maximum	Cases
All observations in current sample					
TURNOUT	72.0145833	13.5128514	39.9000000	96.9000000	60
COMP	.200000000	.403375587	.000000000	1.00000000	60
PR	.516666667	.503939284	.000000000	1.00000000	60
PLUMAJ	.383333333	.490301418	.000000000	1.00000000	60
UNICAM	.466666667	.503097749	.000000000	1.00000000	60
CAT	.450000000	.501692052	.000000000	1.00000000	60
FREQ	3.99717514	1.87982878	1.00000000	11.0000000	59
NUMPART	2.95692290	1.42227560	1.00000000	7.30000000	58
LITERACY	88.3686275	14.9221596	35.4000000	99.5750000	51
LEGEFF	.650000000	.480994732	.000000000	1.00000000	60
CLOSED	.595238095	.496795772	.000000000	1.00000000	42
DEMO	.928333333	1.97176945	.000000000	10.0000000	60
RIOTS	.779166667	2.48994328	.000000000	14.3333333	60
GOVCRIS	.181944444	.390714875	.000000000	2.00000000	60
HDI	.771911765	.136055574	.360000000	.920000000	51
GDP	10672.7300	7182.20454	1342.49500	26426.1200	60
GDP(log)	9.00419775	.792893568	7.20228510	10.1821082	60
POP*	35311.8705	102768.133	42.6100000	753082.067	60
POP(log)	8.50079433	2.35759259	3.75208897	13.5319295	60
WCI	4100.81806	8463.33887	.000000000	40612.3333	60
WCI(log)	4.82445431	4.10618172	.000000000	10.6118271	60
POPDENS	3814.77139	4709.01228	51.5000000	28237.0000	60
POPDEN(log)	7.42876130	1.50641415	3.94158181	10.2483885	60
DISTRICT*	100.544086	199.345974	3.57500000	1383.41333	58
DIST(log)	3.71105604	1.36834686	1.27396518	7.23230915	58
RADIOS**	5305.43722	3540.39472	588.000000	20873.4000	60
RADIO(log)	8.32571985	.779418437	6.37672695	9.94623090	60
TVS***	24497.9454	18386.6656	10.0000000	74669.0000	58
TV(log)	9.49902329	1.72810038	2.30258509	11.2208203	58
NEWSPAPER**	1873.63836	1531.75664	117.000000	5666.66667	53
NEWS(log)	7.13856634	.986314200	4.76217393	8.64235633	53
INFO	10701.6283	7416.36512	303.166667	32714.1333	58
INFO(log)	8.92216374	1.01186596	5.71428271	10.3955625	58

Note:
* scaling 1000
** scaling 0.0001
*** scaling 0.00001

Table 23: *Descriptive statistics – 1990s*

All results based on nonmissing observations.

Variable	Mean	Std.Dev.	Minimum	Maximum	Cases
All observations in current sample					
TURNOUT	67.8786798	15.2811762	21.6100000	96.4000000	89
COMP	.123595506	.330984232	.000000000	1.00000000	89
PR	.505617978	.502801142	.000000000	1.00000000	89
PLUMAJ	.325842697	.471344327	.000000000	1.00000000	89
UNICAM	.561797753	.498977504	.000000000	1.00000000	89
CAT	.415730337	.495639824	.000000000	1.00000000	89
FREQ	3.94288390	1.53168325	1.00000000	14.0000000	89
NUMPART	3.96598423	3.01998041	1.00000000	19.7286350	87
LITERACY	86.5916892	19.5579710	20.4000000	99.8000000	74
LEGEFF	.579545455	.496460828	.000000000	1.00000000	88
CLOSED	.593750000	.495014831	.000000000	1.00000000	64
DEMO	.511363633	.799978226	.000000000	4.50000000	88
RIOTS	.257575754	.612163370	.000000000	3.75000000	88
GOVCRIS	.227272727	.491421829	.000000000	3.00000000	88
HDI	.771267123	.155415656	.300000000	.940000000	73
GINI**	38.6179412	10.2254804	22.5000000	61.0000000	68
GDP	10839.2632	8454.53501	793.555950	40583.1550	87
GDP(log)	8.93393700	.926720255	6.67652405	10.6111084	87
POP*	33627.5641	106580.111	30.5000000	938321.800	88
POP(log)	8.46934611	2.27946857	3.41772668	13.7518482	88
WCI	2230.48097	3208.70232	.000000000	17496.7500	88
WCI(log)	4.76655222	3.88074972	.000000000	9.76977043	88
POPDENS	3687.19299	4734.48783	59.5000000	31010.3333	88
POPDEN(log)	7.44661850	1.43640057	4.08597631	10.3420758	88
DISTRICT*	100.653206	209.427955	2.21500000	1721.69000	84
DIST(log)	3.73673723	1.32654846	.795252403	7.45106165	84
RADIOS***	5544.70900	3647.47158	384.000000	20929.4000	87
RADIO(log)	8.35503668	.821780725	5.95064255	9.94891015	87
TVS****	28861.3900	19601.8822	53.5000000	80164.0000	87
TV(log)	9.76099246	1.42612273	3.97968165	11.2918298	87
NEWSPAPER***	1586.82478	1425.06287	10.0000000	5960.00000	76
NEWS(log)	6.83849210	1.25835525	2.30258509	8.69282576	76
INFO	12512.4817	8084.85297	290.500000	34440.8000	87
INFO(log)	9.05860715	1.09961503	5.67160358	10.4469972	87

Note:
* scaling 1000
** scaling 0.001
*** scaling 0.0001
**** scaling 0.00001

Appendix G – Bivariate regressions, all variables – General Model

Table 24: *Bivariate regressions, all variables – General Model*

Variable	Coefficient	T-stat
Institutional		
COMP	9.1	2.7
FREQUENCY	-1.3	-1.3
CLOSE	3.3	1
MAJPLU	-2.8	-1
PR	5.4	2
UNICAM	-0.7	-0.3
LEGEFF	7.4	2.7
NUMPART	-0.05	-0.13
DISTRICT(log)	-2.2	-2.2
Socio- economic		
GDP(log)	6.7	3.8
POP(log)	-0.6	-1
POPDENS(log)	1	1
LITERACY	0.2	2.6
CAT	-0.8	-0.3
Information circulation		
RADIOS(log)	4	2.2
TVS(log)	1.1	1.5
NEWSPAPER(log)	0.8	1.4
INFO (log)	3	2.1
Activism		
DEMOS	-0.8	-0.6
RIOTS	-1	-1
GOVCRIS	0.6	0.13
WCI(log)	0.1	0.2
EASTERN	-2.4	-0.6