

EUROPEAN UNIVERSITY INSTITUTE
Department of Political and Social Sciences

Beyond Electoralism?
Electoral fraud in third wave regimes
1974-2009

Carolien van Ham

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Political and Social Sciences of the European University Institute

Examining Board:

Prof. Mark Franklin (EUI Supervisor)

Prof. Philippe Schmitter (EUI)

Prof. Staffan Lindberg (University of Florida / University of Gothenburg)

Prof. Petr Kopecky (University of Leiden)

Beyond Electoralism?
Electoral fraud in third wave regimes
1974-2009

Carolien van Ham
European University Institute

Abstract

As the “third wave” of democratization spread across the globe after 1974, more and more citizens got a chance –often for the first time in their lives– to cast their vote in multi-party elections. Since then, the number of countries holding regular elections for executive and legislative offices has sharply increased: over 85% of the world’s states now select their national leaders through elections. Unfortunately, the variety of elections has multiplied concomitantly, ranging from “free and fair” elections with genuine contestation between parties or candidates to “façade” elections that are marred by manipulation and fraud. In light of these empirical developments, research on the quality of elections is increasingly relevant. Not only as a way to clarify the fuzzy boundaries between regime types, particularly electoral autocracy and electoral democracy. But also, and more importantly, to understand the causes of variation in election quality as well as its consequences for the functioning of government and broader democratization processes in these polities. This thesis studies the quality of elections in 97 countries in Southern Europe, Central and Eastern Europe, the Former Soviet Union, Sub-Saharan Africa, South America and Central America from 1974 until 2009. Chapter 1 reviews the literature on democratization and elections and specifies the research questions addressed. Subsequently, chapter 2 proposes a definition of the quality of elections that is grounded both in academic work as well as international legal conventions on human and political rights, and introduces the data collected to ‘measure’ election quality. The resulting database on electoral fraud in third wave regimes contains election quality scores for over 880 elections. Chapter 3 and 4 study variation in election quality across polities, attempting to explain why some new democracies manage to “get their elections right” while others do not. Chapter 5 and 6 ask the “so what” question by investigating the consequences of variation in election quality: do elections of higher quality generate more accountable and responsive governments? Finally, chapter 7 connects the findings in the earlier chapters by inquiring to what degree and how election quality affects broader democratization processes and concludes with suggestions for policy-making and further research.

Acknowledgements

Writing a PhD thesis has been one of the most difficult and yet one of the most enriching experiences in my life, a journey along which I met many inspiring people and learned a lot. Though I won't be able to thank all of them here, I will mention a few. First of all, I would like to thank Mark Franklin for accepting me as his PhD student at EUI and being my supervisor along the way. You were always there in times of need and supported me in all my different projects, and I am very grateful you trusted my capacities sufficiently to agree to my eventual change in thesis-topic, from old to new democracies, at the beginning of the third year. I learned much more from you than can be mentioned here, not only about causal reasoning and quantitative methods, but most importantly what high-quality research is, and of course none of your students will ever forget the importance of contributing to the edifice of knowledge! Secondly, I would also like to thank Petr Kopecky, for his support and encouragement to apply at EUI, and the late professor Peter Mair, for welcoming me at the Department and being always available for a discussion about my research ideas. My thanks also extend to Pippa Norris for hosting me at the Harvard Kennedy School as a Research Fellow in 2010 and for providing excellent comments on my research set-up. In addition, I would like to thank Petr Kopecky, Staffan Lindberg, Jorgen Elklit, Steven Levitsky and Philippe Schmitter for their feedback on research ideas in earlier stages. Also, I would like to thank Sarah Birch, Judith Kelley, Staffan Lindberg, Daniela Donno and Susan Hyde for sharing their data with me, and thank Sarah Birch for also sharing her inspiring book manuscript before it got published. Furthermore, I would like to thank Staffan Lindberg, Petr Kopecky and Philippe Schmitter for agreeing to be part of my thesis jury and for their useful feedback on this manuscript. Finally, I would like to thank friends and colleagues at EUI, colleagues from the colloquium on political behavior were invaluable for learning quantitative methods and great company at the ICPSR summer-school, and feedback from Filipa Raimundo and Marcelo Camerlino contributed to strengthening my theoretical arguments. Lastly, I would like to

thank my new colleagues at the University of Twente, and especially Jacques Thomassen, for their encouragement in the very last phase of the thesis.

Apart from invaluable academic inspiration and discussions, several people were important throughout the process of writing this thesis. At EUI, I would like to thank especially Ana, Filipa, Joergen, Maarit, and Theresa, for their friendship and our wonderful discussions about life and academia. In addition, I would like to thank my parents and brothers, Betty and Jae, Noemi, Anne, Hanna and Jasja for always being there, and Willemijn, Lonneke, and 'de kliek' and 'de paupers' for coming so often to Florence to visit me and enjoy Florence together. Finally, I owe a special thanks to the Capuchin priests in Gemena, who hosted me during the Carter Centre election observation mission to DR Congo which provided such crucial insight into the many faces of fraud, and whose never-ending courage and optimism despite the bleak future of their country provided enormous inspiration. Lastly, I would like to thank Frédéric Renkens, for making this journey together.

Table of Contents

1. Introduction	1
1.1. The third wave of democratization and the role of elections.....	1
1.2. Election quality and election fraud: What do we know and not know (yet)?.....	14
1.3. Elections and democracy	21
1.4. Research questions.....	27
1.5. Outline of the thesis	30
2. How to define and measure election quality?	33
2.1. Conceptualizing election quality.....	33
2.2. Measuring election quality.....	57
2.3. Dataset on Election Quality in Third Wave Regimes 1974-2009	84
3. The quality of elections in third wave regimes 1974-2009	95
3.1. Variation in election quality in third wave regimes 1974-2009: global patterns.....	95
3.2. Variation in election quality in third wave regimes 1974-2009: country patterns	114
4. “Getting elections right”: Explaining variation in election quality	121
4.1. Explaining election quality: incentive structures and norms.....	122
4.2. Hypotheses, data and methods.....	139
4.3. Results.....	144
4.4. Conclusion.....	159
5. Does election quality matter? Consequences for accountability	161
5.1. Elections as instruments of democracy	161
5.2. Consequences of variation in election quality for accountability	171
5.3. Hypotheses, data and methods.....	176
5.4. Results.....	180
5.5. Conclusion.....	201
6. Does election quality matter? Consequences for responsiveness	203
6.1. Consequences of variation in election quality for responsiveness	203
6.2. Hypotheses, data and methods.....	205
6.3. Results.....	211
6.4. Conclusion.....	224
7. Coming full circle: election quality and democratization	227
Bibliography	241
Appendix A – Conceptualizing election quality^a	263
Appendix B – Sample selection	265
Appendix C – Measuring election quality	275
Appendix D – Notes on data & analyses chapter	275
Appendix E – Notes on data & analyses chapter 5	295
Appendix F – Notes on data & analyses chapter 6	298

1. Introduction

1.1. The third wave of democratization and the role of elections

Since the “third wave” of democratization initiated over three decades ago, the number of democracies in the world has more than doubled (Huntington 1991, Diamond 1999)¹. Starting with the transitions to democracy in Portugal and Spain in 1974/1975 and Latin America in the late 1970s and 1980s, subsequent democratization in Central and Eastern Europe and Former Soviet Republics after the collapse of the Soviet Union between 1989 and 1991, followed by the dissolution of Yugoslavia in 1991, and eventually the regime changes in Sub-Saharan Africa and Asia in the course of the 1990s, the total number of democracies in 2009 added up to 116, compared to about 40 at the start of the third wave. Hence, of a total of 194 states in the world, about 60% can currently be considered (electoral) democracies according to Freedom House standards (LeDuc, Niemi, and Norris 2010)².

¹ The “first wave” of democratization stretched over more than a century, and refers to the gradual democratization in Western European nation-states as well as North America, the British dominions and a number of Latin-American countries up until the first World War. The “second wave” of democratization refers to a shorter time-period after the second World War in which several European countries re-democratized and a number of African and Asian countries made transitions to democracy following de-colonization (Huntington 1991). While Huntington’s wave metaphor is commonly used, its empirical accuracy has been questioned, particularly with regard to “reverse waves” that should follow each wave (Doorenspleet 2000). Also, it has been suggested that the post-Cold War regime changes in the former Soviet Republics and Sub-Saharan Africa were quite distinct from those in Southern Europe and Latin America, and hence should be referred to as the “fourth wave” (Doorenspleet 2001, Haerpfer et al. 2009). Whether recent developments in the Middle East should be considered as part of a fourth -or even fifth- wave of democratization remains to be seen in the coming years. For practical reasons I refer to both third and fourth wave regimes as third wave regimes in this thesis.

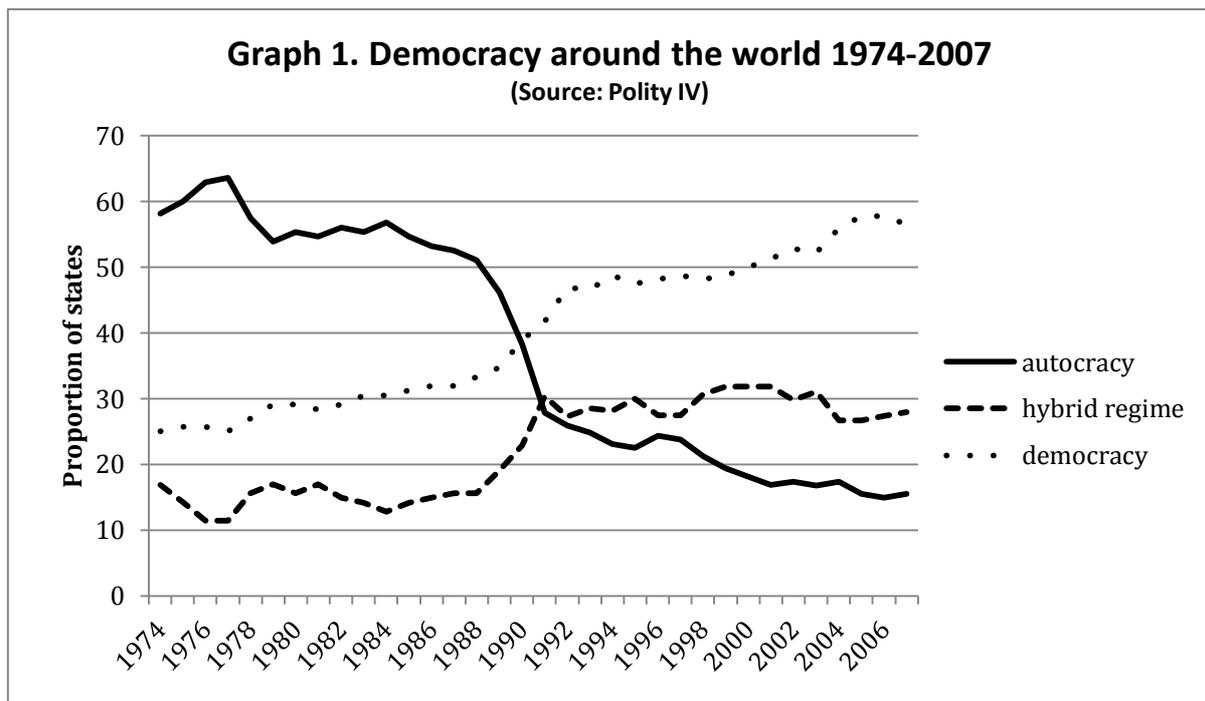
² Freedom House distinguishes between electoral democracies and liberal democracies, with the standards for electoral democracy indicating “the minimal criteria a state needs to meet to be considered democratic”. These criteria are: “(1) A competitive, multiparty political system; (2) Universal adult suffrage for all citizens (with exceptions for restrictions that states may legitimately place on citizens as sanctions for criminal offenses); (3) Regularly contested elections conducted in conditions of ballot secrecy, reasonable ballot security, and in the absence of massive voter fraud, and that yield results that are representative of the public will; (4) Significant public access of major political parties to the electorate through the media and through generally open political campaigning”. Liberal democracies meet the same criteria as electoral democracies in terms of political rights,

However, as the category of ‘democratic regimes’ came to include more countries, the early optimism of scholars and policymakers soon gave way to rising skepticism (Diamond 1999, O’Donnell 1996, 1998, Zakaria 1997, Carothers 2002). Increasingly, observers noted that while holding elections, many new democracies lacked basic guarantees of civil liberties and other aspects commonly considered to be vital elements of democracy, such as institutions of horizontal accountability, and enforcement of the rule of law (O’Donnell 1998, 2004; Diamond 1999, Rose 2001, Zakaria 1997). Rather than new democracies, these seemed to be ‘hybrid’ regimes: regimes that were “neither clearly democratic nor fully authoritarian” (Schedler 2002a: 37, Karl 1995). As an illustration of these trends, the graph below shows the proportion of democratic and autocratic states around the world from 1974 to 2007, with a third category indicating ‘hybrid regimes’³. Clearly, the proportion of autocratic regimes has declined dramatically since 1974, dropping below 50% in 1989 and declining further to less than 30% in 2007. The rise in democratic regimes mirrors this trend in the opposite direction: the proportion of

however in liberal democracies a wider array of civil liberties is present as well (source: http://www.freedomhouse.org/template.cfm?page=351&ana_page=341&year=2008). Freedom House data have important limitations (Munck and Verkuilen 2002), however for our illustrative purposes here are useful.

³ For this graph I use Polity IV data, a well-known democracy index. The Polity IV data provide yearly scores for countries on a 20-point scale from -10 to +10, where scores below 0 indicate degrees of autocracy and scores above 0 degrees of democracy. Both are coded according to 4 components of democracy, i.e. the competitiveness of political participation, the openness and competitiveness of executive recruitment, and constraints on the chief executive (Marshall and Jaggers 2002). For the purpose of Graph 1, these scores were recoded in 3 categories so as to distinguish clearly autocratic regimes and clearly democratic regimes from ‘hybrid’ regimes that are neither clearly autocratic nor clearly democratic. Polities with scores lower than -5 were coded as autocracies, polities with scores higher than +5 score as democracies, and polities with scores between -5 and +5 score as “hybrid regimes”. Several more fine-grained regime typologies exist that differentiate types of autocratic and democratic regimes (cf. next paragraph). However, in order to show the general trends of democracy and autocracy around the world, Polity IV’s threefold typology suffices. This does mean that all three categories have some internal variation: democracies can still receive scores varying from +6 to +10, and autocracies from -6 to -10. For comparisons between Polity and Freedom House data, as well as other aggregate democracy indices, see Munck and Verkuilen (2002) and Hadenius and Teorell (2005).

democratic states approaches 50% in the mid-1990s, and reaches about 60% by the mid-2000s⁴. However, the proportion of democracies rises less quickly than the proportion of autocracies declines. This is due to the fact that the proportion of hybrid regimes, while relatively stable during the 1980s, jumped from about 15% at the end of the 1980s to 30% in the early 1990s, and has remained rather stable since then.



In the wake of these empirical developments, academic responses have been firstly, to describe and classify the new regimes, and secondly, to explain regime changes and subsequent stabilization. As this literature developed, the conception of the role of elections changed as well. The following paragraphs provide a concise overview of this research.

⁴ Note that the number of independent states increased between 1974 and 2006 from about 147 to 194 (based on states included in the Polity IV data), affecting the proportions presented in the graph.

Describing and classifying the variety of new regimes

The unprecedented number of regime transitions in the wake of the third wave, and the concomitant extension of the concept of democracy to new regimes, stimulated the search for new conceptualizations of democracy. The aim was both to develop concepts that could travel better beyond the industrialized democracies as well as to generate concepts that would better describe empirical realities in the new regimes (Sartori 1970, Collier and Mahon 1993, Collier and Levitsky 1997). Following Sartori's advice that in order to apply a concept to a broader number of cases (extension), concepts should have less defining components (intension), scholars aiming to engage in large-N comparative studies of the third wave regimes focused on "minimalist" definitions of democracy⁵. Building on Schumpeter's famous definition of the "democratic method" as "that institutional arrangement for arriving at political decisions in which individuals acquire the power to decide by means of a competitive struggle for the people's vote" (Schumpeter 1975:242, 269; as quoted in O'Donnell 2001:9), other "minimalist" or "procedural" definitions followed. Scholars like Huntington defined democracy as "a political system in which the most powerful collective decision makers are selected through fair, honest, and periodic elections in which candidates freely compete for votes and in which virtually all the adult population is eligible to vote" (Huntington 1991: 7), and proposed the "turnover test of democracy", i.e. whether elections resulted in alternation of power, as a rule of thumb to decide whether regimes had truly made the transition to democracy. Likewise Przeworski and his colleagues claimed that "democracy is a system in which parties lose elections", and that is hence characterized by *ex ante* uncertainty and *ex post* irreversibility (Przeworski et al. 1996:

⁵ Well known large-N datasets on democracy are: the PACL dataset (Przeworski et al. 2000), Hadenius (1992), Polity IV (Marshall and Jaggers 2002), Freedom House (Freedom House 2009). For more detailed discussion of these and other datasets on democracy, see Munck and Verkuilen (2002) and Pemstein, Meserve and Melton (2010).

50-51, 2000)⁶. However, though achieving extension, these procedural conceptualizations did not seem to describe the empirical features of new democracies very accurately. Most notably, the heavy emphasis on elections tended to overlook aspects that perhaps in the older industrialized democracies were taken for granted but were problematic in the new third wave regimes. Hence, scholars turned to, on the one hand, further elaboration of definitions of democracy so as to generate more precise distinctions between regimes, and on the other hand, to creating ‘diminished sub-types’ of democracy that allowed to identify more specifically what aspects new democracies were lacking (Collier and Levitsky 1997).

The elaboration of conceptualizations of democracy lead to more emphasis on “expanded procedural” definitions (O’Donnell 2001; Collier and Levitsky 1997). Here, Dahl (1989) provided inspiration with his definition of polyarchy that consists both of elements referring to elections (i.e. elected officials, free and fair elections, inclusive suffrage, right to run for office) as well as elements that go beyond elections (i.e. freedom of expression, alternative information, and associational autonomy)⁷. This

⁶ Nevertheless, O’Donnell (2001) points out that it is questionable to what degree “minimalist” definitions are truly minimalist. Even Schumpeter (1975) recognizes a number of conditions that need to be met in order for the “democratic method” to work, such as free competition, free votes and a number of other conditions (Schumpeter 1975). Also, a closer look at Huntington’s (1991) definition demonstrates that, though focusing on elections, those elections should take place in a broader context that allows for free competition, a fair and honest electoral process, and alternation in power. Clearly, minimalist definitions are less naïve than often portrayed, as even though over-emphasizing elections, they do specify conditions for the quality of those elections.

⁷ Dahl (1989) defines polyarchy as consisting of the following traits:

1. “Elected officials. Control over government decisions about policy is constitutionally vested in elected officials.
2. Free and fair elections. Elected officials are chosen in frequent and fairly conducted elections in which coercion is comparatively uncommon.
3. Inclusive suffrage.
4. Right to run for office [for] practically all adults.
5. Freedom of expression.
6. Alternative information, [including that] alternative sources of information exist and are protected by law.
7. Associational autonomy. "To achieve their various rights, including those listed earlier, citizens also have a

definition was subsequently elaborated by scholars such as Schmitter and Karl (1991), who added the criteria of absence of reserved power domains (i.e. elected officials must be able to rule) and state sovereignty (i.e. the state must be formally and de facto independent) to their definition of “real existing democracies”.

Conversely, the development of diminished sub-types did not add criteria to define democracy, but rather aimed at identifying those elements that were lacking in order to be considered a ‘full’ democracy. As Schedler (2002a) describes, this research seems to have initially focused on identifying what distinguished the new democracies from old democracies⁸. While new democracies held elections, many lacked other crucial aspects of democracy such as respect for basic civil liberties, institutions of horizontal accountability and rule of law, together often referred to as “liberalism” (O’Donnell 1998, 2004; Diamond 1999, Rose 2001, Zakaria 1997)⁹. Clearly, elections alone appeared

right to form relatively independent associations or organizations, including independent political parties and interest groups.”

⁸ With ‘old’ democracies I refer to those countries that democratized in the first and second wave of democratization (cf. footnote 1). It does not refer however to the countries in Southern Europe, South America and Eastern Europe that had earlier experiences of democratic transitions and subsequently experienced autocratic regimes, to only ‘re-democratize’ in the third wave. These countries, even though they should be considered as re-democratizing rather than democratizing, are considered as part of the third wave, and hence here referred to as new democracies or third wave regimes.

⁹ Liberalism, as a loose label to group these various aspects, is often taken to refer both to the degree to which the state respects citizens’ rights and freedoms, as well as the degree to which state institutions provide guarantees for this through the rule of law, horizontal accountability, etc. As Schneider and Schmitter (2004) and O’Donnell (2001) point out, this is a bit a “can of worms” in the sense that there are many civil and political rights which can be considered crucial for democracy. Nevertheless, there seems to be consensus on a sub-set of “core” rights: “At the level of individuals, these guarantees include the classic elements of the liberal tradition: habeas corpus; sanctity of private home and correspondence; protection against torture and inhuman treatment by authorities; the right to be defended in a fair trial according to pre-established laws; freedoms of movement, of speech, petition, religious conviction and so forth. For social or political groups, these rights have historically covered such things as freedom from punishment for collective expressions of dissent from government policy, freedom from censorship of the means of communication, and freedom to associate voluntarily and peacefully with other persons” (Schneider and Schmitter 2004: 60-61). The link of these civil rights with the rule of law is

to be insufficient conditions for democracy, and the distinction between the older democracies in Western Europe, North America and Oceania and the new post-1974 democracies came to be described as the difference between “liberal” and “illiberal” or “electoral” democracies (Zakaria 1997, Diamond 1999, Karl 2000, Schedler 2002a, Norris 2009)¹⁰. The core assumption in these initial studies hence seems to have been that new democracies succeeded in “getting elections right” but failed to consolidate other dimensions of democracy (Schedler 2002a: 37).

However, as regime transitions spread to the former Soviet Republics, Sub-Saharan Africa and Asia, observers realized that many political regimes that alleged to have made the transition to democracy did not “get elections right” at all. Rather, elections were often marred by severe irregularities, such as opposition intimidation, deficiencies in voter registration and polling procedures, vote-buying, ballot-box stuffing, etc. (Lehoucq 2003, Calingaert 2006). The purpose of these elections seemed to be to allow autocrats to respond to the emerging post Cold War norm of holding multi-party elections, rather than to enable citizens to freely choose their governments (Kelley 2008, Hyde 2011b). As a consequence, academics began to distinguish these regimes as “autocracies in disguise” rather than “democracies with adjectives”, leading to an emerging field of research on “electoral authoritarianism” or “competitive authoritarianism” (Levitsky and Way 2002; 2010; Schedler 2006; Lindberg 2009). Within this literature, various regime-typologies have been developed, some of which bridge the entire spectrum from ‘full’ autocracy to ‘full’ democracy (Munck 2006, Diamond 1999, 2002; Schedler 2002),

that “political liberalization” [involves] “a passive and voluntary connection between individuals and groups who are permitted (but not compelled) by authorities to engage in certain forms of ‘free’ behaviour and a reliable and permanent commitment by authorities not to engage in certain forms of ‘coercive’ behaviour. The shorthand term for this in much of the relevant literature is ‘exercising and respecting the rule of law’ – even if this conjures up a much wider range of connections and commitments, and even if many laws actually in the books are hardly ‘liberal’ in their economic or social content” (Schneider and Schmitter 2004: 61).

¹⁰ Other “diminished sub-types” of democracy are ‘oligarchical’, ‘tutelary’, ‘delegative’ democracies, etc. (cf. Collier and Levitsky 1997 for an overview).

while others concentrate more specifically on autocracies, distinguishing between closed autocracies, hegemonic autocracies, and competitive autocracies (Diamond 2002; Teorell and Hadenius 2009; Levitsky and Way 2010)¹¹. The focus of this literature is hence to distinguish new democracies from new autocracies, based on the insight that “democracy requires elections, but not just any kind of elections” (Schedler 2002a: 37). Concluding, whereas many different regime typologies have been developed to describe the variety of new regimes after 1974, the “liberal” and “electoral” dimension seem to recur in most typologies (Møller 2007, Schedler 2002b).

Explaining regime change and stabilization

Turning to the explanations of regime changes and the subsequent stabilization of those regimes, three broad sub-fields can be distinguished in the democratization literature, each focusing on a different ‘dependent variable’: transitions, consolidation/stabilization and quality (Munck 2004)¹². Starting with in-depth small-N comparative analyses, the early literature focused on transitions from authoritarianism in Portugal, Spain, Greece and Latin America. The transition literature emphasized the highly uncertain nature of periods of regime transition and the –often contingent- role of interactions between conservative and reformist actors within elites as well as the degree of citizen mobilization (O’Donnell, Schmitter and Whitehead 1986; Karl 1990). A key work in the transition literature is the theoretical framework proposed by O’Donnell and Schmitter (1986) that focused on the power struggles involved in transition processes, distinguishing between the actors in power and those in opposition, and studying the internal power balance between conservative and reformist actors

¹¹ While in closed autocracies no elections are held, in hegemonic autocracies elections are held under conditions of very limited opposition ensuring the incumbents’ electoral victory. In electoral authoritarian regimes, competition is more genuinely present, creating a real challenge for incumbents to maintain power (Levitsky and Way 2002, 2010).

¹² For excellent overviews of the democratization literature, theoretical approaches as well as empirical findings, I refer to Geddes (1999), Bunce (2000), Doorenspleet (2001), Casper and Tufis (2003), Munck (2004), Schmitter (2010) and O’Donnell (2010).

in the ruling elite and between moderate and extremist actors in the opposition. Apart from the power struggle between ruling and opposition actors, the degree of elite involvement versus mass mobilization was also perceived to be decisive for flipping the coin either towards democracy or back to authoritarianism (O'Donnell and Schmitter 1986).

As new democracies lasted, research shifted to the consolidation of democracy, emphasizing the gradual attitudinal and behavioral acceptance of democracy by citizens and elites as “the only game in town” to explain post-transition regime stabilization, often using larger-N comparative datasets (Mainwaring, O'Donnell and Valenzuela 1992; Linz and Stepan 1996; Przeworski et al. 1996; Schedler 1998; Diamond 1999; Schneider 2009). A key work in the consolidation literature is the theoretical framework proposed by Linz and Stepan (1996) that focused on the processes of regime stabilization, distinguishing between behavioral, attitudinal and institutional consolidation: “Behaviorally, a democratic regime in a territory is consolidated when no significant national, social, economic, political, or institutional actors spend significant resources attempting to achieve their objectives by creating a non-democratic regime or turning to violence or foreign intervention to secede from the state; Attitudinally, a democratic regime is consolidated when a strong majority of public opinion holds the belief that democratic procedures and institutions are the most appropriate way to govern collective life in a society such as theirs and when the support for anti-system alternatives is quite small or more or less isolated from the pro-democratic forces; Constitutionally, a democratic regime is consolidated when governmental and nongovernmental forces alike, throughout the territory of the state, become subjected to and habituated to the resolution of conflict within the specific laws, procedures, and institutions sanctioned by the new democratic process” (Linz and Stepan, 1996: 6). However, the concept of consolidation remained difficult to define as critiques by for example

Schedler (1998) and Kopecky and Mudde (2000) pointed out (but see Schneider 2009 for an excellent exception)¹³.

Following increasing awareness that not all new democracies evolved towards liberal democracies, but instead ‘consolidated’ as regimes with varying democratic ‘adjectives’, research shifted to defining and measuring the quality of democracy (Diamond and Morlino 2004, 2005; O’Donnell et al. 2004; Beetham 2004). The concept of the quality of democracy defines a number of dimensions that are deemed essential for the functioning of democracy. Diamond and Morlino (2004) propose 8 dimensions by which the quality of democracy can be assessed: freedom, rule of law, vertical accountability, responsiveness, equality, participation, competition, and horizontal accountability. Democracies can score lower and higher on each of these dimensions, allowing for better understanding of trade-offs between dimensions, as well as studying how specific democratic qualities are strengthened or weakened over time. The emerging literature on the quality of elections can be considered an example of this approach (cf. chapter 2).

The role of elections

Along with the developments discussed above, the conception of the role of elections in democracy shifted in two important ways. First of all, elections shifted from being considered a key indicator of successful transitions to democracy to being considered as one of multiple constitutive elements of democracy. Secondly, the notion of elections itself changed from a seemingly straightforward logistical

¹³ The problems pointed out by critics were not only that it is difficult to determine where transition ends and consolidation begins, as well as what the end-point of consolidation should be, it seemed also difficult to separate indicators of consolidation from conditions for consolidation, and hence to separate constitutive components from causal determinants. The shift to studying specific components of democracy and their functioning that is apparent in the study of the “quality of democracy” could be considered partly the result of these difficulties in conceptualization.

operation, to more attention for the quality, and manipulation, of the electoral process (Lehoucq 2003; Elklit and Reynolds 2005a).

In regard to the first element the debate on the “electoralist fallacy” (Karl 1995, 2000, Schmitter and Karl 1991) and the “anti-electoralist fallacy” (Seligson and Booth 1995) is important. As described above, while the transition literature was initially based on in-depth, small-N comparative studies, in the early 1990s and mid-1990s several scholars shifted their attention to mapping and explaining regime changes on a global scale (Huntington 1991; Przeworski et al. 1996, 2000). Attempting to quantify and measure the development of democracy around the world, many of these authors placed heavy emphasis on elections as the key component of democracy, or at least as the minimal institutional requirement of democracy. However, as it became clear that many new democracies might not so easily meet liberal democratic ideals, awareness grew that elections might be necessary, but by themselves insufficient conditions for democracy (Schmitter and Karl 1991, Karl 1995, 2000, Diamond 1999, Carothers 2002). The “electoralist fallacy” hence refers to “the faith that merely holding elections will channel political action into peaceful contests among elites and accord public legitimacy to the winners” – no matter how they are conducted or what else constrains those who win them” (Schmitter and Karl 1991: 78). This was a justified critique addressed to both academics and policymakers, emphasizing that democracy should not be equated with elections and stressing the need to go beyond elections in both the study of democratization as well as democracy promotion efforts. In reaction to this however, several scholars have pointed out that this should not be taken to imply “that elections are meaningless for democratization”, because that would be committing the “anti-electoralist” fallacy (Seligson and Booth, 1995: 16). In fact, it seems that this debate has been magnified somewhat unnecessarily in the past decade, with scholars studying other elements of democracy justifying their choice of subject with reference to the limited importance of elections, while scholars studying elections go to quite some length to explain why elections are important for

democracy¹⁴. It seems time to move beyond this debate, as firstly, clearly elections are an important element of democracy, secondly, clearly democracy is more than just elections, and thirdly, democratic elections are impossible without the prior existence of at least some civil liberties and rights and a certain degree of respect for the rule of law (Elklit 1999, Bratton 1998, Merkel 2004). As Elklit notes “the presence of, respect for, and unhindered use of relevant political and civil rights and freedoms” [...] are prerequisites which must be in place, in other words must be institutionalized, *before we can even think of elections being potentially free and fair.*” (Elklit 1999: 33, italics mine). O’Donnell also notes that: “I do think that fair elections are extremely important. This is not because such elections will necessarily lead to wonderful outcomes. It is because these elections, per se and due to the political freedoms that must surround them if they are to be considered fair, mark a crucial departure from authoritarian rule” (O’Donnell 2001: 9). Hence, the question to be addressed seems to be not whether elections should be considered as important for democracy or not, but rather what kind of elections are important for democracy.

This consideration brings us to the second shift in the literature: the conception of elections itself. In the early days of the third wave, elections were often considered as indicators of democracy, with “founding” elections marking the moment of transition, or the end of the period of transition (O’Donnell and Schmitter 1986). The quality of the electoral process did not attract much attention,

¹⁴ For example, O’Donnell points out in his article on horizontal accountability: “Elections, however, occur only periodically, and their effectiveness at securing vertical accountability is unclear, especially given the inchoate party systems, high voter and party volatility, poorly defined issues, and sudden policy reversals that prevail in most new polyarchies” (O’Donnell, 1998: 113). On the other hand, Pastor notes in his essay on electoral administration: “democracy should be more than just free elections, but it cannot be less” (Pastor 1999: 5), Bratton justifies his analysis of elections in Africa claiming: “It is meaningful to study elections for the simple reason that, while you can have elections without democracy, you cannot have democracy without elections.” (Bratton 1998: 52), and Lindberg stresses “Elections alone are not sufficient to make a democracy, yet no other institution precedes participatory, competitive and legitimate elections in instrumental importance for self-government” (Lindberg 2006a: 1).

probably also due to the fact that most elections held by authoritarian regimes in the mid-1970s and early 1980s were so clearly undemocratic that distinguishing non-democratic from democratic elections was relatively straightforward¹⁵. While the norm of universal suffrage was already widely dispersed and followed at the time, *de jure* restrictions on party competition and reserved power domains were much more common than now (Hermet et al. 1978)¹⁶. Examples are the one-party elections held in the Soviet Union, socialist African states, and several military regimes in Latin America; as well as traditional monarchies, military governments and personalistic regimes that only held elections for legislative office while maintaining an unelected executive.

After the end of the Cold War, the norms of –at least *de jure*- multi-party competition and holding elections for executive office became increasingly widespread (Norris 2009, Hyde 2011b). As a consequence, the vast majority of third wave regimes at present hold formally multi-party elections with universal suffrage for both executive and legislative offices¹⁷. However, while the formal

¹⁵ To be fair, most scholars referring to founding elections do mention that these elections need to be free and fair and that the outcomes should be accepted by the major political actors involved in order to be ‘effective’ founding elections (Schneider and Schmitter 2004). However, it was only with the onset of the ‘fourth’ wave in Former Soviet Republics and Sub-Saharan Africa and the subsequent protracted transitions –making pinpointing a ‘moment’ of transition and hence the ‘founding’ election also more difficult- that lead to increased attention for the quality of elections and their potential role in democratization processes (Lindberg 2006a, 2009).

¹⁶ Even though some third wave regimes had suffrage restrictions for illiterates until well into the 20th century (i.e. Chile until 1970, Ecuador until 1978, Brazil until 1988, etc. Nohlen (2005)), by the time these countries experienced their transitions to democracy, universal suffrage was widely accepted at least regarding income, gender, and –in most cases- also literacy. Limitations on suffrage for young people, non-citizens, detainees and mentally challenged still abound however, as described by Blais et al. (2001).

¹⁷ According to data from the Cross National Time Series archive (CNTS), over 85% of states in the world held regular legislative and executive elections in 2006. Executives were directly elected in 44% of the world’s independent states and an additional 42% of executives were elected indirectly, either via selection by an elected assembly (as for example in parliamentary democracies) or by an elected but uncommitted electoral college (as for example in the United States). Only 13% of states had an executive that was not selected through a direct or

institutions are present, the manipulation of electoral processes seems to have become more diverse and less overt, ranging from unequal media access to administrative barriers to competition and participation, to outright intimidation and coercion of voters and candidates (Schedler 2002ab, Lehoucq 2003). Hence, the presence of elections that meet formal criteria such as *de jure* universal suffrage and multi-party competition is no longer a useful indicator to differentiate elections, and it is not so clear-cut where the line between non-democratic and democratic elections should be drawn now. Clearly, research is needed that goes beyond the “illusion of transparency” of elections (Gueniffey 1993), and instead studies the quality of electoral processes more substantively (Lehoucq 2003; Elklit and Reynolds 2005a). This is necessary both to map the various forms of manipulation that can undermine the quality of elections and better understand the increasing empirical variety of elections around the world, as well as to assess the consequences of such elections for democracy and democratization.

1.2. Election quality and election fraud: What do we know and not know (yet)?

As Lehoucq (2003) and more recently Birch (2011) note, academics were relatively late to pick up the topic of election quality and electoral fraud¹⁸. While the international community invested heavily in election assistance and monitoring as part of its core post-cold war democracy promotion activities and international election observation increasingly became an international norm in the course of the 1990s (Kelley 2008, Hyde 2011b)¹⁹, academics lagged behind. Some scholars did question the

indirect electoral mandate, compared to 40% in 1974. Regarding legislative selection, over 90% of legislatures were selected by direct or indirect popular elections in 2006, compared to 70% in 1974.

¹⁸ I will define these terms more precisely in chapter 2, for now they are used interchangeably, be-it that one is defined ‘positively’ and the other ‘negatively’.

¹⁹ According to Bjornlund (2004) “Between 1989 and 2002, international observers were present in 86% of the national elections in 95 newly democratic or semi-authoritarian countries” (Bjornlund 2004: 43). Considering a larger set of countries, Kelley reiterates that “Monitoring increases from an average below 10 percent of elections

standards used by election observers to evaluate elections (Geisler 1993, Anglin 1998, Carothers 1997), and others initiated attempts to better define and conceptualize the notion of “free and fair” elections (Elklit and Svensson 1997, Pastor 1998), however the real upsurge in scholarly attention seems to have taken place only in recent years (cf. Elklit 1999, Mozaffar and Schedler 2002, Schedler 2002a, Lehoucq 2003, Elklit and Reynolds 2005a, Simpser 2005, Boda 2005, Katz 2005, Calingaert 2006, Hartlyn and McCoy 2006, 2008, Birch 2007, 2011, Alvarez, Hall and Hyde 2008, Collier 2009, Lopez-Pintor 2010, Darnolf 2011).

This is interesting since, as Lehoucq describes, electoral fraud was by no means uncommon in the early years of first wave democracies. In his excellent review, Lehoucq mentions historical accounts of electoral fraud in the US, Latin America and Europe at the end of the 19th and early 20th century. Accounts range from voter intimidation, vote buying and stuffing the voter registry with ineligible voters, to pressure from landlords, employers and the church, in countries ranging from the UK and Ireland, to Germany and Spain (Lehoucq 2003). The fact that these countries by now have had such long experiences of democratic rule, in which electoral processes have come to be seen as straightforward administrative procedures, probably has contributed to the mentioned “illusion of transparency” and the concomitant neglect of the topic of electoral governance in research on third wave regimes (Guennifey 1993, Mozaffar and Schedler 2002, Lehoucq 2003, Pastor 1999)²⁰. However, now the field is developing rapidly, with important advancements made in terms of conceptualizing election quality and electoral fraud, based either on democratic theory or international legal norms for elections (Elklit and Svensson 1997, Elklit 1999, Elklit and Reynolds 2005ab,

from 1975 to 1987, to a high of 81.5 percent of elections in 2004. The most drastic increase occurs between 1988 and 1990” (Kelley 2008: 222).

²⁰ As Pastor (1999) explains: “There are numerous reasons why 'electoral administration' has been the neglected variable, but most of these stem from the fact that people in advanced countries simply take for granted the administrative dimension of elections.” (Pastor 1999: 6).

Schedler 2002a, Mozaffar and Schedler 2002, Lindberg 2006a, Munck 2006, 2009, and Boda 2005, Katz 2005, Goodwin-Gill 2006, Davis-Roberts and Carroll 2010, EC-NEEDS 2007). Also, several cross-national databases have been developed that measure election quality and electoral fraud in third wave regimes (and sometimes also first and second wave regimes) (Elklit and Reynolds 2005a, Lindberg 2006a, Birch 2007, Hartlyn et al. 2008, Munck 2009, Donno 2010, Kelley 2010, Schmeets 2010, Hyde 2011).

Nevertheless, most current research tends to be focused still on conceptualization and measurement issues, as well as in-depth case studies, rather than cross-national accounts of the causes and consequences of electoral fraud. As Lehoucq's and Birch's reviews indicate, most inferential work has been done in the form of case studies, with Lehoucq and Molina's (2002) study of Costa Rica, Eisenstadt's (2004) and Magaloni's (2006) work on Mexico and Nyblade and Reed's (2008) research on Japan as excellent examples (cf. Lehoucq 2003 and Birch 2011 for a more extensive overview of case studies on electoral fraud). Also, more extensive research has been done on specific types of electoral fraud, most notably on vote-buying and to a lesser extent on election violence. Vote buying has been analyzed in a mix of country-studies, qualitative analyses and surveys, as well as experiments (cf. edited volume by Schaffer (2007), Brusco, Nazareno and Stokes (2004), Calvo and Murillo (2007), Vicente (2010), Bratton (2008) and a forthcoming volume by Stokes et al. (2012)). As these studies demonstrate vote-buying seems to occur in most regions of the world, in varying frequencies and at varying costs. As Schaffer (2007) describes, vote-buying appears to be sporadic in some contexts but seems to be quite widespread in others, ranging from an estimated 7% of citizens selling their vote in the Philippines, to 70-75% in Ukraine and Armenia (Schaffer 2007: 3). Bratton reports estimates of vote buying in Nigeria between 12% and 28%, and Brusco, Nazareno and Stokes estimate figures between 12% and 35% in Argentina. The costs of vote-buying seem to differ greatly across countries as well, Schaffer (2007: 2) reports the price of a vote to range from about 60 cents in Manila to about

10,000 dollar in Kuwait. As regards election violence, research seems to have concentrated rather on Sub-Saharan Africa, using both survey and experimental methods to gauge the extent and causes of election violence (Bratton 2008, Collier and Vicente 2011, Von Borzyskowski 2011). As yet, less is known about the frequency of election violence, though Bratton (2008) estimates for the case of Nigeria that about 4% to 6% of voters experienced election violence. Despite the wealth of available case studies and studies on specific types of electoral fraud, cross-national research attempting to explain the causes and consequences of electoral fraud is –as yet– uncommon (with the exception of Birch 2007, 2011, and Hartlyn, McCoy and Mustillo 2008). However, existing research does provide an indication of potential causal factors and consequences of electoral fraud.

Regarding causes of electoral fraud, it has been suggested that lower economic development and economic inequality increases the frequency of vote-buying, election violence, and electoral malpractice in general (Von Borzyskowski 2011, Brusco, Nazareno and Stokes 2004, Bratton 2008). Vote-buying specifically seems to be targeted at poor and low educated voters (Bratton 2008, Birch 2011). It has been suggested as well that electoral fraud would be more common in rural areas than in urban areas (Bratton 2008, Hicken 2007), however Lehoucq (2007) proposes that different types of electoral fraud might be common in rural versus urban areas, as voter intimidation might be relatively easier in rural areas where traditional patron-client relationships might still exist, while in urban areas politicians need to resort to other methods of manipulation such as buying votes. Institutional factors that have been suggested are the type of electoral system and the type of electoral management body. Majoritarian electoral systems appear to be associated with higher levels of electoral malpractice (Birch 2007), while other authors have suggested that candidate-centered electoral systems and electoral systems with small districts (Hicken 2007, Stokes 2007), might also be associated with higher levels of electoral fraud. With respect to electoral management bodies, Hartlyn et al. (2008) find that non-partisan independent Electoral Management Bodies significantly improve the quality of elections.

Apart from socio-economic and institutional aspects, several political factors have been suggested to augment the likelihood of electoral fraud, namely electoral competition and the incumbents' access to state resources (Lehoucq 2003, Hicken 2007, Alvarez and Boehmke 2008). The effect of competitiveness is not straightforward however, as on the one hand, competitiveness might increase parties' incentives to engage in fraud (as Nyblade and Reed 2008 show to be the case in Japan), while on the other hand, competition between parties might lead to more monitoring of the electoral process, both by parties themselves as by civil society organizations. Finally, Birch (2011) finds that media freedom and low levels of corruption are associated with less electoral malpractice, and that the so-called 'resource curse' of resource-rich countries also affects the quality of elections. Also, she finds surprisingly little effect of international election observers, democracy assistance and other international factors (like the geostrategic importance of a country) on electoral malpractice.

Turning to consequences of election quality or electoral fraud, high quality elections have been suggested to have a broad range of positive consequences. As such, Bjornlund (2004) justifies international monitoring of elections in order to improve election quality by mentioning a number of positive consequences, among which (1) spill-over effects to improve respect for other rights, (2) regime change and possibly further democratization, (3) government accountability, and (4) effective and stable governance (p. 34). Birch (2011) mentions similar potential consequences of electoral fraud, proposing that fraud might undermine (1) electoral accountability, (2) regime legitimacy as perceived by voters, (3) have spill-over effects to the state bureaucracy generating higher levels of corruption, (4) entail direct and indirect costs through the inefficient use of state resources, (5) can generate popular protests, violence and ultimately, resumption of civil war or replacement of the incumbent through a coup d'état. That these ideas are common not only among academics but also among policymakers is illustrated by Kohnert (2004), who describes how: "Free elections are considered [by the EU and EIHDR] to be an essential step in the democratization process. They are supposed to promote good

governance, respect for the rule of law as well as a wide range of human rights (Kohnert 2004: 83). Some of these proposed consequences have received empirical support in past years, however, most remain untested. For example, Collier (2009) has shown that electoral fraud indeed increases the likelihood of violent conflict. In addition, Lindberg (2006a) found that successive elections lead to an improvement in civil liberties, and even though not the focus of his study, this effect seemed to be more pronounced in the case of “free and fair” elections. Teorell and Hadenius (2009) found, using a measure of competitiveness of elections, that higher quality elections increase the chances of regime change. Other scholars have found evidence that electoral fraud leads to inefficient provision of public services (Desposato 2007, Baland and Robinson 2007, Stokes 2007, Collier 2009), and lower quality of economic policies (as measured by World Bank definitions, cf. Chauvet and Collier 2009). Also, electoral fraud might have important political consequences, as it deters future opposition (Simpser 2005), or weakens the opposition by making it dependent on the incumbent for access to (state) resources (Desposato 2007). In addition, Mylonas and Roussias (2008) and Roussias (2011) show that electoral fraud undermines strategic coordination between parties and voters, leading to a reduction in the number of parties. Finally, some scholars have also suggested effects of electoral fraud on voter turnout, with Bratton (2008) and Collier and Vicente (2011) finding that election violence depresses turnout, while Birch (2011) suggests a curvi-linear effect of electoral malpractice on turnout, as fraud, when it takes the form of ballot box stuffing, often leads to inflated turnout figures, while expectations of fraud might also keep citizens away from the polls, lowering turnout figures.

Clearly, quite some work remains to be done to uncover the causes and consequences of electoral fraud. Further complicating this task is the fact that electoral fraud may take many forms, and perpetrators of fraud may prefer different combinations of strategies depending on their resources and the context in which elections take place. For example, considering the costs of various fraud strategies, Lehoucq (2007) describes how in early-20th century Spain vote-buying was considered a very

costly and indeed last-resort method of manipulation, as “a market for votes only emerged when parties [...] could not resort to intimidating citizens, manipulating the electoral registry, or manipulating the tally of the vote” (Lehoucq, 2007: 37). On the other hand, nowadays it might very well be the case that the increasingly frequent presence of international election observers has shifted fraudulent practices from election-day to well before election-day. Birch (2011) indeed shows that relatively less visible pre-election malpractice, such as manipulation of the legal framework, is more common than the more visible ‘options’ of voter intimidation and ballot box stuffing on election-day. Conversely, Simpser (2005) demonstrates that for authoritarian incumbents, at least those that are less concerned about the judgment of the international community, overt election-day fraud may be very helpful in demonstrating the incumbents’ strength and discouraging future opposition. Finally, incumbents and opposition may also choose different strategies of manipulation depending on their access to resources. For example, both Bratton (2008) and Collier and Vicente (2011) seem to find in Nigeria that election violence and intimidation is a strategy of the opposition, while vote buying and clientelism is rather pursued by the incumbent.

Concluding, there are different types of electoral fraud as well as different actors engaging in fraud that may choose particular combinations of strategies depending on the institutional, economic, and international context in which the elections take place. The empirical work on such trade-offs and contingent choices is still very limited (but cf. Simpser and Donno 2009, Birch 2011). This realization has nevertheless lead scholars of elections in authoritarian regimes to propose the possibility of different logics of electoral fraud. As Birch (2011) describes: “Heavily manipulated elections typically entail the harnessing of electoral institutions to ends radically different from those envisaged by democratic theory. Far from being mere window-dressing in authoritarian and semi-authoritarian settings, elections are often made to play a key role in maintaining regimes” (Birch 2011: 113). For example, Birch (2000) describes how elections in Ukraine in the Soviet period, were not just show-

cases or masquerades, but served as both a propaganda moment, in which government could explain policies, as well as a possibility to test the state's control capacity by testing its capacity to generate full turnout, and in the post-Stalin period, to provide a certain space of expression of discontent of citizens, with the purpose of maintaining regime stability. Gandhi and Lust-Okar (2009) provide a more extensive discussion of the different functions elections may have in authoritarian regimes, and find that elections might serve to: (a) divide spoils among the elite, (b) to discourage opposition (cf. also Simpser 2005), (c) to co-opt the opposition, (d) to inform the incumbent about his/her own popularity and opposition strongholds, as well as (e) to maintain a system of clientelism-through-elections (Lust-Okar 2009).

Summarizing, research on how to conceptualize election quality and electoral fraud as well as the development of cross-national data in recent years provides an excellent starting point for further research on general causes and consequences of electoral fraud, as well as trade-offs between different types of fraud, actors involved in fraud, the contextual effects on trade-offs and actor strategies as well as different 'logics' of electoral fraud in different regimes. In this thesis we will start with the first step, analyzing and attempting to explain the empirical variety of elections in third wave regimes, as well as investigating the consequences of different elections for democracy and democratization.

1.3. Elections and democracy

Before turning to the research questions that guide this thesis, three issues regarding the relation between elections and democracy need to be briefly addressed.

Regimes and partial regimes

First of all, even though seemingly obvious, in order to differentiate between democracy and elections in our consideration of, for example, the consequences of elections for democratization processes, it is

important to underscore the difference between democracy as a regime and elections as a “partial” regime (Schmitter 1992). Following Lindberg, I take political regime to mean “the rules governing the distribution of power and the relationships between the agents of power in polities” (Lindberg 2006a: 6). Regimes are different from the state and different from government in the sense that: “The state is what one rules, regimes are how one rules, and government is the group of individuals who rule.” (Lindberg 2006a: 6). Elections are then “partial regimes” that form only one component of “a composite of partial regimes” that together form a democratic regime, at least if the elections and other partial regimes meet the criteria for a democratic regime (Schmitter 1992: 427). Hence, elections are conceptually distinct from regimes and can only be a component of regimes (hence, “equating democracy with elections” is, from this point of view, not only empirically, but also conceptually impossible). The unit of analysis in this research is elections, not regimes; and hence in the next chapter, I will attempt to provide a definition of democratic elections, leaving the question of defining democratic regimes to others.

Moreover, since, as described above, elections no longer seem to differ substantially in terms of formal, *de jure* conditions of universal suffrage and multi-party competition, but rather vary in the degree to which informal, *de facto* conditions for participation and competition exist (cf. chapter 2), this research studies elections in different political regimes, both autocracies and democracies. This is possible because, since elections are only components of regimes, theoretically, high quality elections could take place in autocracies (which then would be autocratic because other regime components cannot be considered democratic), and low quality elections could take place in democracies (which would then be democratic because other regime components are democratic). To be fair, considering the importance of other regime components such as respect for civil liberties and the rule of law for the quality of elections, empirically it seems unlikely that we will find high quality elections in autocracies and vice versa (cf. chapter 7). However, theoretically such combinations are possible and

therefore it is important to maintain the distinction between elections as regime components and democracy as –one type of- political regime.

The precise definition of election quality as well as the sample selection requires more in-depth discussion, which is provided in chapter 2. For now it is relevant that elections are the unit of analysis in this research, and that since they are held across different political regimes, to capture the full range of variation in election quality this research studies the quality of elections in political regimes ranging from autocracies to democracies.

Elections as instruments of democracy

Having differentiated elections as components of political regimes, it is important to take a closer look at the notion ‘quality of elections’. Quality of elections is in a way a confusing concept as it can refer both to the electoral process as well as the degree to which elections fulfill their function as defined in democratic theory. The quality of the electoral process refers to the actual election, ranging from pre-election campaigning, to election-day voting and vote tabulation, to post-election day adjudication of electoral disputes (Elklit and Reynolds 2005a). The quality of the electoral regime however, seems to refer to a more functional conception of elections as ‘instruments of democracy’ (Powell 2000). Such a conception derives from the notion that in representative democracy, the alternative for direct rule by the people is rule for the people by elected representatives (Dahl 1989). Hence elections function as a linkage mechanism between citizens and representatives, giving citizens the power to choose representatives to govern in their stead (Miller and Stokes 1963; Thomassen 1994). By doing so, ideally elections not only give citizens the possibility to choose future governments, but also to signal their policy preferences to those governments. In practice, choosing a future government also implies that citizens can evaluate the past government and hold it accountable for its past performance. Hence, elections are opportunities for citizens to (a) render a verdict on the past performance of their

government, (b) select a new government, and (c) establish guidelines for future government actions (Powell 2000; Van der Eijk and Franklin 2009)²¹. While elements (a) and (b) refer to citizen influence on government formation, element (c) also implies citizen influence on government policies²².

The core assumption of elections as instruments of democracy is that citizen influence on government formation generates citizen influence on government policies, as the power of citizens to replace their representatives at elections, i.e. “throwing the rascals out”, creates incentives for representatives to be responsive to citizens’ needs (Powell 2000; Przeworski, Stokes and Manin 1999, Strøm, Müller and Bergman 2003). Citizen influence on government formation is defined in the context of this thesis as accountability, asking to what degree unsatisfactory government performance before elections affects the fate of incumbents at the polls²³. Citizen influence on government policies is defined here as responsiveness, asking to what degree governments formed after elections implement policies that are congruent with the policy preferences of voters. Concluding, elections function as ‘instruments of democracy’ if they generate accountability and responsiveness.

²¹ Admittedly, it is very difficult to do all three with just one vote, rendering elections rather ‘clumsy’ instruments to achieve these goals. Electoral research has indeed demonstrated differences between citizens in the degree to which their vote decision is based on the retrospective evaluation of past performance of incumbents (often perceived to be a decision requiring less information about politics) or based on prospective comparisons of party programs and leaders (often perceived to be a decision requiring more information and stronger party or ideological identification). Despite these individual differences however, at the aggregate level it seems that in established democracies incumbent governments indeed tend to be held accountable in elections when government performance is perceived to be insufficient by sufficient citizens, and governments seem to be responsive to -aggregate- citizens’ policy preferences in between elections (Powell 2000; Soroka and Wlezien 2010; Stimson et al. 1995).

²² More narrow “elitist” conceptions of elections limit the linkage function to the selection of leaders by citizens (Schumpeter 1975), while broader conceptions also consider elections as a mechanism by which citizens signal policy preferences (Powell 2000, Wessels and Schmitt 2008).

²³ Of course, the accountability referred to here is accountability generated through elections, or ‘vertical’ accountability. Other mechanisms exist to hold governments accountable for their actions in the period between elections, among others ‘horizontal’ accountability by institutions that provide checks on executive power, but also ‘diagonal’ accountability by civil society (Schmitter 2004, Smulovitz and Peruzzotti 2000).

Now, the degree to which elections indeed fulfill these functions is likely to depend on the institutional and socio-economic setting in which elections take place. However, the quality of the electoral process is also likely to be a crucial factor, as engaging in electoral manipulation and fraud might be an effective strategy for governments that seek to avoid being held accountable for their past actions, and an incumbent that came to power through fraudulent elections might have little incentive to be responsive to citizens' needs²⁴. Hence, these two notions of election quality need to be clearly separated. In the remainder of this thesis I will therefore refer to the quality of elections, or defined negatively, electoral fraud, when discussing the *electoral process*. When analyzing the functioning of elections qua partial regime (in chapters 5 and 6), I will refer to the quality of the *electoral regime*.

Elections as instruments of democratization?

While classic theories of representation view elections as “instruments of democracy”, generating accountability and responsiveness, recently scholars of democratization have begun to ask whether elections might have spill-over effects on other components of the political regime, thereby triggering broader processes of democratization (Lindberg 2006a, Lindberg 2009)²⁵.

On the one hand, considering democratization-as-regime-change, elections are extraordinary moments of heightened media attention and international scrutiny, that can become “focal points of elite as well as mass coordination” (Schedler, 2009: 306), providing a window of opportunity for regime change even in contexts of limited respect for civil liberties and the rule of law. The electoral revolutions in Serbia, Georgia, Ukraine, Kyrgyzstan and elsewhere illustrate one form such

²⁴ This topic is discussed more in-depth in chapters 5 and 6, where I also test empirically whether electoral fraud indeed undermines electoral accountability and responsiveness.

²⁵ Part of the reason for this was the realization that in many third wave regimes in Sub-Saharan Africa, the first multi-party elections could not really be considered as indicators of completed democratic transitions, but rather seemed to form part of protracted transitions with uncertain destinations (Barkan 2000, Lindberg 2006, Van de Walle 2002).

mobilization might take (Bunce and Wolchik 2006). Moreover, elections tend to provide, even if only momentarily, expanded space of manoeuvre for opposition parties, civil society organizations, journalists and judges that might empower such actors to maintain their newly acquired freedoms also after the election was over. Indeed Lindberg (2006a) found that elections lead to increased civil liberties during and after elections, indicating a potential democratizing effect of elections.

On the other hand, considering democratization-as-regime-stability/consolidation, elections may provide opposition and citizens with a possibility to express their discontent about the government through democratic means rather than through engaging in civil war or a coup d'etat, thereby creating more stable regimes (Birch 2011). Moreover, elections may also confer legitimacy on the government in the eyes of citizens and parties, increasing the gradual acceptance of democracy as “the only game in town”. This might in its turn further generate “democratic lock-in” mechanisms, where the identification with democratic values of citizens creates expectations that leaders should conform to (Lindberg 2006a).

However, what is problematic about assessing the consequences of the quality of elections for democratization is that ultimately, the quality of elections is also an indicator of democratization, leading to circular causal reasoning and endogeneity in empirical analyses. For example, asking whether the quality of elections is associated with regime change is tautological as regime change is measured by changes in overall democracy scores that are heavily based on the quality of elections (cf. chapter 2). One could assess whether high quality elections are more often associated with incumbent turnover (they are, cf. chapter 5), however incumbent turnover by itself does not provide evidence of democratization. Analyses considering the relation between election quality and regime stability as well as regime ‘quality’ suffer from the same problem of endogeneity. Hence, I find it most useful to think of the consequences of elections as spill-over effects of one regime component on other regime

components. Consequently, to ask whether high quality elections are associated with subsequent improvements in respect for civil liberties and rights seems possible (Lindberg 2006a), as does asking whether high quality elections have consequences for subsequent respect for the rule of law, or democratic legitimacy as expressed in the absence of popular protests, civil war or coup d'états.

Summarizing, in this thesis I define elections as partial regimes, i.e. components of regimes, and autocracies, democracies, and hybrids as regimes, i.e. composites of partial regimes. Also, the quality of elections is taken to refer to the quality of the electoral process, or, defined negatively, the extent of electoral fraud and manipulation in the electoral process, whereas the quality of electoral regimes refers to the degree to which elections function as prescribed in democratic theory. Consequences of the quality of elections will be assessed by (1) looking at the functioning of electoral regimes: do elections function the way they should according to democratic theory?; and (2) by considering the spill-over effects of elections on other components of a democratic regime: do high quality elections lead improved respect for civil liberties, the rule of law, and higher legitimacy?

1.4. Research questions

As sketched in the previous paragraphs, the early optimism of scholars and policy-makers about the prospects for democratization in third wave democracies has given way to a more nuanced approach that recognizes first, that democracy takes different forms in different societal, economic and historical contexts (Schmitter and Karl 1991, Carothers 2002, 2007), second, that democracy requires elections, but constitutes more than just elections (Karl 1995, 2000), and third, that democracy requires a certain quality of elections (Schedler 2002a; Elklit and Reynolds 2005a, O'Donnell 2001).

However, exactly what is meant by 'quality of elections' and what degree of quality is needed for elections to be considered democratic is less clear-cut. Also, even though research on democratization

has provided important insights about causal factors driving regime change and stability, the causal dynamics underlying electoral processes and their quality are less well understood. A better understanding of what drives election quality could provide important insights for academics as well as for policymakers and practitioners working in international election assistance and monitoring. Moreover, despite often-heard claims about the positive consequences of elections for governance and democratization, very little empirical evidence exists to support these claims, and the available evidence suggests that potential positive consequences of elections tend to obtain only if elections are of high quality. The objective of this thesis is to contribute to our understanding of these questions, describing the variation in election quality in third wave regimes around the world, attempting to understand its causal dynamics, and analyzing the consequences of such variation for democracy and democratization. These questions are essentially empirical and hence, as O'Donnell aptly notes, not likely to be helped by "either attributing magical properties to all kinds of elections or denying the importance of fair ones" (O'Donnell 2001: p. 9).

The thesis is set-up according to the following research questions:

1. How to conceptualize and measure the quality of elections?
2. What are the causal dynamics that explain variation in the quality of elections?
3. What are the consequences of variation in the quality of elections for the functioning of electoral regimes as 'instruments of democracy', i.e. the degree to which elections generate government accountability and responsiveness?
4. What are the consequences of variation in the quality of elections for the functioning of electoral regimes as 'instruments of democratization', i.e. the degree to which elections have spill-over effects on the quality of other partial regimes of democracy, such as respect for civil liberties and the rule of law?

As noted above, since much work has been done on conceptualization and measurement of election quality in recent years, the purpose of the first question is to provide an overview and evaluation of existing conceptualizations in order to propose our own “systematic concept” used in this study (Adcock and Collier 2001). Also, the aim is to evaluate different measures and their potential biases as well as to carry out reliability and validity tests of these data and propose an improved measure of election quality that can be used for comparative research.

The second question addresses causes of election quality, considering explanatory factors for variation between countries as well as variation within countries over time. Here we build both on the literature on democratization as well as political corruption to derive plausible explanatory factors for election quality and electoral fraud. The third question asks whether election quality or electoral fraud is really that important. Do high quality elections indeed generate accountable and responsive governments, as the literature on elections as “instruments of democracy” suggests? And vice versa, does electoral fraud indeed undermine accountability and government performance?

Also, turning to the fourth question, is it possible that high quality elections have spill-over effects on other partial regimes of democracy? Do high quality elections for example generate improvements in respect for civil liberties and the rule of law? Are high quality elections associated with lower levels of protests and violence, and a diminished likelihood of civil war and coup d’etats? And, finally, what if our assessment of causes of election quality indicates that civil liberties and rule of law prior to elections are associated with higher quality elections, while our evaluation of consequences of election quality finds that higher quality elections are associated with subsequent improvements in civil liberties and the rule of law? Is it possible that there are “virtuous” and “vicious” circles of election quality and other partial regimes, setting some countries on a path of democratization while keeping others

trapped? These last questions together lead to an evaluation of the fourth question, however tentative, regarding the consequences of variation in election quality for democratization.

Investigating these questions has clear relevance beyond academia, as the investment in democratic governance in third wave regimes by the international community is substantial. For example, the United Nations Development Program (UNDP) dedicated 35% of its annual budget in 2009 to democratic governance, representing 1.4 billion US dollars, a substantial part of which was dedicated to electoral assistance and election monitoring. Likewise, organizations such as the European Commission, the OSCE and the OAS, as well as other regional organizations, international NGO's, and individual governments of established democracies dedicate considerable resources to promoting democratic governance. In view of the attention given to electoral assistance and election monitoring in the field of democratic governance, generating better understanding of the causes and consequences of variation in election quality is crucial.

1.5. Outline of the thesis

As the thesis follows the research questions, its outline is rather straightforward. Chapter 2 proposes a definition of the quality of elections that is grounded both in academic work as well as international legal conventions and introduces the data collected to 'measure' election quality. This data is based on existing academic datasets, as well as additional data collected by the author based on coding of academic sources, election observation reports, media and NGO reports. Chapter 3 and 4 study variation in election quality across polities, attempting to explain why some new democracies manage to "get their elections right" while others do not. Chapter 3 is descriptive and aims to identify patterns of development of election quality over time within countries and chapter 4 tests causal explanations of cross-country and over-time variation in election quality statistically. Chapter 5 and 6 ask the "so what" question by investigating the consequences of variation in election quality: do elections of higher

quality generate more accountable and responsive governments? Finally, chapter 7 connects the findings in the earlier chapters by inquiring to what degree and how election quality affects broader democratization processes, and concludes with suggestions for policy-making and further research.

2. How to define and measure election quality?

What does it mean to “get elections right”? This chapter provides a conceptualization and operationalization of election quality and presents the data used to measure election quality in this thesis. Section 2.1 focuses on conceptualizing election quality, by first reviewing existing conceptual work and then proposing a working definition based on Adcock and Collier’s (2001) method of “background” and “systematic” concepts. Subsequently, section 2.2 addresses measurement. I first evaluate existing measures of election quality by examining their content validity and data generating process (Munck and Verkuilen 2002). All empirical studies for which cross-national data were available are examined (i.e. Anglin 1998, Pastor 1999, Van de Walle 2003, Birch 2008, 2011, Donno 2010, Elklit and Reynolds 2005a, Hartlyn, McCoy and Mustillo 2008, Kelley and Kiril 2010, Lindberg 2006a, Munck 2009, Hyde and Marinov 2010). Following the evaluation of existing measures, I present the data used to measure election quality in this thesis.

2.1. Conceptualizing election quality

Existing conceptualizations

As the quality of elections started to attract academic attention in recent years, a number of different conceptualizations emerged, as the overview in table 2.1 shows (Elklit and Svensson 1997, Pastor 1999, Schedler 2002a, Mozaffar and Schedler 2002, Elklit and Reynolds 2005a, Lindberg 2006a, Birch 2008, 2011, Munck 2009, Lopez-Pintor 2010, Davis-Roberts and Carroll 2010)²⁶. These conceptualizations differ, apart from their content to which we will turn shortly, in three important

²⁶ While most of this research focuses on third wave regimes, it is important to note that election quality is also a topic of concern for established democracies, as the case of the 2000 US presidential elections demonstrates. Of the research reviewed here, only Elklit and Reynolds (2005a), Kelley (2010) and Hyde and Marinov (2010) measure election quality across established democracies as well as third wave regimes. Since this thesis focuses on third wave regimes, emphasis lies on the latter, however it is important to note that elections in established democracies can have and do sometimes have flaws too.

aspects: firstly, whether they define election quality positively or negatively, secondly, whether they view election quality as a particular or universal concept, and finally, whether they use process or concept-based definitions.

Starting with the first, authors differ in defining election quality either positively, i.e. by specifying the presence of criteria that render elections democratic, or negatively, i.e. by identifying the norm-violations that will cause elections to be less-than-democratic or plainly un-democratic²⁷. Positive definitions use conceptual labels ranging from free and fair elections (Elklit and Svensson 1997, Anglin 1998, Lindberg 2006a), clean elections (Munck 2009), and democratic elections (O'Donnell 2001, Munck 2009) to election quality (Elklit and Reynolds 2005a, Hartlyn et al. 2008, Kelley and Kiril 2010). Negative definitions on the other hand refer to flawed elections (Pastor 1999), electoral malpractice or misconduct (Birch 2008, Donno 2010), electoral manipulation (Schedler 2002a), or electoral fraud, electoral corruption, or election rigging (Lehoucq 2003, Lopez-Pintor 2010, Simpser 2005, Calingaert 2006). An example of a positive definition is Munck's conceptualization of democratic elections: "First, elections must be inclusive, that is, all citizens must be effectively enabled to exercise their right to vote in the electoral process; second; elections must be clean, in other words, voters' preferences must be respected and faithfully registered; third; elections must be competitive, that is, they must offer the electorate an unbiased choice among alternatives; and fourth; the main public offices must be accessed through periodic elections, and the results expressed through the citizens' votes must not be reversed" (Munck 2009: 88). Examples of negative definitions are Birch's definition of electoral malpractice as "the manipulation of electoral processes and outcomes so as to substitute personal or partisan benefit for the public interest" (Birch 2011: p. 23), Lopez-Pintor's

²⁷ As will become apparent in section 2.2 however, in practice measurement indicators tend to focus on norm-violations, rendering actual measurements of election quality more comparable than the variety of conceptual labels suggests.

notion of electoral fraud as “any purposeful action taken to tamper with electoral activities and election-related materials in order to affect the results of an election, which may interfere with or thwart the will of the voters” (Lopez-Pintor 2010: p. 9), and Lehoucq’s conception of electoral fraud as “clandestine efforts to shape election results” (Lehoucq 2003: 233). Note that both positive and negative conceptualizations identify norms that should be met for elections to be high-quality or norms that should be violated for elections to be lower-quality or fraudulent.

However, while positive definitions focus on defining these norms, negative definitions add emphasis on actors, intentionality and the consequences for election outcomes. For example, Birch’s (and Schedler’s) term “manipulation” implies actor(s) involved in manipulating, as does Lopez-Pintor’s “purposeful action”. The latter also stresses intentionality, i.e. actions have to be purposeful, with the aim to “shape the election results” (Lehoucq 2003) and hence undermine the expression of the will of the voters (Lopez-Pintor 2010). Birch (2011) more broadly defines the goal of “electoral malpractice” to be “the substitution of personal or partisan benefit for the public interest”. Hence, negative definitions seem to emphasize more clearly that “malpractice” or “fraud” should be delimited to those acts that are intentionally perpetuated with the aim to change the election results. Irregularities that result from administrative incapacity are not to be considered as “malpractice” or “fraud” according to these authors, even if they might at times also seriously affect the election results; nor are instances of intentional fraud that nevertheless do not affect the election outcomes. This is problematic, and will be further discussed in section 2.2.

Table 2.1 – Different approaches to conceptualizing election quality

Author(s)	Conceptualization			
	Concept-name	Positive / Negative	Universal / Particular	Process / Concept
Hermet et al. (1978)	elections without choice	negative	universal	concept
Elklit & Svensson (1997)	free and fair elections	positive	universal	concept& process
Anglin (1998)	free and fair elections	positive	universal	concept
Pastor (1999)	flawed elections	negative	particular	concept
O'Donnell (2001)	democratic elections	positive	universal	concept
Schedler (2002a)	electoral manipulation	negative	universal	concept& process
Mozaffar & Schedler (2002)	electoral governance	neutral	universal	process
Van de Walle (2003)	free and fair elections	positive	universal	concept
Elklit & Reynolds (2005a)	election quality	positive	universal& particular	process
Simpser (2005)	electoral corruption	negative	universal	concept
Lindberg (2006a) ^a	free and fair elections	positive	universal	concept
Calingaert (2006)	election rigging	negative	universal	process
Birch (2008, 2011)	electoral malpractice	negative	universal	concept& process
Hartlyn et al. (2008)	election quality	positive	universal	concept
Munck (2009)	clean elections	positive	universal	concept
Munck (2009) ^b	democratic elections	positive	universal	concept
Donno (2010)	electoral misconduct	negative	universal	concept
Kelley and Kiril (2010)	election quality	positive	universal	concept& process
Lopez-Pintor (2010)	electoral fraud	negative	universal	concept

a. Lindberg (2006) provides a broader definition of the “quality of elections” based on participation, competition and legitimacy. However, since his definition considers the outcomes of elections as well, such as turnout and relative strength of incumbent and opposition, I have chosen to use his specific indicator of the free & fairness of the electoral process as data more appropriate to the research-focus in this thesis.

b. Munck (2009) provides both a conceptualization of “electoral democracy” as well as “democratic elections” in his 2009 book. The conceptualization of “democratic elections” is more extensive and precise, and will be discussed in the next paragraph. However, data are not available for “democratic elections”. Instead the data used for this thesis are data from the Electoral Democracy Index that includes an item on “clean elections” (Munck 2009: 83).

Apart from the variety of conceptual labels used, all conceptualizations define a benchmark, either of what constitutes a “high quality”, “clean” or “non-fraudulent” election, or of its counterpart. As Hartlyn and McCoy (2006) discuss, two approaches can be followed to set this benchmark: “quality-based” or “legitimacy-based”. While quality-based approaches define the quality of elections with reference to a universal democratic standard, often based on democratic theory and/or international law and norms, legitimacy-based approaches define the quality of elections with reference to the particular citizens and parties involved. An example of the latter is Pastor’s definition of a flawed

election as “an election in which some or all of the major political parties refuse to participate in the election or reject the results” or Elklit and Reynolds’s assertion that “The quality of an election can [...] be conceptualized as the degree to which political actors at all levels and from different political strands see the electoral process as legitimate and binding.” (Pastor 1999: 15 and Elklit and Reynold 2005b: 189). The argument in favor of a legitimacy-based approach is that elections are different in different contexts, and in the context of founding elections or post-conflict elections, even if the elections do not meet ideal democratic standards, it is ultimately up to domestic stakeholders to judge elections to be acceptable or not (Pastor 1999). Politically, this argument is indeed very convincing, and in fact election observation missions do often take into account the specific context of the election in their final judgment (Donno 2010, Kelley 2010)²⁸. However, academically, and especially for comparative research, it is imperative to use the same assessment criteria for all elections. Moreover, perceptions of legitimacy of the electoral process by citizens and parties might be dependent variables of interest in their own right (Birch 2008, 2011). Hence, in most conceptualizations a “quality-based” or “universal” approach is preferred²⁹.

A third important aspect in the conceptualization of election quality is whether to take a process-based or concept-based approach. While concept-based approaches define the quality of elections with reference to ideal democratic standards that derive from democratic theory or international law, process-based conceptualizations consider the electoral process before, during and after election-day.

²⁸ As such, Donno (2010) and Hartlyn and McCoy (2006) point out that: “founding elections, elections that lead to alternation in power or that establish peace after civil wars, generally receive less international criticism” (Donno 2010: 26).

²⁹ Another example of a “particular” approach is Lehoucq’s proposal of a legalistic conception of electoral fraud, i.e. acts are fraudulent if they break the law of the particular country under consideration. According to Lehoucq, the key advantage of this approach is that “it permits assessing the location of the boundaries between acceptable and unacceptable political activity” (Lehoucq 2003: 235). However, apart from the fact that electoral legislation might itself be flawed, this approach suffers from the same problem of cross-national comparability.

For example, in the months running up to the election, aspects such as party and voter registration, constituency and polling demarcation, and campaign regulation are important determinants of election quality. Common irregularities before election-day are then unequal access to media and campaign resources for opposition parties, inaccurate voter registration lists, opposition intimidation, and ‘creative’ designation of district boundaries so as to favor incumbents (Lehoucq 2003, Calingaert 2006, Birch 2008)³⁰. During election-day, access to polling stations, vote secrecy and access of election monitors are important. Examples of frequently occurring irregularities are lack of polling stations in rural areas, polling stations that open later and close earlier, vote-buying and threats to voters by party representatives, etc. (Calingaert 2006). Finally, after election-day important aspects are counting and tabulating the vote, resolving election related complaints, and publishing the election results. Here, problems that can occur are: biased counting of votes, not allowing for monitoring of vote counting, not publishing disaggregate results, etc. (Lehoucq 2003). Examples of process-based conceptualizations are the frameworks for election quality proposed by Elklit and Reynolds (2005a) and electoral governance by Mozaffar and Schedler (2002)³¹. Of course, process-based approaches still draw on ideal democratic standards to identify irregularities, however these are rather implicit benchmarks against which each step of the electoral process is evaluated.

³⁰ It is important to note that not only incumbents engage in electoral fraud, as opposition parties might just as well engage in such activities as voter intimidation and vote buying. However, incumbents do have a wider range of possibilities to manipulate the electoral process through changing the electoral legislation and (privileged) access to state resources.

³¹ Elklit and Reynolds (2005a) identify 11 sequential steps in the electoral process, i.e. legal framework, electoral management, constituency and polling demarcation, voter education, voter registration, access to and design of ballot paper/party and candidate nomination & registration, campaign regulation, polling, counting and tabulating the vote, resolving election related complaints, and post-election procedures; and measure the quality of each step by a number of sub-indicators. Mozaffar and Schedler (2002) distinguish the 3 phases just mentioned: rule-making (i.e. ex ante choice of rules of electoral competition and rules of electoral governance), rule application (i.e. organizing the electoral ‘game’), and rule-adjudication (i.e. ex post publication of election results and resolving disputes).

A number of scholars combine process- and concept-based approaches. For example, Elklit and Svensson (1997) propose a definition of “free and fair” elections based on democratic theory and then proceed to develop a set of indicators that measures “free- and fair-ness” before, during and after elections. Also, Kelley and Kiril (2010) distinguish between irregularities before and during election-day, and Schedler (2002a) traces the process of electoral choice from the ex ante formation of preferences and availability of choice, to the actual choice, to the consequences of choice, and identifies possible strategies of manipulation at each step. The advantage of process-based conceptualizations is that they allow for precise measurement of election quality. Elections are complex logistical operations, hence ordering the electoral process by the sequential steps taken before, during and after election day, helps to ensure that all relevant aspects are taken into account. However, this approach runs the risk of generating vast “checklists” of indicators by which to judge elections, without a clear notion of the relative importance of certain indicators vis a vis other indicators and how to aggregate such indicators in an overall measure of quality (Mozaffar and Schedler 2002)³². Moreover, data of this sort are extremely labor-intensive to gather, as empirical work by Elklit and Reynolds (2005a) shows. Hence most research takes a concept-based approach, drawing on democratic theory and/or international law to define the quality of elections.

For the purpose of this thesis, I will develop a conceptualization that defines election quality using universal criteria, and a combination of concept- and process-based approaches. As label I will use the positive and negative conceptual labels of election quality and electoral fraud interchangeably. Regarding the use of universal criteria, this is not to deny the fact that ‘logics of fraud’ differ in

³² I do not agree with Mozaffar and Schedler (2002) that the relative weight and aggregation of indicators is the most important problem of these approaches, as both are in fact empirical questions that can be assessed with appropriate latent-structure and item-response analyses. However, the practical aspect of data-gathering poses limitations for comparative research. An example of a good compromise between practical limitations and precise measurement are the data on electoral malpractice gathered by Birch (2008, 2011) and the data on election quality gathered by Kelley and Kiril (2010).

different contexts, and that elections may serve widely varying purposes, as discussed in chapter 1. For example, in a country where clientelist relations between citizens and politicians are common, practices such as vote buying might not be considered problematic (Schaffer 2007). However, to develop a definition that allows for comparison between different countries, we need to set a standard that is the same in all contexts. Since the majority of states around the world, including autocracies-in-disguise, presently hold elections that are claimed –at least by incumbents- to be democratic, I propose such a standard of ‘universal’ criteria should be based on democratic theory, and should specify the criteria that ‘ideal’ democratic elections should meet. By subsequently matching those criteria to ‘real-existing elections’, we can evaluate to what extent they meet these criteria. As mentioned above, in practice this will mean to identify what is lacking, i.e. to identify the presence and extent of irregularities that violate these norms, making the electoral process less-than-ideally-democratic³³. Moreover, I would propose to define electoral malpractice or fraud using a mix of concept- and process-based approaches. The reason for this is that a concept-based approach enables us to identify the criteria that an ‘ideal’ democratic election should meet; while a process-based approach allows us to match each of these criteria with specific elements of the electoral process. The advantage of using a combination is that the benchmark against which elections are evaluated is explicit (the advantage of concept-based approaches), while making sure the entire electoral process is taken into consideration (the advantage of a process-based approach). In the next section I discuss the content of existing conceptualizations and propose a working definition of electoral fraud based on this work.

³³ Note that ‘real existing elections’ never meet the benchmark criteria of democratic elections fully, as even established democracies experience irregularities in their electoral processes, for example with respect to campaign finance regulations or state financial support for parties based on parties’ size in parliament, which often create a non-level playing field between parties (Katz 2005).

Background concepts and systematized concepts

In order to define election quality for the purpose of this thesis, I follow the approach for the formation of concepts proposed by Adcock and Collier (2001), starting with an overview of “background” concepts that sketch the semantic field of existing research, and subsequently making a selection of “systematized” concepts that can then be linked to indicators for measurement. Similarly, Munck and Verkuilen (2002) view concepts as consisting of multiple attributes that, like branches on a tree, define the various parts of the overall concept. They propose that such attributes should meet 2 criteria: on the one hand they should be comprehensive enough so as to not neglect an important attribute of the concept under study, while on the other hand they should be as parsimonious as possible, avoiding redundancy or overlap between attributes. Adcock and Collier (2001) add that the process of concept formation might involve several iterations in order to make sure concepts indeed meet these criteria.

In this section we will map the “background” concepts based on (a) specific definitions of the quality of elections and electoral fraud, as well as (b) conceptualizations of democracy (both minimalist and expanded procedural) that specify criteria for the quality of elections, and (c) relevant international norms on elections. I will then attempt to identify a working concept of election quality that meets the criteria of comprehensiveness and parsimony. Table 2.2 below provides an overview of the different conceptualizations, including background concepts, systematized concepts and measurement indicators proposed in the different studies.

Table 2.2 – Conceptualizing election quality^a

Authors	I. Background concepts	II. Systematized concept		III. Indicators
		Attributes	Components of attributes	
Schumpeter (1975)	Conceptualizations of democracy - Political leaders are elected by citizens - Competition between political leaders for citizens' votes			
Dahl (1989)	- Elected officials - Free and fair elections - Inclusive suffrage - Right to run for office - Freedom of expression, - Alternative information - Associational autonomy	1. Elected officials	1.1. Object of choice: most powerful collective decision-makers selected through elections. 1.2. Consequences of choice: election results are irreversible & elected officials have control over policy-making.	- Elected legislature - Elected executive (directly or indirectly) - Winners of elections assume office - Elected officials have discretion over policy-making, influence of "reserved power domains" is limited.
Huntington (1991)	- Elections for most powerful decision-makers - Fair and honest elections - Periodic elections - Universal suffrage - Free competition	2. Periodic elections	2.1. Elections held at regular intervals.	- Elections held periodically, according to constitution/electoral law.
Przeworski et al. (1996)	- Ex ante uncertainty - Ex post irreversibility - Repeated elections			
Schmitter and Karl (1991)	Dahl's 7 criteria plus 2: - Absence of reserved power domains - State sovereignty	3. Participation	3.1. De jure right to vote 3.2. De facto conditions for the use of the right to vote	- De jure universal & equal suffrage - De facto possibility to participate: - Possibility to form preferences:
Hermet et al. (1978)	Conceptualizations of election quality (concept-based) - Freedom voters - Competition candidates - Effects elections on policies			- Freedom of expression - Alternative sources of information - Voter education - Possibility to cast a vote:
Elklit & Svensson (1997)	- Free elections (before, during & after polling day) - Fair elections (before, during & after polling day)			- Freedom of movement - Freedom from intimidation/coercion
O'Donnell (2001)	- Competitive, Free, Egalitarian, Decisive, Inclusive			- Voter registration (register, registration requirements)
Lindberg (2006a)	- Participation - Competition - Legitimacy			- Polling station access - Secret balloting
Munck (2009)	- Inclusive elections - Clean elections - Competitive elections - Elective public offices	4. Competition	Continued on next page.	- Votes cast are counted: - Accurate vote counting
Lopez-Pintor (2010)	- Political framework: periodic elections, free expression of the will of the people - Free and fair elections: right of the people to vote, guarantees for secret ballot and genuine elections - Right to franchise: right of universal & equal suffrage - Right to participate: right to stand for election & assume office			

Authors	I. Background concepts	II. Systematized concepts		III. Indicators
		Attributes	Components of attributes	
Schedler (2002a)	Conceptualizations of election quality (process-based) <ul style="list-style-type: none"> - Object of choice - Range of choice - Formation of preferences - Agents of choice - Expression of preferences - Aggregation of preferences - Consequences of choice 	4. Competition	4.1. De jure right to run for office 4.2. De facto conditions for the use of the right to run for office	<ul style="list-style-type: none"> - De jure multi-party/candidate elections - De facto possibility to compete: <ul style="list-style-type: none"> - Registration requirements - Ballots accurate - Freedom from intimidation/coercion - Freedom of movement, expression, assembly, association - Alternative sources of information / media coverage - Use of state resources - Campaign finance regulation - Campaign activities - Access to dispute adjudication
Mozaffar & Schedler (2002)	<ul style="list-style-type: none"> - Rule-making (choosing rules of electoral game) - Rule application (organizing electoral game) - Rule adjudication (certifying results & solving disputes) 			
Elklit & Reynolds (2005a)	<ul style="list-style-type: none"> - Legal framework - Electoral management - Constituency & polling district demarcation - Voter education - Voter registration - Ballot paper design & party/candidate registration - Campaign regulation - Polling - Counting and tabulating the vote - Resolving election related complaints, verification & certification - Post-election procedures 	5. Free elections	5.1. Freedom to form preferences and vote free of coercion (for voters) 5.2. Freedom to run in elections and compete for popular support free of coercion (for parties/candidates)	<ul style="list-style-type: none"> - Freedom of movement, expression, information. - Freedom of movement, expression, information, assembly, and association.
Birch (2008, 2011)	<ul style="list-style-type: none"> - Manipulation election rules - Manipulation voting process / election administration - Manipulation of vote choice 	6. Fair elections	6.1. "Unbiased application of rules" of electoral competition & electoral governance before, during and after elections.	<ul style="list-style-type: none"> - Impartial Electoral Management Body - Implementation according to electoral law: <ul style="list-style-type: none"> - Before elections: <ul style="list-style-type: none"> - Constituency & polling district demarcation - Ballot paper design - Registration of voters and parties - Campaign regulation - During elections: <ul style="list-style-type: none"> - Polling - After elections: <ul style="list-style-type: none"> - Counting and tabulating the vote - Resolving election complaints - Post-election procedures
Universal Declaration of Human Rights (UDHR 1948, art. 21, 1 & 3)	International legal conventions ^b <ul style="list-style-type: none"> - Everyone has the right to take part in the government of his country, directly or through freely chosen representatives; - The will of the people shall be the basis of the authority of government, this will shall be expressed in: periodic and genuine elections, which shall be by universal and equal suffrage, and shall be held by secret vote or by equivalent free voting procedures 			
International Covenant on Civil and Political Rights (ICCPR 1966, art.25a&b)	<ul style="list-style-type: none"> - Every citizen shall have the right and the opportunity, [...] without unreasonable restrictions: to take part in the conduct of public affairs, directly or through freely chosen representatives; to vote and to be elected at genuine periodic elections, which shall be by universal and equal suffrage and shall be held by secret ballot, guaranteeing the free expression of the will of the electors. 			

a. The studies by Anglin (1998), Pastor (1999), Van de Walle (2003), Hartlyn et al (2008), Donno (2010), Kelley and Kiril (2010) and Hyde and Marinov (2010) are not included here because they do not provide an elaborate conceptualization of election quality, but rather give an operational definition based on how wide-spread, intentional and consequential irregularities are, or, in the case of Pastor, whether all major parties accept the results. b. For reasons of space, I only report the conditions mentioned for elections in the UDHR and ICCPR. Other instruments are described in the text.

As table 2.2 shows, though authors distinguish different “background” concepts, the main attributes deemed necessary for democratic elections are remarkably similar. Three sets of common attributes can be identified: (1) elected officials and periodic elections, (2) participation and competition, (3) free and fair elections. The first two criteria could be seen as prior conditions of democratic elections, relating to the scope and timing of elections, which are often left implicit in definitions of election quality (with the exception of Munck 2009 and Schedler 2002a), but are more common in broader definitions of democracy. Elected officials implies that “control over government decisions about policy is constitutionally vested in elected officials” (Dahl 1989: 221), i.e. the “most powerful collective decision makers are selected through elections” (Huntington 1991, O’Donnell 2001, Munck 2009, Schedler 2002a, Przeworski et al 2000; Schmitter and Karl 1991). This specification is important because it determines the object of choice in elections (Schedler 2002a), and hence excludes elections in political systems where the discretion of elected officials is limited by “reserved power domains”, such as the monarchy or the army³⁴. It also implies that elections should be irreversible (Przeworski et al. 2000), i.e. that elected officials should be able to assume office after elections. This refers for example to elections that are immediately followed by coups d’etat or civil war, that prevent elected governments from assuming office (Lindberg 2006a). The criterion of periodic elections simply specifies that elections should be held with some regularity, to allow citizens to hold their governments accountable and to generate incentives for governments to be responsive to citizens’ needs (Przeworski, Stokes, and Manin 1999).

³⁴ Examples are present-day monarchies such as Morocco, where parliamentary elections are held, but collective decision-making authority lies primarily with the king, or military systems such as Brazil between 1964 and 1982, which also held parliamentary elections while the core executive consisted of military officials. Another present-day example of limited discretion of elected officials are countries that are not fully sovereign, as international protectorates like Bosnia-Herzegovina since the 1995 Dayton Accords. In such a case, executive discretion is not limited by domestic actors, but by international actors who have considerable influence on government decisions. The latter applied to many former colonies in sub-Saharan Africa too, that often held competitive elections before achieving independence.

The second set of criteria represent the main substantive elements of election quality, and are directly derived from Dahl's (1971) notion of participation and contestation as the two core elements of democracy. Participation refers to who is allowed to cast a vote in elections, i.e. the "agents of choice" (Schedler 2002a). In the context of present-day democracies this implies universal suffrage without restrictions based on income, gender, or literacy³⁵. However, it also implies that voters are in practice able to exercise their right of suffrage, requiring accurate voter registration, access to polling stations, etc. This criterion is also referred to as "inclusive elections" (Munck 2009, O'Donnell 2001, Schedler 2002a). De facto participation also includes "equal suffrage" (Lopez-Pintor 2010, O'Donnell 2001, UDHR 1948 and ICCPR 1996). Equal suffrage refers to the principle of political equality of all citizens, i.e. "one-man-one-vote", indicating that each citizen should have one vote (or, if citizens have multiple votes, all citizens should have the same number of votes) and that citizens' votes should be weighted equally in the counting and tabulation of votes³⁶.

On the other hand contestation, or competition as it is more commonly called, refers to who is allowed to run for office in elections, determining the range of choice available to voters (Wessels and Schmitt 2008, Schedler 2002a). This implies both the legal possibility for multiple candidates and parties to run for office as well as practical possibilities to campaign and compete for popular support. Hence the quality of elections in terms of participation and competition depends both on *de jure* universal suffrage and multi-party competition, as well as on the *de facto* possibilities for citizens to vote in elections and for parties/candidates to compete in elections (Lindberg 2006a). The de facto possibilities for participation and competition depend on a number of "primary" political and civil

³⁵ Though almost all countries around the world nowadays have –at least de jure– universal suffrage, restrictions related to age and occupational groups as police and army officials are still common and generally not considered un-democratic. Blais et al. (2001) provide an overview of suffrage legislation in countries around the world.

³⁶ Malapportionment of electoral systems is an important concern in this respect as it causes the votes of citizens in over-represented districts to be more heavily weighted (Snyder and Samuels 2001).

rights, often described under the label of “free and fair” elections (O’Donnell 2001, Elklit and Svensson 1997, Lindberg 2006a).

This brings us to the third set of attributes, free and fair elections. Elklit and Svensson define freedom as the opposite of coercion, implying choice: “freedom entails the right and the opportunity to choose one thing over another” (Elklit and Svensson 1997: 35). Free elections have been defined more narrowly with respect to the vote choice itself, as well as more broadly, including (a) the formation of preferences prior to the vote choice, and (b) conditions that are necessary for de facto inclusiveness and competitiveness of elections. Conceptualizations that focus on the vote choice stress the importance of the free expression of preferences by citizens through the vote, which implies secret ballots, but also absence of voter coercion through intimidation or vote buying (Birch 2008, Schedler 2002a). Other authors define free elections more broadly, adding conditions necessary for the free formation of preferences, like freedom of expression and alternative information (Dahl 1989, O’Donnell 2001); as well as conditions for de facto inclusiveness, such as freedom of movement, and conditions for de facto competitiveness, such as freedom of association and assembly (Elklit and Svensson 1997). As regards fair elections, most authors define fairness as “the unbiased application of rules” (Elklit and Svensson 1997: 35). This refers to the “impartial application of the rules of electoral competition and electoral governance”, and asks whether the elections were carried out in accordance with those rules (Elklit and Svensson 1997, Mozaffar and Schedler 2002). This involves aspects as vote counting and tabulation procedures, post-election adjudication of disputes, etc. (Birch 2008). Some authors also define “fairness” as the degree to which elections provide a level playing field for competition between parties. Elklit and Svensson (1997) define this element of fairness as “reasonableness”, i.e. the not-too-unequal-distribution of relevant resources among competitors” (pp. 35). This refers to aspects as access to media and campaign resources (Birch 2008, Elklit and Reynolds 2005a).

Undoubtedly, these attributes of election quality cannot always be clearly separated, and actually overlap and interrelate to a certain extent. For example, the “free- and fair-ness” of elections determines to an important extent the de facto possibilities for voters to participate and for parties/candidates to compete in elections, as will be discussed more in-depth in the next section. However, the fact that different conceptualizations consistently refer to similar core attributes indicates that some common agreement exists on the basis of which a definition of election quality can be developed. Such a definition would not only be grounded in democratic theory, but also build on international legal norms for elections, to which we will turn now.

International law

Turning to the international legal framework for elections, we find similar “systematized” concepts as the ones discussed above. International legal norms for elections have been specified in the Universal Declaration of Human Rights (UDHR), the International Covenant on Civil and Political Rights (ICCPR) and numerous regional treaties and political commitments by intergovernmental organizations such as the European Union (EU), the African Union (AU), the Economic Community of West African States (ECOWAS), the Organization of American States (OAS), the Commonwealth of Independent States (CIS), the Council of Europe (CoE), the Organization for Security and Co-operation in Europe (OSCE), the Commonwealth, and the League of Arab States (LAS) (European Commission 2007). Already in 1948, the Universal Declaration of Human Rights specified that “Everyone has the right to take part in the government of his country, directly or through freely chosen representatives” and that “The will of the people shall be the basis of the authority of government, this will shall be expressed in periodic and genuine elections which shall be by universal and equal suffrage and shall be held by secret vote or by equivalent free voting procedures” (UDHR 1948, Article 21, section 1 and 3). Though the word “democracy” is not explicitly mentioned, the elements of elected officials, periodic elections, participation and free elections are clearly included

in this article. Also, the criterion of “genuine” elections may be broadly interpreted as bringing in “political freedoms and rights as the freedom of expression, assembly, association and movement”, and more narrowly interpreted as including “a real choice for voters between political contestants” (European Commission 2007: 6). Hence the element of competition is also included, albeit indirectly.

While the Universal Declaration is a non-binding resolution³⁷, the subsequently formulated International Covenant on Civil and Political Rights (ICCPR 1966) is legally binding for states who have ratified it, and specifies with regard to elections that: “Every citizen shall have the right and the opportunity, without any of the distinctions mentioned in article 2, without unreasonable restrictions: (a) to take part in the conduct of public affairs, directly or through freely chosen representatives; (b) to vote and to be elected at genuine periodic elections which shall be by universal and equal suffrage and shall be held by secret ballot, guaranteeing the free expression of the will of the electors” (Article 25, section a and b). The ICCPR clearly endorses the elements of periodic, inclusive and free elections. The reference to article 2, the non-discrimination article in the ICCPR, implies that article 25 applies to all citizens. Moreover, it clarifies the right to run for elections, making the element of competition a more explicit requirement of elections. The ICCPR specifies less clearly that elections should be held for national government, instead broadening the scope to “public affairs”. This can be interpreted

³⁷ The Universal Declaration of Human Rights is a so-called “non-treaty standard” and some of its provisions are considered as customary international law. However, it seems that article 21 is not part of these provisions yet (European Commission 2007). The General Assembly Resolution on Enhancing the Effectiveness of the Principle of Periodic and Genuine Elections is also a non-treaty standard. However, the International Covenant on Civil and Political Rights is a treaty, and is therefore binding under international law for states that have ratified the treaty. Moreover, article 25 of the ICCPR has received a General Comment by the UN Human Rights Committee (1996), specifying their interpretation of the article. Even if the General Comment only gives recommendations to states about the implementation of article 25, the “general comments are relied upon by the monitoring bodies in evaluating the compliance of states with their obligations under the treaty in question”, and thus states could be held accountable on the basis of the criteria outlined in the General Comment (European Commission, 2007: 27). As Davis-Roberts and Carroll (2010) argue, both non-treaty standards and interpretative documents such as the General Comment can function as evidence of customary international law.

however as implying that “participation should not only take place in relation to the national government, but also in relation to other levels and forms of administration, such as regional and local government levels” (European Commission 2007: 8). Moreover, the reference to “the right and opportunity” of citizens can be taken to imply that states should guarantee both de jure and de facto possibilities to participate for citizens (European Commission 2007: 8). The ICCPR has been ratified by almost 80% of the world’s states (145 states), and hence provides a quite good indication of prevailing international norms with regard to the quality of elections³⁸.

Another important standard is the UN General Assembly Resolution A/RES/46/137 on Enhancing the Effectiveness of the Principle of Periodic and Genuine Elections (1991) that, even though a non-treaty standard, is more specific about the conditions for competition in elections: “determining the will of the people requires an electoral process that provides an equal opportunity for all citizens to become candidates and put forward their political views, individually and in cooperation with others”; and emphasizes that equal and universal suffrage implies the absence of discrimination on the grounds of race and color. Finally, the General Comment (GC) by the UN Human Rights Committee on article 25 of the ICCPR (1996) is an even more detailed interpretation of the provisions of article 25. For example, section 3 of the GC reiterates that “No distinctions are permitted between citizens in the enjoyment of these rights on the grounds of race, color, sex, language, religion, political or other opinion, national or social origin, property, birth of other status”, stressing the reference to article 2 in the preamble of article 25. Also, with respect to elected officials the GC points out that: “Where citizens participate in the conduct of public affairs through freely chosen representatives, it is implicit in article 25 that those representatives do in fact exercise governmental power and that they are

³⁸ Based on European Commission data (2007: 50), of the 184 states within the UN system, 145 had ratified the ICCPR. This does not include states that only signed but did not ratify the treaty, nor those states that ratified the treaty with reservations, i.e. indicating exceptions to the principles outlined in the treaty.

accountable through the electoral process for their exercise of that power” (section 7). Regarding periodic elections, the GC specifies that: “elections must be held at intervals which are not unduly long and which ensure that the authority of government continues to be based on the free expression of the will of electors” (section 9). As regards universal suffrage, the GC clarifies reasonable and unreasonable restrictions on suffrage: “The right to vote at elections [...] must be established by law and may be subject only to reasonable restrictions, such as setting a minimum age limit for the right to vote. It is unreasonable to restrict the right to vote on the ground of physical disability or to impose literacy, educational, or property requirements” (section 10). Suffrage limitations for citizens with “established mental incapacity” are considered reasonable however (section 4). Moreover, with respect to competition, the GC indicates that: “Any restrictions on the right to stand for election, such as minimum age, must be justifiable on objective and reasonable criteria. Persons who are otherwise eligible to stand for election should not be excluded by unreasonable or discriminatory requirements such as education, residence or descent, or by reason of political affiliation” (section 15). Finally, the GC also notes that: “elections must be conducted fairly and freely” (section 19) and specifies conditions for elections to be free and fair, such as: “Freedom of expression, assembly and association are essential conditions for the effective exercise of the right to vote and must be fully protected.” (section 12).

Apart from these treaties and non-treaty standards, several regional intergovernmental organizations have made political commitments on standards for elections, such as OSCE, OAS and AU, that are often even more specific in their requirements and explicitly include additional political rights and freedoms as well as criteria relating to the fairness of elections (cf. European Commission, pp. 16-22). Finally, organizations such as the Inter Parliamentary Union and the Venice Commission have done

important work in further specifying requirements for democratic elections³⁹. Concluding, the overview of background concepts results in a systematized concept consisting of 6 core attributes that are derived from democratic theory, international legal conventions and existing conceptualizations of election quality: (1) elected officials and periodic elections, (2) participation and competition, and (3) free and fair elections. In the next section we will build on these notions to propose a working definition of election quality.

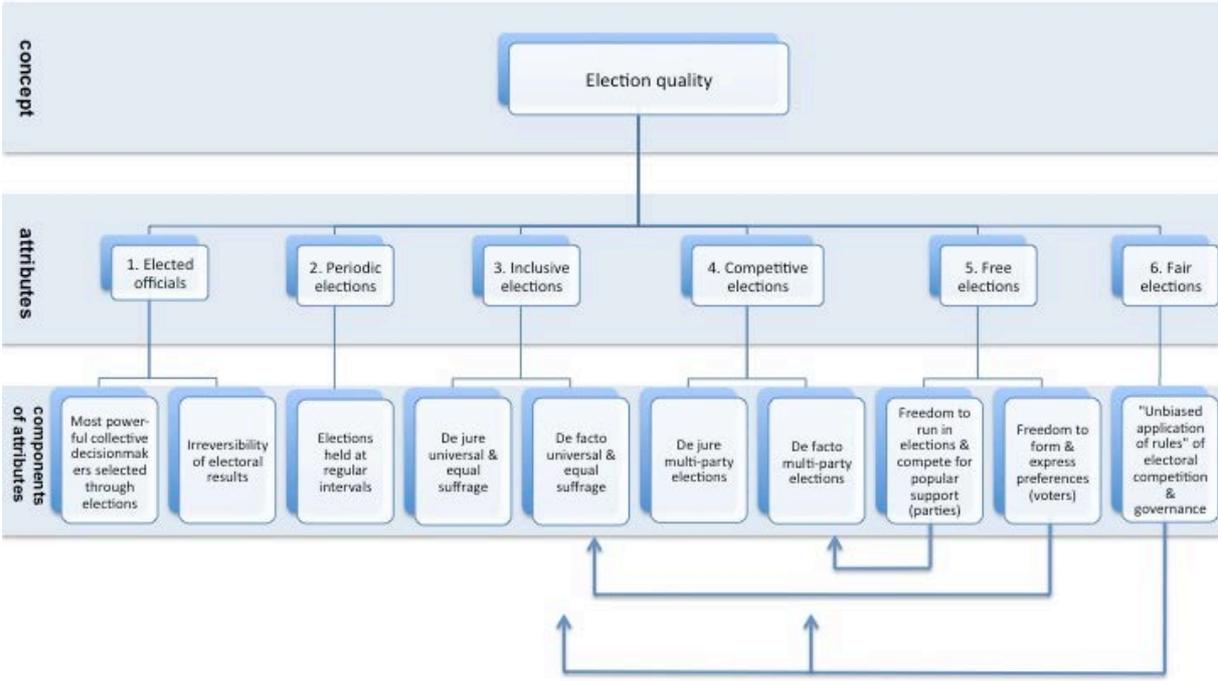
A working definition of election quality

In order to develop a working definition of election quality two issues need to be addressed. In this section I will first consider to what extent the preliminary systematized concept outlined above meets the criteria of comprehensiveness and parsimony. Subsequently, I will address the question of identifying lower and upper boundaries of election quality. As Adcock and Collier (2001) point out, often flaws in concept-formation become apparent when moving from the systematized concept and its attributes to indicators for measurement. Mapping table 2.2 graphically shows how moving from attributes to components of attributes clarifies the overlap between “free and fair” elections and de facto participation and competition.

This becomes especially clear when considering the indicators of measurement (shown in table 2.2.). For example, among the indicators proposed to measure “de facto conditions for the use of the right to vote”, we find freedom of expression, freedom of movement and freedom from intimidation in relation to the voting act, all crucial for the participation of voters in the electoral process. Likewise, among the indicators to measure “de facto conditions for the use of the right to run for office”, we find, in addition to freedom of expression, movement and intimidation for candidates and/or parties, also

³⁹ For a comprehensive overview of all international legal instruments relating to elections, see Davis-Roberts and Carroll (2010), Goodwin-Gill (2006) and Boda (2005).

freedom of assembly and association. These freedoms are all mentioned as indicators for “free elections” as well, and are hence duplicated in this conceptualization. The overlap is apparent even when just considering the components of attributes, as shown in table 2.2 and the figure below.



The attribute “free elections” has 2 components: the freedom to form preferences and vote free of coercion (for voters), and the freedom to run in elections and compete for popular support free of coercion (for parties/candidates). Clearly, freedom of voters refers to the degree to which de facto possibilities for participation exist, while freedom of parties/candidates refers to the de facto possibilities for competition.

Similarly, the “unbiased application of rules of electoral governance and competition”, i.e. fairness of elections, affects the de facto possibilities to participate and compete in many ways, ranging from barriers to registration of parties and voters, to unequal access to media and state resources during the campaign period, to the actual vote and accurate counting of votes, etc. More specifically, fairness as relating to participation refers to voter registration, as well as equal suffrage, implying the principle of

one-man-one-vote and hence the right to have one's vote counted accurately. In relation to competition, fairness implies access of parties/candidates to registration and ballots, a not-too-un-level playing field (in terms of access to resources and media) between parties during the campaign, and impartial adjudication of electoral disputes, both during and after elections.

Hence, the “free & fair” -ness of elections determine to an important extent the de facto possibilities to participate and compete in elections, leaving the question whether “free & fair” should be conceptualized separately or as integral parts of participation and competition. I propose to do the latter, considering “free” and “fair” as qualities that electoral processes can possess in greater and lesser degree and that determine the extent to which elections allow for de facto participation and competition. This creates a sharper conceptualization of election quality with less overlap between components, by grouping the specific aspects of free- and fair elections that affect the de facto possibility to participate and compete in elections under participation and competition respectively⁴⁰.

A second problem with the conceptualization proposed is that some of the attributes seem to be of an “either-or” nature, while other attributes seem to be rather a “matter of degree”. For example, the attribute of elected officials is of an ‘either/or’ nature, as legislatures and executives are either selected

⁴⁰ With respect to fairness, this is relatively straightforward, since “fairness” refers to specific aspects of the electoral process that can easily be related to participation and competition (view Appendix A). For example, impartial voter and party registration, accuracy of voter registers and ballots, and campaign regulation can all be related relatively straightforwardly to participation or competition. On the other hand, freedom is more generic and relates to the general environment of civil and political liberties that is present also in between elections. We propose to measure freedom more specifically, by measuring the specific aspects of the electoral process that will be affected if freedoms are not fully guaranteed. For example, “polling station access” is a more specific indicator of de facto inclusive elections than “freedom of movement”, which refers to generic freedom of movement, also when no elections are held. In the same way, for competition, we could measure “party registration requirements” to see whether there were severe obstacles to party registration and include that in the measure of “competitive elections”; instead of measuring the more generic “freedom of association”. This is important if we would like to test whether generic civil liberties are indeed pre-conditions for high quality elections, and also if we would like to test the consequences of elections of varying quality for subsequent levels of civil liberties, as has been done by Lindberg (2006a; 2009).

through popular elections or by other methods, such as hereditary succession or appointment. The same applies to de jure participation and competition, as the constitution or electoral law either allows for universal and equal suffrage and the right for multiple candidates and/or parties to run for office, or not. Of course, these attributes could also be measured in a continuous way, as legal restrictions on suffrage and the right to run for office are present in all democracies to varying degrees; and the discretion of elected officials over policy-making can certainly be operationalized as a “matter-of-degree” indicator as well. However, for these attributes, it seems that both democratic theory and international legal conventions provide clear guidelines to define a minimal threshold below which elections can be considered non-democratic.

On the other hand, the de facto conditions for participation and competition seem to be rather a matter of degree, for which it is more difficult to determine when violations are severe enough to render elections undemocratic. When measuring these matter-of-degree attributes, most authors classify elections in 3 categories, distinguishing between elections with severe irregularities, elections with less severe irregularities and elections without irregularities (Munck 2009, Lindberg 2006a, Donno 2010; Hartlyn et al 2008). Clearly, in this case, it is rather more difficult to provide a clear-cut distinction between non-democratic and democratic elections: perhaps elections with severe irregularities can be considered non-democratic and elections without irregularities as democratic, but it is less clear where the elections with moderate irregularities would fit. Since the aim of this thesis is to study the full variety of election quality in third wave regimes, defining an a-priori theoretical threshold for these attributes seems unproductive. Rather, such matter-of-degree attributes are more logically conceived of along a continuum running from low to high election quality. Using the “either-or” attributes identified above, I would then propose to set a lower bound, or minimal threshold, that determines whether elections have the potential to be democratic (cf. Lindberg 2006a). This follows Lindberg’s (2006a) and Elklit and Reynold’s (2005a) propositions that election quality comprises both

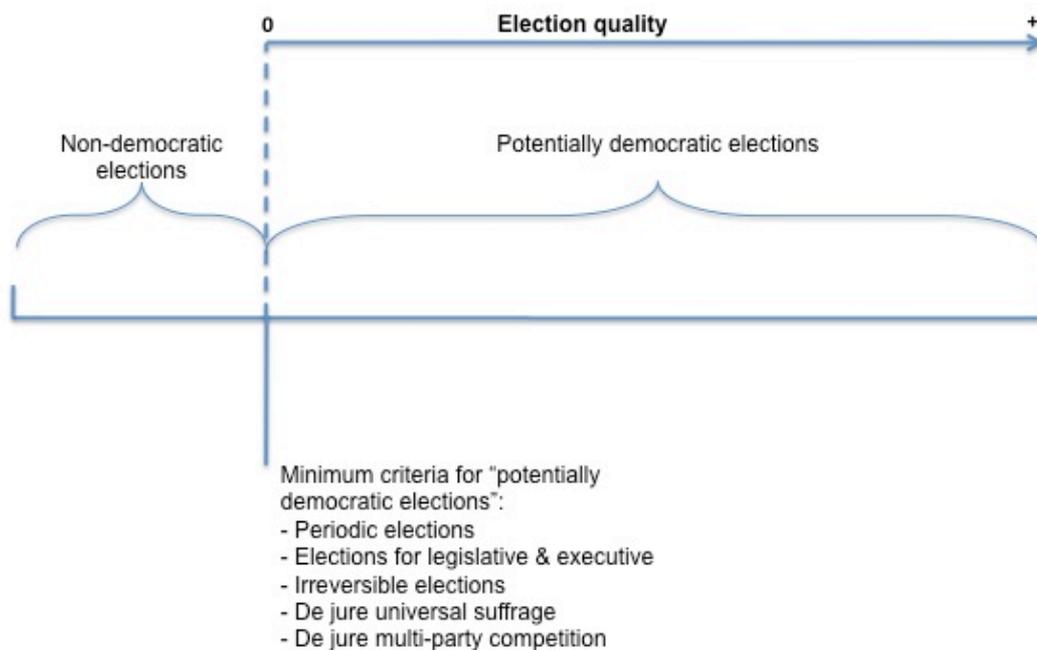
the formal institutions of electoral governance and competition as well as the actual conduct of elections, i.e. the “de jure and de facto rules which shape elections, the written laws and the practical realities” (Elklit and Reynolds 2005a: 150).

Considering elected officials, periodic elections and de jure participation and competition to be the “either-or” attributes of democratic elections, I propose to use these attributes to define the difference between “potentially democratic” elections and clearly non-democratic elections. Hence, ‘potentially democratic elections’ meet the minimal criteria that: elections are held periodically, elections are held for legislative and executive offices (directly or indirectly), winners of elections are allowed to assume office, elected officials have discretion over policy-making, and elections are held under de jure universal suffrage and de jure right to run in elections for multiple parties and/or candidates. The variation in the quality of elections within potentially democratic elections then depends on the de facto possibilities for participation and competition⁴¹. In this sense, the attributes of elected officials, periodic elections and de jure participation and competition can be seen as being higher on the ladder of abstraction, defining the difference of *kind* between categories; while de facto participation and competition define the differences of *degree* within one of these categories (Sartori 1970).

Summarizing, using the “either-or” attributes of election quality we can set a lower threshold, separating clearly non-democratic elections from “potentially democratic elections”. Within

⁴¹ Note that, contrary to Lindberg (2006a) and Elklit and Reynolds (2005a), de facto possibilities to participate and compete do not include the outcomes of elections, i.e. eventual turnout levels and electoral vote and seat shares indicating actual competition between parties. I prefer to follow Sartori (1976) and Hyde and Marinov (2010) here, as in my view, what makes an electoral process high quality is its potential for competition, rather than its actual competition. There are many reasons for this, not only because flawed elections might lead to flawed data about actual participation and competition, but also because the relation between the electoral process and these outcomes might be of interest in itself (i.e. was fraud effective), and finally, because a highly competitive outcome is not necessarily good for democracy.

“potentially democratic elections”, variation in election quality is determined by the “matter of degree” attributes, i.e. de facto possibilities for voters to participate and parties/candidates to compete. Note that within the group of “potentially democratic” elections, there will still be low quality elections that are so severely flawed that many scholars would consider them undemocratic. However, in order to study the full variety of election quality in third wave regimes, and lacking theoretical justification, I preferred not to set an a priori cut-off point. Finally, while we have defined a lower threshold for clearly non-democratic elections, and we can differentiate lower-quality and higher-quality elections as a matter of degree, the upper bound of fully democratic elections has the flavor of an ideal type and seems unlikely to have an empirically existing referent. Even in long established democracies, the playing field between parties is never fully level, nor is suffrage ever fully universal in practice, etc. (Katz 2005). However, taking ideal democratic norms for elections as a benchmark to measure deviation from those norms seems a useful way ahead to measure variation in de facto election quality in “potentially democratic” elections. The figure below shows the minimal threshold graphically.



The cut-off line is indicated on the lower left-side of the graph: these are the either/or criteria that

differentiate non-democratic elections from “potentially democratic” elections. The upper right-side of the graph shows how, within the group of “potentially democratic” elections, election quality varies from low (0) to high (1) as the matter-of-degree criteria, i.e. de facto possibilities for voters to participate and parties/candidates to compete, increase.

2.2. Measuring election quality

In this section we turn to the measurement of election quality, comparing 12 academic datasets that include measures of the quality of elections in various regions of the world (i.e. Anglin 1998, Pastor 1999, Van de Walle 2003, Elklit and Reynolds 2005a, Simpser 2005, Lindberg 2006a, Birch 2008, 2011, Hartlyn et al. 2008, Munck 2009, Donno 2010, Kelley and Kiril 2010, Hyde and Marinov 2010). The data sources are compared following the approach for evaluating the validity of democracy indices used by Munck and Verkuilen (2002). Subsequently, I assess the validity of the different measures by firstly assessing the indicators and measurement level used by authors (from concept to indicators), and secondly, evaluating the data generating process (from indicators to data). The former involves a comparison of data based on overall judgments of election quality with data based on more specific indicators of irregularities before, during and after election-day. The latter involves an evaluation of sources used and their potential biases, the coding process, as well as comparisons between different data-sources coding the same elections to assess disagreement between data-sources. In section 2.3 I then present the data used for this thesis.

Evaluating existing measures

In this section, I examine existing datasets on election quality (i.e. Anglin 1998, Pastor 1999, Van de Walle 2003, Elklit and Reynolds 2005a, Simpser 2005, Lindberg 2006a, Birch 2008, 2011, Hartlyn et al. 2008, Munck 2009, Donno 2010, Kelley and Kiril 2010, Hyde and Marinov 2010). The datasets considered are all cross-national, allowing for comparative research. Survey-data measuring citizens’

perceptions of election quality are excluded, due to their relatively limited coverage of countries and, more importantly, limited coverage of time-periods⁴². Assessments of electoral fraud based on “election forensics”, i.e. statistical analyses of vote and turnout results at local and regional levels within a country to detect deviations that might point to fraud, are also excluded since these methods tend to be applied to single countries and are hence less useful for comparative research (Mebane 2007, Alvarez et al. 2008, Myagkov et al. 2009). Finally, a number of comparative datasets are excluded due to very limited time coverage or too narrow conceptualization of electoral fraud. As such the Freedom House sub-score on the ‘electoral process’ as well as the Economist Intelligence Unit’s (EIU) sub-score on ‘electoral process and pluralism’ are excluded since both are available only as of 2006. The fraud indicator in the World Bank’s Database of Political Institutions (Kefer 2009) is excluded since it employs a rather narrow notion of electoral fraud as “vote fraud or candidate intimidation” and has been found to have a large number of false negatives (Birch 2011, Simpser 2005). Only one cross-national dataset was excluded for practical reasons, as it was not available yet at the time of writing of this thesis, Schmeets’ (2010) analysis of 81 elections in the OSCE region based on OSCE election observation reports. Table 2.3 below provides an overview of the empirical studies examined.

In order to evaluate measurement validity, i.e. whether the measurements used actually capture the phenomenon of interest (cf. Adcock and Collier 2001 for a discussion), we assess (1) the link between systematized concepts and measurement indicators and (2) the link between measurement indicators and data scores, i.e. the data generating process. For the first, the question is which and how many indicators to use, and whether those indicators adequately capture the concept one wants to measure. Related to this is the question of how to weight indicators and aggregate them into an overall score of

⁴² Surveys that ask citizens about perceptions of election quality include the Comparative Study of Electoral Systems, The International Social Survey Project, the Afrobarometer and Latinobarometer (Birch 2011).

election quality. For the second, we need to consider level of measurement, as well as the sources from which data are drawn and the coding procedures used to gather the data.

Table 2.3 – Data sources measuring election quality

Author(s)	Concept-name	Measurement			Aggregation
		Indicators	Scale	Sources ^a	Rule
Anglin (1998)	free& fair elections	2 main components	Ordinal, 3 categories	Author's coding	-
Pastor (1999)	flawed elections	1, overall indicator	Dichotomous	Journal of Democracy Election Watch Reports	-
Van de Walle (2003)	free& fair elections	1, overall indicator	Ordinal, 3 categories	Author's coding	-
Elklit and Reynolds (2005a) ^b	election quality	54 (11 main components & 54 sub-components)	Ordinal, 4 categories	Polling evidence, voter surveys, stakeholder surveys, register data, election data, national legislation. Election observation reports Election/country experts	Addition, but with weights (partially compensatory)
Simpser (2005)	electoral fraud	1, overall indicator	Dichotomous	Secondary data Database of Political Institutions & Pastor (1999); plus author's coding of academic data handbooks, EIU country profiles, JoD Election Watch Reports	-
Lindberg (2006a)	free& fair elections	1, overall indicator	Ordinal, 4 categories	Election observation reports News/media sources Historical information Country experts	-
Birch (2008, 2011)	electoral malpractice	14 (3 main components & 14 sub-components)	Ordinal, 5 categories	Election observation reports	Mean of means (partially compensatory)
Hartlyn, McCoy and Mustillo (2008)	election quality	1, overall indicator	Ordinal, 3 categories	Election observation reports, if no election mission, then news sources Validated by country experts Country experts	-
Munck (2009) ^c	clean elections	1, overall indicator	Ordinal, 3 categories	Country experts	-
Donno (2010)	electoral misconduct	1, overall indicator	Dichotomous & Ordinal, 3 categories	Election observation reports, if no election mission, then news & historical sources	-
Kelley and Kiril (2010 QED)	election quality	9 (4 main components & 9 subcomponents)	Ordinal, 3 and 4 categories	US State Department Country Reports on Human Rights	-
Hyde and Marinov (2010) ^d	-	-	Dichotomous	Election observation reports Academic data-handbooks News and historical information	-

a. Note that Anglin (1998) and Van de Walle (2003) did not specify which sources they used, indicated as “author’s coding”.

b. Elklit and Reynolds (2005a) use many different sources, most of which are mentioned here. However, for an overview of the sources used for each indicator we refer to the original article.

c. Munck (2009) provides a definition of “democratic elections” and a coding framework to code election quality, but no data are available yet. The data used here are data from his Electoral Democracy Index, item “clean elections” (p. 83). Note that, though officially coded in 3 categories, pluses and minuses are used to record intermediary situations.

d. While we will use data from Hyde and Marinov (2010), their dataset is not primarily aimed at measuring election quality or electoral fraud, but rather focuses on “competitive elections” and includes a number of variables relating to election quality. These variables include one direct question about the evaluation of the election by Western international monitors (were there allegations by Western monitors of significant vote-fraud?) and 6 questions that relate to aspects of election quality such as opposition harassment, media bias, and electoral violence. View Appendix C for an overview.

From systematic concept to measurement indicators

Starting with indicators, the quality of elections is not easy to measure, both due to the complexity of electoral processes and due to the covert nature of electoral manipulation. Ideally one would like to break up the electoral process into multiple indicators as discussed in section 2.1, and measure each of these indicators separately. Such an approach would allow one to investigate the full range of irregularities that could undermine the electoral process and thereby reduce measurement error⁴³. Indeed, many scholars have pointed out that overall judgments of election quality (i.e. single overall indicator) by international election observers tend to be focused on election-day, neglecting irregularities in the period before the elections. Both Anglin (1998) and Bratton (1998) stress that in Sub-Saharan Africa, electoral fraud occurred more often before elections, taking the form of serious campaign irregularities, vote buying and political intimidation, combined with relatively peaceful and orderly elections on the day itself. If, as Carothers (1997) and Simpser (2008) have warned, the increased frequency and effectiveness of election monitoring has led “efforts by entrenched leaders to manipulate electoral processes [...] to become more subtle” (Carothers, 1997: 22), shifting irregularities to the period before the elections, data that does not take into account the pre-election environment might seriously underestimate the extent of electoral fraud. However, empirical analyses by Kelley (2009a) seem to demonstrate that overall judgments of elections by international election observers are significantly related to cheating in the pre-election period, implying that this bias towards election day might not be as prominent (anymore)⁴⁴. At the same time, she does find that

⁴³ Such data would also allow one to assess the relative frequencies with which different types of irregularities occur, which irregularities are more likely under which circumstances, etc. and hence allow for more precise investigation of the dynamics of electoral fraud.

⁴⁴ As Bjornlund (2004) and Schmeets (2010) describe, in the course of the 1990s international election observation missions, most notably those carried out by the OSCE, EU, and international NGO's such as the Carter Centre, have significantly professionalized and improved their methodology of observation, including the deployment of long-term observers to observe irregularities before and after election day. We will further discuss the implications of changes in observation methodology for data quality when discussing sources.

administrative irregularities such as inaccurate voter registration lists are not significantly related to overall judgments of elections (Kelley 2009a: 777)⁴⁵. On the other hand, Schmeets (2010) finds that in the OSCE missions, observers' overall assessments of the electoral process reflect the particular irregularities quite well, leaving this question open for further testing.

Apart from the question of whether single indicator measurement of electoral fraud covers the entire electoral process, some scholars have proposed that certain irregularities should weigh more heavily than others in the overall evaluation of election quality (Schmeets 2010, Elklit and Svensson 1997). For example, Elklit and Svensson (1997) argue that in judging the fairness of elections, the unbiased application of electoral rules is more important for overall election quality than creating a level playing field. If such is the case, scores on media access and campaign resources should have a smaller impact on aggregate scores than the scores for unbiased application of electoral rules. Judgments about weighting can be justified theoretically, as in the example just given, but can also be based on empirical assessment of the frequency with which irregularities occur. For example, Schmeets (2010) argues that the severeness of irregularities should be judged by how frequently they have occurred in previous elections in the region. In this vein, he shows that in the OSCE missions, even though all individual irregularities are significantly related to overall judgments of election quality, ballot box stuffing is

⁴⁵ Kelley distinguishes 5 groups of specific irregularities in this study: A variable called structural problems captures the degree of problems in the legal framework for elections. This includes restrictions on the scope of the elected office, restrictions on who can vote and stand for office, and rules and regulations guiding the supervision, funding, and conduct of the election. Pre-election capacity problems includes assessment of four administrative areas: problems in voter lists or registration, complaints about electoral commission conduct, voter information problems and procedural problems, and technical difficulties. Pre-election cheating includes assessment of four categories: intimidation, media, freedom to campaign, and improper use of public funds. Election-day capacity problems captures the degree of irregularities in four categories: informational insufficiencies, administrative insufficiencies, problems in voter lists, and complaints about electoral commission conduct. Finally, election-day cheating assesses three categories: vote processing, voter fraud, and intimidation" (Kelley 2009a: 774). She then related these specific irregularities to the overall judgment of elections by observers to assess the degree of potential bias in these overall evaluations.

considered as a more serious violation, which occurs less frequently and affects the overall evaluation of elections more heavily than for example family voting (i.e. elderly men voting for the women and sometimes also younger men in their families). However, what is problematic about weighting, either based on theoretical justifications or frequency, is that it assumes that certain irregularities violate the norms of inclusiveness and competitiveness more severely than others, sometimes even suggesting a ranking of norms that seems a bit arbitrary. For example, even if family voting occurs frequently in the OSCE region, it is a fundamental violation of ballot secrecy that might disenfranchise a substantial part of the (female) citizenry, violating the principles of universal and free suffrage. Likewise, as Levitsky and Way (2010) have shown, a severely tilted playing field due to for example incumbents' privileged access to state resources and media is what keeps many incumbents in electoral authoritarian regimes in power, representing a subtler and hence easier-to-get-away-with form of manipulation that potentially even eliminates the need for election-day fraud. Hence, I would argue that specific indicators of election quality are "non-compensatory" (Munck 2009), meaning that higher scores on certain indicators cannot compensate for lower scores on other indicators, and that indicators should be weighted equally.

A third question that needs to be addressed when evaluating whether measurement indicators measure the concept they are supposed to measure is whether election quality is a uni-dimensional or multi-dimensional concept. If specific irregularities co-vary in the same way across elections this indicates they measure a single underlying concept, in which case, measuring election quality with a smaller set of indicators or even a single indicator might decrease precision of measurement, but will nevertheless still tap part of the concept of interest. If however certain irregularities co-vary in opposite directions (as when there is a trade-off between irregularities before and during election day), this indicates election quality might be a multi-dimensional concept, in which case indicators are needed for each dimension in order to measure the concept accurately. Measurements of election quality based on

overall judgments of quality assume uni-dimensionality by default, but if multiple indicators are measured, this is something that can be assessed empirically.

Concluding, if election quality is measured by a single overall indicator one cannot evaluate empirically to what extent such overall evaluations are based on specific aspects of the electoral process, neither which aspects are taken into account nor how they are weighted in the overall scores. It seems likely that in practice, overall judgments are based on a variety of indicators that are considered by the coders but not coded explicitly, i.e. the aggregation of data is done in the coder's mind. If different coders have different indicators in mind and/or assign different weights to indicators, this would lead to coder-specific variation in overall coding scores of election quality. For example, if one coder considers election violence to be more serious than another coder, or ballot box stuffing or any other particular irregularity, this will lead to systematic biases in overall scores of election quality. If data-gathering is done by various coders, such as the data gathered by Elklit and Reynolds (2005a), Hartlyn et al. (2008), Birch (2008, 2011), Kelley and Kiril (2010) and Hyde and Marinov (2010), inter-coder reliability scores give an indication of differences between coders and provide a margin of error that can be taken into account in subsequent analyses. However, if data is coded by a single researcher, as the data gathered by Anglin (1998), Pastor (1999), Van de Walle (2003), Simpser (2005), Lindberg (2006a) and Donno (2010), or single country experts, as the data gathered by Munck (2009), there is no way to assess potential measurement error⁴⁶.

Using multiple indicators to measure election quality is hence preferable to decrease measurement error and more accurately measure irregularities in various parts of the electoral process. However, this

⁴⁶ It must be said however that measurement error can also be reduced by using multiple data-sources, which Simpser (2005) and Lindberg (2006a) do, and confidence in the coding process is also increased by careful documentation of the coding procedure, which Simpser (2005), Lindberg (2006a), Donno (2010), and Munck (2009) provide.

requires a higher investment of resources in data gathering (and data-sources that provide precise enough information to enable researchers to gather data on these specific components of the electoral process)⁴⁷. An excellent example of a multi-indicator approach is the work done by Elklit and Reynolds (2005a). Unfortunately, the resources required to gather data in such a precise manner are such that thus-far only 19 elections have been coded, rendering the data more useful for in-depth country studies than for comparative work⁴⁸. Other examples of multi-indicator approaches that use fewer indicators and hence are more practically feasible are the data gathered by Birch (2008, 2011)⁴⁹, Kelley and Kiril (2010)⁵⁰, Hyde and Marinov (2011) and, as an early example, Anglin (1998). All the other datasets evaluated here use single overall judgments of election quality.

From measurement indicators to data

Turning to the next step in evaluating measurement, i.e. the link between indicators and actual data, three issues need to be considered: (1) measurement level or precision, (2) sources, and (3) the coding process.

As regards *measurement precision*, table 2.5 shows that scales used vary from dichotomous distinctions between flawed and non-flawed elections to more fine-grained distinctions of 3 to 5 ordinal categories. When multiple indicators are measured, the scales used are ordinal scales with 4 to 5 categories that range from low to high presence of irregularities or qualities, depending on the authors'

⁴⁷ Arguably, if data-sources do not provide precise information on electoral conduct, using a crude coding scheme probably decreases measurement error. While for many elections since the 1990s more precise information is available in sources such as election observation reports, human rights reports, and news media, this is not always the case for elections that took place longer ago. Hence datasets that go back to the start of the third wave or earlier might prefer to code elections with overall indicators for this reason.

⁴⁸ <http://www.democracy-assessment.dk/start/page.asp?page=1>

⁴⁹ <http://www.essex.ac.uk/government/electoralpractice/>

⁵⁰ <http://www.duke.edu/web/diem/>

conceptualization. For example, coders in the Birch (2008, 2011) data coded indicators on a 5-point scale ranging from “lowest degree of malpractice” to “highest degree of malpractice”, Elklit and Reynolds (2005a) coded each indicator on a 4-point scale from “very poor” to “very good” and codes in the Kelley and Kiril (2010) data vary on a 4-point scale from “no problems” to “major problems”.

When overall judgments of election quality are used, authors often make a distinction between elections with severe irregularities, elections with less severe irregularities and elections without irregularities (Lindberg 2006a, Hartlyn et al 2008, Munck 2009, Donno 2010)⁵¹. The severeness of irregularities depends in most cases on whether the irregularities were widespread, intentional and affected the election outcome. Elections are considered to have “severe irregularities” (or be “unacceptable” or “rejected”) when irregularities are widespread, when they are the result of intentional actions rather than mere organizational incapacity, and when the irregularities are likely to have affected the election outcome.

What is problematic about these codings is that they consider intentionality and the effects of irregularities on the election outcome in judging the ‘severeness’ of irregularities. As to intentionality, not only is distinguishing between intentional actions and organizational incapacity quite difficult in many cases, non-intentional irregularities such as inaccurate voter registration can have significant consequences for election quality. As Pastor (1999) notes: “the boundary line separating political manipulation and technical incapacity is rarely surveyed, and elections can fail for one or both reasons” (p. 2). In addition, speculating whether the irregularities affected election outcomes, or “thwarted the

⁵¹ Three of these 4 authors use very similar 3-point ordinal scales, coding elections as (1) unacceptable/reject/major irregularities; (2) flawed/irregular/significant irregularities; or (3) acceptable/clean/lack of significant irregularities (Hartlyn et al., Donno and Munck respectively). Lindberg uses a very similar scale, but has an additional category identifying cases when “no elections have been held or those held have been wholly unfair” (see Appendix C for the precise coding descriptions).

popular will of the citizenry” (Geisler 1993: 617), seems even more tricky, for the simple reason that the only way to gauge voters’ preferences is through the electoral process and if the latter was flawed, observers have little possibility to know what voters’ preferences were. As Mair (1994) points out: “However understandable it is to avoid the apodictic free and fair, it is even more problematic to speculate on voter intentions and election results. How else can the voter’s intention be registered than by an election, and how can this take place correctly other than in a free and fair election?” (as quoted in Bjornlund, p. 119). Hence, I would propose to measure election quality purely on the basis of how often irregularities occurred, i.e. how widespread they were. However, even if the frequency of irregularities is coded, and regardless of the scale used, in all cases assigning data scores involves a degree of subjectivity. In this case, measurement error can be reduced by specifying clearly to coders what each score on the scale means. The coding schemes of Munck (2009), Birch (2008, 2010), Lindberg (2006a) and Donno (2010) are very clear in this respect⁵².

Another important factor determining the quality of data are the *sources* of information used. Sources vary from election observation reports to news media and historical sources, official records such as complaints filed by political parties, surveys and ethnographic research, and evaluations by country experts (Lehoucq 2003, Birch 2011). Distinctions can be made between “partisan” and “non-partisan” sources (Lehoucq 2003), or between “subjective” and “objective” sources (Hartlyn and McCoy 2006). Partisan sources are for example complaints filed by political parties or newspaper accounts. Assessments of election irregularities by citizens, country experts or election observation missions are considered non-partisan by Lehoucq (2003). However, though probably less partisan than accounts of political parties, all these sources are likely to have some degree of bias. In the case of citizens, they are

⁵² The only slight confusion in Munck (2009) is the addition –a posteriori– of pluses and minuses to the 4-point ordinal scale to distinguish cases that fall in between one of the 4 categories. Fortunately, very little use is made of the pluses and minuses.

participants in the electoral process and have a stake in elections depending on their political orientation. Hence surveys and ethnographic research should include citizens from different political orientations in order to obtain balanced information. In the case of country experts too, researchers are probably well advised to consult multiple experts and be aware of their political orientations. Finally, election observation missions have been found to suffer from biases as well, that we will discuss more in-depth shortly. The distinction between “subjective” and “objective” sources is rather based on the distinction between actors that participate in the electoral process (i.e. citizens and parties) and observers of the electoral process (i.e. news media, country experts, election observers). The argument for not using subjective sources is that they relate to the perceptions of citizens and parties, which might be dependent variables of interest, rather than part of the indicators of election quality. Moreover, since electoral manipulation is often legally and socially sanctioned, it is questionable to what degree citizens (and parties) will report irregularities such as vote-buying in surveys. However, “objective” sources will also have a hard time gathering reliable information about such practices, and as mentioned above, are probably not entirely free from bias either.

Apart from partisanship or subjectivity, other sources of bias may be incomplete geographical or temporal coverage. For example, citizens are only able to report on irregularities that occurred in their surroundings, requiring good geographical coverage of surveys in order to obtain a complete picture of citizens’ perceptions of fraud. Likewise, news media might only have limited access in certain parts of the country and election observation missions cannot be physically present in each and every polling station (and in fact often only cover a fraction of the total number of polling stations in a country)⁵³.

⁵³ In some exceptional elections international observer presence is extensive, such as in the transitional elections in Nicaragua in 1990, where observers were able to visit over 70% of polling stations (Pastor 1990) or the elections in Macedonia in 2002, where the OSCE sent 800 observers to monitor polling in 3000 polling stations (Bjornlund 2004). Normally however, observer presence is much more limited, as in the Indonesian elections in 1999, where 600 international observers were present to monitor about 300.000 polling stations (Bjornlund

Though less obvious perhaps, temporal coverage of sources is also a problem. For example, coding elections that occurred in the 1970s and 1980s makes reliance on election observation mission reports less viable since missions were not that common at the time, and collecting good information on electoral conduct based on news resources in that period may also be more challenging than collecting data for recent years⁵⁴.

Since international election observation missions have become so common in the course of the 1990s (with 80% of worldwide elections nowadays receiving international observer missions (Hyde 2011a)), many cross-national datasets on election quality rely on data from these missions (5 out of the 12 datasets examined here). Moreover, often news media reports and academic sources will build, even though only in part, on the assessments of international election observers in their own analyses of elections. Hence, examining potential biases of this source a bit more in-depth seems worthwhile. Considering partisan bias, it has been demonstrated that missions from intergovernmental organizations (IGO's) with less democratic member states tend to be more lenient in their assessments than IGO's with more democratic member states and NGO's, which is why most researchers give

2004: 143). Schmeets (2010) describes how in OSCE missions, “on average 331 observers were on the ground, ranging from less than 100 in early elections and for example municipal elections, to over 700 in for example the 1996 Bosnian elections under the Dayton agreement” (Schmeets 2010: 13). Comparative data on coverage of election missions does not yet exist as far as I am aware (coverage meaning number of observers relative to number of polling stations), however clearly coverage can vary substantially from one mission to another. Domestic observers tend to achieve much greater geographical coverage than international observers, however, these data are not often used for academic purposes (Bjornlund 2004).

⁵⁴ Moreover, as noted before, criticism of early observation missions (Geisler 1993, Carothers 1997, Anglin 1998) noted the unprofessionalism of election observers and their focus on election-day as significant obstacles to high quality observation. In the course of the 1990s however, this changed, and election observation missions by organizations such as OSCE, EU, OAS, and Carter Centre, now use standardized well-developed methodologies to assess election irregularities before, during and after elections (Hyde 2011). However, this begs the question to what extent election observation reports of elections in the 1980s and early 1990s provide data that are less accurate, and whether an observed increase in election irregularities in recent years is due to an actual increase or rather to improved observer methodology.

more credibility to sources that apply stricter norms in their election assessments such as the OSCE, the EU, OAS and the Carter Centre, NDI, or IFES (Kelley 2009a, Donno 2010)⁵⁵. An indication of the degree to which different international observer missions come to different conclusions is Kelley's finding that "for 577 election monitoring missions that took place between 1980 and 2004, in 9.7% of these elections monitoring groups disagreed" (Kelley 2009b: 10). Apart from differences between international observers, certain types of elections are sometimes also judged differently. As such, international election observers seem to be less critical in "founding elections, elections that lead to alternation in power or that establish peace after civil wars" (Donno 2010: 26, Hartlyn and McCoy 2006). Kelley also finds that "elections that demonstrate an improvement compared to previous elections are also likely to be judged more positively compared to their actual level of irregularities" (Kelley 2009a: 779). This suggests, according to Birch that: "the overall assessment of the election may in some cases be less informative than the sub-components whose evaluation is typically embedded in the reports' details" (Birch 2011: 71), and that are less prone to such 'political' or partisan bias.

A related issue is whether observers actually observe irregularities at all, given their limited coverage and little time spent at polling stations. As Bjornlund (2004) notes: "Rarely will international observers actually witness electoral manipulation or lawlessness, either on election day or during other parts of the process. Neither do observers typically find documentary or physical evidence of wrongdoing." (Bjornlund 2004: 150). Instead, he proposes: "the job of international observers is to get good information from others, including election officials, party representatives, and domestic monitors, as

⁵⁵ Kelley (2010) finds that "in general IGOs are nearly twice as likely to endorse elections as are NGOs, but that this tendency declines as an IGO's membership grows more democratic" (Kelley 2010: 165). More specifically, Kelley (2009a) shows that "the least critical organizations are IGOs with less-democratic member states such as the Commonwealth Secretariat, the Commonwealth of Independent States, and the South African Development Community, [as well as] the Economic Community of West African States and the African Union (Kelley 2009a: 778).

well as to get a feel for the process themselves” (Bjornlund 2004: 144). However, given how difficult it is to observe irregularities it is quite amazing how many irregularities have still been observed in recent years, indicating that even if observers only gauge the proverbial ‘tip of the iceberg’, they do register part of the irregularities. An easy solution to measure electoral fraud in a non-partisan, accurate way is unlikely to exist, due to the inherently covert nature of election fraud and the complexity of electoral processes. However, a possibility to reduce source bias in measurement would be to use multiple data sources. This might be a good way of ‘balancing’ sources so as to limit partisan bias, and this is indeed the strategy followed by Elklit and Reynolds (2005a), Simpser (2005), Lindberg (2006a), and Hyde and Marinov (2010)⁵⁶.

Finally, evaluating the data generating process involves, apart from measurement level and sources, an assessment of the *coding process*. Here, the use of multiple coders, clear coding scales and documentation of coding procedures as well as access to replication data and files is important. If data-gathering is done by various coders, inter-coder reliability scores give an indication of differences between coders and provide a margin of error that can be taken into account in subsequent analyses. This was done by 4 out of 12 datasets (Elklit and Reynolds (2005a), Hartlyn et al. (2008), Birch (2008, 2011), and Kelley and Kiril (2010)). In terms of coding scales, scores along the scale are generally made explicit and explained well. Moreover, most studies provide clear documentation of the coding process and access to replication data and files.

Summarizing, most cross-national empirical datasets measure election quality based on a single overarching indicator using a single source of information. Only 5 out of 12 studies measure election

⁵⁶ Donno (2010) and Hartlyn et al (2008) use multiple sources, but only if election observation reports are not available. Hence, the majority of elections are coded using one type of source, either election observation reports or news reports.

quality with multiple and precise indicators (Anglin 1998, Birch 2009, Elklit and Reynolds 2005a, Kelley and Kiril 2010, Hyde and Marinov 2010); 4 out of 12 studies use multiple sources to measure election quality (Elklit and Reynolds 2005a, Simpser 2005, Lindberg 2006a, Hyde and Marinov 2010); and 5 out of 12 studies use multiple coders to code elections and report inter-coder reliability scores (Elklit and Reynolds 2005, Birch 2008, 2011, Hartlyn et al. 2008, Kelley and Kiril 2010, Hyde and Marinov 2010).

Validity tests

Before turning to the data on election quality used in this thesis, in this section a number of validity tests will be carried out to assess the degree to which the issues discussed above are in fact problematic. I first considered the use of single overall indicators of election quality versus multiple specific indicators, and tested to what degree single overall indicators might be biased. Subsequently, I attempt to analyze what the consequences are of using different data sources and coding scales by comparing the scores assigned to elections by different authors⁵⁷.

In order to test empirically whether overall judgments of election quality cover the entire electoral process, I tested to what extent specific irregularities predict overall judgments of election quality and whether certain irregularities are likely to weigh more heavily in the overall judgment of election

⁵⁷ For the validity tests and our database, 3 datasets will not be included: Pastor (1999), Elklit and Reynolds (2005a) and Simpser (2005). The reason for excluding Pastor is that his conceptualization of election quality is rather narrow, since it is purely based on all parties accepting the elections and does not measure irregularities that might have occurred. Also, Pastor's data as reported in his 1999 article only include the flawed elections, omitting the remainder of the sample, rendering testing of the quality of his data impossible. As regards the data gathered by Elklit and Reynolds (2005a), these refer to 19 elections, of which only 9 match our sample of third wave regimes, and so their data were excluded for this reason. Finally, Simpser's (2005) data on electoral fraud are excluded because he also uses a more limited conceptualization of electoral fraud, not including aspects such as unequal access to media or unfair laws, and because his data are partly based on the Database of Political Institutions fraud variable, which, as noted before, is limited in its conceptualization of fraud and has many false negatives.

quality. I also tested whether specific irregularities map onto a single latent trait, so as to evaluate whether election quality can be considered a uni-dimensional concept. The results of these analyses are reported fully in Appendix C. Overall judgments of election quality appeared to be quite comprehensive, as most specific irregularities were significantly related to overall scores; and specific irregularities seem to be measuring a single underlying latent concept. However, some irregularities are weighted more heavily than others, implying that the legal framework, pre-election cheating (including such aspects as misuse of state resources, campaign and media restrictions and candidate intimidation) and election-day cheating (such as ballot box stuffing and vote counting) might more strongly influence overall election quality scores than administrative irregularities in for example voter registration and polling procedures. Also, forms of electoral manipulation like vote buying and electoral violence stand out as quite different from other irregularities, and seem to be less relevant to observers' judgments of overall election quality. Moreover, it seems that when the overall judgment of election quality is more open to political influence, like the overall statements on the election by the US State Department, or allegations of fraud by Western international observers, they provide a less comprehensive reflection of the actual irregularities that happened.

Comparing datasets: different sources, scales, and disagreement?

In this section, we first compare the scores assigned to the same elections by different data-sources, which allows identification of systematic bias in measures (are mean scores systematically lower or higher). If there are systematic differences between datasets, we can possibly relate such differences to the different sources or coding procedures used. Also, with this procedure we could identify elections on which high disagreement exists between different measures, closer scrutiny of which might indicate what type of elections are difficult to code and consequently, have less reliable election quality scores. Finally, assessing the disagreement between sources coding the same elections enables us to assign a 'margin of error' to each election quality score, which can be used in subsequent analyses.

Differences between datasets can be assessed considering the correlation between the different data-sources or by comparing the standardized mean values assigned to the same elections by different authors. Our data present a problem here however. First of all, not only do the different data-sources not overlap for all elections, the data-sources that do overlap do so for different sub-sets of elections. To solve this we first examine the pair-wise correlations between each data-source, and then compare the mean scores assigned by authors for each sub-set of elections. Since authors use different scales to score election quality, we recoded the election quality scores so that all scores ran from low election quality to high election quality, and vary between 0 and 1⁵⁸.

Starting with the correlations, table 2.4 shows the Pearson correlations between different election quality measurements. Overall, correlations seem to be quite high, however, there are some notable exceptions. The Lindberg data is quite strongly correlated to the Kelley, Anglin and Van de Walle data, and less strongly to Birch and Hyde's data scores. This might be related to the fact that both latter data are based on election observer evaluations, while the former are rather based on (presumably more 'neutral') academic and news sources as well as human rights reports.

⁵⁸ Ideally, we would standardize scores based on the entire sample coded by each author, generating comparable scales while also correcting for the variance of scores assigned by different authors (i.e. generating variables that have a mean of 0 and a variance of 1). This would not be problematic if authors would code the same sample of elections. However, since authors code different regions, countries and different elections within those countries, standardization would lead to variables that are in fact less comparable. For example, if Munck happens to have chosen a sample of elections that has high election quality, his variable will be very skewed, and hence the original score of "1" will become something like 0.3 in the standardized variable. If Donno has a sample of elections with more low quality elections then her original "1" will stay a "1" in the standardized version. The only way to assure that the distribution of codings is not due to sample selection and only due to ways of coding by authors is that authors need to have coded the same sample. An alternative would be to impute missing data for all different datasets and compare the resulting variables. However, since many authors focus on a single region, many countries and regions are missing entirely from their datasets and one would impute election quality scores for Sub-Saharan African countries using a model based on South American countries, which I doubt will produce useful data (see also Honaker and King 2010). Hence, I've preferred to do a simple recoding of variables running from low election quality to high election quality and vary between 0 and 1. An overview of the exact codings of each database as well as the recoding carried out by me is provided in Appendix C.

Table 2.4. Correlations between different election quality measures

	Lindberg	Birch	Donno	Hartlyn	Kelley	Munck	Hyde	Anglin	Vd Walle
Lindberg	1								
N	257								
Birch	0.37+	1							
N	23	182							
Donno	.	0.82***	1						
N	0	114	196						
Hartlyn	.	0.76***	0.35+	1					
N	0	16	30	73					
Kelley	0.63***	0.72***	0.68***	0.40***	1				
N	208	130	183	72	644				
Munck	.	0.24	0.36**	0.59***	0.44***	1			
N	0	38	70	63	153	163			
Hyde	0.44***	0.63***	0.66***	0.42**	0.53***	0.09	1		
N	163	159	174	42	388	101	459		
Anglin	0.70***	.	.	.	0.69***	.	0.74**	1	
N	21	0	0	0	20	0	14	21	
Vd Walle	0.69***	.	.	.	0.63***	.	0.49***	0.65+	1
N	81	0	0	0	77	0	54	8	82

All original variables recoded to vary from low election quality to high election quality on a scale from 0-1. Hyde 2 categories, Donno, Hartlyn, Munck, Anglin, Van de Walle 3 categories, Lindberg and Kelley 4 categories, Birch 5. See Appendix C for a precise overview.

The Birch data seems however to be quite well correlated with other datasets, with high correlations with the Donno, Hartlyn, Kelley and Hyde data. The exceptions are the Munck data and to a lesser extent the Lindberg data, that have low correlations with the Birch data (in Munck's case not even significant). Moving further right, the Donno and Kelley data also seem to have quite high correlations with most other datasets, with the exception of the Hartlyn and Munck data. The relatively lower correlations of the Hartlyn data with some of the other datasets are surprising, given the thorough coding procedure followed (i.e. scores based on election observation reports, or, if not available, news sources, coded by multiple coders and validated by country experts). However, what stands out most is the overall low strength of correlation between the Munck data and the other measurements. Since Munck's coding procedures and scales are very well explained and clear, the only possible explanation might be the source used for this data: country experts, who in this case seem to have given more positive evaluations to the conduct of elections than international observers and other sources. However, whereas correlations give some indication as to what degree measurements score

elections in the same way, a better way of assessing differences between measurements is by same-sample mean comparisons. In order to do these comparisons, we selected each sub-set of elections that had overlapping scores from different authors, resulting in subsets that were scored 6 times, 5 times, etc. Table 2.5 shows the results of these comparisons.

Table 2.5 - Same sample mean comparisons of election quality measures

Comparison	Lindberg	Birch	Donno	Hartlyn	Kelley	Muncuk	Hyde	Anglin	VdWalle	N	Dif
6 (B, D, H, K, M, N)		0.80	0.68	0.73	<i>0.51</i>	0.86	0.91			11	0.40
5 (L, K, N, A, W)	<i>0.44</i>				<i>0.44</i>		0.83	0.54	0.50	6	0.39
5 (B, D, K, M, N)		0.82	0.80		<i>0.59</i>	0.97	1.00			15	0.41
5 (D, H, K, M, N)			0.68	1.00	<i>0.66</i>	0.96	0.86			14	0.34
4 (L, K, N, W)	0.52				<i>0.47</i>		0.87		0.57	46	0.40
4 (L, K, N, A)	<i>0.45</i>				0.50		0.75	0.50		8	0.30
4 (L, B, K, N)	0.55	0.36			<i>0.33</i>		1.00			9	0.67
4 (B, D, K, N)		0.57	<i>0.41</i>		0.44		0.57			72	0.16
4 (D, K, M, N)			<i>0.58</i>		0.63	0.97	0.89			19	0.39
4 (H, K, M, N)				0.82	<i>0.73</i>	1.00	0.91			11	0.27
3 (L, K, N)	0.51				<i>0.42</i>		0.85			66	0.43
3 (L, K, W)	0.60				<i>0.57</i>				0.60	21	0.03
3 (L, B, N)	0.63	<i>0.56</i>					0.75			12	0.19
3 (B, D, N)		0.44	<i>0.31</i>				0.38			8	0.13
3 (D, K, N)			<i>0.47</i>		0.49		0.75			32	0.28
3 (H, K, M)				0.92	<i>0.74</i>	0.89				18	0.18
3 (K, M, N)					<i>0.59</i>	0.86	0.95			22	0.36
2 (L, K)	<i>0.51</i>				0.52					44	0.01
2 (L, N)	<i>0.62</i>						0.93			15	0.31
2 (B, N)		0.72					0.88			17	0.16
2 (D, K)			0.55		<i>0.46</i>					10	0.09
2 (H, K)				0.83	<i>0.66</i>					6	0.17
2 (K, M)					<i>0.73</i>	0.91				29	0.18
2 (K, N)					<i>0.53</i>		0.72			39	0.19
										550	0.27
Lowest value (%)	4/9 = 44%	2/7 = 29%	4/8 = 50%	0/5 = 0%	15/20 = 75%	0/8 = 0%	0/18 = 0%	0/2 = 0%	0/3 = 0%		
Highest value (%)	1/9 = 11%	1/7 = 14%	1/8 = 13%	3/5 = 60%	1/20 = 5%	3/8 = 38%	14/18 = 78%	0/2 = 0%	1/3 = 33%		

Note that the Hyde & Marinov (2010) data is indicated as "N", after the name of their dataset, NELDA, to prevent confusion with "H" for Hartlyn et al (2008). Note also that subsets of elections with less than 5 observations are not reported. The highest mean value within a subset is indicated in bold, the lowest mean value in italic.

As the right column shows, the variation between measurements when assessing the same sample of elections ranges from 0.01 to 0.67 (!), with an average of 0.27. Of the 601 elections that were coded by two or more sources, about 30% had scores with deviations higher than the average of 0.27⁵⁹.

⁵⁹ Note that the difference of 51 elections (601 – 550 as reported in the table), lies in sub-sets of overlapping elections that had 5 or less observations and are not reported in the table.

However, some measurements have mean election quality scores that are consistently higher or lower than the other measurements. This becomes clear when looking at the two lower rows of the table, that indicate how often a given measure had the lowest and highest mean election quality in all subsets of elections for which it had data. The Kelley, Donno and Lindberg data seem to have systematically lower election quality scores than the other datasets, while the Hyde, Hartlyn and Munck data have systematically higher election quality scores. The extremes are the data by Kelley that had the lowest mean value in 75% of the subsets of elections, and the data by Hyde, that gave the highest mean value in 78% of the subsets of elections.

Closer scrutiny of the coding rules used by these authors reveals that the low average scores for Kelley might be due to qualitative differences in scale categories. Considering Kelley's 4 coding categories, the two highest election quality categories are "no problems" and "minor problems". Most other authors score elections with minor irregularities in their highest category of election quality (cf. Donno, Munck, Hartlyn et al. and Anglin, Appendix C). Hence it seems reasonable substantively to recode Kelley's scores to a 3-point scale, combining "no problems" and "minor problems" in the category of "high quality elections"⁶⁰. Doing so lowers the average variation between measurements from 0.27 to 0.22, ranging between 0.03 and 0.64, and solves the systematic downward bias of the original Kelley measure. It leaves Donno and Lindberg as the systematically lower measures however. This seems unlikely to be the result of differences in sources, as though Lindberg uses a variety of sources, Donno uses election observation reports as an important source of information and should hence, if source would have been a significant reason for bias, be more alike to the measures of Birch, Hyde and Hartlyn et al. Considering the scale categories again, the difference rather seems to be due to qualitative differences in scale categories. While the highest election quality score in the Munck

⁶⁰ Kelley's "moderate problems" is coded as the intermediate category, and "major" problems is coded as the lowest election quality category.

data still allows for “technical” administrative irregularities, as long as these don’t affect the election results, in the Donno data elections with administrative irregularities are more likely to be coded in the intermediate category of election quality. Hence the 3-point ordinal categories used by Donno seem to be somewhat stricter in their assessment of election quality. The same holds for Lindberg, who only assigns the highest election quality score to elections that were “free and fair” and assigns a lower score to elections with “deficiencies that did not affect the result”. Since this latter category encompasses both minor deficiencies as well as moderate deficiencies, we have no good substantive reason to recode Lindberg’s 4-point scale, and instead continue to use the original data scores.

Turning to the measurements with systematically higher scores for election quality, the Hyde data seem to have a systematic upward bias that might be problematic. In this case, this might be due both to the sources used as well as coding scales. Regarding the source, Western election observers’ judgments about the elections, our previous analyses clearly showed that this measure of fraud was not comprehensive and excluded important specific irregularities such as election violence and opposition harassment (see Appendix C). Hence, data based on election observers’ overall evaluation of elections might have an upward bias (for reasons discussed before). However, if this was the whole story we would expect to find similar bias in Birch’s measure, and to a lesser extent Hartlyn’s and Donno’s measures. Here, the coding scales are important. The Hyde data are coded as a dichotomous variable, indicating whether Western observers alleged the elections to be fraudulent or not. It seems likely that if observers were critical about certain elements of the elections, but did not condemn the elections entirely, these elections would be coded as “not” having received allegations of fraud and hence end up in the “clean” category. Most likely in the other measurements such elections would have been coded in an intermediate category. Hence, the scales used result in a rather strong upward bias that might affect an overall measure of election quality, and it might be preferable to leave this measure out altogether. Excluding the Hyde data indeed leads to a further lowering of the average variation

between measurements from 0.22 to 0.15, ranging from 0.03 to 0.39. The measurement with a marked upward tendency remains Munck.

Turning to the final validity test, we analyzed the factor structure of the election quality scores assigned by different datasets to test statistically whether the different measurements actually gauge the same underlying concept. Table 2.6 demonstrates the results of factor analyses and alpha reliability tests for each sub-set of overlapping election scores.

Table 2.6 - Same sample mean comparisons: unidimensionality

Comparison	Alpha	Factor analyses			N
		Number of factors	% variance explained	Factor loadings min/max	
5 (B, D, K, M, N) ^a	0.79	1	61% ^{**}	0.50-0.92	15
5 (D, H, K, M, N) ^b	0.71	2	55% [*]	0.55-0.90	14
4 (L, K, N, W)	0.76	1	63% ^{***}	0.65-0.87	46
4 (L, K, N, A)	0.95	1	88% ^{***}	0.88-0.98	8
4 (L, B, K, N)	0.76	1	82% [*]	0.90-0.90	9
4 (B, D, K, N)	0.90	1	80% ^{***}	0.88-0.92	72
4 (D, K, M, N) ^c	0.61	2	48% ⁺	0.50-0.84	19
4 (H, K, M, N)	0.92	1	81% ^{**}	0.85-0.92	11
3 (L, K, N)	0.74	1	67% ^{***}	0.80-0.85	66
3 (L, K, W)	0.84	1	81% ^{***}	0.87-0.94	21
3 (L, B, N)	0.71	1	75% ^{***}	0.73-0.94	12
3 (B, D, N)	0.83	1	86% ^{**}	0.89-0.96	8
3 (D, K, N)	0.79	1	70% ^{***}	0.80-0.87	32
3 (H, K, M)	0.91	1	80% ^{***}	0.79-0.96	18
3 (K, M, N) ^d	0.33	2	47% n.s.	0.34-0.88	22
2 (L, K)	0.75	1	86% ^{***}	0.92-0.92	44
2 (L, N)	0.60	1	73% ⁺	0.85-0.85	15
2 (B, N)	0.86	1	91% ^{***}	0.96-0.96	17
2 (D, K)	0.84	1	87% [*]	0.93-0.93	10
2 (K, M)	0.89	1	84% ^{***}	0.92-0.92	29
2 (K, N)	0.91	1	88% ^{***}	0.94-0.94	39

Cronbach's alpha and factor analyses, principal components analysis. Reported variance explained refers to the first factor. Note that for sample size, a ratio of 2:1 cases/variables was taken as the rule-of-thumb lowest sample size. This means that for 5 variables, N >= 10, for 4 variables, N >= 8, for 3 variables, N >= 6, and for 2 variables, N >= 4. Significance of models is indicated by p-values: + 0.1, * 0.05, ** 0.01, *** 0.001.

a. Hyde is constant in this subset, and was consequently dropped from the analyses.

b. Hartlyn is constant in this subset, and was dropped from analyses. All variables loaded on the first factor, except for Munck, loading on the second factor.

c. Munck loads on the second factor.

d. Munck loads on the second factor.

Considering Cronbach's alpha, presented in the second column, as an indication of scale reliability, most measurements seem to form strong scales, with scores between 0.71 and 0.95. Only three subsets of election have a lower reliability score, 0.60, 0.61 and 0.33 respectively. Turning to the factor analyses, again the large majority of analyses result in a single factor, providing evidence that the different measurements indeed measure the same underlying factor. Moreover, not only are most models highly significant, the single factor explains between 61% and 91% of variance of the individual measures. Factor loadings (indicating the relation between individual variables and the latent factor) are also quite high, ranging from 0.50 to 0.96. Three models jump out as an exception to this pattern however, and are highlighted as bold in the table. All these models resulted in 2 factors and had insignificant or lower significance levels. In all three cases, only one variable appeared to load strongly on the second factor, and in all three cases, this variable was Munck's measure of election quality. Even though in the other subsets that included Munck's measure as a variable the factor solution was one-dimensional and significant, this is an indication that perhaps Munck's election quality variable is measuring a slightly different concept than the other measurements.

Concluding, with the exception of the data by Munck, the correlations between data-sources are substantial. Moreover, examination of mean values assigned by different authors to the same samples of elections demonstrated that deviation in mean scores between authors was on average 0.27. The systematic tendency of Kelley's measurement to assign lower election quality scores was corrected by recoding her data scores to a 3-point ordinal scale, which is substantively reasonable and lowers the average deviation to 0.22. The systematic tendency of Hyde's measurement to assign higher election quality scores could not be corrected however, which could lead to an upward bias in election quality scores if this measure is included in an overall measure of election quality. Given that in the earlier analyses, this measurement was also found to have limited comprehensiveness, an overall measure of election quality would probably be more reliable excluding this measure. Indeed, when Hyde is

excluded, the average deviation in mean scores between authors further declines to 0.15. Finally, the analyses of dimensionality demonstrated that all measurements load strongly on a single factor, providing evidence that the different measurements in fact measure a single underlying concept. However, the measurement by Munck does not seem to fit entirely, as in several analyses it loads on a separate second factor, indicating that it might measure a somewhat different concept than the other measurements. Based on these findings, I also exclude Munck's data for our overall measure of election quality. Doing so lowers average deviation further to 0.13.

Based on these tests, I constructed 2 overall measures of election quality: one measure that includes all 9 datasets, and apart from recoding from 0 to 1, maintains the original ordinal scales used by the authors; and one measure that includes 7 datasets, excluding the data from Hyde and Munck, and using the recoded Kelley variable (3-point ordinal). In both cases, the overall variable runs from 0 to 1, from low election quality to high election quality. The overall measures are constructed by calculating the mean of scores assigned to each election. Apart from that, I also calculated the standard deviation of each overall score, which provides an indication of disagreement between the different data sources. For the variable that includes all 9 datasets, 601 elections were coded by multiple authors, while for the variable including 7 datasets, 474 elections were coded by multiple authors. Of these elections, only 19% and 26% respectively received exactly the same scores by all authors. To assess the degree of disagreement for the remaining elections, we consider election quality scores to disagree only weakly if a majority of authors, i.e. 2 out of 3, 3 out of 4, 3 / 4 out of 5, and 4 / 5 out of 6, agree on the election quality score to be assigned, depending on by how many authors the election was coded. We only count a majority of identical scores however if the different authors scored the elections in adjacent categories, as for example an election scored twice as "clean" (1) and once as "flawed" (0.5). An election coded for example twice as "clean" (1) and once as "severely flawed" (0) will be considered as an election on which authors disagree strongly. Finally, if there was no majority of authors agreeing on

the score to be assigned to the election, but authors gave scores in adjacent categories, these elections will also be considered as agreed upon (for example an election with 4 scores, twice 1 and twice 0.66). While the exact standard deviation that is acceptable varies by the number of times the election was scored, the standard deviation for “agreed upon” elections is always below 0.30. As soon as disagreement is stronger however, we consider the election score to be problematic, and in need of closer scrutiny. For the variable including all 9 datasets, of the 601 elections, authors disagree strongly about the election quality of 28% of elections. For the variable including 7 datasets disagreement is lower as authors disagree strongly about the election quality of only 15% of elections. This implies that for 72% and 85% respectively of elections coded by multiple authors, election quality scores can be derived that are fairly robust, and for the remaining elections closer examination is needed to reconcile the differences in election quality scores.

Differences between elections

I carried out several statistical tests to find out what differentiates the 28%/15% of elections from the elections on which authors agreed, reasoning that high disagreement might be due to methodological reasons or due to characteristics of the elections. Methodological reasons could be simply that election scores might deviate more strongly if they are coded by more sources, or that elections with medium election quality might be elections on which there is more disagreement, since differences in coding between authors might become most apparent in this middle-category of not-clearly-rigged but not-entirely-clean elections. However, it might also be the case that certain types of elections are more difficult to code. For example first elections that took place just after the transition to (at least) de jure multi-party elections, might receive more disagreement in coding since it has been documented that sources like election observation reports tend to report more positively on elections that mark the potential beginning of a transition to democracy, while codings based on human rights reports might not have this bias. Likewise, elections that took place after civil war or coup d’etat might stir more

disagreement among researchers as to how to categorize their quality. Also, if a country has been holding elections for a longer period of time, those elections might be better documented and also researchers might know the country better, easing the assessment of election quality and reducing variation in election quality scores. Finally, we checked whether there was any difference in disagreement according to region and between executive and legislative elections. Table 2.11 shows the results of these analyses.

Table 2.7 Explaining disagreement about election quality scores

Independent variables	Disagreement (9 datasets)	Disagreement (9 datasets)	Disagreement (7 datasets)	Disagreement (7 datasets)
Number of scores (2-6)	0.668*** (0.160)	0.630*** (0.162)	-1.005* (0.398)	-0.867* (0.404)
Medium election? (0-1)	2.586*** (0.273)	2.523*** (0.298)	0.443 (0.363)	0.723+ (0.427)
First election? (0-1)	-0.219 (0.462)		-0.118 (0.480)	
Election sequence (1-18)	-0.040 (0.060)	-0.079 (0.060)	-0.104 (0.088)	-0.124 (0.106)
Election type leg/ex (0-1)	-0.009 (0.170)	0.018 (0.177)	-0.602** (0.223)	-0.698** (0.241)
Civil war (0-1)	1.038* (0.491)		0.487 (0.429)	
Coup d'etat (0-1)	2.038* (0.835)		-0.356 (0.820)	
Civil war lag1 (0-1)		1.313** (0.439)		-0.041 (0.544)
Coup d'etat lag1 (0-1)		-0.475 (0.505)		0.111 (1.042)
Region – FSR/CIS ^a (0-1)	0.211 (0.467)	0.284 (0.510)	-0.280 (0.618)	0.513 (0.785)
Region – SSA (0-1)	0.318 (0.322)	0.486 (0.436)	-0.913+ (0.511)	-0.134 (0.741)
Region - SA (0-1)	-0.392 (0.514)	-0.138 (0.616)	0.312 (0.483)	1.053 (0.681)
Region - CA (0-1)	0.010 (0.498)	0.087 (0.514)	0.924 (0.805)	1.645+ (0.969)
Constant	-3.898*** (0.714)	-3.627*** (0.748)	1.492 (1.112)	0.605 (1.200)
N	600	510	473	397
Pseudo-squared	0.28	0.27	0.09	0.09

Logit regression. Robust standard errors clustered by country. Standard errors in parentheses. P-values: + 0.1, * 0.05, ** 0.01, *** 0.001.

a. Region category Central and Eastern Europe is reference category. The last region, Southern Europe, was omitted because of collinearity.

The dependent variable in these analyses is a dummy indicating whether the different datasets disagreed strongly about the election quality score for this particular election or not. If disagreement

was high, the dependent variable takes on a value of 1, and if disagreement was low or absent, the dependent variable takes on a value of 0. The first 2 models predict disagreement when taking into account all 9 datasets, and the last 2 models predict disagreement when taking into account 7 datasets.

Confirming our assessment of the different datasets earlier in this section, and our proposal to exclude the data from Munck (2009) and Hyde and Marinov (2010), we find that disagreement is systematic when combining 9 datasets, while we find little evidence for systematic variation in disagreement when combining 7 datasets. This is demonstrated by the substantially higher explained variance of the models explaining disagreement between the 9 datasets, 27 and 28% respectively, while the model only explains 9% of disagreement between the 7 datasets. Moreover, we find when comparing the 9 datasets, that indeed more scores imply more disagreement, and “middle” category elections generate more disagreement between datasets. Also, elections taking place in a country with ongoing civil war (model 1) or civil war in the year prior to elections (model 2) generate more disagreement between datasets. The effect for coup d’etats is only significant if the coup d’etat took place in the year of the election. Regional differences are not significant.

While when comparing 9 datasets disagreement can significantly be explained by structural factors, either methodological or relating to the nature of the elections, this is much less the case when comparing the 7 selected datasets. In these models, only the number of scores and the type of election are found to be significantly related to disagreement. The effect of number of scores is the reverse here: the less times an election was scored, the higher disagreement. Also, legislative elections appear to generate more disagreement, which might be due to the fact that legislative elections are –in presidential systems at least- lower profile elections and hence less information might be available about electoral conduct, increasing disagreement. Concluding, since the 7 datasets seem to have less disagreement about election scores and this disagreement seems to be less systematic than

disagreement between 9 datasets, we propose to use the 7 datasets as input for the election quality data used in this thesis. The data used in this thesis builds on the data collected in these studies, merging these various data-sources in a single time-series cross-section dataset, re-coding election quality scores on which high disagreement exists between data-sources, and coding missing election quality scores until 2009. The next section further discusses sample selection and the construction of the dataset.

2.3. Dataset on Election Quality in Third Wave Regimes 1974-2009

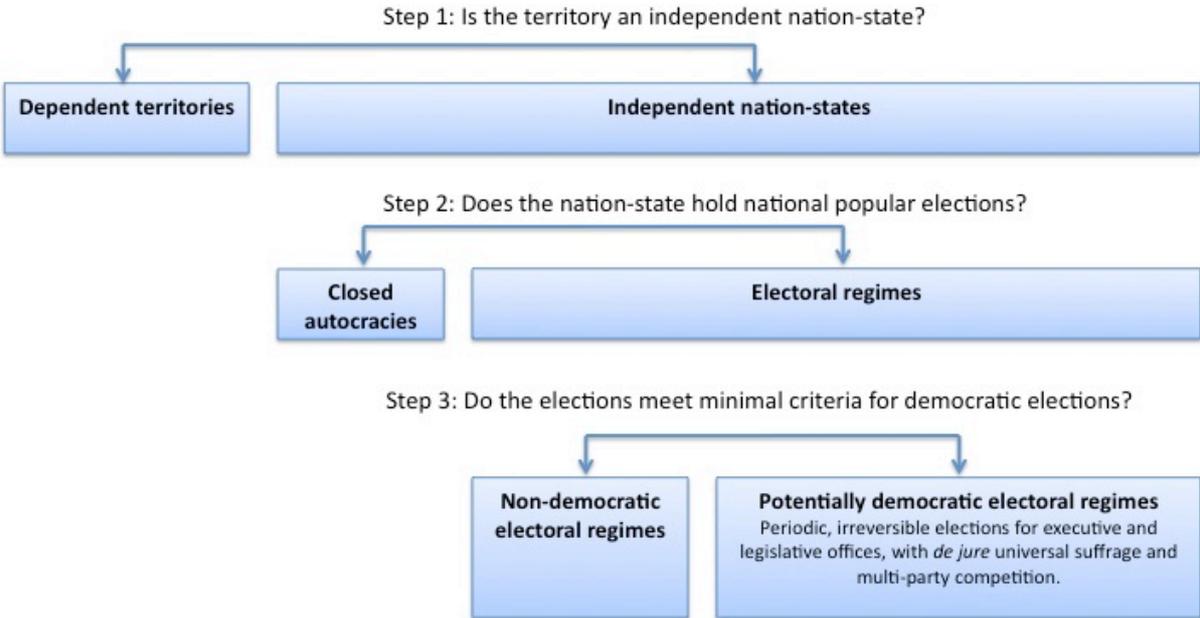
In this section we turn to discussing how election quality is measured for the purpose of this thesis. We first describe the selection procedure followed to generate our sample of ‘potentially democratic elections’, then proceed to combine the existing data on election quality discussed above, and finally present the dataset used in this thesis.

Sample selection

In line with the conceptualization of election quality presented in section 2.1, only countries and time-periods in which elections were held that meet the minimal criteria for “potentially democratic elections” have been included in our sample. Schematically, the sample selection can be presented as follows (cf. LeDuc, Niemi and Norris 2010: 11, table below is my adaptation). Step 1 and step 2 refer to the level of regimes and step 3 refers to the level of elections. The first step eliminates countries that are not recognized as independent nation-states by the United Nations. This means that elections held before independence in former colonies or protectorates, even when taking place under conditions of *de jure* multi-party competition and universal suffrage, are excluded from our sample, as are elections held in international protectorates, such as Bosnia-Herzegovina after the 1995 Dayton Accords. The second step distinguishes electoral regimes that hold popular elections at the national level from closed autocracies where no elections are held. Note that the category of electoral regimes includes both

electoral autocracies as well as electoral and liberal democracies. Step 3 distinguishes electoral regimes that hold clearly non-democratic elections from electoral regimes that hold “potentially” democratic elections. Since variation in the quality of elections occurs only in the group of “potentially” democratic elections, our sample includes only those elections.

Table 2.8 Sample selection of “potentially democratic” elections



‘Potentially’ democratic elections were selected on the basis of the following criteria (cf. paragraph 2.1): (a) elections held for both legislative and executive (directly or indirectly), (b) irreversible elections, (c) legally possible for multiple parties to participate in elections (i.e. de jure multi-party elections), (d) legally possible for all adult citizens to participate in elections (i.e. de jure universal suffrage). The first criterion implies that both the executive and the legislative should be selected through elections, either directly or indirectly. This means that elections held under military governments holding only parliamentary elections, such as Brazil between 1964 and 1982, or elections held under monarchy governments holding only parliamentary elections, such as Swaziland, are excluded. The second criterion means that elections immediately followed by a coup d’etat and/or civil

war that prevent the elected government from taking power are excluded. Elections held under conditions of civil war where nevertheless the elected government was able to assume power are included, as are elections held just after a coup d'état. The third criterion implies that elections should be held under de jure multi-party conditions, i.e. it should be legally possible for multiple parties or candidates to contest elections. This means that elections held under a constitution forbidding the formation of political parties, such as Sudan between 1989 and 1998, or allowing for only one state party, like Malawi between 1966 and 1993, are excluded. Finally, the fourth criterion implies that elections should allow de jure universal suffrage, i.e. it should be legally possible for all adult citizens to participate in elections. This means that elections where suffrage was restricted on the basis of gender, race, income or literacy were excluded. The absence of a de jure secret ballot is also considered as a suffrage restriction, as public voting limits the possibilities for opposition supporters to cast a free vote. Elections where secret ballots were not legally required were hence also excluded. Elections with suffrage restrictions for non-adults, detainees, and specific occupations (such as police or army officials) were included⁶¹.

The sample was subsequently limited to elections having taken place in the period of the third wave, i.e. 1974-2009⁶². In order to track how election quality improves with successive elections, each

⁶¹ These criteria are meant to capture the full range of variation in election quality, thus merely specifying the minimal conditions that need to be met for elections to be “potentially democratic”. This means that elections held under conditions of de facto limitations on party competition and voter participation are included, as long as the de jure conditions specified above hold. The only exception made in this regard are those cases where, despite de jure provisions for multi-party competition, only one party or one candidate contests the elections. The reason for excluding de facto one party or single candidate elections is that such elections do not provide voters with a choice, not even a pseudo-choice, and should therefore be considered non-democratic rather than “potentially democratic”.

⁶² There are several South American countries and a couple of African countries that held “potentially democratic” elections in the period before 1974, as many South American countries experienced periods of democratic government before the third wave and some African countries have been stable democracies ever

election was assigned a sequence number, starting from the first de jure multi-party, universal suffrage, and legislative/executive election, and counting until the last elections held before 2010. Finally, elections for constituent assemblies were included but treated separately in the analyses, because constituent assemblies are generally elected with the specific mandate to draft a new constitution, and hence are not expected to engage in ‘normal’ legislative policy-making. Appendix B provides an overview of the sample, listing the countries and time-periods included.

The sample selection was made using data from the Electoral Handbooks on Africa, South and Central America, Central and East Asia and Europe edited by Dieter Nohlen (Nohlen 1999, 2001, 2005, 2010), as well as the data handbook by Rose and Munro on Central and Eastern Europe (2003). Data for most recent elections, if not included in the Data Handbooks, were gathered from the Parline Database of the Inter-Parliamentary Union (IPU), The Election Guide of the International Federation for Electoral Systems (IFES), and the US State Department country profile information. For each country, a short historical file was made tracking the country’s authoritarian and democratic past, explaining coding decisions and referencing sources if needed (available upon request from the author). While the selection of elections is theory-driven, our selection of regions was more pragmatic, and based on the availability of election quality data. For this reason we exclude third wave regimes in Asia from our analysis. The regions that are included are: Central and Eastern Europe, the Former Soviet Union, Sub-Saharan Africa, South America, Central America and Southern Europe. The sample selection resulted in a sample of 97 countries and a total of 886 elections⁶³. Table 2.13 below shows the total number of legislative and executive elections per region.

since decolonization. For practical purposes of gathering data on election quality for those earlier time-periods however, I excluded these elections.

⁶³ Note that countries are included in the dataset for varying time-periods, depending on the timing of their transition to democracy, and the stability of democracy after the transition. Countries that experienced interludes of autocratic rule between 1974 and 2006 hence have “gaps” in their time-series of elections. In order to identify

Table 2.9 – Number of “potentially democratic” elections 1974-2009

Region	Number of states	Legislative elections	Executive elections	Total elections
CEE	17	84	44	128
FSR/CIS	12	54	45	99
SSA	45	184	136	320
SA	12	104	72	176
CA	8	71	50	121
SE	3	35	7	42
Total	97	532	354	886

CEE = Central and Eastern Europe, FSR/CIS = Former Soviet Republics/Commonwealth of Independent States, SSA = Sub-Saharan Africa, SA = South America, CA = Central America, SE = Southern Europe (Portugal, Spain, and Greece).

Combining existing datasets

To create the dataset on election quality in third wave regimes, I build on the 7 datasets discussed in the previous section and subsequently coded the election quality data for missing elections myself as well as recoded the elections on which there was strong disagreement. The 7 datasets created by Anglin (1998), Van de Walle (2003), Lindberg (2006a), Birch (2008/2011), Hartlyn et al.(2008), Donno (2009, 2010), and Kelley and Kiril (2010) together provided data for 735 legislative and executive elections in Central and Eastern Europe (CEE), Former Soviet Republics (PSR), Sub-Saharan Africa (SSA), South and Central America (SA and CA), and Southern Europe (SE)⁶⁴. This data was matched to our original database including all 886 “potentially democratic elections” in the 6 regions between 1974 and 2009. This resulted in election quality data for 83% of the total number of

such gaps, each election within a country received a sequence number, with the sequence re-starting from 0 after each interruption of democracy. Following Lindberg (2006a), coup d’etats and civil wars are considered as interruptions if they affect the condition of “irreversibility of elections” (i.e. elections that are followed by a coup d’etat and/or civil war that prevents the elected government from taking power). After such events the sequence of elections is re-set to start again. An overview of our sample selection is provided in Appendix B.

⁶⁴ On average, about 10% of election quality data was dropped. Reasons to drop elections were (1) differences in sample (i.e. Bosnia-Herzegovina was excluded due to its status as international protectorate and Caribbean countries were excluded because the election quality datasets only covered a very limited number of elections in countries in this region), (2) differences in types of elections included (i.e. when elections were not legislative or executive elections, but for example referenda, they were not included; and concurrent elections were matched separately as legislative and executive elections); (3) coding of 1st and 2nd rounds in presidential elections as separate elections (i.e. I coded those as single elections). Every election that was not matched has been recorded in a file available upon request from the author. Tables 1 and 2 in Appendix C show the number of elections originally coded by authors and the number of elections matched.

elections. Table 2.14 below shows the number and proportion of elections for which election quality data from the 7 datasets was available.

Table 2.10 – Number of elections with election quality scores 1974-2009

Region	Legislative elections	Executive elections	Total
CEE	69 (82%)	33 (75%)	102 (80%)
FSR/CIS	47 (87%)	35 (78%)	82 (83%)
Africa	155 (84%)	117 (86%)	272 (85%)
SA	84 (81%)	63 (88%)	147 (84%)
CA	58 (82%)	42 (84%)	100 (83%)
SE	27 (77%)	5 (71%)	32 (76%)
Total	440 (83%)	295 (83%)	735 (83%)

Clearly, the election quality data available from the 7 datasets is spread rather evenly over regions and type of elections. The 17% of elections for which no data were available in any of the 7 datasets do seem to be concentrated in recent years however, as three quarters of these missing elections took place between 2006 and 2009. Considering the amount of times elections were scored by the different datasets, 17% of elections were not coded by any of the datasets, 30% was coded once, and 53% of elections were coded by 2 or more data-sources. Elections were never coded more than 4 times, due to the fact that some of the datasets have a specific regional focus (i.e. Anglin 1998, Van de Walle 2003 and Lindberg 2006a only code elections in Sub-Saharan Africa, while Hartlyn et al. 2008 only code elections in Latin America). For the 17% of elections for which election quality data was missing, I coded the elections based on election observation reports, academic articles, human rights reports, and news media reports. For these scores I used a hierarchy of sources, with the academic and election observation reports with well-established methodologies and professional observer teams given priority over reports in news media⁶⁵. I coded elections along a 3-point ordinal scale, based exclusively on the

⁶⁵ The hierarchy of sources used is:

1. Election observation reports by IGO's with many democratic member states and strict standards of election quality: OSCE, European Commission, OAS.
2. Election observation reports by NGO's with strict standards of election quality: Carter Centre, NDI, IRI, IFES.

frequency and spread of irregularities, rather than intentionality or perceived consequences for the outcome⁶⁶. Each election score was documented in a file comprising the information collected from sources as well as references to those sources (available upon request from the author). During the coding process I also noted in an additional variable for which elections limited or contradictory information was available. Of the 151 elections for which no election quality data was available, I was able to code 150 elections⁶⁷. Of these elections, for 27% limited information was available, and only for 1% I found conflicting information. In the subsequent chapters, we will use this information to check whether our results hold when excluding these elections.

Turning to the elections on which there was strong disagreement, i.e. 8% of our sample, I recoded these elections. To do so, I followed the same procedure as for the missing elections, i.e. gathering different sources of information about the elections if available, applying the hierarchy of sources, and subsequently scoring each election along a 3-point ordinal scale ranging from widespread irregularities to none or isolated irregularities. I would then consider the different scores assigned and reach a final reconciled score. The scores for these elections are also recorded in the documentation file. Finally, in order to check my own election quality coding, I coded an additional sample of 150 elections.

-
3. Academic articles about the elections in Electoral studies (articles / research notes) and if no information, academic articles in other journals.
 4. Journal of Democracy Election Watch, Inter-Parliamentary Union and Data Handbook reports/notes on the conduct of the elections.
 5. Information about elections in US State Department Human Rights Reports or other Human Rights reports.
 6. Information about elections in news sources via Lexis Nexis.

⁶⁶ Elections were coded using the following categories: “1 = widespread irregularities, whether intentional or not, that seriously affected the conduct of the elections in a large part of the country; 2 = considerably frequent irregularities, but not so widespread that the election is outright rejected by observers; 3 = no irregularities or some isolated, non-frequent irregularities.”

⁶⁷ The one election for which information was so limited that I felt I could not reliably assign a score was the constituent assembly election in Ecuador in 1997. Hence this score was left missing.

Correlations of the election quality scores with the other 7 election datasets were all positive, significant, and quite strong, ranging between 0.65 and 0.89. Evaluating disagreement by comparing the 7 datasets to my election quality scores only resulted in 9 elections with a standard deviation higher than 0.30⁶⁸.

Having merged the data from the 7 datasets, recoded the elections on which there was strong disagreement and gathered election quality data for those elections that were missing, I combined the different election quality scores into a single overall election quality score. Combining the scores was done by first recoding election quality scores of each author so as to generate variables that ran from low to high election quality. Following the analyses reported in section 2.2, I left all data in their original scales (either 3-point, 4-point or 5-point ordinal), with the exception of the data by Kelley that seemed to perform better if the categories of “no problems” and “minor problems” were recoded to a single “high quality” score. Table C3 in Appendix C shows the original coding by each author as well as the way the variables were recoded. Using the recoded variables, I constructed 2 overall election quality variables: a continuous variable that is based on the average election quality score for the elections coded by the 7 datasets, and my own score if data were missing or in disagreement, and a categorical 3-point variable that recodes the continuous variable into the 3 categories: (1) low quality, (2) medium quality, and (3) high quality⁶⁹.

This leaves a dataset with 885 election quality scores (of our total sample of 886), of which 41% is coded once, and 59% are coded by 2 or more datasets, with a maximum of 5 scores. Considering the

⁶⁸ I coded these elections originally because they had missing data, however with the updates from Birch and Lindberg multiple scores came available in a later stage of the research.

⁶⁹ I also constructed a variable that takes the average of 8 datasets, i.e. the 7 datasets plus my own codings for overlapping elections, and my own codings for missing and disagreed elections, to use as an alternative dependent variable for robustness checks. This variable has a correlation of 0.992*** with the original dependent variable.

fact that 17% of elections was originally missing, it is logical that these elections are now coded once. However, even without these elections, still 24% of elections only received a single score, which might render those data less reliable. Analyzing these data more closely shows that the majority of these elections (88%) are coded by Kelley and Kiril (2010). This is probably due to the fact that this dataset is the only dataset with global coverage, and hence countries and regions that tend to be overlooked by other coders were coded by these authors. As such, in Southern Europe 76% of all elections are coded once, and most often by Kelley and Kiril (2010). Moreover, also in Central and South America we find somewhat higher percentages of elections that were coded only once (35% and 39% respectively), mostly in the years before 1990 (60% and 50% of the elections in those years received single scores). Ideally I would have coded all these elections myself to enable all elections to be coded at least twice, and be able to calculate a margin of error for the election quality scores. However, this would have required coding of an additional 214 elections, which was logistically not feasible. On a positive note however, our validity analyses in the previous sections have shown the data collected by Kelley and Kiril (2010) to be related to a broad set of specific irregularities and to be more reliable than overall election judgments by election observers (see analyses Appendix C), and hence, of relatively higher quality. In order to control for possible biases occurring as the result of using these data, we will use the number of times an election was coded as a control variable in our subsequent analyses.

Concluding, in this section we have carried out validity tests on 9 datasets on election quality, which resulted in excluding two datasets from our research (Munck 2009, Hyde and Marinov 2010) as well as recoding the core variable from one dataset (Kelley 2010). The remaining 7 datasets appear to be highly comparable and measuring a single underlying latent concept. To improve measurement validity, we hence decided to combine these 7 measures in an overall score measuring average election quality. Using the standard deviation of these scores, elections could be identified on which high disagreement exists. Though this disagreement appeared to be hardly systematic for the 7 datasets

concerned, these elections were still checked and recoded independently by the author. Finally, missing elections were also coded by the author based on multiple sources of information. The resulting election quality variable is a quasi-continuous variable ranging from 0 (low election quality) to 1 (high election quality) that provides scores for 885 elections in third wave regimes from 1974 to 2009⁷⁰.

⁷⁰ As a final test of the resulting measurement of election quality, I test whether it co-varies with broader democracy indices. Since election quality is generally one of the components of such indices, I would expect our overall election quality measurement to correlate highly with these data. Even though not providing a strict test of (criterion) validity, if we would find the combined measure of election quality varies as expected across levels of democracy, this would increase confidence in our measure. For this purpose I used a “categorical” and “continuous” measure of democracy, i.e. the Freedom House classification of not free, partly free and free regimes and the Polity IV autocracy/democracy scores. Correlations between these measures are indeed quite high and significant, for Freedom House varying between 0.67*** and 0.71*** and for Polity between 0.68*** and 0.70***. Tables 2.15 and 2.16 in Appendix C show this relationship between election quality and democracy indices visually.

3. The quality of elections in third wave regimes 1974-2009

As the third “wave” of democratization swept the globe after 1974, the number of countries holding regular elections for executive and legislative offices increased sharply: over 85% of the world’s states now select their national leaders through elections. Unfortunately, the variety of elections has multiplied concomitantly, ranging from “free and fair” elections with genuine contestation between parties or candidates to “façade” elections that are marred by manipulation and fraud. The dataset presented in chapter 2 allows us to study this variation in election quality more precisely: which countries manage to “get their elections right” and which countries do not? How does election quality develop over time? Are there common patterns across countries and regions? In this chapter we will address the descriptive side of this question, considering the cross-country and temporal variation in election quality in 97 countries in Southern Europe, Central and Eastern Europe, the Former Soviet Union, Sub-Saharan Africa, South America and Central America from 1974 to 2009. The set-up of the chapter is as follows: section 1 presents the regional and temporal variation of election quality for the sample as a whole, indicating global patterns in election quality; and section 2 analyses variation of election quality within countries over time, identifying common patterns. The last section summarizes the results.

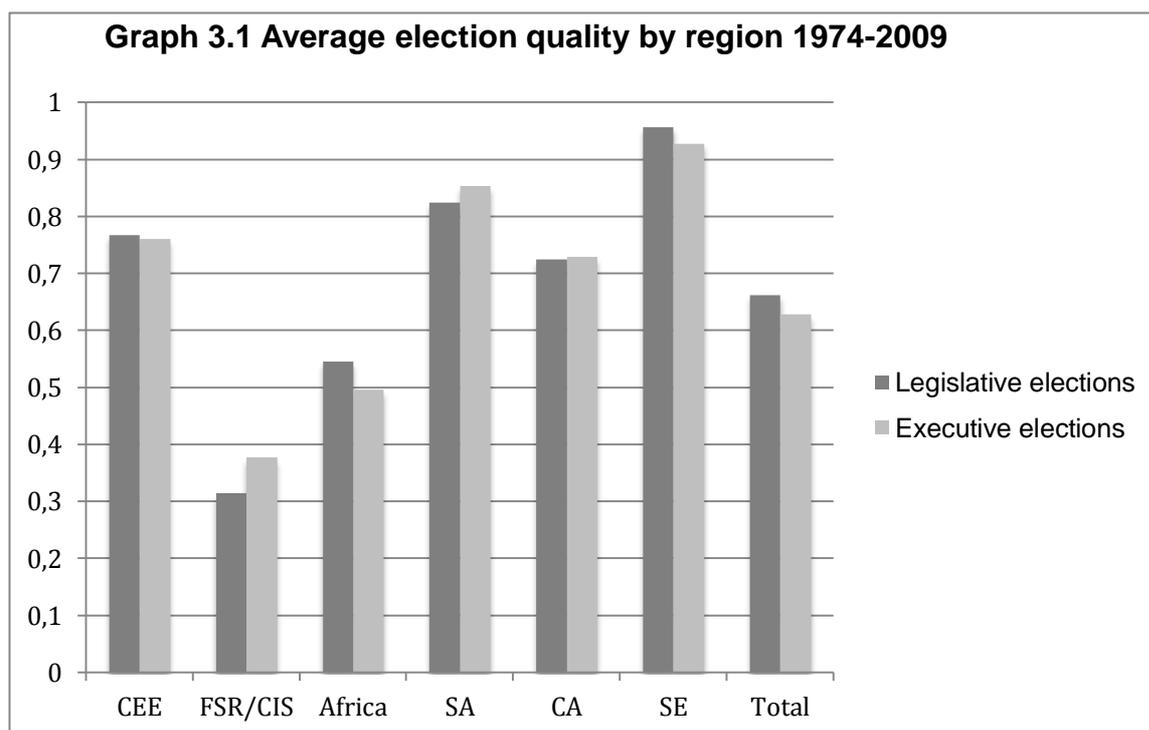
3.1. Variation in election quality in third wave regimes 1974-2009: global patterns

Regional variation

Leaving aside more precise causal dynamics, to which we will turn in chapter 4, in terms of regional variation we would expect election quality to at least roughly correspond to variation in the quality of democracy as a whole. Hence, we would expect election quality to be higher in regions where democracy is faring relatively well compared to regions where democratic development has been halted

or deteriorated (Birch 2011, Pastor 1999). In other words, in regions where democratization has given rise to diminished sub-types of democracy and hybrid regimes, we would expect election quality to be lower. This implies that the third wave democracies in Southern Europe and South America, as well as in Central and Eastern Europe and Central America should have overall higher levels of election quality than third wave democracies in Former Soviet Republics and Sub-Saharan Africa. The table below shows that the regional variation is indeed as expected. Using the ordinal average election quality variable, which runs from 0 (low quality) to 1 (high quality), with 0.5 indicating intermediate situations of medium quality elections, it becomes clear that average election quality is highest in Southern Europe and South America, followed by Central and Eastern Europe and Central America. In all these regions, average election quality lies between 0.73 and 0.95 (and between 0.72 and 1 if we exclude less reliable election quality scores), indicating on average high quality elections⁷¹. On the other hand, the Former Soviet Republics and countries in Sub-Saharan Africa have substantially lower average election quality scores. In these regions, average election quality varies from 0.34 to 0.53 (and from 0.27 to 0.52 if excluding the less reliable election quality scores), indicating on average low quality elections in the Former Soviet Republics and medium quality elections in Sub-Saharan Africa.

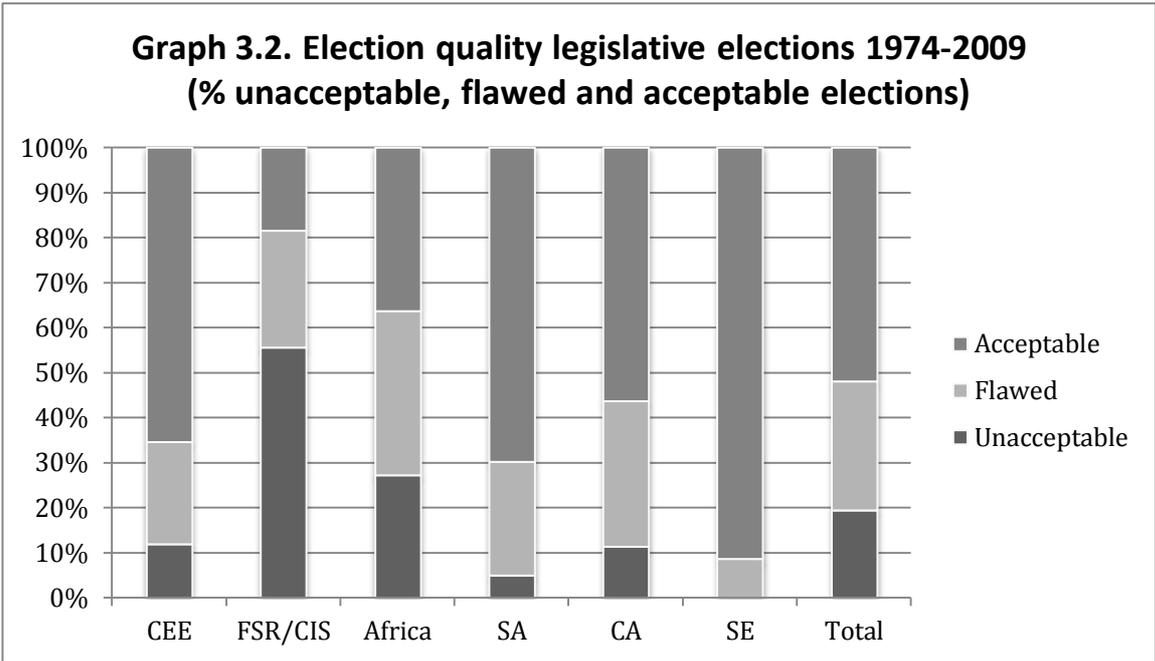
⁷¹ Elections with less reliable election quality scores are elections that were coded only once by other authors as well as elections that were coded only once by me and for which my reliability score indicates little or conflicting information.

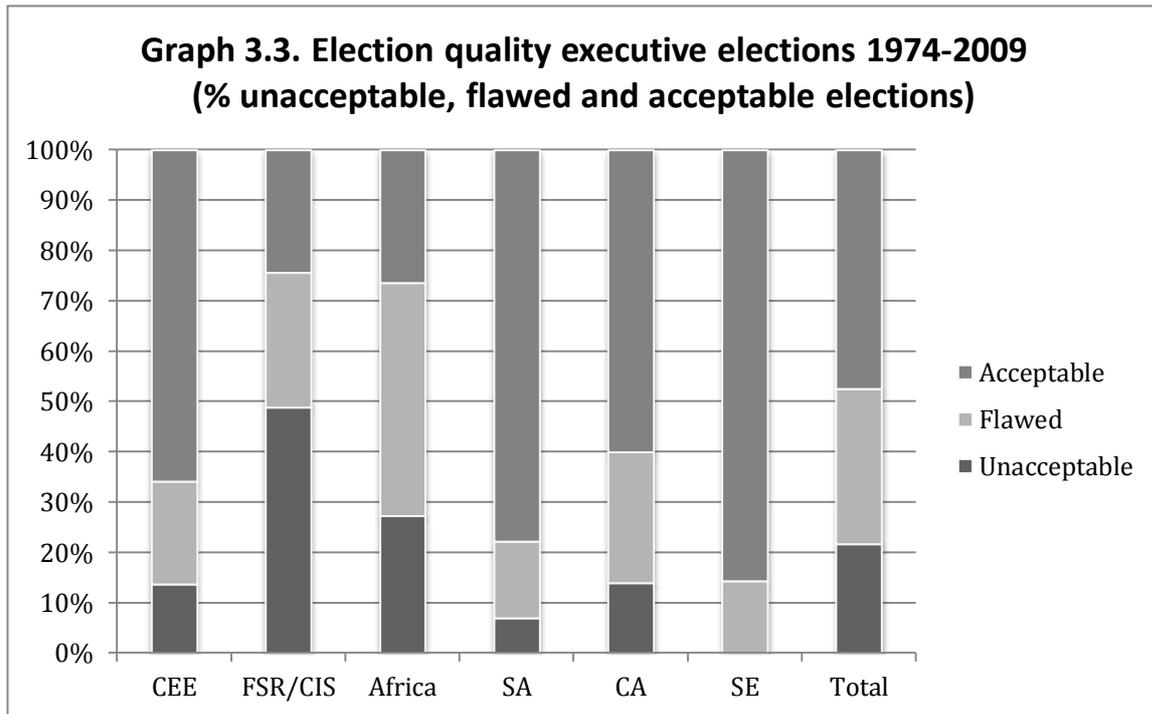


To get a better idea of the proportion of unacceptable, flawed and acceptable elections per region, graphs 3.2 and 3.3 below show the relative frequency of each type of election in each region. Starting with elections in the dataset as a whole, about 50% of our 885 elections were coded as high quality elections (52% of legislative and 48% of executive elections respectively), while 30% received an intermediate score, and 20% were scored as elections with low quality. Compared to this overall picture, elections in Southern Europe clearly score highest election quality ratings, with an average of 90% acceptable elections (between 86% and 91% for executive and legislative elections respectively)⁷². Next follows South America, where about 73% of elections are classified as acceptable, and Central and Eastern Europe and Central America, where about 66% and 58% of elections respectively are classified as acceptable. At the same time, the proportion of flawed and unacceptable elections in Central and Eastern Europe and Central America is still substantial. In Central and Eastern Europe, about 23% of legislative elections and 20% of executive elections are coded as flawed, while 12% of

⁷² Note that the figures for executive elections in Southern Europe are exclusively based on Portugal, since only Portugal has a presidential system.

legislative and 14% of executive elections are coded as unacceptable. In Central America, about 32% of legislative elections, and 26% of executive elections are coded as flawed, and about 11% of legislative elections and 14% of executive elections experienced such grave irregularities that they were coded as unacceptable. In Central and Eastern Europe, most of these low quality elections took place in countries of the former Yugoslav Republic, in Central America these low quality elections are more geographically dispersed. Compared to the Former Soviet Republics and Sub-Saharan Africa however, these regions are still doing fairly well. The pattern of acceptable elections seems to be almost reversed in the latter two regions: only about 32% and 21% of elections in Sub-Saharan Africa and Former Soviet Republics respectively were coded as acceptable. In Sub-Saharan Africa, 41% of the potentially democratic elections experienced irregularities, and were subsequently coded as “flawed” (ranging between 36% of legislative elections and 46% of executive elections), while for Former Soviet Republics this proportion was 26% (26% and 27% of legislative and executive elections respectively). Finally, the proportion of unacceptable elections is over 50% for elections in the Former Soviet Republics (between 56% and 49% respectively), and lies at 27% for both legislative and executive elections in Sub-Saharan Africa.

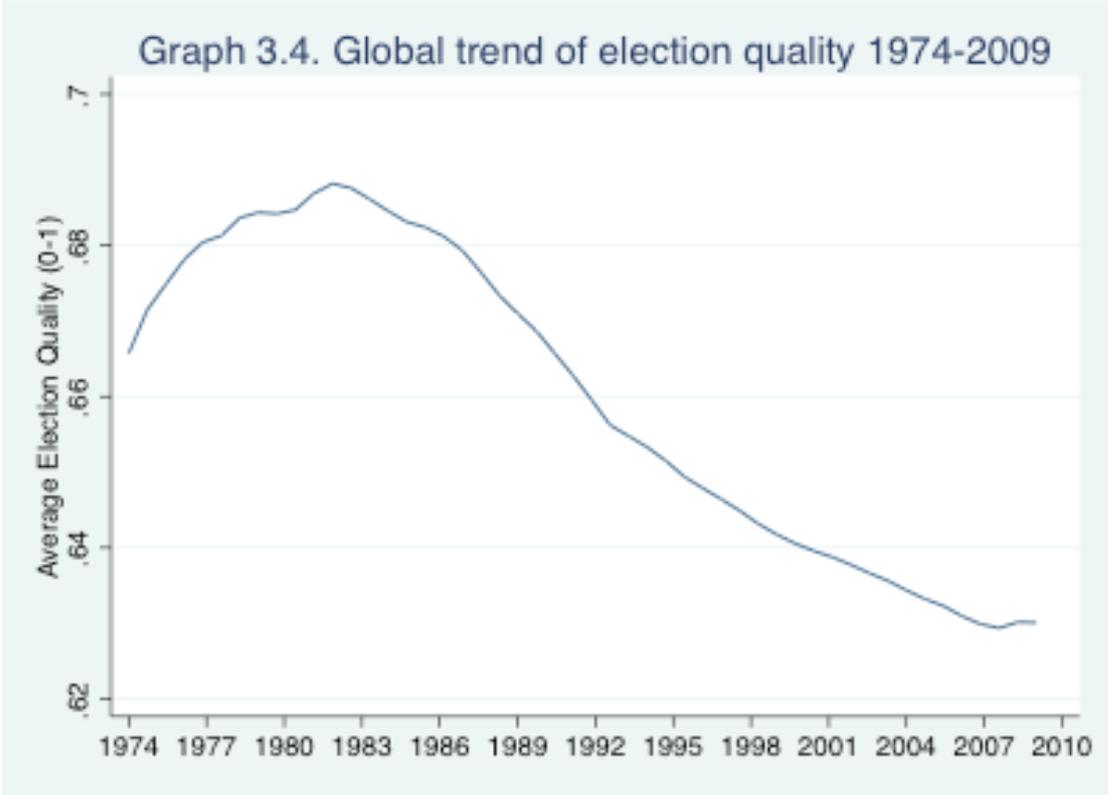




Development over time

Turning to the changes in election quality over time, several expectations can be formulated. First of all, considering global trends, one might expect election quality to have steadily improved in the past decades. The development of the international norm to hold multi-party elections (Kelley 2008, Hyde 2011) and the concomitant increase in international pressure and monitoring of elections (Hartlyn, McCoy and Mustillo 2003, Bjornlund 2004) might have helped to deter fraud and thus increase the quality of elections in third wave regimes. In addition, the broader democratization processes taking place in many third wave regimes might have positively influenced the partial regime of elections as well, contributing to a trend of increasing election quality over time. On the other hand, increasing international pressure for multi-party elections might have generated a reverse effect in hybrid regimes, as leaders attempted to butter both sides of their bread: holding multi-party elections to please the international community while manipulating them to hold on to power. As Bratton (1998) describes with respect to Sub-Saharan Africa: “As the 1990s progressed, leaders became adept at accommodating the international norm for competitive elections, while at the same time learning to

manipulate them to their own ends.” (p. 55). Moreover, the increased frequency and quality of election monitoring might have caused irregularities to be detected relatively easier, potentially also contributing to an observed decrease in election quality over time⁷³. In all likelihood, both trends occurred, and these might cancel each other out when considering global trends. Table 3.4 below shows the development of election quality over time for all the third wave regimes in our sample.



Kernel-weighted local polynomial smoothing.

The graph shows a slight increase in average election quality in the 1970s and the early 1980s, mostly referring to the elections that took place in third wave regimes in Southern Europe, South America and Central America⁷⁴. However, even though demonstrating an increase in election quality, it is

⁷³ However, this potential bias is mitigated by the fact that our overall election quality data is based on multiple sources, including election observation reports, but also alternative sources such as human rights reports, academic analyses and news media.

⁷⁴ Some Sub-Saharan African regimes held elections in this period as well, such as Botswana, Gambia, Mauritius and Madagascar.

questionable to what degree this was driven by international pressure and election monitoring, as both developed only in the course of the 1990s. Rather, increasing election quality in these regions seems to have been the result of the broader democratization processes taking place in these regions at the time that might have positively influenced the partial regime of elections as well. The turning point in average election quality seems to come already before the onset of the “fourth” wave of democratization in 1989, as election quality starts to fall already in the mid-1980s, most likely driven by elections in Central America that remained rather flawed during the 1980s and only started improving by the early 1990s (as noted by observers such as Karl 1990). From this point on however, election quality continues to decline, the decline leveling off somewhat by 2007. This observed decline seems to lend more support to propositions by Bratton and others that increased international pressure and monitoring resulted in incumbents holding multi-party elections while simultaneously attempting to undermine the integrity of the process.

However, again here the overall trend most likely conceals much greater and opposite trends in different regions and regimes. For example, it could very well be that the observed overall decline is driven by the fourth wave democratizers that change the composition of the overall sample. Since these regimes started holding their first “potentially democratic” elections later in time, and these early elections tended to be of lower election quality than subsequent elections, the global ‘decline’ in election quality might be caused by changes in the composition of the overall sample of regimes, rather than by a generalized decline in election quality within countries. Moreover, considering the graph closely reveals that the observed global “trend” in election quality only varies between 0.63 and 0.69, less than 10% of the overall scale of election quality that ranges from 0 to 1. In fact, the global trend is not even significant when regressing time on election quality.

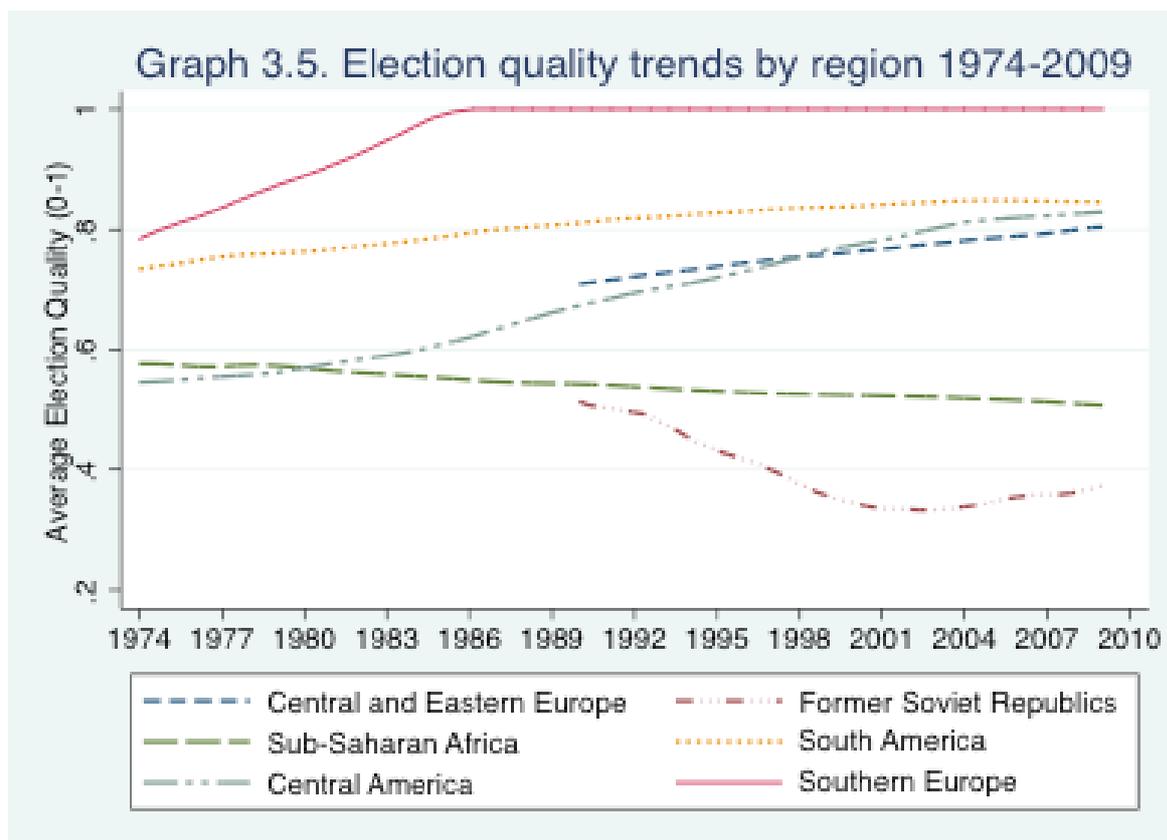
Taking a closer look at regional trends shows why the overall image might be misleading. As graph 3.5

on the next page shows, on average elections in regions as Southern Europe and South America improved in quality over time. Elections in Central America and Central and Eastern Europe started off at a somewhat lower election quality but followed the same trend of gradual improvement over time. However, elections in Sub-Saharan Africa demonstrate a slow but steadily declining trend of election quality, while elections in the Former Soviet Republics appear to decline in quality in the course of the 1990s and then more or less stabilize at a rather low average level of election quality⁷⁵.

However, regional differences do not add substantially to our understanding of why election quality might follow such different trends. Moreover, the explanations referring to increased international pressure and election monitoring mentioned above seem unlikely to provide much traction for explaining global trends either, as all fourth wave regimes experienced pressure and monitoring, and yet average election quality improved in Central and Eastern Europe while deteriorating in Former Soviet Republics and remaining relatively stable in Sub-Saharan Africa⁷⁶.

⁷⁵ Note however that even within regions composition effects are likely to play a role. In Sub-Saharan Africa for example, the late democratizers tend to be those countries with lower overall levels of election quality, leading to a regional ‘trend’ of seeming decline while in fact simply the pool of countries now includes new countries with lower quality elections.

⁷⁶ As Levitsky and Way (2010) demonstrate however, the effectiveness of international pressure depends importantly on the leverage that democracy-promoting states have over third wave regimes, which in the case of Central and Eastern Europe was clearly much stronger than in the Former Soviet Republics and many Sub-Saharan African regimes. Hence a full account of the impact of international pressure and monitoring regarding electoral conduct would have to take these factors into account.

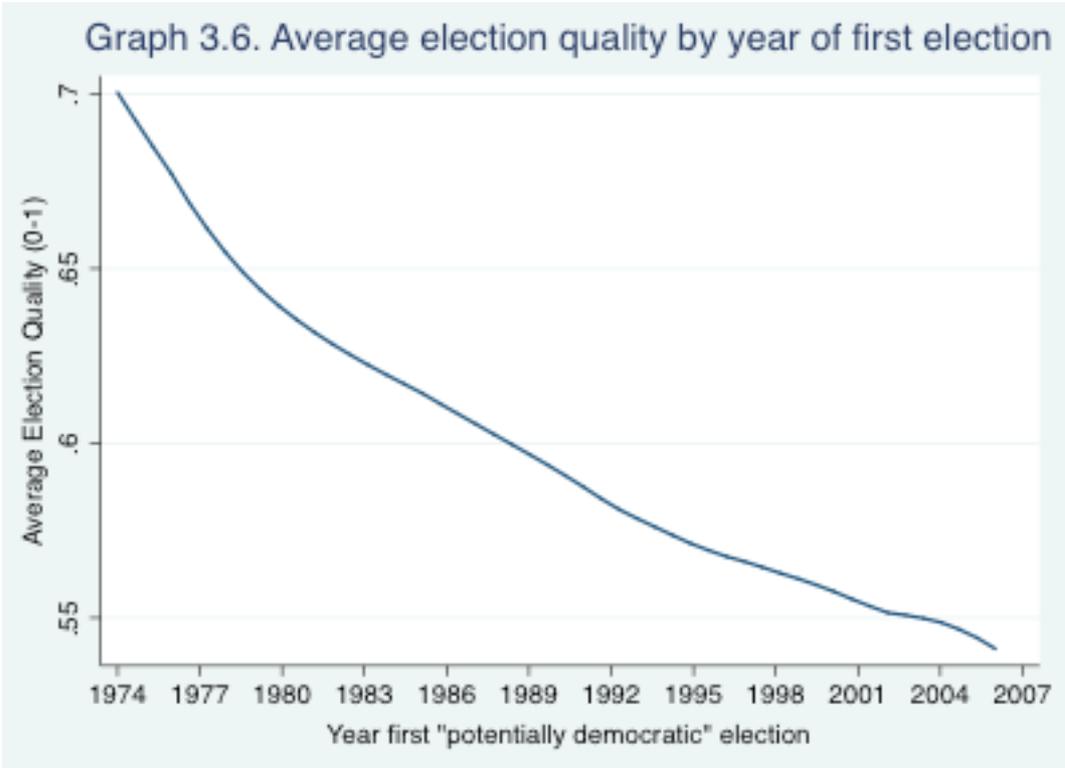


Kernel-weighted local polynomial smoothing.

While we will turn to more in-depth explanations of variation in election quality in chapter 4, when we analyze the incentive structures in which incumbents and opposition make decisions about whether or not to engage in fraud, at this point of analyzing aggregate trends some broader historical explanations might help as a first step to better understand the trends mapped above. Following the notion that democratization is a gradual process in which political elites and citizens learn and eventually come to accept the ‘democratic method’ as ‘the only game in town’, it seems likely that countries that started holding potentially democratic elections at an earlier stage, by now hold elections that are on average of higher quality than countries that have only recently started holding elections. Hence, we would expect early democratizers to have higher levels of election quality. A similar argument could be made with regard to past experience with elections: if a country had experienced potentially democratic elections earlier in its history before 1974, the bureaucracy, political elites and citizens, at least the older ones, might have experience in organizing elections and have

expectations about the conduct of elections, that might contribute positively to the quality of elections held post-1974. A related but different proposition is the notion that holding multi-party elections, even if initially fraudulent, might provide a window of opportunity that, even if only momentarily, stimulates the formation of opposition parties, mobilizes citizens, broadens freedom of journalists, etc., which at the next election might help strengthen the integrity of the electoral process. Lindberg refers to this process as the “self-reinforcing” or “democratizing” power of elections (Lindberg 2006ab, 2009), and has shown in the case of Sub-Saharan Africa that election quality tends to improve as successive elections are held, elections being increasingly free and fair after the third election. Of course, the frequency with which elections are held is itself a function of the broader democratization processes going on in a given country, however, Lindberg finds that even in countries with shorter sequences of elections, the trend of increasing election quality holds (even if average election quality is lower in these countries). This implies that instead of considering trends in election quality over absolute time, we should consider trends over relative time, measured by the sequence of elections within each country. Hence, it is more instructive to compare all first de jure multi-party and universal suffrage elections, and second and third elections and so on, than to compare all elections that took place in 1989, 1990 and so on. The expectation we derive from these findings for Sub-Saharan Africa is thus that election quality improves with each successive election, and we will analyze whether there is ground to believe that these dynamics play out in other regions as well. However, Lindberg also noted that political events, such as coup d’etats and civil wars, might interrupt this “self-reinforcing” sequence. Hence, we also consider to what degree election quality increases over successive elections in countries with and without coup d’etats and civil wars. Statistical testing of these propositions will be carried out in chapter 4, however in the subsequent sections we will explore whether these notions of the development of election quality over time bear some relation to our empirical data.

Starting with the timing of the transition to holding de jure multi-party universal suffrage elections, the expected trend appears clearly⁷⁷. As graph 3.6 shows, countries that held their first post-1974 ‘potentially’ democratic elections earlier, indeed seem to have on average higher quality elections (the trend is also statistically significant). Third wave regimes that held their first potentially democratic elections in the 1970s have high average election quality scores, between 0.65 and 0.7, while regimes holding their first elections in the 1980s have average election quality scores between 0.6 and 0.65.

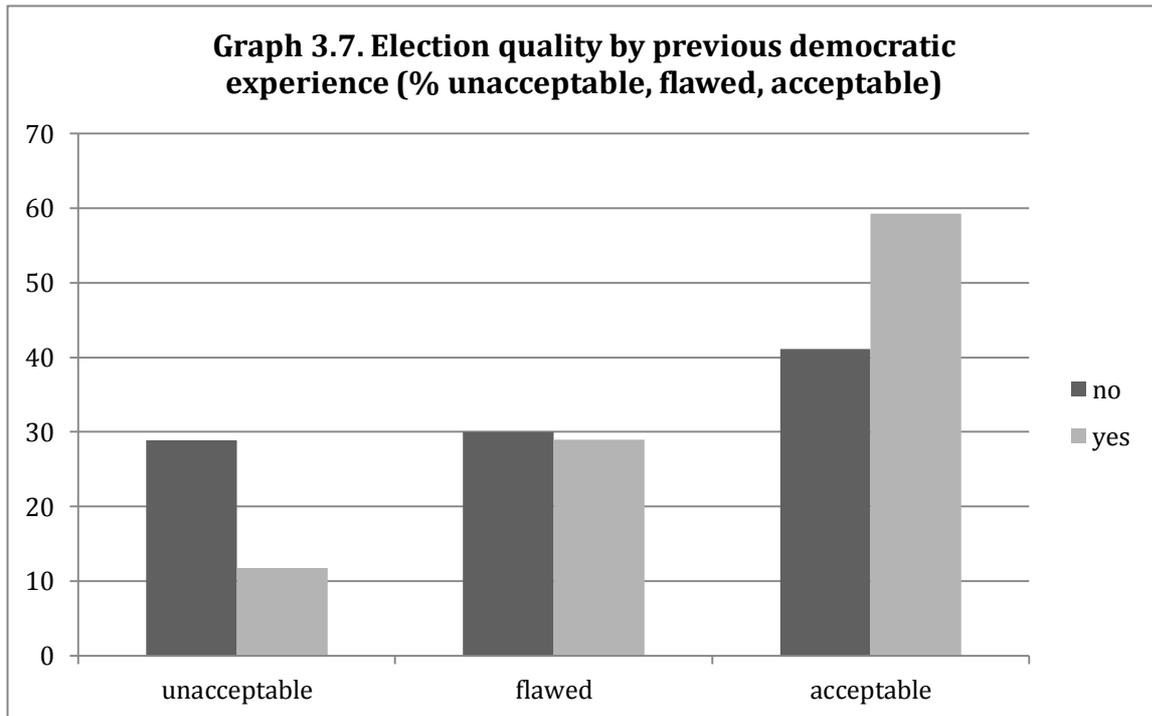


Kernel-weighted local polynomial smoothing.

⁷⁷ Note that what is coded here is the year in which the first ‘potentially democratic election’ took place according to our criteria outlined in chapter 2, i.e. the year in which the first de jure multi-party and universal suffrage election took place, starting from 1974. Since this date only concerns elections, it can take place earlier than the moment indicated by comparative democratization scholars as ‘democratic transition’. For example, the –controversial– moment of democratic transition in Mexico is often dated in 2000, since this is the first election in which the dominant party, the PRI, lost power. However, Mexico held de jure universal suffrage and multi-party elections since 1955, and hence the year of its first potentially democratic election after 1974 is 1976. Appendix B provides an overview for each country of the elections that were selected according to our criteria as ‘potentially democratic elections’.

Third wave regimes that held their first potentially democratic elections in the 1990s come closer to intermediate election quality scores, ranging between 0.55 and 0.6. This might indicate that holding democratic elections is indeed a gradual process in which political elites and citizens learn and eventually come to accept the 'democratic method', thereby improving election quality as more and more successive elections are being held; alternatively it might also indicate that the earlier democratization processes were qualitatively different than the regime changes that occurred later, resulting in differences in election quality.

Now, does the notion of learning also apply to previous democratic experience? In order to analyze this, we coded countries that had held de jure multi-party and universal suffrage elections prior to 1974 as having 'previous democratic experience' and countries that did not hold potentially democratic elections prior to 1974 as having no 'previous democratic experience'. Again, we sought to have a more precise indicator of previous experience with democracy that focused specifically on elections, rather than democracy as a whole, following the criteria outlined in chapter 2. Graph 3.7 below shows the differences in average election quality. As expected, there is a clear difference between countries with and without previous experience with potentially democratic elections. In the countries that had held de jure multi-party and universal suffrage elections prior to 1974, almost 60% of elections after 1974 were of sufficiently high quality to be scored acceptable, 29% were flawed and only 12% had such low election quality scores to be labeled as unacceptable. In countries that had not held potentially democratic elections before 1974, these figures were much less positive: almost 30% of elections after 1974 were unacceptable, and only slightly over 40% was coded as acceptable.

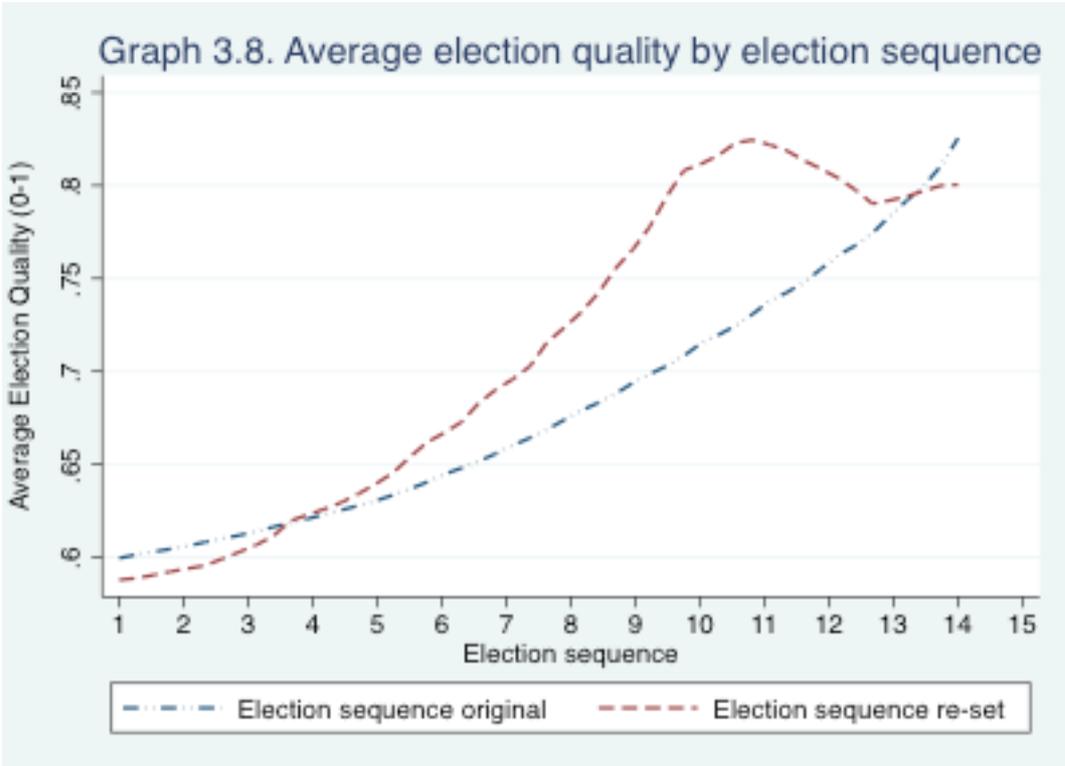


Finally, if there is indeed a dynamic of learning in repetitive elections that leads to a gradual increase in election quality over time, we should find that as more elections are held, the election quality of each successive election increases. To analyze this I created an election sequence variable that counts each successive election from the first ‘potentially democratic election’ between 1974 and 2009. Since concurrent legislative and executive elections take place on the same date, these elections receive the same sequence number. However, events such as coup d’etats and civil war that lead to regime breakdown might set back the democratization process considerably and also affect the election quality of subsequent elections. Hence, I also coded an election sequence variable that re-sets after regime breakdown. As such, first elections after a period of civil war or military rule will be coded again as first elections, which seems to give a more precise basis of comparison of elections⁷⁸. As an example: the Sudanese elections in 1986 were coded as first elections, and the 2000 elections as well. Moreover,

⁷⁸ Note that this assignment of a sequence number of elections is done for the period 1974-2009, as are “re-sets” of sequences after civil war or coup d’etat. This is not to be confounded with “previous democratic experience” that refers exclusively to the period before 1974.

for the few countries that started holding potentially democratic elections already before 1974, such as Costa Rica, this election sequence variable started counting at those first potentially democratic elections (i.e. the Costa Rican 1974 elections are the 7th elections in the sequence of potentially democratic elections).

Graph 3.8 shows the development of average election quality by both of these sequence variables. Both sequences are limited at less than 15, since the number of cases per election sequence above 15 becomes lower than 5, rendering mean values less reliable. Both variables were significantly related to election quality in regression models, even when correcting for country-differences using fixed effects. The findings of Lindberg (2006a) for Sub-Saharan Africa seem to apply to our broader sample of third wave democracies as well: clearly, as successive elections are held, their quality increases.



Kernel-weighted local polynomial smoothing.

The original election sequence variable, even if not corrected for regime breakdown, shows a gradual upward trend, election quality slowly improving until about the 6th/7th election and then increasing more markedly. The election sequence variable that takes regime breakdown into account however, seems to be more in line with Lindberg's earlier finding that election quality starts to improve significantly after the 3d election, sloping upwards more steeply from that moment on⁷⁹.

The tables on the next page give a clearer picture of the proportion of acceptable, flawed and unacceptable elections for first, second, third, and successive elections. In the first three elections, the proportion of acceptable elections remains rather stable around 45%, while the proportion of unacceptable elections only decreases slightly and then remains rather stable. From the 4th election onwards however, the proportion of acceptable elections increases to about 50% and continues to increase gradually in elections after that. This seems to suggest that indeed election quality increases as the number of elections increases. However, these findings might be driven, not by the "self-reinforcing" potential of elections, but rather by the fact that in countries where democracy consolidated, more elections tend to have been held. Indeed, countries that have held the longest sequences of elections are the older third wave democracies in South and Central America and Southern Europe, and hence the results presented below might be driven by these cases of successful democratization (again, a composition effect).

Indeed, the decline in unacceptable elections does not follow the same pattern, and decreases only very slightly, remaining at between one fifth and one quarter of elections until the 7th election, decreasing sharply only at the 8th election and higher. The fact that the median country in our sample experienced about 7 elections, indicates first of all that the higher average election quality from the 8th election

⁷⁹ Note also that this variable is more unreliable at high values, since the problem of small N occurs relatively earlier than for the other variable, i.e. values above 9 should be taken with a grain of salt.

onwards is due to the fact that those countries that have managed to hold more than 7 elections in the period under study have probably experienced more successful democratization processes, explaining both the higher number of elections and their higher quality. On the other hand, the fact that about a quarter of elections remains at very low levels of election quality, even at the 6th and 7th election, indicates that there might be a group of countries in our sample that have succeeded in holding successive low-quality elections for a considerable period of time, indicating that the “self-reinforcing” effects of elections do not occur in all countries.

Table 3.1. Proportion of unacceptable, flawed, acceptable elections by election sequence

Election sequence	Unacceptable	Flawed	Acceptable	N elections	N countries
1 st	30%	25%	45%	128	1
2 nd	22%	33%	45%	123	3
3 ^d	25%	33%	42%	119	6
4 th	25%	24%	51%	109	9
5 th	23%	27%	51%	93	10
6 th	18%	34%	48%	82	11
7 th	22%	21%	57%	68	15
8 th	8%	38%	54%	50	13
9 th	8%	38%	54%	37	10
10 th	0%	43%	57%	21	7
11 th	0%	31%	69%	16	4
12-20 th	0%	23%	77%	39	8
Average/Total	20%	30%	50%	885	97

Election quality scores between 0 and 0.33 are coded as unacceptable, between 0.33 and 0.66 as flawed, and between 0.66 and 1 as acceptable. Election sequence is coded from the first potentially democratic election since 1974 onwards, concurrent elections receive the same sequence number.

Table 3.2. Proportion of unacceptable, flawed, acceptable elections by election sequence (re-set)

Election sequence reset	Unacceptable	Flawed	Acceptable	N elections	N countries
1 st	28	31	41	187	5
2 nd	16	38	47	128	3
3 ^d	28	33	40	120	5
4 th	25	27	49	109	11
5 th	19	29	52	89	18
6 th	16	27	57	67	11
7 th	24	24	53	51	14
8 th	9	34	57	35	10
9 th	5	23	73	22	6
10 th	6	6	88	17	4
11 th	8	31	62	13	3
12 th -20 th	4	19	77	47	7
Average/Total	20	30	50	885	97

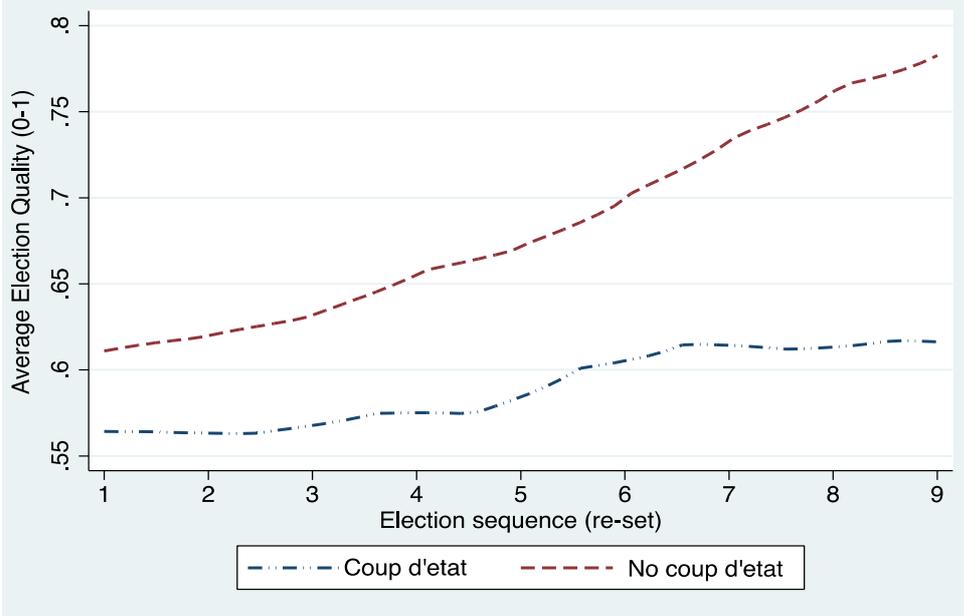
Election quality scores between 0 and 0.33 are coded as unacceptable, between 0.33 and 0.66 as flawed, and between 0.66 and 1 as acceptable. Election sequence is coded from the first potentially democratic election, also if this took place before 1974, and the sequence is re-set each time a coup d'état, civil war or popular revolution lead to regime breakdown. Concurrent elections receive the same sequence number.

The fact that the same trends are observed in both tables indicates that the way of counting election sequences does not significantly alter these conclusions. The question thus remains whether the gradual improvement of election quality as successive elections are held is a general pattern. We will return to this question in section 3.2, when analyzing patterns of election quality within countries.

Before turning to the patterns of election quality at the country level however, we consider whether the aggregate trend of increasing election quality as the number of successive elections increases might be disrupted by events such as coup d'états or civil war, as suggested by Lindberg (2006a, 2009). To consider how these events play out in our sample, we coded all coup d'états and civil wars that took place between 1974 and 2009 in the third wave democracies in our sample. Graph 3.9 and 3.10 show average election quality by election sequence, differentiating between countries in which coup d'états and civil wars took place, and countries in which such events did not occur. As in the other graphs, these graphs only include values based on 5 or more elections.

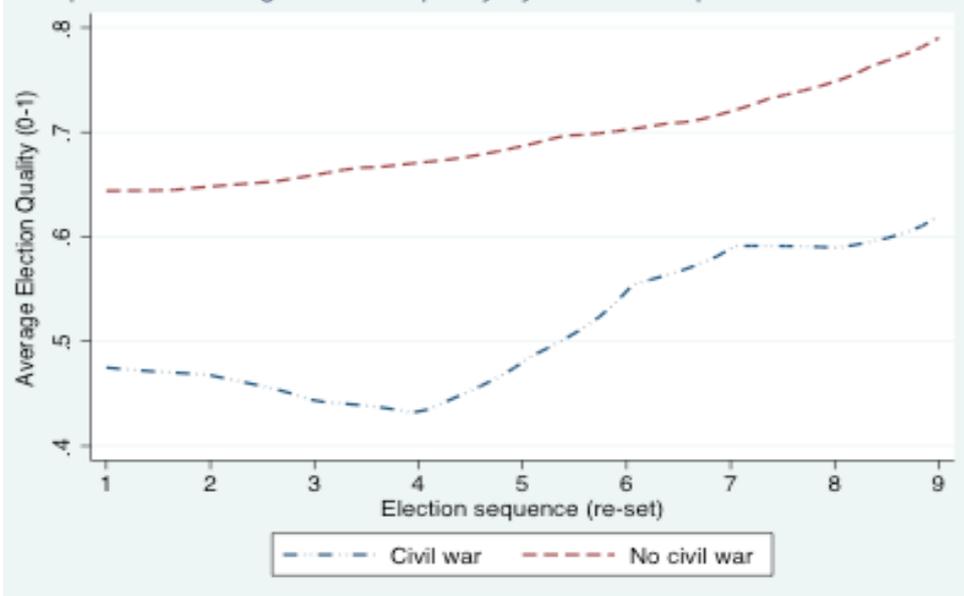
As graph 3.9 shows, in countries where coup d'états disrupted 'normal politics', the trend of increasing election quality as successive elections are held does not seem to occur. Though there is still a very slight gradual increase, the trend seems to be more erratic and average election quality remains at much lower average level. In fact, statistical analyses demonstrate that the trend is not significant for countries that experienced a coup d'état, while for countries that did not experience coup d'état(s) the trend is highly significant. Turning to civil wars, as shown in graph 3.10, the effect seems to be more disruptive in the short run, however election quality seems to increase more sharply in later elections. While average election quality remains at a lower level in countries having experienced civil war than in countries that did not experience civil war, after the 4th election a trend of increasing election quality becomes apparent. Nevertheless, statistical analyses show that this trend is not significant.

Graph 3.9. Average election quality by election sequence and coup d'etat



Kernel-weighted local polynomial smoothing.

Graph 3.10. Average election quality by election sequence and civil war



Kernel-weighted local polynomial smoothing.

Conclusion

Summarizing the global overview of regional and temporal trends in election quality presented above, it is clear that regions where democracy as a whole is faring relatively well, election quality, as a partial regime, is also higher. Thus, election quality is highest in the third wave democracies in Southern

Europe and South America, followed by those in Central and Eastern Europe and Central America, where the majority of elections are coded acceptable. Levels of election quality are substantially lower, as expected, in the third wave regimes of Sub-Saharan Africa and the Former Soviet region, where a majority of elections were coded rather as flawed or even unacceptable. This seems to be at least partly due to the later onset of democratization in these regions, as election quality appears to improve over time as countries gain experience with holding successive elections. Indeed, the notion that democratization is a process that requires learning seems to be sustained by the descriptive data. Hence, average election quality is higher for countries that started holding potentially democratic elections earlier, and election quality increases as the number of successive elections held goes up. As a consequence, for the sample as a whole, the later democratic transitions of third wave regimes in Central and Eastern Europe, the Former Soviet Republics and Sub-Saharan Africa, by some scholars also referred to as the fourth wave, results in a slight overall decline of election quality over the past three decades. However, apart from time and the resulting accumulation of experience with organizing elections, past experiences and political events also influence the level of election quality. As such it seems that pre-1974 experience with potentially democratic elections is associated with higher levels of election quality in the post-1974 period, even if such experiences dated back to before autocratic regimes. Conversely, experiences with coup d'états and civil wars, though not eliminating the pattern of increasing election quality over time, still lower average levels of election quality substantially. Nevertheless, in a sample as diverse as the sample of third wave democracies, such global and regional patterns may conceal considerable variation at the country level. It is this variation to which we turn in the next section⁸⁰.

⁸⁰ In chapter 4 we will subsequently turn to multivariate analyses of election quality to disentangle these different causal factors and attempt to explain the regional differences and temporal trends observed here.

3.2. Variation in election quality in third wave regimes 1974-2009: country patterns

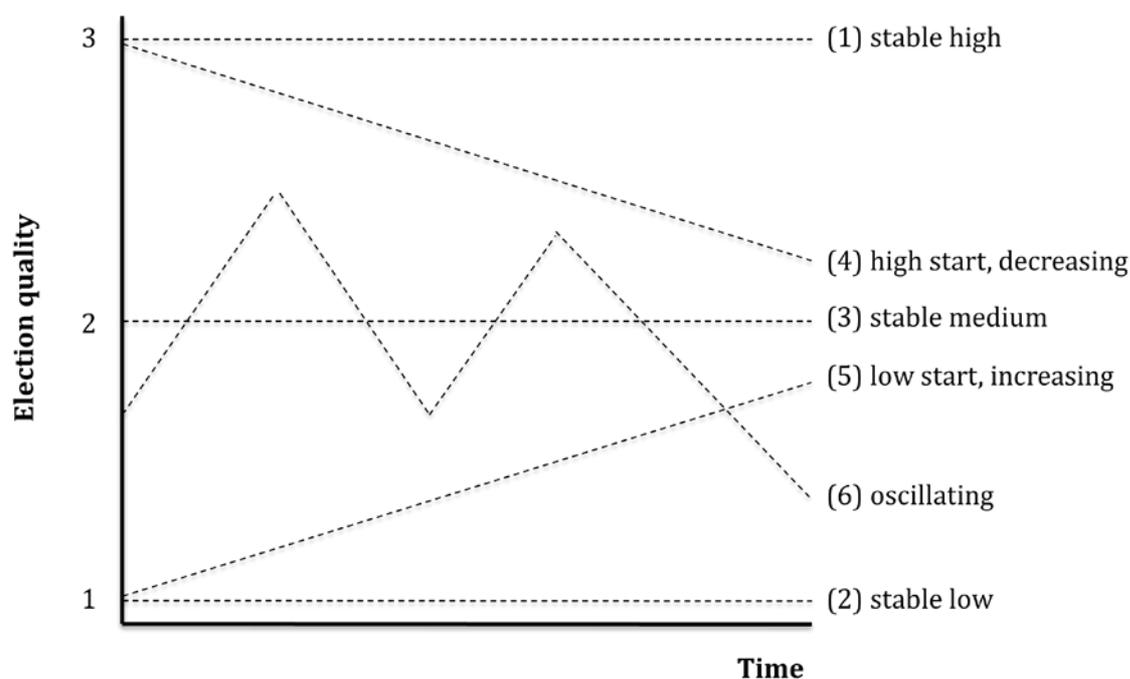
Do the aggregate patterns of election quality identified above hold for individual countries as well, or is the reality within countries more varied? In order to get an idea of the variation within countries, we plotted election quality over time for legislative and executive elections for each country in our sample. Based on the patterns that emerged from these plots, I identified 6 recurring patterns of development of election quality over time, and subsequently classified the patterns of legislative and executive elections for each country according to these patterns. Simplifying the trends somewhat, the table below shows how election quality could develop within a country over time.

Based on an ordinal classification of elections as “unacceptable” (1), “flawed” (2) and “acceptable” (3), election quality could be stable over time at any of these levels, i.e. elections could be of continuously high, continuously low, or continuously medium quality⁸¹. Alternatively, election quality could change over time, either starting high and gradually decreasing, or starting low and gradually increasing. Finally, it could be the case that election quality does not follow a clear trend over time, and changes continuously, oscillating between high, low and medium values. Clearly, this is a stylized representation, i.e. a high start could also be a start between medium and high, a low start could also be a start between low and medium, and oscillation could go in all directions. Based on these 6 stylized patterns, I classified election quality trends for each country. The following coding rules were used: elections that started off at medium election quality and decreased towards low were also coded as (4) high start, decreasing; elections that started off at medium election quality and increased towards high were also coded as (5) low start, increasing. Furthermore, election quality oscillating between 0.66 and 1 is coded as (1) stable high; election quality oscillating between 0 and 0.33 is coded as (2) stable low; and election quality oscillating between 0.33 and 0.66 is coded as (3) stable medium. Only more

⁸¹ As before, we coded elections as unacceptable if the average election quality score was between 0 and 0.33, as flawed if the score was between 0.33 and 0.66, and as acceptable if the score was between 0.66 and 1.

pronounced oscillations of election quality, i.e. those that cross boundaries between high, medium and low are coded as (6) oscillating.

The development of election quality over time: stylized patterns



Considering the trends in election quality within each country over time, and coding these trends in the categories explained above, I find the following results. In contrast to the aggregate trends of increasing election quality over time, patterns at the country level are much more varied. As tables 3.3 and 3.4 show, in Southern Europe and Central and Eastern Europe the median trend is stable high election quality, indicating these countries started off at high levels of election quality (between 0.66 and 1) in their first potentially democratic elections and maintained high quality in successive elections⁸².

⁸² The trends were re-coded using only the elections that were coded more than once or were coded by me. The results do not change substantively when doing so, and the median categories remain the same for all regions.

Table 3.3. Patterns of election quality within countries 1974 – 2009

Category	CEE	FSR	SSA	SA	CA	SE	Total
(1) stable high	9		2	4		2	19
(2) stable low		3	2				5
(3) stable medium	1		1				2
(4) high start, decreasing		2	9				12
(5) low start, increasing	1		5	2	3	1	12
(6) oscillating	6	7	25	6	3		46
(7) only one election			1				1
Total	17	12	45	12	8	3	97

Table 3.4. Patterns of election quality within countries 1974 – 2009 (%)

Category	CEE (%)	FSR (%)	SSA (%)	SA (%)	CA (%)	SE (%)
(1) stable high	52.94		4.44	33.33	25.00	66.67
(2) stable low		25.00	4.44			
(3) stable medium	5.88		2.22			
(4) high start, decreasing		16.67	20.00			
(5) low start, increasing	5.88		11.11	16.67	37.50	33.33
(6) oscillating	35.39	58.33	55.56	50.00	37.50	
(7) only one election			2.22			
Total	100.00	100.00	100.00	100.00	100.00	100.00

However, even in these regions “stable high” represents only 67% and 53% of countries respectively, as other patterns also occur. Considering the other two regions where election quality is on average relatively high and where election quality also seems to have improved over time, South America and Central America, we find a surprisingly low proportion of countries had an election sequence of continuously high quality. Rather, in Central America many countries follow a trend of starting at a medium or low level of election quality and subsequently improve over time. In South America, and in Central America to a lesser extent as well, the trend of oscillating election quality appears to be most common. This is also the case in the Former Soviet Republics and Sub-Saharan Africa, where 58% and 56% of countries respectively demonstrate a trend of oscillating election quality over time. Interestingly, these two regions also have a relatively high proportion of countries in which election quality starts off high and then gradually deteriorates over time. In Former Soviet Republics, the trend of continuously low election quality is also not uncommon.

Given that oscillating election quality occurred as such a common trend across regions, we considered this trend somewhat more in-depth. According to our coding rules, election quality trends that cross boundaries between low, medium and high election quality are coded as oscillating. However, this means in practice that countries that oscillate in election quality between medium and high, countries that oscillate between low and medium, and countries that vary more widely in the quality of their elections between low and high, are all grouped together. However, a country that experiences a series of elections with medium or high quality, even if apparently experiencing recurring irregularities in elections, is on average doing better than countries that experience series of elections with low and medium quality. We recoded the oscillating trends so as to better distinguish these differences, the result of which is shown in table 3.13 below.

Now it becomes clear that 33% of countries in South America have indeed trends of oscillating election quality, however, these variations are always within medium and high election quality, representing a series of elections with on average relatively high quality. The same counts for about 25% of countries in Central America. Now, considering all positive trends in election quality together, i.e. considering “stable high” election quality, “low start and increasing” election quality and “oscillating high” as positive trends, gives a clearer picture of developments within regions. Clearly, the majority of countries in South America and Central America, even if not having continuously high quality elections, show positive trends of election quality, either starting low and continuously increasing, or oscillating between medium and high election quality. In Central and Eastern Europe, positive trends occur in almost 60% of countries, while still about 24% of countries have patterns of election quality that vary strongly over time between low and high election quality⁸³.

⁸³ Most of these are in the Balkan region, i.e. Croatia, Serbia and Montenegro and Albania, however Romania is also part of this group.

Table 3.5. Patterns of election quality within countries 1974 – 2009 (%)

Category	CEE (%)	FSR (%)	SSA (%)	SA (%)	CA (%)	SE (%)	Total
(1) stable high	52.94		4.44	33.33	25.00	66.67	19.59
(2) stable low		25.00	4.44				5.15
(3) stable medium	5.88		2.22				2.06
(4) high start, decreasing	5.88	16.67	20.00				12.37
(5) low start, increasing	5.88		11.11	16.67	37.50	33.33	12.37
(6) oscillating:							
oscillating low	5.88	16.67	13.33				9.28
oscillating high		8.33	20.00	33.33	25.00		16.49
oscillating low-high	23.53	33.33	22.22	16.67	12.50		21.65
(7) only one election			2.22				1.03
Total		100.00	100.00	100.00	100.00	100.00	100.00
Positive trends (%)	58.82	8.33	35.55	83.33	87.5	100	48.45
Negative trends (%)	11.76	58.34	37.77	0	0	0	26.80

At the same time, considering all negative trends in election quality together, i.e. considering “stable low”, “high start and decreasing” and “oscillating low” as negative trends, gives a clearer picture of trends in Sub-Saharan Africa and the Former Soviet Republics. Here trends in election quality are much more varied than in the other regions. Especially in Sub-Saharan Africa, the proportion of countries with positive trends is almost as high as those with negative trends, 36% and 38% respectively. While stable sequences of high, medium or low election quality are quite uncommon, about 11% of countries start with low election quality and improve over time, and another 20% of elections oscillate between medium and high levels of election quality. On the other hand, about 20% of countries start off with relatively clean elections, only to see election quality deteriorate over time, while another 13% oscillate between low and medium quality elections. Hence, considering trends within countries seems to prove both Bratton (1998) and Lindberg (2006a) wrong, election quality is neither consistently decreasing nor consistently increasing over time in Sub-Saharan Africa, rather the emerging picture is one of strong variation. Turning to Former Soviet Republics, negative trends in election quality are clearly dominant in this region, however, even here only 58% of countries experience clearly negative trends in election quality. Indeed, what stands out is the relatively high proportion of countries that hold series of election with low election quality or oscillate between low

and medium quality over a long period of time. However, a substantial proportion of countries also experience strong variation over time in election quality, oscillating between low, medium and high election quality without a clear direction.

Finally, considering trends for all countries together, a positive picture emerges: about half of the countries in our sample show positive trends of election quality, either holding continuously high quality elections, starting off low but improving over time, or following a more instable path of variation between medium and high quality elections. What does become clear from these analyses however, is that gradually increasing election quality over time, though appearing from our global analyses, is certainly not the dominant trend in election quality, indicating that if there is a self-reinforcing effect of elections, this effect does not always occur.

Turning to the elections with clearly negative trends, what stands out is that the proportion of countries experiencing negative election quality trends comprises only slightly more than a quarter of our sample. Moreover, holding series of successively low quality elections, or even elections that oscillate between low and medium quality does not seem to occur frequently. Apparently, there is something about the “institutionalized uncertainty” of elections that makes holding successive series of manipulated elections difficult to sustain over longer periods of time. Finally, for almost a quarter of countries no clear development of election quality could be distinguished as elections oscillated between unacceptable and acceptable status without reaching equilibrium.

Of course, it should be noted that the majority of third wave democracies in our sample (i.e. about 70%), had held 8 or fewer successive elections between 1974 and 2009, and hence the ‘trends’ reported here are often based on a relatively limited number of elections. This means that trends will become clearer, and these results will be subject to change, as time passes and more successive elections are

being held. However, what is important about these data is that the aggregate image of increasing levels of election quality is nuanced considerably when considering trends at the country level, and that development of election quality is much more varied than initially expected. In chapter 4 we will further attempt to explicate this variation in election quality, taking into account the time-series-cross-sectional nature of the data.

Conclusion

Summarizing the descriptive findings when considering trends in election quality over time within countries, variation appears to be much greater than expected. In some regions, as Southern Europe, Central and Eastern Europe, and South and Central America, election quality trends are clearly positive, with continuously high levels of election quality, gradual improvement over time, or variation between medium and high election quality constituting the dominant trends. However, in the Former Soviet Republics development of election quality over time is often negative or trendless. In Sub-Saharan Africa all trends can be found: over a third of countries experience positive trends, while another third experiences negative trends, and a little over 20% are stuck in trendless variation between unacceptable and acceptable elections. In chapter 4 we will further assess the sources of this variation.

4. “Getting elections right”: Explaining variation in election quality

Why do some third wave regimes manage to “get their elections right” while others do not, and why does election quality improve over time in some polities, while deteriorating in others? As illustrated in chapter 3, average election quality differs markedly between third wave regimes, as does the development of election quality over time within those regimes. Even in regions with common historical and cultural backgrounds, such as the Former Soviet Republics or South America, where common patterns of election quality emerge, nevertheless important outliers exist. For example, the stable high quality elections of Mongolia stand out compared to the persistent low quality elections in most of the other countries in the FSR/CIS region, whereas the strong decline in election quality in Peru during the 1990s contrasts with the rather consistent high quality elections in most other countries in South America. In the case of Mongolia, the role of the reform-minded members of the former Communist party leadership as well as the presence of moderate opposition committed to democracy, in combination with Mongolia’s aid and diplomatic relationships with Western Europe and the United States (actively promoted by Mongolia’s political leadership as a way to maintain independent statehood vis a vis its neighbors Russia and China), have been mentioned as explanations for Mongolia’s deviant path of democratization (Doorenspleet and Kopecky 2008; Fritz 2008). In the case of Peru, the decline of election quality in the 1990s is clearly associated with the period in which former President Fujimori was in office. As these examples illustrate, part of the variation in election quality in third wave regimes is likely to be driven by idiosyncratic factors, that are contingent on specific political actors involved at a certain point in time, such as a particular president bent on maintaining power through electoral manipulation, or, conversely, the choice of a political party elite to mark the transition to democracy with clean multi-party elections. However, as important as such factors may be in explaining election quality within specific third wave regimes, the emergence of common patterns within regions seems to indicate that there are also structural factors at play. The

aim of this chapter is to assess the importance of such structural factors in explaining variation in election quality, both between countries as well as over time.

4.1. Explaining election quality: incentive structures and norms

Why would incumbents, and more broadly, political actors engage in electoral fraud? Assuming that politicians are to some degree rational actors, this decision would depend on a comparison of benefits that are to be reaped from electoral fraud versus the costs of engaging in electoral fraud. Benefits of electoral fraud are first of all, the degree to which fraud increases one's chances to win at the polls, and, secondly, the spoils of office. The latter include both direct benefits, such as access to political power and patronage, and legitimacy of power in the eyes of citizens, opposition parties and international observers (if the fraud remained covert), as well as potential future benefits in the form of an incumbency advantage to commit electoral fraud in the next elections⁸⁴. The estimation of potential benefits of fraud is furthermore likely to be affected by the nature of the electoral race, as this determines the share of direct and future benefits that will be obtained when winning the elections (and conversely, the costs of losing). If the electoral race is a zero-sum game, where winners win access to office and losers are marginalized, the trade-off between winning and losing is very sharp, potentially increasing the incentives to engage in fraud. However, if the formal and informal rules of electoral competition allow for more opposition influence, even losing office might still provide some influence over policy-making, or at least chances for delivering pork to one's constituency, as is the case in clientelist systems (Lust-Okar 2009).

⁸⁴ It is important to re-iterate that both incumbents and oppositions can and do engage in electoral fraud. However, incumbents' access to state resources and media provides them with more extensive means to manipulate elections, which might add to the perks of office.

As for costs of electoral fraud, we could differentiate between implementation costs and legitimacy costs (Birch 2011). Implementation costs are direct costs, i.e. the costs to buy votes, select and pay people to stuff ballot boxes, bribe judges, journalists and election officials, etc., and are incurred in any case by politicians wanting to commit electoral fraud. Conversely, legitimacy costs are potential costs that are only incurred if electoral fraud were to be discovered. These costs can be quite high however. For example, if electoral fraud is obvious, the new incumbent(s) lacks domestic legitimacy which might lead to lower compliance with legislation by organizations and citizens, civilian protests and even armed rebellions or attempted coup d'états. Obvious fraud might also affect international legitimacy, and possibly lead to political pressure and/or economic sanctions by international organizations and states. Note that the degree to which legitimacy is of any concern to regimes might differ. For example, in the context of an electoral autocracy with a strong internal security apparatus compensating for lack of internal legitimacy towards citizens, and trade relations with predominantly other electoral autocracies that do not require legitimacy of government as a condition for trade, as in the case of Belarus for example, the costs of electoral fraud, even if overt, can be limited, as the 2010 elections have shown⁸⁵. However, for most regimes in our sample, it seems safe to assume that some degree or at least semblance of domestic and international legitimacy will be considered important, or at least preferable to none at all, and therefore perpetrators of fraud will attempt to keep their actions covert. It follows that what determines the potential costs of fraud in these cases is the likelihood of detection. Hence, potential costs refer both to the risk of electoral fraud being detected as well as the legitimacy costs if detected.

Summarizing, the incentives to engage in electoral fraud are likely to be shaped by a consideration of (a) the chances of winning in 'clean' elections versus the chances of winning in fraudulent elections, (b)

⁸⁵ In fact, Simpser (2005) argues that in those cases, overt fraud is a useful tool for incumbents to demonstrate their strength to voters and opposition, deterring opposition participation –and votes– in future elections.

the nature of the electoral race and (c) the benefits of winning, versus (d) the direct costs to engage in fraud, (e) the chances of detection and (f) potential costs if detected. To take a closer look at these incentives, I follow the instructive example outlined by Collier (2009), and propose to imagine impersonating a politician considering whether or not to engage in fraud in the upcoming elections. The first question is whether he or she is more likely to win with fraud than without. If the politician is an incumbent, and the government has performed well in the past years, our model-politician can be reasonably confident that citizens will re-elect him or her. Alternatively, if the opposition is weak and fragmented, even if government performance was less than satisfactory, there is no viable alternative to replace the incumbent(s). However, even though good government performance in the years before elections and lack of viable governing alternatives might decrease this politician's chances of defeat at the polls, one can never be certain. It follows that the comparison of chances of winning with or without fraud will always turn out in favor of engaging in fraud, because fraud will always increase one's chances to win compared to running in a clean election.

A similar argument holds with respect to the second consideration, the nature of the electoral race. If elections are less of a zero-sum game, losing the electoral race is relatively less costly. For example, if losing the elections will still provide access to some of the perks of office (for example representation in parliament and access to pork-barrel deals in exchange for support for the incumbent), if losing the election does not lead to persecution by the government after the elections, if there is a chance of participating and perhaps even winning in the next elections, and if winning office is not the only way to gain access to resources, incentives to engage in electoral fraud might be relatively less strong than in a zero-sum race. Conversely, if the stakes of an election increase, the costs of losing also increase, and political actors might choose to mitigate the risks of losing by engaging in electoral fraud. Note that the stakes of an election are influenced by the social and economic structure of a polity just as much as the nature of political institutions. While political institutions such as majoritarian electoral

systems or presidentialism might increase concentration of power after the elections and enhance the zero-sum character of the electoral race; salient ethnic or religious cleavages, or social cleavages as the result of civil war, as well as an economic structure where economic activity is highly concentrated in the state might have similar effects. While we would expect such factors to increase the incentives for political actors to engage in fraud, also here we can assume that, regardless of the nature of the race, all political actors will prefer winning over losing. The risks involved in losing are higher in zero-sum electoral races and hence might enhance incentives to engage in electoral fraud, but this doesn't necessarily mean that political actors have no incentives to engage in electoral fraud in lower-stake electoral races.

Finally, we should also expect that if the benefits of winning are higher in a given political system or at a given moment in time, political actors will be more willing to engage in electoral fraud. This might be the case in regimes where governments have control over natural resources, or where economic activity in general is concentrated in the state, and in systems with concentrated power of incumbents and relatively little checks and balances.

It follows that, since winning elections will always be more likely with electoral fraud than without (Collier 2009), and political actors will always prefer to win the elections over losing, incentives to engage in electoral fraud could be considered present under all circumstances. The incentives might be relatively stronger if incumbents are unpopular, if the electoral race is a zero-sum game, and when the benefits of office are high, however even without these conditions, a 'rational' office-seeking politician will engage in fraud. So, if there is always an incentive to engage in fraud, this means that we should find variation in the level of electoral fraud to be influenced most importantly by the costs of fraud. For direct costs this implies that if costs of fraud are lower, we should find more of it. Indeed, in line with this idea, vote buying has been found to occur more often when votes are 'cheap', i.e. poor voters

tend to be targeted relatively more often (Birch 2011, Bratton 2008, Brusco, Nazareno and Stokes 2004). The question is whether even 'cheaper' methods such as voter intimidation and electoral violence are also targeted more frequently at poor voters. Effective electoral fraud also requires organization however, and hence the direct costs of fraud are also likely to be lower if political parties are well-organized and have strong linkages to society, as is illustrated in the case of Taiwan by Wang and Kurzman (2007). Organizational costs of electoral fraud are also likely to be lower, at least for the incumbent, if for example the electoral administration is government-controlled or partisan, or if the bureaucracy can be used to engage in fraud. Indeed, Hartlyn et al (2008) find that non-partisan independent electoral management bodies (EMBs) significantly improve the quality of elections. Direct costs of electoral fraud might also be lower in majoritarian electoral systems as well as electoral systems with small districts and high malapportionment, since in those cases relatively less votes are needed to win (Birch 2007, Hicken 2007).

Turning to potential costs, factors affecting the likelihood of detection are expected to play a role, as well as factors affecting the legitimacy costs of fraud. Starting with the likelihood of detection, fraud is more likely to be denounced if there is an independent media, when there are active civil society organizations and vigilant opposition parties monitoring the elections, and when there are international actors present such as election monitors, but also the presence of election assistance or even peace-keeping missions could increase the perceived risks of detection and thereby lower fraud (Birch 2011, Lopez-Pintor 2010). Apart from detection, subsequent punishment of fraud is likely to be important as well, and hence I would expect independence of the judiciary and its adjudication of electoral disputes to increase costs of electoral fraud as well. Regarding legitimacy costs then, these are likely to shape actors' decisions regarding whether to engage in fraud (as well as which types of fraud to engage in, but that goes beyond the scope of this thesis). Domestic legitimacy costs are likely to always be high, as citizens in new democracies often have high expectations of -at least their first- de

jure multi-party elections, and I can imagine few places where citizens would happily endorse electoral fraud. Having said that, in countries where patron-client relationships are relatively common, practices such as vote-buying might be perfectly acceptable and even expected from politicians (Schaffer 2007). Likewise, in patriarchal societies, family voting, i.e. often elderly men voting for women and sometimes also younger men, might be common and accepted (cf. Schmeets 2010). Hence, certain types of electoral fraud might be relatively less costly in terms of domestic legitimacy than other types of fraud, and the costs of fraud for domestic legitimacy may vary in different countries depending on what is culturally acceptable or expected. Turning to international legitimacy costs, following Levitsky and Way (2010) I expect these to be higher if countries have strong linkage with Western democracies and if the leverage of Western democracies on those countries is higher. More concretely, this implies that if regimes have intense political, social and economic links to Western states, losing international legitimacy as the result of electoral fraud is relatively more costly than if these links are less tight. Also, regimes that depend on international investment and aid incur higher costs of losing international legitimacy than regimes that are economically powerful, dispose of natural resources, or are otherwise geo-strategically important.

Concluding, I expect electoral fraud to be shaped by a comparison of incentives, i.e. (a) the chances of winning in 'clean' elections versus the chances of winning in fraudulent elections, (b) the nature of the electoral race, and (c) the benefits of winning; versus disincentives or costs, i.e. (d) the direct costs to engage in fraud, (e) the chances of detection and (f) potential costs if detected. I also expect costs to affect variation in electoral fraud more strongly than benefits, as the latter presumably speak in favor of electoral fraud under all circumstances. Of course, this discussion is merely meant as an illustration of the idea that political actors are likely to make (bounded) rational calculations about the benefits and costs of engaging in electoral fraud. Undoubtedly, decisions by incumbents and opposition parties to engage in electoral fraud are based on more considerations than just subjective benefits and costs and

naturally, we cannot measure the actual subjective benefits and costs that incumbents and opposition parties expect to derive from electoral fraud (unless one happens to stumble upon a particularly frank politician *ex ante* the elections). Nevertheless, thinking about electoral manipulation in this way allows us to identify structural factors that might shape the incentives of actors to engage in or refrain from electoral fraud. In the next section we discuss these structural factors more in-depth, grouping them - for reasons of clarity- in institutional/political, economic, socio-cultural, international and historical factors⁸⁶.

Institutional and political factors

Concerning political institutions, I hypothesize that presidential political systems and winner-take-all electoral systems increase the benefits of electoral fraud as compared to parliamentary and proportional electoral systems. Presidential systems are likely to increase the “size of the prize” to be won by elections, as a powerful presidency that concentrates power in one person might provide more possibilities to deliver benefits to the president’s constituency, be it through political influence or pork⁸⁷. This concentration of power in the presidency might also render engaging in electoral fraud easier, at least for incumbents, especially if presidential power is unchecked by institutions as the media and the judiciary. Moreover, the high benefits of presidential office and incumbency advantages in terms of ease of future electoral fraud make the electoral race for the presidency a high-risk game, increasing the zero-sum nature of the race, which further augments incentives to engage in fraud.

⁸⁶ As discussed in chapter 2, election quality can be defined both as a sub-category of democratic quality and as a sub-category of corruption. All variables proposed for the theoretical model in this chapter derive from causal factors identified in the broader literature on political corruption and the quality of democracy.

⁸⁷ The case studies of Kenya and the Democratic Republic of Congo by Michela Wrong (2009, 2000) provide excellent illustrations of the idea that presidential systems are more prone to such forms of patronage (which in Africa often takes the form of ethnic favoritism) than parliamentary systems (see also Collier 2009).

A similar, though less pronounced, dynamic might be at work in winner-takes-all electoral systems. Winner-take-all systems are also likely to increase the zero-sum nature of the race, and even more so in single-member districts than in multi-member districts, as the winner(s) of the elections gains access to legislative office to represent the district constituency as a whole, while the loser(s) remain unrepresented until at least the next election. Moreover, in a majoritarian electoral system generally much fewer votes are needed to win an election than in proportional electoral systems, lowering the costs of electoral fraud (Birch 2007). However, the effects of both presidentialism and majoritarian electoral systems are likely to be moderated by the existence of institutions or actors that provide checks on executive power in between elections, such as federal states or autonomous regions, an independent judiciary and an independent media. Federalism generates a division of power between the national executive and regional or local executives, which might render engaging in electoral fraud more costly, even if only due to increased costs of coordination, and might also render benefits to be reaped from electoral fraud more limited, by constraining the discretion of the national executive.

Moreover, the independence of the judiciary and the media provide important checks on executive power in between and during elections that greatly increase the costs of electoral fraud. Clearly, electoral fraud becomes more difficult to organize if independent judges oversee the electoral process, as the final defeat of the PRI in Mexico in 2000 (and already partly in 1997) demonstrates. Conversely, electoral fraud becomes considerably easier if the judiciary is partisan and appointed by the executive, as elections in Uzbekistan (and many of its CIS neighbors) show. As regards the independence of the media, the costs of electoral fraud are increased because the risk of electoral manipulation being discovered and publicized to citizens (and foreign observers) is higher. This is problematic because most actors engaging in electoral fraud will have a stake in keeping their manipulation of the electoral process covert, since a clean election provides election winners with legitimacy in the eyes of both citizens and opposition parties, as well as foreign governments, lowering

internal costs of repression and increasing the chances for sustained trade and aid relationships with foreign governments (Lindberg 2009). An independent media will make hiding electoral fraud more difficult, and hence engaging in fraud more costly as other ways of manipulating the electoral process will have to be sought. For example, instead of stuffing ballot boxes or buying votes at the voting booths, political actors might have to change their strategies to more subtle methods such as wholesale vote-buying through intermediaries, voter intimidation through employers and local chiefs, etc⁸⁸. Finally, in addition to the checks on executive power provided by federalism, the judiciary and the media, the strength of opposition parties also seems to be an important factor in deterring electoral fraud. A strong and united opposition, even without holding a majority in parliament, might provide closer scrutiny of the way in which the incumbent government organizes the elections, increasing the costs of electoral fraud. At the same time, a strong and united opposition also increases the incentives for incumbents to engage in electoral fraud as the risk of losing the elections becomes higher, so this variable might affect election quality in both directions, rendering its effect in an overall statistical model potentially insignificant⁸⁹.

Economic factors

In addition to institutional and political factors, economic conditions are likely to shape the incentives to engage in electoral fraud as well. The literature on corruption suggests that corruption is higher in

⁸⁸ In addition to these ‘horizontal accountability’ mechanisms that might limit electoral fraud, ‘diagonal accountability’, or accountability enforced by citizens, also exists (Schmitter 2004). The electoral revolutions that took place in Serbia, Georgia, Ukraine and Kyrgyzstan are examples of social movements in reaction to electoral fraud that subsequently forced governments to hold new elections and improve the quality of future elections. We will return to citizen mobilization as an explanatory factor later.

⁸⁹ Another problem with including opposition strength as an explanatory factor for election quality is that opposition strength is to a certain extent endogenous to election quality: in clean elections, opposition parties are more likely to gain votes than in fraudulent elections. Hence, if we would measure opposition strength by the outcome of the elections, endogeneity would be a problem. This will be addressed in the statistical analyses by testing the effect of lagged opposition strength, i.e. opposition strength in the previous elections.

countries with low levels of economic development, high levels of income inequality, and an economy that thrives to a large extent on natural resources (Treisman 2007, Montinola and Jackman 2002). These factors have also been identified as hampering democratic consolidation and democratic quality (Przeworski et al. 2000). How could such economic conditions shape incentives for electoral fraud? In terms of incentives to engage in electoral fraud, low economic development might lower the cost of engaging in electoral fraud through such techniques as vote-buying, since more citizens will be willing to sell their vote in return for direct financial gain. In fact, as has been illustrated by Collier (2009) and Stokes (2007), in subsistence economies the future is so uncertain that citizens often view elections as opportunities for immediate and safe gain, rather than a chance to choose a government that might – or might not- provide public goods in the future. In addition, economic inequality might increase possibilities of engaging in wholesale vote-buying or –even cheaper- voter intimidation via local strongmen or employers. Also, the concentration of economic wealth in a small elite renders political cooptation of these elites easier, thus weakening political opposition and consequently scrutiny of the quality of the electoral process, further lowering the costs of electoral fraud. The dependence of the economy on natural resources, and ownership by the government of these resources, also referred to as the ‘resource curse’, might have similar effects of creating a small economic (and political) elite that is dependent on the government for its wealth and hence easier to co-opt, while at the same time providing access to funds to finance electoral fraud. Hence, economic dependence on natural resources might render electoral fraud less costly, while at the same time increasing the benefits of office. Birch (2011) indeed finds the “resource curse” to significantly increase electoral malpractice.

Social and cultural factors

Social and cultural factors that could shape the incentives to engage in electoral fraud have also been proposed in the literature on corruption and democratization. Research on corruption has found that in countries with higher levels of education and/or literacy rates, corruption is less frequent (Ades and

Di Tella 1997, Svensson 2005). Indeed, an informed citizenry seems essential to democratic quality, as more highly educated and especially literate citizens are more likely to be able to inform themselves about government conduct via the media, and subsequently hold government accountable through elections, protests or other means. Applied to electoral fraud, informed citizens might be less open to vote buying or voter intimidation and might protest when elections are blatantly rigged, as happened in the electoral revolutions in several Former Soviet Republics. Apart from education, cultural and religious values have been put forward as explanations of democratic quality and corruption as well. As such, traditional religious and cultural values seem to be associated with higher levels of corruption, as compared to more secular and 'modern' cultural values (Montinola and Jackman 2002; Sandholtz and Taagepera 2005). In line with this finding election quality might be lower in societies with stronger traditional values as elections might be more prone to serve 'clientelist' purposes under such circumstances. Whereas findings relating religious and cultural values to corruption and democratic consolidation are quite contested, a related –and more consistently empirically substantiated– argument is that social fragmentation might contribute to electoral fraud. While fragmentation can occur also along linguistic or religious lines, in many third wave regimes social fragmentation is ethnic. Ethnic fragmentation has been found to undermine democratic consolidation and increase levels of corruption, depending on the mobilization and salience of ethnic identities (Alesina et al. 2003, Norris 2008, Reynolds 2011). Related to this claim, in societies that are ethnically fragmented electoral fraud might be more common as well, as votes can be more easily mobilized along ethnic lines (lowering the costs of fraud), and the spoils of office can be distributed along ethnic lines either through patronage or, if ethnic groups are regionally concentrated, through area-specific spending. Wrong's description of the case of Kenya shows that once such a dynamic is initiated, it might become difficult to break the pattern, even if there is turnover in power, as the incoming ethnic group will consider it 'their turn' to benefit from the spoils of office (Wrong 2009).

International factors

Related to the economic conditions mentioned are international factors, both in terms of trade and aid relations as well as the geo-strategic importance of countries. Despite the international diffusion of democracy in the immediate aftermath of the Cold War, and the concomitant cascade of multi-party elections in Central and Eastern Europe, Former Soviet Republics, and Sub-Saharan Africa in the early 1990s, more recent developments in these polities suggest that many of these initial moves to democracy might have been 'cosmetic' changes driven by international pressure rather than the genuine desire of incumbents to transit to a democratic system of government. In relation to electoral fraud this implies that whether or not electoral fraud is costly for international legitimacy depends on the geo-strategic position of a country and its trade and aid relations (Levitsky and Way 2010). While the costs of lacking domestic legitimacy are likely to be similar for all governments (i.e. higher costs of repression), the external costs depend on whether the countries' trade and aid relationships are predominantly with autocratic regimes or with democratic regimes. Moreover, even if trade and aid relationships are predominantly maintained with democratic regimes, an additional question is whether the countries' geo-strategic importance trumps concerns for democratic governance or not (Kelley 2010, Donno 2010). Hence, for many countries in the Commonwealth of Independent States, which have trade and aid relationships predominantly with Russia, holding flawed elections might not be that costly, as clean elections are not a condition for these relationships to thrive. The increasing economic influence of China in Sub-Saharan Africa might have a similar effect on third wave regimes there (Levitsky and Way 2010). On the other hand, for countries in Central and Eastern Europe, holding clean elections is one of the many conditions that the European Union requires to be met for accession, and continuously emphasizes in trade and aid relationships. Other international factors that are likely to be important for election quality are whether international organizations are involved in election assistance or the monitoring of elections, or whether there is a peacekeeping operation present in the country at the time of the elections (Bjornlund 2004).

Historical factors

In addition to the factors discussed above, historical factors might also play a role in shaping the incentives for political actors to engage in electoral fraud, as historical experiences might shape the expectations of citizens and political actors as to how the electoral process should function. Here, turning our initial question upside down can help: Why would politicians strive to hold high-quality elections? Clearly, holding high-quality elections can also bring about benefits for political actors, most importantly legitimacy in the eyes of citizens and opposition, which in its turn might help to ease policy implementation and broaden government discretion for decision-making. After all, even if disagreeing with the government, citizens and opposition parties will know the government came to power legitimately and that there will be a fair chance for their candidate to win the elections next time round. However, this only works if the rules of the game are sufficiently institutionalized that election losers can accept the results based on the expectation that they will in fact have a chance to win in the next elections. This is what Bunce (2000) described as the key difference between established and new democracies: where the former have uncertain outcomes, but certain procedures; the latter have uncertain outcomes and uncertain procedures. The question is then how to arrive at the state in which the procedures are certain. Here, the literature points to the importance of socialization and learning in shaping the norms and expectations on the basis of which political elites (and citizens) interact. For example, if elites come to power that are strongly committed to democracy, as in the case of Portugal after the 1974 revolution, these political actors initiate a cycle of regular democratic elections, which generate gradual acceptance of norms of democratic elections and eventually active implementation of such norms also by new elites, creating a reinforcement mechanism based on social norms and existing practices. These norms and practices can to some extent also be influenced by historical experiences with democracy. For example, when Greece started holding multi-party elections again in 1974, the older part of the electorate and political elite had experienced democratic elections in the period between 1956 and 1964 and hence most likely had norms and expectations of

democratic elections based on these earlier experiences. While such reinforcement mechanisms create inertia, i.e. high election quality reinforces itself, and low election quality reinforces itself, change also occurs, and might be triggered by events like a civil war, a coup d'état or an (electoral) revolution, that suspend the existing "rules of the game" and allow those rules to be "re-set" or "re-defined". More gradual change can also occur if political actors gradually 'erode' existing rules of the game, as Fujimori in Peru and Chavez in Venezuela, or if political actors gradually push for reforms, as in the case of Mexico.

Hence, in addition to the causal factors mentioned earlier, I would expect election quality in previous elections, historical experiences with democracy and autocracy and events like civil wars, coup d'états and electoral revolutions to influence election quality as well. As regards the influence of the quality of previous elections, this simply refers to the reinforcement mechanism discussed above generating a degree of inertia in election quality over time. In addition to this aspect, we will also test the opposing hypothesis that election quality increases as successive elections are being held, based on the notion that over time organizational experience and knowledge about how to organize elections is built up and norms of democratic elections gain acceptance. This hypothesis, put forward by Lindberg (2006a) in the case of Sub-Saharan Africa, does not seem to be born out by our descriptive analyses of trends of election quality in chapter 3, as election quality certainly does not only trend upwards over time in all cases, however this might be due to other explanatory variables not yet taken into account.

As for historical experiences with democracy and autocracy, I expect previous experience with democratic elections to influence election quality positively. As in the example of Greece, if countries have held democratic elections in their past, citizens and political actors (at least the older ones) are likely to have a notion of democratic elections that influences the quality of current elections. The degree to which such past experiences have an impact seems likely to depend on how long the

experience(s) lasted and how long ago they are. As regards autocratic legacies, the degree to which pluralism existed under autocracy may be relevant. Here, a crucial difference may exist between personalist and single-party regimes on the one hand and military regimes other hand, with the latter more often experiencing factions and splits, i.e. some degree of pluralism, while personalist and single-party regimes seem to be more effective at limiting dissent and maintaining the central leaders' or party control (Geddes 1999). Hence, we might expect former military regimes to have relatively higher quality elections. As to colonial legacies, another type of institutional memory may be relevant, namely the degree to which state institutions were built up under colonial rule. In relation to democratic consolidation it has been argued that there is an important difference between so-called "settler" colonies (many former British colonies) and "extraction" colonies (many former French, Portuguese and Belgian colonies) (Acemoglu et al. 2001, Acemoglu and Robinson 2006, Norris 2008). Whereas in settler colonies colonial powers often built up more extensive institutions to govern the colony, institutions that remained after decolonization, in extraction colonies the institutional infrastructure remained confined to what was needed to optimize production and transport of resources, leaving very little behind in terms of broader government institutions at decolonization⁹⁰.

Finally, in addition to the self-reinforcing dynamic of election quality and historical legacies, events such as civil war or a coup d'état, that interrupt the normal functioning of politics and thereby provide a chance to change the rules of the game, for good or for worse, are also likely to influence election

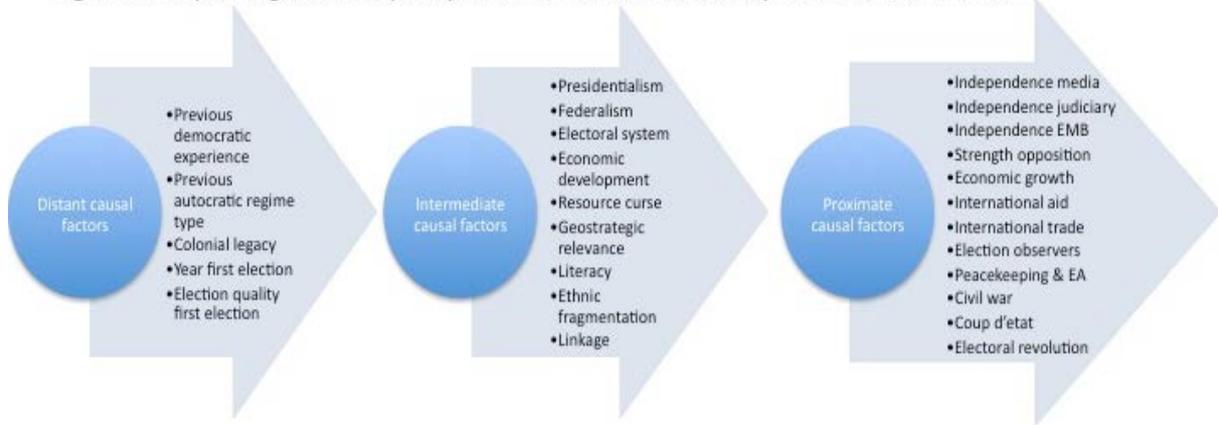
⁹⁰ The debate is still ongoing whether these differences are due to natural conditions in former colonies that either stimulated or discouraged settlement or rather due to the specific characteristics of the colonizing powers. As such, it has often been argued that former British colonies tend to have achieved higher levels of democratic quality compared to former French, Spanish and Portuguese colonies, due to the development of democratic political institutions, the traditions of British common law, and the development of a 'democratic culture' under British rule. In our analyses we will test whether differences in colonial origin affect election quality as well as "settler mortality" an indicator that measures the rate of mortality of settlers in colonies, and that is found to explain well the difference between settler and extraction colonies.

quality. In the case of civil war the effect on election quality is likely to be negative, since not only does civil war negatively affect respect for civil and political liberties and rule of law, but it also leaves a highly polarized citizenry and political arena after the conflict that might be unwilling to accept defeat in elections, creating a zero-sum electoral race that increases the incentives for electoral fraud. In principle, we could expect a similar negative effect of coup d'état on election quality, since political actors that have gained power in an undemocratic way might be more inclined to engage in electoral fraud if they subsequently organize elections. While there are certainly many examples of coup-leaders subsequently 'winning' flawed elections, there are also examples of coup-leaders that actually intervened as a protest against flawed elections or out of dissatisfaction with the rate of democratic reforms. Examples of the latter are the military coup d'état in Niger in 1999 or in Portugal in 1974, both of which resulted in higher quality elections after the coup than before. Hence the effect of coups on election quality might be less straightforward. Finally, a specific event that has become more common in recent years are so-called "electoral revolutions", i.e. the mobilization of citizens in protest against flawed elections that occurred in Serbia in 2000, Georgia in 2003, Ukraine in 2004 and Kyrgyzstan in 2005, that may have been successful in improving the quality of subsequent elections.

Summarizing, election quality is likely to be driven by institutional/political, economic, socio-cultural, international and historical factors shaping the incentives and disincentives to engage in electoral fraud. However, before proceeding to the empirical analyses, it needs to be noted that not all these causal factors are likely to be equally closely related to election quality. For example, whereas different historical experiences might make certain countries more likely to experience lower levels of election quality than other countries, these aspects are likely to be "distant" causes, that themselves affect other aspects of these polities that are more "proximate" causes of election quality. For example, countries with previous democratic experience might have greater levels of media freedom and a more independent judiciary than countries with no previous experience with democracy, which in turn

enhances election quality. Hence, the effect of distant causal factors like previous democratic experience might work through more proximate causal factors such as media freedom and judicial independence, in which case more distant causal factors are likely to turn out as insignificant once more proximate causal factors are included in our model⁹¹. The figure below organizes the causal factors identified in terms of their “distant”, “intermediate” and “proximate” effect on election quality. One could also think of the factors organized here as stable versus time-changing causal factors: whereas the distant factors indicate non-changing variables, the intermediate factors indicate slowly changing variables, and the proximate factors are variables that change faster and can potentially change between each election. Organizing the independent variables this way also helps to think through which variables are likely to explain differences in election quality between countries (i.e. the non-changing or slower changing variables) and which might explain differences in election quality over time within countries (the faster changing variables).

Figure 4.1 Explaining election quality: distant, intermediate, and proximate causal factors



⁹¹ Of course, if the association is too strong we will not be able to include all factors in the model as multi-collinearity will bias our estimates, but we have carefully checked for this in all our models and associations between independent variables were never sufficiently strong as to cause multi-collinearity.

4.2. Hypotheses, data and methods

Having formulated our theoretical expectations of what drives election quality I turn to presenting the hypotheses as well as discussing the data and methods used to test the hypotheses. The table below provides an overview of the explanatory factors discussed in the previous section along with the direction of the hypothesized effect. In the subsequent analyses, these are the hypotheses that will be tested. Note that for a number of variables, the values tested are indicated as lagged values ($t-2$), indicating that these variables will be measured 2 years before the elections. This is to prevent endogeneity or reversed causality. For example, an increase in international aid in the year of the election might be the result of a clean election, rather than a cause of it. Taking the level of international aid some time before the elections as an indicator helps to diminish such reverse causality⁹².

The time lag of 2 years is chosen to also reduce the likelihood of potential anticipation effects. For example, if elections will take place in the following year, and are expected to be fraudulent or problematic, foreign investors might already retreat in anticipation and hence international trade might drop before elections, as a consequence of the anticipated low quality of the electoral process, rather than causing such lower quality (and in fact I expect trade to have a positive effect). Finally, a time lag of 2 years also reduces the likelihood of temporal overlap between the measurement of independent variables and the electoral process, which, counting the campaign period prior to elections as well as adjudication of complaints and the final publication of results, can take up to anywhere between 3 to 6 months, and sometimes even longer (Lindberg 2006a).

⁹² It does not entirely exclude the possibility of endogeneity however, as the election quality of the previous election could be driving both the level of aid and the election quality of the current election. For this reason, I also test whether the results hold when including a lagged dependent variable in the models.

Table 4.2. Overview of hypotheses

Causal factors	Expected effect on election quality
Institutional factors	
Presidentialism	Negative
Federalism	Positive
Electoral system majoritarian	Negative
Strength incumbent (lagged previous election)	Negative
Strength opposition (lagged previous election)	Positive
Fragmentation opposition (lagged previous election)	Curvilinear
Independence judiciary	Positive
Independence EMB	Positive
Independence media	Positive
Economic factors	
Economic development (t-2)	Positive
Economic growth (t-2)	Positive
Economic inequality	Negative
International trade (t-2)	Positive
Resource curse (t-2)	Negative
Socio-cultural factors	
Ethno-linguistic fractionalization	Negative
Religious fractionalization	Negative
Literacy/education (t-2)	Positive
Active civil society: domestic election monitors	Positive
Active civil society: electoral 'revolution'	Positive
International factors	
Linkage (economic, social, political) (t-2)	Positive
Aid dependence (t-2)	Positive
International election monitors	Positive
Election assistance	Positive
Peacekeeping operations	Positive
Historical factors	
Previous democratic experience	Positive
Previous autocratic experience (personalist & single party)	Negative
Colonial legacy (extraction)	Negative
Civil war (t-2)	Negative
Coup d'etat (t-2)	-
Year founding election	Negative
Quality founding election	Positive
Control variables	
Election type	-
Election sequence number	Positive
Election quality previous election	Positive
Reliability election quality score	-
Region	-

Turning to the data used for these analyses, these derive partly from the election quality dataset as presented in chapter 2, and partly from the Quality of Government time-series cross-sectional dataset (Teorell et al. 2011). The data for the dependent variable election quality were partly merged from different sources and partly collected by the author, as described in chapter 2. Independent variables such as previous democratic experience, civil war, coup d'etat, and electoral revolutions, as well as data

on opposition and incumbent vote and seat shares were gathered by the author on the basis of election data handbooks and historical sources. The data for the other independent variables are taken from the Quality of Government dataset (version April 6th 2011) and some additional data was collected from sources such as IDEA, the UN, the World Bank's Development Indicators, and Bjornlund (2004). Appendix D gives an overview of the independent variables and information on coding procedures.

Turning to statistical techniques, since the sample includes 97 third wave regimes with multiple elections held at different time-points within each regime, this allows us to study variation in election quality between countries as well as over time. The advantage of between country comparisons is that it allows us to assess the effects of factors that do not change over time, or at least do not change rapidly, such as political institutions or social structure. The disadvantage is however that, for variables that do change over time, it is more difficult to assess the direction of causality in such analyses: since we aggregate to the country-level, a statistically significant association between economic development and election quality might indicate that economic development is driving higher election quality, or the other way around.

Over-time analyses within the same country allow for a more precise assessment of causality: does economic development in the year(s) before an election lead to a higher quality election (correcting for other time-changing independent variables)? The power of such analyses is that it holds non-changing variables constant (by virtue of the within-country analyses, 'fixing' the country-level variation), allowing for a 'ceteris paribus' test of the effect of variables that do change over time. However, the bottleneck of these analyses is that inferences are based on the number of observations within each country, and hence the number of time-points (T) needs to be large enough to allow for sufficient degrees of freedom to draw statistical inferences. Moreover, by holding non-changing variables constant, the effects of such variables are corrected for, not tested.

Ideally, one would have data on many countries and on many elections within each country, in order to be able to both model the average level of election quality for each country as well as the over-time development in election quality within countries. Time-series cross-sectional analyses (TSCS) typically allow for such a combination of ‘between-effects’ (country-level effects) and ‘within-effects’ (over-time effects) analyses. However, in the case of third wave regimes it is questionable to what degree TSCS models will provide reliable estimates, for a number of reasons. First of all, T is typically limited and varying across countries, resulting in unreliable estimates for countries with small T as well as highly unbalanced panels generating heteroskedastic errors. Moreover, since the number of elections held in any third wave regime is generally not arbitrary but related to the level of democratic consolidation (i.e. more stable democracies will have held more elections), excluding those countries with limited observations in T to create a sample with a more balanced panel-structure will lead to a biased sample. In addition, whereas in more common political economy TSCS data, time observations will be successive years and all countries will have been observed over the same period, in our case time observations are election years, which are not the same in all countries, and moreover, are held at different intervals in different countries (the reasons of which are often not arbitrary either, such as when elections have not been held for several years due to civil war). Finally, even if we were to solve these issues, T would remain rather low for a substantial number of countries in our sample, especially the countries of the so-called ‘fourth’ wave of democratization.

Table 4.3. below shows the number of legislative, executive and total elections per country in our sample. It seems that the median T is between 4 to 6 legislative elections (with a long tail of countries that have held more legislative elections) and about 3 to 5 executive elections (with a group of countries that have held only one executive election, or none in the case of the 14 parliamentary systems in our sample). When counting legislative and executive elections as forming part of a single

'sequence' of elections, the median lies between 5 and 9 elections, with long tails to both sides of the distribution⁹³.

Table 4.3. Number of elections per country 1974-2009

Number of elections	Legislative elections	Executive elections	Total
1	1	10	1
2	9	3	3
3	7	17	6
4	25	21	9
5	19	13	10
6	15	8	11
7	4	4	15
8	4	2	13
9	3	4	10
10	2	1	7
11	2		4
12	3		3
13	1		-
14	1		-
15	1		1
16			1
18			2
20			1
Total	97	83	97

Table 4.4. Average number of elections per country per region 1974-2009

Region	Average number of elections per country T	Number of countries N
CEE	6.71 (7)	17
FSR/CIS	8.17 (8)	12
Africa	5.87 (6)	45
SA	9.83 (10)	12
CA	9.63 (10)	8
SE	14 (14)	3
Total	7.35	97

An alternative way of modeling data with this structure as proposed by Beck (2008) is to think of the data as having a hierarchical structure, with elections (level 1) being nested within countries (level 2). Time-changing variables, i.e. variables that take on different values with each election, can then be considered as level 1 variables, and non-changing variables, i.e. variables that are the same for each

⁹³ Note that for this overall sequence, concurrent elections receive the same sequence number, so they are not counted double.

election within a country, as level 2 variables. This would also fit nicely with the notion of ‘distant’ and ‘intermediate’ causal factors on the one hand and ‘proximate’ causal factors on the other hand. Distant and intermediate causal factors (i.e. non-changing and slowly changing variables) could be thought of as the country-level variables that explain the ‘structural’ propensity of a country to have high or low levels of election quality, while proximate causal factors are the election-level variables that explain the election quality of particular elections, once the structural factors have already been accounted for, i.e. why election quality is higher or lower than the ‘base-level’ or ‘structural’ election quality that one would expect on the basis of the country-level variables. Of course, also with this modeling approach one needs to test for serial autocorrelation and heteroskedastic errors, but the approach seems to fit our theoretical set-up better. In the next section we will discuss the results of the analyses considering distant variables first, followed by intermediate factors and proximate factors. To make sure our results are robust, we will first carry out the proposed analysis using the full sample including both executive and legislative elections, and then test whether the results hold in TSCS analyses of legislative and executive elections separately.

4.3. Results

Table 4.5 below shows the results of the statistical analyses. The models presented are based on the full sample of elections for which election quality was available from 1974 to 2009. This results in a total of 836 elections in 90 countries⁹⁴. The results were re-estimated for samples limited to include only countries that had held minimally 5 elections, and limited to those elections with reliable election

⁹⁴ For 7 countries, data were missing that could not be imputed. For Serbia and Montenegro (until 2005), Montenegro (since 2006) and Serbia (since 2006) data were missing for many independent variables. For Belize, Seychelles, Sao Tome and Principe, and Suriname data were missing concerning independence of the judiciary. Since these data were missing for the entire time-period in which these countries are included in our sample, I chose not to impute these values but leave these countries missing from the analyses. Appendix D provides an overview of all independent variables, the sources from which they were gathered as well as minimum and maximum values, and which data are missing.

quality scores⁹⁵. The results for these models were substantively the same as the ones presented below⁹⁶. I present the results again in groups of variables (i.e. institutional/political, economic, socio-cultural, international and historical factors), to maintain a clear overview. Apart from our expectations concerning the effects of independent variables, as specified in the hypotheses in table 4.2., we also expect “distant” causal factors to explain less variance in election quality than “proximate” causal factors. Moreover, since the number of elections per country varies considerably, between 1 and 20, we expect to perform less well in explaining level 1 (election-level) variation than explaining level 2 (country-level) variation. These expectations are in fact born out by the results of our analyses. First of all, the null-model indicates the proportion of variance unexplained at the country-level (level 2) and the election-level (level 1). As expected, 56% of unexplained variance in election quality is country-level variance, whereas 44% is election-level variance. Presumably, if we would have longer election-sequences, the amount of variance to be explained at election-level would increase. Comparing the decrease in unexplained variance as we add explanatory variables to the null-model gives us an indication of how much variation the models explain. In the table I indicate for each model how much variance is explained at level 1 (elections) and at level 2 (countries). Clearly, the model is rather successful at explaining country-level variance in election quality, as the full model (model 6) leaves only about 18% of variance unexplained, which amounts to 68% explained variance at the country-level. However, the model is much less successful at explaining variance within countries over time, as the full model (model 6) leaves 40% of variance unexplained, amounting to only 9% explained variance at the election-level. Including all control variables improves the model fit, as the last model in table 4.6 will show, to an overall 61% of variation explained.

⁹⁵ Elections with reliable election quality scores are elections that were coded more than once, or coded by the author on the basis of reliable information (excluding scores based on little or conflicting information) (see chapter 2).

⁹⁶ Only the variables “education” and “peacekeeping” appear to be less robust, as education becomes insignificant in one of the models, while peacekeeping becomes significant in one of the models. Otherwise the results were substantively the same as those presented in table 4.5.

For now, turning to our expectations regarding the effects of independent variables on election quality, table 4.5 shows the results for each group of variables. Starting with historical factors, previous democratic experience enhances election quality significantly, confirming our hypotheses in this regard. The results for previous autocratic regime type and colonial legacy are also in line with what we expected, with former personalist and single-party regimes showing on average lower election quality than military regimes and French and Belgian former colonies having lower levels of election quality than countries with different or no colonial legacies⁹⁷. However, both variables have p-values just above what would be needed to reach one-tailed statistical significance. Historical events such as civil war and coup d'etat are significantly related to election quality. If a country experienced civil war in the 2 years prior to elections, election quality is about 10% lower, a dynamic that was also shown in the descriptive tables in chapter 3. The effect of a coup d'etat taking place in the 2 years prior to elections is less intuitive: on average, coup d'etats tend to be followed by higher quality elections. I checked whether this might be due to regional differences, however this appears not to be the case: the coefficient is positive in all regions in our sample. Finally, the timing of the first de jure multi-party and universal suffrage election is negative as expected, later founding elections being associated with lower election quality, but not significantly so.

Turning to institutional variables (model 2), as expected presidentialism and majoritarian electoral systems are negatively associated with election quality. Clearly, zero-sum races increase the odds of electoral fraud: in presidential systems election quality is about 15% lower than in parliamentary systems, and in majoritarian electoral systems elections quality is consistently lower as well. However, the effect of electoral system appears to be stronger for legislative elections than for presidential

⁹⁷ It seems likely that the result found for colonial legacy is in fact due to the difference between extraction and settler colonies, as the majority of French and Belgian colonies were extraction colonies. Indeed similar results were found when using settler mortality as an indicator, but since this indicator has slightly more missing data I preferred to use colonial origin in the final analyses.

elections. This is not so surprising though, as the electoral system variable for legislative elections compares majoritarian electoral systems to all other electoral systems, including proportional systems, while the electoral system variable for presidential elections compares absolute majority systems to plurality systems, which differ less from each other than the various legislative electoral systems. Turning to federalism, while the coefficient appears to be in the expected direction, with federal systems having higher quality elections, the effect is not significant (and remains so under various different specifications). How about institutional factors influencing the detection costs of electoral fraud? Here we find that, as expected, costs of detection are strongly and significantly related to election quality. Media independence 2 years before elections increases election quality with more than 20% compared to countries where the media are under complete government censorship. Independence of the judiciary is also important, though its effect on election quality is less pronounced than independence of the media. Finally, the only surprising result in these analyses is the effect of EMB independence on election quality. Following Hartlyn et al. (2008), I expected independent electoral management bodies to have a significant positive effect on election quality, however, in our analyses we find a consistently insignificant (and negative) effect. The reason for this finding is probably that the variable measures EMB independence in 2006 (IDEA - Wall et al. 2006). Not only does this generate static data, it also leads to a variable that is measured after most of the elections in our sample took place, making assessing the effect of EMB independence on election quality difficult. Moreover, in recent years many third wave regimes have moved towards greater (formal) EMB independence, with the result that the majority of countries in our sample have formally independent EMBs, while those countries that still have EMBs that are not or only partly independent, are the regimes that held on average higher quality elections and consequently experienced less need for reform of EMBs (i.e. Greece, Spain, Portugal, Czech Republic, Hungary). The group of institutional/political variables also included opposition strength and fractionalization,

however since these are measured as lagged variables, all first elections are lost when including these variables in the analyses, and hence their effects are tested at a later stage (cf. table 4.6.).

Now, considering socio-cultural variables (model 3), ethnic and religious fractionalization indeed appear to have a strong negative effect on election quality. Surprisingly however, given the literature on ethnic fractionalization and democratization, ethnic fractionalization does not seem to significantly affect election quality, even when testing different specifications and excluding possible outliers. Religious fractionalization does appear to have a significant and negative effect on election quality. However, this result only holds when correcting for the proportion of muslim population in a country. This 'suppressor' effect seems to be due to the fact that in countries with a high proportion of muslims (and hence low religious heterogeneity), election quality is on average lower, obfuscating the effect of religious fractionalization. Including both clarifies their effects on election quality. The effects of proportions of Catholics and Protestants were also tested but had no significant effects, and were hence left out of the model. Note that the size of the effects of religious fractionalization and proportion of Muslims are approximately equally large, as the proportion of Muslims is measured on a scale from 0 to 100. In addition to religious fragmentation, literacy also appears to have the expected positive association with election quality. Finally, an active civil society as expressed in domestic monitoring of elections and electoral revolutions as protests against electoral fraud appear to enhance election quality, as expected. Though the p-value for domestic observers is close to one-tailed significance, electoral revolutions do not seem to have a significant effect on election quality in the years after the revolutions, which is line with the recent literature questioning whether these revolutions lead to sustained effects on democratization and election quality (Kalandadze and Orenstein 2009).

Turning to economic variables (model 4), economic development has the expected effect on electoral fraud: higher GDP per capita tends to be associated with higher election quality. This effect is quite strong as well, moving from the lowest per capita income to the highest per capita income in our sample is associated with an increase in election quality of 50%. The effect of economic growth is not as expected however, I expected higher economic growth to increase incumbent popularity and thereby decrease the incentives to engage in electoral fraud, however the effect seems to be insignificant (and reversely signed). In addition to economic development and growth, the effect of economic inequality (gini-coefficient) was also tested. However, economic inequality appeared consistently insignificant in relation to electoral fraud, which is probably due to a combination of low-quality data and a high number of missing values, and was hence left out of the model. Turning to trade, we expected countries with large volumes of exports to and imports from advanced industrial democracies to experience relatively stronger pressures to hold clean elections. To measure this, I used data on the exports and imports to high-income economies, all of which (except Singapore) are also established democracies. As table 4.5 shows however, while both were indeed positively related to election quality, only imports turned out to be significant. This is not so surprising as dependence on imports from high-income economies probably gives those countries more leverage over the recipients of such imports, rendering international legitimacy more valuable for the latter regimes. Finally, considering the exports of natural resources, in third wave regimes with extensive natural resources (such as Azerbaijan, Congo and Nigeria but also Venezuela and Bolivia), election quality is significantly lower, about 20%, than in countries with no such resource ‘curse’⁹⁸.

⁹⁸ The non-significant effect of exports to high-income economies could be due to the fact that the “resource curse” is also measured as a proportion of total exports. However, exports to high-income economies refer to merchandise exports and the two variables are not correlated, so this seems unlikely to have affected the results.

Table 4.5 Explaining election quality in third wave regimes 1974-2009

	Null-model	Model 1 Historical	Model 2 Institutional	Model 3 Socio-cultural	Model 4 Economic	Model 5 International
Previous democratic experience (0-1)	0.143*					
Previous autocratic regime – single party & personalist (0-1)		(0.059) -0.084 (0.052)				
Colonial legacy – French & Belgian (0-1)		(0.065) -0.101 (0.065)				
Civil war: occurred in 2 years before elections? (0-1)		(0.033) -0.102**				
Coup d'etat: occurred in 2 years before elections? (0-1)		(0.046) 0.098*				
Year first election (1974-2006)		(0.004) -0.004 (0.004)				
Presidentialism (0-1)				-0.147** (0.049)		
Electoral system majoritarian legislative (0-1)				-0.103*** (0.028)		
Electoral system absolute majority executive (0-1)				-0.040+ (0.021)		
Federalism (0-1)				0.047 (0.054)		
Independence media t-2 (0-1)				0.218*** (0.031)		
Independence judiciary t-2 (0-1)				0.129*** (0.027)		
Independence EMB (0-1)				-0.040 (0.050)		
Ethnic fractionalization (0.02-0.88)					-0.010 (0.141)	
Religious fractionalization (0.01-0.86)					-0.373** (0.138)	
Proportion of muslims (0-99.7)					-0.004*** (0.001)	
Domestic observers (0-1)					0.040 (0.027)	
Electoral revolution (0-1)					0.063 (0.093)	

	Null-model	Model 1 Historical	Model 2 Institutional	Model 3 Socio-cultural	Model 4 Economic	Model 5 International
Average years of education men & women 25+ t-2 (0.25-13.2)				0.017** (0.006)		
Economic development - GDP per capita t-2 (118-26193)					0.000019*** (0.000)	
Economic growth - % change GDP/cap 2 yrs before elections (-22.37 - 55.91)					-0.001 (0.002)	
Exports to high income economies as % of total merchandise t-2 (0 - 99.75)					0.000 (0.001)	
Imports from high income economies as % of total merchandise t-2 (0 - 96.01)					0.003** (0.001)	
Resource curse - Fuel, ores, metals exports as % of total exports t-2 (0 - 99.70)					-0.002** (0.001)	
Import missing data (0-1)					0.202* (0.081)	
Resource curse missing data (0-1)					-0.073* (0.030)	
Linkage (economic, social, political globalization) t-2 (19.8-85.3)						0.005*** (0.001)
Aid received as % of GDP t-2 (0-141.8)						-0.000 (0.001)
International observers (0-1)						0.031 (0.024)
Election assistance (0-1)						0.046 (0.032)
Peacekeeping mission (0-1)						-0.037 (0.043)
Constant	0.602*** (0.029)	9.303 (7.886)	0.634*** (0.064)	0.754*** (0.085)	0.407*** (0.060)	0.364*** (0.056)
LL	-68.56	-67.24***	-39.56***	-63.28***	-83.11***	-78.14***
N level 1	836	836	836	836	836	836
N level 2	90	90	90	90	90	90
Variance explained at level 1 (elections)	0% (44%)	1%	2%	1%	1%	0%
Variance explained at level 2 (countries)	0% (56%)	12%	29%	15%	15%	11%
Total variance explained	0%	13%	31%	16%	16%	11%

Standard errors in parentheses. P-values: ^ 0.1, * 0.05, ** 0.01, *** 0.001.

The final model in table 4.5 presents the effects of international factors (model 5). The effect of linkage appears to be positive and in the expected direction: countries that are more strongly globalized (economically, socially and politically) hold higher quality elections. The effects of aid are less straightforward, while I expected a positive effect on election quality, the coefficient is negative and clearly insignificant. This might be due to a selection effect, as many countries in our sample that have higher levels of economic development and receive less aid, are also those countries where democracy is relatively more consolidated and election quality is on average higher, potentially blurring the effect of aid (i.e. Southern Europe, Latin America). A similar selection effect seems to occur with respect to peacekeeping missions, that appear to be associated with lower levels of election quality, while it is of course more likely that missions take place in fragile post-conflict situations, where election quality is likely to be lower as well. The presence of international observers and election assistance however seems to influence election quality positively as expected. These variables, though not significant, are close to one-tailed significance and will hence be included in the overall model.

What happens if all explanatory variables are combined in a full model? Table 4.6 presents these results. Only those variables that were significant in the analyses in table 4.5 are included, with the exception of the variables that were very close to statistical significance: previous autocratic regime type, colonial legacy, domestic and international observers and election assistance. Clearly, distant and intermediate factors loose in strength once proximate causal factors are added, but several nevertheless remain significant. As such, of the distant causal factors, previous democratic experience remains significant, though the coefficient size is cut in half. As regards intermediate causal factors, presidentialism, the electoral system for legislative elections, religious fractionalization as well as imports and the resource curse remain significant. Surprisingly, economic development and linkage both drop out, while they were rather strongly associated with election quality in previous models. This appears to be caused by the strong correlation between these two variables, which, though not

generating multi-collinearity, might nevertheless cause both variables to become insignificant in an overall model. Indeed, when including linkage but not economic development in the parsimonious model (model 7), linkage is significant. Finally, the proximate explanatory factors all remain significant in the overall model, with the exception of domestic and international observers that become insignificant⁹⁹.

Now let's see if these results still hold when including the missing political variables, as well as additional controls. Model 8 presents the results when I include incumbent strength, opposition strength and opposition fragmentation in the previous elections as explanatory variables, as well as election quality of the founding elections. All these variables belong to the explanatory variables for which I specified hypotheses in table 4.2., however since inclusion of these variables leads to a sample excluding the first elections, I preferred to add them after estimating the overall model. Incumbent strength is measured as the vote share of the winning candidate (for executive elections) or the seat share of the winning party (for legislative elections) in the previous elections, and opposition strength is measured as the vote of the largest opposition candidate (for executive elections) or the seat share of the largest opposition party (for legislative elections) in the previous elections¹⁰⁰. Since opposition strength refers to the largest opposition candidate or party, this variable is related to incumbent strength but not collinear. Opposition fragmentation was measured by the number of opposition candidates or parties that won over 1% of votes in the previous elections. Since I expect a curvilinear

⁹⁹ It is not unthinkable that this result for observers is due to data limitations, as the data for these two variables were derived from Bjornlund (2004), who includes a more limited sample of elections. On the other hand, Birch (2011) doesn't find the presence of international observers to be significantly related to electoral malpractice, even when controlling for selection effects. Due to space limitations, the coefficients for both domestic and international observers are not reported in the table.

¹⁰⁰ This was done because data on vote shares for legislative elections were often missing in the data handbooks used by the author to collect these data. Seat shares were more often available and are hence used for legislative elections. In the next section the analyses are done for legislative and executive elections separately to make sure the overall results are robust.

effect of fragmentation, i.e. dominant party systems as well as highly fragmented party systems leading to lower election quality, this variable was recoded to analyze the effects of moderate multi-party systems versus dominant party systems and extremely fragmented party systems. Election quality of the founding elections is simply the election quality score of the first ‘potentially democratic’ elections in our sample. Since the political variables are lagged with one election, these first elections are dropped from the analyses, and hence including the quality of the first election as an explanatory variable no longer leads to endogeneity. As model 8 in table 4.6 shows, the full model specified in model 7 seems to be robust to the inclusion of these variables, as all variables remain significant, with the exception of imports. Election assistance appears as insignificant in the table but has a p-value very close to one-tailed significance (0.109).

Regarding the effects of the political variables, these are all in the expected direction, with incumbent strength being negatively associated with election quality and opposition strength and moderate multi-party/candidate systems positively associated with election quality. Nevertheless, only incumbent strength and opposition fragmentation appear to be significant, which might be explained by the fact that incumbent strength and the structure of the party system are likely to be more persistent over time, while opposition support as measured in the previous elections might change considerably between one election and the next, rendering our measurement of opposition support less precise. Furthermore, as regards the quality of the first election, this is positive as expected, but not significant. Hence, it seems that the first election is not necessarily a “lock-in” moment that defines the subsequent development of election quality. Perhaps the effect of election quality is rather more persistent over successive elections, each previous election shaping the quality of the next to a certain extent. Model 9 tests this idea by including lagged election quality as an explanatory variable¹⁰¹.

¹⁰¹ Including the lagged dependent variable was not only substantively but also statistically necessary as model 8 appeared to suffer from serial autocorrelation, i.e. correlated errors over time. Including the lagged dependent

Lagged election quality appears to have a large significant effect on election quality, indicating that there is indeed persistence over time. The relation is however certainly not perfect, leaving room for change and additional effects of other variables. In fact, while improving the explained variance of the model to 60% overall, and lowering the coefficient sizes of the other independent variables somewhat, all independent variables remain significant. The only exception is incumbent strength, which is no longer significant when lagged election quality is included. This is not so surprising though, as incumbent strength in the previous election may be partly the result of the election quality of that previous election, and hence including lagged election quality wipes out this effect.

Turning to model 10, I added control variables to test whether the results found are robust to inclusion of these variables. The variables included are: election sequence, election type, constituent assembly elections, concurrent elections, reliability of the election quality score (when the election was coded by the author), the number of election quality scores received, as well as regional dummies. The overall results appear very robust, as all independent variables remain significant even when including these control variables, with the exception of linkage, that becomes insignificant (again, election assistance has a p-value that is as good as significant one-tailed, i.e. 0.104). Turning to the effects of the control variables, the expectation for election sequence was that as a country holds successive elections, election quality should increase (i.e. the “self-reinforcing” power of elections). Note that this is a distinct idea from the idea of persistence in election quality, whereas the latter posits that low election quality in election 1 is associated with low election quality in election 2 and vice versa, the former posits that election quality should increase over time, even when starting off low. The results clearly speak in favor of persistence rather than gradual increase. However, both effects may be

variable solved this. I also tested for heteroskedasticity, and this appeared not to be a problem. Finally, I also tested whether the dependent variable is stationary over time, and this was also not the case, suggesting that using multi-level models to correct for the clustering of elections within countries is appropriate (Beck 2008).

different in different regions¹⁰². Turning to election type, I hypothesized that legislative elections might have higher quality than presidential elections since the latter are higher stake races, and I also hypothesized that concurrent elections might have lower election quality, for the same reason. Moreover, I expected constituent assembly elections to be different from ‘normal’ elections, without a clear expectation about their differences in terms of election quality. Nevertheless, none of these control variables appeared to be significant, with the exception of constituent assembly elections that show worse election quality. Reliability and number of scores also are not significant, and of the 6 regional dummies, only the Former Soviet Republics remain as having significantly lower election quality than the other regions.

Model 12 shows the parsimonious full model, in which linkage (economic, social and political globalization) is again significant in the expected direction. The full model explains 61% of the overall variation in election quality, which amounts to 85% of country-level variation (i.e. 57% explained of 67% of total country-level variation in this sample), and 12% of over-time variation (i.e. 4% explained of 33% of total election-level variation). Clearly, the fact that our sample includes countries with very different numbers of elections renders explaining variation in election quality over time more difficult. The model nevertheless performs rather well in explaining between-country differences. A question that remains to be addressed is whether there are substantive variables missing that could “explain away” the significantly lower election quality found in the Former Soviet Republics (the fact that this regional dummy remains significant indicates we might have missed an explanatory variable of particular importance to that region).

¹⁰² However interaction effects with regional dummies and election sequence or election quality lagged are not significant when included in model 10, so these findings seem to be similar across different regions.

Table 4.6. Explaining election quality in third wave regimes 1974-2009 (continued)

	Model 6 Full model	Model 7 Pars model	Model 8 Pars model political vars	Model 9 Pars model w lagged DV	Model 10 Pars w lag DV and controls	Model 11 Pars model w sign.controls
Previous democratic experience (0-1)	0.066+ (0.038)	0.065+ (0.038)	0.083* (0.039)	0.068* (0.031)	0.060+ (0.033)	0.058+ (0.032)
Previous autocratic regime – single party & personalist (0-1)	-0.046 (0.039)					
Colonial legacy – French & Belgian (0-1)	0.036 (0.054)					
Civil war: occurred in 2 years before elections? (0-1)	-0.081* (0.032)	-0.077* (0.032)	-0.094** (0.035)	-0.088** (0.034)	-0.086* (0.034)	-0.085* (0.034)
Coup d'état: occurred in 2 years before elections? (0-1)	0.142** (0.044)	0.141** (0.044)	0.131** (0.041)	0.126** (0.040)	0.135*** (0.041)	0.135*** (0.041)
Presidentialism (0-1)	-0.137** (0.046)	-0.145** (0.045)	-0.147** (0.049)	-0.111** (0.041)	-0.111* (0.044)	-0.102* (0.043)
Electoral system majoritarian legislative (0-1)	-0.083** (0.028)	-0.071** (0.025)	-0.070** (0.026)	-0.062* (0.024)	-0.069* (0.028)	-0.059* (0.024)
Electoral system absolute majority executive (0-1)	-0.022 (0.020)					
Independence media t-2 (0-1)	0.180*** (0.030)	0.182*** (0.030)	0.164*** (0.030)	0.121*** (0.030)	0.119*** (0.030)	0.114*** (0.030)
Independence judiciary t-2 (0-1)	0.112*** (0.027)	0.112*** (0.027)	0.105*** (0.027)	0.084*** (0.025)	0.085** (0.027)	0.081** (0.026)
Religious fractionalization (0.01-0.86)	-0.240** (0.090)	-0.289*** (0.084)	-0.379*** (0.086)	-0.301*** (0.068)	-0.252** (0.084)	-0.260** (0.081)
Proportion of muslims (0-99.7)	-0.002** (0.001)	-0.003*** (0.001)	-0.002** (0.001)	-0.002** (0.001)	-0.001* (0.001)	-0.001* (0.001)
Average years of education men & women 25+ t-2 (0.25-13.2)	-0.000 (0.007)					
Economic development - GDP per capita t-2 (118-26.193)	0.000 (0.000)					
Imports from high income economies as % of total merchandise t-2 (0 - 96.01)	0.001** (0.000)	0.001** (0.000)	-0.000 (0.001)			
Resource curse - Fuel, ores, metals exports as % of total exports t-2 (0 – 99.70)	-0.001** (0.000)	-0.001** (0.000)	-0.002** (0.000)	-0.001** (0.000)	-0.001** (0.000)	-0.001** (0.000)
Linkage (economic, social, political globalization) t-2 (19.8-85.3)	0.002 (0.001)	0.002* (0.001)	0.003** (0.001)	0.002* (0.001)	0.001 (0.001)	0.002* (0.001)

	Model 6 Full model	Model 7 Pars model	Model 8 Pars model political vars	Model 9 Pars model w lagged DV	Model 10 Pars w lag DV and controls	Model 11 Pars model w sign.controls
Election assistance (0-1)	0.061* (0.030)	0.066** (0.025)	0.042 (0.026)	0.047+ (0.026)	0.044 (0.027)	0.042+ (0.026)
Election quality founding election (0-1)			0.070 (0.050)			
Strength incumbent (lagged previous election) (5.5-100)			-0.001+ (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Strength opposition (lagged previous election) (0-66.67)			0.000 (0.001)			
Fractionalization opposition (dummy moderate fractionalization vs low and high fractionalization)			0.048* (0.020)	0.044* (0.020)	0.041* (0.020)	0.042* (0.020)
Election quality ((lagged previous election) (0-1)				0.259*** (0.035)	0.243*** (0.035)	0.246*** (0.035)
Election sequence number (1-20)					0.005 (0.004)	
Election type - executive (0-1)					-0.013 (0.018)	
Constituent assembly election (0-1)					-0.118+ (0.067)	-0.133* (0.065)
Concurrent election (0-1)					0.025 (0.025)	
Reliability election quality score (0-2)					-0.009 (0.021)	
Number of election quality scores received (1-5)					-0.001 (0.009)	
Regional dummy (FSR)					-0.161** (0.060)	-0.152** (0.058)
Constant	0.603*** (0.091)	0.607*** (0.087)	0.667*** (0.115)	0.532*** (0.092)	0.620*** (0.108)	0.578*** (0.101)
LL	-55.53***	-29.82***	40.13***	74.96***	55.48***	70.7***
N level 1	836	836	667	667	667	667
N level 2	90	90	89	89	89	89
Variance explained at level 1 (elections)	4%	4%	3%	4%	4%	4%
Variance explained at level 2 (countries)	38%	38%	48%	56%	57%	57%
Total variance explained	42%	42%	51%	60%	61%	61%

Standard errors in parentheses. P-values: ^ 0.1, * 0.05, ** 0.01, *** 0.001.

Finally, these results were double-checked in a sample including only countries that had held at least 5 elections, and in a sample including only the elections that received multiple election quality scores or were scored by the author. No substantive differences were found, except for the fact that in both restricted samples, election assistance is no longer significant, and in the model with at least 5 elections per country previous democratic experience and opposition fragmentation are no longer significant. Also, we re-ran these analyses in time-series-cross-sectional models with only legislative and only executive elections respectively. In the models with these smaller samples, again previous democratic experience and opposition fragmentation no longer appear as significant, neither do presidentialism and economic linkage. Moreover, while past experience with civil war and incumbent strength appear to be only significant in predicting election quality in executive elections, election assistance is only significant in predicting election quality in legislative elections. Robust effects on election quality across legislative and executive elections are the positive effects of independence of the media, independence of the judiciary in the year prior to elections, and election quality in the previous election, as well as the positive effect of a coup d'état having taken place in the 2 years before the election. The negative effects of religious fractionalization and the resource curse also appear robust in these smaller samples, as does the consistently lower level of election quality in former Soviet Republics.

4.4. Conclusion

Viewing these results in light of our discussion of incentives and disincentives to engage in electoral fraud, I expected costs of electoral fraud to affect variation in electoral fraud more strongly than benefits, as the latter presumably speak in favor of electoral fraud under all circumstances. This does not entirely seem to be born out by the results however. To be sure, independence of the media and judiciary are very important factors increasing the chances of detection of electoral fraud, and thereby the costs to engage in electoral fraud, as do factors as linkage and trade that increase the potential costs

of losing international legitimacy if electoral fraud is detected. However, factors that could lower direct costs of electoral fraud such as economic inequality and economic development do not seem to be related as strongly to election quality (at least not in aggregate-level analysis as carried out here). Instead, explanatory variables relating to the stakes of the electoral race appear to be very important in explaining election quality as well. As such, a recent history of civil war negatively affects election quality, as do institutions that increase the zero-sum nature of the electoral race such as presidentialism and majoritarian electoral systems. Social fractionalization appears to increase the zero-sum nature of elections as well, with especially religious fragmentation leading to lower election quality. Also, factors that increase the benefits of winning in elections, such as extensive natural resources, add to the probability of electoral fraud occurring. Moreover, historical experiences and events appear to have some effect on election quality as well: previous democratic experiences tend to lead to higher quality elections, and, a surprising finding, coup d'états tend to lead to higher quality elections a posteriori.

Finally, while the models presented here confirm part of our hypotheses, it also seems the model is better at explaining variation in election quality between countries than at explaining variation within countries over time. Clearly, a step for future research would be to further think through which omitted variables might need to be added to the model to improve specification of over-time developments in election quality. Finally, it might also be the case that part of the variation over time in election quality is in fact idiosyncratic as it is shaped by events that are specific to each country. The fact that even when including the lagged dependent variable, there is still 88% of unexplained variance at the election-level (accounting for 29% of overall variance in election quality) is telling in this respect. Such idiosyncrasy, or unpredictability might be higher for democracies that just started holding elections (i.e. with shorter sequences of elections held), or for very low quality democracies (or rather electoral autocracies).

5. Does election quality matter? Consequences for accountability

5.1. Elections as instruments of democracy

In this chapter we link the quality of the electoral process to the quality of the electoral regime, asking to what extent electoral fraud affects the functioning of elections as ‘instruments of democracy’. Does electoral fraud indeed allow governments to avoid being held accountable for past actions? And are governments that came to power through fraudulent elections indeed less responsive to citizens’ needs once the elections are over?

As briefly discussed in chapter 1, addressing this question requires a separate conception of the quality of the electoral process and the quality of the electoral regime. While the quality of the electoral process refers to the actual election, ranging from pre-election campaigning, to election-day voting and vote tabulation, to post-election day adjudication of electoral disputes (Elklit and Reynolds 2005a), the quality of the electoral regime refers to a more functional conception of elections as ‘instruments of democracy’ (Powell 2000). Such a conception derives from the notion that in representative democracy, the alternative for direct rule by the people is rule for the people by elected representatives (Dahl 1989). Hence elections function as a linkage mechanism between citizens and representatives, giving citizens the power to choose representatives to govern in their stead (Miller and Stokes 1963; Thomassen 1994). Ideally, elections should give citizens the possibility to choose future governments and to signal their policy preferences to those governments. In practice, choosing a future government also implies that citizens can evaluate the past government and hold it accountable for its past performance. Hence, elections are opportunities for citizens to (a) render a verdict on the past performance of their government, (b) select a new government, and (c) establish guidelines for future government actions (Powell 2000; Van der Eijk and Franklin 2009). While elements (a) and (b) refer to citizen influence on government formation, element (c) also implies citizen influence on

government policies. The core assumption of elections as instruments of democracy is that citizen influence on government formation generates citizen influence on government policies, as the power of citizens to replace their representatives at elections, i.e. “throwing the rascals out”, creates incentives for representatives to be responsive to citizens’ needs (Powell 2000; Przeworski, Stokes and Manin 1999, Strøm, Müller and Bergman 2003). Hence, elections function as ‘instruments of democracy’ if they generate both accountability and responsiveness.

Now, the degree to which elections indeed fulfill these functions is likely to depend, not only on the quality of the electoral process, but also on the broader institutional and socio-economic setting in which elections take place. To clarify the other factors that we expect to be important in generating accountability and responsiveness, it is useful to examine the linkage mechanism of elections a bit further. Following Powell (2000, 2004), the linkage that elections generate between citizens and governments can be represented schematically as a sequence of linkages that connect preferences to votes, votes to governments and governments to policy. The flowchart below shows this sequence of linkages, and demonstrates the institutional aspects that influence the quality of each linkage: preferences of citizens are translated into votes (depending on the quality of elections and the choices the party system offers), votes are translated into governments (depending on the electoral rules and rules/practices of government formation), and governments subsequently implement policies (depending on the institutional division of power and state capacity). The conjunction of linkages is also referred to as the ‘electoral regime’ (Van der Eijk and Franklin 2009)¹⁰³.

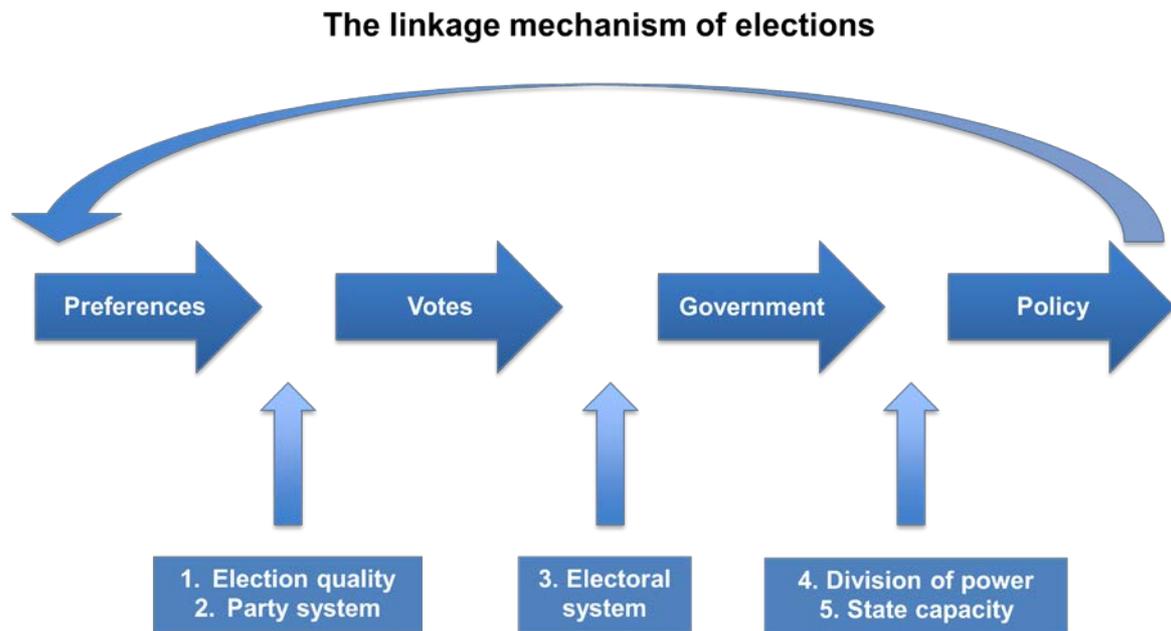
¹⁰³ Clearly, this schematic representation over-simplifies the electoral regime and the more detailed dynamics that take place within each ‘link’. Just as the quality of elections is a complex concept that merits detailed study in its own right, so are the structure of the party system and its interactions with voter preferences and electoral rules; the electoral system and its effect on the aggregation of votes into election outcomes and parliamentary seats; the process of government formation, shaped by the electoral system and the relative size and ideological differences between parties; and the process of policy-implementation, shaped by the institutional structure of the political system, most notably the division or concentration of power, the state’s bureaucratic and economic

As the flowchart below shows, the quality of the electoral regime is likely to depend not only on the quality of the electoral process, but also on the degree to which citizens are presented with meaningful choices in elections, and the degree to which such choices affect subsequent government formation and policy-making. The extent to which citizens are presented with meaningful choices in elections depends on the structure of political supply, i.e. the parties and/or candidates that run in elections and the policy choices they offer to citizens¹⁰⁴. Assuming that voters have divergent interests and hence divergent political preferences, what is important for voters to have a meaningful choice in elections is not only that there is more than one candidate or party running in the elections, but also that the different parties/candidates present different policy platforms, and that the relative power balance between parties/candidates is such that there is more than one viable governing alternative (Wessels and Schmitt 2008, Anderson 2000).

capacity, etc. Moreover, while schematically represented as a linear process, there are many feedback loops, either temporally or due to anticipation. First of all, over time policy implementation is likely to feed back to citizens' preferences, as research on dynamic representation indeed demonstrates, which should then be reflected in the policy positions of parties running in elections as well. In addition, strategic actors are likely to anticipate for example the electoral rules or the degree to which elections are consequential for government formation and adapt their behavior accordingly. Apart from internal dynamics, feedback loops and anticipation, another important aspect not reflected in this schema is the difference between presidential and parliamentary systems. Parliamentary systems have a single linkage mechanism as the one presented, where citizens vote for parliaments, and parliaments subsequently appoint the executive. In presidential systems however, the linkage mechanism presented above is duplicated, as both presidents and parliaments are popularly elected, and subsequently have to negotiate about policy implementation with each other (Strøm 2000). In this chapter we further consider how parliamentary versus presidential systems affect accountability and responsiveness differently. Despite its simplicity, this schematic representation does provide a somewhat clearer overview of the elements that should be relevant for explaining variation in accountability and responsiveness across third wave regimes.

¹⁰⁴ The degree to which party systems present voters with meaningful choices is certainly related to the quality of elections, as the possibilities for opposition parties to take part in elections can be limited quite considerably in low-quality elections. However, the structure of the party system is likely to depend on a number of other factors as well, such as societal cleavages, the electoral system, elite divisions, etc.; and hence it seems unlikely that the two factors would be collinear in a statistical analysis. I therefore specify the number and differentiation of parties participating in elections as a factor separate from the quality of elections.

Figure 5.1. A schematic representation of elections as linkage mechanisms



For example, if parties present similar policy proposals, voting either one or the other is not likely to make much of a difference. Likewise, it is questionable to what degree voters have a meaningful choice in extremely fragmented party systems or in dominant party systems. On the one hand, in extremely fragmented party systems, while voters are perhaps more likely to find a party that is close to their preferences, government formation after elections will be challenged by the need to form a coalition with many parties. On the other hand, in dominant party systems, while there might be other parties/candidates running in elections, there is no viable governing alternative to the dominant party, rendering it difficult for voters to change their government if they would wish to do so. Hence, while a multi-party system is required to have a choice at all, a party system that is neither too fragmented nor too dominant, in conjunction with policy differentiation between parties, is probably the best guarantee to provide voters with choices that can generate accountable and responsive governments (Wessels and Schmitt 2008)¹⁰⁵.

¹⁰⁵ Another aspect that is relevant in this respect is the stability of the party system. In order for citizens to select future governments and signal policy preferences, they need to know which parties there are and what they stand

Moving from votes to government formation, this linkage step, also called “institutional aggregation”, is influenced most importantly by the electoral system that determines how citizens’ votes are aggregated to election outcomes (Powell 2004). Here, the most important distinction is between proportional electoral systems that aim for representation of all voters in the aggregated outcome, and majoritarian electoral systems that aim for representation of the majority of voters in the aggregated outcome (Lijphart 1999; Powell 2000)¹⁰⁶. Hence, while proportional electoral systems are likely to allow for a broader representation of voters in parliament and power-sharing coalition governments, majoritarian systems are more likely to generate power-concentrating single-party governments backed by majority support in parliament. The consequences of different electoral systems on accountability and responsiveness have been subject to much debate, and are discussed in relation to the division of power below.

Turning to the link between governments and policy-implementation, here two elements seem to be crucial: (a) the division of power, and (b) state capacity. The division of power refers to the political system as a whole, and the degree to which competences for policy-making and implementation are divided between different institutions (Lijphart 1999). The division of power determines what is at stake in elections or rather the “relative power of elected officials”, defining the leverage that elected officials have over policy-making, and thereby the influence that citizens can have through elections (Coppedge 2002)¹⁰⁷. On the other hand, state capacity refers to the degree to which the state

for. It has been demonstrated that citizens are less informed about young parties and about parties in party systems that change rapidly (Dahlberg 2009; Van Ham 2009). Hence, a stable party system should contribute to citizen’s capacity to choose parties that reflect their preferences, strengthening the link between preferences and votes. Unfortunately however, since I don’t have data on party system stability for the 97 third wave regimes, the effect of this variable was not tested.

¹⁰⁶ I refer to voters here, instead of citizens, to underline the fact that non-voters are not represented.

¹⁰⁷ “Some authors have suggested that merely considering whether offices are elected is not sufficient to get at the essential question at stake—who exercises power?” (Munck and Verkuilen 2002). Labeled in various ways,

bureaucracy is capable of implementing policies, which is likely to depend both on state resources as well the effectiveness of the bureaucracy. While state capacity is relatively straightforward, the division of power is more complex and hence discussed in somewhat more detail here. Following the literature, I distinguish two elements of division of power: horizontal division of power and vertical division of power (O'Donnell 1998). The horizontal division of power refers to institutions that do not fall under the discretion of elected officials. These institutions are also often denoted as “horizontal accountability institutions”. Examples are independent judiciaries, independent central banks, and non-majoritarian policy implementation institutions. The vertical division of power refers to multi-level governance, i.e. the delegation of competences to smaller or larger territorial units, or ‘lower’ or ‘higher’ levels of government. Examples are de-centralization of government or federalism, as well as policy-making at the level of the European Union or agreements within other supra-national regional organizations.

Apart from these formal institutional limits on the discretion of elected officials, there are a number of additional factors that are likely to limit the relative power of elected officials, especially in the case of new democracies and hybrid regimes. The first has been referred to as “external control”, and defines the degree to which elected governments are in fact sovereign, an example being the limits on government discretion in terms of economic policy when economies are strongly dependent on aid and loans and governments are under close supervision of international financial institutions such as the IMF or the World Bank. The second is a set of institutions that create horizontal division of power, but that tend to be referred to not as “horizontal accountability institutions”, but rather as “reserved

ranging from “effectiveness of elections” (Hadenius 1992); “agenda-setting power of the executive” (Munck 2004); “how much do rulers rule” (Snyder 2006); or the “relative power of elected officials” (Coppedge 2002); these authors postulate that what makes elections meaningful is not only contestation within the electoral arena, but also the “space of contestation”, i.e. what is decided by elections and what not, that is determined by the institutional configuration of the extra-electoral arena.

power domains". Examples are (formal or informal) influence on policy-making of the military or the monarchy, such as the military in Turkey or the monarchy in Thailand (or the Netherlands). Finally, the literature on power-sharing versus power-concentration has not provided a clear answer yet on how to consider presidential versus parliamentary systems. While a presidential system creates division of power by allowing for two parallel electoral linkage mechanisms, where both president and parliament are directly elected, the constitutional division of competencies between president and parliament determines to what degree power is indeed divided or concentrated in the office of the president.

Turning to expectations for the consequences of the division of power for accountability and responsiveness, dispersion of power versus concentration of power is a major point of contention in the literature on comparative political institutions (Lijphart 1999, Powell 2000, Norris 2008, Van der Eijk and Franklin 2009). The proponents of dispersion of power, or 'power-sharing' argue that power-sharing strengthens the linkage function of elections due to its ability to represent all diverging interests in society (i.e. representing the citizenry as a whole instead of just the majority of citizens, as is the case under power-concentrating or majoritarian systems). The representation of diverging interests, power-sharing advocates believe, does not only strengthen the link between votes and government, but also strengthens the link between government and policy, as policy implementation is likely to be more effective when all interests affected by the policies have been included in the decision-making process (Lijphart 1999, Norris 2008). On the other hand, the proponents of power-concentration argue that power-concentration improves the linkage mechanism of elections, as it stimulates the formation of broad parties that bring together divergent interests, and it tends to generate governments with a clear mandate from the majority of the citizenry, enhancing effective policy implementation (Gerring et al. 2005). A third view is that power-sharing and power-concentration allow for different forms of citizen influence (Powell 2000, Van der Eijk and Franklin

2009). While power-sharing systems allow for ex ante representation of all interests in government, with the subsequent formation of government and implementation of policies being a somewhat more fuzzy process, power-concentration systems allow for ex post accountability of governments, due to high clarity of responsibility, while allowing for less broad representation of interests¹⁰⁸. In terms of accountability and responsiveness, this literature suggests that majoritarian systems generate better accountability, whereas proportional systems generate better responsiveness (Powell 2000). The hypotheses formulated in this chapter derive from this literature. Table 5.1 below gives an overview of the aspects that affect the functioning of elections as linkage mechanisms, i.e. the quality of the electoral regime.

Table 5.1 Factors influencing the functioning of elections as linkage mechanisms

Functioning of elections as linkage mechanisms influenced by:	Which linkage affected?
Election quality:	From preferences to votes
- De facto possibilities to participate freely and have ones' vote counted fairly (voters)	
- De facto possibilities to compete freely and on a fair playing field for citizens' votes (parties)	
Party system:	From preferences to votes
- Number of parties	
- Policy differentiation parties	
- Relative size parties	
- Stability party system	
Electoral system:	From votes to governments
- Type of electoral system	
Division of power:	From governments to policy
- Political system (presidential versus parliamentary)	
- Horizontal division of power:	
o Bicameralism	
o Independence judiciary	
o Independence central bank	
o Non-majoritarian policy implementation agencies	
o Reserved power domains	
- Vertical division of power:	
o Federalism / decentralization	
o Supra-national regional integration	
- External control:	
o Aid / debt conditionality	
- State capacity:	From governments to policy
o State resources	
o Effectiveness bureaucracy	

¹⁰⁸ Note that, as Van der Eijk and Franklin (2009) and Powell (2000; 2004) point out, most political systems are neither fully power-sharing nor fully power-concentrating, but combine elements of dispersion and concentration, hence 'real existing democracies' are likely to combine features of both ideal-types.

Summarizing, in representative democracies elections function as linkage mechanisms between citizens and governments, providing opportunities for citizens to render a verdict on the past performance of their government, select a new government, and establish guidelines for future government actions (Powell 2000; Van der Eijk and Franklin 2009). Hence, when elections function as “instruments of democracy”, they should generate both citizen influence on government formation as well as citizen influence on government policies (Przeworski, Stokes and Manin 1999; Powell 2000). Citizen influence on government formation is defined here as accountability, asking to what degree unsatisfactory government performance before elections affects the fate of incumbents at the polls. Citizen influence on government policies is defined here as responsiveness, asking to what degree governments formed after elections implement policies that are congruent with the policy preferences of the voters.

Before turning to the specification of hypotheses and empirical testing, two aspects relating to these definitions need to be clarified. A first aspect is that of preference heterogeneity among voters. Both the definitions of accountability and responsiveness assume the possibility of aggregating citizen’s preferences into an aggregate “public opinion” that represents the median preference of citizens. A considerable research literature exists that questions to what degree voter preferences can indeed be aggregated, especially when multiple issues are at stake (Arrow 1963; Przeworski, Stokes and Manin 1999). In the case of accountability, this is relevant to determine when government performance is “unsatisfactory”. Most empirical work on accountability defines unsatisfactory performance using economic indicators, while, especially in third wave regimes, citizens’ might consider other issues such as corruption, safety, or respect for human rights and civil liberties as equally important or more important for evaluating government performance (Barreiro 2008). Assessing accountability would thus imply assessing which issues are most important to citizens and whether unsatisfactory government performance on these issues indeed leads to vote losses for incumbents, and eventually,

incumbent turnover. In the case of responsiveness, this is relevant to determine which policies are preferred by voters. Here too, most research focuses on governments' economic policies or public expenditures, while giving less consideration to the salience of different policy areas to citizens, or to citizens' preferences on other policy terrains, such as health care, education, or foreign policy. Hence, when evaluating the degree to which elections generate accountability and responsiveness at the aggregate level, it is important to consider the degree to which aggregate preferences match the distribution of individual citizens' preferences, as well as the potential differences in the extent of accountability and responsiveness in different policy domains. Nevertheless, such fine-grained analyses require detailed individual level data on citizens' opinions on issue salience, performance evaluations of government and policy preferences, which renders such analyses more suitable for within-country analyses than for cross-national research.

In this chapter, I therefore take an aggregate approach, assuming citizens' aggregate preferences for certain policy outcomes, rather than actually measuring those preferences. The degree to which elections generate accountability and responsiveness then depends on the whole sequence of linkages connecting citizens' preferences to votes to governments to policy, as represented schematically in figure 5.1. This "electoral regime" refers to the elections, their quality and the choices they offer to voters, but also to the wider institutional context in which elections take place, that determine how votes are aggregated into outcomes, and to what degree election outcomes are consequential for policy-making and implementation. Since our interest lies in assessing the relative importance of the quality of elections for accountability and responsiveness, this aggregate 'system'-approach seems justified. However, as noted, an aggregate analysis of this type cannot address preference heterogeneity among voters, nor does it address the more specific dynamics within the sequence of linkages.

5.2. Consequences of variation in election quality for accountability

Research on electoral accountability tends to focus on economic performance, i.e. whether governments are rewarded or punished in elections for their record on economic growth, unemployment or inflation during their years in office (Maravall and Sanchez-Cuenca 2008, Cheibub and Przeworski 1999, Van der Brug et al. 2007). In advanced industrial democracies, effects found are typically small and contingent, which might partly be attributed to the use of different economic performance indicators and different dependent variables (i.e. incumbent support, incumbent vote share, or incumbent turnover), as well as to differences in sample selection (Lewis-Beck and Paldam 2000, Anderson 2007, Samuels and Hellwig 2010). Recent years have brought increased research on electoral accountability in third wave regimes as well, which seems to demonstrate the same variability in results, some studies finding strong and consistent effects of economic performance on incumbent performance in elections (Jhee 2006, Posner and Simon 2002, Zielinski et al 2005), while other authors have found smaller or conditional effects (Pacek and Radcliff 1995, Duch 2001).

This 'contingency dilemma' has led to more careful reconsideration of factors that might affect electoral accountability (Lewis-Beck and Paldam 2000, Anderson 2007). In individual-level studies, more precise research developed on how voters gather information about government performance, how accurately voters' perceive government performance and how voters' subjective economic perceptions are formed, as well as whether and how voters attribute responsibility for economic performance (Anderson 2007, Van der Brug et al. 2007, Duch 2001). In aggregate-level studies, research started to consider how the political and institutional context could affect accountability, finding that electoral accountability is more likely in political systems that have high 'clarity of responsibility' (Powell and Whitten 1993, Duch and Stevenson 2005, Anderson 2000, Samuels 2004), in party systems with 'viable governing alternatives' (Anderson 2000, Maeda 2009), and in electoral systems where electoral losses lead to decisive incumbent replacement (Van der Eijk and Franklin

2009). In this chapter, we build on the findings in this literature, connecting those to our theoretical framework of the electoral regime as a set of linkages connecting voters to governments.

Starting with the main effect of electoral accountability, we would expect government performance in the year(s) before elections to affect incumbent's electoral support and eventually, survival in office¹⁰⁹.

It is key to note that 'the rascals' have to be thrown out for a reason, i.e. electoral accountability in the narrow sense in which we define it here, only occurs if elections serve as a punishment or reward mechanism for past government performance. Hence, an incumbent government being replaced even though it performed well in the years before the elections would not indicate accountability, nor would an incumbent government staying in office despite dreadful performance. This leads to the second point, often noted but not often followed up in empirical analyses: the fact that citizens may consider performance of their government in many more policy terrains than just the economy. As Pacek and Radcliff (1995), Jhee (2006), Tucker (2006) and others have noted, in many third wave regimes other policy domains may be considered much more important than the economy, such as safety, respect for human rights and civil liberties, and the direction and speed of political reforms. On the other hand, in poorer third wave regimes we might still expect economic performance to be important for incumbent electoral support as the economy has a stronger direct effect on citizens' lives in these

¹⁰⁹ I am aware that this is a very limited operationalization of accountability, as the process of political accountability doesn't stop at elections. As Schmitter points out: "The exchanges of information, justification, and judgment that make up the ordinary cycle of accountability are less obtrusive than the "big bang" of "throwing the rascals out", but no less real and significant for all that (Schmitter 2004: 49). This implies that the process of accountability occurs also in between elections and that the most accountable rulers might be the ones that don't need to be punished at elections, since they are already responsive to citizens' interests on beforehand. Hence, measuring accountability as incumbent electoral performance and turnover is admittedly an imperfect way of measuring accountability, but in my view can be seen as a "minimal requirement test", indicating whether, if governments really perform bad in citizens' views, citizens are able to replace them. Hence, analyzing electoral accountability in this way allows us to assess at least the minimal function of democratic elections: the ability to "throw the rascals out", if they behave like rascals.

countries¹¹⁰. Hence, we would expect both economic performance as well as performance on other policy terrains to affect incumbents' success at the polls.

Now, this main 'electoral accountability' effect is likely to be mediated at each step of the linkage mechanism that connects voter's preferences to policy outcomes. First of all, electoral accountability is likely to be affected by the quality of elections: fraudulent elections should help incumbents to avoid electoral accountability and vice versa. Note that this is a conditional effect: we expect the correlation between government performance and incumbent success in elections to be weakened in fraudulent elections, and strengthened in clean elections. A second factor conditioning electoral accountability is the structure of the party system. As demonstrated by Anderson (2000) and Maeda (2009), governments are more likely to be held accountable for their economic performance in the years before the elections if there is at least one party or candidate that provides a "viable governing alternative". This is unlikely to be the case in dominant party systems or fragmented party systems, as in both cases, citizens' might be dissatisfied with the incumbent government and yet might not consider competing parties as viable governing alternatives. A third factor likely to condition the degree to which elections generate electoral accountability is the electoral system. Here, the idea is simply that in order to generate accountability, incumbent loss of votes should also lead to incumbent turnover, or what Van der Eijk and Franklin (2009) and Powell (2000) refer to as 'decisive incumbent replacement'. The effect of the electoral system is not straightforward however. While proportional electoral systems are more sensitive to small changes in electoral support, i.e. small changes in vote shares are reflected in seat shares, the consequences of changes in electoral support for government formation are less direct. While an informal rule in coalition formation is often that electoral winners should lead coalition

¹¹⁰ Jhee (2006) finds the opposite effect however of stronger economic accountability in countries with higher levels of economic development. This might be due to the fact that electoral fraud, which Jhee doesn't take into account, is relatively more common in poor countries, undermining electoral accountability.

negotiations, the viability of coalition governments depends on policy differences and relative size of parties, and hence coalition governments can also –and quite often do–include parties that lost elections. On the other hand, in majoritarian electoral systems generally a greater electoral loss is needed before the decline in electoral support becomes visible in the division of parliamentary seats. At the same time however, a decline in seat shares is more directly associated with incumbent replacement. Hence, as a guiding hypothesis I assume that incumbent replacement might be more decisive in majoritarian systems, and hence electoral accountability might be stronger in majoritarian electoral systems.

This leads us to the fourth factor conditioning electoral accountability: the institutional division of power that shapes clarity of responsibility (Powell and Whitten 1993, Duch and Stevenson 2005, Anderson 2000, Samuels 2004). Here the intuition is that in order to evaluate the performance of the incumbent government, it needs to be clear to voters who can be held responsible for policy outcomes. In the case of coalition governments or political systems with high division of power, it might not be all that clear which party in the coalition was responsible for policy outcomes, or if policy outcomes fell under the discretion of the government. Clarity of responsibility can be influenced by the domestic institutional structure, for example it is likely to be higher for single-party majority governments, unicameral and unitary systems, as opposed to multi-party coalition governments, bicameral and federal or decentralized systems. However, international political and economic factors can weaken clarity of responsibility as well, as is the case for –for example– member states of the European Union, or for economies that are highly integrated into the global economy, leaving governments with limited influence on economic performance (or at least with an argument to shift the blame for weak performance to the whims of the international economy), weakening electoral accountability (Hellwig and Samuels 2007a). Whether presidential systems generate better accountability is an empirical question, as on the one hand, the visibility of a president should enhance clarity of responsibility, but

on the other hand, the dual linkage mechanism in presidential systems also divides policy-making competencies over the president and the parliament, obfuscating clarity of responsibility. Hellwig and Samuels (2007b) and Samuels (2004) found stronger economic accountability in presidential systems in established democracies and hence as a working hypothesis I will assume the same for our sample.

Finally, related to the notion of clarity of responsibility is state capacity. If state capacity is weak, governments might not have the economic or bureaucratic capacity to generate policy outcomes that are evaluated positively by citizens, undermining the possibility to hold governments accountable. For example, in poorer developing countries, or countries that just emerged from civil war, the economy might be in such bad shape that governments might have very little capacity to actually generate policy outcomes that might satisfy citizens. Whether citizens will hold their governments to account under these circumstances is an empirical question, but at least governments will have the availability of compelling circumstances to shift blame, which might weaken the extent to which citizens hold their governments to account.

Concluding, elections are more likely to generate accountability if elections are clean, there is clarity of responsibility, decisive incumbent replacement, and (at least one) viable governing alternative to the incumbent government. Clarity of responsibility is most likely influenced by the institutional division of power and state capacity; decisive incumbent replacement is most likely to be influenced by the electoral system; and the presence of viable governing alternatives depends on the structure of the party system. In addition to these macro-level factors conditioning the degree to which elections generate accountability, in the case of third wave regimes the role of availability of information seems a crucial aspect to take into account as well. Jhee (2006) found that electoral accountability is strengthened over successive elections in transitional democracies, indicating a possible learning process of voters becoming more informed about government performance and starting to hold their

rulers to account as they experience more elections. Likewise, Van der Brug et al. (2007) and Duch (2001) have pointed to the important role of information for electoral accountability, suggesting that independence of the media might be an important aspect conditioning electoral accountability as well.

5.3. Hypotheses, data and methods

Table 5.2 below gives an overview of the hypotheses. As indicated above, the only direct effect of interest for electoral accountability is the effect of government performance in the years before elections on incumbent success in elections. Incumbent success is measured as incumbent turnover and performance as economic performance as well as performance in terms of respect for civil liberties (the exact data used is discussed in the next paragraph). Since I expect positive performance to be associated with lower turnover, this coefficient is hypothesized to be negative. All other factors of interest are conditional factors that strengthen or weaken the electoral accountability link. Starting with election quality, I expect a positive effect on accountability, i.e. at higher levels of election quality, the relation between performance and incumbent turnover should be stronger. Turning to the availability of viable alternatives in the party system, I expect the link between economic performance and incumbent turnover to be weaker in dominant party systems and highly fragmented party systems, and stronger in party systems that have viable governing alternatives, such as two-party systems and moderate multi-party systems¹¹¹. Considering the electoral system then, we hypothesized that majoritarian systems should strengthen the relation between economic performance and incumbent turnover. Turning to clarity of responsibility, I expect electoral accountability to be weaker in political systems with higher institutional division of power, i.e. in parliamentary systems, bi-cameral systems, federal systems, and systems with many veto-points. I also consider coalition government as a factor

¹¹¹ However, since governments in moderate multi-party systems tend to be coalition governments, the positive effect of the availability of governing alternatives might be offset by the negative effect of lack of clarity of responsibility. For this reason, I include coalition government as a separate variable as well.

potentially weakening clarity of responsibility and hence electoral accountability. Continuing, considering state capacity it seems that lower levels of economic development, greater dependence on development aid, a history of civil war and a strongly globalized economy should weaken the link between economic performance and incumbent turnover, while an effective bureaucracy should strengthen electoral accountability. Finally, the availability of information should affect accountability as well, in this case strengthening the performance-turnover connection, and hence we hypothesize independence of the media, concurrent elections and democratic experience to have a positive effect.

Table 5.2. Overview of hypotheses

Independent variables		Hypotheses
Direct effect		Expected effect on incumbent turnover
Economic performance		Negative
Civil liberties performance		Negative
Conditional effect		Expected effect on coefficient performance and incumbent turnover
Election quality		Positive
Availability of viable alternatives		
	Dominant party system	Negative
	Two-party or moderate multi-party system	Positive
	Fragmented party system	Negative
Decisive incumbent replacement		
	Majoritarian electoral system	Positive
Clarity of responsibility		
	Parliamentarism	Negative
	Bicameralism	Negative
	Federalism	Negative
	Coalition government	Negative
	Number of veto-points	Negative
State capacity		
	Level of economic development	Negative
	Aid as proportion of GDP	Negative
	History of civil war	Negative
	Bureaucratic effectiveness	Positive
Availability of information		
	Independence media	Positive
	Concurrent elections	Positive
	Democratic experience	Positive
Control variables		
	Election type	-
	Election sequence number	-
	Region	-

The data used for the empirical analyses originate partly from the Quality of Government dataset (version April 6th 2011) from the University of Gothenburg (Teorell et al. 2011) and partly from data collected by the author. The data for election quality are based on our aggregate variable as discussed in chapter 2. Data for which no precise indicators were available in the QoG dataset, such as the dependent variable incumbent turnover, coalition governments, the structure of the party system, legislative electoral systems, and experiences with civil war were gathered by the author on the basis of information from the Electoral Handbooks on Africa, South and Central America, Central and East Asia and Europe edited by Dieter Nohlen (Nohlen 1999, 2001, 2005, 2010), as well as the data handbook by Rose and Munro (2003) on Central and Eastern Europe, and the Parline Database of the Inter-Parliamentary Union (IPU). Appendix E presents an overview of the variables used.

The dependent variable is measured as incumbent turnover. I originally constructed a variable that differentiates wholesale turnover from partial turnover, with partial turnover indicating legislative elections where coalition governments changed composition only partly (for example one party continuing in the new government with new coalition parties), and executive elections where a new president was elected that belonged to the incumbent political party. However, in practice most of these partial turnovers appear to present rather continuity of government, as in the case of legislative elections, more often than not the majority party continued in government with new minor partners, and in the case of presidential elections, often the incumbent president did not run due to illness/old age or due to age limits, and the new president was the hand- or party-picked successor. Hence, I recoded partial turnovers as no turnover, leaving a dichotomous dependent variable for our analyses. Economic performance was measured with growth or decline in real GDP per capita in the 2 years

before the elections¹¹². Alternative performance was measured with respect for civil liberties 2 years before the elections¹¹³.

As for the conditional variables, election quality is measured using our overall election quality variable that runs from 0 to 1, with 1 indicating high quality. To ease interpretation of the conditional effects, I recoded election quality to a dummy variable that differentiates severely flawed and flawed elections from clean elections. The structure of the party system is also a dummy variable differentiating dominant party systems (number of opposition parties/candidates ≤ 2) from other party systems. Several specifications for party system structure were tested, including dummies for two-party, multi-party, and fragmented party systems, but none of these variables turned out significant. The electoral system is a dummy distinguishing majoritarian (1) from proportional systems (0) in the case of legislative elections and absolute majority (1) from plurality systems (0) in the case of presidential elections. The division of power variables are all dummy variables indicating whether the political system is parliamentary, bicameral, federal and whether there was a coalition government. We also included a quasi-continuous variable counting the number of veto-players in the political system, derived from the Database of Political Institutions, which runs from 1 to 8 (Keefe 2009). As indicators of state capacity, level of economic development was measured as GDP per capita 2 years before the elections (continuously and as a dummy separating the “bottom billion” countries from

¹¹² I took economic growth instead of the average degree of economic development in the years before elections because it seems to make more sense intuitively that people would blame their government for changes in economic conditions rather than the state of the economy as such. In fact, preliminary analyses using a variety of GDP and inflation measures showed that only growth/decline of GDP significantly affected incumbent turnover.

¹¹³ Change in respect for civil liberties in the 2 years before the elections was tested as well, but not found to have a significant effect on incumbent turnover. Rather, static indicators of respect for civil liberties, i.e. 2 years, 3 years, and the average of 2/3 years before the elections were found to affect incumbent turnover. For the civil liberties variable I took the average level 2 and 3 years before elections, leaving a lag of one year between the measurement of civil liberties and the election year, to limit endogeneity.

middle-income and richer countries, cf. Collier 2009), aid was measured as the proportion of GDP consisting of official development aid 2 years before the elections, and bureaucratic effectiveness was measured by the World Bank's government indicator of "government effectiveness" that measures the effectiveness of the bureaucracy. History of civil war is a dummy variable indicating whether or not the country had experienced civil war in the 2 years before elections. Finally, as indicators of the availability of information we include a variable on independence of the media 2 years before the elections from the CIRI dataset (Cingranelli and Richards 2010), as well as a dummy variable indicating whether elections were concurrent or not, the sequence number of the election, and an indicator of previous democratic experience. Finally, control variables were the type of the election (legislative or executive), the presence of term limits for presidential elections and geographical region.

5.4. Results

The tables below show the result of the empirical analyses using our incumbent turnover dummy as the dependent variable. Following my definition of "potentially democratic elections", the full sample includes 886 elections in 97 countries between 1974 and 2009. The sample that can be used in our analyses is however limited slightly by missing data on the independent variables, leaving an N of 847 elections and 92 countries¹¹⁴. Though the structure of the data lends itself for time-series-cross-sectional analyses, as explained in chapter 4, with this sample, for many countries the time-series of elections are too short to do TSCS models (5 elections per country is the rule-of-thumb minimum for TSCS models). This is mainly due to the fact that many countries in the sample (re-)democratized later than 1990, or had democratic periods interrupted by periods of autocratic rule or civil war,

¹¹⁴ Four countries drop out due to missing data for the variable counting the number of veto-players (from the Database of Political Institutions) and one country drops out due to missing data for the variable average GDP growth. These countries are Sao Tome and Principe, Seychelles, Montenegro, Serbia (since 2006) and Serbia and Montenegro (before 2006). Since these data are missing for all years in these countries, imputation was not possible. However, a rerun of the analyses without this independent variable gives near-identical results.

lowering the number of elections held. Moreover, since I analyze legislative and executive elections together, concurrent elections generate double time-observations, rendering TSCS analyses difficult to run on a pooled dataset. Hence, I follow the approach outlined in chapter 4, and analyze the data using multi-level models, correcting for the clustering of elections within countries. As a check on our results, I also re-run the final models with a sample limited to $T > 5$, and re-run the models using TSCS analyses on separate samples of legislative and executive elections. I first considered the main accountability effect, i.e. the effect of economic performance on incumbent turnover, as shown in model 1, and then look at the effects of conditional variables by 'linkage mechanism', resulting in model 2 - 7 demonstrated in table 5.3 (election quality, party system, electoral system, division of power, state capacity). In each model, we include both main effects as well as the interaction effect with economic performance to test whether the independent variables weaken or strengthen the impact of economic performance on incumbent turnover.

As table 5.3 shows, the effect of economic performance is significant and in the expected direction: economic growth in the 2 years before elections is associated with lower turnover and vice versa. Note that, since this main accountability link is a negative coefficient (economic growth leads to lower turnover and vice versa), a conditional effect that strengthens this link is indicated by a negative coefficient for the interaction term. Model 2 confirms the strong conditional effect of election quality: in clean elections, economic growth prior to elections has a strong negative effect on incumbent turnover, while in flawed elections this effect is much weaker and only significant at the 0.01 level. Election quality has a considerable main effect as well, i.e. regardless of economic performance, clean elections are significantly more likely to result in incumbent turnover than flawed elections. Turning to the conditional effect of the structure of the party system, i.e. the availability of viable alternatives, I found no significant effects of two-party and multi-party systems strengthening accountability, nor of fragmented party systems weakening accountability. The only marginally significant effect was the

interaction with dominant party systems, which seems to weaken the accountability link between economic performance and incumbent turnover. In other words, in dominant party systems (i.e. when equal or less than 2 other parties or candidates run in elections), economic decline in the years before elections does not lead to higher levels of incumbent turnover. This finding confirms (though only weakly) the idea that citizens should have viable alternatives to the incumbent president or party in order for elections to generate accountability. Turning to model 4, I expected majoritarian electoral systems to result in more decisive incumbent replacement, and hence strengthen the performance-turnover link. However, though the negative main effect is in line with expectations (in majoritarian electoral systems generally a greater vote loss is needed for incumbent turnover), the interaction effects turn out insignificant and in the opposite direction, both for legislative and executive elections. The conditional effect of electoral systems remains insignificant in the time-series cross-sectional models for legislative and executive elections separately, hence our hypothesis with regard to electoral systems needs to be rejected.

Turning to clarity of responsibility, model 5 reports the results for our division of power variables. I expected accountability to be weaker in political systems with higher division of power since those systems tend to have less clarity of responsibility. Hence, we should expect a weaker link between economic performance and incumbent turnover in parliamentary, bicameral and federal systems, as well as in elections where previous governments were coalition governments¹¹⁵. While a conditional effect that strengthens the accountability link is indicated by a negative interaction effect, a conditional effect that weakens the accountability link, as we expect for high division of power systems, should be indicated by positive interaction effects. As model 5 shows, only bicameralism and federalism seem to weaken electoral accountability, while the coefficients for parliamentarism and coalition governments

¹¹⁵ Note that model 5 has a lower number of cases since the indicator of previous coalition government is a lagged variable, leading a number of elections to drop out.

seem to strengthen accountability. Moreover, none of these effects are significant, which seems to reject the clarity of responsibility hypothesis in our sample. However, perhaps clarity of responsibility is better measured by a composite variable that comprises all of these systemic characteristics (i.e. parliamentarism, bicameralism, federalism) and also includes a more precise indication of the division of power between chambers and between the president and the parliament (i.e. whether the governing party holds a majority in both houses, and whether president and the parliamentary majority belong to the same party). The Database of Political Institutions includes a variable that provides such a composite measurement of division of power by counting the number of veto-players in a political system (Keefe 2009)¹¹⁶.

Model 6 shows the conditional effect of the number of veto-players, and this model seems to provide more support for the clarity of responsibility hypothesis. Not only does the inclusion of this variable strengthen the main accountability effect, i.e. the link between economic performance and incumbent turnover, the conditional effect also seems to confirm the idea that in high division of power political systems, the link between economic performance and turnover is weaker.

¹¹⁶ The variable is constructed as follows: “Equals one if the Legislative Index of Political Competitiveness (dpi_lipc) or the Executive Index of Political Competitiveness (dpi_eipc) is less than 6. In countries where dpi_lipc and dpi_eipc are greater than or equal to 6, dpi_checks is incremented by one if there is a chief executive, by a further one if the chief executive is competitively elected, and by a further one if the opposition controls the legislature. In presidential systems, dpi_checks is incremented by one for each chamber of the legislature (unless the president’s party has a majority in the lower house and a closed-list system is in effect), and by one for each party coded as allied with the president’s party and which has an ideological (left-right) orientation closer to that of the main opposition party than to that of the president’s party. In parliamentary systems dpi_checks is incremented by one for every party in the government coalition as long as the parties are needed to maintain a majority, and by one for every party in the government coalition that has a position on economic issues closer to the largest opposition party than to the party of the executive.” (The Quality of Government Dataset, April 6 2011).

Turning to our argument about state capacity, here we expected that the presence of credible mitigating circumstances allowing governments to shift blame for economic performance would weaken accountability. Hence, I expect that governments in poor countries, countries that receive high proportions of aid, governments in countries with strongly globalized economies, and countries with a recent history of civil war to have (or at least be able to argue to have) comparatively less influence over economic policies. Hence I would expect the economic performance – turnover link to be weaker in such contexts. I expected the opposite effect for bureaucratic effectiveness. As the results in model 7 show, economic development indeed has the expected effect, though the coefficient is not statistically significant. In richer countries (those with average GDP per capita in the 2 years before elections higher than 2000 dollar) economic growth seems to be more strongly associated with incumbent turnover than in poorer countries. The coefficient for bureaucratic effectiveness runs in the opposite direction as expected, but is not significant. The effect of aid appeared to be significant only when including two highly influential outliers of election years in which countries had received aid amounting to more than 100% of their GDP. Since this lead to the unnecessary loss of cases, and the effect of aid is clearly insignificant once these cases were excluded, it is not reported in the models here. The only state capacity variable that appears to be significantly related to economic accountability is civil war, however, the effect is contrary to what we hypothesized. In countries that experienced civil war in the 2 years before the elections the performance–turnover link seems to be stronger. Apparently, rather than providing governments with a chance to shift blame for economic performance, economic performance becomes more important for incumbent survival after civil war (which makes sense given the detrimental economic effects of civil war, cf. Collier 2009).

As a final set of conditional hypotheses, we postulated that the performance – turnover link should be influenced by the availability of information: after all, in order for voters to use their vote to hold incumbents accountable for past performance, voters need to know how the economy is doing and

have a notion of the governments' track record of policy decisions with regard to the economy in the past years. Of course, many citizens, especially in poorer countries, will not need media to inform them about the state of the economy as economic growth or decline affects citizens daily lives more strongly than in more developed countries. However, in order to attribute responsibility for economic performance to the government, information might still be important, even in those contexts.

Table 5.3. Electoral accountability: Explaining incumbent turnover in elections 1974-2009

	Model 1 Performance	Model 2 Election quality	Model 3 Party system	Model 4 Electoral system	Model 5 Division power	Model 6 Veto-points	Model 7 State capacity	Model 8 Information
Average growth GDP per capita (t-2 %)	-0.061*** (0.015)	-0.032+ (0.019)	-0.068*** (0.016)	-0.067** (0.021)	-0.056+ (0.032)	-0.126*** (0.034)	-0.038 (0.036)	-0.081* (0.035)
Election quality (flawed vs clean 0-1)		1.921*** (0.210)						
Interaction election quality * growth		-0.102** (0.035)						
Dominant party/candidate system (0-1)			-0.498 (0.355)					
Interaction party system * growth			0.100+ (0.061)					
Electoral system maj leg (0-1)				-0.762** (0.277)				
Interaction electoral system leg * growth				0.016 (0.043)				
Electoral system abs maj ex (0-1)				-0.163 (0.205)				
Interaction electoral system ex * growth				0.005 (0.035)				
Parliamentary system (0-1)					-0.043 (0.442)			
Interaction parliamentarism * growth					-0.008 (0.058)			
Bicameralism (0-1)					-0.422 (0.356)			
Interaction bicameralism * growth					0.002 (0.040)			
Federalism (0-1)					0.159 (0.554)			
Interaction federalism * growth					0.037 (0.058)			
Coalition government (lag 1 0-1)					-0.084 (0.292)			
Interaction coalition * growth					-0.082 (0.060)			

	Model 1 Performance	Model 2 Election quality	Model 3 Party system	Model 4 Electoral system	Model 5 Division power	Model 6 Veto-points	Model 7 State capacity	Model 8 Info
Average growth GDP per capita (t-2 %)						-0.126*** (0.034)	-0.038 (0.036)	-0.081* (0.035)
Number of veto players (1-7)						0.154* (0.073)		
Interaction veto players * growth						0.024+ (0.013)		
Economic development (poor vs rich 0-1)							0.655* (0.302)	
Interaction development rich * growth							-0.012 (0.038)	
Civil war (av t-2 0-1)							0.331 (0.324)	
Interaction civil war * growth							-0.083+ (0.050)	
Effectiveness bureaucracy (-1.8 – 1.4)							0.263 (0.251)	
Interaction bureaucracy * growth							0.006 (0.029)	
Independence media t-2 (0-1)								0.588* (0.296)
Interaction media * growth								0.035 (0.049)
Election sequence number (1-20)								0.067* (0.031)
Interaction sequence * growth								-0.002 (0.006)
Constant	-0.595*** (0.152)	-1.454*** (0.175)	-0.558*** (0.155)	-0.464* (0.187)	-0.345 (0.230)	-0.987*** (0.235)	-0.927*** (0.270)	-1.179*** (0.242)
LL	-524***	-477***	-522***	-521***	-455^	-518***	-516***	-519***
N level 1	847	847	847	847	730	847	847	847
N level 2	92	92	92	92	86	92	92	92
Pseudo-Rsq / % correctly predicted	27.8% / 72.4%	32.6% / 75.1%	28.3% / 72.6%	27.6% / 72.7%	29.0% / 73.4%	28.6% / 74.0%	28.2% / 73.1%	27.8/73.6

Standard errors in parentheses. P-values: + 0.1, * 0.05, ** 0.01, *** 0.001.

Hence I would expect the availability of information to strengthen the performance-turnover link, both with respect to economic outcomes as well as with regard to respect for civil liberties or other policy areas citizens might hold their government accountable for. Model 9 shows the results. The effects of concurrent elections and previous experience with democracy are not shown since neither main nor interaction effects were significant. Regarding independence of the media, surprisingly the interaction effect does not appear to be significant either, while the main effect is significant. So independent media are associated with higher levels of incumbent turnover, however this effect does not run via strengthened accountability for economic performance. A similar result appears with respect to the sequence number of the elections. This is an alternative indicator of democratic experience, and I would expect the link between economic performance and incumbent turnover to become stronger as citizens experience successive elections. However, this idea does not seem to be supported by the data, as I only find a significant main effect (i.e. turnover does increase as a country holds more elections), and no significant interaction effect (i.e. economic accountability does not necessarily strengthen over time). Table 5.4 on the next page shows the full models.

Model 9 shows the results of a full model including all the independent variables that were significant or close to one-tailed significance in the previous models. Model 9 demonstrates how crucial election quality is for electoral accountability to function: in clean elections, the association between economic growth and incumbent turnover is strong and significant, while in flawed elections, there is no such association. Availability of alternatives also appears to be important in shaping the link between economic performance and incumbent turnover: in dominant party systems, the performance-turnover link is significantly weaker. Clarity of responsibility seems to have less explanatory power in this sample: while a higher number of veto-players weakens the performance-turnover link, this interaction effect is no longer significant in the full model. Nor are the variables that had strong main effects on incumbent turnover, as the electoral system, economic development and independence of the media.

Considering state capacity, the main effect and the conditional effect of civil war remain significant in a full model. Again, civil war doesn't seem to lead to a weakening of the performance-turnover link as hypothesized, but rather strengthen the relation between economic growth and incumbent turnover. Given the importance of economic recovery after civil war this might not be so surprising however: economic growth simply becomes much more relevant for the electoral success of incumbents when the need for economic recovery is high, which is in line with arguments made by Collier (2009). To test whether these results hold when including control variables, I reran a more parsimonious model, including only those variables that remained significant in the full model 9, and added control variables for the type of election (legislative or executive), constituent assembly elections, and region.

The effect of concurrent elections was already tested in model 8 and didn't turn out significant, so it is left out here as well¹¹⁷. Election sequence was already included. For all control variables I tested the interaction effects with economic growth as well, but since none of these interactions turned out significant, they are not reported in model 10. Inclusion of the control variables doesn't substantially change the results, if anything, it seems to strengthen the significance of several independent variables. The conditional effects of election quality, dominant party system and civil war remain very similar to the previous model. Including region as an additional control variable shows that some unexplained variance remains in our model, which is picked up by the regional dummies: incumbent turnover appears to be lower in Sub-Saharan Africa. As for type of elections, executive elections appear to have somewhat higher levels of turnover.

¹¹⁷ Another control variable tested was the presence of term limits for presidential elections. Since this variable turned out insignificant in the full model it is not reported here.

Table 5.4. Electoral accountability: Explaining incumbent turnover in elections 1974-2009

	Model 9 Full model	Model 10 Pars model w controls	Model 11 Pars model T >= 5	Model 12 Pars model lagged DV	Model 13 Pars model legislative	Model 14 Pars model executive
Average growth GDP per capita (t-2 %)	-0.060 (0.045)	-0.007 (0.022)	-0.006 (0.022)	-0.023 (0.029)	-0.003 (0.029)	-0.018 (0.029)
Election quality (flawed vs clean 0-1)	1.989*** (0.233)	2.000*** (0.230)	2.025*** (0.232)	1.781*** (0.286)	1.974*** (0.291)	1.857*** (0.308)
Interaction election quality * growth	-0.159*** (0.039)	-0.163*** (0.039)	-0.168*** (0.039)	-0.163** (0.050)	-0.167*** (0.048)	-0.144* (0.059)
Dominant party/candidate system (0-1)	-0.481 (0.385)	-0.589 (0.388z)	-0.559 (0.391)	-0.675 (0.542)	-0.749 (0.650)	-0.469 (0.484)
Interaction party system * growth	0.127+ (0.065)	0.139* (0.064)	0.139* (0.064)	0.164+ (0.092)	0.157 (0.105)	0.132 (0.085)
Electoral system maj leg (0-1)	-0.439 (0.276)					
Number of veto players (1-7)	0.054 (0.084)					
Interaction veto players * growth	0.020 (0.015)					
Economic development (poor vs rich 0-1)	0.231 (0.271)					
Independence media t-2 (0-1)	-0.043 (0.325)					
Civil war (av t-2 0-1)	0.709* (0.348)	0.691* (0.338)	0.633+ (0.354)	1.281** (0.455)	0.777+ (0.437)	0.346 (0.408)
Interaction civil war * growth	-0.097+ (0.054)	-0.120* (0.049)	-0.117* (0.050)	-0.120+ (0.071)	-0.108+ (0.061)	-0.127+ (0.074)
Election sequence number (1-20)	0.044 (0.030)	0.060* (0.030)	0.055+ (0.030)	0.097** (0.037)	0.078* (0.037)	0.032 (0.042)
Election type (legislative vs executive 0-1)		0.412* (0.180)	0.370* (0.182)	0.380+ (0.205)		
Constituent assembly elections		0.999+ (0.563)	0.478 (0.609)	-0.059 (0.835)	1.100+ (0.573)	
Lagged dependent variable				-0.502* (0.246)		
Region – Former Soviet Republic ^a		-0.907+ (0.505)	-1.038* (0.498)	-1.518* (0.627)	-0.710 (0.574)	-1.151* (0.572)
Region – Sub-Saharan Africa		-1.015** (0.386)	-1.101** (0.391)	-1.576** (0.493)	-0.830+ (0.424)	-1.158** (0.447)

	Model 9 Full model	Model 10 Pars model w controls	Model 11 Pars model T >= 5	Model 12 Pars model lagged DV	Model 13 Pars model legislative	Model 14 Pars model executive
Region – South America		-0.686 (0.440)	-0.813+ (0.434)	-1.092* (0.547)	-0.889+ (0.478)	-0.478 (0.458)
Region – Central America		0.027 (0.484)	-0.104 (0.474)	-0.136 (0.598)	-0.026 (0.521)	-0.109 (0.487)
Region – Southern Europe		-1.295+ (0.676)	-1.414* (0.660)	-1.890* (0.843)	-1.218+ (0.681)	-2.094* (0.994)
Constant	-1.956*** (0.339)	-1.320*** (0.383)	-1.137** (0.382)	-0.814+ (0.477)	-1.413** (0.433)	-0.517 (0.459)
LL	-462***	-456***	-440***	-374***	-279***	-183***
N level 1	847	847	814	682	510	337
N level 2	92	92	82	91	92	78
Pseudo-Rsq / % correctly predicted	34.9 / 76.3%	34.6 / 76.5%	34.1 / 75.7	37.7 / 78.3	19.1 / 69.6%	24.9 / 73.0%

Standard errors in parentheses. P-values: ^ 0.1, * 0.05, ** 0.01, *** 0.001. a. Central and Eastern Europe is the reference category.

Model 11 shows the results if we limit our sample to those countries that experienced at least 5 elections, which should give more robust estimates. The results do not appear to change substantially. I then proceeded to test if the model suffers from serial autocorrelation, in which case Beck (2008) suggests to re-estimate the model including a lagged dependent variable. There is in fact very limited serial autocorrelation, the Lagrange multiplier test is just significant¹¹⁸. Hence, to make sure that our results are robust I re-estimated the parsimonious model with controls including a lagged dependent variable. Model 12 shows the results. Even though the lagged dependent variable does appear to be significant and negative, implying that turnover in the previous elections renders turnover in the current elections somewhat less likely, inclusion of the lagged dependent variable does not affect our findings for election quality. Even under these model restrictions clean elections generate economic accountability whereas flawed elections do not. The standard errors of the interaction effects of dominant party system and civil war do seem to be affected somewhat, however both remain significant at the one-tailed level. Finally, I re-ran the model using time-series cross-section logit analyses on separate samples of legislative and executive elections, the results of which are shown in models 13 and 14. Both models show that the conditional effect of election quality is about equally strong in legislative as in executive elections: in clean elections, economic performance affects incumbents' success at the polls, while in flawed elections it does not. The effect of dominant party systems is no longer significant in these two smaller samples, but otherwise the results of the separate models are very similar to the pooled model 10.

It seems hence that our findings are rather robust to different model specifications: clearly clean elections are a prerequisite for electoral accountability, and conversely, electoral fraud undermines the accountability function of elections. Nevertheless, thus far we have only considered one topic for

¹¹⁸ I also tested whether the dependent variable has a unit root and whether there was significant heteroskedasticity, and this was not the case.

which governments might be held accountable by citizens: the economy. While we would expect the economy to be very important, especially in many of the developing countries in our sample, other topics such as security, the rule of law and respect for civil liberties are likely to be of concern to citizens as well. To test this I ran a set of models taking respect for civil liberties as the performance variable. Now if elections also generate accountability in an alternative domain of performance like the governments' respect for civil liberties, limited respect for civil liberties by the government should be associated with punishment at elections and hence higher probabilities of turnover. However, naturally, it is also likely that limited civil liberties negatively affect the chances of turnover in elections, as opposition parties' possibilities to organize and assemble might be limited, media might be constrained to freely express criticism of the government, etc. I expect the quality of elections to determine the direction of this effect: in clean elections, citizens will be able to hold their government to account for a poor record on civil liberties, whereas in flawed elections, low respect for civil liberties will increase governments' hold on power.

Now of course, since civil liberties like the ones just mentioned are part of our measurement of election quality in the period of elections, the question arises whether these factors can in fact be separated. I argue that it is possible to separate these factors, for the following reasons. First of all, the measure of election quality focuses on the period of the elections, which taking into account pre-election campaigning and post-election dispute adjudication, is roughly taken to range from about half a year before the elections to about half a year after. The measure of civil liberties used here is an average of the governments' respect for civil liberties in the 3^d and 2nd year before the elections, leaving a one-year gap to ensure there is no overlap with the start of the campaign period. Moreover, considering the correlation between election quality and this measure shows a substantial but not very high correlation (0.49***), and there is strong variation in levels of respect for civil liberties before both clean elections as well as flawed elections. Hence, of course there is an association between civil

liberties in periods before elections and subsequent election quality (as the effects of rule of law and independent media on election quality in chapter 4 demonstrated), but the relationship is certainly not deterministic. Finally, the measure for civil liberties used here is the Freedom House score of civil liberties, which comprises a much wider range of civil liberties in addition to freedom of assembly and freedom of speech, and is intentionally coded separately from Freedom House's well-known political rights indicator, which includes a measure of the quality of the electoral process. Summarizing, the expected direct effect of respect for civil liberties could run in both directions, as argued above. We expect the conditional effect of election quality however to clarify the civil liberties-turnover link, where clean elections should result in a negative association between civil liberties and incumbent turnover (a worse civil liberties record leading to higher turnover), while flawed elections will have the opposite effect (a worse civil liberties record aiding the incumbent to stay in power). The expected effects of the other variables are the same conditional effects as those expected for economic performance: the availability of viable governing alternatives, decisive incumbent replacement, clarity of responsibility, state capacity and the availability of information should strengthen the negative civil liberties – turnover link.

Table 5.5 below shows the results. Indeed as expected the main effect of civil liberties is not conclusive: the coefficient is positive but only significant one-tailed. However, as election quality and its interaction with pre-election period civil liberties is added in model 2, the expected opposite effects become clear. In flawed elections, the coefficient of civil liberties and incumbent turnover equals the main effect of civil liberties, i.e. 0.311. In clean elections, the coefficient of civil liberties and incumbent turnover equals the main effect of civil liberties plus the interaction effect, resulting in a coefficient of $0.311 - 0.617 = -0.306$, an almost equally strong and clearly negative relation. Election quality also has a strong and significant main effect on the likelihood of incumbent turnover. Turning to the conditional effect of the availability of alternatives in the party system, even though the

interaction coefficient seems to suggest that dominant party systems lower electoral accountability, neither the interaction effect nor the main effect are significant. The effects of multi-party systems, fragmented party systems and two-party systems were also tested, however none resulted significant. Turning to the electoral system, as in the economic performance models, majoritarian electoral systems appear to have exactly the opposite effect as expected. While such systems should be associated with more decisive incumbent replacement and higher clarity of responsibility, it appears that majoritarian electoral systems rather weaken the accountability relationship. However, neither of the two interaction effects are significant, so these results are certainly not conclusive. The only significant coefficient is the main effect for legislative electoral systems, where majoritarian systems are associated with lower levels of incumbent turnover.

As for the division of power variables, all conditional effects are in the expected direction: parliamentary systems seem to strengthen accountability, while bicameralism and federalism weaken the civil liberties – turnover link. However, only the effect of bicameralism is significant. The effect of the number of veto-players was tested, but since it resulted insignificant as well, this alternative measure of division of power was left out of the model¹¹⁹. Apparently, clarity of responsibility seems relatively less important for accountability in our sample. State capacity on the other hand appears to be very relevant, as model 6 shows. As in the economic performance models, the effect of civil war is opposite to what we expected. As such, civil war doesn't provide an opportunity to shift blame as hypothesized, but rather seems to heighten the impact of government performance on turnover, also if performance in the years before the elections is measured as respect for civil liberties. The same seems to hold for aid, as receiving higher proportions of aid is not only associated with higher incumbent

¹¹⁹ Lagged coalition government, i.e. whether the previous elections resulted in a coalition government, was also not significant and since it substantially lowers the number of cases, left out of model 5. Note that the number of cases in model 5 is slightly lower than in the other models, this is due to missing data on the variable federalism. See Appendix E for descriptive information about the variables used.

turnover, it also strengthens the association between the government's civil liberties record and turnover. The effect of bureaucratic effectiveness is as expected: if the bureaucracy is more effective, the association between civil liberties and turnover is stronger. Economic development is not significant, and will be left out in the comprehensive models reported on the next page.

Table 5.5. Alternative electoral accountability: Explaining incumbent turnover in elections 1974-2009

	Model 1 Performance	Model 2 Election quality	Model 3 Party system	Model 4 Electoral system	Model 5 Division power	Model 6 State capacity	Model 7 Information
Civil liberties (av t-2/t-3 0-7)	0.121+ (0.071)	0.311** (0.119)	0.111 (0.073)	0.033 (0.092)	0.037 (0.099)	0.148 (0.188)	0.204 (0.163)
Election quality (flawed vs clean 0-1)		4.476*** (0.708)					
Interaction election quality * civil liberties		-0.617*** (0.154)					
Dominant party system (0-1)			-0.841 (1.033)				
Interaction party system * civil liberties			0.115 (0.228)				
Electoral system maj leg (0-1)				-1.518+ (0.803)			
Interaction electoral system leg * civil liberties				0.220 (0.183)			
Electoral system abs maj ex (0-1)				-0.716 (0.649)			
Interaction electoral system ex * civil liberties				0.130 (0.140)			
Parliamentary system (0-1)					1.071 (1.021)		
Interaction parliamentarism * civil liberties					-0.280 (0.199)		
Bicameralism (0-1)					-1.953** (0.752)		
Interaction bicameralism * civil liberties					0.341* (0.154)		
Federalism (0-1)					-0.282 (0.902)		
Interaction federalism * civil liberties					0.153 (0.187)		
Economic development (poor vs rich 0-1)						-0.068 (0.884)	
Interaction development * civil liberties						0.106 (0.200)	
Aid received as % of GDP (t-2 0-79%)						0.070** (0.026)	

	Model 1 Performance	Model 2 Election quality	Model 3 Party system	Model 4 Electoral system	Model 5 Division power	Model 6 State capacity	Model 7 Information
Interaction aid * civil liberties						-0.018** (0.007)	
Civil war (av t-2 0-1)						2.387** (0.907)	
Interaction civil war * civil liberties						-0.476+ (0.255)	
Effectiveness bureaucracy (-1.8 – 1.4)						2.563*** (0.659)	
Interaction bureaucracy * civil liberties						-0.545*** (0.135)	
Independence media t-2 (0-1)							2.885** (0.990)
Interaction media * civil liberties							-0.501* (0.213)
Concurrent election							-0.994 (0.692)
Interaction concurrent election * civil liberties							0.311* (0.151)
Election sequence (1-20)							-0.001 (0.113)
Interaction election sequence * civil liberties							0.007 (0.022)
Previous democratic experience (0-1)							1.062 (0.749)
Interaction democratic experience * civil liberties							-0.193 (0.155)
Constant	-1.275*** (0.333)	-2.777*** (0.500)	-1.200*** (0.342)	-0.704 (0.446)	-0.738+ (0.445)	-1.348 (0.824)	-2.239** (0.698)
LL	-528***	-478***	-527***	-525***	-520***	-512***	-518***
N level 1	855	855	855	855	851	855	855
N level 2	96	96	96	96	94	96	96
Pseudo-Rsq / % correctly predicted	27.2% / 72.9%	32.9% / 76.7%	27.2% / 72.9%	27.2% / 72.9%	28.1% / 74.3%	27.6% / 74.7%	29.1% / 73.7%

Standard errors in parentheses. P-values: ^ 0.1, * 0.05, ** 0.01, *** 0.001.

Finally, of the variables meant to capture information and learning effects on accountability, only independence of the media appears as significant, strengthening the civil liberties – turnover link. The interaction effect of concurrent elections was also significant, but curiously in the opposite direction as expected. Previous democratic experience and election sequence were not significant.

Turning to the full model, reported in table 5.6 on the next page, here we included only the variables that resulted significant in models 1 until 7. The effect of election quality appears robust to the inclusion of other independent variables: in clean elections the relation between pre-election civil liberties and incumbent turnover is negative, while in fraudulent elections the relation remains positive. The conditional effects of civil war and bureaucratic effectiveness also remain significant: both strengthen electoral accountability. The effect of aid is no longer significant however, nor is the effect of independence of the media. The conditional effect of concurrent elections is close to one-tailed significant ($p = 0.103$) and is hence included in the parsimonious model 10. How robust are these results to the inclusion of control variables? Model 9 demonstrates the results for a parsimonious model including the significant predictors from model 8, as well as adding control variables for the type of election and region. Election sequence was not included here since it was already tested in model 7 and did not turn out to be significant. Interaction effects were tested for all regional dummies, however since none of them turned out to be significant they are not reported in model 9. It appears that including the control variables leaves both the conditional effect of civil war and of bureaucratic effectiveness insignificant (even though the main effect of civil war is unaffected).

Election quality continues to have a significant main and interaction effect. The control variables capturing the type of election appear not to be significant, and with regard to the regional dummies, again Sub-Saharan Africa appears to have lower levels of incumbent turnover than predicted by our model.

Table 5.6. Alternative electoral accountability: Explaining incumbent turnover in elections

	Model 8 Full model	Model 9 Pars w controls	Model 10 T > 5	Model 11 Lagged DV	Model 12 Legislative	Model 13 Executive
Civil liberties (av t-2/t-3)	0.346+ (0.203)	0.231 (0.173)	0.247 (0.162)	0.319 (0.210)	0.193 (0.192)	0.378 (0.247)
Election quality (0-1)	4.164*** (0.855)	4.681*** (0.816)	4.537*** (0.843)	3.384** (1.070)	4.017*** (1.002)	5.826*** (1.257)
Interaction election quality * cl	-0.520** (0.188)	-0.636*** (0.180)	-0.618*** (0.186)	-0.411+ (0.229)	-0.473* (0.221)	-0.953*** (0.277)
Electoral system maj leg (0-1)	-0.315 (0.271)					
Bicameralism	-1.525* (0.772)	-0.714** (0.272)	-0.710* (0.281)	-0.832** (0.322)	-0.749* (0.293)	-0.838* (0.366)
Interaction bicameralism * cl	0.181 (0.159)					
Aid received as % of GDP (t-2 0-79%)	0.025 (0.028)					
Interaction aid * civil liberties	-0.010 (0.007)					
Civil war (av t-2 0-1)	2.963** (1.029)	2.519* (0.985)	2.257* (1.081)	3.089* (1.507)	3.186* (1.241)	2.013 (1.383)
Interaction civil war * cl	-0.520+ (0.275)	-0.424 (0.265)	-0.392 (0.284)	-0.473 (0.402)	-0.612+ (0.329)	-0.373 (0.360)
Effectiveness bureaucracy (-1.8 – 1.4)	1.107 (0.683)	0.091 (0.630)	0.162 (0.646)	-0.666 (0.932)	0.092 (0.740)	-0.221 (0.906)
Interaction bureaucracy * cl	-0.299* (0.142)	-0.141 (0.126)	-0.150 (0.129)	-0.039 (0.179)	-0.181 (0.148)	-0.019 (0.179)
Independence media t-2 (0-1)	0.764 (1.177)					
Interaction media * cl	-0.177 (0.243)					
Concurrent elections	-0.886 (0.722)	-1.286+ (0.726)	-1.135 (0.734)	-0.901 (0.907)	-1.196 (0.890)	-1.828+ (1.063)
Interaction concurrent elect * cl	0.250 (0.154)	0.328* (0.154)	0.302* (0.154)	0.290 (0.188)	0.332+ (0.187)	0.394+ (0.227)
Constituent assembly elections		0.638 (1.768)				
Interaction assembly elect * cl		-0.084 (0.450)				
Election type (0-1)		-0.042 (0.633)				
Interaction election type * cl		0.064 (0.132)				
Lagged dependent variable				-0.285 (0.234)		
Region ^a – Former Soviet Republic		-1.111* (0.491)	-1.318** (0.491)	-1.427* (0.574)	-0.995+ (0.540)	-1.532* (0.621)
Region – Sub-Saharan Africa		-1.695*** (0.400)	-1.900*** (0.414)	-2.153*** (0.477)	-1.664*** (0.428)	-1.916*** (0.539)
Region – South America		-0.696 (0.451)	-0.913* (0.446)	-1.001+ (0.514)	-1.003* (0.487)	-0.309 (0.564)
Region – Central America		-0.511 (0.464)	-0.737 (0.460)	-0.647 (0.534)	-0.663 (0.486)	-0.682 (0.594)
Region – Southern Europe		-0.268 (0.624)	-0.572 (0.620)	-0.580 (0.712)	-0.072 (0.592)	-1.197 (1.011)
Constant	-2.574** (0.906)	-1.537+ (0.875)	-1.241 (0.818)	-1.562 (1.081)	-1.413 (0.939)	-1.358 (1.261)
LL	-461***	-454***	-437***	-381***	-275***	-180***
N level 1	855	855	811	697	513	342
N level 2	96	96	82	95	96	81
Pseudo-Rsq / % correct predicted	33.0 / 76.5	32.7 / 75.9	31.5 / 75.6	33.5 / 77.3	21.8 / 73.1	26.9 / 73.1

Standard errors in parentheses. P-values: ^ 0.1, * 0.05, ** 0.01, *** 0.001. a. Central and Eastern Europe is the reference category.

Continuing, model 10 reports the results if the sample is limited to those countries having experienced at least 5 elections, the results are substantively the same as those reported in model 9, indicating that analyzing our full sample of countries is not affecting the results. Moreover, I tested for heteroskedasticity and serial autocorrelation, and while heteroskedasticity appeared not to be a problem, some serial autocorrelation of the residuals was found. To assess to what extent this affects the results found so far, model 11 reports the results if a lagged dependent variable is included. Though the coefficient size of election quality and its interaction effect is somewhat decreased, the results remain very similar. Finally, models 12 and 13 report the results of time-series-cross-section analyses of separate legislative and executive samples. In both samples the main effect as well as the conditional effect of election quality remain robust. Interestingly however, the coefficient sizes appear quite much stronger in executive elections than in legislative elections.

5.5. Conclusion

This chapter studied the consequences of variation in election quality for electoral accountability. Does fraud help incumbents to avoid accountability, and do clean elections help citizens to punish incumbents that performed badly in the years prior to elections? For the third wave regimes in our sample, I indeed found a very strong effect of election quality on electoral accountability: economic decline in the years prior to election was strongly associated with incumbent turnover only in clean elections. Of other factors found to be important in research on electoral accountability in established democracies, the availability of alternatives appeared to play a role in third wave regimes as well: in dominant party systems the economic performance – incumbent turnover link was significantly weakened. However, the effect of institutional variables enhancing or obfuscating clarity of responsibility appeared to be of little significance in our sample. Finally, an interesting finding was that civil war strengthens the economic performance –turnover link, presumably because the detrimental economic consequences of civil war render economic performance more important for

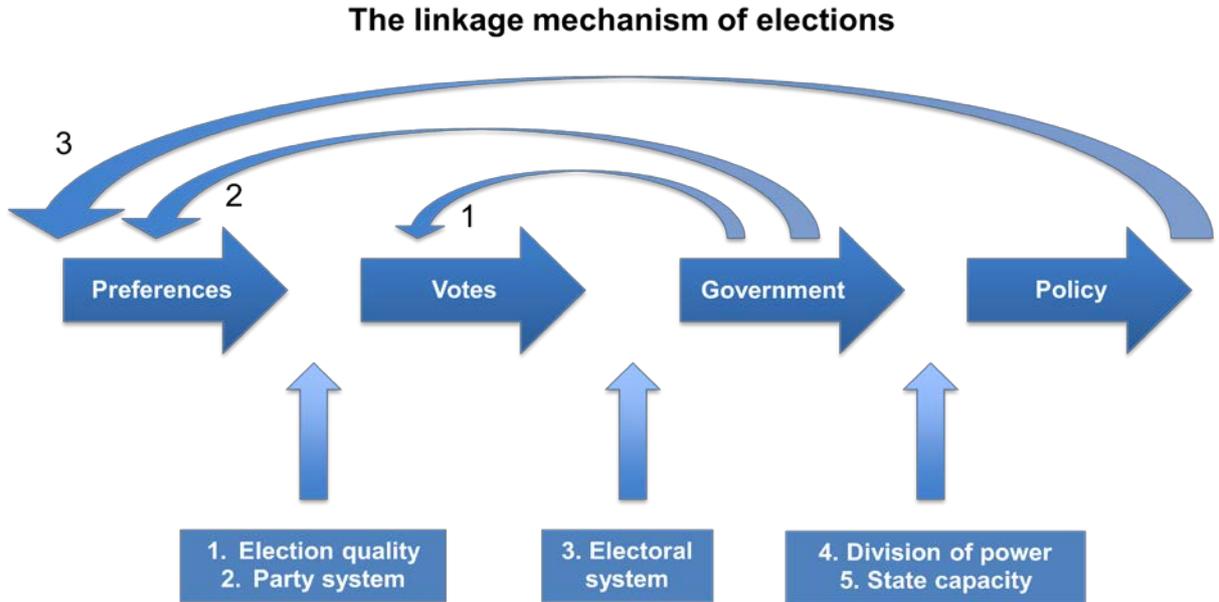
citizens' evaluations of incumbents. Tests of electoral accountability for an alternative government performance indicator, i.e. the governments' record for respecting civil liberties in the years prior to elections, demonstrated that here also the conditional effect of election quality was important. In clean elections, a worse civil liberties record in the years before elections is associated with higher incumbent turnover in elections, while in flawed elections, there is no association between the two variables. Again civil war appears to strengthen the accountability link, however this effect is less robust as in the economic performance models. Concluding, indeed it seems that election quality has important consequences for electoral accountability.

6. Does election quality matter? Consequences for responsiveness

6.1. Consequences of variation in election quality for responsiveness

In section 5.1, we defined the function of elections as ‘instruments of democracy’ to be to generate accountability and responsiveness. Accountability was defined as citizen influence on government formation, asking to what degree unsatisfactory government performance before elections affects the fate of incumbents at the polls. Responsiveness was defined as citizen influence on government policies, asking to what degree governments formed after elections implement policies that are congruent with the policy preferences of voters. The same steps of the linkage mechanism connecting voter preferences to policy outcomes as outlined in the flowchart in chapter 5 (figure 5.1) are likely to be important here: preferences translate into votes at elections, votes translate into governments and governments implement policies. The various aspects that influence the quality of the electoral regime, i.e. election quality, the party system, the electoral system, the division of power and state capacity, should hence also strengthen or weaken the link between voters’ preferences and policy outcomes.

Figure 6.1. A schematic representation of elections as linkage mechanism: responsiveness



The existing literature on responsiveness, which is mostly focused at established democracies, has either looked at part of the linkage mechanism, studying the link between (1) citizens' votes and government composition (McDonald and Budge 2005) or (2) citizens' preferences and government composition (measured by congruence between citizens' ideological left-right positions and average government position) (Golder and Stramski 2010, Powell 2000); or looked at the linkage mechanism as a whole, studying the link between (3) citizens' preferences and government policy-making (measured by congruence between citizens' preferences for spending on policy domains and actual government spending on those domains, cf. Stimson et al. 1995; Soroka and Wlezien 2010). Figure 6.1 below shows the flowchart discussed in chapter 5, and adds the various responsiveness links that have been subject to empirical research thus-far.

Ideally, I would follow the third approach to assess to what degree elections generate a link between citizens' preferences and governments' policies, as it allows for an evaluation of the functioning of the electoral regime as a whole. However, given the limited availability of data on citizens' policy and ideological preferences in third wave regimes, I do not have direct measures of voters' policy preferences for all the countries in our sample. Certainly, data on citizen's preferences could be found in the CSES and ISSP surveys as well as in a number of regional Barometers, such as the Latinobarometer and the Afrobarometer. However, since these surveys do not provide the time and country coverage covered in this thesis, this would limit the analyses to a very small sub-set of the sample. An alternative approach, and the approach that I have preferred here, is to consider government performance, instead of responsiveness strictly defined, on a number of policy areas or public goods, for which we might reasonably assume citizens' preferences to be rather homogeneous. Examples of such government performance indicators are economic growth, human development, public spending on health and education, and respect for human rights and civil liberties. The advantage of such an aggregate analyses of policy outcomes is that this could also be considered as an

“objective” way of assessing the quality of the electoral regime, i.e. under which conditions does the electoral regime generate such desirable policy outcomes as human development, respect for human rights and civil liberties. Moreover, it also allows closer scrutiny of the “democratizing” effect of elections, as it allows us to investigate under what conditions election quality generates increased respect for civil liberties ex post the elections. Hence, I selected a set of government performance indicators for which we might reasonably assume citizens’ preferences to be rather homogeneous, and for which sufficient data was available to carry out aggregate analyses. These indicators are economic growth and respect for human rights and civil liberties¹²⁰. The dependent variable(s) are thus policy outcomes after elections, measured using the indicators just mentioned, and the question to be assessed empirically is to what degree the various aspects of the electoral regime, i.e. election quality, the party system, the electoral system, the division of power and state capacity, strengthen or weaken such policy outcomes. Setting up the analyses in this way implies also that, contrary to our analyses of accountability, we do not expect interactive effects: the various aspects of the electoral regime are expected to have direct effects on policy outcomes.

6.2. Hypotheses, data and methods

Table 6.1 gives an overview of the hypotheses. With respect to election quality, I expect the effect to be positive: high election quality is likely to generate better government performance after elections than low election quality. The features of the party system are likely to generate two opposing effects:

¹²⁰ The QoG dataset and the Worldbank Development Indicators also provide data on public spending on health and education and the human development index (an index that combines information about life expectancy, education and average income (UNDP 2004), which would have been very interesting indicators for government performance as well. However, for health and education spending data are missing for a large number of countries and years, leaving us with only about 25% of our sample, which is why I don’t include these as performance indicators. The human development index has less missing data but appeared stationary, i.e. practically non-changing over time, implying that I could only predict between country-differences, in which case the direction of causality would be difficult to assess (i.e. does election quality lead to higher levels of human development or the other way around). For these reasons I focus on the two performance indicators mentioned.

on the one hand, party systems that generate single-party governments are likely to be more effective in policy-implementation, on the other hand, party systems that generate coalition governments are likely to represent the interests of a broader part of the citizenry, which might strengthen responsiveness. Party systems that combine the best of these two might generate the best policy outcomes. Hence, I expect that if the party system is dominant or fragmented (as compared to two-party or moderate multi-party), this will be associated with lower government performance after elections. Moreover, apart from the number of parties or candidates I also expect the relative balance of power between incumbents and opposition to influence government performance after the elections. On the one hand, incumbents that have a weak power base or face a strong opposition might be less effective at implementing policies, resulting in worse government performance. On the other hand, incumbents that won elections with a large majority or face a weak opposition might have no incentives to be responsive to the needs of their citizens and hence might also score worse on our performance indicators. Again, governments that are somewhere in the middle, having a reasonably effective majority while facing a healthy opposition, might perform best. Hence, I expect that the margin of victory in elections (i.e. the difference between the vote/seat share of the incumbent and the vote/seat share of the largest opposition candidate or party) should have a curvi-linear effect on government performance, with governments' performing worse at very low and very high margins of victory¹²¹.

Considering the effects of the electoral system and institutional division of power, I expect power-sharing systems to generate better policy-outcomes (Lijphart 1999; Norris 2008). The literature on

¹²¹ In terms of measurement, for presidential elections the margin of victory refers to the difference in vote shares between the winning candidate and the largest opposition candidate, while for parliamentary elections the margin of victory refers to the difference in seat shares between the winning party (or coalition of parties if a coalition was formed) and the largest opposition party. I used seat shares for parliamentary elections since data on vote shares for parliamentary elections had a lot of missing values. The analyses are re-run for legislative and executive elections separately however and results are very similar, so it does not seem to affect the results.

comparative democratic institutions proposes that power-sharing systems should generate better policy outcomes because these systems allow for broader interest representation than power-concentrating systems, providing more opportunities for post-election bargaining and opposition influence on policy-making. This representation of diverging interests, power-sharing advocates believe, does not only strengthen the link between votes and government, but also strengthens the link between government and policy, as policy implementation is likely to be more effective when all stakeholders have been included in the decision-making process (Lijphart 1999, Norris 2008). Hence I expect proportional electoral systems to be associated with higher government performance after elections and I expect parliamentary systems, bicameral and federal systems as well as systems with more veto-points to generate better government performance after elections.

Finally, state capacity is likely to influence responsiveness directly because it determines the degree to which states actually have the financial means and organizational capacity to implement policies. As such, I expect responsiveness to be higher in states with higher levels of economic development and a more effective bureaucracy, and lower in states that derive a higher proportion of GDP from development aid and that have recently experienced civil war. Finally, since it is not unthinkable that policy outcomes might be influenced by many other variables than the ones taken into account in this analysis, a crucial control variable is the lagged value of the dependent variable in the year before the elections. In this way we ensure that any additional explained variance by factors such as election quality is robust. As further control variables I include the type of election (legislative or executive, constituent assembly elections, and concurrent elections), sequence number of the election, and region.

Table 6.1. Overview of hypotheses

Independent variables		Hypotheses
		Expected effect on policy outcomes
Election quality		Positive
Party system		
	Dominant party system	Negative
	Two-party or moderate multi-party system	Positive
	Fragmented party system	Negative
	Margin of victory (squared)	Negative
Electoral system		
	Majoritarian electoral system	Negative
Division of power		
	Parliamentarism	Positive
	Bicameralism	Positive
	Federalism	Positive
	Coalition government	Positive
	Number of veto-points	Positive
State capacity		
	Level of economic development	Positive
	Aid as proportion of GDP	Negative
	History of civil war	Negative
	Bureaucratic effectiveness	Positive
Control variables		
	Lagged dependent variable	-
	Type of election	-
	Election sequence number	-
	Region	-

The data used for the empirical analyses originate partly from the Quality of Government dataset from the University of Gothenburg (Teorell et al. 2011, version April 6th 2011) and partly from data collected by the author. The data for election quality are based on our aggregate variable as discussed in chapter 2. Data for which no precise indicators were available in the QoG dataset, such as coalition governments, the structure of the party system, incumbent vote and seat share and opposition vote and seat share, legislative electoral systems, and experiences with civil war were gathered by the author on the basis of information from the Electoral Handbooks on Africa, South and Central America, Central and East Asia and Europe edited by Nohlen (Nohlen 1999, 2001, 2005, 2010), as well as the data handbook by Rose and Munro (2003) on Central and Eastern Europe, and the Parline Database of the Inter-Parliamentary Union (IPU). Also, if more recent data were available in the original data-sources, as was the case for economic growth as measured by the World Banks World Development

Indicators, these data were added to the dataset. Appendix F presents an overview of the variables used.

There are three dependent variables: economic growth, respect for civil liberties and respect for human rights. Economic growth was measured with data from the World Bank's Human Development Indicators, and measures the percentage of annual growth in GDP per capita. Respect for civil liberties was measured using the data from the Freedom House civil liberties scale (Freedom House 2009)¹²². Since Freedom House data have received criticism in recent years, mainly related to lack of clarity in coding rules and changes in coding rules over time, as an additional test I also estimate models using respect for human rights as the dependent variable (these models are reported in Appendix F). Respect for human rights was measured with the political terror scale developed by Gibney, Cornett and Wood (2010) on the basis of Amnesty International reports and US State Department Human Rights reports. This scale classifies the level of human rights violations in a wide range of countries on a yearly basis¹²³. The human rights and civil liberties variables were recoded to

¹²² "Civil liberties allow for the freedoms of expression and belief, associational and organizational rights, rule of law, and personal autonomy without interference from the state. Countries are scored on a 7-point scale ranging from 1 (most free) to 7 (least free)" (QoG Codebook v6 April 2011: 42/43).

¹²³ Scores for this scale are assigned on a 5-point ordinal scale as follows:

- Level 1: Countries under a secure rule of law, people are not imprisoned for their view, and torture is rare or exceptional. Political murders are extremely rare.
- Level 2: There is a limited amount of imprisonment for nonviolent political activity. However, few persons are affected, torture and beatings are exceptional. Political murder is rare.
- Level 3: There is extensive political imprisonment, or a recent history of such imprisonment. Execution or other political murders and brutality may be common. Unlimited detention, with or without a trial, for political views is accepted.
- Level 4: Civil and political rights violations have expanded to large numbers of the population. Murders, disappearances, and torture are a common part of life. In spite of its generality, on this level terror affects those who interest themselves in politics or ideas.
- Level 5: Terror has expanded to the whole population. The leaders of these societies place no limits on the means or thoroughness with which they pursue personal or ideological goals." (QoG Codebook v6 April 2011: 46).

vary from low to high respect for human rights or civil liberties respectively. All these variables were measured taking the average value in the 2 years after elections, leaving a one-year gap between the elections and the year of measurement to limit endogeneity¹²⁴.

As for the independent variables, election quality is measured using our overall election quality variable that runs from 0 to 1, with 1 indicating high quality. To ease interpretation of the coefficients, I used the recoded election quality dummy that differentiates severely flawed and flawed elections from clean elections. The structure of the party system is a categorical variable differentiating dominant party systems (number of opposition parties/candidates < 2), two-party systems (2-3 opposition candidates/parties), multi-party systems (4-15 opposition candidates/parties), and fragmented party systems (16-28 opposition candidates/parties). Note that this variable is based purely on the number of opposition candidates and parties and does not take into account their vote or seat shares. This aspect is measured by the margin of victory variable, that measures the difference between the vote / seat share of the incumbent and the vote / seat share of the largest opposition party. The electoral system is a dummy distinguishing majoritarian (1) from proportional systems (0) in the case of legislative elections and absolute majority (1) from plurality systems (0) in the case of presidential elections. The division of power variables are all dummy variables indicating whether the political system is parliamentary, bicameral, federal and whether a coalition government was formed after the elections. We also included a quasi-continuous variable counting the number of veto-players in the political system, derived from the Database of Political Institutions, which runs from 1 to 8 (Keefer 2009). As indicators of state capacity, level of economic development was measured as GDP per capita in the

¹²⁴ Concretely, this means that if the election took place in year t , the dependent variables were measured as the average value at $t+2$ and $t+3$, i.e. 2 years and 3 years after the elections. Resulting Pearson correlations between election quality and human rights and civil liberties are 0.40 and 0.58 respectively (and 0.43 and 0.59 in the election year), indicating that they are related, as we should expect, but not sufficiently to suspect that they are measuring the same thing.

year the elections, aid was measured as the proportion of GDP consisting of official development aid in the year of the elections, and bureaucratic effectiveness was measured by the World Bank's government indicator of "government effectiveness" that measures the effectiveness of the bureaucracy. Civil war is a dummy variable indicating whether or not the country experienced civil war in the year of the elections. Finally, control variables were the type of the election (legislative or executive, constituent assembly, concurrent elections), election sequence number and geographical region.

6.3. Results

Table 6.2 shows the results for economic growth. The dependent variable is average GDP per capita growth in the 2nd and 3^d year after the elections. As set out in chapter 4, given the structure of the data, with elections being clustered within countries, I use multi-level models to correct for the clustering of observations within countries. However, since the observations within each country are over-time observations, additional tests are needed to ensure that there is no serial autocorrelation of errors and that the dependent variable is not stationary (Beck 2008). In addition, tests for heteroskedasticity are also important to make sure that the error variances are similar for all countries in the model. If this is not the case, using multi-level models to correct for clustering of observations is appropriate¹²⁵. None of the three dependent variables used in this chapter appear to be stationary. Results for tests of serial autocorrelation and heteroskedasticity will be reported when discussing the models. Now, multi-level models separate the explained variance of the dependent variable into a random and fixed component, where the random component provides an indication of the amount of

¹²⁵ Ideally, I would use time-series-cross-sectional methods to analyze the data, however, given the limited number of elections in a substantial part of the countries included in this dataset, as well as the variation in number of elections held in each country, time-series cross-sectional analyses could only be applied to a sub-set of the countries in our sample (those that had held at least 5 elections), and only to analyses of legislative and executive elections separately. For this reason, I use the more general multi-level models, following Beck's (2008) suggestions for testing the robustness of these models, and also run the final models using time-series-cross-section methods on legislative and executive elections separately to see if the results hold.

variance to be explained at the higher level, and the fixed component of the amount of variance explained at the lower level. This allows us to identify how much variance can be explained at either the country or the election level. Economic growth appears to vary mainly over time within countries: 77% of variance is at the election level, and 23% of variance at the country level. This implies that slow-changing variables such as institutional variables will most likely not have a very strong effect, while variables that change over time might explain economic growth better.

Turning to the models, model 1 until 5 present the results for each group of independent variables. Election quality appears to have a counter-intuitive effect on economic growth, at least when modeled without control variables: clean elections are associated with lower economic growth in the years after elections and fraudulent elections are associated with higher economic growth! The coefficient is not very large and election quality explains only 1% over over-time variation in growth, however, this finding is contrary to the positive effect we expected. Turning to party system characteristics, it appears that the hypothesis about a curvi-linear effect of margin of victory needs to be rejected as well. In analyses not reported here, it appeared that margin of victory was clearly linearly related to economic growth after the elections, so model 2 shows the result for margin of victory (instead of margin of victory squared). Here too, the coefficient runs in the opposite direction as expected: higher margins of victory (i.e. stronger incumbents and weaker opposition) are associated with higher levels of growth in the years after elections. These results do not change when extremely low and high values of margin of victory were excluded. Turning to the number of candidates or parties, this does not seem to affect economic growth much, only fragmented party systems appear to be associated with higher economic growth after elections (contrary to expectations). Nevertheless, also model 2 does not explain a large amount of variation in the dependent variable: only 1% of over-time variation is explained by party system characteristics.

Turning to electoral systems, we find the expected negative effect, however both coefficients are far from being significant. Model 4 present the findings for the institutional division of power, and here we find the expected positive effect of power-sharing institutions on government performance: the number of veto-players in the political system is positively associated with economic growth after the elections. All the other institutional variables do not reach statistical significance however, and the model as whole explains only 2% of over-time variation.

Considering model 5 then, civil war has a strong and negative effect on economic growth in the years after elections, which is in line with other research (Collier 2009). Also, bureaucratic effectiveness appears to be important for generating economic growth, as its effect is positive and highly significant. The level of economic development in the year of the elections appears to have a negative effect on economic growth in the years after the elections, which in fact is not as surprising as it might seem. It seems likely that this variable captures a ceiling effect of growth, i.e. countries that have a higher GDP are less likely to experience very high growth rates than poorer countries¹²⁶. The proportion of aid received does not appear to affect economic growth, even if the sample excludes the elections for which data on aid were imputed and even if outliers such as countries receiving extremely high or low proportions of aid were excluded. Since including the proportion of aid does lead some cases to drop out, I didn't include this variable in the final results reported here. Model 6 present the full model including the variables that were significant in earlier models.

¹²⁶ Also, since the dependent variable is measured as the annual *proportion* of GDP per capita increase or decrease, poorer countries need relatively lower levels of de facto GDP per capita increase to achieve a similar growth rate as richer countries (simply because the denominator is smaller).

Table 6.2. Explaining economic growth in 2 & 3 years after elections (average)

	Model 1 Election quality	Model 2 Party system	Model 3 Electoral system	Model 4 Division power	Model 5 State capacity	Model 6 Full model
Election quality (0-1)	-0.734+ (0.402)					-0.898* (0.429)
Margin of victory		0.015* (0.007)				0.012 (0.007)
Two-party systems ^a		0.515 (0.759)				0.225 (0.738)
Multi-party systems		0.546 (0.722)				0.307 (0.704)
Fragmented party systems		2.017+ (1.066)				1.712+ (1.037)
Electoral system maj leg			-0.502 (0.546)			
Electoral system abs maj ex			-0.476 (0.413)			
Parliamentary system				0.380 (0.802)		
Bicameralism				-0.007 (0.681)		
Federalism				0.456 (0.640)		
Coalition government				-0.451 (0.465)		
Number of veto players				0.411** (0.144)		0.518*** (0.145)
Economic development (gdp/cap)					-0.0004*** (0.00009)	-0.0004*** (0.00009)
Civil war					-2.454*** (0.726)	-2.336** (0.723)
Bureaucratic effectiveness					1.679** (0.578)	1.778** (0.604)
Constant	2.556*** (0.364)	1.102 (0.797)	2.454*** (0.366)	1.141* (0.551)	4.296*** (0.558)	2.794** (1.019)
N level 1	783	783	783	782	783	783
N level 2	93	93	93	93	93	93
Variance explained level 2 (country)	0%	1%	0%	0%	0%	0%
Variance explained level 1 (election)	1%	0%	0%	2%	4%	6%
Total variance explained	1%	1%	0%	2%	4%	6%

Standard errors in parentheses. P-values: ^ 0.1, * 0.05, ** 0.01, *** 0.001. a. Reference category is dominant party systems.

Election quality remains significant and negative, and the effects of fragmented party systems, number of veto-players, economic development, civil war and bureaucratic effectiveness remain significant as well. Margin of victory is practically significant one-tailed ($p = 0.104$), so I included this variable in further models as well. Now, even if these models do not explain a large amount of variation in economic growth, the negative effect of election quality is contrary to what we would expect and contradicts assumptions about the positive consequences of clean elections. This might be a spurious effect however. Hence, before jumping to conclusions, it is important to test how robust these findings are if control variables are included and if we check and control for potential time-series-cross-section data problems such as serial autocorrelation and panel heteroskedasticity.

Model 7 in the table on the next page shows the results if controlling for type of election, election sequence number and region. The coefficient for election quality does not seem to be affected very much by the inclusion of these controls, it remains negative and significant. The party system characteristics are no longer significant, both margin of victory and fragmented party systems drop out in this full model with controls. The effect of veto-players is also substantially reduced, it is still significant but the coefficients size is almost halved. The coefficient size for economic development also declines when including these controls, but the effect continues to be highly significant. The effects for civil war and bureaucratic effectiveness remain significant and about as strong as in earlier models. Considering the controls, the type of election does not seem to affect post-election performance much. However, election sequence appears to have a positive and significant effect on economic growth in the years after elections. This is in line with findings by Chauvet and Collier (2009) that countries having experienced a higher number of elections tend to perform better economically. Finally, several regional dummies are also significant in expected directions, with Sub-Saharan Africa and to a lesser extent South and Central America experiencing lower economic growth than Central and Eastern Europe (the base category). Model 8 shows the result for a sample that only

includes those countries that experienced at least 5 elections, as estimating the multi-level model on a sample with more balanced panels should give more robust results. The results are very similar to those presented in model 7, indicating that including also the countries with less than 5 elections in our analyses is not problematic.

I then tested for serial autocorrelation using the residuals of model 7, which demonstrated that indeed the model suffers from correlated errors over time. Beck (2008) proposes to deal with serial autocorrelation by including a lagged dependent variable. This makes sense substantively as economic growth during the elections might very well be related to economic growth after the elections, and, even if changes over time are also explained by other factors, we should expect some continuity over time. Hence, model 9 reports the result when the lagged dependent variable is included. This model no longer suffers from serial autocorrelation implying that the results are more robust. It appears that all independent variables that were significant before remain significant and in the same direction, except election quality which is now insignificant. We subsequently tested for heteroskedasticity and found no significant differences in error variance between countries, except one outlier: Equatorial Guinea. Excluding this country from the model does not significantly alter the results however. As a final check of the results, I ran the full model on a sample with only legislative and only executive elections, which enables the use of time-series-cross-section methods. Models 10 and 11 show the results.

Clearly, also in these more robust models election quality is no longer significant, even though its coefficient is still negative. The models were re-run (a) using a sample of countries with at least 5 elections, (b) including the lagged dependent variable –to correct for serial autocorrelation, and (c) using fixed effects –to correct for heteroskedasticity- and in all models election quality remained

insignificant¹²⁷. Concluding, election quality does not have an effect on government performance if measured by economic growth after elections. If anything, the effect seems to be negative, even though in more comprehensive and robust analyses it becomes clear that this is a spurious effect: election quality is not significantly related to economic growth. Of course economic growth is a very crude measure of government performance. More analyses would need to be done on more precise indicators of government performance relating more directly to public service provision (public spending on health, education, etc.) and public service functioning (absence of corruption, etc.). At present however, data availability was too limited to analyze the effects of election quality on such aspects. For now, let's turn to the other indicator for government performance: respect for civil liberties.

¹²⁷ Finally, I reran these models using a different measurement of economic growth (the indicator collected by the United Nations Statistics Division in the QoG dataset (Teorell et al. 2011), and found identical results as the ones reported here.

Table 6.3. (continued) Explaining economic growth in 2 & 3 years after elections (average)

	Model 7 Pars model w controls	Model 8 Pars model T > 5	Model 9 Pars model lagged DV	Model 10 Pars model legislative	Model 11 Pars model executive
Election quality (0-1)	-0.784+ (0.422)	-0.751+ (0.435)	-0.394 (0.409)	-0.447 (0.548)	-0.872 (0.714)
Margin of victory	0.011 (0.007)	0.011 (0.008)	0.007 (0.007)	0.007 (0.009)	0.045** (0.014)
Two-party systems ^a	0.046 (0.722)	0.102 (0.756)	0.109 (0.693)	0.529 (1.099)	0.138 (1.144)
Multi-party systems	-0.054 (0.694)	-0.014 (0.729)	-0.083 (0.664)	0.085 (1.053)	0.860 (1.080)
Fragmented party systems	1.218 (1.030)	1.458 (1.078)	1.466 (0.987)	1.482 (1.337)	2.313 (2.749)
Number of veto players	0.256+ (0.153)	0.249 (0.157)	0.175 (0.147)	0.192 (0.193)	0.138 (0.260)
Economic development (gdp/cap)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001** (0.000)
Civil war	-2.083** (0.715)	-1.921* (0.778)	-1.857** (0.679)	-0.787 (0.898)	-3.929*** (1.129)
Bureaucratic effectiveness	2.026* (0.789)	1.949* (0.893)	2.140** (0.698)	1.365+ (0.720)	1.972* (0.974)
Election type	0.011 (0.327)	0.040 (0.337)	0.052 (0.315)		
Constituent assembly elections	-0.353 (1.044)	-0.742 (1.139)	0.409 (1.005)	-0.341 (1.072)	
Concurrent elections	-0.354 (0.510)	-0.511 (0.533)	-0.361 (0.483)	-0.445 (0.596)	0.456 (0.845)
Election sequence	0.404*** (0.070)	0.415*** (0.072)	0.305*** (0.067)	0.230** (0.085)	0.385*** (0.116)
Lagged dependent variable			0.171*** (0.019)		
Region - SSA ^b	-3.674** (1.359)	-4.097** (1.541)	-3.264** (1.195)	-3.022* (1.186)	-3.178+ (1.702)
Region - SA	-3.385* (1.451)	-3.533* (1.575)	-2.981* (1.267)	-2.984* (1.239)	-2.633 (1.654)
Region - CA	-3.564* (1.638)	-3.709* (1.765)	-3.170* (1.430)	-2.991* (1.405)	-3.030+ (1.831)
Constant	6.300*** (1.449)	6.522*** (1.558)	6.046*** (1.330)	5.088** (1.650)	3.307 (2.099)
N level 1	783	745	780	471	312
N level 2	93	81	93	93	79
Variance explained (level 2 – country level)	0%	0%	0%		
Variance explained (level 1 – election level)	11%	12%	17%		
Total variance explained	11%	12%	17%	7%	12%

Standard errors in parentheses. P-values: ^ 0.1, * 0.05, ** 0.01, *** 0.001. a. Reference category is dominant party systems. b. Central and Eastern Europe is the reference category, coefficients for FSR and SE not reported since insignificant in all models. c. Coefficients for country dummies not reported.

Table 6.3 shows the results of models predicting the average level of respect for civil liberties in the 2nd and 3^d year after the elections. Model 1 shows the effect of election quality on respect for civil liberties in the years after the elections: the effect is positive and highly significant as expected. Clean elections are associated with increased respect for civil liberties after the elections, and vice versa. Moreover, election quality appears to explain a substantial amount of variance, though mostly at the country level. Turning to party system characteristics, as in the previous models, margin of victory appeared to be linearly related to the dependent variable rather than curvi-linear as hypothesized: apparently, relatively stronger incumbents are associated with higher levels of civil liberties in the years after elections. The number of opposition parties and candidates does seem to affect government performance as expected: levels of civil liberties are somewhat higher in two-party and multi-party systems, however these effects do not reach conventional levels of statistical significance. Model 3 shows that the effects of legislative and presidential electoral systems are both in the expected direction and significant. However in terms of explained variance the electoral system only explains about 3% of variation in post-election levels of civil liberties. Turning to model 4, all institutional variables are in the expected positive direction: power-sharing systems are associated with higher levels of civil liberties. However, only parliamentarism and the number of veto-players reach statistical significance.

Note that our control variable for missing data on the number of veto-players is also significant in this model, however the results do not change when excluding the elections for which veto-players data was missing, and this variable is also no longer significant in a full model (model 6), so the results for

veto-players appear robust. Turning to model 5 that shows the results for state capacity, it appears these variables explain an important amount of variance in post-election levels of civil liberties. The effects of economic development, civil war and bureaucratic effectiveness are in line with our expectations: higher levels of economic development and bureaucratic effectiveness are associated with higher civil liberties, while civil war is associated with lower civil liberties. The effect of aid is contrary to expectations, though probably good news for donors: countries that receive higher proportions of aid in the year of elections tend to have higher levels of respect for civil liberties in the years after.

Finally, model 6 shows the result of a full model including all independent variables that were significant or close to one-tailed significance in the previous analyses. The effect of election quality continues to be positive and significant, though the coefficient size is diminished somewhat by the inclusion of other explanatory variables. The effects of margin of victory also appears stable. The electoral system and parliamentarism variables are no longer significant in a full model, however the positive effect of the number of veto-players remains significant. Finally, the state capacity variables all remain significant in the full model. How robust are these results to the inclusion of controls and robustness checks? Table 6.5 shows the results of these checks.

Table 6.4. Explaining respect for civil liberties in 2 & 3 years after elections (average)

	Model 1 Election quality	Model 2 Party system	Model 3 Electoral system	Model 4 Division power	Model 5 State capacity	Model 6 Full model
Election quality (0-1)	0.514*** (0.069)					0.357*** (0.066)
Margin of victory		-0.006*** (0.001)				-0.005*** (0.001)
Two-party systems ^a		0.118 (0.121)				
Electoral system maj leg			-0.241* (0.099)			-0.102 (0.088)
Electoral system abs maj ex			-0.144* (0.068)			-0.099 (0.062)
Parliamentary system				0.502* (0.200)		0.066 (0.162)
Bicameralism				0.153 (0.265)		
Federalism				0.150 (0.118)		
Coalition government				0.045 (0.077)		
Number of veto players				0.154*** (0.024)		0.113*** (0.022)
Veto players missing				-0.225* (0.101)		-0.123 (0.095)
Economic development (gdp/cap)					9.91 ^{e-05***} (0.00001)	7.16 ^{e-05***} (0.00001)
Aid received as % of GNI					0.012*** (0.003)	0.011*** (0.003)
Civil war					-0.458*** (0.119)	-0.399*** (0.114)
Bureaucratic effectiveness					1.337*** (0.135)	1.127*** (0.129)
Constant	4.162*** (0.132)	4.510*** (0.184)	4.488*** (0.146)	3.848*** (0.179)	4.508*** (0.126)	4.291*** (0.148)
N level 1	763	763	763	763	763	763
N level 2	93	93	93	93	93	93
Variance explained (level 2 – country level)	18%	7%	3%	16%	54%	61%
Variance explained (level 1 – election level)	1%	0%	0%	1%	1%	3%
Total variance explained	19%	7%	3%	17%	55%	64%

Standard errors in parentheses. P-values: ^ 0.1, * 0.05, ** 0.01, *** 0.001. a. Reference category is dominant party systems, coefficients multi-party and fragmented party systems not significant, hence not reported in model.

Model 7 reports the results if controls are included for the type of elections, election sequence number and region. The inclusion of these controls does not seem to affect our findings much, with the exception of the coefficients for the number of veto-players and economic development that are smaller than in previous models. As for the control variables, the type of election does not seem to affect post-election civil liberties much. Election sequence number has a consistent positive effect, also in later models, indicating that average levels of civil liberties increase as successive elections are held. Of the regional dummies, Former Soviet Republics and Southern European countries appear to have lower levels of civil liberties than could be accounted for by the explanatory variables in our model¹²⁸. Model 8 reports the same model using a smaller sample including only those countries that experienced 5 or more elections, as this should give more robust estimates. The results are very similar to the results reported in model 7, indicating that using the full sample in our analyses is not problematic. Turning to model 9, this model reports the results when estimating the model using fixed effects, i.e. including country dummies to correct for heteroskedasticity. Such a model only estimates over-time variance, and hence non-changing variables like bureaucratic effectiveness drop out of the model¹²⁹. Continuing, I tested for serial autocorrelation, which resulted to be a problem in these data, and hence I corrected the model by including a lagged dependent variable. The results are reported in model 10 (including the lagged dependent variable indeed solved the serial autocorrelation).

¹²⁸ Note that in fact, the Southern European third wave regimes have on average the highest levels of civil liberties as compared to other regions. However, given the high levels of economic development and election quality in this region, our model would predict the level of civil liberties to be even higher, hence the negative coefficient, which in this case merely indicates that there is something about the 3 Southern European regimes that explains their relatively lower-than-expected level of civil liberties, that is not included as a variable in our model.

¹²⁹ Heteroskedasticity did not appear to be very problematic in these models, as only 3 countries turned out to have very high error variance, i.e. Belize, Cape Verde, and Serbia & Montenegro. Excluding these 3 countries from the analyses did not change the results. As an additional test of the robustness of results to potential heteroskedasticity however, I report the fixed effects model.

Table 6.5. (continued) Explaining respect for civil liberties in 2 & 3 years after elections (average)

	Model 7 Pars model w controls	Model 8 Pars model T > 5	Model 9 Pars fixed effects ^c	Model 10 model lagged DV	Model 11 Pars model legislative	Model 12 Pars model executive
Election quality (0-1)	0.365*** (0.065)	0.381*** (0.067)	0.315*** (0.066)	0.154* (0.061)	0.424*** (0.091)	0.428*** (0.109)
Margin of victory	-0.004*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.002+ (0.001)	-0.005** (0.002)	-0.004* (0.002)
Number of veto players	0.071** (0.024)	0.074** (0.024)	0.068** (0.024)	0.009 (0.021)	0.099** (0.031)	0.144*** (0.037)
Economic development (gdp/cap)	3.68 ^{e-05*} (1.72 ^{e-05})	4.17 ^{e-05*} (1.75 ^{e-05})	5.94 ^{e-05**} (1.93 ^{e-05})	8.30 ^{e-08} (1.43 ^{e-05})	6.1 ^{e-05**} (1.78 ^{e-05})	4.23 ^{e-05} (2.87 ^{e-05})
Aid received as % of GNI	0.013*** (0.003)	0.012*** (0.003)	0.012*** (0.003)	0.007** (0.003)	0.010** (0.004)	0.014** (0.005)
Civil war	-0.374*** (0.112)	-0.294* (0.122)	-0.309** (0.118)	-0.070 (0.101)	-0.413** (0.154)	-0.552** (0.176)
Bureaucratic effectiveness	1.239*** (0.139)	1.171*** (0.149)		0.764*** (0.094)	1.083*** (0.135)	1.160*** (0.169)
Election type	-0.049 (0.050)	-0.049 (0.051)	-0.046 (0.050)	-0.027 (0.046)		
Constituent assembly elections	0.102 (0.158)	0.094 (0.171)	0.123 (0.160)	0.214 (0.148)	0.022 (0.173)	
Concurrent elections	0.164* (0.080)	0.166* (0.083)	0.180* (0.087)	0.112 (0.068)	0.064 (0.101)	0.209 (0.136)
Election sequence	0.051*** (0.011)	0.047*** (0.011)	0.043*** (0.012)	0.045*** (0.010)		
Lagged dependent variable				0.503*** (0.031)		
Region – FSR ^b	-0.655* (0.283)	-0.619* (0.289)		-0.464** (0.175)	-0.532* (0.270)	-0.640* (0.325)
Region - SSA	-0.374 (0.241)	-0.209 (0.256)		-0.336* (0.151)	-0.281 (0.229)	-0.338 (0.299)
Region - SE	-0.927* (0.420)	-0.820* (0.416)		-0.612* (0.254)	-0.826* (0.383)	-0.164 (0.645)
Constant	4.523*** (0.216)	4.434*** (0.222)	3.429*** (0.279)	2.478*** (0.202)	4.508*** (0.232)	4.412*** (0.305)
N level 1	763	724	763	749	455	308
N level 2	93	80	93	93	92	79
Variance explained (countries)	64%	61%	-	75%	-	-
Variance explained (elections)	3%	4%	-	6%	-	-
Total variance explained	67%	65%	80%	81%	64%	65%

a. Reference category is dominant party systems. b. Central and Eastern Europe is reference category, coefficients SA and CA not reported as insignificant in all models.

c. Coefficients for country dummies not reported.

In this model the effects of economic development and civil war are no longer significant. Election quality remains significant and positive, however its coefficient size is considerably diminished. Apparently part of the effect of election quality in the previous models was driven by the level of civil liberties in the election year, rather than by the quality of the electoral process per se. Nevertheless, even controlling for the level of civil liberties in the election year, the quality of the elections still has an independent, additional effect on post-election civil liberties, an important finding for proponents of the “democratizing” effects of elections. Finally, I estimated the full model using time-series-cross-sectional methods on separate samples of legislative and executive elections, as reported in model 11 and 12. The separate models for legislative and executive elections were re-run with (a) using a sample of countries with at least 5 elections, (b) including the lagged dependent variable –to correct for serial autocorrelation, and (c) using fixed effects –to correct for heteroskedasticity- and in all models election quality remained significant and positive. Again, also in these separate models of legislative and executive elections, inclusion of the lagged dependent variable reduces the coefficient size of election quality by about half, but a significant and positive effect remains even under these stricter conditions. Clearly, even if clean elections do not enhance post-election economic growth, they do have a real and substantial positive effect on respect for civil liberties in the years after elections. Analyses using respect for human rights in the years after elections as dependent variable re-iterate this finding, hence these results are not driven by our choice of dependent variable (see Appendix F for the human rights models).

6.4. Conclusion

In this chapter the aim was to assess consequences of variation in election quality for responsiveness, operationalized here as government performance in the years after the elections. Government performance was measured using two indicators: economic growth in the years after elections and respect for civil liberties. With respect to economic growth, election quality appeared to have no effect

at all on post-election growth. Institutional division of power, measured as the number of veto-players affected economic growth positively, and state capacity, measured as bureaucratic effectiveness affected growth positively. Civil war had the expected negative effect on post-election growth rates. Nevertheless, the models explained a relatively low proportion of variance in post-election economic growth, and election quality remained insignificant across models. Clearly, even if election quality increases the possibilities for citizens to hold incumbents accountable for economic decline prior to the elections, clean elections do not lead to higher economic growth after elections. Naturally, economic growth is a crude government performance indicator and hence future research should investigate whether election quality affects other governance indicators, such as public service provision or public service functioning.

Turning to our alternative government performance indicator, respect for civil liberties in the years after elections, this appeared to be strongly and positively influenced by election quality. Even if respect for civil liberties is quite persistent over time, in models including the lagged dependent variable, election quality still exerted an independent additional effect. Hence, clean elections are associated with higher post-election respect for civil liberties as compared to fraudulent elections. In addition to the effect of election quality, state capacity variables as bureaucratic effectiveness and the proportion of aid received appear to strengthen post-election respect for civil liberties as well. Clearly, even if clean elections do not enhance post-election economic growth, they do have a real and substantial positive effect on respect for civil liberties in the years after elections.

7. Coming full circle: election quality and democratization

At the time of writing this conclusion, Egyptians were getting ready for their first presidential elections with multiple candidates in over 30 years. After the post-1974 third wave of democratization, and the post-1989 fourth wave, the Arab spring might come to constitute the fifth wave of transitions away from authoritarianism. At the same time however, just a couple of months earlier, fourth wave president Joseph Kabila managed to stay in power in Congo after an election that was so deeply flawed and so obviously rigged that it could hardly be called an exercise in democracy. Nor is Kabila alone in knowing his way around the toolbox of electoral fraud as recent elections in Russia, Belarus and Cote d'Ivoire have shown.

Clearly, while holding elections has become the global norm in the vast majority of countries around the world, the quality of these elections varies strongly, ranging from free and fair elections with genuine contestation between parties or candidates to façade elections that are marred by manipulation and fraud. While election quality, or conversely electoral fraud, has been a topic of increasing concern among policymakers and practitioners in the field of democratic governance, academics have been relatively late to pick up and only very recently begun to study electoral fraud. As such, existing research on electoral fraud mostly consists of conceptual work and data-collections on election quality, as well as historical and case-study research. However, while excellent historical studies and case studies have been carried out on this topic, comparative research on electoral fraud is still relatively scarce (with the exception of Birch 2011).

In an attempt to contribute to this literature, this thesis has addressed the following questions:

1. How to conceptualize and measure electoral fraud?
2. What are the causal dynamics that explain variation in electoral fraud?

3. What are the consequences of electoral fraud for the degree to which elections generate government accountability and responsiveness?
4. What are the consequences of electoral fraud for broader democratization processes?

The conceptualization of electoral fraud was grounded both in academic work as well as international legal conventions on human and political rights and proposed a working definition including four elements: elected officials, periodic elections, participation and competition. A distinction was further made between *de jure* and *de facto* participation and competition, where *de facto* participation and competition includes the often noted requirements of “free and fair” elections. The criteria of elected officials, periodic elections and *de jure* participation and competition were considered as indicating differences *in kind* between elections, allowing the differentiation of ‘clearly non-democratic elections’ from ‘potentially democratic elections’. Conversely, the criteria of *de facto* participation and competition rather indicate differences *in degree* between elections, allowing for gradual variation between elections from low to high quality. Since election quality only varies in the group of ‘potentially democratic elections’, this thesis focuses on elections that meet the minimal criteria for ‘potentially democratic elections’, i.e. elected officials, periodic elections and *de jure* participation and competition. Concretely, this means elections were included if: (a) taking place in an independent nation-state, (b) elections were held for both legislative and executive (directly or indirectly), (c) elections were held periodically, (d) elections were irreversible (elected office-holders could take office), (e) it was legally possible for multiple parties to participate in elections (i.e. *de jure* multi-party elections), and (f) it was legally possible for all adult citizens to participate in elections (i.e. *de jure* universal suffrage) (see chapter 2 and Appendix A and B for conceptualization and sample selection).

The final sample of elections spans a total of 885 elections in 97 third and fourth wave regimes in Southern Europe, Central and Eastern Europe, the Former Soviet Union, Sub-Saharan Africa, South

America and Central America, having taken place between 1974 and 2009¹³⁰. For each of these 885 ‘potentially democratic elections’, data was collected on the quality of the election, based on the degree to which the elections allowed for *de facto* participation and competition of parties/candidates and voters. Data was partly collected from existing datasets (i.e. Anglin 1998, Van de Walle 2003, Lindberg 2006, Birch 2008/2011, Hartlyn et al. 2008, Donno 2009, and Kelley and Kiril 2010), and partly collected by the author (missing elections and elections on which existing datasets strongly disagreed). Each election thus received an election quality score indicating whether the election was low, medium or high quality (see chapter 2 and Appendix C for measurement)¹³¹.

Based on the conceptualization and measurement of electoral fraud, the thesis subsequently addressed the question how electoral fraud is distributed and has developed over time in third and fourth wave regimes from 1974-2009 (see chapter 3). According to our data, about 50% of the elections in our sample were clean, while about 30% of elections were moderately flawed and another 20% severely flawed. Regional variation in election quality matched our expectations: election quality is highest in the third and fourth wave regimes of Southern Europe and South America, followed by Central and Eastern Europe and Central America. In all these regions, average election quality lies between 0.73 and 0.95, indicating on average high quality elections. On the other hand, the third and fourth wave regimes in Sub-Saharan Africa and especially the Former Soviet Republics have substantially lower

¹³⁰ Asian third and fourth wave regimes were excluded due to limited data availability on election quality in the cross-national datasets used for this thesis.

¹³¹ The low, medium and high quality scores correspond to how widespread irregularities were during the election (see chapter 2 and Appendix C), as follows:

- Low quality = widespread irregularities, whether intentional or not, that seriously affected the conduct of the elections in a large part of the country;
- Medium quality = considerably frequent irregularities, but not so widespread that the election was outright rejected by observers;
- High quality = no irregularities or some isolated, non-frequent irregularities.

Elections were scored on a scale running from 0-1, where low quality = 0, medium quality = 0.5 and high quality = 1.

quality elections, with average election quality varying between 0.34 and 0.53. Considering the development of election quality over time, in the sample as a whole there seemed to be a slight global trend of decreasing election quality between 1974 and 2009. However, closer analysis revealed that this apparent trend was generated by the changing composition of the sample of third and fourth wave regimes. In fact, strong differences were found between regions in development of election quality over time, and even when considering trends in election quality over time within regions, variation appeared to be much greater than expected. Coding trends within countries over time, the dominant pattern is still positive: in almost half of the 97 regimes (i.e. 48%), election quality was either high from the start and remained high (20%), started at medium quality and gradually improved over time (12%), or followed a less predictable pattern, fluctuating between medium and high quality (16%). However, a smaller but substantial proportion of regimes appears to be stuck between moderately flawed and severely flawed elections, that do not seem to be improving over time as successive elections are being held (i.e. 26%). While trends of consecutively flawed elections are relatively uncommon (5%), almost one tenth of regimes held elections that oscillate between low and medium quality (9%), and another 12% started off with relatively clean elections, only to see election quality deteriorate over time. Hence, even if elections are moments of “institutionalized uncertainty” that provide a window of opportunity for change, in over a fourth of the regimes studied in this thesis elections apparently did not create such a momentum for change. Finally, apart from the overall positive trends in about half of the regimes in our sample and negative trends in about a fourth, in slightly over 20% of regimes election quality did not appear to follow any trend over time at all. Instead, elections varied from high quality elections with no or minor irregularities to severely flawed elections with numerous irregularities, without a clear pattern emerging¹³². Hence, the question that

¹³² In terms of regional variation in trends, in Southern Europe, South America and Central America, election quality trends are predominantly positive (between 83% and 100%), with continuously high levels of election quality, gradual improvement over time, or variation between medium and high election quality constituting the most common trends. Positive trends are also most common in Central and Eastern Europe, however to a much

remains is why in some regimes election quality appears to be at a sustainable level of high quality, while in other regimes elections remain stuck at low or medium quality, and yet in other regimes we find tremendous variation in election quality from one election to the next?

In chapter 4 addresses this question by studying the causal dynamics explaining variation in electoral fraud. Attempting to explain differences in “getting elections right” both between countries as well as over time lead to a number of findings. First of all, in terms of over-time development of election quality, it appeared learning dynamics were less important than expected. Even though previous experience with democracy was associated with higher levels of election quality, election quality does not appear to improve over time as successive elections are held. At least this is not a pattern that holds cross-nationally (it does hold for a subset of countries, as illustrated in chapter 3). Rather, election quality appears to have a degree of persistence over time: election quality in previous elections affects election quality in the next elections. While this indicates a lock-in mechanism, the effect runs both ways: clean elections reinforce future clean elections, but rigged elections also increase the likelihood of future election fraud. Fortunately, this effect is certainly not deterministic, as other factors appeared to affect election quality as well, generating potential for change over time.

Testing the effects of incentives and disincentives to engage in electoral fraud, I found that (a) factors shaping the costs of electoral fraud (or rather: risk of detection), (b) the stakes of the electoral race, and (c) the benefits of winning the elections were very important for explaining variation in the quality of elections. Starting with factors that increase the likelihood of detection of fraud, independence of the media and the judiciary in the years prior to elections appeared to be strongly associated with election

lesser extent (59%). However, in the Former Soviet Republics development of election quality over time is often negative (58%) or trendless (33%). In Sub-Saharan Africa all trends can be found: over a third of countries experience positive trends, while another third experiences negative trends, and a little over 20% are stuck in trendless variation between low and high quality elections.

quality. Moreover, globalization or linkage, measured as the intensity of economic, social and political exchanges a country has with other countries, appears to have a consistently positive effect on election quality as well, which seems to indicate that costs of electoral fraud are higher if regimes have a stake in maintaining their international reputation¹³³. Apart from costs of fraud and risks of detection, factors affecting the stakes of the electoral race appear to be important in explaining election quality as well. These factors are not only institutional, but can reflect broader historical experiences and social structure. As such, a recent history of civil war negatively affects election quality, as do institutions that increase the zero-sum nature of the electoral race, such as presidentialism and majoritarian electoral systems. Social fractionalization appears to increase the zero-sum nature of elections as well, with especially religious fragmentation leading to lower election quality. Finally, factors that increase the benefits of winning in elections, such as the presence of natural resources, add to the probability of electoral fraud occurring. Finally, a surprising finding that future research will have to validate was that coup d'etats tend to lead to higher quality elections a posteriori.

International factors such as election assistance, the presence of international election observers and trade and aid relationships appeared to be much less robust in explaining election quality, which is in line with other research on electoral fraud (Birch 2011). However some of these results may also have been due to selection effects, the limited quality of the data available, and perhaps even because the impact of some international factors might be different for different regimes (Levitsky and Way 2010).

¹³³ However, it is not entirely clear whether this effect is indeed due to linkage or rather to economic development, which is strongly correlated to this measure (even if it drops out in the full statistical model). In the latter case, the positive effect on election quality would be generated not by concerns for a regime's international reputation but rather because economic development raises the costs of fraud. Moreover, it seems likely that the effects of linkage differ depending on the countries or regions linkage is strongest with. The growing influence of non-democratic regimes like China in for example Sub-Saharan Africa might generate important cross-pressures to the influence of democratic regimes such as the United States or Europe, weakening the aggregate effect found here.

In the case of international election observers for example, the non-significant effect might very well be due to a selection effect (observers being sent more often to elections that are expected to experience irregularities). In the case of aid and trade relationships and the effects of linkage and leverage (Levitsky and Way 2010), much depends on which countries these relationships are with and which counter-pressures a regime experiences as a result. For example, if a regime's economic and political exchanges are predominantly with non-democratic regimes, as is the case with for example Belarus, aid, trade and other international ties might have no effect or even a negative effect on election quality. On the other hand, if a regime's international exchanges are both with non-democratic and democratic regimes, as is the case with for example DR Congo, this might generate cross-pressures that result in a zero net effect. More precise data would be needed to disentangle these effects. However, in addition to better data, more precise causal reasoning should on the one hand clarify the different mechanisms of international influence (i.e. reputation, learning, conditionality, etc.), and on the other hand specify how these mechanisms might play out differently for different regimes.

Another question that remains to be addressed by future research is, building on the found effects of independent media and judiciary, the effects of domestic checks and balances on electoral fraud. While the analyses in this thesis found a limited impact of factors such as the strength and fragmentation of opposition parties, this might very well be due to the difficulty of measuring opposition strength in aggregate comparative research as carried out here. Of course, many things change between elections and especially the popularity of political actors and parties. Hence an opposition party or candidate that might have done well in the previous elections (which is how opposition strength was measured in this thesis) might have lost its support in the meantime, while a new party or candidate might have arisen that poses a serious challenge to the incumbent party or candidate. In cross-national datasets, precise data on these dynamics are difficult to find, however it seems quite likely that strong and

unified opposition helps in deterring (or at least detecting) fraud, as the example of Tsvangirai in Zimbabwe shows. Other important domestic checks and balances limiting electoral fraud may be domestic observers, an aspect on which good data is also still lacking.

Nevertheless, these results do suggest that policymakers seeking to promote clean elections might want to consider investing in stimulating the availability of independent media and strengthening the independence of the judiciary well in advance of the elections, rather than confining efforts to election monitoring and assistance. Also, in addition to stimulating independent media and judiciary as a way of strengthening domestic checks and balances, other initiatives aimed at creating stronger domestic checks and balances might enhance election quality as well (such as domestic monitoring groups, opposition parties, civil society ‘watch-dog’ organizations that monitor transparency and corruption more generally, etc.). Finally, institutional reforms that lessen the zero-sum nature of the electoral race (such as more proportional electoral systems) as well as initiatives to reduce social polarization (such as post-conflict reconciliation after civil war) might limit electoral fraud. Of course, all such initiatives require more long-term investment (though not necessarily more costly), however the consequences for the quality of elections and broader democratization are likely to be more sustainable. For academics, these results underscore the importance of the presence of civil liberties and rule of law as preconditions for elections to be “free and fair” and reassert the claims by Elklit (1999) and O’Donnell (2001) about the interdependence between democratic elections and civil liberties¹³⁴.

¹³⁴ As Elklit notes “the presence of, respect for, and unhindered use of relevant political and civil rights and freedoms” [...] are prerequisites which must be in place, in other words must be institutionalized, before we can even think of elections being potentially free and fair.” (Elklit 1999: 33). O’Donnell also notes that: “I do think that fair elections are extremely important. This is not because such elections will necessarily lead to wonderful outcomes. It is because these elections, per se and due to the political freedoms that must surround them if they are to be considered fair, mark a crucial departure from authoritarian rule” (O’Donnell 2001: 9).

Apart from the causes of electoral fraud, this thesis also investigated the consequences of electoral fraud. Chapters 5 and 6 address the question to what extent fraud undermines the functioning of elections as ‘instruments of democracy’ (Powell 2000). According to democratic theory, elections should generate governments that can be held accountable and that are responsive to citizens’ interests. Engaging in electoral fraud might be a very effective way for incumbents to avoid being held accountable for their past performance and incumbents that came to power through fraudulent elections might have little incentive to be responsive to citizen’s needs.

As regards accountability, indeed clean elections were found to strengthen electoral accountability while accountability was absent in fraudulent elections: economic decline in the years prior to election was strongly associated with incumbent turnover only in clean elections (see chapter 5). The effect is quite strong, and given the varying results found in previous research on economic electoral accountability, this suggests that election quality may have been a crucial omitted variable in these studies. Of other factors found to be important in research on electoral accountability in established democracies, the availability of alternatives appeared to play a role in third wave regimes as well: in dominant party systems the economic performance – incumbent turnover link was significantly weakened. However, the effect of institutional variables enhancing or obfuscating clarity of responsibility appeared to be of little significance in our sample. Finally, an interesting finding was that civil war strengthens the economic performance –turnover link, presumably because the detrimental economic consequences of civil war render economic performance more important for citizens’ evaluations of incumbents. Tests of electoral accountability for an alternative government performance indicator, i.e. the governments’ record for respecting civil liberties in the years prior to elections, demonstrated that here also the conditional effect of election quality was important. In clean elections, a worse civil liberties record in the years before elections is associated with higher incumbent turnover in elections, while in flawed elections, there is no association between the two variables.

Again civil war appears to strengthen the accountability link, however this effect is less robust as in the economic performance models. Concluding, indeed it seems that election quality has important consequences for electoral accountability, with electoral fraud undermining electoral accountability and thereby distorting one of the core functions of elections in democracies: giving citizens the power to “throw the rascals out” if they are dissatisfied with their performance. However, the effects found were stronger for economic performance, and concerns of endogeneity limit the robustness of the effects found for alternative performance domains. Future research should hence focus on investigating the consequences of electoral fraud for accountability in non-economic policy domains such as good governance and corruption, quality of public services, etc.

Turning to responsiveness, the consequences of electoral fraud were analyzed as the effect of election quality on government performance in the years after the elections (see chapter 6). Government performance was measured using two indicators: economic growth in the years after elections and respect for civil liberties in the years after elections. With respect to economic growth, election quality appeared to have no effect at all on post-election growth. Clearly, even if election quality increases the possibilities for citizens to hold incumbents accountable for economic decline prior to the elections, clean elections do not lead to higher economic growth after elections. Future research should investigate whether election quality does affect other governance indicators, such as public service provision or public service functioning. Even if election quality did not affect post-election economic growth however, respect for civil liberties and human rights in the years after elections appeared to be strongly and positively influenced by election quality. Clean elections are associated with higher post-election respect for civil liberties and human rights as compared to fraudulent elections, generating further evidence for the spill-over effects of election quality on other partial regimes of democracy.

These findings on spill-over effects of election quality on other partial regimes of democracy, leads to the question of the consequences of election quality for broader democratization processes. Apart from accountability and responsiveness, high quality elections are often assumed to have a wide variety of wonderful consequences, ranging from effective and stable governance, to regime legitimacy, to spill-over effects on other partial regimes of democracy such as the rule of law and respect for civil liberties, etc. However, most of these claims have not (yet) been subject to empirical testing. The scope of this thesis did not allow thorough empirical analyses of the consequences of electoral fraud for broader processes of democratization either, however, based on the findings presented in earlier chapter some directions for future research can be proposed.

As outlined in chapter 1, assessing the consequences of elections for democratization is challenging due to the fact that elections are partial regimes that -together with other partial regimes- form composite regimes that, depending on the quality of all partial regimes together, may or may not be considered democracies. Hence, when democratization is studied as consolidation of democracy or quality of democracy, the quality of the elections forms one of the elements that shape the quality of the regime as a whole. In these cases, differentiating cause from consequence becomes difficult if not impossible. In this thesis it was therefore proposed that studying the consequences of election quality for broader democratization processes can be done in two ways: first, by considering democratization-as-regime-stability and second, by considering democratization-as-spill-over. Considering democratization-as-regime-stability, the data appear to indicate that coup d'états and civil war are less common after clean elections, suggesting that the absence of fraud is indeed associated with more stable regimes. While these are bi-variate findings that would require more in-depth analyses, they are in line with findings of other authors (Collier 2009), and hence further investigation of these questions could be a fruitful direction for future research. Considering democratization-as-spill-over, this implies that the quality of elections could affect the quality of other partial regimes that are considered

important for democracy. Here, indeed our findings in chapter 6 seem to suggest that higher quality elections generate increased respect for civil liberties and human rights in the years after elections. This effect was robust to the inclusion of a variety of control variables and even when correcting for levels of civil liberties in the years before the elections, the quality of the elections continued to exert an independent and significant effect on respect for civil liberties. Moreover, chapter 5 demonstrated that in clean elections citizens hold governments accountable for their past record on civil liberties as well. Hence, if clean elections help citizens to hold their governments accountable for unsatisfactory levels of civil liberties in the years prior to elections, and if clean elections have spill-over effects on respect for civil liberties in the years after elections, this suggests that the quality of elections may have important consequences for democratization, improving the quality of other partial regimes as successive elections are held.

However, since election quality is itself, to a certain extent, driven by aspects of civil liberties such as freedom of speech, and rule of law as expressed in independence of the judiciary, the causal connection seems to run both ways. Indeed, the empirical analyses found that independence of the media and judiciary prior to elections are associated with higher quality elections, while the analyses of consequences of election quality demonstrated that higher quality elections are associated with subsequent improvements in civil liberties and respect for human rights. A question to address in future research is hence how to further disentangle these causal dynamics. Is it possible that there are regimes that find themselves in a “virtuous” circle of democratization, with high quality elections generating increased respect for civil liberties that in turn further enhance the quality of the next elections, etc.? And vice versa, could there be regimes that find themselves in a “vicious” circle of flawed elections leading to further declining respect for civil liberties which in turn undermines the quality of the next elections, etc.? Could such dynamics set some countries on a path of democratization while keeping others trapped? And if so, what explains these different trajectories and

changes in these trajectories? Analyzing partial regimes, and the interplay between partial regimes, may generate better understanding of processes of democratization.

Concluding, this thesis has provided some important insights into the causal dynamics underlying electoral fraud and has shed light on (some of) the consequences of electoral fraud. However, several important questions remain to be addressed in future research. First of all, more specific data is needed on types of fraud, the actors involved and within-country variation. Fraud can take many different forms and have many different perpetrators, hence analyzing its empirical variety with more precise and disaggregate indicators is very important. Secondly, better theorizing and subsequent empirical testing is needed on the conditional nature of fraud: under what conditions can different actors be expected to engage in fraud, and what types and combinations of fraud are more likely under what conditions. As the Economist noted in its recent article on “how to rig an election”: “the sophisticated vote-rigger sets about rigging the election well in advance” (Economist March 3^d 2012). However, not all actors engaging in fraud might want fraud to remain covert, not all actors might have access to possibilities to rig the elections well in advance, and no matter how carefully planned, even sophisticated vote-riggers might realize after election day that their efforts were not sufficiently successful, requiring more risky methods of manipulation such as ‘creative’ vote tabulation. Such disaggregate and conditional analyses would not only advance academic understanding of this topic, but would also generate more useful advice for practitioners, both in terms of mapping the ‘menu of manipulation’ as well as gathering more concrete insights in how to counteract fraud. Finally, future research should consider how electoral fraud affects broader democratization processes, not only by studying its effects on other partial regimes, but also by seeking to disentangle causal links between different partial regimes and better understand over-time changes in the quality of elections and eventually, the quality of democracy.

Bibliography

- Acemoglu, D., Johnson, S., and Robinson, J.A. (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*, 91(5), 1369-1401.
- Acemoglu, D., and Robinson, J.A. (2006). *Economic origins of dictatorship and democracy*. Cambridge, New York: Cambridge University Press.
- Adcock, R. and Collier, D. (2001). Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *American Political Science Review*, 95(3), 529-546.
- Ades, A. and Di Tella, R. (1997). The New Economics of Corruption: A Survey and Some New Results. *Political Studies*, 45, 496-515.
- Akkerboom, H. and Schmeets, H. (1998). Survey data collection methodology for election missions: How professional form design helps to control a survey. *International Journal of Public Opinion Research*, 10(3), 257-274.
- Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S. and Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, 8 (June), 155-194.
- Altman, D. and Perez-Linan, A. (2002). Assessing the Quality of Democracy: Freedom, Competitiveness and Participation in Eighteen Latin American Countries. *Democratization*, 9(2), 85-100.
- Alvarez, R.M., Hall, T.E. and Hyde, S.D. (eds.). (2008). *Election Fraud: Detecting and Deterring Electoral Manipulation*. Washington, D.C.: Brookings Institution Press.
- Anderson, C. J. (2000). Economic voting and political context: a comparative perspective. *Electoral Studies*, 19, 151-170.
- Anderson, C. J. (2007). The end of economic voting? Contingency dilemmas and the limits of democratic accountability. *Annual Review of Political Science*, 10, 271-296.
- Anderson, L. and Dodd, D.C. (2005). *Learning democracy – Citizen engagement and Electoral choice in Nicaragua 1990-2001*. Chicago: University of Chicago Press.

- Anglin, D. G. (1998). International Election Monitoring: The African Experience. *African Affairs*, 97(389), 471-495.
- Arrow, K. J. (1963). *Social choice and individual values* (Vol. 2nd edition). New Haven: Yale University Press.
- Baland, J. M. and Robinson, J.A. (2007). How Does Vote Buying Shape the Economy? In F. C. Schaffer (Ed.), *Elections for Sale: The Causes and Consequences of Vote Buying*. Boulder: Lynne Rienner Publishers.
- Barkan, J. D. (2000). Protracted Transitions Among Africa's New Democracies. *Democratization*, 7(3), 227-243.
- Barreiro, B. (2008). Explaining the electoral performance of incumbents in democracies. In J. M. Maravall, Sanchez-Cuenca, I. (Ed.), *Controlling governments: voters, institutions, and accountability*. New York: Cambridge University Press.
- Beck, N. (2008). Time-series Cross-section Methods. In J. M. Box-Steffensmeier, Brady, H.E. and Collier, D. (Ed.), *The Oxford Handbook of Political Methodology*. New York: Oxford University Press.
- Beetham, D. (2004). Towards a Universal Framework for Democracy Assessment. *Democratization*, 11(2), 1-17.
- Berg-Schlosser, D. (Ed.). (2004). *Democratization: the State of the Art*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Birch, S. (1997). Nomenklatura democratization: Electoral clientelism in post-soviet Ukraine. *Democratization*, 4(4), 40-62.
- Birch, S. (2000). *Elections and democratization in Ukraine*. New York: St. Martin's Press.
- Birch, S. (2007). Electoral Systems and Electoral Misconduct. *Comparative Political Studies*, 40(12), 1533-1556.

- Birch, S. (2008). Electoral Malpractice and Electoral Manipulation in New and Semi-Democracies. from <http://www.essex.ac.uk/government/electoralmalpractice/data.htm>:
- Birch, S. (2011). Electoral Malpractice. Unpublished manuscript.
- Bjornlund, E. C. (2004). *Beyond Free and Fair: Monitoring Elections and Building Democracy*. Washington D.C.: Woodrow Wilson Center Press.
- Blais, A., Massicotte, L. and Yoshinaka, A. (2001). Deciding who has the right to vote: a comparative analysis of election laws. *Electoral Studies*, 20, 41–62.
- Boda, M.D. (2005). Judging elections by public international law: a tentative framework. *Representation*, 41(3), 208-229.
- Boda, M.D. (2005). Reconsidering the 'free and fair' question. *Representation*, 41(3), 155-160.
- Bratton, M. (1998). Second Elections in Africa. *Journal of Democracy*, 9(3), 51-66.
- Bratton, M. (2008). Vote buying and violence in Nigerian election campaigns. *Electoral Studies*, 27, 621-632.
- Brusco, V., Nazareno, M. and Stokes, S.C. (2004). Vote Buying in Argentina. *Latin American Research Review*, 39(2), 66-88.
- Bunce, V. J. (2000). Comparative Democratization: Big and Bounded Generalizations. *Comparative Political Studies*, 33(6/7), 703-734.
- Bunce, V.J., Wolchik, S.L. (2006). Favorable conditions and electoral revolutions. *Journal of Democracy*, 17(4), 5-18.
- Calingaert, D. (2006). Election Rigging and How to Fight It. *Journal of Democracy*, 17(3), 138-151.
- Calvo, E. and Murillo, M.V. (2007). *How Many Clients does it take to win an election? Estimating the Size and Structure of Political Networks in Argentina and Chile*. New Haven, United States: Yale University.
- Carothers, T. (1997). The Observers Observed. *Journal of Democracy*, 8(3), 17-31.
- Carothers, T. (2002). The End of the Transition Paradigm. *Journal of Democracy*, 13(1), 5-21.

- Carothers, T. (2006). *Confronting the Weakest Link: Aiding Political Parties in New Democracies*. Washington D.C.: Carnegie Endowment for International Peace.
- Carothers, T. (2007). The "Sequencing" Fallacy. *Journal of Democracy*, 18(1), 12-27.
- Casper, G. and Tufis, C. (2003). Correlation versus Interchangeability: The Limited Robustness of Empirical Findings on Democracy using Highly Correlated Datasets. *Political Analysis*, 11(2), 196-203.
- Chauvet, L. and Collier, P. (2009). Elections and economic policy in developing countries. *Economic Policy*, 24(59), 509-550.
- Cheibub, J.A. and Przeworski, A. (1999). Democracy, elections and accountability for economic outcomes. In: A. Przeworski, Stokes, S.C., Manin, B. (Ed.), *Democracy, Accountability, and Representation*. New York: Cambridge University Press.
- Cingranelli, D.L. and Richards, D.L. (2010). The Cingranelli-Richards (CIRI) Human Rights Dataset. Version 2010.05.17. <http://www.humanrightsdata.org>:
- Collier, D. and Levitsky, S. (1997). Democracy with Adjectives: Conceptual Innovation in Comparative Research. *World Politics*, 49(3), 430-451.
- Collier, D. and Mahon, J.E. (1993). Conceptual "Stretching" Revisited: Adapting Categories in Comparative Analysis. *American Political Science Review*, 87(4), 845-855.
- Collier, P. (2009). *Wars, Guns and Votes*. New York: Harper.
- Collier, P. and Vicente, P.C. (2011). *Votes and Violence: Evidence from a Field Experiment in Nigeria*. <http://www.pedrovicente.org/violence.pdf>,
- Coppedge, M. (2002). Democracy and Dimensions: Comments on Munck and Verkuilen. *Comparative Political Studies*, 35(1), 35-39.
- Cox, G.W. and Thies, M.F. (2000). How Much Does Money Matter? "Buying" Votes in Japan, 1967-1990. *Comparative Political Studies* 33(1), 37-57.
- Dahl, R. A. (1971). *Polyarchy: participation and opposition*. New Haven: Yale University Press.

- Dahl, R. A. (1989). *Democracy and its critics*. New Haven: Yale University Press.
- Dahlberg, S. (2009). Political parties and perceptual agreement: The influence of party related factors on voters' perceptions in proportional electoral systems. *Electoral Studies*, 28, 270-278.
- Dalton, R.J., Shin, D.C. and Jou, W. (2007). *Popular Conceptions of the Meaning of Democracy: Democratic Understanding in Unlikely Places*. UC Irvine: Center for the Study of Democracy (<http://escholarship.org/uc/item/2j74b860>).
- Darnolf, S. (2011). *Assessing Electoral Fraud in New Democracies: A New Strategic Approach*. Washington, D.C. International Foundation for Electoral Systems: White Paper Series Electoral Fraud.
- Davis-Roberts, A. and Carroll, D.J. (2010). Using international law to assess elections. *Democratization*, 17(3), 416-441.
- Desposato, S.W. (2007). How Does Vote Buying Shape the Legislative Arena? In F. C. Schaffer (Ed.), *Elections for Sale: The Causes and Consequences of Vote Buying*. Boulder: Lynne Rienner Publishers.
- Diamond, L. (1999). *Developing democracy: toward consolidation*. Baltimore: Johns Hopkins University Press.
- Diamond, L. (2002). Elections without democracy: Thinking about hybrid regimes. *Journal of Democracy*, 13(2), 21-35.
- Diamond, L. and Morlino, L. (2004). The Quality of Democracy: An Overview. *Journal of Democracy*, 15(4), 20-31.
- Diamond, L. and Morlino, L. (Ed.). (2005). *Assessing the quality of democracy*. Baltimore: Johns Hopkins University Press.
- Donno, D. (2010). Who Is Punished? Regional Intergovernmental Organizations and the Enforcement of Democratic Norms. *International Organization*, 64(593-625).

- Doorenspleet, R. (2000). Reassessing the Three Waves of Democratization. *World Politics*, 52(3), 384-406.
- Doorenspleet, R. (2001). The fourth wave of democratization: identification and explanation (PhD dissertation). University of Leiden, Leiden.
- Doorenspleet, R. and Kopecky, P. (2008). Against the Odds: Deviant Cases of Democratization. *Democratization*, 15(4), 697-713.
- Dreher, A., Gaston, N. and Martens, P. (2008). *Measuring Globalization – Gauging its Consequences*. New York: Springer.
- Duch, R.M. (2001). A Developmental Model of Heterogeneous Economic Voting in New Democracies. *The American Political Science Review*, 95(4), 895-910.
- Duch, R.M. and Stevenson, R. (2005). Context and the Economic Vote: A Multilevel Analysis. *Political Analysis*, 13, 387-409.
- Eisenstadt, T.A. (2002). Measuring Electoral Court Failure in Democratizing Mexico. *International Political Science Review / Revue Internationale de Science Politique*, 23(1), 47-68.
- Eisenstadt, T.A. (2004). *Courting democracy in Mexico : party strategies and electoral institutions*. Cambridge; New York: Cambridge University Press.
- Elklit, J. (1999). Electoral institutional change and democratization: You can lead a horse to water, but you can't make it drink. *Democratization*, 6(4), 28-51.
- Elklit, J. and Reynolds, A. (2005a). A framework for the systematic study of election quality. *Democratization*, 12(2), 147-162.
- Elklit, J. and Reynolds, A. (2005b). Judging elections and election management quality by process. *Representation*, 41(3), 189-207.
- Elklit, J. and Svensson, P. (1997). What Makes Elections Free and Fair? *Journal of Democracy*, 8(3), 32-46.

- European Commission, NEEDS. (2007). *Compendium of International Standards for Elections* (2nd edition). Brussels: European Commission.
- Fritz, V. (2008). Mongolia: The Rise and Travails of a Deviant Democracy. *Democratization*, 15(4), 766-788.
- Gandhi, J. (2008). What Are We Missing? Electoral Data in Democracies and Non-democracies. In: *Committee on Concepts and Methods - Working Paper Series*.
- Gandhi, J. and Lust-Okar, E. (2009). Elections Under Authoritarianism. *Annual Review of Political Science*, 12, 403-422.
- Garber, L. (1993). The OAU and Elections. *Journal of Democracy*, 4(3), 55-59.
- Geddes, B. (1999). What do we know about democratization after twenty years? *Annual Review of Political Science*, 2, 115-144.
- Geisler, G. (1993). Fair? What Has Fairness Got to Do with It? Vagaries of Election Observations and Democratic Standards. *The Journal of Modern African Studies*, 31(4), 613-637.
- Gerring, J., Thacker, S.C., Moreno, C. (2005). Centripetal democratic governance: a theory of global inquiry. *American Political Science Review*, 99, 567-581.
- Gibney, M., Cornett, L., Wood, R. (2010). Political Terror Scale 1976-2008. Retrieved April 21, 2010 from the Political Terror Scale web site: <http://www.politicalterroryscale.org>.
- Golder, M. and Stramski, J. (2010). Ideological Congruence and Electoral Institutions. *American Journal of Political Science*, 54, 90-106.
- Golder, M. and Wantchekon, L. (2004). Africa: Dictatorial and Democratic Electoral Systems since 1946. In: J. Colomer (Ed.), *Handbook of Electoral System Design*. London: Palgrave.
- Goodwin-Gill, G.S. (2006). *Free and Fair Elections*. New Expanded Edition. Geneva: Inter-Parliamentary Union.
- Gueniffey, P. (1993). *Le nombre et la raison: la Révolution française et les élections*. Paris: École des Hautes Études en Sciences Sociales.

- Hadenius, A. (1992). *Democracy and development*. Cambridge: Cambridge University Press.
- Hadenius, A. and Teorell, J. (2005). *Assessing Alternative Indices of Democracy*. In: Committee on Concepts and Methods - Working Paper Series.
- Haerpfer, C. W., Bernhagen, P., Inglehart, R.F., and Welzel, C. (Ed.). (2009). *Democratization*. Oxford: Oxford University Press.
- Hartlyn, J., McCoy, J. and Mustillo, T.J. (2003). The 'Quality of Elections' in Contemporary Latin America: Issues in Measurement and Explanation. Paper presented at the Paper prepared for the XXIV International Congress of the Latin American Studies Association.
- Hartlyn, J., McCoy, J. and Mustillo, T.M. (2008). Electoral Governance Matters: Explaining the Quality of Elections in Contemporary Latin America. *Comparative Political Studies*, 41(1), 73-98.
- Hartlyn, J. and McCoy, J. (2006). Observer paradoxes: how to assess electoral manipulation. In A. Schedler (Ed.), *Electoral authoritarianism: the dynamics of unfree competition*. Boulder: Lynne Rienner Publishers.
- Hellwig, T. and Samuels, D. (2007a). Voting in Open Economies: The Electoral Consequences of Globalization. *Comparative Political Studies*, 40(3), 283-306.
- Hellwig, T. and Samuels, D. (2007b). Electoral Accountability and the Variety of Democratic Regimes. *British Journal of Political Science*, 38, 65-90.
- Hermet, G., Rose, R., Rouquie, A. (Ed.). (1978). *Elections without choice*. New York: Macmillan.
- Hicken, A. D. (2007). How Do Rules and Institutions Encourage Vote Buying? . In F. C. Schaffer (Ed.), *Elections for Sale: The Causes and Consequences of Vote Buying*. Boulder: Lynne Rienner Publishers.
- Hicken, A. D. (2007). How Effective Are Institutional Reforms? In F. C. Schaffer (Ed.), *Elections for Sale: The Causes and Consequences of Vote Buying*. Boulder: Lynne Rienner Publishers.

- Honaker, J. and King, G. (2010). What to do About Missing Values in Time Series Cross-Section Data. *American Journal of Political Science*, 54(2), 561-581.
- House, F. (2009). Freedom in the World: <http://www.freedomhouse.org>.
- Huntington, S.P. (1991). *The third wave: democratization in the late twentieth century*. Norman: University of Oklahoma Press.
- Hyde, S. and Marinov, N. (2010). *Which Elections Can Be Lost?* Yale University.
- Hyde, S. D. (2007). The Observer Effect in International Politics. Evidence from a Natural Experiment. *World Politics*, 60(1), 37-63.
- Hyde, S. D. (2011a). International Dimensions of Elections. In N. Brown (Ed.), *The Dynamics of Democratization: Dictatorship, Development, and Diffusion*. Baltimore: Johns Hopkins University Press.
- Hyde, S. D. (2011b). Catch Us If You Can: Election Monitoring and International Norm Diffusion. *American Journal of Political Science*, 55(2), 356-369.
- Jhee, B. K. (2006). *Economic voting in new democracies*. PhD Dissertation, University of Missouri-Columbia.
- Kalandadze, K. and Orenstein, M.A. (2009). Electoral Protests and Democratization Beyond the Color Revolutions. *Comparative Political Studies*, 42(11), 1403-1425.
- Karl, T.L. (1990). Dilemmas of Democratization in Latin America. *Comparative Politics*, 23(1), 1-21.
- Karl, T.L. (1995). The Hybrid Regimes of Central America. *Journal of Democracy*, 6(3), 72-86.
- Karl, T.L. (2000). Electoralism. In R. Rose (Ed.), *International encyclopedia of elections*. London: Macmillan.
- Katz, R.S. (2005). Democratic principles and judging tree and fair. *Representation*, 41(3), 161-179.
- Keefer, P. (2009). *Database of Political Institutions: Changes and Variable Definitions: Development Research Group: World Bank*.

- Kelley, J. (2008). Assessing the Complex Evolution of Norms: The Rise of International Election Monitoring. *International Organization*, 62, 221–255.
- Kelley, J. (2009a). D-Minus Elections: The Politics and Norms of International Election Observation. *International Organization*, 63(Fall 2009), 765–787.
- Kelley, J. (2009b). The More the Merrier? The Effects of Having Multiple International Election Monitoring Organizations. *Perspectives on Politics*, 7, 59–64.
- Kelley, J. (2010). Election observers and their biases. *Journal of democracy*, 21(3), 158–172.
- Kelley, J. (2011). Do International Election Monitors Increase or Decrease Opposition Boycotts? *Comparative Political Studies*, 44.
- Kelley, J. and Kiril, K. (2010). Election Quality and International Observation 1975–2004: Two New Datasets. Duke University (available at SSRN: <http://ssrn.com/abstract=1694654>).
- Kohnert, D. (2004). Election Observation in Nigeria & Madagascar: Diplomatic vs. Technocratic Bias. *Review of African Political Economy*, 31(99), 83–101.
- Kopecky, P. and Mudde, C. (2000). What has Eastern Europe taught us about the democratisation literature (and vice versa)? *European Journal of Political Research*, 37(4), 517–539.
- Kumar, K. (Ed.). (1998). *Postconflict Elections, Democratization, and International Assistance*. London.
- LeDuc, L., Niemi, R.G., and Norris, P. (Ed.). (2010). *Comparing Democracies 3: Elections and Voting in the 21st Century*. Los Angeles: SAGE.
- Lehoucq, F. (2003). Electoral Fraud: Causes, Types, and Consequences. *Annual Review of Political Science*, 6, 233–256.
- Lehoucq, F. (2007). When does a Market for Votes Emerge? In F. C. Schaffer (Ed.), *Elections for Sale: The Causes and Consequences of Vote Buying*. Boulder: Lynne Rienner Publishers.
- Lehoucq, F. and Molina, I. (2002). *Stuffing the ballot box: fraud, electoral reform, and democratization in Costa Rica*. Cambridge; New York: Cambridge University Press.

- Levitsky, S. and Way, L. (2002). Elections Without Democracy: The Rise of Competitive Authoritarianism. *Journal of Democracy*, 13(2), 51-65.
- Levitsky, S. and Way, L. (2010). *Competitive Authoritarianism. Hybrid Regimes after the Cold War*. New York: Cambridge University Press.
- Lewis-Beck, M.S. and Paldam, M. (2000). Economic voting: an introduction. *Electoral Studies*, 19, 113–121.
- Lewis-Beck, M.S. and Stegmaier, M. (2008). The economic vote in transitional democracies. *Journal of Elections, Public Opinion and Parties*, 18(3), 303-323.
- Lijphart, A. (1999). *Patterns of democracy: government forms and performance in thirty-six countries*. New Haven; London: Yale University Press.
- Lindberg, S. (2006a). *Democracy and elections in Africa*. Baltimore: Johns Hopkins University Press.
- Lindberg, S. (2006b). The Surprising Significance Of African Elections. *Journal of Democracy*, 17(1), 139-151.
- Lindberg, S. (Ed.). (2009). *Democratization by Elections. A New Mode of Transition*. Baltimore: John Hopkins University Press.
- Linz, J.J. and Stepan, A. (1996). *Problems of democratic transition and consolidation: Southern Europe, South America, and post-communist Europe*. Baltimore: Johns Hopkins University Press.
- López-Pintor, R. (2010). *Assessing Electoral Fraud in New Democracies: A Basic Conceptual Framework*. Washington D.C. International Foundation for Electoral Systems: White Paper Series Electoral Fraud.
- Lust-Okar, E. (2009). Opposition versus dictators: explaining divergent electoral outcomes in post-Communist Europe and Eurasia. In S. Lindberg (Ed.), *Democratization by Elections. A New Mode of Transition*. Baltimore: John Hopkins University Press.

- Lyons, T. (2004). Post-conflict Elections and the Process of Demilitarizing Politics: The Role of Electoral Administration. *Democratization*, 11(3), 36–62.
- Maeda, K. (2009). Divided We Fall: Opposition Fragmentation and the Electoral Fortunes of Governing Parties. *British Journal of Political Science*, 40, 419–434.
- Magaloni, B. (2006). *Voting for autocracy: hegemonic party survival and its demise in Mexico*. Cambridge; New York: Cambridge University Press.
- Mainwaring, S., O'Donnell, G.A. and Valenzuela, J.S. (Ed.). (1992). *Issues in democratic consolidation: the new South American democracies in comparative perspective*. Notre Dame: Published for the Helen Kellogg Institute for International Studies by University of Notre Dame Press.
- Mair, S. (1994). *International Election Observation: One Form of Democratization Assistance*: Ebenhausen: Research Institute for International Politics and Security.
- Maravall, J. M., Sanchez-Cuenca, I. (2008). *Controlling governments: voters, institutions, and accountability*. New York: Cambridge University Press.
- Marshall, M.G. and Jaggers, K. (2002). *Polity IV Project: Political Regime Characteristics and Transitions, 1800–2002: Dataset Users' Manual*. Maryland: University of Maryland.
- Massicotte, L. (2005). Judging elections by their outcome? *Representation*, 41(3), 180–188.
- McCoy, J., Garbar, L. and Pastor, R.A. (1991). Pollwatching and Peacemaking. *Journal of Democracy*, 2(4), 102–114.
- McDonald, M.D. and Budge, I. (2005). *Elections, parties, democracy: conferring the median mandate*. Oxford: Oxford University Press.
- Mebane, W.R. (2007). *Election Forensics: Statistics, Recounts and Fraud*. Paper presented at the Annual Meeting of the Midwest Political Science Association.
- Merkel, W. (2004). Embedded and defective democracies. *Democratization*, 11(5), 33–58

- Miller, W. E. a. S., D.E. (1963). Constituency Influence in Congress. *American Political Science Review*, 57(1), 45-56.
- Møller, J. (2007). The Gap between Electoral and Liberal Democracy Revisited. Some Conceptual and Empirical Clarifications. *Acta Politica*, 42, 380-400.
- Montinola, G.R. and Jackman, R.W. (2002). Sources of Corruption: A Cross-Country Study. *British Journal of Political Science*, 32, 147-170.
- Mori McElwain, K. (2008). Manipulating Electoral Rules to Manufacture Single-Party Dominance. *American Journal of Political Science*, 52(1), 32-47.
- Mozaffar, S. and Schedler, A. (2002). The Comparative Study of Electoral Governance—Introduction. *International Political Science Review / Revue Internationale de Science Politique*, 23(1), 5-27.
- Munck, G.L. (2004). Successes and failures of the new democracies. In D. Berg-Schlosser (Ed.), *Democratization: the state of the art*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Munck, G.L. (2006). Drawing boundaries: how to craft intermediate regime categories. In A. Schedler (Ed.), *Electoral authoritarianism: the dynamics of unfree competition*. Boulder: Lynne Rienner Publishers.
- Munck, G.L. (2009). *Measuring democracy: a bridge between scholarship and politics*. Baltimore: Johns Hopkins University Press.
- Munck, G.L. and Verkuilen, J. (2002). Conceptualizing and Measuring Democracy: Evaluating Alternative Indices. *Comparative Political Studies*, 35(1), 5-34.
- Myagkov, M. G., Ordeshook, P.C. and Shakin, D. (2009). *The forensics of election fraud: Russia and Ukraine*. Cambridge; New York: Cambridge University Press.
- Mylonas, H. and Roussias, N. (2008). When Do Votes Count? Regime Type, Electoral Conduct, and Political Competition in Africa. *Comparative Political Studies*, 41(11), 1466-1491.

- Nevitte, N. and Canton, S.A. (1997). The Role of Domestic Observers. *Journal of Democracy*, 8(3), 47-61.
- Nohlen, D., Krennerich, M. and Thibaut, B. (Ed.). (1999). *Elections in Africa: a data handbook*. Oxford: Oxford University Press.
- Nohlen, D., Grotz, F. and Hartmann, C. (Ed.). (2001). *Elections in Asia and the Pacific: a data handbook*. Oxford: Oxford University Press.
- Nohlen, D. (Ed.). (2005). *Elections in the Americas: a data handbook (part 1 & 2)*. Oxford: Oxford University Press.
- Nohlen, D. and Stöver, P. (Ed.). (2010). *Elections in Europe: a data handbook*. Baden-Baden, Germany: Nomos.
- Norris, P. (2008). *Driving democracy: do power-sharing institutions work?* Cambridge, New York: Cambridge University Press.
- Norris, P. (2009). All Elections Are Not the Same: Why Power-Sharing Elections Strengthen Democratization. In S. Lindberg (Ed.), *Democratization by Elections. A New Mode of Transition*. Baltimore: John Hopkins University Press.
- Nyblade, B. and Reed Chuo, S.R. (2008). Who Cheats? Who Loots? Political Competition and Corruption in Japan, 1947–1993. *American Journal of Political Science*, 52(4), 926-941.
- O'Donnell, G.A., Schmitter, P.C. and Whitehead, L. (Ed.). (1986). *Transitions from authoritarian rule: prospects for democracy*. Baltimore: Johns Hopkins University Press.
- O'Donnell, G.A. (1996). Illusions About Consolidation. *Journal of Democracy*, 7(2), 34-51.
- O'Donnell, G.A. (1998). Horizontal Accountability in New Democracies. *Journal of Democracy*, 9(3), 112-126.
- O'Donnell, G.A. (2001). Democracy, Law, and Comparative Politics. *Studies in Comparative International Development*, 36(1), 7-36.

- O'Donnell, G.A. (2002). In Partial Defense of an Evanescent "Paradigm". *Journal of Democracy*, 13(3), 6-12.
- O'Donnell, G.A. (2004). Why the Rule of Law Matters. *Journal of Democracy*, 15(4), 32-46.
- O'Donnell, G.A., Vargas Cullell, J. and Iazzetta, O.M. (Ed.). (2004). *The quality of democracy: theory and applications*. Notre Dame: University of Notre Dame Press.
- O'Donnell, G. A. (2010). Schmitter's Retrospective: A Few Dissenting Notes. *Journal of Democracy*, 21(1), 29-32.
- Pacek, A. and Radcliff, B. (1995). The political economy of competitive elections in the developing world. *American Journal of Political Science*, 39(3), 745-759.
- Pastor, R.A. (1990). The Making of a Free Election. *Journal of Democracy*, 1(3), 13-25.
- Pastor, R.A. (1998). Mediating elections. *Journal of Democracy*, 9(1), 154-163.
- Pastor, R.A. (1999). The role of electoral administration in democratic transitions: Implications for policy and research. *Democratization*, 6(4), 1-27.
- Pemstein, D., Meserve, S.A., and Melton, J. (2010). Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type. *Political Analysis*, 18, 426-449.
- Posner, D.N. and Simon, D.J. (2002). Economic Conditions and Incumbent Support in Africa's New Democracies: Evidence from Zambia. *Comparative Political Studies*, 35(3), 313-336.
- Powell, G.B. (2000). *Elections as Instruments of Democracy: Majoritarian and Proportional Visions*. New Haven: Yale University Press.
- Powell, G.B. (2004). The Chain of Responsiveness. *Journal of Democracy*, 15(4), 91-105.
- Powell, G. B. Jr. and Whitten, G.D. (1993). A Cross-National Analysis of Economic Voting: Taking Account of the Political Context. *American Journal of Political Science*, 37(2), 391-414.
- Przeworski, A., Alvarez, M., Cheibub, J.A. and Limongi, F. (1996). What Makes Democracies Endure? *Journal of Democracy*, 7(1), 39-55.

- Przeworski, A., Stokes, S.C., Manin, B (Ed.). (1999). *Democracy, accountability, and representation*. New York: Cambridge University Press.
- Przeworski, A., Alvarez, M., Cheibub, J.A. and Limongi, F. (2000). *Democracy and Development: Political Institutions and Well-Being in the World, 1950-1990*. Cambridge: Cambridge University Press.
- Rakner, L. and Van de Walle, N. (2009). Opposition Weakness in Africa. *Journal of Democracy*, 20(3), 108-121.
- Regalia, M. (2009). Conceptualizing and Measuring Elections. Paper presented at the Paper presented at Second Italian Graduate Conference.
- Reynolds, A. (2011). *Designing Democracy in a Dangerous World*. Oxford: Oxford University Press.
- Rose, R. (2008). Evaluating Democratic Governance: A Bottom-up Approach to European Union Enlargement. *Democratization*, 15(2), 251-271.
- Rose, R. and Munro, N. (2003). *Elections and parties in new European democracies*. Washington, D.C.: CQ Press, A Division of Congressional Quarterly.
- Rose, R. and Shin, C.D. (2001). Democratization Backwards: The Problem of Third-Wave Democracies. *British Journal of Political Science*, 31(2), 331-354.
- Roussias, N. (2011 (June 29-July 1)). How Does Electoral Misconduct Affect Party Systems? An Examination of the Mechanisms of Electoral Fraud. Paper presented at the conference: Electoral fraud, vote buying and clientelism. Juan March Institute, Madrid, Spain.
- Samuels, D. (2004). Presidentialism and Accountability for the Economy in Comparative Perspective. *The American Political Science Review*, 98(3), 425-436.
- Samuels, D. and Hellwig, T. (2010). Elections and Accountability for the Economy: A Conceptual and Empirical Reassessment. *Journal of Elections, Public Opinion and Parties*, 20(4), 393-419.
- Samuels, D. and Snyder, R. (2001). The Value of a Vote: Malapportionment in Comparative Perspective. *British Journal of Political Science*, 31, 651-671.

- Sandholtz, W. and Taagepera, R. (2005). Corruption, Culture and Communism. *International Review of Sociology/Revue Internationale de Sociologie*, 15(1), 109-113.
- Sartori, G. (1970). Concept Misformation in Comparative Politics. *The American Political Science Review*, 64(4), 1033-1053.
- Schaffer, F.C. (Ed.). (2007). *Elections For Sale. The Causes and Consequences of Vote Buying*. London: Lynne Rienner Publishers.
- Schedler, A. (1998). What is Democratic Consolidation? *Journal of Democracy*, 9(2), 91-107.
- Schedler, A. (2002a). The Menu of Manipulation. *Journal of Democracy*, 13(2), 36-50.
- Schedler, A. (2002b). The Nested Game of Democratization by Elections. *International Political Science Review / Revue Internationale de Science Politique*, 23(1), 103-122.
- Schedler, A. (Ed.). (2006). *Electoral authoritarianism. The dynamics of unfree competition*. Boulder Lynne Rienner Publishers.
- Schedler, A. (2009). A theory of elections as a mode of transition. In S. Lindberg (Ed.), *Democratization by Elections. A New Mode of Transition*. Baltimore: John Hopkins University Press.
- Schmeets, J.J.G. (2010). *The Assessment of Elections in the OSCE region, 1996-2010*. Paper presented at the ECPR-Conference.
- Schmitter, P.C. (1992). The Consolidation of Democracy and Representation of Social Groups. *American Behavioral Scientist*, 35(4/5), 422-449.
- Schmitter, P.C. (2004). The ambiguous virtues of accountability. *Journal of Democracy*, 15(4), 47-60.
- Schmitter, P.C. (2010). Twenty-Five Years, Fifteen Findings. *Journal of Democracy*, 21(1), 17-28.
- Schmitter, P.C., & Brouwer, I. (1999). *Conceptualizing, Researching and Evaluating Democracy Promotion and Protection (No. 9)*. Florence: European University Institute.
- Schmitter, P.C. and Karl, T.L. (1991). What Democracy is ... and is not. *Journal of Democracy*, 2(3), 75-88.

- Schneider, C.Q. (2009). *The consolidation of democracy: comparing Europe and Latin America*. New York: Routledge.
- Schneider, C.Q. and Schmitter, P.C. (2004). Liberalization, transition and consolidation: measuring the components of democratization. *Democratization*, 11(5), 59-90.
- Schumpeter, J. A. (1975). *Capitalism, socialism, and democracy*. New York: Harper.
- Seligson, M.A. and Booth, J.A. (Ed.). (1995). *Elections and Democracy in Central America - Revisited*. Chapel Hill: The University of North Carolina Press.
- Simpser, A. (2005). *Making votes not count: Strategic incentives for electoral corruption*. Stanford University.
- Simpser, A. (2008). Unintended Consequences of Election Monitoring. In R. M. Alvarez, Hall, T.E. and Hyde, S.D. (Ed.), *Election Fraud. Detecting and Deterring Electoral Manipulation*. Washington, D.C.: Brookings Institution Press.
- Simpser, A. and Donno, D. (2009). *Can International Election Monitoring Harm Governance?* University of Chicago and University of Pittsburgh.
- Smulovitz, C. and Peruzzotti, E. (2000). Societal Accountability in Latin America. *Journal of democracy*, 11(4), 147-158.
- Snyder, R. (2006). Beyond electoral authoritarianism: the spectrum of nondemocratic regimes. In A. Schedler (Ed.), *Electoral authoritarianism: the dynamics of unfree competition*. Boulder: Lynne Rienner Publishers.
- Snyder, R. and Samuels, D. (2001). Devaluing the Vote in Latin America. *Journal of Democracy*, 12(1), 146-159.
- Soroka, S.N. and Wlezien, C. (2010). *Degrees of democracy : politics, public opinion, and policy*. New York: Cambridge University Press.
- Stimson, J.A., Mackuen, M.B., Erikson, R.S. (1995). Dynamic Representation. *American Political Science Review*, 89(3), 543-565

- Stokes, S., Dunning, S., Nazareno, M. and Brusco, V. (2012 –expected–). *Buying Votes: Distributive Politics in Democracies*. Unpublished manuscript.
- Stokes, S.C. (2007). *Is Vote Buying Undemocratic?* In F. C. Schaffer (Ed.), *Elections for Sale: The Causes and Consequences of Vote Buying*. Boulder: Lynne Rienner Publishers.
- Strøm, K. (2000). Delegation and accountability in parliamentary democracies. *European Journal of Political Research*, 37, 261-289.
- Strøm, K., Müller, W.C., Bergman, T. (2003). *Delegation and accountability in parliamentary democracies*. Oxford: Oxford University Press.
- Svensson, J. (2005). Eight Questions About Corruption. *The Journal of Economic Perspectives*, 19(3), 19-42.
- Teorell, J. and Hadenius, A. (2009). Post-Cold War political regimes: when do elections matter? In S. Lindberg (Ed.), *Democratization by elections : a new mode of transition*. Baltimore: John Hopkins University Press.
- Teorell, J., Charron, N., Samanni, M., Holmberg, S.; Rothstein, B. (2011). *The Quality of Government Dataset, version 6Apr11*. University of Gothenburg: The Quality of Government Institute, <http://www.qog.pol.gu.se>.
- Thomassen, J. (1994). Empirical research into political representation: Failing democracy or failing models? In M. K. a. M. Jennings, T.E. (Ed.), *Elections at home and abroad: Essays in honour of Warren E. Miller*. Ann Arbor: University of Michigan Press.
- Treisman, D. (2007). What have we learned about the causes of corruption from ten years of cross-national empirical research? . *Annual Review of Political Science*, 10, 211-244.
- Tucker, J. A. (2006). *Regional economic voting: Russia, Poland, Hungary, Slovakia and the Czech Republic, 1990--1999*. New York: Cambridge University Press.
- Turner, A. W. (1993). Postauthoritarian Elections: "Testing Expectations About "First" Elections". *Comparative Political Studies*, 26(3), 330-349.

- UNDP. (2004). Human Development Report 2004: Cultural Liberty in Today's Diverse World.
- Van de Walle, N. (2002). Africa's Range of Regimes. *Journal of Democracy*, 13(2), 66-80.
- Van de Walle, N. (2003). Presidentialism and Clientelism in Africa's Emerging Party Systems. *The Journal of Modern African Studies*, 41(2), 297-321.
- Van der Brug, W., Van der Eijk, C., Franklin, M.N. (2007). *The Economy and the Vote: Electoral Responses to Economic Conditions in 15 Countries*. New York: Cambridge University Press.
- Van der Eijk, C. and Franklin, M. (2009). *Elections and Voters*. New York: Palgrave Macmillan.
- Van Ham, C. (2009). Choice or complexity? The impact of party system characteristics on voter uncertainty in European democracies. Paper presented at the European Consortium of Political Research (ECPR) 2009 Joint Sessions.
- Van Schuur, W. H. (2003). Mokken Scale Analysis: Between the Guttman Scale and Parametric Item Response Theory. *Political Analysis*, 11, 139-163.
- Vicente, P. C. (2010). Is Vote-buying Effective? Evidence from a Field Experiment in West Africa. <http://www.pedrovicente.org/banho.pdf>,
- Vicente, P. C. and Wantchekon, L. (2009). Clientelism and vote buying: lessons from field experiments in African elections. *Oxford Review of Economic Policy*, 25(2), 292-305.
- Von Borzyskowski, I. (2011). Election Violence: International and Domestic Triggers. Paper presented at the Presented at the Colloquium on Political Behavior.
- Wall, A., Ellis, A., Ayoub, A., Dundas, C., Rukambe, J., Staino, S. (2006). *Electoral Management Design: The International IDEA Handbook*. Stockholm: Sweden: International Institute for Democracy and Electoral Assistance.
- Wang, C. S. and Kurzman, C. (2007). The Logistics: How to Buy Votes. In F. C. Schaffer (Ed.), *Elections For Sale. The Causes and Consequences of Vote Buying*. London: Lynne Rienner Publishers.

- Wessels, B. and Schmitt, H. (2008). Meaningful choices, political supply, and institutional effectiveness. *Electoral Studies*, 27(1), 19-30.
- Wrong, M. (2000). *In the Footsteps of Mr. Kurtz*. New York: Harper Collins.
- Wrong, M. (2009). *It's Our Turn to Eat: The Story of a Kenyan Whistle-Blower*. London: Harper.
- Zakaria, F. (1997). The rise of illiberal democracy. *Foreign Affairs*, 76, 22-41.
- Zielinski, J., Slomczynski, K., Shabad, G. (2005). Electoral control in new democracies: the perverse incentives of fluid party systems. *World Politics*, 57(April), 365-395.

Appendix A – Conceptualizing election quality^a

Authors	I. Background concepts	II. Systematized concept		III. Indicators
		Attributes	Components of attributes	
Schumpeter (1975)	<ul style="list-style-type: none"> Conceptualizations of democracy - Political leaders are elected by citizens - Competition between political leaders for citizens' votes 			
Dahl (1989)	<ul style="list-style-type: none"> - Elected officials - Free and fair elections - Inclusive suffrage - Right to run for office - Freedom of expression, - Alternative information - Associational autonomy 	1. Elected officials	<ul style="list-style-type: none"> 1.3. Object of choice: most powerful collective decision-makers selected through elections. 1.4. Consequences of choice: election results are irreversible & elected officials have control over policy-making. 	<ul style="list-style-type: none"> - Elected legislature - Elected executive (directly or indirectly) - Winners of elections assume office - Elected officials have discretion over policy-making, influence of "reserved power domains" is limited.
Huntington (1991)	<ul style="list-style-type: none"> - Elections for most powerful decision-makers - Fair and honest elections - Periodic elections - Universal suffrage - Free competition 	2. Periodic elections	2.1. Elections held at regular intervals.	- Elections held periodically, according to constitution/electoral law.
Przeworski et al. (1996)	<ul style="list-style-type: none"> - Ex ante uncertainty - Ex post irreversibility - Repeated elections 			
Schmitter and Karl (1991)	<ul style="list-style-type: none"> Dahl's 7 criteria plus 2: - Absence of reserved power domains - State sovereignty 	3. Participation	<ul style="list-style-type: none"> 3.1. De jure right to vote 3.2. De facto conditions for the use of the right to vote 	<ul style="list-style-type: none"> - De jure universal & equal suffrage - De facto possibility to participate: <ul style="list-style-type: none"> - Possibility to form preferences: <ul style="list-style-type: none"> - Freedom of expression - Alternative sources of information - Voter education - Possibility to cast a vote: <ul style="list-style-type: none"> - Freedom of movement - Freedom from intimidation/coercion - Voter registration (register, registration requirements) - Polling station access - Secret balloting - Votes cast are counted: <ul style="list-style-type: none"> - Accurate vote counting
Hermet et al. (1978)	<ul style="list-style-type: none"> Conceptualizations of election quality (concept-based) - Freedom voters - Competition candidates - Effects elections on policies 			
Elklit & Svensson (1997)	<ul style="list-style-type: none"> - Free elections (before, during & after polling day) - Fair elections (before, during & after polling day) 			
O'Donnell (2001)	<ul style="list-style-type: none"> - Competitive, Free, Egalitarian, Decisive, Inclusive 			
Lindberg (2006a)	<ul style="list-style-type: none"> - Participation - Competition - Legitimacy 			
Munck (2009)	<ul style="list-style-type: none"> - Inclusive elections - Clean elections - Competitive elections - Elective public offices 	4. Competition	Continued on next page.	
Lopez-Pintor (2010)	<ul style="list-style-type: none"> - Political framework: periodic elections, free expression of the will of the people - Free and fair elections: right of the people to vote, guarantees for secret ballot and genuine elections - Right to franchise: right of universal & equal suffrage - Right to participate: right to stand for election & assume office 			

Authors	I. Background concepts	II. Systematized concepts		III. Indicators
		Attributes	Components of attributes	
Schedler (2002a)	<ul style="list-style-type: none"> Conceptualizations of election quality (process-based) - Object of choice - Range of choice - Formation of preferences - Agents of choice - Expression of preferences - Aggregation of preferences - Consequences of choice 	4. Competition	<ul style="list-style-type: none"> 4.1. De jure right to run for office 4.2. De facto conditions for the use of the right to run for office 	<ul style="list-style-type: none"> - De jure multi-party/candidate elections - De facto possibility to compete: <ul style="list-style-type: none"> - Registration requirements - Ballots accurate - Freedom from intimidation/coercion - Freedom of movement, expression, assembly, association - Alternative sources of information / media coverage - Use of state resources - Campaign finance regulation - Campaign activities - Access to dispute adjudication
Mozaffar & Schedler (2002)	<ul style="list-style-type: none"> - Rule-making (choosing rules of electoral game) - Rule application (organizing electoral game) - Rule adjudication (certifying results & solving disputes) 			
Elklit & Reynolds (2005a)	<ul style="list-style-type: none"> - Legal framework - Electoral management - Constituency & polling district demarcation - Voter education - Voter registration - Ballot paper design & party/candidate registration - Campaign regulation - Polling - Counting and tabulating the vote - Resolving election related complaints, verification & certification - Post-election procedures 	5. Free elections	<ul style="list-style-type: none"> 5.1. Freedom to form preferences and vote free of coercion (for voters) 5.2. Freedom to run in elections and compete for popular support free of coercion (for parties/candidates) 	<ul style="list-style-type: none"> - Freedom of movement, expression, information. - Freedom of movement, expression, information, assembly, and association.
Birch (2008, 2011)	<ul style="list-style-type: none"> - Manipulation election rules - Manipulation voting process / election administration - Manipulation of vote choice 	6. Fair elections	6.1. "Unbiased application of rules" of electoral competition & electoral governance before, during and after elections.	<ul style="list-style-type: none"> - Impartial Electoral Management Body - Implementation according to electoral law: <ul style="list-style-type: none"> - Before elections: <ul style="list-style-type: none"> - Constituency & polling district demarcation - Ballot paper design - Registration of voters and parties - Campaign regulation - During elections: <ul style="list-style-type: none"> - Polling - After elections: <ul style="list-style-type: none"> - Counting and tabulating the vote - Resolving election complaints - Post-election procedures
Universal Declaration of Human Rights (UDHR 1948, art. 21, 1 & 3)	<p>International legal conventions^b</p> <ul style="list-style-type: none"> - Everyone has the right to take part in the government of his country, directly or through freely chosen representatives; - The will of the people shall be the basis of the authority of government, this will shall be expressed in: periodic and genuine elections, which shall be by universal and equal suffrage, and shall be held by secret vote or by equivalent free voting procedures 			
International Covenant on Civil and Political Rights (ICCPR 1966, art.25a&b)	<ul style="list-style-type: none"> - Every citizen shall have the right and the opportunity, [...] without unreasonable restrictions: to take part in the conduct of public affairs, directly or through freely chosen representatives; to vote and to be elected at genuine periodic elections, which shall be by universal and equal suffrage and shall be held by secret ballot, guaranteeing the free expression of the will of the electors. 			

a. The studies by Anglin (1998), Pastor (1999), Van de Walle (2003), Hartlyn et al (2008), Donno (2010), Kelley and Kiril (2010) and Hyde and Marinov (2010) are not included here because they do not provide an elaborate conceptualization of election quality, but rather give an operational definition based on how wide-spread, intentional and consequential irregularities are, or, in the case of Pastor, whether all major parties accept the results. b. For reasons of space, I only report the conditions mentioned for elections in the UDHR and ICCPR. Other instruments are described in the text.

Appendix B – Sample selection

Elections were selected on the basis of the following criteria (cf. paragraph 2.3):

- taking place in an independent nation-state
- elections held for both legislative and executive (directly or indirectly)
- irreversible elections
- legally possible for multiple parties to participate in elections (i.e. de jure multi-party elections)
- legally possible for all adult citizens to participate in elections (i.e. de jure universal suffrage)

The first criterion implies that elections should be for a sovereign national government that can govern without intervention of international actors or other states. This means that elections held before independence in former colonies or protectorates, even when taking place under conditions of de jure multi-party competition and universal suffrage, are excluded from our sample, as are elections held in international protectorates, such as Bosnia-Herzegovina after the 1995 Dayton Accords. The second criterion implies that both the executive and the legislative should be selected through elections, either directly or indirectly. This means that elections held under military governments holding only parliamentary elections, such as Brazil between 1964 and 1982, or elections held under monarchy governments holding only parliamentary elections, such as Swaziland, are excluded. The third criterion means that elections immediately followed by a coup d'état and/or civil war that prevent the elected government from taking power are excluded. Elections held under conditions

of civil war where nevertheless the elected government was able to assume power are included, as are elections held just after a coup d'état. The fourth criterion implies that elections should be held under de jure multi-party conditions, i.e. it should be legally possible for multiple parties or candidates to contest elections. This means that elections held under a constitution forbidding the formation of political parties, such as Sudan between 1989 and 1998, or allowing for only one state party, like Malawi between 1966 and 1993, are excluded. Finally, the fifth criterion implies that elections should allow de jure universal suffrage, i.e. it should be legally possible for all adult citizens to participate in elections. This means that elections where suffrage was restricted on the basis of gender, race, income or literacy were excluded. The absence of a de jure secret ballot is also considered as a suffrage restriction, as public voting limits the possibilities for opposition supporters to cast a free vote. Elections where secret ballots were not legally required were hence also excluded. Elections with suffrage restrictions for non-adults, detainees, and specific occupations (such as police or army officials) were included. As elaborated in chapter 2, these criteria are meant to capture the full range of variation in election quality, thus merely specifying the minimal conditions that need to be met for elections to be “potentially democratic”. This means that elections held under conditions of de facto limitations on party competition and voter participation are included, as long as the de jure conditions specified above hold. The only exception made in this regard are those cases where, despite de jure provisions for multi-party competition, only one party or one candidate contests the elections. The reason for excluding de facto one party or one candidate elections is that such elections do not provide voters with a choice, not even a pseudo-choice, and should therefore be considered non-democratic rather than “potentially democratic”.

The sample selection was made using data from the data handbooks by Nohlen et al. on Africa (1999), Central Asia (2001), South and Central America (2005), and Europe (2010), as well as the data handbook by Rose and Munro on Central and Eastern Europe (2003). Data for more recent elections, when not included in the Data Handbooks, were gathered from the Parline Database of the Inter-Parliamentary Union (IPU), The Election Guide of the International Federation for Electoral Systems (IFES), and US State Department country profile information. For each country, a short historical file was made tracking the country's authoritarian and democratic past, explaining coding decisions and referencing sources if needed (available upon request from the author). The sample was subsequently limited to elections having taken place in the period of the third wave, i.e. 1974-2009. In order to track how election quality improves with successive elections, each election was assigned a sequence number, starting from the first de jure multi-party, universal suffrage, and legislative/executive election, and counting until the last elections held before 2010. If potentially democratic elections had been held in the period before 1974, this is indicated as 'previous democratic experience' and the sequence number of the first election after 1974 is adapted accordingly. Finally, elections for constituent assemblies were included but treated separately in the analyses, because constituent assemblies are generally elected with the specific mandate to draft a new constitution, and hence are not expected to engage in 'normal' legislative policy-making.

Southern Europe (N = 3)							
Country	Independent nation state	Legislative elected	Executive elected	De jure multi-party & universal suffrage	Constituent assembly	Irreversibility elections	Previous democratic experience
Greece	1830	Yes, directly	Yes, indirectly	1974-present	-	Yes	Yes (i.e. 1956-1964 elections, Mackie and Rose, p. 186)
Portugal	1140	Yes, directly	Yes, directly	1975-present	1975	Yes	No
Spain	1512	Yes, directly	Yes, indirectly	1977-present	1977	Yes	Yes (i.e. 1931-1936 elections, Mackie and Rose, p. 387)

Central and Eastern Europe (N = 16)

Country	Independent nation state	Legislative elected	Executive elected	De jure multi-party & universal suffrage	Constituent assembly	Irreversibility elections	Previous democratic experience
Albania	1912	Yes, directly	Yes, indirectly	1991-present	1991	Yes	No
Bulgaria	1908	Yes, directly	Yes, directly	1990-present	1990	Yes	No
Croatia	1991	Yes, directly	Yes, directly	1992-present	-	No, civil war 1991-1995	No
Czech Rep.	1993	Yes, directly	Yes, indirectly	1996-present	-	Yes	Yes (1920-1935 Rose and Munro, p. 84)
Estonia ^a	1991	Yes, directly	Yes, indirectly ^b	1992-present	-	Yes	Yes (1920-1933 Rose and Munro, p. 129)
Hungary	1920	Yes, directly	Yes, indirectly	1990-present	-	Yes	Yes (1945-1949 Rose and Munro, p. 144/145)
Latvia ^a	1991	Yes, directly	Yes, indirectly	1993-present	-	Yes	Yes (1922-1934 Rose and Munro, p. 157)
Lithuania ^a	1991	Yes, directly	Yes, directly	1992-present	-	Yes	Yes (1920-1926 Rose and Munro, p. 173)
Macedonia	1991	Yes, directly	Yes, directly	1994-present	-	No, civil war 2000/2001	No
Montenegro	2006	Yes, directly	Yes, directly	2006-present	-	Yes	Yes (1992-2006 Serbia & Montenegro)
Poland	1918	Yes, directly	Yes, directly	1990-present	-	Yes	Yes (1922-1926 Rose and Munro, p. 191)
Romania	1878	Yes, directly	Yes, directly	1990-present	-	Yes	No
Serbia	2006	Yes, directly	Yes, directly	2006-present	-	Yes ^c	Yes (1992-2006 Serbia & Montenegro)
Serbia-Montenegro	1992	Yes, directly	Yes, indirectly until 2000, directly since	1992-2005	-	No, civil war 1992-1995 & 1998-1999 & 2000 "Bulldozer Revolution"	No
Slovakia	1993	Yes, directly	Yes, indirectly until 1999, directly since	1994-present	-	Yes	Yes (1920-1935 Rose and Munro, p. 84)
Slovenia	1991	Yes, directly	Yes, directly	1992-present	-	Yes	No

a. The three Baltic states are Former Soviet Republics. However, since they now are members of the European Union, I group them under Central and Eastern Europe.

b. Note that the 1992 presidential elections were based on a direct popular vote, as the Estonian constitution requires "the first president to be elected for a four-year term by a popular vote" [...] "and subsequent elections to be by parliament for a fixed five-year term" (Rose and Munro, 2009: 131).

c. However, there is continued disagreement over the status of Kosovo after the latter declared independence in 2008.

Former Soviet Republics (N = 12) ^a							
Country	Independent nation state	Legislative elected	Executive elected	De jure multi-party & universal suffrage	Constituent assembly	Irreversibility elections	Previous democratic experience
Armenia	1991	Yes, directly	Yes, directly	1991-present	-	No, Nagorno-Karabach conflict 1991-1994.	Yes (1919 parliamentary elections, Nohlen: p. 319 & 322)
Azerbaijan	1991	Yes, directly	Yes, directly	1992-present	-	No, Nagorno-Karabach conflict 1991-1994 & coup d'etat 1993	No
Belarus	1991	Yes, directly	Yes, directly	1994-present	-	Yes	No
Georgia	1991	Yes, directly	Yes, directly (except 1992-1995 indirectly)	1991-present	-	No, coup d'etat 1992 & civil war Abkhazia 1992-1993 & conflicts Abkhazia and South Ossetia ongoing & "Rose Revolution" 2003	Yes (1919 constituent assembly elections, Nohlen p. 374)
Kazakhstan	1991	Yes, directly	Yes, directly	1994-present	-	Yes	No
Kyrgyzstan	1991	Yes, directly	Yes, directly	1995-present	-	No, "Tulip Revolution" 2005	No
Moldova	1991	Yes, directly	Yes, directly before 2000, indirectly since	1994-present	-	Yes	No
Mongolia	1921	Yes, directly	Yes, indirectly in 1990, directly since	1990-present	-	Yes	No
Russian Federation	1991	Yes, directly	Yes, directly	1993-present	-	Yes	No
Tajikistan ^b	1991	Yes, directly	Yes, directly	1991-present	-	No, civil war 1992-1997	No
Turkmenistan	1991	Yes, directly	Yes, directly	excluded	-	-	No
Ukraine	1991	Yes, directly	Yes, directly	1991-present	-	No, "Orange Revolution" 2004	No
Uzbekistan	1991	Yes, directly	Yes, directly	1991-present	-	Yes	No

a. Elections in most Former Soviet Republics were de jure multi-party elections, but in practice limitations on opposition are substantial, with incumbent strategies ranging from outright banning of parties to participate in elections (even if not banning all parties), to hindering opposition party participation through legal procedures as not accepting candidates, taxation problems, illegal campaign finance, etc. (IPU). Examples are Azerbaijan, Belarus, Georgia, Kazakhstan, Kyrgyzstan, Russia after 1999, Uzbekistan and Tajikistan. The only country in this sample where elections remained de jure single party contests after independence from the Soviet Union is Turkmenistan. In 2007 the constitution was changed to allow for multiple parties to run in elections, however since then only one election has been held (2008), hence this country is excluded from our sample.

b. There has been an interruption of de jure multi-party elections in Tajikistan between 1993 and 1995 (Nohlen, p. 456).

Sub-Saharan Africa (N = 45)

Country	Independent nation state	Legislative elected	Executive elected	De jure multi-party & universal suffrage	Constituent assembly	Irreversibility elections	Previous democratic experience
Angola	1975	Yes, directly	Yes, directly	1992 and 2002-present ^a	-	No, civil war 1992-2002	No
Benin ^b	1960	Yes, directly	Yes, directly	1991-present	-	Yes	Yes (1960-1964 Nohlen, p. 80)
Botswana	1966	Yes, directly	Yes, indirectly	1974-present	-	Yes	Yes (1969 parliamentary elections, Nohlen, p. 109)
Burkina Faso ^c	1960	Yes, directly	Yes, directly	1978 1992-present	-	Yes	Yes (1970 parliamentary elections, Nohlen, p. 124)
Burundi	1962	Yes, directly	Yes, directly in 1993 & indirectly in 2005	1993 2005-present	-	No, civil war 1993-2006, coup d'etat 1996	No
Cameroon	1961	Yes, directly	Yes, directly	1992-present	-	Yes	Yes (1964 legislative elections, Nohlen, p. 168)
Cape Verde	1975	Yes, directly	Yes, directly	1991-present	-	Yes	No
Central African Republic	1960	Yes, directly	Yes, directly	1981 1993-present	-	No, civil war in 1996, coup d'etat in 2003, and civil war 2006-2008	No
Chad	1960	Yes, directly	Yes, directly	1996-present	-	No, civil war/war with Sudan 2003-2010	No
Comoros	1975	Yes, directly	Yes, directly	1990-present	-	No, coup d'etat in 1995 and 1999 ^d	No
Congo (Brazzaville)	1960	Yes, directly	Yes, directly	1992-present	-	No, civil war in 1993 and 1997, coup d'etat in 1997, civil war in 1998/1999.	No
Congo (Democratic Republic)	1960	Yes, directly	Yes, directly	2006-present	-	Yes	No
Côte d'Ivoire	1960	Yes, directly	Yes, directly	1990-present	-	No, coup d'etat in 1999, civil war from 2002-2009	No
Djibouti ^e	1977	Yes, directly	Yes, directly	1992-2003	-	No, civil war 1991-1994	No
Equatorial Guinea	1968	Yes, directly	Yes, directly	1993-present	-	Yes	No
Eritrea	1993	Yes, directly ^f	Yes, indirectly	excluded	-	-	No
Ethiopia	n.a. ^g	Yes, directly	Yes, indirectly	1994-present	1994	No, war with Eritrea 1998-2000	No
Gabon	1960	Yes, directly	Yes, directly	1990-present	-	Yes	Yes, 1964 legislative elections (Nohlen p. 388)
Gambia ^b	1965	Yes, directly	Yes, indirectly until 1982 & directly since	1977-present	-	No, coup d'etat 1994	Yes, 1966 and 1972 legislative elections (Nohlen, p. 418)
Ghana	1957	Yes, directly	Yes, directly	1979 1992-present	-	No, coup d'etat 1981	Yes, 1960 presidential elections, 1969 legislative elections (Nohlen, p. 424)
Guinea	1958	Yes, directly	Yes, directly	1993-present	-	No, war in 2000 and coup d'etat in 2008	No
Guinea-Bissau	1974	Yes, directly	Yes, directly	1994-present	-	No, civil war in 1998 and coup d'etat in 1999 & 2003	No
Kenya	1963	Yes, directly	Yes, directly	1992-present	-	Yes	Yes, 1966 legislative elections (Nohlen, p. 485)

Sub-Saharan Africa (N = 45)

Country	Independent nation state	Legislative elected	Executive elected	De jure multi-party & universal suffrage	Constituent assembly	Irreversibility elections	Previous democratic experience
Lesotho ^h	1966	Yes, directly	Yes, indirectly	1993-present	-	No, coup d'etat in 1994 and civil war in 1998-1999	No
Liberia	1847	Yes, directly	Yes, directly	1985 1997-present	-	No, civil war 1989-1997 & 2003	No ⁱ
Madagascar ^l	1960	Yes, directly	Yes, directly	1977-2009	-	No, civil war 2001 and coup d'etat 2009	Yes, legislative elections 1960, 1965, 1970 (Nohlen, p. 541/542)
Malawi	1964	Yes, directly	Yes, directly	1994-present	-	Yes	No
Mali	1960	Yes, directly	Yes, directly	1992-present	-	Yes	No
Mauritania	1960	Yes, directly	Yes, directly	1992-present	-	No, coup d'etat 2005 & 2008	No
Mauritius	1968	Yes, directly	Yes, indirectly	1976-present	-	Yes	No
Mozambique	1975	Yes, directly	Yes, directly	1994-present	-	Yes	No
Namibia	1990	Yes, directly	Yes, directly	1994-present	-	Yes	No
Niger	1960	Yes, directly	Yes, directly	1993-present	-	No, coup d'etat 1996 & 1999 (& 2010)	No
Nigeria	1960	Yes, directly	Yes, directly	1979-1983 1998-present	-	No, coup d'etat 1983 & 1993 & 1998	No
Rwanda	1962	Yes, directly	Yes, directly	2003-present	-	Yes	No
Sao Tome & Principe ^k	1975	Yes, directly	Yes, directly	1991-present	-	No, coup d'etat 1995 & 2003	No
Senegal	1960	Yes, directly	Yes, directly	1993-present	-	Yes	Yes, 1963 legislative elections (Nohlen, p. 765)
Seychelles	1976	Yes, directly	Yes, directly	1992-present	1992	Yes	No
Sierra Leone	1961	Yes, directly	Yes, directly	1977 1996-present	-	No, one-party rule from 1978-1985 & coup d'etat in 1997 & civil war 1998-2002	Yes, 1962 and 1967 legislative elections (Nohlen, p. 795)
Somalia ^l	1960	-	-	excluded	-	-	Yes, 1964 and 1969 legislative elections (Nohlen, p. 811)
South Africa ^m	1910	Yes, directly	Yes, indirectly	1994-present	1994	Yes	No
Sudan	1956	Yes, directly	Yes, directly	1986 2000	1986 ⁿ	No, coup d'etat 1989 & civil war 1983-2005	Yes, constituent assembly elections in 1965 & 1968 (Nohlen, p. 851)
Swaziland ^o	1968	Yes, directly	No	excluded	-	-	No
Tanzania	1961	Yes, directly	Yes, directly	1995-present	-	Yes	Yes, 1962 presidential elections (Nohlen, p. 883)
Togo	1960	Yes, directly	Yes, directly	1993-present	-	Yes	No
Uganda ^p	1962	Yes, directly	Yes, directly (indirectly in 1980)	1980 2006-present	-	No, coup in 1985 & 1986	No
Zambia	1964	Yes, directly	Yes, directly	1991-present	-	Yes	Yes, 1968 presidential & legislative elections (Nohlen, p. 940)
Zimbabwe	1980	Yes, directly	Yes, directly	1990-present	-	Yes	No

- a. De jure multi-party and universal suffrage direct elections for parliament and president, however only elections held were parliamentary elections in 2008, presidential elections still postponed.
- b. Our sample starts at the year indicated as the start of the “third wave”, i.e. 1974. In the case of Benin, this means that the de jure multi-party and universal suffrage parliamentary and presidential elections that took place in 1960 and 1964 are not included. Instead, “previous democratic experience” is coded as yes, and since the 1964 elections were interrupted by a coup d’etat in 1965, the sequence number of the elections in 1991 starts again at 1. Idem for, for example Burkina Faso, but since in that case elections were not interrupted by civil war or coup d’etat (but by the president dissolving parliament, redrafting the constitution and holding new elections), the sequence number of the 1978 parliamentary elections starts at 2. Likewise, for Gambia the sequence number of the 1977 parliamentary elections is 3.
- c. The presidential elections in 1991 in Burkina Faso were de jure multi-party elections, but due to the opposition parties boycotting the elections, resulted in de facto one-candidate elections (Nohlen, p. 125). For this reason these elections were excluded from the sample.
- d. The 1995 coup d’etat was unsuccessful due to military intervention by France. However, President Djohar was “de facto deprived of power” and new presidential elections were organized in 1996 (Nohlen, p. 245).
- e. The 2005 presidential elections and 2008 parliamentary elections were de facto single candidate and single party elections due to opposition boycotts. Hence these elections were excluded from the sample.
- f. The constitution of Eritrea has been ratified in 1997, but not been implemented. Hence, even though the constitution foresees a directly elected National Assembly and an indirectly elected president, elections have not been held since independence, i.e. Eritrea is not included in our sample.
- g. With the exception of 1935-1941, when Ethiopia was occupied by Italy, Ethiopia was never dominated by external powers (Nohlen, p. 373).
- h. According to US State Department, Lesotho is a constitutional monarchy, with the monarch serving as head of state and the prime minister as head of government and cabinet. (source: <http://www.state.gov/r/pa/ei/bgn/2831.htm>). Hence Lesotho is included in our sample.
- i. Liberia has had an elected legislative and executive for many years before the coup d’etat in 1980. However, competition was limited and even though not de jure single-party, the True Wig Party had a de facto monopoly on power. The reason pre-1980 Liberia is not included in our sample and not considered democratic is however that universal suffrage, though extended to women, was restricted to landowners, thereby excluding a large part of the citizenry from casting a vote.
- j. The 2001 events were coded as civil war because considerable violence was involved, but should be rather considered as a mix between an electoral revolution and a coup d’etat. Following the 2001 presidential polls, as IPU describes, a “heated dispute over the election returns rapidly developed into a serious political crisis during the first half of 2002. On the strength of his broad-based support, Mr. Marc Ravalomanana was able to take over as Head of State, forcing outgoing President Didier Ratsiraka into exile” (source: http://www.ipu.org/parline-e/reports/arc/2193_02.htm). The international community expressed preference for holding new elections, which however did not take place.
- k. Both coup d’etats that occurred in Sao Tome and Principe were reversed by international and national mediation shortly after the coups, thus restoring constitutional order.
- l. Somalia has held multi-party elections since independence only between 1960 and 1969, and is hence not included in our sample.
- m. South Africa has a rather long history of multi-party elections, however since suffrage was limited to whites and colored people until 1994, the criterion of universal suffrage was only met in 1994 (Nohlen, p. 817 & 820).
- n. According to Nohlen (p. 851) elected both as constituent and legislative Assembly.
- o. Even though since 1993 Swaziland has a (mostly) directly elected legislature, executive power lies with the unelected monarchy and political parties were constitutionally banned until 2006, hence Swaziland is not included in our sample.
- p. The 1996 and 2001 presidential and parliamentary elections were all de facto multi-candidate elections, however, since candidates were forbidden to run under a political party label, and this implies that de jure multi-party competition was not allowed, these elections are not included in our sample.

South America (N = 12) ^a							
Country	Independent nation state	Legislative elected	Executive elected	De jure multi-party & universal suffrage	Constituent assembly	Irreversibility elections	Previous democratic experience
Argentina	1816	Yes, directly	Yes, directly	1983-present	-	Yes	Yes (i.e. 1951, 1958, 1963, 1973 presidential and parliamentary elections, 1957 constituent assembly elections, Nohlen p. 67)
Bolivia	1825	Yes, directly	Yes, directly	1978-present	-	No, coup d'etat 1978, 1979, 1980, 1981, 1982	Yes (i.e. presidential and parliamentary elections 1956-1964 and 1966, Nohlen, p. 133)
Brazil	1822	Yes, directly	Yes, directly	1989-present	-	Yes	No
Chile	1818	Yes, directly	Yes, directly	1989-present	-	Yes	Yes (i.e. parliamentary elections 1973, Nohlen, p. 254)
Colombia	1819	Yes, directly	Yes, directly	1974-present	1990	No, civil war 1974-present (Nohlen, p. 297) ^b	Yes (i.e. presidential and parliamentary elections 1957-1974, Nohlen, p. 305) ^c
Ecuador	1830	Yes, directly	Yes, directly	1978-present	1997 & 2007	No, coup d'etat 1997, 2000 and 2005 ^d	No
Guyana	1966	Yes, directly	Yes, indirectly	1980-present	-	Yes	Yes (i.e. parliamentary elections 1968 & 1973, Nohlen, p. 363)
Paraguay	1811	Yes, directly	Yes, directly	1978-present	1991	No, coup d'etat 1989	Yes (i.e. presidential and parliamentary elections 1963, 1968, 1973 & constituent assembly elections 1967, Nohlen, p. 425)
Peru	1821	Yes, directly	Yes, directly	1980-present	1992	No, coup d'etat 1992 ^e	No
Suriname	1975	Yes, directly	Yes, indirectly	1977-present	-	No, coup d'etat 1980 & 1990	No
Uruguay	1825	Yes, directly	Yes, directly	1984-present	-	Yes	Yes (i.e. presidential and parliamentary elections from 1934-1973, Nohlen, p. 494)
Venezuela	1830	Yes, directly	Yes, directly	1978-present	1999	Yes	Yes (i.e. presidential and parliamentary elections 1947 & constituent assembly elections 1946 & presidential and parliamentary elections from 1958-1973 (Nohlen, p. 537)

a. Many Latin American countries started their first democratic experiences early, and countries like Argentina, Chile, Colombia and Uruguay were part of the first long “wave” of democratization between 1828 and 1926 (Nohlen, 2005: 7). However, as in many European countries at the time, suffrage restrictions were widespread. As we only take into account elections that are de jure competitive (i.e. multi-party) and de jure inclusive (i.e. universal suffrage), the period of elections is delimited by universal suffrage for men and women. Moreover, since we only include the time-period of the third wave, only elections as of 1974 were included.

b. Colombia has experienced political violence since the creation and consolidation of guerilla groups in the course of the 1960s, and the subsequent formation of paramilitary groups and drug cartels in the 1980s (Nohlen, p. 297). While peace agreements with some groups have been signed and the 1991 constitution was changed with the aim of including demobilized guerilla groups in politics, the two most important organization, the FARC and ELN have thus far refused to participate and negotiations and fighting are still ongoing (US State Department <http://www.state.gov/r/pa/ei/bgn/35754.htm>).

c. Though these elections were de jure multi-party and universal suffrage elections, their consequentiality was limited due to the Frente Nacional agreement between the two main political parties, that agreed on alternation in government between them every 4 years (Nohlen, p. 296).

d. In all three occasions, Congress removed the president from office, either after institutional deadlock or in reaction to popular protests. However, according to the Data Handbooks, in all cases the support of the military was crucial in enabling Congress to remove the president, and hence these events are coded as coup d'etats.

e. On April 5, 1992, President Fujimore dissolved the parliament in an “auto-golpe”, suspended the constitution, and scheduled elections for a constituent assembly which was to draft a new constitution, which was subsequently approved by voters in a 1993 referendum (Nohlen, 448/449).

Central America (N = 8)							
Country	Independent nation state	Legislative elected	Executive elected	De jure multi-party & universal suffrage	Constituent assembly	Irreversibility elections	Previous democratic experience
Belize	1981	Yes, directly	Yes, indirectly	1984-present	-	Yes	No
Costa Rica	1821	Yes, directly	Yes, directly	1974-present	-	Yes	Yes (i.e. presidential and parliamentary elections from 1949-1970, Nohlen, p. 155)
El Salvador	1821	Yes, directly	Yes, directly	1974-present	1982	No, coup d'etat 1979 & civil war from 1979-1992	Yes (i.e. presidential and parliamentary elections 1964-1972 Nohlen, p. 176)
Guatemala	1821	Yes, directly	Yes, directly	1974-present	1984	No, coup d'etat 1982, 1983 & civil war 1982-1996	Yes (i.e. presidential and parliamentary elections from 1958-1961 & from 1964-1970, Nohlen, p. 318)
Honduras	1821	Yes, directly	Yes, directly	1980-present	1980	No, coup d'etat 2009	Yes, (i.e. presidential and parliamentary elections in 1954, 1971 & constituent assembly elections in 1956, 1957, 1965, Nohlen, p. 407) ^a
Mexico	1822	Yes, directly	Yes, directly	1976-present	-	Yes	Yes (i.e. presidential and parliamentary elections from 1955-1973, except 1976 presidential elections, Nohlen, p. 453).
Nicaragua	1821	Yes, directly	Yes, directly	1974 1984 1990-present	1984	No, revolution 1979 & civil war 1980-1990	Yes (i.e. presidential and parliamentary elections from 1963-1972, Nohlen p. 489)
Panama	1903	Yes, directly	Yes, directly	1984-present	-	No, US intervention 1989	Yes (i.e. presidential and parliamentary elections from 1946-1968, Nohlen, p. 518)

a. Though many of these elections were fraudulent and took place in a context of oppression of political opposition and frequent coup d'etats by the military, they were de jure multi-party and universal suffrage (Nohlen, p. 410/411).

Appendix C – Measuring election quality

The data on election quality used for the context of this thesis is based on the 7 datasets created by Anglin (1998), Van de Walle (2003), Lindberg (2006), Birch (2008/2011), Hartlyn et al. (2008), Donno (2009), and Kelley and Kiril (2010). This data was matched to our original database including all 886 “potentially democratic elections” in the 6 regions between 1974 and 2009, resulting in election quality scores for 735 legislative and executive elections in Central and Eastern Europe (CEE), Former Soviet Republics (FSR), Sub-Saharan Africa (SSA), South and Central America (SA and CA), and Southern Europe (SE), i.e. 83% of the total number of elections. Data from Munck (2009) and Hyde and Marinov (2011) were originally also matched, but dropped after the validity tests reported in chapter 2. The remaining 17% of missing elections were coded by the author in the period from February-April 2011. Of the 7 datasets, on average about 10% of election quality data was dropped. Reasons to drop elections were: (1) differences in sample (i.e. Bosnia-Herzegovina was excluded due to its status as international protectorate and Caribbean countries were excluded because the election quality datasets only covered a very limited number of elections in countries in this region), (2) differences in types of elections included (i.e. when elections were not legislative or executive elections, but for example referenda or Senate elections, they were not included; and concurrent elections were matched separately as legislative and executive elections); (3) coding of 1st and 2nd rounds in presidential elections as separate elections (i.e. I coded those as single elections). As such, the relatively higher proportion of elections dropped for Hartlyn et al. (2008) are due to the fact that they coded 1st and 2nd rounds of presidential elections as separate elections, and the higher proportion of elections dropped for Kelley (2010) are due to the fact that a number of elections in this dataset did not match my criteria for ‘potentially democratic elections’ and because Kelley (2010) includes elections as of 1960. Every election that was not matched has been recorded in a file

available upon request from the author. Tables 1 and 2 below show the number of elections originally coded by authors and the number of elections matched.

Table C1 – Datasets measuring the quality of elections in new democracies

Dataset	CEE	FSR	SSA	SA	CA	SE	Total	Years
Birch 2008*	48	46	26	28	19		167 (136)	1995-2006
Birch 2010	51	51	26	37	22		187 (155)	1995-2007
Donno 2010* (dummy)	102	64		80	53		299 (265)	1990-2005
Hartlyn et al 2008				65	29		94 (104)	1980-2003
Lindberg 2006			232				232	1969-2003
Lindberg update			50				50	2004-2007
Munck 2009				99	65		164	1984-2005
Kelley and Kiril 2010	131	73	343	141	95	32	815 (661)	1978 - 2004
Hyde & Marinov 2010	132	95	395	164	111	38	935	1960-2006
Anglin 1998			25				25	1996
Van de Walle 2003			87				87	1989-2000

* Concurrent elections are coded as legislative and presidential elections, hence I include them twice in my dataset. Between brackets is the original number of elections coded (including elections in the Carribean in the case of Birch, Donno and Hartlyn et al). Donno coded in total 313 elections, 48 of which are excluded because they were referenda.

Table C2 – Datasets measuring the quality of elections in new democracies – data included.

Dataset	CEE	FSR	SSA	SA	CA	SE	Total	% dropped
Birch 2008	48	46	25	25	19		163	2.4%
Birch update 2011	51	51	25	33	22		182	2.7%
Donno 2009	92	61		78	52		283	5.4%
Hartlyn et al 2008				48	25		73	22.3%
Lindberg 2006			208				208	10.3%
Lindberg update (2004-2007)			49				49	2%
Munck 2009				98	65		163	0.6%
Kelley 2010	92	68	226	135	93	32	646	20.7%
Hyde & Marinov 2011	105	80	253	156	106	38	738	20.1%
Anglin (1998)			21					16%
Pastor (1999)	8	8	37	1	2		56	9.7%
Van de Walle (2003)			82				82	5.7%
Average dropped								9.8%
Total elections w EQ score	102	82	272	147	100	32	735	

After merging the data, the different election quality scores were combined into a single overall election quality score. Combining the scores was done by first recoding election quality scores of each author so as to generate variables that ran from low to high election EQ quality, and second by standardizing the scores to as to all vary from 0 to 1. Following the analyses reported in chapter 2, I left all data in their original scales (either 3-point, 4-point or 5-point ordinal), with the exception of the data by Kelley that seemed to perform better if the categories of “no problems” and “minor problems” were recoded to a single “high quality” score. For each election with multiple scores, the election quality score used for analyses is the mean. Elections with high disagreement were recoded by the author. Table C3 below shows the original coding by each author as well as the way the variables were recoded.

Table C3 – Election quality scores

Author	Original coding	Original coding	Recoding
Birch 2008, 2011 (reversed coding for analyses)	1 – Lowest degree of malpractice 2 3 4 5 – Highest degree of malpractice Answer to the question “How does the [election observation mission] report describe the probity of the election overall?”, where the election is coded as 1 if the election was deemed “substantially free and fair”, as 3 if “there were some violations of the freeness and/or fairness of the election, but not enough to alter the outcome”, and as 5 if “the freeness and/or fairness of the election was severely compromised, such that the validity of the outcome was in doubt.” The scores 2 and 4 were used for intermediate situations ¹³⁵ . (Methodology for the Construction of the Index of ElectoralMalpractice, http://www.essex.ac.uk/government/electoralmalpractice/methodology.htm#Methodology_Construction)	From high to low EQ 1 – Lowest degree of malpractice 2 3 4 5 – Highest degree of malpractice	From low to high EQ 1 = 1 “high quality” 2 = 0.75 3 = 0.5 “medium quality” 4 = 0.25 5 = 0 “low quality”

¹³⁵ “In general, isolated problems suggest a score of ‘2’, widespread problems with pockets of good conduct suggest a score of ‘4’ and problems that are on a scale in between ‘isolated’ and ‘widespread’ suggest a score of ‘3’. If there are no problems or only minor problems you should code ‘1’, whereas comprehensive failure to observe the norms of transparency, impartiality, and rule of law suggest a score of ‘5’.” (Methodology for the Construction of the Index of ElectoralMalpractice, http://www.essex.ac.uk/government/electoralmalpractice/methodology.htm#Methodology_Construction)

Author	Original coding	Original coding	Recoding
Donno 2010 (reversed coding for analyses)	Dichotomous variable 0 – Not flawed 1 – Flawed Ordinal variable: 0 – Clean 1 – Irregular 2 – Reject “Election observers’ verdict is coded as clean if the election was deemed to be without misconduct or, at most, isolated cases of procedural flaws were found in some districts or polling stations. The verdict is coded as irregular if systematic misconduct was found, but was attributed to weak administrative capacity or disorganization, or could not be definitely linked to intentional government manipulation. The observers’ verdict is coded as reject if they reported clear evidence that misconduct was intentionally perpetrated by the government, or if misconduct altered the outcome of the election.” (Donno 2009: 22-23)	From high to low EQ Ordinal variable: 0 – Clean 1 – Irregular 2 – Reject	From low to high EQ 0 = 1 “high quality” 1 = 0.5 “medium quality” 2 = 0 “low quality”
Hartlyn, McCoy, Mustillo 2008	0 – Unacceptable 1 – Flawed 2 – Acceptable “If the basic elements for procedural fairness and technical soundness are present to an important degree, then the election is deemed acceptable. If the process is deemed by observers to be egregiously deficient procedurally or technically, or sufficiently deficient to prevent the election results from reflecting the will of the voters freely expressed at the ballot box, the election is coded as unacceptable. The intermediate category of flawed is reserved for elections in which the process experienced significant procedural or technical deficiencies (whether due to incompetence, abuse of state resources or power or other extralegal resources, fraud or violence), but these are not clearly sufficient to affect the outcome of the voting for the presidential election”. (Hartlyn et al. 2008: 77).	From low to high EQ 0 – Unacceptable 1 – Flawed 2 – Acceptable	From low to high EQ 0 = 0 “low quality” 1 = 0.5 “medium quality” 2 = 1 “high quality”
Kelley and Kiril 2010 (reversed coding for analyses)	Overall election quality variable: 0 - Acceptable 0.5 - Ambiguous 1 - Unacceptable “The measure captures whether the State Department report, notwithstanding the level of problems (captured in the problems variable), considered the election acceptable.” The variable is coded 0 if “the report indicates that the election results represent the will of the people”, coded 0.5 if “the report is ambiguous about whether the elections represented the will of the people” and coded 1 if “the report indicates that the election results do not represent the will of the people” (However, cf. Quality of Elections Data codebook, June 2010, that seems to indicate a scoring from 0 to 2, which, given the description of codes in the text seems to be a mistake).	From high to low EQ Extent of problems variable: 0 – No problems 1 – Minor problems only 2 – Moderate problems 3 – Major problems	From low to high EQ 0 + 1 = 1 “high quality” 2 = 0.5 “medium quality” 3 = 0 “low quality”

Author	Original coding	Original coding	Recoding
Kelley and Kiril 2010 (reversed coding for analyses) (continued from previous page)	Extent of problems variable: 0 – No problems 1 – Minor problems 2 – Moderate problems 3 – Major problems “This variable assesses the extent of the problems in the election. This is a combined assessment that considers problems in the legal framework, political and administrative problems in the pre-election period, and then the integrity of the election day itself. A code of 0 is only used when no problems at all are mentioned, and a code of 0 or 1 can only be used in conjunction with a 0 for overall election quality. A code of 3 can only be used in conjunction with a 1 for overall election quality” (Quality of Elections Data codebook, June 2010).	From high to low EQ Extent of problems variable: 0 – No problems 1 – Minor problems only 2 – Moderate problems 3 – Major problems	From low to high EQ 0 + 1 = 1 “high quality” 2 = 0.5 “medium quality” 3 = 0 “low quality”
Munck 2009	0 – Major irregularities in the voting process that have a determinative effect on the result of elections (i.e. alteration in the election for the national executive and/or the balance of power in parliament); 1 – Significant irregularities in the voting process (e.g. intimidation of voters, violence against voters, electoral fraud) which do not, however, have a determinative effect on the results of elections; 2 – Lack of significant irregularities in the voting process (e.g. elections that might include “technical” irregularities but not any systematic bias of considerable weight).	From low to high EQ 0 – Major irregularities in the voting process that have a determinative effect on the result of elections; 1 – Significant irregularities in the voting process which do not, however, have a determinative effect on the results of elections; 2 – Lack of significant irregularities in the voting process.	From low to high EQ 0 = 0 “low quality” 1 = 0.5 “medium quality” 2 = 1 “high quality”
Anglin 1998 (reversed coding for analyses)	“Freeness and Fairness: classification made on a judgmental basis following a close study of available reports and opinions” (Anglin 1998: 481) Pre-Election: A. Substantially free and fair: only minor lapses (coded as 1) B. Freeness and/or fairness qualified: absence of substantially level playing field and/or fraudulent practices (coded as 2) C. Neither free nor fair: process deeply flawed (coded as 3) Election and Post-Election: 1. Substantially free and fair: minor administrative irregularities only 2. Not fully free and fair: significant administrative irregularities and/or fraudulent practices 3. Neither free nor fair: serious electoral abuses (Anglin 1998: 476)	From high to low EQ Pre-Election: 1. Substantially free and fair 2. Freeness and/or fairness qualified 3. Neither free nor fair Election and Post-Election: 1. Substantially free and fair 2. Not fully free and fair 3. Neither free nor fair	From low to high EQ (averaged both scores to generate one overall score) 1 = 1 “high quality” 2 = 0.5 “medium quality” 3 = 0 “low quality”
Van der Walle 2003	Were the elections free and fair? 0 – No 1 – Partly	From low to high EQ 0 – No 1 – Partly	From low to high EQ 0 = 0 “low quality” 1 = 0.5 “medium quality”

	2 - Yes	2 - Yes	2 = 1 "high quality"
Author	Original coding	Original coding	Recoding
Lindberg 2006	<p>0 – “No” when no elections have been held or those held have been wholly unfair; 1 – “Irregular” when there were serious defects that influenced the results; 2 – “Yes, irregularities not significant” when there were deficiencies but they did not affect the result; 3 – “Yes” when elections were free and fair</p>	<p>From low to high EQ 0 – “No”, wholly unfair; 1 – “Irregular” serious defects that influenced the results; 2 – “Yes, irregularities not significant” deficiencies but they did not affect the result; 3 – “Yes”, free and fair</p>	<p>From low to high EQ 0 = 0 “low quality” 1 = 0.33 2 = 0.66 3 = 1 “high quality”</p>
Hyde and Marinov (2010)	<p>The National Elections Across Democracy and Autocracy (NELDA) dataset includes 7 variables relating to election quality, all coded as: 1 – No 2 – Unclear 3 – Yes</p> <ol style="list-style-type: none"> (N11) Before elections, are there significant concerns that elections will not be free and fair? “A "yes" indicates that there was evidence of domestic or international concern that the election process was not going to be free or fair. A "yes" is also coded when the elections were widely perceived to lack basic criteria for competitive elections, such as more than one political party. (source: http://hyde.research.yale.edu/nelda/#variables) (N13) Were opposition leaders prevented from running? “A "yes" was coded when at least some opposition leaders were prevented from running and contesting the elections. A decision to boycott the election was coded "yes" here only if it was in response to the government preventing opposition figures from running. Cases where opposition was not allowed were also coded as "yes." (source: http://hyde.research.yale.edu/nelda/#variables) (N15) Is there evidence that the government harassed the opposition? “If there was evidence of intentional government harassment of the opposition, a "yes" was coded. Harassment may include detaining opposition leaders, disrupting opposition political rallies with state forces, and shutting down opposition newspapers and offices. If opposition was banned, or if there was no opposition, then an "N/A" was coded.” (source: http://hyde.research.yale.edu/nelda/#variables) (N16) In the run-up to the election, were there allegations of media bias in favor of the incumbent? “If there were reports by either domestic or outside actors of media bias in favor of the incumbent or ruling party, it is coded as a "yes." In 	<p>From high to low EQ 1 – No 2 – Unclear 3 – Yes</p>	<p>From low to high EQ 1 = 1 “high quality” 2 = . no data 3 = 0 “low quality”</p>

	<p>cases where the media is totally controlled by the government, and/or no opposition is allowed, the answer is "yes." It is possible that the answer is "no" even if the political system is tightly controlled." (source: http://hyde.research.yale.edu/nelda/#variables)</p> <p>5. (N30) If [there were riots and protests after the election], did they involve allegations of vote fraud? This variable "was coded as "yes" if the riots or protests are backed with allegations of vote fraud. If there are no allegations of vote fraud fueling the riots or protests, a "no" was coded." If there were no riots or protests, "N/A" was coded. (source: http://hyde.research.yale.edu/nelda/#variables)</p> <p>6. (N33) Was there significant violence involving civilian deaths immediately before, during, or after the election? "If there was any significant violence relating to the elections that resulted in civilian deaths, a "yes" is coded." (source: http://hyde.research.yale.edu/nelda/#variables)</p> <p>7. (N47) If [Western international monitors were present], were there allegations by Western monitors of significant vote-fraud? "If there were no Western monitors, or no international monitors, "N/A" was coded. If there were Western monitors present and there were allegations of significant vote-fraud by any Western monitors, then "yes" was coded. If there were no allegations of fraud, "no" was coded." (source: http://hyde.research.yale.edu/nelda/#variables)</p>		

Additional validity tests

Single indicators versus multiple indicators

For this purpose we used the datasets with disaggregate indicators of election quality, i.e. Kelley and Kiril 2010, Hyde and Marinov 2010 and Birch 2008, 2011¹³⁶. Table 2.11 below shows the models using data from Kelley and Kiril (2010). This dataset is based on information from US State Department Human

¹³⁶ Anglin (1998) only uses 2 indicators of election quality, i.e. before and during election-day, and coded only 25 elections, so is left out for these validity tests.

Right reports and includes 2 overall measures of election quality: one variable indicating whether the report considered the overall election acceptable, ambiguous or unacceptable based on references to “free- and fair-ness” of elections and/or whether elections “represented the will of the people”; and one variable indicating the extent of problems during the elections mentioned in the report, running from no problems to major problems. I ran ordinal logit regressions with all specific irregularities on each of these dependent variables, and since elections were clustered within countries, re-ran these models with fixed effects dummies for countries. Since all variables are scaled in the same direction, if all specific irregularities are taken into account in the overall judgment of election quality, we should expect all independent variables to be significant and have positive coefficients. This is indeed the case for the dependent variable that measures the extent of problems, even in a fixed effects model. However, the overall election quality variable seems to be less comprehensive, as pre-election and election-day violence and unrest and administrative capacity are not significant. Clearly the overall election quality score is less comprehensive as it is driven by fewer specific irregularities. Kelley (2010) herself indicates that indeed this variable might be more prone to political influence as it relies on the State Departments’ evaluation of the quality of elections, rather than considering the actual problems that occurred before and during the election. The second dependent variable, that measures the extent of problems in elections, seems hence to perform better in providing a comprehensive measurement of election quality and will be used in the remainder of this thesis¹³⁷. Considering differences in weighting of specific irregularities, since the scaling of all the variables is the same, the coefficients provide an indication of the magnitude of the effect of independent variables on

¹³⁷ Since these data have a time-series cross-sectional structure, with elections clustered in countries, tests for unit roots, serial autocorrelation and heteroskedasticity need to be done to make sure results for pooled models are robust. Neither of the two dependent variables appears to have a unit root, nor did the models suffer from significant heteroskedasticity. Serial autocorrelation of errors did appear to be a problem, however this could be solved by including the lagged dependent variable. Since the results are not substantially different if a lagged dependent variable is included in the model however, the simpler pooled model results are reported here.

the dependent variable. The quality of the legal framework, pre-election cheating and cheating on election-day seem to affect the overall election quality score more strongly than pre-election administrative capacity. This seems to be the case for both dependent variables.

Table 2.11. Overall judgments of election quality and specific irregularities

Specific irregularities	DV overall election quality (0-2)	DV extent of problems (0-3)	DV overall election quality (fixed effects)	DV extent of problems (fixed effects)
quality of legal framework (0-3)	1.842*** (0.145)	2.661*** (0.184)	2.418*** (0.324)	3.197*** (0.304)
pre-election cheating (0-3)	1.027*** (0.129)	1.921*** (0.150)	1.378*** (0.266)	2.161*** (0.222)
pre-election violence & unrest (0-3)	-0.051 (0.119)	1.480*** (0.111)	-0.423+ (0.241)	1.452*** (0.163)
pre-election administrative capacity (0-3)	0.453+ (0.254)	0.562* (0.232)	0.296 (0.362)	0.650+ (0.333)
election day cheating (0-3)	1.292*** (0.163)	1.714*** (0.164)	1.657*** (0.268)	1.965*** (0.235)
election day violence and unrest (0-3)	-0.134 (0.203)	0.839*** (0.184)	-0.107 (0.362)	1.352*** (0.267)
election day administrative capacity (0-3)	0.168 (0.190)	0.806*** (0.182)	0.445 (0.334)	1.263*** (0.262)
N	812	812	812	812
Pseudo-Rsquared	0.59	0.65	0.77	0.76

Ordinal logit analyses. Coefficients of country dummies for fixed effects not reported. Standard errors in parentheses. P-values: + 0.1, * 0.05, ** 0.01, *** 0.001. Data from QED dataset, Kelley and Kiril (2010).

Turning to the data collected by Hyde and Marinov (2010), overall election quality was measured by whether there had been allegations of significant vote-fraud by Western international monitors. The disaggregate indicators of election quality are based on broader sources such as academic data handbooks, and news and historical information. The variables were all dichotomous (yes/no) and coded in the same direction. Fixed effects were not possible due to limited within-country variation in the dependent variable, and hence clustered standard errors were used to correct for country-clustering.

Table 2.12. Overall judgments of election quality and specific irregularities

Specific irregularities	DV overall election quality (0-1)	DV overall election quality (0-1) (robust s.e.)
before elections, significant concerns elections not free & fair? (0-1)	2.371*** (0.450)	2.371*** (0.491)
opposition leaders prevented from running? (0-1)	0.709 (0.439)	0.709 (0.505)
evidence government harassed opposition? (0-1)	0.732+ (0.401)	0.732 (0.453)
before elections, allegations of media bias for incumbent? (0-1)	1.971*** (0.367)	1.971*** (0.421)
violence before, during or after elections? (0-1)	0.597 (0.376)	0.597 (0.514)
N	408	408
Pseudo-Rsquared	0.44	0.44

Logit regression. Robust standard errors clustered by country. Standard errors in parentheses. P-values: + 0.1, * 0.05, ** 0.01, *** 0.001. Data from NELDA dataset, Hyde and Marinov (2010). NELDA includes elections for all rounds of elections; for these analyses I selected the first rounds of presidential and legislative elections.

If the overall measure of election quality is comprehensive, we should again expect positive and significant coefficients for all the independent variables. Surprisingly however, the logit regressions below show that, even if all coefficients are positive, not all are significant. While prior concerns about the conduct of elections, allegations of media bias, and to a lesser extent government harassment of the opposition, are significantly related to the chances of Western international monitors alleging fraud, neither limitations on opposition access to the elections nor election violence is significantly associated with monitor's evaluations. Apparently, as has been suggested by previous research, violence reduces the chances of condemning flawed elections due to observers' fear that negative evaluations might fuel further violence (Bratton 2008, Collier and Vicente 2011, Kelley 2010). However, why limitations on opposition access to elections do not significantly affect evaluations of elections is a puzzle to me. Whether or not these omissions are understandable however, for the use of election observer evaluations as data about election quality these findings are problematic, as both exclusion of opposition and violence can severely undermine the de facto competitiveness and inclusiveness of an election. As regards the relative weighting of indicators, concerns about electoral conduct before the elections and allegations of media bias appear to be most strongly associated with observers' allegations of fraud.

Turning to the data gathered by Birch (2008, 2011) based on international observation mission reports, there is a rather high proportion of missing data, implying that regressing all specific irregularities variables on the dependent variable of overall election quality would lead to a loss of almost 2/3ds of the cases. Hence, I ran bi-variate correlations and regressions analyzing the effect of specific indicators on the overall indicator one by one. Though less robust, the results seem to indicate that all specific irregularities are significantly related to the overall election quality score, indicating that the overall score is quite comprehensive. Nevertheless, in terms of weighting clearly some irregularities seem to affect the overall score more strongly than others. Like with the Kelley

data, polling arrangements (pre-election administrative capacity) seem to be less strongly related to overall quality scores than factors like voter and candidate intimidation, vote counting and tabulation, etc. What stands out is the low coefficient for vote buying, which Birch explains by the probable under-reporting of vote buying by citizens and its subsequent under-estimation by election observers.

Table 2.13. Overall judgments of election quality and specific irregularities

Specific irregularities	Bivariate correlations overall election quality	Bivariate regressions overall election quality	N
<i>Manipulation of legislative framework:</i>			
• Legal framework	0.5756*** (123)	0.824***	123
<i>Manipulation of vote choice:</i>			
• media coverage	0.6972*** (134)	0.661***	134
• misuse of resources	0.5572*** (109)	0.640***	109
• vote buying	0.2514*** (145)	0.333**	145
• voter intimidation/obstruction	0.6812*** (145)	0.676***	145
• candidate intimidation/ obstruction	0.7275*** (142)	0.664***	142
<i>Manipulation of electoral administration:</i>			
• impartiality electoral authorities	0.7951*** (149)	0.738***	149
• contestation (ballot paper access)	0.6203*** (134)	0.655***	134
• voter registration	0.6120*** (131)	0.735***	131
• polling arrangements	0.5452*** (128)	0.756***	128
• voting process	0.7454*** (147)	0.942***	147
• counting, tabulation & reporting results	0.7653*** (146)	0.675***	146
• dispute adjudication	0.7430*** (106)	0.737***	106
• observer access	0.6462*** (149)	0.775***	149

Pearsson correlations and OLS regression respectively (r-squared values between 0.3 and 0.6 respectively, except vote-buying 0.1). P-values: + 0.1, * 0.05, ** 0.01, *** 0.001. Data from Index of Electoral Malpractice, Birch 2008 and 2011 forthcoming.

Finally, to evaluate the dimensionality of the data, I carried out factor analyses on those data that had specific irregularities variables. Using the Kelley and Kiril (2010) data, a factor analysis including the 7 specific variables results in a single factor explaining 92% of variation, and a model that is highly significant ($p < 0.000$). However, factor loadings show that violence and administrative capacity are relatively less important for the overall scores than the legal framework and the cheating variables, so even if measuring a single latent variable, the weighting of indicators is different. As for the Hyde and Marinov (2010) data, since variables were dichotomous, Mokken scaling was used instead of factor analysis. An analysis including the 5 specific variables results in a single factor of 4 variables with Loevinger coefficients between 0.40 and 0.64, well above the rule-of-thumb value of 0.30 (Van Schuur 2003)¹³⁸. One variable was excluded from the scale, election violence. Finally, for the Birch (2011) data, a factor analysis of the reduced sample with all specific irregularities also results in a single factor solution, that explains 76% of variance, and a model that is highly significant ($p < 0.000$). These analyses demonstrate that, even if election quality is a complex concept with many specific aspects, all these specific irregularities tap into a single underlying latent concept.

Overall judgments of election quality appeared to be quite comprehensive, as most specific irregularities were significantly related to overall scores; and specific irregularities seem to be measuring a single underlying latent concept. However, some irregularities are weighted more heavily than others, implying that the legal framework, pre-election cheating (including such aspects as misuse of state resources, campaign and media restrictions and candidate intimidation) and election-day cheating (such as ballot box stuffing and vote counting) might more strongly influence overall election quality scores than administrative

¹³⁸ Note that N for these analyses is much higher (759), because data for the specific irregularities is more widely available in the NELDA dataset than data on the evaluations of Western international monitors. The model was significant at each iteration with p-values < 0.01 .

irregularities in for example voter registration and polling procedures. Also, forms of electoral manipulation like vote buying and electoral violence stand out as quite different from other irregularities, and seem to be less relevant to observers' judgments of overall election quality. Moreover, it seems that when the overall judgment of election quality is more open to political influence, like the overall statements on the election by the US State Department, or allegations of fraud by Western international observers, they provide a less comprehensive reflection of the actual irregularities that happened.

Criterion validity

As a final test of the resulting measurement of election quality, I tested whether it co-varies with broader democracy indices. For this purpose I used a “categorical” and “continuous” measure of democracy, i.e. the Freedom House classification of not free, partly free and free regimes and the Polity IV autocracy/democracy scores. Correlations between these measures are indeed quite high and significant, for Freedom House varying between 0.67*** and 0.71*** and for Polity between 0.68*** and 0.70***. Tables 2.15 and 2.16 below show this relationship between election quality and democracy indices visually.

Table 2.14 – Mean election quality by Freedom House score and region 1974-2009

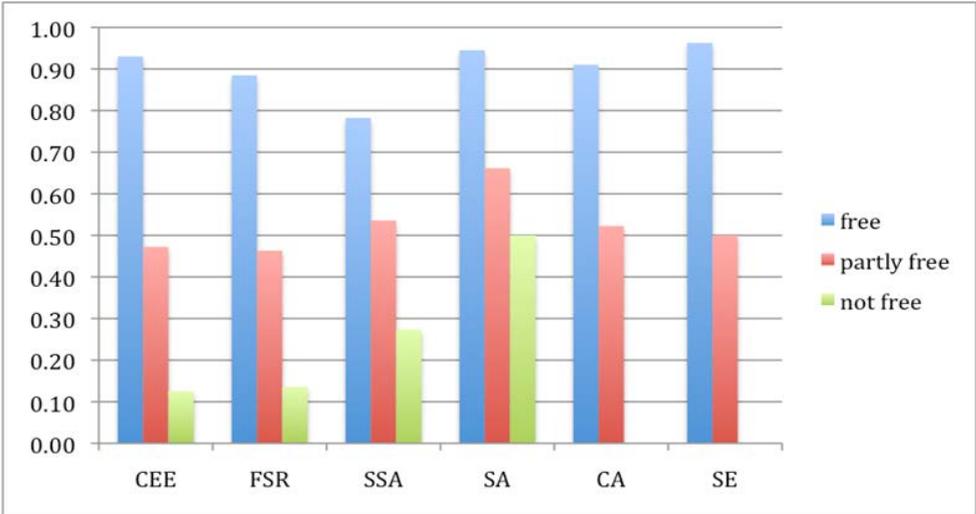
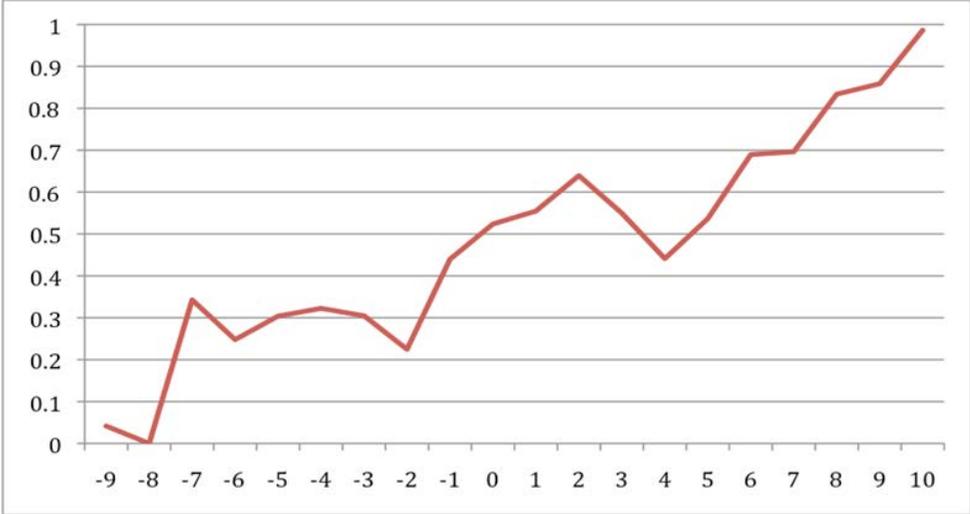


Table 2.15– Mean election quality by Polity IV scores 1974-2009



Appendix D – Notes on data & analyses chapter 4

Variables	Data source	Measurement	N	Min	Max	Missing	
Dependent variable							
Election quality	Anglin (1998), Van de Walle (2003), Lindberg (2006), Birch (2008/2011), Hartlyn et al. (2008), Donno (2009), Kelley and Kiril (2010) and Author's data	Qualitative classification of election quality: how frequent/widespread were irregularities during the electoral process? (0-1) (average)	885	0	1	1 missing election	
Independent variables							
Institutional factors							
	Presidentialism	Author's data	Presidential system yes/no: dichotomous	886	0	1	None
	Federalism	QoG dataset (DPI, Henisz 2002, Persson and Tabellini 2003, Regan and Clark 2010, Norris 2009)	Federal system yes/no: dichotomous (average)	882	0	1	Montenegro & Serbia
	Electoral system majoritarian ^a	Author's data (leg)	Proportional, mixed or majoritarian: categorical	532	1	3	None
		QoG dataset (ex) (Golder 2005, Regan and Clark 2010, IFES)	Plurality or absolute majority: dichotomous	354	1	2	None
Strength opposition (lagged previous election)		Author's data	Vote share largest opposition candidate in previous elections (ex) (%)	269	.7	49.6	First elections (N = 180, 83 1 st executive elections & 97 1 st legislative elections) plus 5 missing data.
Fragmentation opposition (lagged previous election)		Author's data	Seat share largest opposition party in previous elections (leg) (%)	432	0	66.67	First elections
		Author's data	Number of opposition candidates (ex) or parties (leg) in previous elections that won > 1% of the vote	706	0	28	First elections
Strength incumbent (lagged previous election)		Author's data	Vote share incumbent candidate in previous elections (ex) (%)	270	21.8	97.6	First elections plus 4 missing data.
		Author's data	Seat share incumbent party in previous elections (leg) (%)	432	5.5	100	

Variables	Data source	Measurement	N	Min	Max	Missing
Independence judiciary (lagged 2 years before election)	QoG dataset (Henisz 2002, Cingranelli and Richards 2010, Kaufmann et al 2009, Freedom House)	Judiciary independent yes/no: dichotomous (average)	857	0	1	Belize, Sao Tome & Principe, Seychelles, Suriname
Independence EMB	IDEA (2006)	EMB governmental, mixed or independent: categorical (country constant)	886	0	1	-
Independence media (lagged 2 years before election)	QoG dataset (Cingranelli and Richards 2010, Freedom House)	Complete censorship, some censorship, or free media: categorical (average)	886	0	1	-
Economic factors						
Economic development (lagged 2 years before election)	QoG dataset (World Development Indicators, United Nations Statistics Divisions 2009, Penn World Tables 2009)	GDP per capita: continuous (average)	877	118.03	26193.90	Serbia & Montenegro plus 1 missing data.
Economic growth (average 2 years before election)	QoG dataset (World Development Indicators, United Nations Statistics Divisions 2009, Penn World Tables 2009)	Growth GDP in 2 years before elections (average) (%)	877	-22.50	52.78	Serbia & Montenegro plus 1 missing data.
Economic inequality	QoG dataset (Galbraith 2009, Deininger & Squire 1996, United Nations University 2008, World Development Indicators)	GINI-coefficient of income inequality (average, country constant)	873	0	64.34	Equatorial Guinea, Rwanda, Sudan
International trade (lagged 2 years before election)	World Development Indicators	Merchandise imports from high income economies as % of total merchandise	886	0	96.01	-
		Merchandise exports from high income economies as % of total merchandise	886	0	99.75	
		Merchandise trade as % of GDP	886	0	986.65	
Resource curse (lagged 2 years before election)	QoG dataset (World Development Indicators)	Fuel, ores and metals exports as % of total merchandise exports	886	0	99.70	-
Socio-cultural factors						
Ethno-linguistic fractionalization	QoG dataset (Alesina et al 2003, Fearon 2003, Roeder 2001)	Index of ethno-linguistic fractionalization (0-1) (average, country constant)	880	0.02	0.88	Serbia, Montenegro

Variables	Data source	Measurement	N	Min	Max	Missing
Religious fractionalization	QoG dataset (Alesina et al 2003, La Porta et al 1999)	Index of religious fractionalization (0-1) (country-constant)	872	0.01	0.86	Serbia and Montenegro, Serbia, Montenegro
		Catholics and Muslims as percentage of population in 1980 (country-constant)	880 880	0 0	96.9 99.7	
Literacy/education (lagged 2 years before election)	QoG dataset (Gakidou et al. 2010)	Average years of education of men and women aged 25 and older (average)	877	0.25	13.2	Serbia and Montenegro plus 1 missing data.
Active civil society: domestic election monitors	Bjornlund (2004)	Domestic observers present: yes/no	886	0	1	-
Active civil society: electoral 'revolution'	Author's data	Electoral revolution: yes/no (coded 1 in year of revolution and all years after)	886	0	1	-
International factors						
Linkage (economic, social, political) (lagged 2 years before election)	QoG dataset (KOF Index of Globalization, Dreher et al 2008)	Average economic, political and social globalization (average) (0-100)	873	11.10	85.30	Serbia and Montenegro, Montenegro, plus 2 missing data.
Aid received (t-2)	QoG dataset (World Development Indicators, United Nations Statistics Divisions 2009)	Net development assistance and aid as proportion of GDP	878	0	141.77	Serbia and Montenegro
International election monitors	Bjornlund (2004)	International observers present: yes/no	886	0	1	-
Election assistance	Bjornlund (2004)	Country received election assistance: yes/no	886	0	1	-
Peacekeeping operations	UN	Peacekeeping mission present: yes/no	886	0	1	-
Historical factors						
Previous democratic experience	Author's data	Previous experience with 'potentially democratic' elections: yes/no	886	0	1	-
Previous autocratic experience	QoG dataset (Wright 2008)	Autocratic regime type military, single-party or personalist in year before first 'potentially democratic election' (country constant)	886	0	3	-

Variables		Data source	Measurement	N	Min	Max	Missing
	Colonial legacy	QoG dataset (Teorell and Hadenius 2005, Acemoglu et al 2001)	Former colonial ruler of the country: categorical (country constant)	886	0	8	-
	Civil war (average 2 years before election)	Author's data & QoG (UCDP/PRIO Armed conflict database)	Civil war in 2 years before election: yes/no	877	0	1	9 missing data.
	Coup d'etat (average 2 years before election)	Author's data	Coup d'etat in 2 years before election: yes/no	886	0	1	-
	Year founding election	Author's data	Year in which first 'potentially democratic' election after 1974 were held (country constant)	886	1974	2006	-
	Quality founding election	Author's data	Quality of first 'potentially democratic' elections (country constant)	758	0	1	All 128 first elections ^b
Control variables							
	Election type	Author's data	Legislative or executive election? Concurrent elections: yes/no	886	0	1	
			Constituent assembly election: yes/no	886	0	1	
	Election sequence number	Author's data	Sequence number of election starting from first 'potentially democratic election'	886	1	20	-
	Reliability election quality score	Author's data	Author's score based on ample, little or conflicting information	886	0	2	-
	Number of election quality scores received	Author's calculation	Number of times election received election quality score from the 7 datasets plus the author's coding	886	0	5	-
	Sample check	Author's calculation	Smaller sample of elections that were coded more than once, or coded by the author on the basis of reliable information (excluding scores based on little or conflicting information)	613	1	1	-
	Region	QoG dataset (Teorell and Hadenius 2005)	Region (based geographic proximity and area specialists)	886	1	6	-

a. In models including both legislative and executive elections, legislative elections are scored 0 for the presidential electoral system variable and vice versa. In the separate models for legislative and executive elections, only the appropriate electoral system variable is included in the model. Missing data for the presidential electoral system were updated with IFES data.

b. There are 97 countries in our sample and hence one would expect to lose only 97 cases. However, since some first elections were concurrent elections, these legislative and executive elections both received an election sequence number of "1", and hence in total 128 first elections are excluded from the sample when testing for the effect of election quality in the first election (this is necessary to avoid endogeneity).

Missing data were treated as follows: For each variable that had missing data, I created a dummy indicating if data were missing or not. Subsequently, if data were missing for, for example, the last year in the dataset, or the first year a country was included in our dataset (which was the case for most of the missing data), I updated these missing data with the value of the closest year within that country. For example, if the export and import data for 2009 were not available, I took the value of exports and imports in 2008 in that country as the value for 2009. I updated missing data in this way for the following variables: presidential electoral system, independence media, independence judiciary, international trade (export & imports), resource curse, globalization, aid dependency, previous autocratic regime type and coup d'état. Data missing for an entire country, i.e. over all elections included in our sample within that country, were left missing, except for international trade, resources, aid and previous autocratic regimes where missing countries were coded as 0. We ran full analyses always including a dummy indicating originally missing values for variables, if these were *not* significant I consider the overall results unproblematic. For some variables, I took the average value of various available variables if these separate variables were measured in the same way and highly correlated. These variables are: federalism, independence judiciary, GDP per capita, gini coefficient of income inequality, ethno-linguistic fractionalization and globalization.

Appendix E – Notes on data & analyses chapter 5

Variables	Data source	Measurement	N	Min	Max	Missing	
Dependent variable							
Incumbent turnover	Author's data	Ordinal classification of wholesale turnover, partial turnover (i.e. other candidate but same political party in presidential elections; part of governing coalition incumbent parties in parliamentary elections), and no alternation. Recoded to dichotomous: turnover yes/no. (partial turnover coded as no turnover).	885	0	1	1 missing election	
Independent variables							
Government performance							
Economic growth (average in 2 years before elections)	QoG dataset (United Nations Statistics Divisions 2009)	Growth GDP per capita in 2 years before elections (average) (%)	872	-22.6	52.8	Serbia & Montenegro plus 5 elections missing	
Civil liberties (average of 3 ^d and 2 nd year before elections)	QoG dataset (Freedom House)	Freedoms of expression and belief, associational and organizational rights, rule of law, and personal autonomy without interference from the state. Scored from 1 (most free) to 7 (least free). Scores reversed to range from low respect to high respect for civil liberties.	865	1	7	20 elections missing	
Election quality	Election quality	Anglin (1998), Van de Walle (2003), Lindberg (2006), Birch (2008/2011), Hartlyn et al. (2008), Donno (2009), Kelley and Kiril (2010) and Author's data	Qualitative classification of election quality: how frequent/widespread were irregularities during the electoral process? Recoded original continuous variable to a dummy separating flawed elections (election quality 0-0.66) from clean elections (0.66-1).	885	0	1	1 missing election
Party system	Structure party system	Author's data	Number of opposition candidates (ex) or parties (leg) that won > 1% of the vote in the elections. Recoded to: (0) dominant party system (0-1 opposition candidates/parties); (1) two-party system (2-3 opposition candidates/parties); (2) multi-party system (4-15 opposition candidates/parties); (3) fragmented party system (16-28 opposition candidates/parties).	885	0	3	1 missing election
Electoral system	Electoral system legislative elections ^a	Author's data	Proportional, mixed or majoritarian: categorical. Recoded to dichotomous variable (proportional & mixed (0) vs majoritarian (1))	532	0	1	None

Variables	Data source	Measurement	N	Min	Max	Missing
Electoral system executive elections ^a	QoG dataset (Golder 2005, Regan and Clark 2010, missing updated by author using data from IFES)	Plurality or absolute majority: dichotomous	354	0	1	None
Division of power						
Parliamentary system	Author's data	Parliamentary system yes/no: dichotomous	886	0	1	None
Bicameralism	QoG dataset (Johnson & Wallack 2006, missing updated by author using data from IPU)	Bicameral legislative yes/no: dichotomous. (country constant)	886	0	1	None
Federalism	QoG dataset (DPI, Henisz 2002, Persson and Tabellini 2003, Regan and Clark 2010, Norris 2009)	Federal system yes/no: dichotomous (average, recoded to 0-1)	881	0	1	Montenegro & Serbia, 4 elections
Coalition government (lagged)	Author's data	Coalition government formed after the previous elections yes/no: dichotomous (presidential elections coded as 0). Lagged value.	787	0	1	98 elections missing (1 st legislative elections)
Number of veto-players	QoG dataset (DPI, (Beck et al 2000 & 2001, Keefer 2009)	Number of veto-players in political system ^b	862	1	8	Montenegro, Serbia, Sao Tome and Principe and Seychelles missing
State capacity						
Economic development	QoG dataset (World Development Indicators, United Nations Statistics Divisions 2009), missing updated by author with data from Worldbank WDI website.	GDP per capita in year of election: continuous (average). Recoded to dichotomous separating "bottom billion" (Collier 2006) poor countries from richer countries, i.e. < 2000 USD per capita = 0, and > 2000 USD per capita = 1	876	0	1	10 elections missing
Aid received (t-2)	QoG dataset (World Development Indicators, United Nations Statistics Divisions 2009)	Net development assistance and aid in year of election as proportion of GDP (%) in 2 years before elections.	875	0	78.7	11 elections missing
Civil war (t-2)	Author's data & QoG (UCDP/PRIO Armed conflict database)	Civil war in 2 years before election: yes/no	877	0	1	9 missing data.

Variables	Data source	Measurement	N	Min	Max	Missing
Bureaucratic effectiveness	QoG dataset (Kaufmann et al 2009)	Index of "government effectiveness: combines data on quality of public service provision, quality of the bureaucracy, competence of civil servants, independence of civil service from political pressures, and credibility of government's commitment to policies. (QoG Codebook v6 April 2011, p. 69). Scores normally distributed with mean of 0 and standard deviation of 1, i.e. most scores between -2.5 and +2.5, higher indicating more effective government. (country constant)	886	-1.8	1.4	None
Availability of information						
Independence media	QoG dataset (Cingranelli and Richards 2010, Freedom House)	Complete censorship, some censorship, or free media: categorical (average). Measured in 2 years before elections.	886	0	1	None
Previous democratic experience	Author's data	Previous experience with 'potentially democratic' elections: yes/no	886	0	1	-
Concurrent elections	Author's data	Concurrent elections: yes/no	886	0	1	None
Control variables						
Election type	Author's data	Legislative or executive election?	886	0	1	None
		Concurrent elections: yes/no	886	0	1	None
		Constituent assembly election: yes/no	886	0	1	None
Election sequence number	Author's data	Sequence number of election starting from first 'potentially democratic election'	886	1	20	None
Region	QoG dataset (Teorell and Hadenius 2005)	Region (based geographic proximity and area specialists designation)	886	1	6	None

a. In models including both legislative and executive elections, legislative elections are scored 0 for the presidential electoral system variable and vice versa. In the separate models for legislative and executive elections, only the appropriate electoral system variable is included in the model. Missing data for the presidential electoral system were updated with IFES data.

Appendix F – Notes on data & analyses chapter 6

Variables	Data source	Measurement	N	Min	Max	Missing	
Dependent variables							
Economic growth (proportion of change in GDP per capita, average 2 nd and 3 ^d years after elections)	QoG dataset (World Development Indicators), updated with data from: http://dataworldbank.org/data-catalog	Annual percentage growth rate of GDP per capita based on constant local currency.	806	-20.1	36.7	79 missing elections (72 last elections, 7 missing)	
Respect for civil liberties (average 2 nd and 3 ^d years after elections)	QoG dataset (Freedom House)	Freedoms of expression and belief, associational and organizational rights, rule of law, and personal autonomy without interference from the state. Scored from 1 (most free) to 7 (least free). Scores reversed to range from low respect to high respect for civil liberties.	811	1	7	74 missing elections (72 last elections, 2 missing)	
Respect for human rights (average 2 nd and 3 ^d years after elections)	QoG dataset (Political terror scale - US State Department, Gibney, Cornett and Wood 2010; Gibney and Dalton 1996)	Human rights score from level 1 to 5, ranging from secure rule of law to widespread human rights violations. Scores reversed to range from low respect to high respect for human rights.	770	1	5	115 missing elections (109 last elections, 6 missing)	
Independent variables							
Election quality	Election quality	Anglin (1998), Van de Walle (2003), Lindberg (2006), Birch (2008/2011), Hartlyn et al. (2008), Donno (2009), Kelley and Kiril (2010) and Author's data	Qualitative classification of election quality: how frequent/widespread were irregularities during the electoral process? Recoded original continuous variable to a dummy separating flawed elections (election quality 0-0.66) from clean elections (0.66-1).	885	0	1	1 missing election
Party system	Margin of victory	Author's data	Strength incumbent party/candidate minus strength largest opposition party/candidate. Strength measured as last round vote share for presidential candidates and seat share for legislative parties.	860	0	97	5 missing, plus 20 outliers excluded (< 0 and 100)
	Fragmentation opposition / Structure party system	Author's data	Number of opposition candidates (ex) or parties (leg) that won > 1% of the vote in the elections. Recoded to: (0) dominant party system (0-1 opposition candidates/parties); (1) two-party system (2-3 opposition candidates/parties); (2) multi-party system (4-15 opposition candidates/parties); (3) fragmented party system (16-28 opposition candidates/parties).	885	0	3	1 missing election
Electoral system	Electoral system legislative elections ^a	Author's data	Proportional, mixed or majoritarian: categorical. Recoded to dichotomous variable (proportional & mixed (0) vs majoritarian (1))	532	0	1	None

Variables	Data source	Measurement	N	Min	Max	Missing
Electoral system executive elections ^a	QoG dataset (Golder 2005, Regan and Clark 2010, missing updated by author using data from IFES)	Plurality or absolute majority: dichotomous	354	0	1	None
Division of power						
Parliamentary system	Author's data	Parliamentary system yes/no: dichotomous	886	0	1	None
Bicameralism	QoG dataset (Johnson & Wallack 2006, missing updated by author using data from IPU)	Bicameral legislative yes/no: dichotomous. (country constant)	886	0	1	None
Federalism	QoG dataset (DPI, Henisz 2002, Persson and Tabellini 2003, Regan and Clark 2010, Norris 2009)	Federal system yes/no: dichotomous (average, recoded to 0-1)	881	0	1	Montenegro & Serbia
Coalition government	Author's data	Coalition government formed after the elections yes/no: dichotomous (presidential elections coded as 0)	883	0	1	3 elections missing
Number of veto-players	QoG dataset (DPI, (Beck et al 2000 & 2001, Keefer 2009)	Number of veto-players in political system ^b	862	1	8	Montenegro, Serbia, Sao Tome and Principe and Seychelles miss
State capacity						
Economic development	QoG dataset (World Development Indicators, United Nations Statistics Divisions 2009), missing updated with data from Worldbank WDI website.	GDP per capita in year of election: continuous (average)	879	194	30657	7 elections missing
Aid received	QoG dataset (World Development Indicators, United Nations Statistics Divisions 2009)	Net development assistance and aid in year of election as proportion of GDP (%)	873	0	84.5	13 elections missing
Civil war	Author's data & QoG (UCDP/PRIO Armed conflict)	Civil war in year of election: yes/no	886	0	1	None
Bureaucratic effectiveness	QoG dataset (Kaufmann et al 2009)	Index of "government effectiveness: combines data on quality of public service provision, quality of the bureaucracy, competence of civil servants, independence of civil service from political pressures, and credibility of government's commitment to policies. (QoG Codebook v6 April 2011, p. 69). Scores normally distributed with mean of 0 and standard deviation of 1, higher indicating more effective government. (constant)	886	-1.8	1.4	None

Variables	Data source	Measurement	N	Min	Max	Missing
Control variables						
Election type	Author's data	Legislative or executive election?	886	0	1	None
		Concurrent elections: yes/no	886	0	1	None
		Constituent assembly election: yes/no	886	0	1	None
Election sequence number	Author's data	Sequence number of election starting from first 'potentially democratic election'	886	1	20	None
		Region	QoG dataset (Teorell and Hadenius 2005)	Region (based geographic proximity and area specialists designation)	886	1

a. In models including both legislative and executive elections, legislative elections are scored 0 for the presidential electoral system variable and vice versa. In the separate models for legislative and executive elections, only the appropriate electoral system variable is included in the model. Missing data for the presidential electoral system were updated with IFES data.

b. DPI measures the number of veto-players in the political system as follows: "Equals one if the Legislative Index of Political Competitiveness (dpi_lipc) or the Executive Index of Political Competitiveness (dpi_eipc) is less than 6. In countries where dpi_lipc and dpi_eipc are greater than or equal to 6, dpi_checks is incremented by one if there is a chief executive, by a further one if the chief executive is competitively elected (dpi_eipc greater than six), and by a further one if the opposition controls the legislature. In presidential systems, dpi_checks is incremented by one for each chamber of the legislature (unless the president's party has a majority in the lower house and a closed-list system is in effect), and by one for each party coded as allied with the president's party and which has an ideological (left-right) orientation closer to that of the main opposition party than to that of the president's party. In parliamentary systems dpi_checks is incremented by one for every party in the government coalition as long as the parties are needed to maintain a majority, and by one for every party in the government coalition that has a position on economic issues closer to the largest opposition party than to the party of the executive. (The prime minister's party is not counted as a check if there is a closed rule in place.)" (QoG Codebook v6 April 2011, p. 86).

Missing data were treated as follows: For each variable that had missing data, I created a dummy indicating if data were missing or not. Subsequently, if data were missing for, for example, the last year in the dataset, or the first year a country was included in our dataset, I updated these missing data with the value of the closest year within that country. For example, if the proportion of aid received in 2009 was not available, I took the value of aid received in 2008 in that country as the value for 2009. I updated missing data in this way for the following variables: number of veto-players and proportion of aid received. Data missing for an entire country, i.e. over all elections included in our sample within that country, were left missing, except for aid where the 3 Southern European countries were missing and all coded as 0. We ran full analyses always including a dummy indicating originally missing values for variables, if these

were *not* significant I consider the overall results unproblematic. For two variables, missing data were so common that I decided to use the country-average instead. This was the case for government effectiveness and bicameralism.

Table 6.4. below shows the results when using respect for human rights after the elections as government performance indicator (as coded by Gibney, Cornett and Wood (2010) on the basis of US State Department human rights reports), rather than the Freedom House civil liberties data. The results are essentially similar as for civil liberties: under a wide range of different model specifications and including control variables, clean elections always appear to have a significant and positive effect on the level of respect for human rights in the years after elections. This holds also when we use a variable measuring respect for human rights based on Amnesty International reports (also from Gibney, Cornett and Wood (2010)). Model 1 shows the effect of election quality on respect for human rights in the years after the elections: the effect is positive and highly significant as expected. Clean elections are associated with increased respect for human rights after the elections, and vice versa. Moreover, election quality appears to explain a substantial amount of variance, though mostly at the country level. Turning to party system characteristics, as in the previous models, margin of victory appeared to be linearly related to the dependent variable rather than curvi-linear as hypothesized. The effect is rather small however and does not reach one-tailed significance in model 2 (it does become significant once election quality is corrected for however, as shown in model 6). The number of opposition parties and candidates does seem to affect government performance as expected: respect for human rights is somewhat higher in two-party and multi-party systems, however also these effects do not reach conventional levels of statistical significance. Model 3 shows that the effects of legislative and presidential electoral systems are both in the expected direction, however neither is significant.

Table 6.4. Explaining respect for human rights in 2 & 3 years after elections (average)

	Model 1 Election quality	Model 2 Party system	Model 3 Electoral system	Model 4 Division power	Model 5 State capacity	Model 6 Full model
Election quality (0-1)	0.340*** (0.057)					0.326*** (0.057)
Margin of victory		0.001 (0.001)				0.002* (0.001)
Two-party systems ^a		0.140 (0.103)				0.190* (0.097)
Multi-party systems		0.067 (0.099)				0.078 (0.093)
Fragmented party systems		0.001 (0.148)				0.002 (0.139)
Electoral system maj leg			-0.007 (0.083)			
Electoral system abs maj ex			-0.086 (0.057)			
Parliamentary system				0.560*** (0.155)		0.269+ (0.141)
Bicameralism				-0.142 (0.177)		
Federalism				-0.173+ (0.103)		-0.126 (0.095)
Coalition government				0.069 (0.066)		
Number of veto players				0.023 (0.020)		
Economic development (gdp/cap)					-0.00003* (0.00001)	-0.00003* (0.00001)
Aid received as % of GNI					0.009*** (0.003)	0.009*** (0.003)
Civil war					-0.675*** (0.101)	-0.667*** (0.098)
Bureaucratic effectiveness					0.926*** (0.109)	0.756*** (0.113)
Constant	3.259*** (0.089)	3.311*** (0.138)	3.443*** (0.097)	3.330*** (0.123)	3.836*** (0.106)	3.430*** (0.152)
N level 1	726	726	726	726	726	726
N level 2	93	93	93	93	93	93
Variance explained (level 2 – country level)	11%	0%	1%	10%	35%	37%
Variance explained (level 1 – election level)	1%	0%	0%	0%	2%	3%
Total variance explained	12%	0%	1%	10%	37%	40%

Standard errors in parentheses. P-values: ^ 0.1, * 0.05, ** 0.01, *** 0.001. a. Reference category is dominant party systems.

Turning to model 4, parliamentary systems appear to be associated with higher respect for human rights, as hypothesized, but the other institutional variables do not appear to be significantly related to government performance as measured here. Model 5 shows the results for state capacity variables that explain an important amount of variance in the level of respect for human rights after elections. The effects of civil war and bureaucratic effectiveness are in line with our expectations: civil war is associated with lower respect for human rights while bureaucratic effectiveness has a positive effect. The effect of aid is contrary to expectations: countries that receive higher proportions of aid in the year of elections tend to have higher levels of respect for human rights in the years after.

The level of economic development is also contrary to expectations: it seems that higher levels of economic development are associated with lower levels of respect for human rights. Closer inspection of the data reveals that this result is probably driven by the presence of a substantial number of poor countries in our dataset that nevertheless have relatively high levels of respect for human rights. In fact, the effect of economic development is not robust to the inclusion of control variables and stricter model specifications, as models 7 – 12 will show. Finally, model 6 shows the result of a full model including all independent variables that were significant or close to one-tailed significance in the previous analyses. The effect of election quality continues to be positive and significant, and the effects of party system variables are now also significant. The effect of parliamentarism is diminished in the full model and the effects of the state capacity variables remain rather stable. How robust are these results to the inclusion of controls and robustness checks?

The table on the next page shows the results of these checks. Model 7 reports the results if controls are included for the type of elections, election sequence number and region. Model 8 reports the same model using a smaller sample including only those countries that experienced 5 or more elections. Model 9 reports the results when estimating the model using fixed effects, i.e. including country dummies to correct for heteroskedasticity. Such a model only estimates

over-time variance, and hence non-changing variables like bureaucratic effectiveness drop out of the model¹³⁹. Continuing, I tested for serial autocorrelation, which resulted to be a problem in these data as well, and hence I corrected the model by including a lagged dependent variable. The results are reported in model 10. Finally, I estimated the full model using time-series-cross-sectional methods on separate samples of legislative and executive elections, reported in model 11 and 12. The separate models for legislative and executive elections were re-run with (a) using a sample of countries with at least 5 elections, (b) including the lagged dependent variable –to correct for serial autocorrelation, and (c) using fixed effects –to correct for heteroskedasticity- and in all models election quality remained significant and positive. Clearly, even if clean elections do not enhance post-election economic growth, they do have a real and substantial positive effect on respect for human rights in the years after elections.

¹³⁹ Heteroskedasticity did not appear to be very problematic in these models, as only 3 countries turned out to have very high error variance, i.e. Congo (Kinshasa), Uganda and Guinea Bissau. Excluding these 3 countries from the analyses did not change the results. As an additional test of the robustness of results to potential heteroskedasticity however, I report the fixed effects model.

Table 6.4. (continued) Explaining respect for human rights in 2 & 3 years after elections (average)

	Model 7 Pars model w controls	Model 8 Pars model T > 5	Model 9 Pars model fixed effects	Model 10 Pars model lagged DV	Model 11 Pars model legislative	Model 12 Pars model executive
Election quality (0-1)	0.339*** (0.056)	0.364*** (0.056)	0.318*** (0.057)	0.216*** (0.054)	0.348*** (0.076)	0.433*** (0.097)
Margin of victory	0.002+ (0.001)	0.002+ (0.001)	0.002+ (0.001)	0.003** (0.001)	0.002+ (0.001)	0.003 (0.002)
Two-party systems ^a	0.202* (0.096)	0.207* (0.098)	0.236* (0.097)	0.208* (0.089)	0.311+ (0.159)	0.090 (0.155)
Multi-party systems	0.107 (0.093)	0.102 (0.095)	0.163+ (0.094)	0.122 (0.085)	0.126 (0.153)	-0.046 (0.147)
Fragmented party systems	0.031 (0.140)	0.036 (0.143)	0.111 (0.142)	0.162 (0.127)	-0.072 (0.193)	0.452 (0.422)
Parliamentary system	0.204 (0.142)	0.333* (0.147)	0.387* (0.193)	0.056 (0.115)	0.156 (0.148)	0.491 (0.423)
Economic development (gdp/cap)	9.86 ^{e-07} (0.00001)	-4.18 ^{e-07} (0.00001)	9.29 ^{e-06} (0.00001)	-4.45 ^{e-06} (0.00001)	-4.76 ^{e-05**} (0.00001)	-6.44 ^{e-06} (0.00002)
Aid received as % of GNI	0.008** (0.003)	0.007* (0.003)	0.006* (0.003)	0.006* (0.003)	0.008* (0.003)	0.015** (0.005)
Civil war	-0.705*** (0.097)	-0.770*** (0.103)	-0.677*** (0.102)	-0.338*** (0.095)	-0.628*** (0.131)	-0.860*** (0.157)
Bureaucratic effectiveness	0.554*** (0.138)	0.534*** (0.125)		0.344*** (0.099)	0.650*** (0.133)	0.614*** (0.158)
Election type	-0.007 (0.044)	-0.011 (0.044)	-0.009 (0.044)	0.020 (0.040)		
Constituent assembly elections	-0.112 (0.137)	-0.163 (0.147)	-0.051 (0.137)	-0.028 (0.132)	-0.108 (0.143)	
Concurrent elections	-0.110 (0.071)	-0.121+ (0.072)	-0.051 (0.076)	-0.095 (0.064)	-0.149+ (0.088)	-0.139 (0.128)
Election sequence	-0.047*** (0.009)	-0.048*** (0.009)	-0.053*** (0.010)	-0.023** (0.009)		
Lagged dependent variable			0.401*** (0.030)			
Region - SSA ^b	-0.357 (0.235)	-0.148 (0.211)		-0.291+ (0.160)	-0.486* (0.218)	-0.301 (0.276)
Region - SA	-0.402 (0.254)	-0.394+ (0.216)		-0.280 (0.171)	-0.518* (0.233)	-0.454+ (0.274)
Constant	3.716*** (0.224)	3.701*** (0.209)	3.539*** (0.790)	2.160*** (0.210)	3.792*** (0.261)	3.602*** (0.305)
N level 1 / level 2	726 / 93	684 / 79	726 / 93	705 / 93	430 / 92	296 / 79
Variance explained (level 2 – country level)	36%	37%	-	57%	-	-
Variance explained (level 1 – election level)	4%	6%	-	7%	-	-
Total variance explained	40%	43%	72%	64%	47%	42%

a. Reference category is dominant party systems. b. Central and Eastern Europe is the reference category. Coefficients for FSR, CA and SE not reported since insignificant in all models.