1

Introduction: How Electoral Systems Matter – for Politics and for the Scientific Study Thereof

Who governs? Electoral systems matter in democracies because they affect the answer to this question Robert Dahl (1961) posed in a different context. In democracies, the answer might be "those who win elections." However, it is rarely so simple. We might rephrase the question as "who wins elections?" One might immediately respond with, of course, those who win the most votes. Yet again, it is not always so straightforward. In the US presidential election of 2016, Hilary Clinton obtained over 65.8 million votes, which exceeded those of Donald Trump by almost 2.9 million (and 2.1 percentage points of the national total). Yet it was Trump, not Clinton, who became President, due to the way the electoral system takes "who got votes" and turns it into "who governs."

The US in 2016 is one recent and notable example. A reversal of which candidate won the nationwide votes and which one became president also happened in the US in 2000.[1] Yet maybe these US cases are simply aberrations. Maybe in other democracies it is always simple to say that "who governs is whoever wins the most votes." Actually it is not so simple; it is often the case that the way the electoral system works makes a difference in how the votes get turned into the important positions of power that determine who governs. A couple of other examples will demonstrate the point.

New Zealand had two consecutive elections (1978 and 1981) in which the Labour Party won the most votes nationwide, yet the rival National Party formed the government. Why? Because the Nationals won more *seats* in the national parliament. This, of course, invites the question, why did the Nationals win more seats – a majority[2], in fact – when Labour won more votes? The answer lies in the way the *electoral system* worked. There were no controversies in either of these elections over whether the outcome was correctly or "legitimately" decided. Yet these elections helped begin a process

[1] The significance of the US election controversy of 2000 for understanding electoral systems is discussed further in Chapter 3.

[2] We use the term, majority, strictly in the American sense as "at least one more than half." If we mean the largest share, whether or not more than half, we will say "plurality."

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

that, by 1996, would see the country implement a completely different electoral system.

We could also think of the case of Denmark in 2015. In this election, the Social Democrats won the most votes *and* seats. Yet when a government was formed after the election, the Social Democrats were in the opposition. The party with the second most votes and seats, the Danish People's Party, also was not in the government. Instead, the third largest party, the Venstre, provided the country's prime minister, Lars Løkke Rasmussen. His government obtained the parliamentary backing of some other parties, including the Danish People's Party, but not the Social Democrats. This example illustrates that sometimes many parties obtain representation – the three parties just mentioned combined for only slightly more than two-thirds of the votes and a similar share of seats. Six other parties obtained representation. Thus "who governs" depended on bargaining between various parties, because the electoral system made it possible for many parties to receive seats as well as votes.

Each of these examples shows that it is not enough to say there are elections. We need to know something about how votes get converted into governing power. It is the electoral system that is a key element in this conversion process. If the country has a "parliamentary" form of government, like New Zealand and Denmark, the electoral system only turns votes into seats for parties in the parliament. If one party has a majority – as in the 1978 and 1981 New Zealand examples – it forms the government (regardless of whether the party also had earned the most votes). If there is a "multiparty" system, as in the Danish example, then determining who governs involves another step – coalition bargaining. This is a step we will not consider at length in this book; many other works cover it as a topic in its own right. Instead, this book focuses on how votes become seats, and how this process varies from country to country, depending on the *electoral system*.

By electoral *system* we mean the set of rules that specify how voters can express their preferences (ballot structure) and how the votes are translated into seats. The system must specify at least the number of areas where this translation takes place (*electoral districts*), the number of seats allocated in each of these areas (*district magnitude*), and the seat allocation formula. All this will be discussed in more detail later.

This book deals only with elections that offer some choice. It bypasses fake elections where a single candidate for a given post is given total or overwhelming governmental support, while other candidates are openly blocked or covertly undermined. It also largely overlooks pathologies of electoral practices such as malapportionment and gerrymander, except for pointing out which electoral systems are more conducive to such manipulation.

The physical conditions of elections matter, such as ease of registration of voters and candidates, location and opening times of polling stations, and the

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

timing of elections. It is presumed in this book that such conditions are satisfactory. Our only concern is to explain, in what are considered fair elections, how electoral systems affect the translation of votes into seats, how the results also affect the distribution of the votes in the next elections, and what it means for party systems. Moreover, the book largely limits itself to first or sole chambers of legislatures, although Chapters 11 and 12 will include analysis of presidential elections, and we will offer an occasional example from an elected second chamber (upper house).

This scope may look narrow, but translation of votes into seats by different electoral systems can lead to drastically different outcomes, both within a country (one election to another, or one elected body versus another) and across countries. For example, Green parties committed to more environmentally sound policies have emerged in many established democracies. The Green Party of Canada even received 6.8 percent of the votes in the national election for the House of Commons in 2008. Yet it won exactly zero seats. The Green Party of Germany in 1998 had obtained almost the same national vote share for the Bundestag, 6.7 percent, and not only won seats in parliament (forty-seven of the 669, or 7.0 percent) but also obtained seats in the cabinet including the Minister of Foreign Affairs. The difference in the outcomes, despite similar vote shares, is due to the electoral systems these countries use.

Thus, electoral systems affect party strengths in the representative assembly[3] and, if the political system is parliamentary, the resulting composition of the governing cabinet. They can encourage the rise of new parties, bringing in new blood but possibly leading to excessive fractionalization, or they can squeeze out all but two parties, bringing clarity of choice but possibly leading to eventual staleness. It is well worth discovering in quantitative detail how electoral systems and related institutions affect the translation of votes into seats. We now offer one more detailed example that introduces several themes of the book.

The Polish Election of 2015 and How the Rules Mattered

Poland has been a stable democracy since the 1989–1990 transition to democracy as the Soviet Communist bloc collapsed. Since that time and through 2015, Poland has had eight elections for Sejm and Senate, the two chambers of the national parliament, and six for its politically powerful presidency. Nonetheless, the Sejm election of 2015 stands out and led to

[3] Throughout this book, we will generally prefer the term, *assembly*, rather than legislature, parliament, congress, or other terms. We thus avoid any implicit commentary on the institution's precise role in a given democratic system and call attention instead to its essential feature as a plural body in which elected representatives assemble for their various tasks.

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

significant international concern. For the first time in the post-Communist era, a single party won a majority of seats in the Sejm, the first chamber (and the more powerful one). This party, the Law and Justice Party (known by its Polish initials, PiS), already held the presidency and thus was able to set the nation's policy agenda essentially unilaterally. This agenda was controversial, particularly a law passed near the end of 2015 that changed the procedures for the Polish Constitutional Court. This change drew an official rebuke from the European Commission as a "systemic risk to the rule of law in Poland."[4]

It is useful to situate the Polish events that troubled its neighbors into the context of political institutions, including the electoral system, which made PiS's Sejm majority possible. Poland uses one of the many examples of an electoral system that is typically called "proportional representation" (PR). Usually, under PR, no party wins a majority in the assembly, because the system is designed to make it feasible for many parties to win seats. It tends to support bargaining among parties after the election (as in the Danish case mentioned earlier). Indeed, this had been the case in Poland since the 1990s. So what changed in 2015 to allow a single-party majority?

First of all, the *timing* of the election was critical. Poland had just elected a new president in May 2015. It was a close contest. Like many countries that elect a politically powerful presidency by direct vote, Poland votes in "two rounds": in the event that no candidate has a majority in the first round, there is a runoff between the top two candidates two weeks[5] later. The winner, Andrzej Duda of PiS, won 51.6 percent to 48.5 percent, over a candidate backed by the then-governing party, Civic Platform (PO). Six months later, when the Sejm election came up, it was within the "honeymoon" of the newly elected PiS president. Presumably aided by this honeymoon period, PiS won 37.6% of the vote, which was more than Duda himself had received in the first round of his election (34.8%) and also was a 7.7 percentage-point increase in the party's vote over the previous (2011) Sejm election.

Ordinarily, the timing of assembly elections relative to presidential elections is not considered part of the "electoral system"; however, as we show in Chapter 12, such timing does indeed have systematic effects on the performance of parties. The Polish pattern in 2015, whereby the newly elected president's party enjoyed a surge in votes, is a common pattern (Shugart 1995). So, the timing of elections – an institutional feature of Polish democracy – may have helped PiS gain the most votes. How did the electoral system for Sejm turn those into a majority of seats?

[4] "Commission adopts Rule of Law Opinion on the situation in Poland," European Commission press release, June 1, 2016 (http://europa.eu/rapid/press-release_IP-16-2015_en.htm, accessed July 13th 2016).

[5] The time between rounds varies across countries; in Poland the rounds were May 10 and 24. We discuss some implications of two-round elections in Chapter 3.

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

TABLE 1.1 *Polish Sejm election result, October, 2015 (national figures)*

Party (and name or abbreviation in Polish)	Votes	Votes (%)	Seats	Seats (%)
Law and Justice (PiS)	5,711,687	37.58	235	51.09
Civic Platform (PO)	3,661,474	24.09	138	30.00
Kukiz'15 (K'15)	1,339,094	8.81	42	9.13
Modern (Nowoczesna)	1,155,370	7.6	28	6.09
United Left (ZL)	1,147,102	7.55	0	0.00
Polish People's Party (PSL)	779,875	5.13	16	3.48
KORWiN	722,999	4.76	0	0.00
Together (Razem)	550,349	3.62	0	0.00
German minority (MN)	27,530	0.18	1	0.22
others	105,191	0.69	0	0.00
Total	15,200,671		460	

The row for "others" includes no single party with more than 0.28 percent of the vote.

Table 1.1 shows the detailed results of the 2015 Sejm election. The first puzzle is the one already mentioned – the fact that the PiS won more than half the seats on only 37.6 percent of the vote. How could that be? Poland, after all, uses "proportional representation." Yet 51 percent of the seats on under 38 percent of the votes is not very "proportional." As is further shown in Table 1.1, part of the answer lies in the *thresholds*.[6] Two parties have over 3.5 percent of the votes apiece, yet no seats. This is because the electoral law required 5 percent of the nationwide votes to win seats, unless the party represents a national minority. The latter provision explains why a party for the German Minority has a seat on only 0.18 percent of the vote.[7] However, the parties called KORWiN and Together are not ethnic-minority parties. Thus the threshold excluded them from representation.

It is further visible from the table the United Left had 7.55 percent of the votes – easily clearing the 5 percent party threshold – yet no seats. How can that be? It is due to yet a further feature of Polish electoral law: if two or more parties jointly contest the election, the threshold is 8 percent, rather than 5 percent, for them. The provision is presumably intended to prevent parties from making "marriages of convenience" just to pass the threshold jointly. Yet in this case, the parties in question are a set of ideologically proximate parties that came close, but not close enough, to clearing the threshold.

[6] For details on Polish election laws, see Hardman (n.d.)

[7] Separate provisions for ethnic minorities will not be a theme of this book. See the excellent and detailed treatment by Lublin (2014).

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

We will discuss these features of the system in subsequent chapters, as we discuss electoral alliances in Chapters 6, 7, and 14 (and in passing in other chapters); we discuss thresholds in Chapter 15.

After the exclusion of parties that fell below the thresholds, PiS had around 45 percent of the total vote cast for parties that were eligible to win seats. That means its assembly majority still gives it a degree of over-representation. This is due in part to the electoral formula used – something called the D'Hondt divisor method, which will be explained in Chapter 2. A different "proportional" formula might have netted PiS less than half the seats, just as a lower threshold, or no separate threshold for alliances, might have meant more seats for smaller parties and hence fewer for PiS. An election not in the president's honeymoon, or a different outcome of the close presidential race itself, might have meant fewer votes for PiS in the first place.

While it may seem right now as if we are just making a "laundry list" of obscure provisions and Polish political idiosyncrasies, in fact, *all of these electoral rules mattered to the outcome*. With a different set of rules, then, there might not have been a PiS majority government, and the resulting international controversy that the government became embroiled in may never have occurred. In other words, electoral systems have consequences for how a country is governed, and by whom.

In Table 1.1 we saw the national outcome of Poland's 2015 Sejm election. However, Poland actually has forty-one electoral districts. With an assembly size of 460, that means each district elects *on average* just over eleven seats. As we explain further in Chapter 2, this means we can speak of Poland having an average *district magnitude* of eleven. Thus the electoral system is not really national in scope; indeed, few are. The results in Table 1.1 show the nationwide aggregation of votes and seats, but there are in fact forty-one different contests playing out, each one electing anywhere from seven to twenty members of the Sejm. Unlike many other books on electoral systems, this one will analyze not only nationwide aggregate patterns, but also district-level dynamics. In fact, Chapter 10 is devoted entirely to developing models of patterns in data disaggregated to the district level. We even go one step farther than this; in Chapters 13 and 14, we look at the *intraparty dimension* of representation, whereby individual candidates compete for votes against others of their party under certain electoral systems. Poland has one of these systems of intraparty competition, as we explain briefly here.

In **Table 1.2**, we see partial results from one of Poland's forty-one electoral districts in the 2015 election. The district is Konin, and it has a district magnitude of nine. In Poland, as in many other (but by no means all) electoral systems, the voter casts her ballot by placing a mark by the name of a single candidate. The table shows only the votes of the top fifteen candidates, including the nine winners. How is it possible that *the winners were not simply the top nine in votes*? More specifically, why

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

TABLE 1.2 *Votes for the leading candidates for election for members of Polish Sejm from the district of Konin, October 2015*

Number elected: 9 Candidate	Party	Votes	Votes	Elected?
Wojciech Witold Czarnecki	Law and Justice (PiS)	26399	9.52%	Yes
Zbigniew Dolata	Law and Justice (PiS)	18060	6.51%	Yes
Paul Anthony Arndt	Civic Platform (PO)	17925	6.46%	Yes
Tadeusz Tomaszewski	United Left	15350	5.53%	No
Tomasz Piotr Nowak	Civic Platform (PO)	11820	4.26%	Yes
Jan Krzysztof Ostrowski	Law and Justice (PiS)	9443	3.40%	Yes
Bartosz Jozwiak	Kukiz'15	8747	3.15%	Yes
Ryszard Bartosik	Law and Justice (PiS)	8163	2.94%	Yes
Leszek Richard Galemba	Law and Justice (PiS)	7708	2.78%	Yes
Agnieszka Mirecka-Katulska	Law and Justice (PiS)	7520	2.71%	No
Paulina Hennig-Klóska	Modern	7306	2.63%	Yes
Zofia Mariola Itman	Law and Justice (PiS)	6913	2.49%	No
Eugene Thomas Grzeszczak	Polish People's Party (PSL)	6609	2.38%	No
Kazimierz Czeslaw Broadsword	United Left	5174	1.87%	No
Maria Bychawska	Civic Platform (PO)	5053	1.82%	No

All candidates who obtained at least 5000 votes (about 1.75 percent) are shown; there were 120 additional candidates who are not shown.
Source: Authors' compilation from http://parlament2015.pkw.gov.pl/349_Wyniki_Sejm/0/0/37/3062.

did Tadeusz Tomaszewski and Agnieszka Mirecka-Katulska not win seats even when they had more votes than did other candidates who were elected?

The answer lies in an important detail that is typical of many proportional representation electoral systems, and which we explain in detail in Chapters 2, 5, and 6: the system uses "party lists." Parties, or alliances of parties, present lists of candidates. The country's system first allocates the seats to these lists.[8] Only then do the candidate votes come into play, with the top vote-earners in each list getting the seats that each list has obtained. Thus, not shown in the table, the PiS ran nine candidates, whose combined votes amounted to 37.4 percent of the district's total (quite close to the party's nationwide percentage, as Table 1.1 showed). This entitled them to five of the nine seats (55.6 percent); its winners are the five *on the PiS list* with the highest

[8] A vote for any candidate is also counted as a vote for the list on which the candidate is running for office.

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

vote totals. Similarly, the PO was collectively entitled to two seats – its top two candidates winning them – and K'15 and Modern to one each.

Tomaszewski and Mirecka-Katulska did not win a seat despite vote totals that were in the top nine because their *lists* did not have enough votes for them to win. In the case of Tomaszewski, this was due to the list of the United Left falling short of the *nationwide threshold*, as we saw in Table 1.1. Without the threshold, this list would have had sufficient votes to win a seat in the district, but in this case a rule that is applied on nationwide votes interfered. In the case of Mirecka-Katulska, she did not win even though her own votes ranked her ninth in the district overall, once Tomaszewski was excluded. Critically, however, she ranked only *sixth in her list*, which was entitled to five seats. Under a list system of proportional representation, the votes for the list of candidates are the first criterion in allocating seats. Various list systems are the most common of all electoral systems, and not just an unusual feature of Poland.[9]

Thus from Table 1.2 we see that at the district level, and even at the level of individual candidates, the electoral system affects who wins representation. This book is about all of these various ways that electoral systems matter.

How Electoral Systems Constrain and How Science Walks on Two Legs

Politics takes place in time and space – both the immutable physical space and the institutional space that politics can alter, but with much inertia. Institutions place constraints on politics. For instance, in a five-seat electoral district, at least one party and at most five parties can win seats. Within these bounds, politics is not predetermined, but the limiting frame still restricts the political game. It is rare for one party to win all seats in a five-seat district, while such an outcome is inevitable in a single-seat district. This observation may look obvious and hence pointless, but it leads to far-reaching consequences.

A key method followed in this book is the building of *logical quantitative models*. Much of contemporary social science is quantitative, in the sense of working with numbers, running and reporting statistical regressions, and so on. However, too little social science work builds its quantitative edifice on a foundation of *logic*. In this book, we will report many a regression result, but most of these are reported as tests of logical models that we derived before going to the statistical program and asking what the coefficients and standard errors (etc.) are.

In building logical models, we first ask, *what do we expect* the relationship to be between A and B? This means *thinking* about how A shapes B (and maybe vice versa). It means thinking about the shape of the relationship. Do not just

[9] Open lists, where candidates' votes determine the winners from the list, are less common. They are by no means rare, as we shall see in later chapters.

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

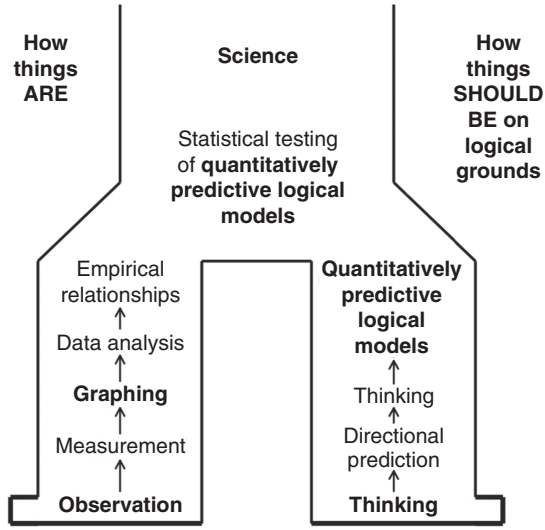

FIGURE 1.1 Science walks on two legs: observation and thinking
Source: Modified from Taagepera (2015).

run to the computer program and find out what a basic linear regression result is, because what if the relationship is not linear? We will display a lot of data graphs in the book, because it is important to see the scatterplot. This will tell us if our logic is on the right track, and whether our data need to be transformed – for instance, taking logarithms – before we enter them into regression equations.

The most important reason for thinking before you regress is that science walks on two legs, as shown schematically in **Figure 1.1**. As with any walk, the process involves *taking alternating steps on each leg*. However, the two legs of science represent different aspects of what science is. The walker can't reach the destination without using both legs. Hopping on one leg is highly inefficient!

One leg (the left in the schematic) deals with determining how things *are*. This involves careful observation, measurement, graphing of data, and statistical summaries of patterns in the data. The other leg deals with asking how things *should be*, on logical grounds? That question guides the first one. If it does not – if we jump to running statistical regressions first – we run the risk of seeing what we want to see. Or, worse, running numerous slightly different specifications of the regression equation, or different regression commands, until we see what we want to see. It is in thinking about "How things *should be*" that we come to understand *what* to look for before we use statistics. The two legs come together when expectations produced by logical modeling are tested with data, mostly using statistics. We will explain our use of statistics

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

later, but first – because it should be first – we discuss how we start with logical model building.

Let's take the Polish district of Konin, shown previously in Table 1.2. We saw that four parties won seats in this district in the 2015 election. There were a total of nine seats available. Is four parties a lot or few? To know, we might look at other districts in Poland in 2015, and also at districts in other countries. We might see that in the UK, every district elects only one party. Again, that's obvious – there is only one seat! Yet starting with the obvious is exactly how we start to build a logical model. If there is one seat, there can be only one winning party. If there are more seats, like the nine in Konin, we expect there to be more winning parties. We can look further, perhaps at Israel and find out that ten parties won seats, in a district that has 120 seats available (see Chapter 5 for details). So, here we are dealing with *observation*: Election districts with many seats available tend to have many parties win seats in them. The second step is *thinking* about this observation. This leads to a *directional prediction*: If there are more seats in the district, the number of winning parties increases. *Measurement* of the number of seats – what we call district magnitude – and the number of parties confirms this prediction.

But a merely *directional* prediction is of limited value. Any Toscana peasant could have told Galileo in which direction things fall. They fall down! What else do you need to know? Galileo also wanted to know how fast they fall, and why. If we want to be taken seriously as scientists with results of value to offer the world of practitioners, we must ask similar questions about the number of parties, and about every other directional relationship. Yet, far too many works published in political science journals neglect to venture beyond the directional hypothesis. We should not be like the Toscana peasant who might have said, "I see which way things fall, and that is all I need to know."[10] Whenever researchers can go beyond the merely directional, they should. What is the meaning of this abstract advice? A specific example follows.

An essential step is to *graph the data*. Then really *look* at this graph and *ponder* what it wishes to tell us. In Figure 1.2, we use district-level data, from many elections around the world. We see two panels, both of which plot on the x-axis the number of seats in a district (or its magnitude, designated M) and on the y-axis the number of parties (of any size) that win in the district (designated by the strange looking label, N'_{S0}, for reasons that will become clear in later chapters). The difference in the two panels is the way the scales are drawn.

[10] We will accept that there are applications in which the directional hypothesis is the best a social scientist can do, and even where confirming such a hypothesis adds considerable value. However, in many applications – especially those that are the substantive topics of this book – we really must strive to be more specific in our expectations.

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

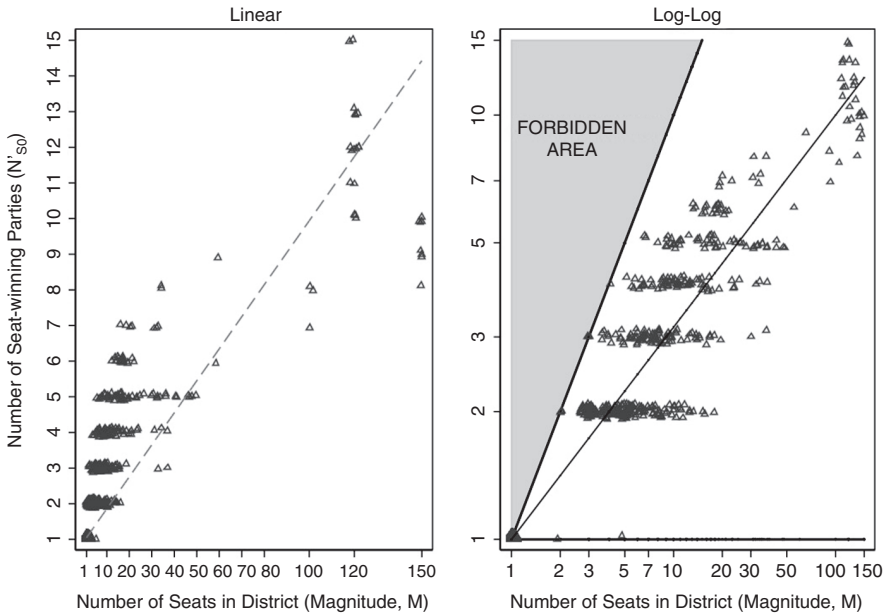

FIGURE 1.2 Two ways of visualizing the relationship between district magnitude (M) and the number of seat-winning parties (N'_{S0})

When graphed on regular scales, as in the left panel of Figure 1.2, the pattern is not very impressive. We use elections in more than eleven thousand districts all over the world. It may not look like that many, in part because only 451 of them are districts with more than one seat. All the one-seat districts just jumble up at the lower left of the graph. (We will explain the dashed line later.) We might notice that there seems to be some curvature in the data pattern, but it is hard to tell in this format what that might be.

In the right panel of Figure 1.2, we have exactly the same data points graphed with both axes being on logarithmic scales ("log–log"). In this graph, we have added two thick dark lines to help us further our logical thinking. We should graph the data, but we should do more: *we should graph where the data cannot possibly be*. It is logically impossible for the number of winning parties to be greater than the number of seats in the district. You cannot have ten parties win in your district if there are only nine seats available, for instance. The thick diagonal line thus shows a hard constraint: $N'_{S0} \leq M$. The area above and left of this line is shaded and marked as a "forbidden area" because there simply cannot be data points here, and marking a plot area in this way can be useful for thinking through logical relationships. The thick horizontal line is the minimum: it is impossible to have fewer than one party win in the district. Perhaps that seems too obvious to say, but sometimes the obvious is the clue to a logical quantitative model of the relationship.

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

Further *thinking* leads to the prediction that the number of parties winning in a district should be about the square root of the district magnitude,

$$N'_{S0} = \sqrt{M},$$

which also can be written as

$$N'_{S0} = M^{0.5}. \qquad \text{(Eq.1.1)}$$

We will explain the logic behind Equation 1.1 later, in Chapter 7. The short version is that this is simply the average of the maximum, $N'_{S0} = M$, and the minimum $N'_{S0} = 1$, when both variables are expressed in logarithmic format – technically known as the geometric average. Without graphing, seeing the shape of the data distribution, and thinking logically, we could not have arrived at Equation 1.1, which we first derived in one of our earlier collaborations (Taagepera and Shugart 1993). This equation is a fundamental building block for most of what comes in later chapters of this book.

Now – and only now – *statistical approaches* enter. If we ran simple linear regression on the two variables, we would get the dashed line in the left panel. It simply plots the output of a regression. Without graphing the data, we might have said our directional hypothesis is confirmed: N'_{S0} goes up as M goes up. We would be cheered that the relationship was "statistically significant" and that the R-squared measure of how well the equation fits was almost unbelievably strong for a political-science relationship, at R^2=0.74. How wonderful! Meanwhile, if we had not graphed it, we would not even realize that the fit is not nearly as good as the R^2 would lead us to believe. Deviation from straight line is not random but systematic: at low M almost all data points are above the line, while at high M more data points are below the line. Visibly, the straight line is not the best-fit curve, even though the statistical program suggests it is!

Instead, after having studied the graphs and done our thinking, we could run linear regression on the *logarithms* of number of parties and district magnitude – *not* on the quantities themselves. When we do that, we get Equation 1.1![11] In other words, we confirm the logical model that the number of parties winning at least one seat in a district tends to be the square root of the district magnitude. Even better, the R^2 on the log-log regression is 0.95.

The end point of the process is a *quantitatively predictive logical model*. This model is "quantitatively predictive" because it predicts not only the direction of change but also the quantity of parties, on average, at any given district magnitude. The model is "logical" in that the square-rooting of the district magnitude comes from logical considerations. Note that we took alternating steps with each of the legs on which science walks. We started with observation, the left leg of Figure 1.1, followed by directional thinking, the right leg.

[11] The actual coefficient is very slightly bigger than 0.5, but trivially so. The result is discussed (and the regression table displayed) in Chapter 7.

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

Graphing involved the observation leg. Then quantitative logical modeling took us to the thinking leg. This is the type of interaction we have in mind when saying that *whenever researchers can go beyond the merely directional, they should*. For a broader perspective on such logical approaches, see Taagepera (2008).

This section has demonstrated the interplay of logical and empirical work in finding the relationship between the number of parties and number of available seats in a district (its magnitude). It is just one example. Each of the features of how electoral institutions matter, and that we reviewed in our sketch of the Polish election result in the preceding section, can be understood through a systematic, quantitative relationship. In later chapters of this book, we develop models of, among other topics: how the relative timing of elections (such as the "honeymoon" election in Poland 2015) shapes changes in the president's party vote share; how assembly size and average district magnitude shape how many important parties there are in the assembly as a whole; how the magnitude of an individual district shapes the sizes of parties there; and how candidates' personal vote totals (i.e., what we saw in Table 1.2) are distributed in a district.

The Use of Statistics in This Book: Think Before You Regress

In the preceding subsection, we discussed the importance of undertaking graphing and logical thinking before running regressions. Ultimately, testing our quantitative logical models through statistical regression is a critical part of the enterprise of this book. However, our approach to presenting regression results is different from many standard works, as we explain here.

Statistical "analysis" is a term often used in an overly broad way. The process should be thought of as having two utterly different functions. One is statistical *description* of data: best fit to whatever is deemed a suitable mathematical format, including the values of constants in this format, measures of lack-of-scatter around this best fit (such as R^2), etc. The other is statistical *testing* of preconceived models: how well the prediction agrees with data *average*. In model testing, R^2 is far more marginal as a factor in assessing a result than it is when we do not have a logically grounded expectation before running the statistical commands. Measures of goodness of agreement with expectation, such as *F*-test, take precedence.

Strictly speaking, we will call it *analysis* only when we are testing a logical model, not any time we run a regression. For similar reasons, when we display results of two or more statistical tests in a table, we will use the expression, "Regression One" (etc.), rather than the common "Model One." Only when we have a prior expectation of at least some of the coefficient values will we say that we have a "model."

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

In addition, precisely because most of our regressions are testing a specific quantitative expectation from a logical model, we will dispense with the usual "stars" that clutter up most published regression tables. Too often most authors are interested only in whether the sign of a coefficient is "right" (i.e., the expected direction) and whether the coefficient is "significant," meaning statistically distinguishable from zero. Most regression coefficients printed in journals and books are thus "dead on arrival" (Taagepera 2008: Ch. 7). The numbers in regression tables are rarely used for any further inquiry.

By contrast, we are interested in the coefficients' values, and whether a given coefficient is distinguishable from *the value our logical model says it should be*. Thus in many of our regression tables, we will run *F*-tests not of whether the coefficient might actually be zero, but whether its value is the logically expected value, or is too far off to consider the model supported.[12] On the other hand, sometimes our logical model actually demands that a constant term (intercept) be zero[13]; in such cases, "statistical insignificance" is actually what we want to see! We will revert to the practice of reporting conventional tests of significance (from zero) only in cases – very rare in this book – in which we do not have a specific quantitative expectation, but only directional.

Many of our regression results are collected in chapter appendices, in order to avoid breaking of the flow of text and the all-important visual test offered by our graphs. Only some regressions most critical to demonstrating the success of our modeling will be reported in tables located within the main chapter text. Some other specific decisions we make regarding setting up regressions and transforming data beforehand are discussed in chapters (or their appendices) where the matter comes up. For a lengthier treatment of the principles, see Taagepera (2008).

What Does "Quantitative Prediction" Mean?

The term "prediction" is subject to different meanings. When we use it, it usually means "quantitative prediction": if these factors have these values, then, *ceteris paribus* (all other conditions being the same), this other factor has that value, within some range of likely variation. The following example from the realm of electricity clarifies what this means.

Ohm's law, $I=V/R$, implies that if a potential difference of $V=25$ volts is applied across a wire with known resistance $R=5$ ohms, then a current of $I=5$

[12] Similarly, we will report 95 percent confidence intervals on these coefficients, to enable the reader to see at a glance whether the expected value is within this interval.

[13] For example, in Equation 1.1, the regression on the logarithms of N'_{S0} and M must yield a constant of zero, in order that we get a prediction of $N'_{S0}=1$ when $M=1$. (The log of 1 is equal to 0).

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

amperes will flow, with a possible variation of plus or minus 1 ampere. Here a variation range of 0.7 amperes emerges already because 25 volts could mean anything between 24.5 and 25.5 volts, while 5 ohms could mean anything between 4.5 and 5.5 ohms. The *ceteris paribus* provision presumes a conducting material with no semiconductor elements mixed in, and so on. So a possible variation range of ±1 ampere could be expected, until V and R can be measured with more precision. Repeated measurements at various V and R could also establish an empirically observed range of variation.

Our Equation 1.1, $N'_{S0}=M^{0.5}$, has the same broad format $y=cx^k$, so frequent in physics. (Many more will follow in this book.) It implies that if a district has $M=25$ seats, then $N'_{S0}=5$ parties will win seats, with a likely variation range of several parties. Empirical data in Fig. 1.2 indicate that the likely range extends from three to seven parties. More generally, if one wants to be on the safe side, one might say that the number of seat-winning parties is the square root of district magnitude multiplied or divided by no more than 2. (This is equivalent to saying that $\log N'_{S0}=0.5\log M\pm 0.3$.) The *ceteris paribus* provision presumes that some PR allocation rule is used. The playfield such rules allow to democratic politics does not exclude the possibility that once in a (very long) while $N'_{S0}=1$ or $N'_{S0}=25$ could materialize. It just says that $N'_{S0}=5$ is the most likely outcome for $M=5$, and that $N'_{S0}=5\pm 2$ covers pretty much all the observed outcomes.

In sum, when we say "quantitatively predictive," we mean nothing more and nothing less than what physicists attribute to statements like Ohm's law: a specific value of one factor leads to a specific value of the other.

How Electoral Systems and Party Politics Are Related

In a previous subsection, we showed an exercise in how we arrived at a logical model, using the example of how the number of parties is connected logically to the district magnitude. This insight in turn allowed Taagepera (2007) to hit upon a significant breakthrough in how electoral systems and party politics are related; in this book, we take those findings several steps further.

When we are interested in party politics, we are usually interested not just in the seats, but also in the votes. Moreover, while the number of parties – of any size – turns out to be a useful building block for logical model building, it is not *intrinsically* very useful. In any real legislature or national electorate, there are bound to be some parties that are bigger than others. Do we count even the tiniest party as being just as important as the biggest? When we are looking at small districts, it may not matter. There is something intrinsically interesting about the difference between a five-seat district with four parties represented and another with just two.

When we go to the nationwide level (or a bigger district), it is not so straightforward. Take the UK parliament following the 2015 election. The biggest party had over half the seats in parliament, which is obviously

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

consequential because this allowed it to form the governing cabinet on its own. At the same time, there were eleven other parties winning seats; six of these had just one to four seats each, in the 650-seat house. Do we really think the raw count is of interest? Well, probably not, as the relative sizes of the parties are more politically relevant than their mere number. This is where a concept called the *effective* number of parties comes in, but we will not explain what this means until Chapter 4. (Readers already versed in works on electoral and party systems will be familiar with this concept, as it has become a standard measure since its introduction by Laakso and Taagepera 1979.)

The relative sizes of parties become even more important when we turn to the votes. Take the UK 2015 election again. The largest party had about 37 percent of the votes, or more than eleven million out of over thirty million total. But how many parties won votes? It is not so easy to answer. Even if we counted only parties that ran candidates for a minimum of 5 percent of all seats – an arbitrary cutoff – we would find that this included at least one party with barely over 6,500 votes (0.02 percent). In other words, some vote-earning parties really are too small to "count." Hence the need for a measure that considers their differing sizes. Even more than for the sake of measuring, however, the greater difficulty of specifying just how many "vote-earning parties" there are becomes a fundamental insight to thinking about how electoral systems shape party politics. We go into this question in depth in Chapter 8.

Most political scientists who have attempted to connect electoral systems to the number of parties have been satisfied with directional hypotheses. Their method has been to test via statistical regression whether the ("effective") number of parties goes up, as expected, with district magnitude and other input variables posited to have either a "restrictive" or "permissive" effect on increases in the number of parties. In some influential works,[14] additional input variables that are held to explain how many parties earn votes are things like how many ethnic groups there are and how many presidential candidates ran (zero if the system is parliamentary – which is itself a problematic "measure" for reasons we explain in Chapters 7 and 12).

These standard works then take the number of vote-earning parties and enter it into a regression in which they seek to explain the number of seat-winning parties via electoral rules.[15] That is, a common scheme regarding how electoral systems and party systems are related looks something like **Figure 1.3**. The figure represents a highly stylized version of the relationship as seen by proponents of the now-standard approach: the

[14] For instance, Ordeshook and Shvetsova (1994), Amorim Neto and Cox (1997), Cox (1997), Hicken and Stoll (2011), and Clark and Golder (2006).

[15] In these regressions, it is again standard to use the "effective" number, not a raw count.

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

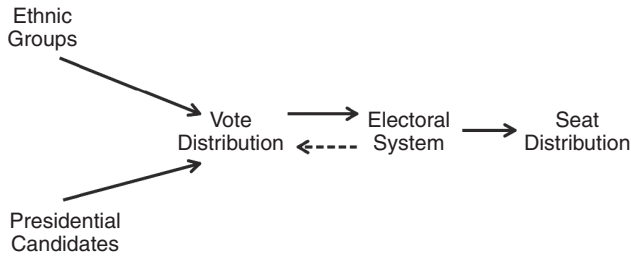

FIGURE 1.3 A common way of seeing the impact of the electoral system on votes and seats

distribution of votes is shaped by a societal feature, such as the number of ethnic groups, as well as by the number of competitors for the presidency (if there is one). The votes are also affected by the electoral system, but only indirectly, according to this standard view – hence it is depicted with a dashed line. That is, according to many authors, the electoral system effect is only fully realized in *interaction* with social factors. For instance, an electoral system favoring many parties nonetheless may result in relatively few parties unless the number of ethnic groups is large, but an electoral system that makes it hard for many parties to win may reduce the number of parties even if there are many ethnic groups. The problem with the approach depicted in Figure 1.3 is that it results in estimates that lack logical foundation, and sometimes are odd or even absurd. We discuss these problems further in Chapters 7 and 12.

The main flaw in this conventional thinking is to try to estimate the seats only after the votes have been estimated. The reason this is flawed is that votes are the *less constrained* of the two quantities. *Thus authors who do this are left with only the ability to make a directional hypothesis*: the number of vote-earning parties goes up as ethnic groups go up, if the electoral system is permissive of such increases, or similar phrasing. However, a more promising approach is to come at it from the opposite side. As we saw in the preceding section, *seats are constrained by district magnitude*. We can thus derive a logical model of the relationship (Equation 1.1). At the national level, they are further constrained by the size of the assembly, a critical consideration that we introduce in Chapter 7 (see also Taagepera 2007).

Figure 1.4 shows the opposite impacts of electoral systems and party politics on the distribution of seats and votes among parties. *Electoral systems restrict directly the way seats can be distributed*. In particular, when single-member districts are used, only one party can win a seat in the given district. *The impact on votes is more remote*. When a party fails to obtain seats in several elections, it may lose votes because voters give up on it, or it may decide not to run in the given district. The impact on party

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

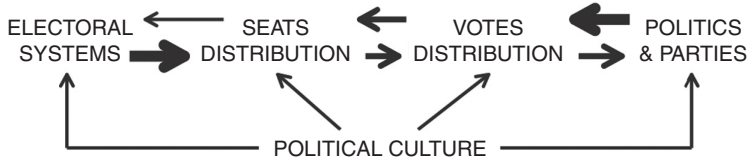

FIGURE 1.4 The opposite impacts of electoral rules and current party politics
Source: Adapted from Taagepera (2007).

systems[16] and hence on politics in general is even more remote. Still, if a party fails to win seats all across the country, over many elections, it may fold, reducing the number of parties among which the voters can choose.

The impact of the existing party system and current politics is attenuated in the reverse direction. The total number of meaningful parties may be limited by the workings of the electoral system, but current politics determines which parties obtain how many votes. The impact of current politics on the seats distribution is weaker, as the electoral system may restrict the *number* of parties that can win seats. Still, current politics determines *which* parties win seats. Finally, current politics has no impact at all on the electoral system, most of the time. Yet, infrequently, it has a major impact, when a new electoral system is worked out from scratch (for instance, in New Zealand in the 1990s – see Denemark 2001).

At all stages, political culture plays a role. The same electoral laws play out differently in different political cultures, shaping different party systems. Along with the initial party system, political culture shapes the adoption of electoral laws. If stable electoral and party systems succeed in lasting over a long time, this experience itself can alter the initial political culture – a connection not shown in Figure 1.4. This book mentions political culture rarely – we return to the issue briefly in Chapter 7. Our infrequent reference to the term is not because we underestimate it. We just do the relatively easy things first – institutions and their constraints; political culture is harder to tackle.

In the study of current politics, votes come first, and seats follow – the arrows at the top of Figure 1.4. This direction may look natural, and that is presumably why conventional approaches frame the problem this way (as in Figure 1.3). However, crucially, *the direction is reversed when we study the impact of electoral systems*. Now seats are restricted directly (as in Equation 1.1), and restrictions on votes follow in a more diffuse way – the arrows at the bottom of Figure 1.4. This is precisely why this book is entitled ***Votes from Seats***. In Chapter 8, we offer a novel logical model for understanding how the votes

[16] In this book, when we say, "party system" it is just shorthand for how many parties there are, and their relative sizes. For a caution on the use of the word, system, see Taagepera (2007: 5–7).

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

follow from the seats. Recognition of such reversal of the direction of effects is essential for elucidating the impact of electoral systems.

Plan of the Book

The remainder of this book consists of five parts. In Part I, we introduce **Rules, Tools, and Context**. This section consists of five chapters. Chapter 2 introduces the main components of "simple" electoral systems, including the main rules for proportional representation. Chapter 3 delves into more complex rules, including ballots that give voters the ability to rank candidates, systems with "runoffs," and composite systems that combine two or more different components. In Chapter 4, we introduce two key measuring tools (and some others as well): the *effective number of parties* and *deviation from proportionality*. These are tools for assessing how rules work in their context. Then we have two chapters that situate rules and tools more deeply in context by exploring examples of electoral systems in countries chosen to illustrate how systems work.

The second part consists of four chapters on the **Interparty Dimension of Assembly Politics: The Seat Product Model**. When we are counting parties or measuring deviation from proportionality we are analyzing the interparty dimension, because we are concerned with how many parties there are, their relative sizes, and the relationship between their votes and seats. These concerns have been fundamental to the electoral systems literature, because they are crucial to so much of democratic politics. They get right to the question of how voters are represented and whether one party governs alone or several must cooperate.

The chapters in Part II develop the Seat Product Model, which states that the effective number of parties in the assembly has a systematic relationship to the product of a country's assembly size and mean district magnitude. This model was first proposed by Taagepera (2007), but in this book we are able to expand its scope in several ways. First, in Chapter 7, focused on nationwide assembly party systems, we test it on a large multi-country dataset of individual elections. Then in Chapter 8, we offer a novel logic that allows us to predict the effective number of vote-earning parties from the same input factors. Chapter 9 shows how the preceding chapters have allowed for the development of four *basic laws of party seats and votes*. It also offers an application of the laws to the question of how electoral systems shape one of the most important outcomes of electoral systems – their degree of proportionality – which, we find, also is affected by a country's seat product.

While Chapters 7–9 are about nationwide effects, in Chapter 10 we focus our lens on the district level. It has long been recognized that it is at the district level where the impact of rules is most felt – for exactly the reason we showed earlier in this chapter in our discussion of Equation 1.1. Of course, it is true, but we show that the reverse is true, too, and more fruitful for deriving quantitative

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

logical predictions: *the district-level effect depends on the size of the nationwide electoral system* (i.e., the number of assembly seats) in which the district is *embedded*. The notion of embeddedness, which we extend from an earlier concept that we developed but barely were able to test at the time (Taagepera and Shugart 1993), becomes essential as a building block for several further analyses in later chapters.

The entire second part is focused principally on parliamentary systems, because those are where the number and sizes of parties directly shape government formation. However, we also are able to show that presidential systems – where there is a politically powerful executive elected separately from the assembly – are not fundamentally different from parliamentary in the way electoral systems shape assembly party systems. This is a finding at variance with the conventional literature, which claims that we must know about the number of candidates competing for the presidency (as depicted in Figure 1.3) in order to understand assembly party systems in presidential democracies.

In Part III, we explore the consequences of **Bringing the President In**. Many works on electoral systems and party systems have treated the vote for president as a factor conditioning the vote for assembly (as depicted in Figure 1.3). While such an understanding is surely correct in some sense, it is also limiting. It would be more useful if we could tie all these various party system outcomes – *including the distribution of presidential candidates' votes* – to an institutional input. In Chapter 11, we show that we can do just that. We make the shocking claim that the Seat Product Model devised for assembly party systems also can predict the effective number of presidential candidates. That is, assembly electoral systems have primacy, even in presidential systems.

A key institutional feature of presidential systems is how elections are timed relative to one another (recall our example of Poland, 2015, earlier in this chapter). Other authors have attempted to use this factor to help explain the number of parties in the assembly. Yet they have done so with only minimal theoretical logic for the notion. We show instead, in Chapter 12, that the effect of election timing is on the *votes won by the president's party*, thereby confirming an effect first shown by one us some years ago (Shugart 1995). Except when an assembly election is very late in a president's term, knowing the timing of the elections does not help us predict the number of parties in the assembly, but it critically shapes the support the president obtains in the assembly.

Part IV turns attention to the **Intraparty Dimension of Competition**. Some electoral systems pit candidates of the same party against one another, meaning there is another dimension of competition aside from that between parties – an intraparty dimension. This dimension has been less studied than the interparty. Nonetheless, we show that it, too, follows predicable patterns, and is constrained by the electoral rules. Part IV consists of two chapters. In the first of these, Chapter 13, we develop logical models of how two different types of

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

electoral system that let voters select a candidate within a party shape the distribution of votes across multiple candidates. In Chapter 14, we look at how different rules for intraparty competition shape how parties "manage" the votes distribution across their candidates at the district level. We also address how parties divide up the nominations and the votes when forming alliances in one type of proportional representation system.

Part V is devoted to the question, **What Can We Expect From Models of Electoral Systems?** While most of the book has focused on "simple" electoral systems, many of the world's electoral systems actually are complex. So can our models help predict average outcomes in complex systems? In Chapter 15 we extend the Seat Product Model to "upper tiers."[17] This allows us to expand the range of countries to which the Seat Product Model applies, including a common class of "complex" system. We also consider ethnic diversity, which has been suggested by other works on electoral-system effects. We find that adding ethnicity does not improve much on the Seat Product predictions.

In Chapter 16, we ask whether predictions for other types of complex systems can be made using the Seat Product Model. We introduce another logical model of the impact of one common complexity – a nationwide threshold. We show that for some systems it predicts well, but most of the time our predictions from the Seat Product Model work better. The predictive power of the Seat Product Model even holds for other forms of complexity, such as ranked-choice ballots and runoffs, with some caveats. This is good news for electoral system designers, because it means that introducing some minimal complexity – sometimes desirable for specific purposes – does not mean we are flying blind in understanding the effects. We also offer some examples of overly complex systems, and the problems encountered by them.

Finally, Chapter 17 assesses what we can expect from models of votes and seats, and how they offer a baseline for evaluating country-specific deviations from model predictions. It further points out how the book's methods can contribute to advance the science of politics. We conclude with some basic recommendations for electoral-system design, drawn from our findings.

[17] An upper tier is a set of seats allocated based on votes accumulated across the various districts of the "basic tier."

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002

Downloaded from https://www.cambridge.org/core. Balfour Library (Pitt Rivers Museum), on 31 Aug 2021 at 14:07:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108261128.002