Votes fr

Logic

MATTH
University of C

REIN TAAGEPE
University of California, I

CAMBRIDGE
UNIVERSITY PRESS

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.

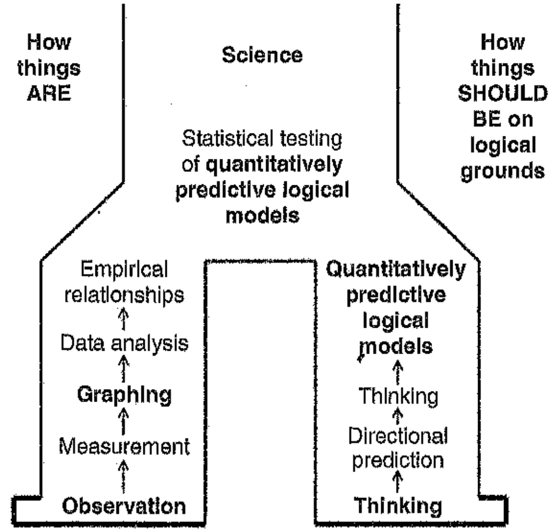

FIGURE 1.1 Science walks on two legs: observation and thinking
Source: Modified from Taagepera (2015).

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 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'_{SO}, 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.

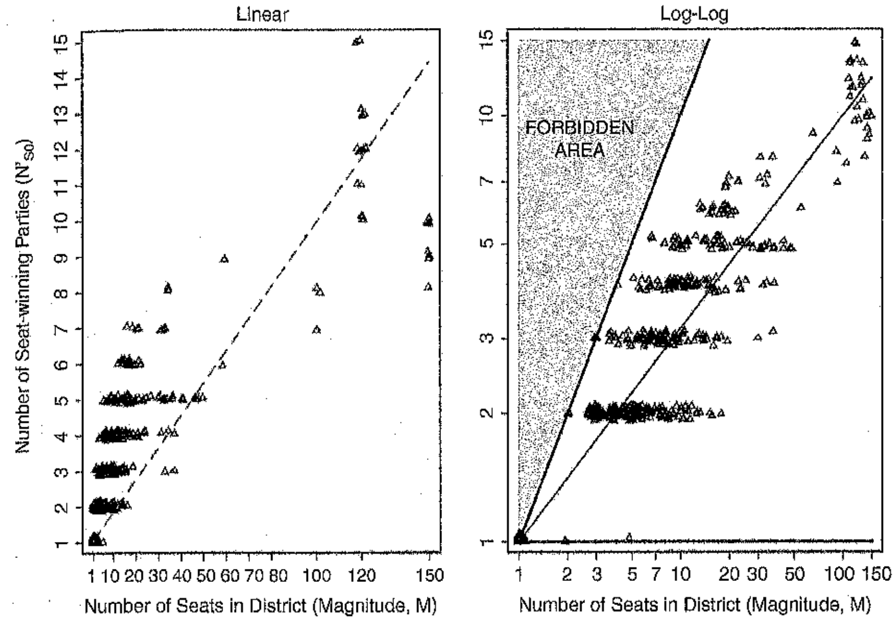

FIGURE 1.2 Two ways of visualizing the relationship between district magnitude (M) and the number of seat-winning parties (N'_{SO})

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: $N'_{SO} \leq M$. 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.

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.

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."

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).

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 $\log N'_{S0}=0.5\log M\pm0.3$.) 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 $N'_{S0}=5\pm2$ 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

7

The Seat Product Model of the Effective Number of Assembly Parties

How do electoral systems shape party systems? Here we begin to address this question seriously, meaning quantitatively, walking on the two legs pictured in Figure 1.1 at the start of this book: observation and thinking. We begin with nationwide assembly parties. These are more directly affected by the institutional framework, compared to electoral parties, which will come in Chapter 8.

The tradition of such study is long in political science. Yet Clark and Golder (2006: 682) summed up the prevailing view when they concluded that "the so-called institutionalist approach does not produce clear expectations" about the number of parties, and that "everything depends on the presence or absence of social forces." Such a pessimistic claim about the "institutionalist approach" is not justified. Once the institutions themselves are more fully specified, we can have clear expectations for worldwide average patterns, which in turn can offer guidance to practitioners.[1] Demand for such guidance has been demonstrated: many political scientists have engaged in consulting missions in countries experiencing transitions from authoritarianism (Carey et al. 2013).

As a preview to what this chapter is about, we offer Figure 7.1. It deals with the effective number of seat-winning parties (N_S) in the first or sole chamber of national assemblies. This graph uses logarithmic scales and graphs N_S against a possibly surprising quantity: the product of mean district magnitude (M) and the size of the assembly (S). We can see that the relationship looks pretty tight. Our graph further shows the line that expresses the Seat Product Model for N_S:

$$N_S = (MS)^{1/6}. \tag{7.1}$$

This line visibly expresses the average trend well. We explain its derivation later in this chapter. Dashed lines indicate values that are twice, or half, the value predicted from Equation 7.1. With presidential and parliamentary

[1] It is not that social forces of various types such as ethnic divisions are irrelevant. Rather, their impact on the nationwide party system can be felt only within a range set by the institutional rules. Analysis of the (rather limited) impact of a common measure of ethnic diversity will be one of our themes in Chapter 15.

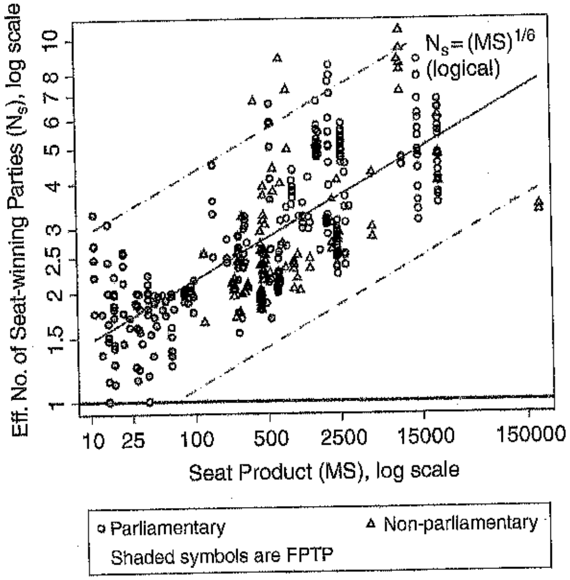

FIGURE 7.1 Relation of the nationwide effective number of seat-winning parties (N_S) to the seat product (MS)
The central line is the logical model $N_S=(MS)^{1/6}$, *not* a statistical best fit (see text). Dashed lines mark values that are double or half the logical model's predictions.

systems shown with different symbols, we can see that the scatter is greater for presidential (triangles), but the average pattern remains the same. Yet it should be apparent that, in general, presidential systems do not require a fundamentally different means of explanation – some conventional literature that we review later notwithstanding.[2]

In Figure 7.1 we also differentiate FPTP cases from PR, with the former symbols being shaded. While almost all cases of $MS<100$ are FPTP countries, there is an intermixing of PR and FPTP when MS is greater than 100 up to 650, beyond which all systems are PR. The graphing of systems in this way allows us to see that *we do not need separate models to account for "FPTP versus PR,"* as many others might frame it. The Seat Product, MS, is more important than these two old categories. This amplifies a point from Chapter 2: *FPTP is just PR reduced to its minimum possible magnitude,* $M=1$. If a PR system has MS around 600, perhaps due to a mean magnitude of five and $S=120$, its impact on N_S is about the same as if it were FPTP with $S=600$.

[2] Our "parliamentary" subsample in this and subsequent chapters includes Austria, Finland, Iceland, Portugal, and Switzerland, which actually have hybrid executive formats. Our results do not depend on this choice. More details on our classification of presidential systems and data results for these systems can be found in Chapter 11 and its appendix.

A crucially important point of emphasis is that the central line in Figure 7.1 is *not* a statistical best-fit line. This line represents a model (Equation 7.1) derived logically, *without using any data.* It fits remarkably well. In fact, we can summarize its average fit not only via visual inspection of Figure 7.1, but also with a series of ratios for different subsets of the data. These ratios are a given election's observed value of N_S, divided by the prediction from Equation 7.1, averaged over all elections in the given subset: r=(value observed)/(value expected). Our ideal would be $r=1.00$.

If we consider all parliamentary systems, the mean is $r=1.07$ (standard deviation, 0.34, median 0.988). For parliamentary and PR, we get $r=1.125$ (0.35, 1.05) and for parliamentary FPTP $r=1.029$ (0.33, 0.958). Turning to nonparliamentary systems, for the full subsample we get $r=0.930$ (0.41, 0.774). Note how these data reflect the greater scatter of presidential systems. They do not fit as well, for reasons we take up in detail in Chapter 11. On the other hand, on average, they are not wildly off the predicted values.[3]

How is Equation 7.1 obtained? This is the topic of the next section. Thereafter we test the degree of validity of this central part of the Seat Product Model, using standard statistical means. Finally, we place this model in the historical context of the "Duvergerian agenda" and comment on some other factors.[4]

THE SEAT PRODUCT MODEL FOR ASSEMBLY PARTIES: THEORY

The Seat Product Model (SPM), introduced by Taagepera (2007) and foreshadowed by Taagepera and Shugart (1993), has the important advantage of relying strictly on institutional input variables, which are in principle subject to (re-)design. It does not include rather more immutable societal variables or any variables that are themselves the product of behavior that may be shaped by the institutions, such as competition for a presidency (as in several well-known works reviewed later). The model is also unique within the broader literature of the Duvergerian agenda, to be described later on, in that it has a stable and logically supported coefficient

[3] We could further break the presidential cases down by electoral system, but the combination of FPTP and presidentialism is heavily dominated by one country, the US. For the several countries that combine presidentialism and PR, we find a mean $r=1.025$, but again evidence of greater variability from a standard deviation of 0.47 and a median value of 0.855.

[4] In Chapter 8, we develop a new logical model to explain the votes. The reason for starting with seats is that seats are most directly constrained by institutions; it is to the constraints on seats that voters and other actors adapt. Focusing first on the seats is also justified by the importance of the seats for determining who governs, if the system is parliamentary, or for who can pass legislation (in either presidential or parliamentary systems). Strictly speaking, then, seats are more important than votes to understanding a democratic polity.

for the key institutional variable. In this section, we offer a condensed overview of the logic. For the full details the reader should consult Taagepera (2007).

The Logic Underlying the Number of Seat-Winning Parties

We already introduced in Chapter 1 the fundamental building block of the relationship of the number of parties that win at least one seat in a district and the district magnitude (M). Equation 1.1 expresses this relationship:

$$N'_{SO} = M^{0.5},$$

seat-earning

where N'_{SO} is the actual number of parties, of any size. The apostrophe in N'_{SO} designates a district-level quantity (in contrast to nationwide); subscript S refers to seats (in contrast to votes), and 0 indicates raw count of parties (in contrast to the effective number). This is a systematic notation that we shall adopt throughout this book. In the introductory chapter, Figure 1.2 (specifically, its right panel) displayed this equation and showed it fits well. However, Equation 1.1 is something we can derive purely from logic, *without using any data!*

The model starts at the level of a single district that elects a certain number of legislators, designated as M for the district's magnitude. For any value of M seats to be allocated, the lower bound on the number of seat-winning parties is obviously one (single party wins all M seats), while the upper bound is M (each seat won by a different party). Thus the possible range of N'_{SO} is

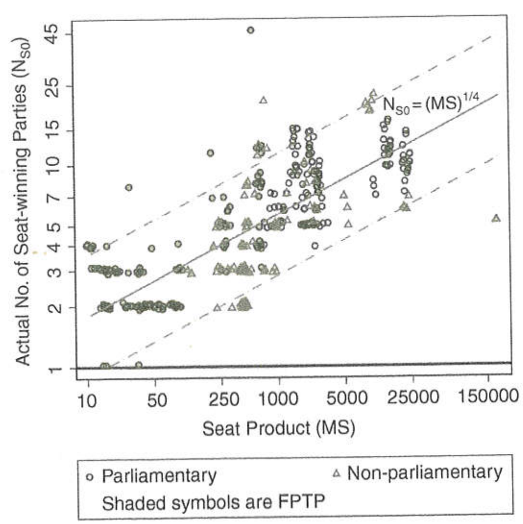

FIGURE 7.2 How the actual number of seat-winning parties (N_{SO}) relates to the seat product (MS), national level

$1 \leq N'_{SO} \leq M$. The most likely actual (not effective) number of parties to win at least one seat is the geometric mean of these boundaries (Taagepera and Shugart 1993; Taagepera 2007, 119).[5] The geometric mean of 1 and M yields what we already identified as Equation 1.1: $N'_{SO} = M^{0.5}$.

This approach does not rule out the possibility that considering other information may render a better estimate. Rather, it simply means that the geometric average of extreme values is our best estimate *in the absence of other information* – "the worst possible prediction one could make – except for all others" (Taagepera (2007: 12). We will test this district-level prediction below, but first, how do we go from the district level to the national level, thereby arriving at the Seat Product Model that yields Equation 7.1?

Suppose a country, using a nonmajoritarian formula, elects its legislators in districts of magnitude M. If it elected all its legislators in a single nationwide district, then the magnitude of this one district would be the same as the assembly size, S. By Equation 1.1, we thus would have $N'_{SO}=S^{1/2}$, when $M=S$. When $M<S$, the number of parties cannot exceed this number because that system cannot be more permissive than a national-district system. This means $S^{1/2}$ is an *upper* bound of the number of parties for any given M and S.

At the same time, the number of parties winning at least one seat *nationwide* (N_{SO}, without the apostrophe) cannot be smaller than the district-level number of parties, already estimated as $M^{1/2}$, which therefore is a theoretical *lower* bound. Thus the possible range of N_{SO} is $M^{1/2} \leq N_{SO} \leq S^{1/2}$. Then, for any given M and S, in the absence of other information we should expect the geometric mean of $M^{1/2}$ and $S^{1/2}$:

$$N_{SO} = [(M^{1/2})(S^{1/2})]^{1/2} = (MS)^{1/4}. \qquad (7.2)$$

The same logic applies for any value of M and S, provided all seats are allocated within districts and the formula is nonmajoritarian – the key features defining "simple" electoral systems (Chapter 2). Figure 7.2 shows the scatterplot of the data, with the solid line representing Equation 7.2. As in Figure 7.1, dashed lines indicate values that are twice, or half, the predicted values. The fit is visually good for worldwide average, albeit with a few cases quite scattered. Later in the chapter we offer a regression test of Equation 7.2.

[5] The geometric mean of two quantities is the square root of their product. But why the geometric mean rather than good old arithmetic? Ask the original question in a slightly different way: How many seats are seat-winning parties likely to win, on the average? If M=25, it could be from one to twenty-five, depending on the number of such parties. The geometric mean offers five parties at an average of five seats each, which multiplies to the twenty-five seats we have. In contrast, the arithmetic mean would offer thirteen parties at an average of thirteen seats each, for a total of 169 seats! Why doesn't the arithmetic mean work? See Taagepera (2007, 119, and 2008, 120–127).

Of course, most of the time we are unlikely to be interested in the raw count of parties in a legislature, which tells us nothing about their relative strengths. It was for precisely this reason that the effective number of parties was devised (Laakso and Taagepera 1979), and this has become by far the standard index of party-system fragmentation. In order to derive N_S from N_{S0}, the next step in the chain is to derive the largest party's seat share (s_1).

The Logic Underlying the Largest Seat Share

We can deduce the boundaries of this quantity's range from N_{S0}. For any given number of parties represented, the smallest possible value of s_1 is when all parties are equal-sized: $s_1=1/N_{S0}=N_{S0}^{-1}$. The largest is as close to 1.0 as feasible to still allow the remaining $N_{S0}-1$ parties to have one seat each.[6] Simplifying a bit, we can again try the geometric average – in this case of 1 and N_{S0}^{-1}:

$$s_1 = (1 * N_{S0}^{-1})^{1/2} = N_{S0}^{-1/2}.$$

If the largest seat share fits, then we must also have, substituting $(MS)^{1/4}$ for N_{S0} (cf. Equation 7.2),

$$s_1 = [(MS)^{1/4}]^{-1/2} = (MS)^{-1/8}. \quad (7.3)$$

Figure 7.3 shows scatterplots of the data related to both models for s_1. In the left panel, we graph s_1 against N_{S0}. We see a good fit to the expectation, $s_1 = N_{S0}^{-1/2}$, for the world average. The right panel graphs s_1 against the Seat Product, *MS*. Here, too, we see a good fit to our expectation, i.e., Equation 7.3. Now we are ready to turn to the final link of the chain leading us to Equation 7.1, the core expression of the Seat Product Model for the effective number of seat-winning parties.

The Logic Underlying the Effective Number of Seat-Winning Parties

The largest seat share (s_1) is the single most important component in the calculation of the effective number of parties (N_S). This is due to the way in which N_S is calculated, whereby it is a weighted index in which each party share is weighted by itself, through squaring (see Chapter 4). Because of this calculation of N_S, once we have s_1, we have tight limits on what N_S can be. The derivation of the relationship is more involved than for the previous steps – see Taagepera (2007: 160–164). Nonetheless, an expected average relationship between N_S and s_1 is quite simple:

[6] This additional complication is set aside, but further research on its impact might be needed. For discussion, see Taagepera and Shugart (1993) and Taagepera (2007:135).

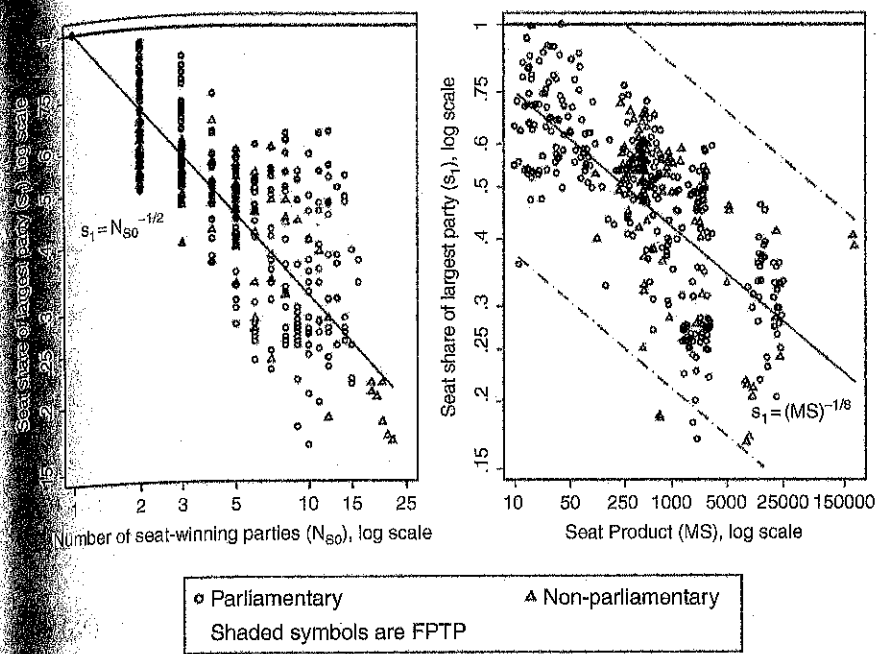

FIGURE 7.3 How the largest seat share (s_1) relates to the number of seat-winning parties (N_{S0}), left panel, and the seat product (*MS*), right panel

$$N_S = s_1^{-4/3}.$$ ← Check why

Figure 7.4 shows a strong fit of this expression; there is remarkably little scatter. Forbidden areas are marked and correspond to $N_S<1/s_1$ (the limit when N_{S0} parties are equal sized) and $N_S>1/s_1^2$ (a limit for the situation when all parties but the largest are infinitesimally small).[7] Now we can take the final step. Substituting $(MS)^{-1/8}$ for s_1 (cf. Equation 7.3) results in the equation we already identified as Equation 7.1:

$$N_S = [(MS)^{-1/8}]^{-4/3} = (MS)^{1/6}.$$

Conventional approaches (to be discussed at end of this chapter) are quite different. Their first step typically is to estimate the effective number of *vote-earning* parties (N_V) based on several inputs (including two that occur only in presidential systems). Then they take N_V as an input to a second equation in which the effective number of seat-winning parties (N_S) is the output variable,

[7] As explained by Taagepera (2007: 161–162), the actual limits for most values of s_1 are narrower still, leading to the average approximation, $N_S= s_1^{-4/3}$. Further details may be found in our online appendix, www.cambridge.org/votes_from_seats

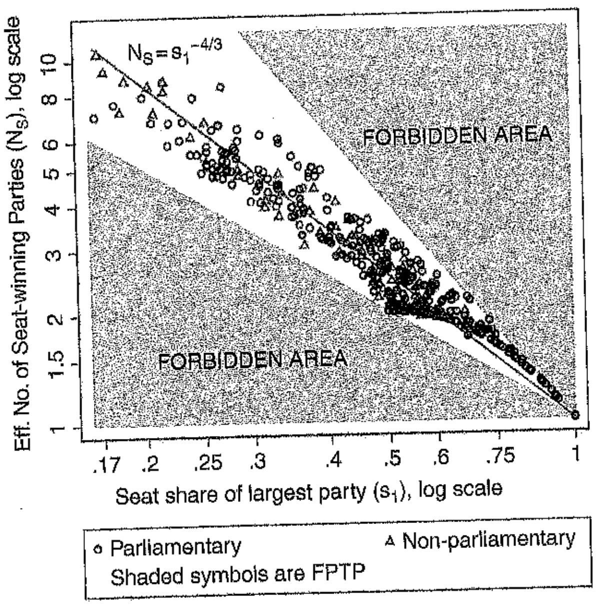

FIGURE 7.4 Relationship of the effective number of seat-winning parties (N_S) to the largest seat share (s_1)

and in which electoral-system variables are also included as further inputs shaping the votes–seats conversion.

By contrast, the SPM is based on a logic by which *votes come from seats*. The logical basis for doing so is what we demonstrated in this section – namely, that we can stipulate mandatory lower and upper bounds of seat-winning parties (ranging from 1 to M in a district and 1 to S in the assembly). From these logical starting points, we can deduce largest-party seat share (s_1) and effective number of seat-winning parties (N_S), as demonstrated here. The votes, on the other hand, are not directly constrained; that is why they do not enter into our models until Chapter 8. We now discuss one further implication of the SPM, and then devote the rest of the chapter to regression tests of the logic and comparison to the more conventional approaches.

Seat Product and the Duration of Government Cabinets

A major consequence of the effective number of assembly parties is how long government cabinets tend to last, if the executive type is parliamentary. With two major parties, single-party cabinets form and tend to last (upon re-election) around ten years. With five major parties, coalition cabinets are needed, and they tend to last less than two years.

A logical model (Taagepera 2007: 165–175) was finalized in Taagepera and Sikk (2010): C=42 years/N^2. It follows from $N=(MS)^{1/6}$ that C=42 years/$(MS)^{1/3}$. In most countries, the average cabinet duration over several decades is within a factor of two of this model (see graph in Taagepera 2007: 171).

This may be hard to believe, but the average cabinet duration is largely predetermined once a country chooses a parliamentary system with a given assembly size and district magnitude, through the intermediary of the number of parties. This is how deeply and predictably the consequences of electoral rules reach into politics. Important as this extension of the SPM is, it is outside the direct focus of the present book.

Such sets of mathematical expressions are not the typical style of the discipline, although such chains of simple bivariate expressions, one building on the other, are common in other sciences (Colomer 2007). We ask at this point that readers not object to the style, but judge it on its results. We have shown evidence in the form of graphs. Our next section carries out statistical testing of the models based on the Seat Product.

REGRESSION TESTS OF THE BASIC SPM

To test the Seat Product Model on our dataset, we first have to turn Equation 7.1 into a testable linear relationship. This is obtained by taking the decimal logarithms on both sides:

$$\log N_S = 0 + 0.167\log(MS).$$

When regressing $\log N_S$ on $\log(MS)$, we expect a simple straight line: $\log N_S=\alpha+\beta\log(MS)$. But not any such line will do; we further expect that α=0 and β=0.167. Note that the constant of zero is equivalent to (unlogged) N_S=1.00, which must be the case when MS=1. That is, if there were only a single seat filled in a given national election, then there could be only one seat-winning party – as is the case in direct presidential elections.

Before we test the final link in the logical chain, let us see if the steps upon which the chain depends are accurate: Is it true that the physical number of parties, N'_{S0}, at the district level, nationwide N_{S0} in the assembly, and the largest party's vote share (s_1) can all be predicted from the Seat Product as claimed in Equations 7.1–7.3? Using logarithmic transformation, we turn these equations into their respective testable linear forms:

$$\log N'_{S0} = 0 + 0.5\log M \quad \text{(District level)};$$
$$\log N_{S0} = 0 + 0.25\log(MS) \quad \text{(Nationallevel)};$$
$$\log s_1 = 0–0.125\log(MS);$$
$$\log N_S = 0 + 0.167\log(MS).$$

To test Equation 1.1 we use the Belden and Shugart (n.d.) district-level dataset. All other models in this chapter and in Chapters 8 and 9 require nationwide data, and thus use Li and Shugart (n.d.).

The regressions in this chapter are based on parliamentary systems, because these are the systems where the connection between features of the assembly electoral system and the party system is most direct. In such systems it is the balance of parties in the assembly that determines the national executive (hence the connection to cabinet durability, mentioned in the preceding section). We already saw in Figures 7.1–7.4 that presidential systems do not stand out as requiring a different approach. We also will discuss below, in our review of prior literature, reasons why presidential-specific variables should not be included in regressions that pool the executive types. A detailed consideration of presidential systems will await Chapters 11 and 12.

Testing the District-Level Expectation

Table 7.1 reports regression results on Equation 1.1, which, to our knowledge, had never been tested prior to Li and Shugart (2016) except on a very limited set of districts by Taagepera and Shugart (1993). The unit of analysis is the district-election, and we use Ordinary Least Squares (OLS) regression,[8] with observations clustered by country.

In Regression One[9] in Table 7.1, we see a strong confirmation of the logic of the district-level foundation to the Seat Product Model (Equation 7.2). This model predicts $\log N'_{S0}=0+0.5\log M$, and Table 7.1 shows an actual result of $\log N'_{S0}=0.00037+0.510\log M$. Regression Two restricts the test to districts that elect more than one member. We run this model because a reader might be reasonably skeptical that we have "stacked the deck" by including all of those single-seat districts that appear in Regression One (note differing sample sizes).

Does the model hold within proportional systems, or is the good fit in the full sample due to artificially anchoring the intercept at the logical $N'_{S0}=1$ when $M=1$? We see from Regression Two that the constant (0.06) is only slightly greater than the expected zero, even without the single-seat districts to force the regression to yield the mandatory result of $N'_{S0}=1$ when $M=1$. The coefficient, 0.458, is a bit low, compared to the expected 0.5. However, an F test shows that we are unable to reject the null hypothesis that the

[8] Because the outcome variable for Equation 1.1 is a count variable, one might propose using a tool such as Poisson regression. If we do so, the results are almost identical to the models reported here, as shown in the replication materials of Li and Shugart (2016).

[9] As noted in Chapter 1, we refrain from calling a statistical test a "model," contrary to most authors. In this book, we generally perform regressions to determine whether the statistical pattern supports a *logical model*, devised prior to running the regression.

TABLE 7.1 *District magnitude and the number of seat-winning parties*

	(1) Parl.	(2) Parl. M>1 only
log$M$	0.510	0.458
Expected: 0.5	(0.0243)	(0.0332)
Constant	0.000373	0.0607
Expected: 0.0	(0.000304)	(0.0239)
F test that coefficient on log$M$ = 0.5	0.693	0.23
Observations	11,654	453
R-squared	0.955	0.738

Robust standard errors in parentheses.
Dependent variable: actual number of seat-winning parties (N'_{S0}), logged

correct value is indeed 0.5.[10] Both regressions strongly support the district-level logic that, based only on boundary conditions and taking the geometric average "in the absence of further information," the number of seat-winning parties tends to be the square root of district magnitude.

We already saw the scatterplot of N'_{S0} and M, in Figure 1.2, in which it is visible that the data points cluster mostly around that figure's gray diagonal line, which corresponds to $N'_{S0}=M^{1/2}$. Although there is scatter, the more remarkable thing is just how closely the data cloud hews to Equation 1.1. This equation, *derived without the data*, using very sparse reasoning "in the absence of other information," turns out to describe the actual relationship exceedingly well. We carry out further testing of relationships at the district level in Chapter 10. Now we turn to tests of the nationwide model.

From Districts to National Effects

At the nationwide level, first we test Equation 7.2 for the actual number of seat-winning parties, then Equation 7.3 for the size of the largest party, and finally Equation 7.1 for the main dependent variable of interest, the effective number of seat-winning parties in the national assembly. Consistent with other regression tests of party systems in the literature, but in contrast to Taagepera (2007), our unit of observation for national-level tests is the individual election. We pool the observations and use OLS, with

[10] The equation based on Regression Two would be $N'_{S0}=1.150M^{0.458}$, which is a very minor deviation from the logical model, $N'_{S0}=M^{1/2}$ (Equation 1.1). When we include M=1 districts, Regression One yields $N'_{S0}=1.0009M^{0.510}$.

TABLE 7.2 *Nationwide effects of the Seat Product, parliamentary democracies*

	1 No. of parties (of any size) $\log(N_{S0})$	2 Seat share, largest party $\log(s_1)$	3 Effective No. of seat-winning parties $\log(N_S)$
MS, logged	**0.242**	**-0.1255**	**0.164**
	(0.0210)	(0.0135)	(0.0163)
Expected coeff.	**0.250**	**–0.125**	**0.167**
F test	0.712	0.960	0.873
Constant	0.0601	–0.00205	0.0181
	(0.0614)	(0.0274)	(0.0386)
Observations	265	298	298
R-squared	0.674	0.544	0.610

Robust standard errors in parentheses.

cluster-robust standard errors,[11] thereby following a methodological approach similar to that of the standard works on party-system fragmentation, including Clark and Golder (2006).

In Table 7.2 we see that all three nationwide models are supported.[12] In Regression One, we see that the coefficient on the log of the Seat Product (*MS*) is 0.242 (as compared to the expected 0.25) when the dependent variable is the log of the actual number of parties. In Regression Two, we see that the coefficient is –0.126 (as compared to the expected –0.125) for the log of the seat share of the largest party. Finally, for Regression Three we obtain a coefficient of 0.164 (as compared to the expected 0.167) in the estimated equation for the effective number of seat-winning parties (N_S).

In each case, these coefficients are almost precisely what the logical model predicts. *F* tests show that all of the reported coefficients are statistically indistinguishable from their logically expected values. Moreover, in all three regressions, the constants are indistinguishable from zero, as the logical models require. We have already seen the plots of the data for the relationship between *MS* and N_S (Figure 7.1), N_{S0} (Figure 7.2) and s_1 (Figure 7.3).

A final question concerns the electoral formula. Would considering the different PR allocation formulas (see Chapter 2), in addition to assembly size and district magnitude, affect the results? The precise answer depends on

[11] We define a cluster as a consistent set of electoral rules using the criteria of Lijphart (1994). The results are substantively the same if we simply use country as our cluster variable.

[12] These results differ trivially from those reported in Li and Shugart (2016), because we have excluded STV and SNTV elections (for reasons explained in Chapter 3), whereas Li and Shugart included them.

TABLE 7.3 *How assembly parties and seat product connect*

$N_{S0} = (MS)^{1/4}$		
$s_1 = (MS)^{-1/8}$	$s_1 = N_{S0}^{-1/2}$	
$N_S = (MS)^{1/6}$	$N_S = N_{S0}^{2/3}$	$N_S = s_1^{-4/3}$

a coding decision that is discussed in the appendix. The general answer is, effectively no. The PR formula has no systematic effect at the national level once we know the Seat Product.

Table 7.3 summarizes the relationships of N_{S0}, s_1 and N_S to *MS* and to each other. The one that has not previously been discussed, $N_S = N_{S0}^{2/3}$, follows from the others, through algebra.

BASELINE, NOT A THREAT

Some readers may feel threatened by the Seat Product Model. If this model worked perfectly, would it take politics out of politics? And where would that leave the political scientists? Purely statistical approaches do not pose such an apparent threat, as they only map the relationships produced *by* politics. In contrast, a logical model such as the SPM seems to impose itself *on* politics. Such fears are overblown. First, the SPM's inputs, *M* and *S*, are themselves products of past politics, and are occasionally modified by politics (although they generally remain pretty stable over time). Second, the outputs of the SPM do not freeze in the current politics but only hem them in to some degree. The values of R^2 in Table 7.2 are around 0.60. This means SPM accounts for about 60 percent of variation in the effective number of parties, leaving 40 percent to other factors, including current politics. Most important, the SPM says nothing about *which* parties will get seats. Do not worry, the SPM will not eat out politics!

Actually, the SPM puts political effects into a clearer perspective. It does so by supplying a comparison level – the effective number of parties that we would expect to materialize at a given *MS*, in the absence of any other information. The impact of politics could conceivably place all actual data points above the SPM curve – or place them all below this curve. Then we could tell that the impact of politics is the difference between the model-based expectation and the real-world effective number.

But it turns out even more interesting than that. The data points in Figure 7.1 straddle the curve. Indeed, the SPM line is close to the best statistical fit. What this intimates is that average politics produces average number of parties, given the constraints set by *M* and *S*. Now the difference between the actual and expected number of parties yields information on politics: if this difference is positive, there is something in the politics, society, current events, and history of this country that pushes the number of parties unusually high – and reverse for a negative difference. This is useful information. (During model testing, in contrast, such deviations are just an awkward nuisance.)

Of course, deviations from a statistical best-fit curve also supply a measure of where a country stands, compared to an average country. But there is a difference. Such an average depends on the sample of countries chosen, which can vary. In contrast, a baseline anchored in the seat product MS is stable. If we come across a new sample of elections in which it does not hold, the finding does not invalidate the model. Rather, it should prompt us to ask, what is it about this sample that results in deviation from the baseline?

HISTORICAL BACKGROUND: THE DUVERGERIAN AGENDA

The idea of predictable relationships between electoral systems and party-political consequences has long been around, yet has remained controversial. The study of electoral systems began with advocacy pieces for specific sets of rules, such as those written by Borda (cf. Colomer 2004: 30), Hare (1859), Mill (1861), and Droop (2012 [1869]). This tradition continued up to the mid-twentieth century. (For details, see Taagepera and Shugart 1989a: 47–50 or Colomer 2004.) A major analytical landmark was reached with Maurice Duverger's work (1951, 1954). He highlighted the possibility of predictable relationships between electoral systems and political outcomes. One broad idea underlies the line of inquiry that received a major boost from Duverger's work, although Duverger himself expressed it in a narrower form. When the electoral system is simple,

the average distribution of party sizes depends on the number of seats available.

In any given electoral district, the seats available are determined by the district's magnitude. Single-seat districts restrict the number of parties more than do multiseat districts. However, the total number of seats in the representative assembly matters, because more seats offer more room for variety. It is possible to have more than ten parties in a 500-seat assembly, but not in the ten-seat national assembly of St. Kitts and Nevis. At the same district magnitude, a larger assembly is likely to have more parties, all other factors being the same. The two size effects, district magnitude and assembly size, could in principle act separately, and one might be much stronger than the other. But it turns out, maybe surprisingly, that the logical derivation of Equation 7.2 makes them act through their product, on an equal footing.[13] This is the foundation of the Seat Product Model.

[13] It may seem that M and S have equal impacts on the number of parties, since they act through the product MS, but this is not so. The largest observed M (450) exceeds the smallest (one) 450-fold, while the largest observed S (around 650) exceeds the smallest (about ten) only sixty-five-fold. Thus M impacts MS more than does S.

But how did we reach this stage? This section traces the development of the Duvergerian approach since the mid-twentieth century. It highlights its achievements but also limitations.

Duverger's Propositions, Based on Mechanical and Psychological Effects

Duverger (1951, 1954) was the first to announce clearly what Riker (1982) later anointed as Duverger's "law" and "hypothesis," making a connection between electoral and party systems. Avoiding implications of unidirectional causality, they can be worded as follows:

1) Seat allocation by plurality in single-seat districts tends to go with two major parties ("law").
2) PR formulas in multiseat districts tend to go with more than two major parties ("hypothesis," because more exceptions were encountered).

Note that Duverger's propositions involve only one parameter, district magnitude.[14] They say nothing about the various complex rules that we discussed in Chapter 3 and will discuss again in Chapter 15.[15] Duverger also did not yet have the concept of *effective number of parties* (see Chapter 4), which was introduced by Laakso and Taagepera (1979).

Duverger's propositions imply a sharp break between FPTP and PR. Actually, as district magnitude increases from M=1 to M=S (nationwide single district), the number of parties tends to increase gradually and at a decreasing rate, as first shown graphically in Taagepera and Shugart (1989a: 144). In this light, the discontinuity between the "law" and "hypothesis" should be removed, leading to a single function $N=f(M)$ for the average pattern at a given S. The question remained whether the effective number of electoral parties (N_V) or legislative parties (N_S) should be used. Taagepera and Shugart (1989a: 144, 153) presented rough empirical best-fit equations for both N_V and N_S as a "generalized Duverger's rule." However, these equations no longer should be used, because subsequent work, starting with Taagepera and Shugart (1993) and continuing with Taagepera (2007), Li and Shugart (2016), and this book, show we can do better.

What produces the outcomes noted by Duverger? Low district magnitudes – with M=1 being the lowest possible – arguably put a squeeze on the number of parties in two ways. In any single-seat district with plurality rule, one of the two largest parties nationwide will win, unless a third party has a local concentration of votes quite greater than its nationwide degree of support. This is what Duverger referred to as the *mechanical effect*. Hence third-party votes most often are "wasted" (for the purpose of winning seats), so that these

[14] Rae (1967) was the first to use the term, district "magnitude," and to carry out systematic worldwide analysis of its effects.

[15] In a retrospective essay, Duverger (1986) included mention of the "two-round majority system."

parties are underpaid, nationwide. Correspondingly, the two largest parties will be overpaid in terms of seats. This effect is observed instantaneously, for any given election, once the seat and vote shares are compared. In this sense, it is "mechanical."

In contrast, what Duverger termed the *psychological effect* may develop more slowly, over several elections. The mechanical effect means that votes for third parties are effectively wasted in most districts of low magnitude (and M=1 in particular). In the next election, many voters may abandon such parties – a point already noted by Droop (2012 [1869])[16] – except in the few districts where the third party won or came close. Sometimes regional or ethnic differences allow parties that are small nationwide to persist because they can win many districts in their regional strongholds. At various times, a Bloc Quebecois has been important in Canadian national politics because it could win many districts in the province of Quebec (Massicotte 2018), and the United Kingdom has its Scottish National Party and smaller regional parties in Wales and Northern Ireland (Lundberg 2018). India has a profusion of such parties, as we showed in Chapter 5 and discuss in detail in Chapter 15.

In contrast to parties that draw on regional strongholds, a nationwide third party that wins few seats may see many of its voters bleed away at the next election, causing even further voters to give up on them. Such parties may tend gradually to be reduced to insignificance or even be eliminated, according to Duverger's argument.

The psychological effect is usually presented in terms of voter strategies, but it also works on politicians and contributors. Anticipating another defeat and lacking resources, a third party may desist from running in a district even before its former voters have a chance to abandon it. Financial contributors may be hard to find, and few people may volunteer to campaign for a lost cause.

Duverger's effects apply foremost at the district level.[17] This is where the seat is lost or won and where the votes are wasted or not, regardless of nationwide results. Voters have no direct reason to abandon a third party nationwide who won in their own district – or only narrowly lost and could win in the next election. The extension of the psychological effect to the nationwide scene need

[16] As reported by Riker (1982), Droop (2012 [1869]: 10) had called attention to these effects about eight decades before Duverger:

> an election is usually reduced to a contest between the two most popular candidates or sets of candidates. Even if other candidates go to the poll, the electors usually find out that their votes will be thrown away, unless given in favour of one or other of the parties between whom the election really lies.

[17] Yet, Duverger initially stipulated what came to be known as his "law" as the plurality electoral system tending to produce a two-party system *in the legislature* – see for example the graph in Duverger (1954: 209). Other statements in his rich treatise refer to votes or to individual districts, but his preoccupation was with the nationwide assembly party system. In this chapter, so is ours.

not follow. Sometimes it does, but by no means always. Third parties have minimal presence in the United States, but such parties have survived and even made a comeback in the United Kingdom. Significant national parties other than the top two persist in Canada, and not only (as often claimed) in specific regions (see Gaines 1999).

It is thus important to emphasize that the psychological effect on nonregional third parties is only a tendency. In fact, many smaller parties persist despite winning few seats and continue to enter, and receive votes in, districts that are hopeless. Such minor-party persistence is an anomaly to those wedded to the Duvergerian tendencies, with their near-exclusive focus on district magnitude to the exclusion of assembly size.[18]

Why "Duverger's Law" Does Not Qualify as Law – but Still Is a Useful Tendency

The observation that seat allocation by plurality in single-seat districts tends to go with two major parties has passed into political science literature as "Duverger's law." Yet it does not pass the test as law in the scientific sense. It is too vague, as we will see. Maurice Duverger himself would agree. In retrospect, Duverger (1986) claimed merely a tendency, saying that it's the American authors, especially Riker (1982), who have called it Duverger's *law*. So, instead of "Duverger's law" we should talk of Duverger's *tendency*, until law-like firmness is demonstrated. The question is, how strong is this tendency?

This tendency would have utter firmness if two complementary conditions were satisfied. First, if FPTP always led to two-party systems; and conversely, if all two-party systems originated from FPTP rule. In contrast, suppose only one-half of all FPTP elections led to two-party systems, and only one-half of all two-party systems originated from FPTP rule. This would mean complete randomness, and Duverger's tendency would be completely rejected. Where do we actually stand, between these two extremes? The hard reality is that such a test never has been carried out.

Indeed, the very setting up of such a test runs into problems. How do we recognize a "two-party system" in operational terms? While the effective number of parties serves well in quantifying the core idea of the Duvergerian approach, it would be a poor measure of "two-partyness" for the specific

[18] Given that a district is "embedded" in a nationwide assembly electoral system (that may have many dozens or hundreds of other districts) the FPTP logic means our models actually predict more than two parties receiving substantial vote shares even in M=1 districts, if the assembly is large. This can happen because some voters may actually think nationally even when votes are turned into seats only in districts (Johnston 2017). For instance, they may want to vote for a party that they know wins seats *in other districts*, even though it has no chance in the voter's own district. This point about the national impact on the district will be the theme of Chapter 10. For now the important point is that we see no sharp break between FPTP and PR, like standard Duvergerian works do.

purpose of testing the validity of "Duverger's law." This is so because N=2.00 can originate not only from 50-50-0, which expresses pure two-partyness, but also from constellations such as 70-(five parties at), 4-(five at) 2, which is pure one-party hegemony. Such N=2.00 could also come from 66.6–16.7–16.7, which combines features of one-party hegemony and a three-party constellation that do not meaningfully "average out" to two parties. Gaines and Taagepera (2013) offer better ways to distinguish two-party constellations from one-party and multiparty combinations, but problems remain.

By the way, are we talking about votes or seats? Is it *votes* parties receive in electoral districts, or *seats* they win in the national assembly? Duverger implied both, at different points in his original treatise. First, in each electoral district only two parties emerge to compete for votes. Then, by some quite fuzzy process (what Cox, 1997, termed "linkage"), these parties supposedly turn out to be the *same* two parties in each district, so that assembly seats go only to these two parties. OK, such a connection may exist. But we would have to test the presumed tendency at both levels – district votes and assembly seats.

As one surveys the literature on FPTP systems, several desirable outcomes are claimed for what would be an ideal Duvergerian two-party system. They go beyond just having two major parties. *First*, such an ideal system leads to a comfortable single-party majority, so that the government can act decisively. *Second*, it leads to a single vigorous opposition party that keeps the government on its toes and can take over after new elections. *Third*, FPTP rule favors regular alternation in power, so that neither major party becomes stale. *Fourth*, it even offers proportional representation of sorts in the long run, as the two major parties tend to win an equal number of elections. A test along these lines (Taagepera, 2015) brings mixed results. Indeed, the Australian Alternative Vote would seem more Duvergerian than FPTP (see Chapter 16).

In sum, we are left with a tendency that falls short of qualifying as law. *The SPM is a major step beyond the so-called Duverger's law*, because the latter offers no equation between operationally measurable quantities and hence can offer no quantitative predictions. The SPM does. But this in no ways reduces Duverger's enormous contribution to the field. He was the first one to announce clearly some basic tendencies that have guided much of the work in electoral systems ever since. This is the basis for the Duvergerian agenda, to which we come next.[19]

The Duvergerian Agenda

The "Duvergerian Agenda" (as termed by Shugart 2005a) refers to the scholarly work that builds on Duverger's tendencies regarding institutional effects on

[19] Remarkably, we could develop the Seat Product Model without explicit reference to Duvergerian tendencies and effects. However, they are implicit there. The SPM is very much a product of the Duvergerian agenda.

party system fragmentation. This agenda has evolved into a "mature" and active subfield for an ever-growing set of scholars despite having been considered "underdeveloped" just over three decades ago (Lijphart 1985).

The Duvergerian agenda consists of explaining and predicting the results and causes of Duverger's effects. It includes "micro" and "macro" dimensions. Micro considerations underlie the psychological effect and related strategic considerations and "coordination" (see Cox 1997). Our focus is principally on the macro perspective – the systemic relationship between institutional rules and party-system outcomes.[20] This macroscopic approach tries to make use of the restrictions imposed by electoral rules (low district magnitude and small assembly size, in particular) to explain and predict the number and relative sizes of parties, as well as the degree of disproportionality of seats to votes. In many works, some measure of social diversity or "issue dimensions" is taken into account (Taagepera and Grofman 1985, Ordeshook and Shvetsova 1994, Amorim Neto and Cox 1997, Cox 1997, Clark and Golder 2006, van de Wardt 2017, Moser, et al., 2018). The macro dimension of the Duvergerian agenda has been called the "core of the core" of electoral studies (Shugart 2005a).

For many decades, the understanding of the macro level of electoral-system effects was dominated by the idea that *seats come from votes*. This view, which has retained substantial currency, primarily sees the electoral system as a black box in between the votes and seats (cf. Taagepera and Shugart 1989a: 64 and 202). In one of our earlier collaborations (Taagepera and Shugart 1993), we observed instead that votes and the electoral system *both* affect seats, from opposite directions.

Due to the actual impact on seats from these two directions, the concept of the electoral system as an "intervening control box" between votes and seats is ***wrong***! Progress in furthering the Duvergerian agenda required specifying how the number of available seats constrains electoral outcomes, including the votes. ***Votes come from seats***, although that is not to say that they come *only* from seats. Current politics and deeper cultural and historical factors of the country are important, too, but these are country-specific rather than systematic cross-national factors like institutional variation.

More recent works, starting with Amorim Neto and Cox (1997) and Cox (1997) have explicitly recognized that the electoral system shapes the votes. However, their dominant theoretical approach remains one that sees the distribution of seats entirely as an output of the electoral system, *after* the votes are fed in (recall our depiction of this approach in Figure 1.3).

[20] We will not get into a philosophical debate on whether the macro level can be understood at all without solving the micro level. Let the results speak for themselves. Note that in physics thermodynamics was developed (and continues to serve) much before statistical physics explained its micro level underpinnings.

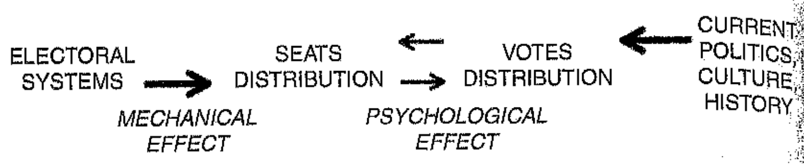

FIGURE 7.5 The opposite impacts of current politics and electoral systems
Source: Adapted from Taagepera (2007).

The more theoretically fruitful approach is the one that sees the seats distribution as affected by both the votes and the electoral system. We already depicted this in Figure 1.4, but we amplify it in Figure 7.5 with the addition of Duverger's notions of mechanical and psychological effects.[21]

For individual elections, votes come first, based on current politics and, more remotely, on the country's historical peculiarities. They will determine the seats, in conjunction with the mechanical effect of the electoral system. But for the average of many elections, and for the purpose of elaborating systematic cross-national models, we need to recognize that the causal arrow reverses direction. Through the mechanical effect, the electoral system pressures the distribution of seats to conform to what best fits in with the total number of seats available. Through the psychological effect, the electoral system eventually also impacts the distribution of votes. In the process, it might even counteract culture and history to some degree by rendering some parties nonviable.

The Duvergerian agenda continues to be salient because of the impact of party fragmentation on the wider pattern of governance, such as the balance between accountable governments and broadly representative political inclusion (Cheibub et al. 1996; Carey and Hix 2011; Lijphart 1999), and because political scientists continue to be commissioned for their advice on electoral-system design (Carey et al. 2013). Although there are still disagreements over how much fragmentation is optimal for a party system, no one will deny that we can hardly make any normative judgment on an electoral system without knowing its impact on the distribution of seats among parties.

SOME OTHER FACTORS OF PARTY-SYSTEM FRAGMENTATION

The Seat Product Model accounts for about 60 percent of the variation in the effective number of assembly parties (cf. Table 7.2). This leaves 40 percent to other causes and randomness. It would be welcome, if further factors could be located. Socioethnic diversity could be one, acting by itself or in conjunction with seat product, as could be presidential factors in the case of such regimes.

[21] For simplicity, it omits political culture. In Chapter 8, we will discuss how it can be brought back in.

Unfortunately, we will see here and in Chapters 11, 12, and 15 that such factors add little explanation and prediction ability, once the seat product has been accounted for.

Socioethnic Diversity

It is reasonable to suspect that the tendency of PR systems to be associated with multipartism is conditional on social cleavages – a point made by Ordeshook and Shvetsova (1994) and extended by Amorim Neto and Cox (1997). Several scholars, including Clark (2006), Clark and Golder (2006), and Hicken and Stoll (2012), have estimated models of these effects. While each study differs in important respects, a key conclusion of all of them is summarized by the statement in Clark and Golder (2006: 682) that "absent any knowledge concerning the social pressure for the multiplication of parties, it is not possible to predict whether multiple parties will actually form in permissive electoral systems."

In other words, they see no direct relationship between permissive (i.e., substantively proportional) electoral systems and fragmented party systems, and only when high social heterogeneity and permissive rules interact do they claim we will find fragmentation. In contrast to the Seat Product Model, these approaches did not include assembly size, using only district magnitude (mean or median) and the size of upper tiers (if any) of seat-allocation as measures of system permissiveness. The question of the impact of ethnic diversity is one that we turn to in Chapter 15. We will spoil the suspense a bit here, however, by noting that the reason we come back to it so late in the book is that ethnic diversity has only very limited impact on the predictions we derive from the Seat Product Model, which uses only institutional inputs.

Previous regression-based works on electoral systems and party systems broke important ground in furthering knowledge about statistical patterns via broad cross-national datasets. Unfortunately, they come up short in offering guides to real-world institutional design. For instance, Table 2 in Clark and Golder (2006: 698) shows that their most preferred regression model has highly unstable coefficients on key independent variables depending on the sample to which it is applied. This may not be a problem if all we are curious about is *which* factors matter. But if we were tasked with offering advice on the likely impact of a proposed electoral system on party-system fragmentation, our most honest answer, based on the existing statistical methods, would be "we can't say." One can only imagine how frustrating it would be if Newton's Laws of Motion showed only *what* can change an object's speed, but without giving a stable formula: most industrial products in our time would not have been possible.

Presidential Impact on Assembly Parties

When it comes to the effective number of assembly parties in presidential democracies, some works include as an input variable the effective number of presidential candidates as well as the temporal "proximity" of legislative and presidential elections (see our depiction in Figure 1.3).[22] On the one hand, this would imply that presidential systems are fundamentally different from parliamentary. On the other hand, it has forced some regression analyses to treat parliamentary systems essentially as if they were *special cases of presidential systems*, with the effective number of presidential candidates being zero.[23] The approach requires us to know the votes distribution in the presidential election in order to understand the number of parties in the assembly or the votes distribution for assembly elections. Yet, Figures 7.1 and 7.2 suggest that assembly party systems in presidential democracies can be explained by the same fundamental institutional input as those in parliamentary – the Seat Product.

The inclusion of variables specific to presidential systems in regressions that pool all democracies has already been called into question. Elgie et al. (2014) show that the conclusions of such models do not hold for presidential systems when the parliamentary systems are removed from the sample. Then, Li and Shugart (2016) showed that the widely accepted interactive effect of social and institutional factors also does not hold for parliamentary systems, when they attempted to replicate Clark and Golder's (2006) regressions.[24] If the conclusions about either executive type hold only when the other is included, along with variables specific only to presidential systems, then the overall conclusions themselves might be due for a serious rethink. Fortunately, once we conceptualize the electoral system via the Seat Product, we do not need to separate the samples by executive format, or incorporate presidential-specific variables, in order to make sense of the statistical patterns.

CONCLUSION

In this chapter, we show the value of the Seat Product Model for predicting the effective number of seat-winning parties in the national assembly (N_S). For simple systems, N_S tends to be around the sixth root of the product of mean district magnitude and assembly size. This formula was introduced by Taagepera (2007). Here we extended its applicability by testing it on a much wider set of cases; in Chapter 15 we will extend it farther to include complex two-tier systems and the role of ethnic diversity (see also Li and Shugart, 2016).

[22] These works include the following: Amorin, Neto, and Cox 1997, Cox 1997, Clark and Golder 2006, Hicken and Stoll 2012. We return to them (and others) again in Chapter 12.

[23] Even though an "effective" number, by definition, cannot be less than one.

[24] More specifically, the finding of Clark and Golder (and others using similar approaches) is not robust to the exclusion of India, by far the case with the highest ethnic diversity. We discuss the Indian case further in Chapter 15. By contrast, the SPM is statistically robust to whether we include or exclude India. See Li and Shugart (2016).

A key message we can deliver at the conclusion of this chapter is that institutional theories should no longer be considered just a thought game played in the academic community that have failed to produce robust expectations. Rather, the Seat Product Model can have genuine real-world impact when applied to the institutional-design process in newer democracies. This is a valuable contribution, because political scientists are often called to advise on electoral system design in emerging democracies (Carey et al. 2013).

Of course, just predicting party fragmentation in the legislature cannot satisfy us; we want to know if the Seat Product can lead us to a prediction of electoral fragmentation. This is the task of Chapter 8. There we offer a completely novel theory of how the SPM applies to nationwide elective party systems.

Appendix to Chapter 7

THE IMPACT OF ALLOCATION FORMULA

Does the precise allocation formula of a PR electoral system (see Chapter 2) affect the relationship of the seat product, *MS*, to the output quantities of interest in this chapter? We can test this by calculating ratios of the actual output to the value expected under a simple system, as follows:

$$N_{S0}/(MS)^{1/4};$$
$$s_1/(MS)^{-1/8};$$
$$N_S/(MS)^{1/6}.$$

We then perform difference-of-means tests on these ratios for PR systems, depending on whether they use D'Hondt or another PR formula. In all cases but two, the allocation formula, if not D'Hondt, is either Hare quota with largest remainders, or Modified Ste.-Laguë. The exceptions are Brazil and Finland, which require further explanation.

As we demonstrated in Chapter 6, many lists in Brazil and Finland contain candidates of more than one party in alliance. Under the rules used in these countries the parties *within the list* win their seats as if the formula were SNTV. That is, each list wins some number of seats, which we can designate *s*, via the application of D'Hondt, then the top *s* candidates on the list are elected, without regard to the party affiliations of the candidates on the list.

Depending on how many different parties have candidates in a given list's top *s* vote totals, the number of parties winning may be two or more per list. Thus the provision for alliances in these countries potentially inflates the number of parties. We go into the impact of these provisions in more detail in Chapter 14, which focuses on intralist allocations at the district level.

TABLE 7.A1 *Impact of formula on ratios of actual values to Seat Product predictions*

	If Brazil and Finland are considered to be D'Hondt		If Brazil and Finland are not considered to be D'Hondt	
Ratio	D'Hondt	Other PR formula	D'Hondt	Other PR formula
$N_{S0}/(MS)^{1/4}$	1.156	1.014	1.119	1.109
$s_1/(MS)^{-1/8}$	0.984	1.086	*1.036*	*0.978*
$N_S/(MS)^{1/6}$	*1.096*	*1.072*	***1.022***	***1.190***

Statistically significant differences (at $p<0.05$) are in bold. Those in the expected direction are in italics.

In the sense that they effectively use D'Hondt for lists and then switch to SNTV within lists, Brazil and Finland do not actually have a "pure" D'Hondt system. The way we code these cases affects our answer to our question of whether the PR allocation formula matters.

Table 7.A1 summarizes the results, which contain some surprises. We expect D'Hondt to lead to lower $N_{S0}/(MS)^{1/4}$ and $N_S/(MS)^{1/6}$ but higher $s_1/(MS)^{-1/8}$. Regardless of our coding of Brazil and Finland, the effect on the nationwide number of parties winning at least one seat, N_{S0}, is opposite of expectation. There is no reason why, in any given district, D'Hondt (without alliances) would permit more parties to win seats than other PR formulas. The result must be due to quirks of cross-district politics.[25]

If, however, we recode the "impure" D'Hondt formulas in Brazil and Finland, we see that the effects on the nationwide effective number of parties, N_S, or on the seat share of the largest party, s_1, are as expected. The difference for N_S is also statistically significant (so is the difference for s_1 if we adopt a $p<0.065$ standard). That is, D'Hondt, without alliances, has a significantly lower N_S and higher s_1 than other PR formulas.

Because it depends on how we treat these two unusual cases, we do not consider the findings on the impact of formula to be robust. We are on safer grounds estimating these quantities from the Seat Product, and considering deviations from expectation as matters of individual country or election politics. One of those country-level (or even election-specific) variations is whether alliances are allowed, and how many lists elect candidates of more than one party. We return to this question in more detail in Chapter 14.

[25] When Brazil and Finland are not considered D'Hondt, the difference between D'Hondt and other PR formulas is insignificant if Spain is excluded. Indeed, Spain's districted PR system features a plethora of regional parties, each of which wins seats in one or a few districts only (as detailed in Hopkin 2005).

8

Winners Plus One: How We Get Votes from Seats

Now we literally deduce votes from seats, in line with the book's title. We take a completely novel step and extend the Seat Product Model to votes distribution. This transition was hard to come by. For instance, the chapter on "The institutional impact on votes" in Taagepera (2007) is in retrospect a total flop, and the approach to estimating votes fragmentation taken by Li and Shugart (2016) lacked a theoretical foundation. Yet the additional assumption needed will be seen to be of utmost simplicity.

When a given number of parties wins seats in a representative assembly, how many more are likely to try their luck? How are these two numbers connected logically? If a logical model can be proposed, does it agree with the empirical average pattern? We can expect, of course, that the actual worldwide data will show variation over a wide range. In fact, we can expect any data on vote-earning parties to show more variation than comparable data on seat-winners, for the same reason that we argue for starting the analysis with the seats and only then extending votes: the seats are ultimately constrained by the institutions, whereas the votes actually are not. Whatever total number of seats comprise the assembly, we can be certain that there are no more parties than this number that gain representation. But how many others may run, and earn votes? The only constraint is that obviously there could not be more parties earning votes than there are voters, but that observation is hardly helpful. The extent to which the number of parties earning votes is greater than that winning seats is softly constrained, and only by the willingness of voters to tolerate "wasted votes" and the elites who form parties to keep running and losing.

Despite the expectation of wide variation in the number of electoral parties, the average value for a given country or a given institutional setup can offer a useful benchmark. When establishing a logically expected number of parties competing, we implicitly also ask: *Given the number of parties that won seats, was the number of parties competing for seats in a particular country unusually high or low, or as expected?* If there is marked deviation from expectation, then one may wish to look for the reason. In the absence of such a benchmark, anything goes and nothing gets explained.