1 Introduction

What can we learn from a regression model? This is the question that motivates *Regression Inside Out* (RIO). Regression is among the most widely used tools for data analysis. A workhorse of both academic and nonacademic research, it is a standard part of curricula in fields ranging from accounting to zoology. It can be used to explore data, test hypotheses, and bring statistical theory, discipline-specific theory, and data into dialogue (Belsley, Kuh, and Welsch 2004). A technical answer to the question, "what can we learn from a regression model?," is that we learn how average values of an outcome vary across subpopulations of observations that are defined by the values of a set of predictors (Cook and Weisberg 1999; Gelman, Hill, and Vehtari 2020). In a more practical sense, we routinely use regression models to learn a great deal about the world around us.

Over the past several decades, however, a growing chorus of scholars in the social sciences have raised concerns about *how* we learn from regression models. An early voice in this chorus was the pioneering quantitative sociologist Otis Dudley Duncan, who described a "syndrome" that he referred to as *statisticism*. Statisticism is

> the notion that computing is synonymous with doing research, the naïve faith that statistics is a complete or sufficient basis for scientific methodology, the superstition that statistical formulas exist for evaluating such things as the relative merits of different substantive theories or the 'importance' of the causes of a 'dependent variable'. (Duncan 1984: 226)

To be clear, Duncan was not against regression analysis (as is evident from his own scholarship; see Goodman 2007). Rather, he was deeply concerned with how regression was used to conduct research and draw conclusions. This concern with how we learn from regression models has only grown in the decades since Duncan's prescient diagnosis (e.g., Abbott 1988; Berk 2004; Emirbayer 1997; Freedman 1991; Ragin 2006; Shalev 2007; Tong 2019).

For some observers, concerns about how we learn from regression are grounded in theory (Abbott 1988; Emirbayer 1997; Ragin 2000, 2006;

Shalev 2007). Regression imposes "homogenizing assumptions" (Ragin 2000: 5) that lead researchers to construe social reality in terms of what Abbott (1988) refers to as a *general linear reality*. Abbott explains that, far from simply summarizing how average values of an outcome vary across subpopulations of observations (as textbooks teach it), regression shapes how we think about, interpret, and understand the social world (Abbott 1988: 169). Ragin (2000) similarly observes that conventional approaches to quantitative research structure how analysts make sense of populations, cases, and causes in ways that constrain dialogue between theory and evidence, thereby limiting discovery. Taken together, these and other theoretically motivated critiques highlight how conventional regression analysis is not a neutral representation of empirical realities. Rather, regression imposes strong philosophical assumptions that guide how we think about the phenomena we study.

For others, concerns about how we learn from regression are grounded in practice (e.g., Berk 2004; Freedman 1991; Tong 2019). While issues with how practitioners use regression are quite varied, they tend to focus on a widespread overemphasis on model outputs (e.g., fitted values, coefficients, and variance) with insufficient attention to model inputs (e.g., data and "the ordering of data in time, space, or other characteristics" [Belsley et al. 2004: 4]). Observers point out that much of the focus of standard regression theory has to do with sampling fluctuations, and regression is routinely used to analyze data that fails to meet the assumptions that render these theories applicable (e.g., a representative sample from a known population). Consequently, conclusions are drawn from regression that are not supported by the data or by the method itself (see Berk 2007). Elaborating on how inattention to model inputs can create significant methodological problems, Tong (2019) argues that "formal, probability-based statistical inference should play no role in most scientific research" (p. 246). He makes the case that data are routinely used to guide model specification, and when they are, the inferential statistics that quantify uncertainty are biased. Other advocates of regression analysis who are concerned with how it is applied similarly cite routine failures to meaningfully consider how model inputs shape what we learn from regression. As Belsley, Kuh, and Welsch (2004) observe in their influential book on regression diagnostics,

> In years past, when multivariate research was conducted on small models using desk calculators and scatter diagrams, unusual data points and some obvious forms of collinearity could often be detected in the process of 'handling the data,' in what was surely an informal procedure. With the introduction of high-speed computers and frequent use of large-scale models, however, the researcher has become ever more

detached from intimate knowledge of [their] data. It is increasingly the case that the data employed in regression analysis, and on which the results are conditioned, are given only the most cursory examination for their suitability. (p. 4)

They make the case that thorough engagement with data and other model inputs is essential to good statistical practice.

To be clear, scholars voicing both theoretical and practical concerns with regression analysis recognize its value as a "formidable and effective method" (Abbott 1988: 169). Nonetheless, they emphasize the need for a more careful consideration of what we can actually learn from a regression model. They ask us to confront difficult questions, such as: How do the philosophical assumptions that undergird regression shape our understanding of the social world? To what extent do summaries of the relationships among variables apply to and inform our understanding of the specific observations in our data? How do we reconcile an intuitive understanding of causality as multiple and complex with an analytic focus on net effects? Or, more simply: What can we actually learn from a regression model?

Our goal in this book is to expand *what* we can learn from regression models by fundamentally rethinking *how* we learn from regression models. We do this by turning regression "inside out." As we elaborate in Chapters 2, 5, and 7, conventional regression analysis renders the cases, their relationships to one another, and their unique characteristics – all of which are key model inputs (Belsley et al. 2004) – invisible (Shalev 2007). By contrast, RIO makes the complexities of the cases (i.e., the rows of the data matrix) visible and puts them in dialogue with the variables. While RIO begins with a generalized linear model (GLM), it allows us to identify each individual observation's additive contribution to model outputs. Because the contributions are additive, we can move seamlessly between individual cases and the net effects produced by the GLM, comparing how individual cases or sets of cases shape the net effect. This clearly situates each case within the broader context represented by the overall model space. As we show throughout this book, this ability has both theoretical and methodological payoffs.

1.1 A Case-Oriented Approach to Regression Models

RIO is designed to allow us to look inside the regression model and gain a deeper understanding of how it represents the data. In doing so, RIO allows us to relax many of the restrictive philosophical assumptions embedded in regression analysis (Abbott 1988), which constrain dialogue between theory

and evidence. This relaxing of assumptions is accomplished by analytically allowing for the complexities that emerge when observations are conceptualized as cases: spatially and temporally delimited phenomena of theoretical interest in their own right (Gerring 2017).

As noted earlier, the thrust of standard regression theory focuses on sample fluctuation. Regression is designed to identify population-level trends based on a representative subset of that population (Berk 2004). When analyzing a sample that meets the baseline assumptions for inference in a regression model, individual observations are entirely interchangeable, or more technically, they are exchangeable (Kutner, Nachtsheim, and Neter 2004). For example, if we are interested in assessing how attitudes toward gerrymandering affect the probability that voters in the United States will elect a Democrat or Republican president, it makes no difference whether we (the three authors) or you (the reader) are personally included in the sample, how we feel, or what we prefer. What matters is how the data were sampled. If the data were sampled properly, we can summarize how the conditional distribution of voter preferences varies based on attitudes toward gerrymandering and use those summaries to draw conclusions about the relationship between these variables in the population as a whole.

Yet, regression is often applied in contexts where the goal is to draw conclusions about a given set of cases rather than a population based on a representative sample. Consider, for example, almost all analyses where countries are the unit of analysis. We might use regression to examine the effects of a nation's regime type on interstate war (e.g., Schultz 1999), the determinants of status in the international system (e.g., Bezerra et al. 2015), or the effects of regional integration on poverty (Beckfield 2006). However, the observations in these analyses are not exchangeable. Consequently, it would be statistically and substantively untenable to take a representative sample of countries and use that sample to make inferences about all countries in the world. Moreover, unlike when we are drawing conclusions about population-level trends, the results of an analysis of countries are meaningful only to the extent that they can be related to tangible outcomes for real cases. Particularly in comparative international research, it is (perhaps surprisingly) common to find zones in a distribution with no observed data (Rosenberg, Knuppe, and Braumoeller 2017). Consequently, the results of a linear regression might produce values that have no observable empirical basis. These same limits and considerations are equally applicable to other, smaller units of analysis. A focus on cases can be found in regression analyses with observations ranging from individuals (Ragin and Fiss 2017) to corporations (McKendall and Wagner 1997), and beyond.

Cases necessarily introduce complexities. When observations are theoretically and substantively exchangeable, differences from one case to the next are simply factored into the error term and are substantively irrelevant beyond any possible trends in the error term. However, when observations are theoretically or substantively meaningful, differences from one case to the next can imply substantively or theoretically important differences in associations between variables, distinct causes of an outcome, or any number of other forms of complexity (Abbott 1988; Mahoney and Goertz 2006; Ragin 2006, 2014b). Consider, for example, an analysis of the relationship between social capital and school achievement. A regression model can show that social capital obtained through families improves children's academic achievement more than social capital obtained through school (Dufur, Parcel, and Troutman 2013). However, attention to case-level variation may reveal that different sources of social capital have different impacts from one student to the next, or that social capital does not matter at all for some students but matters a great deal for others. Similarly, a regression model might show that organizations engaged in political violence are more likely to participate in illicit drug economies when they control territory (Asal, Rethemeyer, and Schoon 2019; Cornell and Jonsson 2014). However, attention to case-level variation may reveal that the reasons groups participate in illicit drug economies are not simply the inverse of the reasons that they do not, such that participation is driven by economic need while lack of participation is driven by an absence of opportunity. These possibilities are rendered invisible (or, theoretically impossible) within the bounds of the general linear reality (Abbott 1988; Rambotti and Breiger 2020; Shalev 2007). However, by accounting for how individual cases contribute to net effects, turning our regression model inside out allows us to explore such discontinuities at the level of cases and subsets of cases. Thus, rather than rejecting "general linear reality," our aim in this book is to show how to get more out of it.

Shalev (2007) offers a similar example in research on comparative social policy. He notes that a well-established finding in comparative welfare state research is that there are two subtypes of European welfare states that are known to spend a great deal: Social Democracies and Christian Democracies (see Kersbergen 2003; Korpi and Shalev 1980). Discussing the effects of regime types on spending, he writes,

> [T]his presents no problem for the standard additive regression model provided that the effects are equivalent and unrelated – if for instance a strong social-democratic party could be expected to have the same effect whether or not it governed in coalition with a Christian-Democratic party. However, the Austrian experience suggests

that this is unlikely since historically, the black half of the 'red-black' [Christian Democratic/Social Democratic] coalition severely constrained its welfare state development. (Shalev 2007: 265)

Here, the unique features of Austria suggest an interactive effect, but it is not clear whether such an interaction would be statistically significant in a net effect model given the unique features of the Austrian case.

It is RIO's ability to account for these kinds of complexities at the level of cases that allows us to relax many of the restrictive assumptions built into the GLM. Most straightforwardly, it allows us to avoid the homogenization of cases that is inherent in conventional regression (Ragin 2000). More broadly, however, this relaxing of assumptions extends further. Consider Abbott's (1988) elaboration of the philosophical assumptions that are embedded in GLMs. As he notes, when we use regression models to represent social reality, we are required to transpose social life onto the algebra of regression models. He continues:

Such representational use assumes that the social world consists of fixed entities (the units of analysis) that have attributes (the variables). These attributes interact, in causal or actual time, to create outcomes, themselves measurable as attributes of the fixed entities. The variable attributes have only one causal meaning (one pattern of effects) in a given study, although of course different studies make similar attributes mean different things. An attribute's causal meaning cannot depend on the entity's location in the attribute space (its context), since the linear transformation is the same throughout that space. For similar reasons, the past path of an entity through the attribute space (its history) can have no influence on its future path, nor can the causal importance of an attribute change from one entity to the next. All must obey the same transformations. (p. 170)

Abbott notes that some methods – such as demographic methods, sequence analysis, and network analysis – relax these basic assumptions of general linear reality. Demographic models, for instance, relax the assumption of fixed entities with variable attributes by allowing entities to move, appear, disappear, merge, or divide over time. Sequence analysis, in contrast, relaxes nearly all the assumptions, while network analysis relaxes assumptions of independence (both independence among observations and independence from context).

RIO's grounding in the GLM binds it to the assumption of fixed entities with variable attributes. However, its orientation toward cases allows us to relax the remaining assumptions. By focusing on how individual cases shape the linear model, we can account for discontinuities in the meanings of

particular variables. These effects depend on an entity's location in the attribute space, which we can see by breaking down and mapping that attribute space (see Chapters 2, 3, 7, and 9), situating individual cases in relation to one another and to the variables. Moreover, a focus on cases allows us to account for the past path of an entity through the attribute space and explore how that history influences its future path forward. This can be done either by incorporating substantive knowledge into our interpretation of the location of cases in the attribute space, or by mapping the trajectory of individual cases across the model space over time. We illustrate this latter possibility in Chapter 7 (see Figure 7.12c), showing how a break in the plotted trajectory of an individual case across the attribute space corresponds with a major historical event that shifted its relationship to the variables in the model. Finally, a focus on cases allows us to explore and account for the possibility that the causal importance of an attribute changes from one entity to the next.

In short, by turning regression models inside out, we are able to get more out of the summaries of conditional distributions that are represented by conventional model outputs and engage with the complexity that often undergirds the social realities that regression represents. RIO is still firmly grounded in regression and statistical thinking (Tong 2019). Yet, by shifting how we learn from a regression model – turning attention toward the empirical relationships among cases rather than limiting our focus to the relationships among variables – we can dramatically expand what we are able to learn from a regression model.

1.2 A Methodological Gateway

By allowing us to relax many of the philosophical assumptions of conventional regression analysis, RIO opens the door to incorporating insights from, and contributing insights to, methodologies that operate under quite different philosophical (i.e., conceptual and epistemological) assumptions. Because of regression's ubiquity in the social sciences, it is routinely used as a benchmark when enumerating defining features of other seemingly disparate methodological approaches. Distinctions between qualitative and quantitative methods in the social sciences typically associate quantitative approaches with the logic of regression and contrast this logic with qualitative approaches that are case-oriented and highly sensitive to the influence of individual observations (see, e.g., Mahoney and Goertz 2006). Similarly, in *The Comparative Method*, Charles Ragin's foundational introduction of

qualitative comparative analysis (QCA), Ragin often contrasts QCA with regression to emphasize key elements of the comparative approach (see also Ragin 2000, 2006, 2009; Ragin and Fiss 2017). In his *Manifesto for a Relational Sociology,* Emirbayer (1997) contrasts relational approaches to what he refers to as substantialist approaches. He identifies methods common to these approaches, situating regression among the substantialist approaches in contrast to methods of network analysis.

In these and other instances, comparisons between regression and other methods often imply technical differences along with conceptual differences. However, the technical differences are often not as great as they appear at first blush. Breiger (2000) illustrates this through his comparison of correspondence analysis and lattice analysis (two fundamentally relational methods) with the quantitative approach developed by James Coleman (1994) in his *Foundations of Social Theory* (a fundamentally substantialist approach). He shows that there is "a remarkable homology – at the level of formal practices, if not indeed in their 'very spirit' – between the mathematical techniques" (p. 95). Subsequent research (e.g., Breiger 2009; Breiger and Melamed 2014; Breiger et al. 2011, 2014; Pattison and Breiger 2002; Rambotti and Breiger 2020) shows how the mathematical techniques associated with network analysis, configurational comparative analysis, and regression all share similar homologies.

Similarities in the formal practices undergirding these methods provide an opportunity to bring them into dialogue and highlight how the barriers that have motivated many to draw distinctions between regression and other analytic tools are more philosophical than methodological. As we illustrate in Chapters 8 and 9, the fact that RIO allows us to relax many of the assumptions of regression provides us with an opportunity to incorporate other philosophies and assumptions into our thinking as we apply and interpret regression models.

In addition to expanding how we interpret and engage with regression, the mathematical homologies between regression and other methodologies also stand to enhance multimethod research that incorporates regression analysis. The standard design for multimethod research relies on triangulation, which involves asking the same question using different methods and comparing the findings of each (e.g., Jick 1979; Tarrow 1995). However, as Seawright (2016) argues, there are no standards for drawing conclusions when two methods yield conflicting answers. He thus advocates for an integration-oriented approach. Rather than using each method to validate the other, he recommends bringing the two into conversation so that each method enhances the other. RIO stands to contribute to such efforts, offering a way of bringing regression into closer dialogue with many of the (typically case-oriented) methodologies commonly

employed to complement regression in multimethod research. Whether this means assessing how case studies fit in relation to an overall regression model (Chapter 7), bridging the gap between set-theoretic and correlational approaches to analysis (Chapter 8), or incorporating insights from field theory (Chapter 9), looking inside a regression model allows us to better assess how results from other methods (which are typically assumed to be disparate) are situated in relation to the results of a regression.

1.3 Understanding versus Improving Models

In the literature on regression models, cases are discussed mostly in the context of regression diagnostics. In that context, the aim is to identify a small number of cases that do not fit the model, and therefore imply a different model (with different cases). While the analytic framework of RIO builds on known methods for regression diagnostics, RIO's intended purpose is quite distinct from their typical goals, which are oriented toward improving regression analysis by formulating new models. As the preceding discussion indicates, the intended purpose of RIO is to provide new ways of learning from – interpreting, engaging with, and thinking via – regression models. We highlight this distinction because it provides a necessary orientation for readers moving forward.

While we view diagnostics as a critical step in any regression analysis, the question of how to fit a better model is quite distinct from the question of how to interpret and understand the model at hand. Over the past decade, standard textbooks on statistical methods have increasingly incorporated thorough discussions of regression diagnostics (e.g., Gelman et al. 2020), and there are many excellent texts devoted entirely to developing and/or explaining methods for improving model fit, detecting collinearities, correcting for biases, and many other necessary tasks for estimating an analytically robust regression model (e.g., Belsley et al. 2004; Berry 1993; Fox 2019; Pregibon 1981; Velleman and Welsch 1981). Despite important and exciting innovations in regression diagnostics specifically, and statistical modeling more generally, making sense of regression outputs is generally treated as well-trod ground and left to introductory texts.

Because RIO builds on the GLM, we assume that any user will have already fit a model, and it is our hope that this will have been done in dialogue with appropriate tests and assessments to ensure that the model itself is the best possible representation of the data. As we show in Chapter 6, turning a

regression model inside out may lead an analyst to respecify their model. However, the value added by turning a regression model inside out extends farther than identifying a better model. Put differently, the aim of regression diagnostics is to learn about problematic cases, while the aim of RIO is to learn more about how the cases and the variables co-constitute the regression output (i.e., the coefficients and standard errors).

Keeping this distinction in mind will help to situate some of the facets of RIO that we discuss. For example, if we identify a single observation as being highly influential using conventional diagnostics like Cook's distance (Cook 1977) or DFBETA (Belsley et al. 2004), RIO will likely identify that observation as having a large additive contribution to one or more regression coefficients, and/or to the variance. However, simply having a large additive contribution to one or more regression coefficients or the variance does not imply that dropping that case will meaningfully alter our model, as it typically does when such cases are identified using conventional diagnostics. The reason for this is, with RIO, each case's contribution is based on the given model. If we drop one case or alter a variable, the model itself changes, and so do the relationships. Thus, RIO does exactly what is advertised: it allows us to look at what is going on inside our given regression model. It is worth noting that understanding what is going on inside our regression model may lead us to revise the model, but that is not our primary goal.

Despite these differences, throughout this book, we often compare RIO with methods employed for the purposes of regression diagnostics. This is because diagnostics is the only area of conventional regression analysis where individual observations are taken seriously. Given that our focus is on cases (which are typically conceptualized as observations in the data matrix, but can be represented by multiple observations, as we show in Chapters 5 and 7), the methods used in regression diagnostics provide a useful counterpoint for us to illustrate how a case-oriented approach to regression contrasts with the treatment of cases in conventional, variable-oriented regression, where individual cases (or sets of cases) are only considered to the extent that they risk violating assumptions used to draw conclusions about the relationships among variables.

1.4 Plan of the Book

As sociologists, we are writing from the perspective of the social sciences. The examples that we use as illustrations throughout the book are all drawn from the social sciences (specifically from sociology and political science), as are many