Political Analysis (2003) 11:255–274
DOI: 10.1093/pan/mpg014

Selection Bias in Comparative Research: The Case of Incomplete Data Sets

Simon Hug
Institut für Politikwissenschaft, Universität St. Gallen,
Dufourstrasse 45, 9000 St. Gallen, Switzerland
e-mail: simon.hug@unisg.ch

Selection bias is an important but often neglected problem in comparative research. While comparative case studies pay some attention to this problem, this is less the case in broader cross-national studies, where this problem may appear through the way the data used are generated. The article discusses three examples: studies of the success of newly formed political parties, research on protest events, and recent work on ethnic conflict. In all cases the data at hand are likely to be afflicted by selection bias. Failing to take into consideration this problem leads to serious biases in the estimation of simple relationships. Empirical examples illustrate a possible solution (a variation of a Tobit model) to the problems in these cases. The article also discusses results of Monte Carlo simulations, illustrating under what conditions the proposed estimation procedures lead to improved results.

1 Introduction

Over the last few years comparative politics has had to swallow some healthy methodological medicine. The symptoms indicated a serious illness of selection bias. Researchers afflicted by this illness were seen choosing their cases as a function of the phenomenon they attempted to explain. While the medicine administrated by, among others, Geddes (1991) and King et al. (1994) shows some positive effects, some ill-justified references to "most similar cases" or "most different cases" designs still appear passingly in the literature.[1]

Author's note: Part of this article draws on coauthored work with Dominique Wisler. I wish to thank him for letting me use some material here. An earlier version of the article was prepared for presentation at the Annual Meeting of the Midwest Political Science Association in Chicago, April 23–25, 1998. Greatly appreciated discussions with Chris Achen, Lars-Erik Cederman, and Thomas Christin on several aspects of the article and helpful remarks by seminar participants at UCLA and UCSD and anonymous reviewers have improved it considerably. Thus, I am quite willing to accept the responsibility for all remaining errors and misinterpretations. Partial financial support by the Swiss National Science Foundation (Grant No. 8210-046545) is gratefully acknowledged. The data used in this article, as well as additional material, is available on the *Political Analysis* Web site.

[1] Recent work on testing necessary conditions emphasizes that selection on the dependent variable can be a useful research strategy (e.g., Dion 1998; Goertz and Braumoeller 2000). The question arises, however, whether scholars are confident enough to state their theoretical results in terms of necessary conditions and can give specific conditions under which these conditions are to be rejected. To avoid this debate I will focus on empirical tests of sufficient conditions in this article.

While this illness related to the selection of cases appears to be close to eradication, another variant of selection bias has hardly been addressed in comparative politics. This other variant appears frequently in cross-national studies when the data sets at hand are themselves afflicted by a selection process. A simple example may help to illustrate this problem. Studies of ethnic violence or discrimination against minorities have to rely on some identification of ethnic groups or "minorities at risk" (Gurr 1993). But how is it possible to identify the relevant groups and minorities? In a comparative framework (and even at the level of a case study) it is simply impossible to determine the complete set of relevant groups. Instead, researchers (e.g. Gurr 1993; Gurr and Moore 1997) rely frequently on some previous discrimination against a group or some previous mobilization by a minority. Obviously, the set of ethnic groups and "minorities at risk" then becomes an incomplete sample of all possibly relevant groups. But researchers interested in finding explanations for the eruption of ethnic violence should have at hand a data set covering all types of ethnic groups and minorities. Relying on their studies on incomplete samples, researchers expose themselves potentially to serious symptoms of selection bias.

The aim of this article is twofold. First, it attempts to demonstrate that such cases of incomplete data sets are hardly rare in comparative research. I discuss three different selection mechanisms that may lead to incomplete data sets, and provide for each mechanism an empirical example which I analyze subsequently in more detail. This demonstrates that such incomplete data sets appear in a wide array of research questions in comparative politics. Second, the article shows under what circumstances ignoring the selection mechanism, which leads to such samples, results in biased estimates. For this I rely on the truncated regression estimator proposed by Maddala (1983, pp. 176ff), Muthen and Jöreskog (1983), and Bloom and Killingsworth (1985). This estimator allows for correcting possible selection biases even if no information is available on the observations missing from that data set. The only information needed is about the mechanism that leads to the incomplete (or truncated) data set. To assess the usefulness of this estimator I carry out Monte Carlo simulations for a range of observed sample sizes that are typical in the empirical examples I discuss. These simulations demonstrate that for the incomplete data sets that normally appear in cross-national research, the truncated regression estimator provides in many circumantances estimates that are less biased than simple OLS estimates. Whether the proposed estimator is preferable in a given empirical context depends on the degree of selection and whether the selection mechanism is related to the substantive empirical model to be estimated. With three empirical examples based on data sets that fall in the range where the truncated regression estimator performs on average well, I demonstrate how selection biases due to incomplete data sets can be corrected.

The remainder of the article is organized in seven sections. In the next section I start by discussing in more detail the circumstances under which selection bias might appear in comparative research. Section 3 discusses the truncated regression model that allows addressing the issue of selection bias. I summarize the results of Monte Carlo simulations and demonstrate under what conditions it allows for correcting the sample biases due to incomplete data sets. In Sections 4 through 6 I present three empirical examples of selection bias in comparative politics. The first example concerns studies of the success of new political parties. These new parties form a "self-selected" sample of groups that might have considered forming a political party. I show that ignoring this "self-selection" may lead to serious biases in the empirical results. The second example addresses studies of new social movements that rely extensively on reports in newspapers for their empirical material. Such reports lead obviously to incomplete data sets, as newspapers select the events they cover. Ignoring this problem may lead to erroneous conclusions. Studies of ethnic conflict provide

the material for the third example. Again, I show that selection mechanisms are at work and that ignoring them can mislead researchers. I discuss these three examples in light of the Monte Carlo simulation results before summarizing my main findings and suggesting avenues for future research in the conclusion.

2 Selection Bias in Comparative Politics

Researchers in comparative politics have become increasingly aware of the dangers of selection bias. Work by Geddes (1991), King et al. (1994), and Przeworski et al. (2000) clearly illustrates the sensitivity of comparative research to issues of selection bias. But selection biases appear in at least three different guises, and I argue that one of these guises is not yet well understood.

In the work of Geddes (1991) and King et al. (1994), the presumption is that there is a clearly defined population from which a sample has to be drawn for deeper study. Both sets of authors warn comparative researchers from selection on the dependent variable, as samples drawn in this fashion may lead to biased results. Strictly speaking, the problem of selection bias in this first guise could also be corrected by studying the whole population from which the sample is drawn.

In the second guise, selection bias is conceived as a more fundamental problem in comparative politics, namely that cases select themselves into particular categories. Again, the whole population can be studied, but, to take an example, whether or not a country is democratic is not a random event. The careful discussion by Przeworski et al. (2000) of selection problems in research on democratization clearly demonstrates the importance of this issue. Most clearly, it appears for the research question of whether economic development increases the likelihood of democratization, or whether economic development hinders the breakdown of democracy. In this perspective, democracies are not a random sample of all countries, but the nature of a regime is endogenously determined. And this endogeneity needs to be addressed in order to assess the effect of democracy, for instance, on economic well-being. Przeworski et al. (2000) address this issue by correcting for the selection of countries into the category of democracies.[2] Compared to the first guise of selection bias, the second cannot be resolved by relying on population data. The problem here is inherent to the way in which the population data are produced. Thus, here there is a tight link with counterfactual analysis (e.g., Fearon 1991; Tetlock and Berlin 1996): would the economic well-being of a particular democratic state be any different if it were nondemocratic?

In its third guise, selection bias appears because we are unable to observe or clearly define the population from which our data set is drawn. This obviously does not occur when our population is the set of nations as in Przeworski et al. (2000) or to some degree in Geddes (1991). However, increasingly comparative research also employs data sets that cover many countries but where the observations do not correspond to countries. For instance, work on ethnic and civil wars relies on data on "minorities at risk" (Gurr 1993). Moreover, research on new social movements studies protest events cross-nationally and studies on new parties compare the success of new contenders across various countries. In all these cases, the populations from which "minorities at risk," protest events, and new political parties are drawn are almost impossible both to observe and to clearly delimit. Thus, while we are still able to create data sets on "minorities," protest events, or new political parties, these are

[2] These corrections are directly related to the two-stage estimators discussed by Achen (1986) and Stolzenberg and Relles (1990, 1997). Stolzenberg and Relles (1997) also present a procedure allowing researchers to determine whether a two-stage estimation is warranted.

hardly random draws from the respective "populations." Given this, as in the second guise of selection bias, our results from empirical analyses may be biased if we do not consider the way in which our data set was produced. But as we only have information on the cases that are in our data set, we cannot employ the tools to correct for selection bias that Przeworski et al. (2000) propose. Similarly, we cannot take the advice of Geddes (1991) and King et al. (1994) for the selection of our cases, because we largely cannot determine the properties of our population.

To illustrate this third guise of selection bias I referred above to three research questions from different research areas. These three questions and the data sets employed in these studies correspond, however, also to three different mechanisms that lead to such "incomplete data."[3] I label these three processes "self-selection by object of study," "selection by third party," and "selection by researcher." For each of these processes I will draw on one of the empirical examples mentioned.

Self-selection by the object of study is the typical problem in survey research. Individuals, who are the objects of survey research, choose freely to participate in surveys and to give answers to particular questions. Given the extraneous information, these selection biases can, however, rather easily be corrected (e.g., Brehm 1993). More complicated is the situation in cases where such extraneous information is absent. One example appears in the study of the electoral success of new political parties. At any point in time many groups or organizations might consider forming a new party and participate in elections. Forming a party, however, is a conscious choice among a series of different options (e.g., Rosenstone et al. 1984; Hug 1996, 2001; Cox, 1997, p. 162). Individuals involved in the formation of a new party are likely to consider the expected success of such an enterprise. Consequently, studies of the electoral success of new parties inevitably rely on self-selected samples. Not taking into account this problem can lead to considerable biases.

Selection by a third party occurs when a researcher relies on secondary sources.[4] Often such reliance on secondary sources is unavoidable, as field research is practically impossible. Given that the "creators" or "authors" of the secondary sources almost by definition select what they wish to report, considerable biases can result. Research on social movements provides an example of data sets that suffer from this type of selection bias. Longitudinal and cross-sectional studies often attempt to characterize social movements by the events they stage. A convenient research strategy is to rely on newspaper or wire reports to identify these events. Unfortunately, it is also well known that media are subject to selectivity, choosing only particular stories to report. Research clearly shows that newspapers disproportionaly report violent and large events (e.g., Fillieule 1996; McCarthy et al. 1996; Barranco and Wisler 1999). Consequently, using data based on newspaper reports can lead to serious selection biases (e.g., Hug and Wisler 1998).

A final type of selection involves the researcher herself. In several research situations the population is itself almost impossible to define. Consequently, the researcher has to impose some criteria to select the cases she judges to be relevant. A topical example appears in

[3]This term stems from similar discussions in survey research where "incomplete data" appears through "unit nonresponse" (e.g., Madow et al. 1983). Brehm (1993) discusses these issues in detail and suggests ways in which the resulting biases can be corrected, provided information on the missing respondents is available. Sigelman and Zeng (2000) discuss similar issues. In later research Brehm (2000) proposes ways in which corrections might be made, when only contextual information for the missing respondents is available. Important to note is that in the cases I discuss here, neither type of information is available, because the population is often impossible to delimit. Breen (1996) discusses the relationship between these different models in detail.

[4]Lustick (1996) discusses this type of "selection bias" in relation to the use of historical "evidence" in political science and sociology. In part his discussion also addresses the third type of selection discussed below.

studies of ethnic conflict. Almost by definition it is impossible to determine all ethnic groups present in the world. Consequently, data sets comprising information on ethnic conflict are inherently problematic, as most often only ethnic groups facing considerable discrimination or being engaged in violent conflict can be identified.[5] While the problem here appears to be mostly of a theoretical nature, namely, how to identify an ethnic group, it has serious empirical consequences.

For each type of selection process a series of other examples could be found. In other disciplines, from biology to economics, similar problems appear. While in these disciplines attempts to address the problem have proved successful, similar attempts are rare in comparative politics research.

3 A Possible Correction for Selection Bias

Ignoring the problem caused by incomplete data sets can lead to serious biases in statistical estimations carried out with the usual tools (e.g., Maddala 1983). But much more is known about possible solutions when the data set is censored, namely, when only the values of the dependent variable are unobserved. Achen (1986) surveys the appropriate models and provides several empirical examples, while Brehm (1993, pp. 93–117) discusses these biases in detail in the context of survey research. Much less is known about the properties and usefulness of a possible correction for selection bias that is appropriate when the data set is truncated, i.e., no information at all is available for the nonselected observations. Nevertheless, the truncated regression proposed by Maddala (1983, pp. 176ff), Muthen and Jöreskog (1983), and Bloom and Killingsworth (1985) allows for a correction, provided we have some theoretical notions about the variables influencing the selection into the observed sample.

In most empirical situations theoretical arguments provide some insights into the nature of the selection mechanism. In the case of the success of new political parties, studies exist that attempt to explain the formation of new political organizations (e.g., Rosenstone et al. 1984; Harmel and Robertson 1985; Kitschelt 1988). In research on new social movements, some cursory evidence exists on the nature of the selection of events by newspaper reporters and editors (e.g., McCarty et al. 1996; Barranco and Wisler 1999). Consequently, often we can identify the variables that are likely to influence the selection of cases into the observed sample. Given this, it is possible to model directly the selection mechanism that leads to the incomplete data set in a selection equation. Together with a simple outcome equation, we get the following system of equations, which has been proposed and discussed by several authors (e.g., Maddala 1983, pp. 176ff; Muthen and Jöreskog 1983; Bloom and Killingsworth 1985):

$$
\begin{aligned}
y_i &= \beta_0 + \sum_{j=1}^{k} \beta_j x_{ji} + \epsilon_i \\
t_i &= \delta_0 + \sum_{j=1}^{l} \delta_j x_{ji} + \theta_i \\
\text{if } t_i &> 0 \qquad y_i \text{ and } x_{ji} \text{ are observed} \\
\text{if } t_i &\leq 0 \qquad y_i \text{ and } x_{ji} \text{ are not observed.}
\end{aligned} \tag{1}
$$

[5] Fearon and Laitin (1997) explore Gurr's data set (Gurr 1993) on "Minorities at Risk" partly from this angle. Cohen (1997), on the other hand, brushes aside the problem of selection bias with a sleight of hand, when relying on the same data set.

In this setup, x_j is a set of independent variables, some of which explain the variable of substantive interest (y), and some of which (possibly some of them the same) explain the selection of cases into the observed sample. The selection mechanism is modeled with t as a latent variable, which reflects in some sense the likelihood that a given case appears in the observed sample. If for a given case, i, $\hat{t}_i$ exceeds 0, it is predicted to appear in the observed sample. Important to note here is that in observations for which $\hat{t}_i \leq 0$, we fail to observe not only y_i and the independent variables of the outcome equation but also the independent variables of the selection equation (thus the whole set of x_{ji}). Hence, the selection equation cannot be estimated separately as in the Heckman (1976) estimator (Achen 1986; Przeworski et al. 2000, pp. 279–289) or the familiar Tobit models (e.g., Sigelman and Zeng 2000). The assumption, however, that ϵ_i and θ_i are multinormally distributed allows deriving a likelihood function of the following form (Maddala 1983, pp. 176ff; King 1989, p. 215)[6]:

$$
\begin{aligned}
L\left(\beta, \delta, \sigma_\epsilon^2, \sigma_\theta^2, \sigma_{\epsilon\theta} \middle| y\right) &= \prod_{i=1}^{n} f_n\left(y_i \middle| \mu_{y_i}, \sigma_\epsilon^2\right)[1 - F_n(0 \mid \lambda_i, \phi^2)] \Big/ \left[1 - F_n\left(0 \middle| \mu_{t_i}, \sigma_\theta^2\right)\right] \\
\text{where } \lambda_i &= \mu_{t_i} + \frac{\sigma_{\epsilon\theta}}{\sigma_\epsilon^2}\left(y_i - \mu_{y_i}\right) \\
\phi^2 &= \sigma_\theta^2 - \frac{\sigma_{\epsilon\theta}^2}{\sigma_\epsilon^2} \\
\mu_{y_i} &= \beta_0 + \sum_{j=1}^{k} \beta_j x_{ji} \\
\mu_{t_i} &= \delta_0 + \sum_{j=1}^{l} \delta_j x_{ji}.
\end{aligned}
\tag{2}
$$

Such a model is underspecified and, therefore, an additional restriction has to be included. King (1989, p. 216) suggests fixing the standard error of the selection equation (σ_θ^2) at 1. While this is a reasonable restriction, it can lead to serious problems in the estimations, as there is no guarantee that ϕ^2 is strictly positive, which is necessary for the likelihood function to be defined.[7] Muthen and Jöreskog (1983), on the other hand, favor the imposition of another restriction enhancing the estimation, namely, setting ϕ^2 to 1. This implies that σ_θ^2 is equal to $1 + \frac{\sigma_{\epsilon\theta}^2}{\sigma_\epsilon^2}$.[8]

Together the system of Eq. (1) and the likelihood function (2) define a truncated regression, in which the parameters of interest are estimated on the basis of the truncated or incomplete data set. Several authors discuss applications of this model, but little is known about the properties of the estimator in small samples typical for comparative research and under what circumstances it performs well. Muthen and Jöreskog (1983) report results based on simulated data for three different models. For each model they use two different data sets, one comprising 1000 observations and the other 4000 observations before selection. They find that the truncated regression results come much closer to the true values of the coefficients. They note, however, that "the large-sample approximation of the standard errors is

[6]The original formulation in King (1989, p. 215) contained some typos and has been corrected in the meantime.

[7]I implemented King's solution (King 1989, p. 215) in Gauss. The relevant command file with a simulated data set is available on the *Political Analysis* Web site. LIMDEP's INCIDENTAL procedure relies on a similar restriction.

[8]This latter restriction proved more practical for the gauss-implementation of the estimation, as the value of ϕ^2 does not have to be restricted.

rather poor in the truncated case for the smaller sample sizes used" (Muthen and Jöreskog 1983, p. 162). I extend their Monte Carlo simulations and explore in more detail under what circumstances the proposed estimator leads to better results. Specifically, I attempt to show how strongly the error terms of the two equations have to correlate for the estimator to perform well, and also how its performance is affected by the degree of selection.[9] The degree of selection reflects the percentage of cases of the complete sample that appears in the data sets that researchers have at hand. Answers to both questions are important for comparative research, as often some prior knowledge on these two crucial parameters exists or can be gauged.

For the Monte Carlo simulations I use the following model to generate the simulated data:

$$y_i = x_{1i} + \epsilon_i \tag{3}$$

$$t_i = \delta_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \theta_i \tag{4}$$

$$x_{1i}, x_{2i}, y_i \text{ observed if } t_i > 0. \tag{5}$$

I set the variances of the two error terms $(\sigma_\epsilon^2, \sigma_\theta^2)$ equal to 1 and sampled the independent variables (x_1, x_2) from two normal distributions $(N(0, 1))$. The two crucial parameters that I propose to study are the correlation coefficient $\rho_{\epsilon,\theta}$ and the intercept δ_0.[10] The correlation coefficient measures how closely the two equations are linked. In the case of a coefficient close to 0, the log-likelihood function (3) collapses to a simple linear regression. The intercept of the selection equation δ_0 determines the degree of selection. With a value of 0 on average, 50% of the original sample appears in the incomplete sample. For lower values fewer cases clear the hurdle of the selection equation, while higher values lead to larger incomplete samples. In addition I use three sets of values for γ_1 and γ_2. In the first setting I let $\gamma_1 = -1$ and $\gamma_2 = 0$, which produces the simplest possible model and corresponds, with a small exception, to one used by Muthen and Jöreskog (1983). The second set uses $\gamma_1 = -\sqrt{0.5}$ and $\gamma_2 = \sqrt{0.5}$, which results in the same proportion of the variance of t being explained as in the first setup. In the third setup $\gamma_1 = -1$ and $\gamma_2 = 1$, in which case the variance explained in t increases from one half to two thirds.

For each set of values of γ_1 and γ_2 I created samples based on population sizes of 100, 250, 500, and 1000. For each combination of population size and values of γ_1 and γ_2 I created observed observations for nine values of δ_0 resulting in $10, 20, \ldots, 90\%$ of the original population being selected and nine values $(0.1, 0.2, \ldots, 0.9)$ for $\rho_{\epsilon,\theta}$. For each combination I created 1000 data sets and jointly estimated selection and outcome equation with a truncated regression and the outcome equation with OLS.

3.1 *Results*

Given the complexity of the estimation procedures of this truncation model, it is not surprising that the estimation procedure failed to converge in some instances.[11] Maddala (1983,

[9] Muthen and Jöreskog (1983) use simulated data where the correlation between the error terms is either equal to 0.5 or -0.5. In one model they impose a degree of selection that corresponds approximately to 50% of the original sample; in the second the respective percentage is approximately 25%. Thus their smallest sample comprised 250 cases. Important to note is that the correlation between the error terms and the degree of selection determine from which part of the joint distribution of the error terms the sample is drawn.

[10] The LIMDEP command file used for these Monte Carlo simulations appears on the *Political Analysis* Web site. I also carried out Monte Carlo simulations with Gauss, but on average convergence in the estimation process was more difficult to achieve in Gauss than in LIMDEP. This is most likely related to the differences in the algorithm employed. Thus, I report only the results from my LIMDEP estimations.

[11] The Monte Carlo simulations were carried out with LIMDEP, using the INCIDENTAL procedure with the default maximization algorithm DFP (Davidon–Fletcher–Powell) rank 1 update. I report the number

p. 177) notes that "it is not known whether or not the log-likelihood function ... is well-behaved in the sense of having a unique global maximum" and thus suggests using various starting values, which is impractical in Monte Carlo simulations. Thus, I used systematically the same but arbitrary starting values and report all the results only for the number of replications where the estimation routine converged.[12] Especially, when the observed sample sizes are very small, convergence rates may drop exceptionally to less than 50%. However, as I discuss below, convergence is almost always achieved in situations where a truncated regression estimation is warranted. In addition, using various starting values normally allows for a much higher percentage of converging estimations. Thus, in empirical situations, as discussed below, it is worth exploring a series of starting values and compare carefully the results of the converging estimations.

As the main thrust of the Monte Carlo results points in the same direction as those obtained by Muthen and Jöreskog (1983), I refrain from reporting them in detail. For interested readers, who might want to assess the likely bias of OLS and the potential for correction with a truncated regression, I report the detailed results of the Monte Carlo simulations in the Web appendix to this article. The mean estimates of the truncated regression are often closer to the true values of the parameters than the mean of the OLS estimates, but the standard errors of the truncation estimates (and the variance of the estimates) are inflated. With respect to the mean of the slope coefficient of the outcome equation, the results suggest that with larger population sizes the average slope estimate of the truncation regression is under most circumstances closer to the true value than the mean estimate of the OLS regression. More precisely, if the population size is 1000 and provided that the correlation between the two error terms exceeds 0.2, under almost all degrees of selection and setups of the selection equation the OLS estimator is on average further away from the true value than the truncated regression estimator. If the population size is 500, this is only the case if the correlation between the two error terms exceeds 0.4. For the still smaller population sizes of 250 and 100, samples which comprise either only a small or a large portion of the population lead to mean estimates of the truncated regression further from the true value than those from the OLS regression. Only if the degree of selection is in a mid-range and the correlation between the two errors is rather large does the truncated regression estimator outperform the OLS estimator in situations where the population sizes are small (i.e., 100 or 250).

Hence, the bias of the truncated regression is most often smaller than the bias of the OLS regression. However, as noted by Muthen and Jöreskog (1983) the standard errors of the estimated coefficients are inflated and the estimates vary widely. For this reason I also focus on the root mean-squared error of the slope estimate of the outcome equation. Also for this summary, the Monte Carlo simulations suggest a considerable impact of the size of the population. In samples stemming from populations of 100 or 250 cases, the truncated regression estimate has most often a larger root mean-squared error than the OLS estimate. Only for samples selected from populations with 500 or 1000 cases does the truncated regression estimator perform better with respect to this criterion.

Thus, the Monte Carlo results suggest that estimating a truncated regression leads in many situations to less biased estimates than OLS, but at the same time the variation around

of converging estimations per 1000 replications in the Web appendix available on the *Political Analysis* Web site.

[12] By default LIMDEP uses OLS estimates of the outcome equation, with all other coefficients set to 0, as starting values. Estimations using these starting values underperformed systematically, lending support to Maddala's suggestion of using a series of starting values (Maddala 1983, p. 177). Admittedly, using only the results from replications that converged might bias the results reported here.

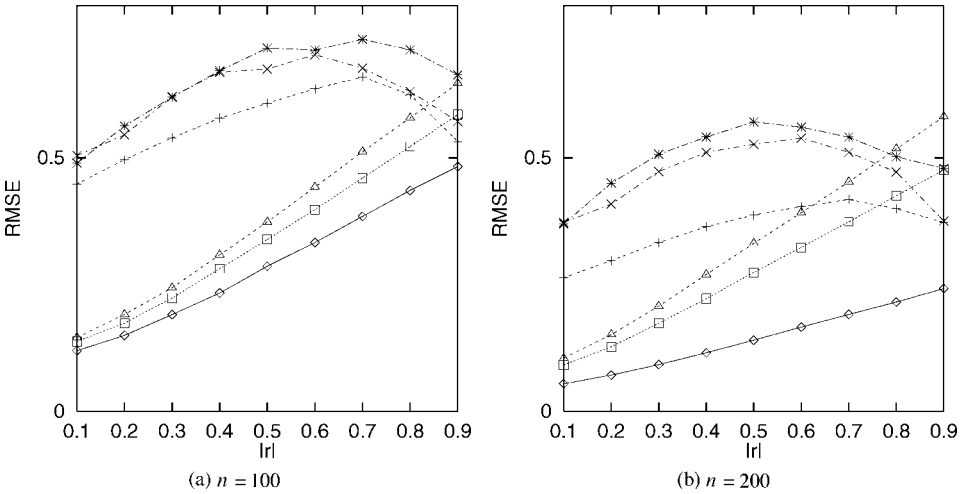

Fig. 1 Root mean-squared error of slope estimate (rmse(β_1)) as a function of degree of selection and correlation among error terms ($\gamma_1 = -1, \gamma_2 = 0$) for an observed sample of $n = 100$ (a), $n = 200$ (b) and various population sizes (N). OLS $N = 250$ $\diamond$; truncated $N = 250$ $\star$; OLS $N = 500$ $\square$; truncated $N = 500$ $\times$; OLS $N = 1000$ $\triangle$; truncated $N = 1000$ $+$.

the true values (i.e., the root mean-squared error) is often higher. The truncated regression is certainly preferable if both the mean bias and the root mean-squared error are smaller than the ones for OLS. This occurs when important omitted variables influence simultaneously the selection into the incomplete data set and the dependent variable of the outcome equation, and the selection is not too extreme (in both directions). The latter aspect is, however, more difficult to assess for the researcher, as she normally has a given sample of a certain size and has to assess whether a truncated regression estimation might be beneficial. For this reason I depict in Fig. 1a and 1b the root mean-squared errors for sample sizes of 100 and 200 for the data generated with $\gamma_1 = -1$ and $\gamma_2 = 0$. Given my Monte Carlo simulations, these samples could come either from a population of 250, 500, or 1000. For each of these population sizes and the nine values for the correlation coefficient I depict the root mean-squared errors.

Figure 1a and 1b show first of all that even with these small sample sizes the truncated regression estimator is not necessarily useless. For an observed sample size of 100, provided that the sample is selected from a population of 1000 and the correlation between the error terms is 0.9, the truncated regression is preferable. Not surprisingly, if the sample size is 200, the range of correlation coefficients and original populations for which the truncated regression estimator is better increases. Thus, except if the sample stems from a population of 250, which implies that only a fifth of the population fails to appear in the sample, the truncated regression yields a smaller root mean-squared error than OLS for high values of the correlation coefficient.[13]

While these Monte Carlo results give us valuable information on the circumstance under which a truncated regression estimation leads to smaller biases, the results underestimate in part the usefulness of this estimator. First, as mentioned above, the estimations of a

[13]It is worth noting that the truncated regression estimates are for most of the depicted conditions in Fig. 1a and 1b on average closer to the true values than the OLS-estimates (see the Web appendix on *Political Analysis* Web site). Thus, using the root mean-squared error as a criterion underestimates in some sense the usefulness of the truncated regression estimator.

truncated regression are sensitive to the starting values. It is quite likely that in empirical applications a truncated regression estimation will perform even better, provided several starting values are used for the estimation. Second, even though the OLS estimates in many circumstances have smaller root mean-squared errors than those of a truncated regression, the former have also often quite narrow confidence intervals. And in many instances, these narrow confidence intervals, given the small standard errors, fail to include the true values of the coefficients. On the other hand, the standard errors of the estimates from a truncated regression are often larger and thus the confidence intervals include more often the true values. Thus, this creates an additional tradeoff that a researcher has to keep in mind in empirical research, to which I turn now. These three empirical cases illustrate selection biases in three different research areas, and also show that the truncated regression may be a better estimating strategy for data sets affected by the three selection mechanisms I discussed above.

4 Studying the Electoral Success of New Political Parties

Studies of the electoral success of newly formed political parties are potentially challenged by problems of selection bias.[14] If we believe Rosenstone, Behr, and Lazarus' argument (Rosenstone et al. 1984), third candidates in presidential races present themselves when their likely success is high. Similarly, theoretical work (e.g., Hug 1996, 2001; Cox 1997, p. 162) suggests that the decision by political entrepreneurs to form a new party is partly dependent on the expected electoral success. Consequently, the third candidates and new political parties that actually appear on the ballot form a self-selected sample of all groups that have considered competing in an election. And the process of selection is likely to be affected by the hurdles that new parties have to cross before they qualify for the ballot (e.g., Rosenstone et al. 1984, pp. 148ff).

Several authors (e.g., Harmel and Robertson 1985; Rootes 1995) suggest that institutional characteristics like ballot access should influence the fortunes of new parties. Harmel and Robertson (1985) expect the ballot structure to affect the electoral success of new parties. Their empirical results, however, hardly support their theoretical claims. If we believe that selection bias plays a role, such discouraging results are hardly surprising. I attempt to illustrate this potential problem with a data set covering 21 Western democracies from 1945 to 1990.[15] The data set covers 225 new parties that have appeared on ballots at national elections in these countries.

If Harmel and Robertson (1985) are correct, important hurdles to get on the ballot should decrease the electoral success of new parties. I measure these hurdles with two crude measures. They reflect the number of signatures and the electoral deposit required to qualify for the ballot.[16] Harmel and Robertson (1985) and Müller-Rommel (1993) also suggest that the electoral system should affect the electoral success of new parties. Thus, I introduce as an additional independent variable the threshold of representation (Lijphart and Gibberd 1977) to reflect the effect of the electoral system.

The dependent variable is the log-transformed vote share the new party obtained at its first participation in national elections. The log transformation was necessary because of a

[14] This section draws on Hug (2000, 2001).

[15] The data set stems from Hug (2001), where it is described in detail. In the Web appendix I report the descriptive statistics of all variables employed in this example as well as those of the two other examples.

[16] Both variables are scaled to make comparisons across time and space meaningful. The signature requirement is expressed as per thousands of the voting population, whereas the electoral deposit is divided by the GDP per capita and adjusted for inflation.

Table 1 Explaining the electoral success of new parties

	Total sample OLS		*Selected sample truncated*	
	b	*SE*	*b*	*SE*
Outcome equation				
Threshold of representation	0.929	0.709	0.187	0.518
Petition requirement (in per mille of electoral body)	0.107	0.047	0.085	0.078
Registration costs	0.734	0.536	−0.420	1.270
Constant	−2.464	0.057	−2.465	0.062
Selection equation (formation of a new party)				
Petition requirement (in per mille of electoral body)			−0.118	0.285
Registration costs			−13.102	3.566
Public party financing			8.212	507.070
Constant			1.533	0.428
$\rho_{\epsilon_o,\epsilon_s}$			0.918	0.113
Standard error of the estimate	0.595		0.611	0.023
Log-likelihood	−200.381		−190.780	
n	225		225	

very skewed distribution of the vote shares. The empirical results (Table 1, column 1) based on a simple OLS regression suggest that the petition requirement has a positive impact on the electoral success of new political parties. This effect reaches statistical significance, and suggests, contrary to the argument of Harmel and Robertson (1984), that as it becomes more difficult to qualify for the ballot, new parties become more successful. The other coefficients fail to reach statistical significance. Interesting to mention is the positive estimate for the threshold of representation. This reflects the mixed results appearing in the literature on the effect of electoral systems on the fortune of new parties (e.g., Harmel and Robertson 1985; Kitschelt 1988; Müller-Rommel 1993).

I reestimate the same explanatory model, assuming that the independent variables measuring the difficulty to get on the ballot also influence the selection, or more precisely, the formation of new parties. In addition I include a variable measuring whether or not parties receive public funding. I refrain from introducing the threshold of representation in the selection equation, as the electoral system primarily translates votes into seats. Thus, the effect on the decision to create a political party is at most indirect, namely mediated through the expected number of votes a new party might expect.

The estimated coefficients of the outcome equation undergo considerable change, and none of them reaches statistical significance anymore (Table 1, column 2). More specifically, I can no longer reject the null hypothesis that increasing the petition requirement fails to increase the predicted electoral success of new parties. On the other hand, the estimated coefficients for the selection equation suggest that both the level of the petition requirement and especially the amount of the electoral deposit influence negatively the selection into the sample. Intuitively this makes considerable sense. The difficulty to get on the ballot should not influence directly the electoral success of new parties. It should influence much more heavily the decision of political entrepreneurs to form a new party or abstain from

such an endeavor.[17] A positive effect, but with an inflated standard error, appears also for the party-financing variable.

This example first of all illustrates a selection mechanism that is due to a decision by the object of study, i.e., a group or organization to form a new party and compete in an election. The main question a researcher has to ask herself in light of the Monte Carlo results discussed above is how large the "unobserved" population is likely to be, from which the sample of 225 new parties was drawn. If the size of the "unobserved" population is at least 500, Fig. 1b and the estimated correlation coefficient between the two error terms would suggest that the truncated regression is preferable to the OLS estimation. Obviously, this depends on whether the structure of the selection and outcome equation in this empirical example is comparable to the much more simple model employed for the Monte Carlo simulations.

5 New Social Movement Research

Newspaper reports form a central source for research on new social movements.[18] Obviously, newspapers select carefully the stories they report on. Thus such research nicely illustrates the problems of using data sets that are incomplete because of "selection by third party." Using only newspaper reports in order to study the activities of new social movements can lead to serious biases. Despite this problem, most recent research in this field continues to rely on newspaper or wire reports.[19] But several recent studies have shown that newspapers seriously underreport nonviolent events and demonstrations with small numbers of participants (e.g., Fillieule 1996; McCarthy et al. 1996; Barranco and Wisler 1999). Neglecting this selectivity of newspapers can potentially damage empirical results.

To illustrate this potentially damaging selection bias in research on new social movements, I rely on a data set covering events staged by such movements in four Swiss cities (Barranco and Wisler 1999). The advantage of this data set is that it relies both on police reports and on newspapers to identify the relevant events. While the events reported by newspapers is obviously an incomplete data set, this is also the case for the events recorded by the police, but to a lesser degree. If we assume (falsely, I would argue)[20] that the events reported by the police are the population of events, we can directly compare the biases of the various estimation strategies in the incomplete sample. Nevertheless, some caution is in order, as police records only approximate a "complete sample."

Thus, the data set allows for investigations into how newspapers select the events they report on compared to those recorded by the police. For the example presented here, I look at how relying only on a national newspaper [*Neue Zürcher Zeitung* (NZZ)] to identify events can bias empirical results. I study a very simple relationship linking the size

[17] This result finds a parallel in Rosenstone, Behr, and Lazarus's analysis, when they employ a two-stage-least-squares approach (Rosenstone et al. 1984). They argue that the proportion of voters that have third candidates on their ballots is indirectly influenced by the likely electoral success of third parties. Consequently, they chose an instrumental variable approach.

[18] This part of the article draws heavily on Hug and Wisler (1998), where a more detailed discussion of the problems in research on new social movements appears.

[19] A review of the most recent research efforts in this field appears in Hug and Wisler (1998).

[20] While it will appear that police reports yield almost twice as many events than a national newspaper, some of the events reported in newspapers fail to appear in police reports (Barranco and Wisler 1999). Hug and Wisler (1998) and Barranco and Wisler (1999) show that larger and more violent events are more likely to be covered in a national newspaper.

of a demonstration with the degree of violence.[21] As both violence and the size of a demonstration are linked to the likelihood of an event to be reported in a newspaper, selection bias is likely to occur. The police reports allow me to compare reports in the newspapers to a much more complete picture of reality.

Table 2 reports in the first column a simple OLS regression using the sample of events covered by the police attempting to explain the degree of violence. The results suggest that as demonstrations attract more people, the more likely they are to turn violent. This effect is rather large and statistically significant. The very same variable, namely the number of participants, affects also the probability that an event covered by the police appears in a report in the national newspaper (Table 2, column 2). Similarly, the city where the event occurs has a considerable influence on the selection. Given that Geneva is the reference category, the results suggest that events in the three other cities are reported significantly more often. This is most likely not unrelated to the fact that the national newspaper (NZZ) appears in German and Geneva is the only French-speaking city covered in the sample.

When considering only the events reported in the national newspaper, the sample is roughly reduced in half. Estimating the same simple relationship between the participation numbers and the degree of violence, one finds a reduced coefficient, which fails to reach statistical significance (Table 2, column 3). Based on these results one would be unable to reject the hypothesis that the size of a demonstration has no influence on the degree of violence. Reestimating this same relationship as a truncated regression alters the picture considerably (Table 2, column 4). The effect of the size of the demonstration increases and becomes highly significant. At the same time the size of the event also largely explains the probability of whether it will be reported or not, while there appear no longer any significant differences among the four cities considered. Again, it appears that the data at hand suffer from considerable selection bias. If we take the results from the truncated regression as sufficiently close to reality, this would also suggest that not only newspapers select the events they report on, but that the same thing also holds for police records.

This second empirical example illustrates another selection mechanism, namely the selection by a third party. The data set on protest events includes two such selections, namely the one by the police, which seems to be less lenient, and the more severe one by the national newspaper. In the latter case, the size of the observed sample is large, but given the size of the police sample and my argument that even the larger sample is incomplete, one would expect that the degree of selection is quite small. In all cases for the national newspaper this degree of selection must be smaller than 50%. Combined with the large estimated correlation coefficient between the two error terms, the Monte Carlo simulations suggest again, that the OLS estimates would be more strongly biased than the estimates of the truncated regression. Two elements also seem to speak in favor of the assumption that even the police sample is far from complete. First, while the cities appear to affect very strongly the selection from the police sample to the sample covered in the national newspaper, these effects almost disappear in the selection equation of the truncated regression. Second, the effect of the number of participations is larger in the outcome equation of the truncated regression than in the simple OLS estimation based on the police sample. Both these differences seem to suggest (not surprisingly) that even the police reports miss some events.

[21] Strictly speaking, the dependent variable is an ordinal scale going from minor violent occurrences (1) to full-blown battles with police forces (5), while 0 indicated no violence. Details appear in Hug and Wisler (1998), where similar results are discussed for a slightly differently defined dependent variable.

Table 2 Explaining the degree of violence of demonstrations in Swiss cities (1965–1994)

	Police sample				*Selected sample (NZZ)*			
	OLS		*Probit*		*OLS*		*Truncated*	
	b	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Outcome equation								
Participants (in 1000)	0.030	0.009			0.020	0.011	0.026	0.009
Constant	0.262	0.022			0.409	0.039	−0.303	0.123
Selection equation (report in national newspaper NZZ)								
Participants (in 1000)			0.281	0.031			0.995	0.425
Cities (Geneva base category)								
Berne			0.892	0.085			0.277	0.394
Basle			0.652	0.155			0.163	0.663
Zurich			1.120	0.078			−0.330	0.354
Constant			−0.800	0.056			0.117	0.376
$\rho_{\epsilon_o,\epsilon_s}$							0.914	0.041
Standard error of the estimate	0.858				1.034		1.066	0.028
Log-likelihood	−2169.309		−1000.715		−1253.682		−1214.921	
n	1714		1714		864		864	

6 Studying Ethnic Conflict

Research on ethnic conflict has proliferated considerably in recent times. Several studies attempt to explain the degree of violence or repression of ethnic groups based on the characteristics of ethnic groups and the features of the political systems in which the latter exist. Such studies face, however, considerable problems. As several authors acknowledge (e.g., Fearon and Laitin 1997), empirical studies of ethnic conflict are potentially subject to considerable selection bias. This stems from the fact that mobilized, rebelling, and violent ethnic groups are simply much easier to identify than peaceful, reclusely, living tribes. Scholars attempt to alleviate this problem by adopting very lenient criteria for including a group as, for instance, a minority at risk (e.g., Gurr 1993). This, however, cannot assure us that the problem of selection bias has been completely addressed. Quite to the contrary, the theoretical difficulty (or even impossibility) of defining ethnic groups makes the presence of selection bias almost a certainty. Given that researchers are unable to define once for ever the "population" of ethnic groups, this problem is likely to stay with us. Thus, research on ethnic conflict based on data on "minorities at risk" nicely illustrates the problem "selection by researcher" may cause.

I attempt to illustrate this problem by replicating a simple analysis proposed by Gurr and Moore (1997). These authors employ one of the most complete data sets on "political communal groups," namely Gurr's data set (Gurr 1993) on "minorities at risk." Gurr (1993, p. 5, emphasis in original) defines the scope of this study in the following way:

> This study is limited to nonstate communal groups that were politically salient during the post-World War II era, that is *political communal groups.* Communal groups are politically salient, for our purposes, if they meet one or both of the two primary criteria: they experience economic or political discrimination, and they have taken political action in support of collective interests.

While being rather lenient in some sense, this definition also excludes many communal groups. In addition, these criteria are also most likely quite closely linked to several research questions one might want to study with this data set. For instance, explaining the degree of political action and the violence thereof is closely linked to the selection criteria. Gurr and Moore's (1997) model, with which they attempt to explain the degree of rebellion,[22] provides an example for this problem. Based on their data set[23] I replicate their findings. In Table 3 I present in the first column the results that Gurr and Moore (1997, p. 1091) report for the last stage of a three-square-least-squares (3SLS) estimation. Estimating only this last equation with OLS leads to only minor substantive changes in the results (Table 3, second column).[24] The results suggest that the degree of rebellion of a particular group is positively affected by the level of mobilization of the group and the presence of rebellion in the same region. Democratic countries appear to reduce the expected level of rebellion, whereas grievances fail to influence the dependent variable in a statistically significant manner.

[22] It has to be noted that the primary aim of their article is to propose a strategy of risk assessment for ethno-political conflict.

[23] Their data set is available via the Web at the ICPSR archive.

[24] For this estimation I extracted from the replication data set of Gurr and Moore (1997) the following variables: reb80s, allgrixx, mobplus, dempow, iconreb8. These correspond to the dependent and the independent variables in the order as they appear in Table 3.

Table 3 Explaining the level of rebellion of "minorities at risk"

	Gurr/Moore in AJPS[a] *3SLS*		*Replication OLS*		*Truncated regression*	
	b	*SE*	*b*	*SE*	*b*	*SE*
Outcome equation						
Grievances	0.14	0.13	0.113	0.055	0.169	0.074
Mobilization	0.86	0.21	0.720	0.126	−0.543	0.337
Democratic power	−0.04	0.01	−0.053	0.012	−0.028	0.011
International rebellion	0.66	0.16	0.653	0.136	0.576	0.101
Constant	0.69	0.64	0.008	0.483	4.342	1.260
Selection equation (definition as "minority at risk")						
Grievances					−0.047	0.032
Mobilization					1.361	0.394
Constant					−0.736	0.463
$\rho_{\epsilon_o,\epsilon_s}$					−0.938	0.0216
Standard error of the estimate			3.304		3.823	0.296
Log-likelihood			−525.4984		−497.8657	
n	202		202		202	

[a]The results in this column reflect those reported by Gurr and Moore (1997, p. 1091) for the last stage of the 3SLS estimation. Definitions for the variables can be found in their article.

Interesting in this setup is that the mobilization of the group, which is partly used to select the relevant groups, appears as an independent variable.[25] Consequently, I propose a model where two independent variables of the last stage of the 3SLS setup of Gurr and Moore (1997) also appear as independent variables for a selection equation. Given the definition of a "communal group," there are good theoretical reasons to expect the degree of mobilization and the level of grievances to affect the selection into the sample. Estimating this truncated regression modifies some of the basic insights of Gurr and Moore's (1997) analysis. The direct effect of the degree of mobilization on the level of rebellion becomes negative and fails to reach statistical significance. The same independent variable, not surprisingly, shows a considerable effect in the selection equation. As the mobilization of a group increases, it is far more likely to appear in the incomplete data set. While other coefficients also differ somewhat, the change of the mobilization estimate is the most striking. That it should change, once the selection mechanism is considered, makes intuitive sense. That it should turn negative is less easy to grasp. One explanation might be that in a complete sample of all relevant groups, highly mobilized groups fail to feel it necessary to resort to rebellious behavior. Conceivably, their opponents might consider it more prudent to oblige to their demands than risk the rebellion of a highly mobilized group.

This final empirical example illustrates the selection of cases by a researcher. The estimated correlation coefficient between the error terms exceeds, like in the the other examples of the truncated regressions, 0.9 in absolute value. To assess whether the truncated regression performs better for the sample size of slightly more than 200 cases, we need to gauge

[25]In their 3SLS model mobilization also appears as the dependent variable. This explains the differences in the estimates between the results of Gurr and Moore (1997) and my replication based on a simple OLS estimation.

the likely size of the "unobserved" population. A recent discussion of a broader data set of ethnic groups (Fearon 2002) suggests that the degree of selection must be quite important, as more than 800 groups are covered. Thus again, based on the results of the Monte Carlo simulations a researcher might find it preferable to rely on a truncated regression estimation that, most likely, will yield less biased results than an OLS regression.

7 Discussion

The three empirical examples illustrate the three selection processes that I sketched at the beginning of the article. They also show that the selection biases appearing in these three cases cannot be addressed in the same way as the selection problems discussed by Geddes (1991), King et al. (1994), and Przeworski et al. (2000, pp. 279–289). It is not possible to choose more carefully the cases from the population, as the latter is not clearly defined. Similarly, the problem is not that some observations have selected themselves (or been selected) into a particular category of our variables. In the three empirical examples, observations either selected themselves into the sample (new political parties), were selected by a third person (protest events), or were selected by the researcher herself ("minorities at risk"). And these selection processes are either impossible to control for the researcher, or a selection process is necessary as the population (e.g., groups that might want to engage in ethnic war) is simply impossible to define.

The three examples also share some aspects that are of importance in the light of the Monte Carlo simulations. In two examples the observed samples comprised roughly 200 observations, whereas in the remaining one the sample consisted of more than 800 events. Thus in all cases the samples exceeded 100, for which the Monte Carlo simulations proved to be crucial. For smaller sample sizes the root mean-squared errors of the truncated regressions are much larger than those of an OLS estimation, even though in many cases the former estimates are less biased than the latter.

Second, in all three examples, the truncated regression yields an estimate for the correlation between the error terms of the selection and outcome equation, which is large and exceeds 0.9. Again, the Monte Carlo estimations suggest that with such a large correlation, almost under all circumstances the truncated regression estimates are less biased than the OLS estimates. In addition, even the root mean-squared errors are almost systematically lower for the truncated regressions than the OLS estimates.

Combined with the fact that the sample sizes are in all three cases larger than 200, this suggests that the estimates from the truncated regression presented in the three examples should be less biased than those of the OLS regression. In all the Monte Carlo simulations where the correlation between the error terms of the selection and outcome equation was 0.9 and the sample size 200, the truncated regression yielded smaller biases and root mean-squared errors than OLS. Obviously, this comparison can only be indicative, as the models estimated for the three examples are more complicated than the ones used in the Monte Carlo simulations. Similarly, this also assumes that in the three examples there are no significant specification problems.

Finally, in all three cases there are theoretical reasons why one might expect one or more independent variables on the outcome equation to influence the sample selection. Given that these independent variables are likely to be supposed to tap into broader theoretical concepts, it is likely that some variables are omitted both on the selection and outcome equation. Thus, and related to the previous point, the error terms of the two equations are likely to be correlated. This suggests again that failing to address the selection problem would lead to considerable biases.

8 Conclusion

Comparative politics has become much more attentive to problems of selection bias in recent years. This increased attention has led scholars to choose their cases more carefully and to spend more effort on justifying their choices. Similarly, scholars have become more attentive to related issues of endogeneity and self-selection into categories of variables, which are closely linked to counterfactual analysis. I argued in this article that despite these improvements, another type of selection bias still remains largely unaddressed in comparative research. This other type of selection bias occurs because of incomplete data sets. In several fields researchers can at best hope to have at hand a data set that is not too heavily afflicted by its incompleteness. In these contexts it is often materially impossible to obtain a complete data set or to learn more about the characteristics of the total population from which the incomplete data set stems.

In these situations, I argue, much more attention has to be paid to the mechanisms that select the observations into the incomplete data sets. Whether newspapers or historians choose to report only certain events, whether researchers have to impose some criteria to define their universe of observations, or whether the objects of study select themselves into a sample, in all cases scholars should gain a better understanding of the underlying mechanisms.

Understanding these mechanisms is crucial whenever one attempts to correct the potentially resulting biases in empirical analyses. I discussed and evaluated one particular possible solution to the problem of incomplete data sets. This solution consists of directly modeling the selection mechanism and estimating a truncated regression. This method works reasonably well under specific conditions. Monte Carlo simulations allowed me to demonstrate the properties of the estimator.

This statistical evaluation can, however, only be a first step. In three empirical examples I attempted to show the potentialities of this approach. While some of them yield rather conclusive results, others do so only partly. Further research has to explore in more detail under what circumstances a truncated regression model performs well in empirical settings. The tradeoff between using biased OLS estimates with small standard errors and using often less heavily biased truncated regression estimates with larger standard errors is likely to be central in this context.

References

Achen, Christopher H. 1986. *Statistical Analysis of Quasi-Experiments*. Berkeley: University of California Press.

Barranco, José, and Dominique Wisler. 1999. "Validity and Systematicity of Newspaper Data in Event Analysis." *European Sociological Review* 15(3):301–322.

Bloom, David E., and Mark R. Killingsworth. 1985. "Correcting for Truncation Bias Caused by a Latent Truncation Variable." *Journal of Econometrics* 27:131–135.

Breen, Richard. 1996. *Regression Models: Censored, Sample Selected or Truncated Data*. Thousand Oaks, CA: Sage.

Brehm, John. 1993. *The Phantom Respondents. Opinion Surveys and Political Representation*. Ann Arbor: University of Michigan Press.

Brehm, John. 2000. "Alternative Corrections for Sample Truncation: Applications to the 1988 and 1990 Senate Election Studies." *Political Analysis* 8:183–199.

Cohen, Frank S. 1997. "Proportional Versus Majoritarian Ethnic Conflict Management in Democracies." *Comparative Political Studies* 30:607–630.

Cox, Gary W. 1997. *Making Votes Count*. Cambridge: Cambridge University Press.

Dion, Douglas. 1998. "Evidence and Inference in the Comparative Case Study." *Comparative Politics* 30:127–145.

Fearon, James D. 1991. "Counterfactuals and Hypothesis Testing in Political Science." *World Politics* 43:169–195.

Fearon, James D. 2002. “Ethnic Structure and Cultural Diversity Around the World: A Cross-National Data Set on Ethnic Groups.” Paper prepared for delivery at the 2002 Annual Meeting of the American Political Science Association, Boston, August 29–September 1, 2002.

Fearon, James D., and David D. Laitin. 1997. *A Cross-Sectional Study of Large-Scale Ethnic Violence in the Postwar Period*. Paper prepared for the conference Cooperation under Difficult Conditions, La Jolla, University of California, San Diego, 1997.

Fillieule, Olivier. 1996. *Police Records and the National Press in France: Issues in the Methodology of Data-Collections from Newspapers*. Florence, European University Institute. EUI working papers of the Robert Schuman Centre; RSC 96/25.

Geddes, Barbara. 1991. “How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics.” In *Political Analysis*, James A. Stimson, ed. Ann Arbor: University of Michigan Press, pp. 131–152.

Goertz, Gary D., and Bear F. Braumoeller. 2000. “The Methodology of Necessary Conditions.” *American Journal of Political Science* 44:844–859.

Gurr, Ted Robert. 1993. *Minorities at Risk. A Global View of Ethnopolitical Conflict*. Washington, DC: United States Institute of Peace Press.

Gurr, Ted Robert, and Will H. Moore. 1997. “Ethnopolitical Rebellion: A Cross-Sectional Analysis of the 1980s with Risk Assessments for the 1990s.” *American Journal of Political Science* 41:1079–1103.

Harmel, Robert, and John D. Robertson. 1985. “Formation and Success of New Parties.” *International Political Science Review* 6:501–523.

Heckman, James J. 1976. “The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models.” *Annals of Economic and Social Measurement* 5:475–492.

Hug, Simon. 1996. “Altering the Electoral Scene. The Emergence of New Political Parties from a Game-Theoretic Perspective.” *European Journal of Political Research* 29:169–190.

Hug, Simon. 2000. “Studying the Electoral Success of New Political Parties. A Methodological Note.” *Party Politics* 6:187–197.

Hug, Simon. 2001. *Altering Party Systems. Strategic Behavior and the Emergence of New Political Parties in Western Democracies*. Ann Arbor: University of Michigan Press.

Hug, Simon, and Dominique Wisler. 1998. “Correcting for Selection Bias in Social Movement Research.” *Mobilization* 3:141–161.

King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Cambridge: Cambridge University Press.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry*. Princeton, NJ: Princeton University Press.

Kitschelt, Herbert. 1988. “Left-Libertarian Parties. Explaining Innovation in Competitive Party Systems.” *World Politics* 40:194–234.

Lijphart, Arend, and Robert W. Gibberd. 1977. “Thresholds and Payoffs in List Systems of Proportional Representation.” *European Journal of Political Research* 5:219–244.

Lustick, Ian S. 1996. “History, Historiography, and Political Science: Multiple Historical Records and the Problem of Selection Bias.” *American Political Science Review* 90:605–618.

Maddala, G. S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Madow, William G., Harold Nisselson, and Olkin Ingram, eds. 1983. *Incomplete Data in Sample Surveys*. New York: Academic Press.

McCarthy, John D., Clark McPhail, and Jackie Smith. 1996. “Images of Protest: Dimensions of Selection Bias in Media Coverage of Washington Demonstrations, 1982 and 1991.” *American Sociological Review* 61:478–499.

Müller-Rommel, Ferdinand. 1993. *Grüne Parteien in Westeuropa. Entwicklungsphasen und Erfolgsbedingungen*. Opladen: Westdeutscher Verlag.

Muthen, Bengt, and Karl G. Jöreskog. 1983. “Selectivity Problems in Quasi-Experimental Studies.” *Evaluation Review* 7:139–174.

Przeworski, Adam, Michael E. Alvarez, José Antonio Cheibub, and Fernando Limongi. 2000. *Democracy and Development: Political Institutions and Well-Being in the World, 1950–1990*. New York: Cambridge University Press.

Rootes, Chris. 1995. “Environmental Consciousness, Institutional Structures and Political Competition in the Formation and Development of Green Parties.” In *The Green Challenge. The Development of Green Parties in Europe*, Dick Richardson and Chris Rootes, eds. London: Routledge, pp. 232–252.

Rosenstone, Steven J., Roy L. Behr, and Edward H. Lazarus. 1984. *Third Parties in America*. Princeton, NJ: Princeton University Press.

Sigelman, Lee, and Langche Zeng. 2000. "Analyzing Censored and Sample-Selected Data with Tobit and Heckit." *Political Analysis* 8:167–182.

Stolzenberg, Ross M., and Daniel A. Relles. 1990. "Theory Testing in a World of Constrained Research Design." *Sociological Methods and Research* 35:101–132.

Stolzenberg, Ross M., and Daniel A. Relles. 1997. "Tools for Intuition about Sample Selection Bias and Its Correction." *American Sociological Review* 62:494–506.

Tetlock, Philip E., and Aaron Berlin, eds. 1996. *Counterfactual Thought Experiments in World Politics*. Princeton, NJ: Princeton University Press.