Chapter 2 on the design of this study contains a section which deals with incomplete data encountered in the analysis of this book in particular. In this section I suggest that existing techniques for the treatment of missing data do not apply for certain variables of this book, namely for leader evaluations and left-right distances. For these two variables I suggest a specific treatment of missing cases in the form of a ‘white-noise imputation’. The objective of this appendix is to give a more general and elaborate argument for such a ‘white-noise imputation’, a method that is applicable not only for problems encountered in the analysis of this book but for kinds of missing data in general. Due to the general applicability of the technique, I do not refer specifically to the analysis of voters’ reasoning in this stand-alone appendix but illustrate the argument on the basis of simple examples of (political) survey research.

This appendix is structured in six sections. First, I will very briefly review some arguments why missing data may become problematic for statistical analysis of survey data and which tools allegedly compensate for such problems. Second, I will explicate the assumptions on which these methods rest and under which circumstances such assumptions do or do not hold. In a third section, I will formally develop the ‘white-noise imputation’ of missing data as an alternative to existing imputation procedures and in a fourth section I will formally demonstrate the consequences of the technique for a
simple OLS model. A fifth section provides an empirical example (voting for the Dutch Greens (GL) in the 1994 Dutch parliamentary election) that illustrates the consequences of different missing data treatments. Finally, I will address the question under which circumstances which method is most suitable.

A1.1 Types of Missing Data and Alleged Proper Ways to Handle Them

Incomplete data are a widespread phenomenon in survey research. On the basis of empirical articles published in leading American political science journals, King et al. (2001) estimate that about 1/3 of original samples is usually deleted from analyses because of missing scores on single variables. It is commonly acknowledged that incomplete data obstruct practical restrictions: one cannot analyse data that have many missing scores, as a listwise deletion of incomplete data will leave too few cases for the substantive analysis. Yet another problem of the listwise deletion of missing data, the possibility of selection bias, is often ignored.

The literature on incomplete data distinguishes three forms of ‘missingness’. Data missing completely at random, MCAR, denotes a situation in which the occurrence of incomplete data is not related to any other variable in the dataset. If the generation process of missing scores is random, a listwise deletion resembles a random draw from the original sample. A listwise deletion is thus a sensible solution to the missing data problem because statistical models based on this random sub-sample will yield approximately the same, unbiased estimates as statistical models based on the complete dataset. However, the smaller sample size makes the estimation less efficient.
The second form of missing scores, data missing at random, MAR, is central for the further discussion. MAR refers to a situation in which the likelihood of missing data is related to other variables in the dataset. This is probably the most frequent form of missing data: the propensity of incomplete answers will typically be higher for certain groups of respondents. For instance, in this study incomplete data is related the context of political systems and education (Chapter 2). As MAR implies a systematic occurrence of missing data, the listwise deletion of such data contains the problem of selection bias. A listwise deletion of missing data in this book would result in a biased sub-sample in which less educated respondents are underrepresented. As education significantly moderates the reasoning of vote choice (Chapter 4), any analysis of the calculus of voting generates bias in parameter estimates if one deletes missing data listwise. According to the literature, imputation techniques handle problems related to MAR most effectively (e.g., Allison 2002). These methods estimate unobserved data using patterns of relationships in observed cases. The literature suggests that of the host of parametric as well as non-parametric methods, multiple multivariate normal specifications and related methods approximate unobserved data most effectively (Schafte 1997; King et al. 2001). A detailed discussion of these methods and their differences is beyond the scope of this book (see e.g., Allison 2002). Yet it is crucial for the further argument to note that these methods in principle belong to a single class of solutions. Applying these methods means replacing missing data by a reasonable guess. In terms of linear models, some sort of $\hat{Y}$ is substituted for missing scores.

The last form of incomplete data, nonignorable missing data, NI, describes a situation in which the likelihood of missing scores is related to the true value that is unobserved. NI can be observed, for example, if wealthy respondents are more reluctant to report their income. Heckman models for selection bias are the alleged proper
way to handle this specific problem of missing data (Heckman 1976). The following paragraphs omit the category of nonignorable missing data, NI, as well as data missing completely at random, MCAR, and focus instead on solutions to the problems of data missing at random, MAR.

A1.2 Basis of Missing Data

As argued in the previous section missing data (MAR) generate practical restrictions (not all variables can be analysed), result in an inefficient estimation (fewer cases), and possibly cause selection bias. In this section I argue that assumptions in imputation techniques about the basis of missing data frequently do not hold empirically. Proposed solutions to these problems are, in my opinion, therefore often not applicable, despite their sophistication. I distinguish two bases for missing data in this section: measurement problems and insufficient information.

A first possible reason for incomplete data is measurement problems. The literature on survey research contains various examples for that. These range from respondents hiding the true answer (income, drug abuse, etc.) to social desirability (church attendance, vote turnout, etc.). Or it may simply be the case that respondents want to get through the questionnaire as fast as possible by giving 'don't know' answers (Tourangeau et al. 2000). Furthermore, analysts often cause measurement problems deliberately by applying split-half techniques: variables are not observe for all cases of the sample but only for a certain group or a random subsample, which consequently produces a considerable amount of incomplete data. If missing data can be traced back to measurement problems it implies that there exists a true score that could not be observed due to these problems. To recapitulate, imputation techniques replace missing scores by a reasonable guess. These methods
therefore apply if one can safely assume that a true value of the unobserved information exists.

The second basis of incomplete data is insufficient information. If respondents are asked to give a statement on a stimulus, this presupposes that respondents are sufficiently familiar with the stimulus. However, a ‘don’t know’ answer may also be a valid response if respondents truly do not have adequate information with regards to the stimulus. In this case a ‘don’t know’ answer does not reflect measurement problems and cannot be treated as unobserved data. The true score is in fact observed, the problem, however, is that this value cannot be related to the metric of the response scale. It does not make sense to approximate a score that does not exist. Consequently, existing imputation techniques do not apply. To illustrate this point, suppose an analyst asks respondents throughout Europe whether or not they like Romano Prodi, the president of the European Commission. It is very likely that a considerable fraction of respondents states never to have heard this name before.\textsuperscript{208} A ‘don’t know’ answer is in this case the true answer. Yet one cannot relate the answer to the like-dislike scale and it is therefore not possible to include it in the analysis.\textsuperscript{209} Keeping in mind

\textsuperscript{208} One may expect less missing data on the question in Italy than in the rest of Europe, MAR. Also, politically interested respondents probably answer this question with ‘don’t know’ less frequently. If incomplete cases are deleted listwise, one introduces selection bias in the sub-sample (Italians and politically interested respondents are over-represented).

\textsuperscript{209} In some applications one finds a solution which Greene (2000: 262) refers to as a modified zero-order regression. Missing cases on an independent variable are filled with a constant (e.g., zero) and one enters an additional dummy-variable in the analysis that differs between observed and unobserved cases on the independent variable. However, as Greene (2000) points out, this procedure is algebraically identical to the simple mean substitution. This simple mean substitution generates approximately the same (possibly biased) parameter estimates than a procedure of listwise deletion and is therefore no solution to the problem. For a formal argument on this point see also Greene (2000) or Appendix 3 of this book.
that incomplete data can be problematic, irrespective of its cause, it is crucial to find a way that allows this information to be taken into account.

A1.3 A ‘Don’t Know’ Answer as Valid Response and Its Operationalisation

What can be done when one encounters incomplete data whose occurrence is related to other variables of the analysis, MAR, and one knows that the basis of missingness is insufficient information? First, a listwise deletion of these data will produce problems as discussed in the previous paragraph. Not all variables can be included in the analysis, the estimation of statistical models is inefficient, and most importantly, parameter estimates are possibly biased. Deletion is therefore no solution to the problem.

Second, to apply existing imputation techniques even though one knows that incomplete data are not due to measurement problems but rather to insufficient information, means to create patterns in the data that do not exist. The method is indeed not robust to violations of its assumptions. Coming back to the example of Romano Prodi, suppose one is interested in the effect the evaluation of Romano Prodi may have on the overall evaluation of the EU. If ‘don’t know’ answers on the Romano Prodi question are imputed by existing imputation techniques, relationships found in the group of informed respondents are imposed on the group of uninformed respondents. This alters the research question. One does no longer study the effect of sympathy towards Romano Prodi for the assessment of the EU in Europe but the

---

210 This is the problem as described in Chapter 2 of this book: many respondents are found not to know all politicians of a political system and, moreover, these incomplete data are linked to certain political systems (e.g., Ukraine, Hong Kong) and certain groups of respondents (less educated).
relevance of Romano Prodi for the evaluation of the EU in a fictional world in which all citizens have full information.

If one nonetheless regards estimated results on the basis of imputed data as a reflection of the real world, it is a biased reflection. Politically ignorant respondents are ‘transformed’ into knowledgeable respondents and therefore underrepresented in the sample. Selection bias of the listwise deletion is replaced by selective misspecification bias of the imputation of incomplete data. The latter usually results in exaggerated effects. If there is an effect between evaluation of Romano Prodi and the overall evaluation of the EU for informed respondents, this relationship will be replicated also for respondents who state not to know Romano Prodi. But in my opinion this result is artificial. How can the evaluation of Romano Prodi have a positive/negative effect on the evaluation of the EU in the group of respondents who never have heard of this politician before?

This section so far establishes that existing imputation techniques do not apply in the described situation in which respondents lack information with regard to the stimulus of a survey question. But if not \( \hat{Y} \), what else to impute? What kind of information is present in valid ‘don’t know’ answer that allows the analyst to relate a ‘don’t know’ answer to the analysable response categories of a survey question and consequently to analyse all cases in a statistical analysis?

I argue that a lack of information (a valid ‘don’t know’ answer) is resembled best by a random answer on the response scale. If one would force respondents who do not know Romano Prodi to give an answer on the sympathy scale anyway, one usually would expect a random answer. Thus, what is in fact unobserved by a ‘don’t know’ answer in this specific case is an arbitrary answer on the survey question. I therefore assume that a substitution of random values for incomplete data is what comes closest to the information provided by a ‘don’t know’ answer that originates from insufficient knowledge about the stimulus of the question.
But how are these random responses most likely distributed? First, it seems plausible to retrain the metric of the answer scale. If respondents had been forced to give an answer they would have done this using the answer categories offered. Second, the likelihood with which they would have chosen certain answer categories is unknown.\textsuperscript{211} Different distributions of the random variable are possible and reasonable, though for practical reasons I propose the empirical univariate\textsuperscript{212} distribution in the complete cases. The

\textsuperscript{211} Preferably, one would like to have information from experimental survey research on response functions for different ‘fake’ questions. In other words, how do respondents in an experiment evaluate a politician who does not exist (and the analyst forces them to give an answer)? Where do respondents place themselves on counterfeit issue scales? The problem of such an approach is, however, that such answers may depend heavily on cues provided in the specific stimulus. For instance, even if respondents do not know Romano Prodi, in a forced evaluation they will presumably consider information that is provided in the question, namely, that it is a male politician with an Italian name. Such information is likely to affect the distribution of responses. If the experimental knowledge on response functions on a like-dislike scale is based on a stimulus that refers, for instance, to a female politician with a Dutch name, the response function may look different. Hence, ‘forced’ responses may be highly sensitive to cues given in the specific wording of the question and experimental knowledge on such response functions may thus be difficult to generalize.

\textsuperscript{212} One could equally well select the conditional, or multivariate, distribution in the observed cases. That is to say, one takes into account the response distribution for ‘related’ cases in the substitution process of missing information. Suppose, for instance, that women evaluate Romano Prodi more negatively than men do, thus the distribution of the evaluation of Romano Prodi differs between both groups. One could use these two distributions for the generation of two random variables and replace missing cases for women and men separately. The difference between univariate and conditional distributions will for all practical purposes be a minor one. Only if subgroups of the sample significantly differ in their response function \textit{and} the proportion of missing cases significantly differs between such groups \textit{and} the proportion of missing cases in the sample is very high, the difference between a univariate and conditional random distribution alters the total variance of an imputed variable. The question is, however, if one wants to use information on response functions in
randomly imputed values thereby become a random draw from the complete cases. This has useful implications for the estimations of effect parameters based on imputed datasets, as I will explicated in the following section. The proposed white-noise imputation of missing scores, $D_{mis}$, that draws random values from the univariate observed distribution, $D_{obs}$, can be formulated as

$$D_{mis} \sim f(D_{obs} | \bar{\mu}, \bar{\sigma})$$  \hspace{1cm} (A1.1)

If one substitutes values for incomplete data that are the result of one random draw, computed standard errors are underestimated (Allison 2002). Randomly imputed scores are treated as if they are observed. Thereby one disregards that a random draw does not deterministically retrieve the properties as defined, but only with a certain probability. The possibility of falsely relying on an outlier-draw introduces uncertainty in statistical models which base on randomly imputed data. To control for this uncertainty and to estimate more adequate standard errors, Rubin (1987) suggests the creation of multiple imputed datasets. This practice is habitually applied in other imputation techniques that rely on a random component (Schafer 1997; King et al. 2001; Allison 2002). In each of these generated multiple datasets, the observed data are identical, as is the algorithm that leads to random draws. However, each random draw generates possibly different scores. Parameters (mean, regression coefficients, etc.) estimated on imputed datasets will also be

specific groups of the sample in the first place. After all, the proposed white-noise imputation rests on the notion that one ought not always transfer patterns in the observed cases to incomplete information. To use conditional distributions presumes a certain response from respondents if they would have been familiar with the stimulus. In the example, women would have evaluated Romano Prodi more negatively than men if they had known him. Hence, it appears from a conceptual point of view more warranted to rely on less information in the observed cases (univariate distribution) when generating random values for incomplete data.
slightly different across multiple datasets. On the basis of these \( m \) simulated datasets one estimates \( m \) parameters \( p \) of interest. The mean point estimate \( \bar{p} \) over all \( m \) simulated datasets can be calculated as

\[
\bar{p} = \frac{1}{m} \sum_{j=1}^{m} p_j
\]  

(A1.2)

To the extent that these results vary over \( m \) datasets, one estimates standard errors of parameters \( p \) in the overall model incorporating a punishment function. Hence, the standard error of the point estimate \( \bar{p} \) derives from the mean estimated variance of point estimates within each dataset and the sample variance in the point estimate across simulated datasets, correcting for small numbers of \( m \).

\[
SE(\bar{p}) = \frac{1}{m} \sum_{j=1}^{m} SE(p_j) + \sqrt{VAR(p_j)(1 + \frac{1}{m})}
\]  

(A1.3)

Although there is of course no limit to the number of generated datasets, five replications have proven to be sufficient for most applications (Allison 2002).\(^{213}\)

A1.4 Consequences of a Multiple White Noise Imputation

What are the consequences of a white-noise imputation of incomplete data for the analysis of substantive research questions? Suppose one aims to explain variable \( Y \) (overall

---

\(^{213}\) Note that this procedure of a white-noise imputation is meant to apply to sufficiently large samples. A case-specific descriptive investigation of imputed datasets becomes meaningless since one substitutes arbitrary values for missing scores.
evaluation of the EU) by variable \( X \) (evaluation of Romano Prodi) in an ordinary least squares regression. One of the main interests will usually be the parameter estimate \( b \) in this regression. Recall that the least squares slope vector \( b \) estimates as

\[
b = [X'X]^{-1}[X'Y]
\]  

(A1.4)

Suppose furthermore that respondents often are not sufficiently familiar with the stimulus of \( X \) (Romano Prodi) and one therefore encounters many missing scores. Let \( X_{\text{obs}} \) denote observed values and \( X_{\text{mis}} \) missing values on \( X \). Splitting between complete and missing scores leaves the slope estimate as

\[
b = \left[\begin{bmatrix} X_{\text{obs}}' \cr X_{\text{mis}}' \end{bmatrix} \begin{bmatrix} X_{\text{obs}} \\ X_{\text{mis}} \end{bmatrix} \right]^{-1} \begin{bmatrix} X_{\text{obs}}' \\ X_{\text{mis}}' \end{bmatrix} \begin{bmatrix} Y_{\text{obs}|x_{\text{obs}}} \\ Y_{\text{obs}|x_{\text{mis}}} \end{bmatrix}
\]  

(A1.5)

These missing values are randomly imputed as suggested (equation A1.1). Hence I substitute \( D_{\text{mis}} \) as a random draw from the observed data as defined for \( X_{\text{mis}} \).

\[
b = \left[\begin{bmatrix} X_{\text{obs}}' \\ D_{\text{mis}}' \end{bmatrix} \begin{bmatrix} X_{\text{obs}} \\ D_{\text{mis}} \end{bmatrix} \right]^{-1} \begin{bmatrix} X_{\text{obs}}' \\ D_{\text{mis}}' \end{bmatrix} \begin{bmatrix} Y_{\text{obs}|x_{\text{obs}}} \\ Y_{\text{obs}|x_{\text{mis}}} \end{bmatrix}
\]  

(A1.6)

Note that \( D_{\text{mis}} \) is a random variable. The covariance between \( D_{\text{mis}} \) and \( Y \) is therefore approximately zero. Note also that I define the distributional properties of \( D_{\text{mis}} \) to follow the univariate distribution in \( X_{\text{obs}} \) (since \( D_{\text{mis}} \) draws from \( X_{\text{obs}} \)). Hence the variance of \( D_{\text{mis}} \) and \( X_{\text{obs}} \) are approximately the same. Equation (A1.6) can thus be reduced to
Consecutive steps from (A1.4) to (A1.7) point up two important implications of a white-noise imputation of missing data. First, imputing random values as proposed results in a parameter estimate that is approximately zero for respondents who do not have sufficient information with regards to $X$. The parameter estimate $b$ between the evaluation of Romano Prodi and the EU is approximately zero for respondents who do not know Romano Prodi. This seems a reasonable implication from what is known about these incomplete data. The evaluation of Romano Prodi cannot have affected the evaluation of the EU if respondents never heard the name of Romano Prodi before. The second implication regards the overall estimate of $b$ in the whole sample. Equation (A1.7) specifies that the estimation solely rests on the covariance in the observed cases but the variance in all cases. Hence, the size of $b$ in the whole sample is reduced according to the number of missing data in the sample. Incomplete data is in fact replaced by white-noise. As a consequence, fit statistics are also reduced according the proportion of missing cases. An example provided in the following section illustrates this point.

More generally, the multiple white noise imputation of incomplete data avoids problems related to data missing at random, MAR, when they are based on insufficient information. Practical limitation can be overcome by including all cases in the analysis. All variables of interest can be incorporated in the analysis, irrespective of how many missing cases one encounters. The estimation of statistical models therefore becomes more efficient. Moreover, the procedure prevents selection bias by generating relationships (white-noise) that are not imposed by patterns in the observed data.
A1.5 Example: Voting for GL in the Dutch Parliamentary Election of 1994

An empirical example of Dutch politics demonstrates the consequences of different missing data solutions (listwise deletion, multile EMis imputation, multiple white noise imputation) for the analysis of a substantive research question. The analysis is based on data from the easily accessible (ICPSR, Steinmetz Archive, ZA Köln) and well-documented (Anker & Oppenhuys 1997) Dutch parliamentary election study 1994 (DPES'94).

In the example, I aim to explain voting for the Green-Left party (Groen/Links, GL) in the parliamentary election of 1994. The dependent variable is a ten-point scale on which respondents indicate how likely it is that they will ever vote for this party (Tillie 1995; van der Eijk 2002; Kroh & van der Eijkk 2003). I include two independent variables in the ordinary least square regression. The first independent variable is the self-placement on a nuclear-plants-opinion-scale that ranges from 1 (more nuclear plants) to 7 (no nuclear plants). As reported in Table A1.1, 95% of the sample has an opinion on the issue of nuclear plants and gives an answer on the survey question. The second explanatory variable of voting GL is the like-dislike of one of the two party leaders of GL at that time, Mohammed Rabbæe. Mohammed Rabbæe was widely unknown in the electorate, 45% of the sample report not to know this politician. The result may not be surprising, keeping in mind that there are usually more than ten parties represented in the Dutch parliament and party leaders of small parties do not get as much medial attention as leaders of large parties. Moreover, Mohammed Rabbæe happened to be especially neglected by media during the election campaign (Brants & van Praag 1995). Asking the same question of like-dislike for the former prime minister (Ruud Lubbers) and the new prime minister elected in the 1994 election (Wim Kok) produces only
marginal non-response. More than 99% of the sample reports an opinion on both politicians. Hence, incomplete data on the second independent variable is in all likelihood caused by insufficient information and not by measurement problems.

Given a simple model that explains voting for GL by a vital issue for environmental parties, nuclear power plants, and the evaluation of the party leader of GL, Mohammed Rabbæe, a reasonable research question could be the following: What was more important for voting GL in 1994, the nuclear plant issue or the leadership of Mohammed Rabbæe? Many scholars would probably not approach the question on the basis of the given data with the argument that 47% of non-response comprises too much slippage to sensibly draw conclusions. Retaining these reservations, an OLS model based on listwise deleted data suggests that Mohammed Rabbæe was more important in voters' reasoning than the issue of nuclear plants (see estimated parameters and standard errors reported in the first column of Table A1.1). This result would probably be considered surprising for observers of Dutch politics. The estimated effects based on listwise deleted data are of course true, as long as one is interested in the group of respondents who happen to know Mohammed Rabbæe. But since one is usually interested in the whole sample, findings are more than doubtful. The results also appear suspicious to someone who is not familiar with Dutch politics. It is implausible to predict a strong effect for the evaluation of Mohammed Rabbæe for the whole sample if this variable causes 45% non-response.
If one includes missing data in the analysis by an advanced imputation technique, one obtains results reported in the second column. This imputation solution is based on the importance weighted Expectation Maximization (EMis) algorithm as proposed by King et al. (2001). I generate multiple datasets using AMELIA software developed by Honaker et al. (1999). Without going into details of the procedure it may be noted that there is some agreement among experts, that this is one of the recommended solutions to the missing data problem (Allison 2002). Besides variables of the model, additional variables (age, sex, education, interest in politics) are included in the imputation process. The presented results are based on five simulated datasets. Results of the EMis imputation illustrate that such imputation methods reduce standard errors of the parameter estimates by including all cases in the analysis. From a substantive point of view, however, one finds rather similar effect parameters than under listwise deletion. The counterintuitive result is confirmed that Mohammed Rabbæ was more important for voting GL than nuclear plants. What one estimates is how the world would look like if all respondents had known Rabbæ.

The third column of Table A1.1 reports results based on a multiple white-noise imputation of incomplete data, which is also based on five simulated data sets. Again, standard errors are smaller compared to the solution based on listwise deleted data but effect parameters and the goodness
of fit statistic are also. According to the number of missing cases (few for the first and many for the second independent variable) estimated parameters decrease in size. These results correspond with what one intuitively would expect: the evaluation of Mohammed Rabbae is less important for voting GL than the issue of nuclear plants in the whole sample. The low $R^2$ also reflects more accurately what one in fact knows about the sample. One has in fact less information than the variance reduction in the first two columns suggests. Given the set of independent variables included in the model, the solution of a multiple white-noise substitution realistically reflects the fact that explanations do not apply to almost half of the sample.

The example illustrates that different treatments of missing data considerably affect substantive findings. Clearly, results are different, yet one cannot judge from expectation and prior knowledge which result is correct. The question, which approach to choose, has to be decided already before the imputation of missing data. The basis of this decision lies in the applicability of assumptions underlying different methods.

A1.6 Conclusion: Choosing a Method

Each solution to the missing data problem rests on assumptions about these data. The application of different methods therefore hinges upon the correctness of assumptions, something that has to be judged for each variable in the dataset separately. To delete all missing data without considering consequences is likely to produce false results. But the idea to throw everything in the imputation software and to produce thereby a ‘cleaned’ dataset is also problematic.

A listwise deletion of incomplete data is a sensible strategy if these scores are missing completely at random,
MCAR. If, however, one finds that the likelihood of incomplete data depends on variables in the statistical models, MAR, one generates selection bias by listwise deleting all missing cases. The listwise deletion of incomplete data becomes inappropriate because the underlying assumption does not hold.

If one encounters data missing at random, MAR, and one is confident that measurement problems cause such missing data (a true score for each missing value exists), existing imputation techniques are the proper choice. Such methods apply in all those cases in which respondents are asked to report facts like demographics or past behaviour. Irrespective of whether respondents answer questions on their age or education, they nonetheless have a certain age and education. Also, when respondents state not to know whether they attended religious services, they do or do not go to church.\(^{214}\) For a number of variables that do not concern facts, there exists established evidence that measurement problems often generate non-response. The question of intended electoral participation can be named in this respect (cf. Tourangeau 2000). It seems plausible to assume that all respondents have an idea what is meant by this question and whether or not they consider to cast a ballot. Missing values on this question most likely reflect measurement problems.

The multiple white noise imputation is the preferable choice if one knows that the assumption of insufficient information holds. When respondents are asked to report their opinion (rate/evaluate/prefer/etc.) on an external stimulus one has to take into consideration that the stimulus of the question is possibly unknown to many respondents. If the likelihood of measurement problems is low but many respondents are in all probability truly not familiar with the stimulus, a multiple-

\(^{214}\) Accordingly, I impute incomplete data on the social background of respondents, such as age, church attendance, rural/urban residence, union membership, and education applying an EM algorithm in the analysis of this study. See Appendix 3.
white noise imputation is a sensible solution. In the CSES dataset the variables of leader evaluation and the self and the perceived party placements on the left-right scale fall into this category. It requires a considerable amount of political information to evaluate each party leader in a political system or to place all parties on the left-right scale. In most cases, respondents report a leader evaluation or a party placement for governmental or larger parties but fail to do so for minor parties. This seems to reflect a legitimate absence of political information and not measurement problems. I therefore impute missing scores on these two variables applying a white-noise imputation.

In many situations one will not know for sure what caused certain scores to be missing. Research on survey response should indicate how likely measurement problems are for different variables. If one does not have enough information to reject or support either the application of existing imputation techniques or the application of a multiple-white noise imputation, the latter may be regarded as the more conservative method. As illustrated in the empirical example, statistical models based on multiple white-noise imputations tend to decrease effect parameters. Methods of data augmentation, conversely, tend to increase effect parameters (or at least to generate similar ones than in the observed cases). Hence, the error one makes when applying falsely existing imputation techniques is to overestimate a relationship that does not exist. The other case, where a multiple white-noise imputation is applied although there is a relationship underlying incomplete data, means to underestimate relationships in the whole sample. The choice for a white-noise imputation is therefore the more conservative approach since it tends to underestimate effect parameters.