Missing values in exploratory factor analysis

A ‘best of all possible worlds’ approach to imputation for incomplete survey data

Bas Bosma^{*,1} and Arjen van Witteloostuijn^{1,2}

¹ VU Amsterdam, School of Business and Economics, the Netherlands

² University of Antwerp / Antwerp Management School, Belgium

January 2021

* Corresponding author: De Boelelaan 1105, 1041 HV Amsterdam, the Netherlands. Email:

b.bosma@vu.nl.

Missing values in exploratory factor analysis

A ‘best of all possible worlds’ approach to imputation for incomplete survey data

Abstract. In the social sciences, multi-item scales and factor analyses are standard tools in survey research. In the social sciences, such tools are omnipresent, as are, unavoidably, nonresponses. The question is how to handle missing values when an exploratory factor analysis is intended. Deletion methods will result in — oftentimes substantial and damaging — reduction of power. The seemingly obvious alternative is to keep all respondents and apply imputation to missing values. However, with the true factor structure unknown, theoretically recommendable multiple imputation methods cannot simply be applied. Instead of declaring an entire method unsuitable for exploratory analysis, we propose an approach that keeps the relevant aspects of various methods and combines these by sacrificing less relevant aspects. Doing so, we keep understanding and ease of use in mind, aiming for an approach that is more rigorous and ‘correct’ than what is commonly used in practice, whilst still being straightforward enough to actually be used.

Keywords. Imputation, missing data, survey, exploratory factor analysis.

1 Introduction

A shared goal in the empirical social sciences is the discovery of concepts by which our social world can be understood. When concepts cannot be directly observed, like assumed abilities, researchers typically try to design an indirect measurement tool. Arguably one of the most common of these in the applied social sciences is the survey. Generalizing, there are two research strategies in which surveys figure prominently. The first is confirmatory. It starts by designing a survey containing questions believed to be related to the assumed ability, to then tests whether responses by individuals are indeed consistent with the hypothesized underlying structure. The second is more exploratory. It starts from the multidimensional responses to a survey, and subsequently tries to infer whether one or more underlying dimensions could account for the response pattern actually observed. Whilst the first strategy builds upon theory and tests the theory's prediction, the second strategy starts from an exploratory position without a priori hypotheses regarding the (number and meaning of) factors that might be present and their structure (Bartholomew, 1980).

Whilst surveys are a ubiquitous instrument in the applied social sciences, the analysis of survey data is still clouded by widely varying methodological recommendations (e.g. van der Eijk & Rose, 2015). On top of that, the applied researcher using survey data is often confronted with challenges even before the analysis can start, especially in exploratory research where two of the most familiar pre-processing challenges are small sample size and missing data. Often performing their exploratory analyses on primary data, the applied researcher probably has to make due with small sized samples. With classic recommendations on sample size perhaps best captured by Comrey and Lee (1992, p. 217), “50 – very poor; 100 – poor; 200 – fair; 300 – good; 500 – very good; ≥ 1000 – excellent,” the rule of thumb for applied researchers seems to have become that a sample size of 200 suffices for most descriptive

analyses, while reaching 500 or more respondents is strongly recommended (MacCallum et al., 1999).¹

Russell (2002), however, investigated all articles appearing in the *Personality and Social Psychology Bulletin* over a three-year period, and found that 27% of all articles explore underlying latent structures in survey data, of which 39% on sample sizes of 100 or smaller and another 23% on sample sizes between 100 and 200. Fabrigar et al. (1999), in their review of contributions to the top-ranking psychology outlets *Journal of Applied Psychology* and *Journal of Personality and Social Psychology*, comparably found that 50% of the studies exploring latent structure in survey data had sample sizes below 200, whilst Conway and Huffcutt (2003) report very similar statistics for organizational research outlets. In organizational research, gaining access to large samples is often even more problematic.

Struggling with the recommended sample sizes even when the data would be complete, the applied researcher additionally has to come to terms with survey data being plagued by missing data. King et al. (2001) conclude that, on average, half the respondents to surveys do not answer one or more questions. In much organizational research, we suspect, this 50% would be regarded to be very good. Obviously, addressing this issue of missing data in the already small sized samples is of key importance. It can, quite literally, make the difference between reaching insightful conclusions and having to conclude that no insights of any significance can be extracted from the data collected. And yet, although few will disagree with the previous statement, it is precisely the standard practice of many applied researchers in addressing missing data that seems to be below par.

¹ Note that the use of sample size or sample size to estimated parameter ratios can never be more than a very rough indication. The strength of the relations among variables is known to be far more salient in determining the trustworthiness of the estimated factor structure in relation to the population model. See, e.g., the simulation results in McNeish (2017).

Notwithstanding the extensive literature on missing data techniques (cf. Allison, 2003; Graham, 2009; Little, 1988; Little & Rubin, 2014; Raghunathan et al., 2001), the most common methods of dealing with missing data remain the very straightforward and thoughtless listwise and pairwise deletion strategies. Reviewing over 800 articles in political science journals, King et al. (2001) find 94% of the studies dealing with missing data to do so by listwise deletion, possibly after partially replacing missing values by ‘educated guesses.’ The loss of information in doing so is, on average, about a third of the information collected. Furthermore, even assuming that the educated guesses are correct, on average, the certainty with which we know the values we substitute for missing data is still grossly overestimated. The common way of dealing with missing data can thus be characterized as “a loss of valuable information at best and severe selection bias at worst” (King et al., 2001, p. 49).

Our first aim is to review and assess the suboptimal practices of applied social scientists in dealing with missing data. We specifically focus on survey data consisting of Likert-type items on which exploratory factor analysis (EFA) is performed in search for latent structure.² We will explicitly take account of both the practical aspects of doing applied survey research, as well as the benefits of addressing missing data using the more advanced techniques available in the statistical sciences. Our second aim is to suggest a new approach that aims to be ‘best of all possible worlds’ – that is, more sound than current practices by leveraging more advanced missing data methods, but also adapting these to the actual context and way of working of the applied social scientist, including the organizational researcher. We hereby hope to contribute to a common and significant problem by offering an approach that can guide and improve the quality of future applied social science research by being used in practice.

² Our focus should not be read as a methodological recommendation of our own; it merely reflects the fact that EFA is very popular in applied social science research, and missing data is an especially acute problem in this context.

2 Factor analysis

Many surveys (partly) consist of sets of Likert-type items, and although data of this kind can be analyzed by exploratory variants of Structural Equation Modelling (SEM) or Item Response Theory (IRT), to name just a few alternative techniques, the more popular method in search for latent structure seems to be EFA (see, e.g., Alexandre et al., 2020; de Boeck & Wilson, 2004).³ Before introducing our approach to handling missing data in the context of EFA, some background and technical detail regarding factor analysis will help in understanding the challenges of addressing missing data in the context we focus on.

The roots of factor analysis can be traced to the work of Spearman (1904). In his study of human intelligence, he administered a battery of mental tests to a group of individuals. Analyzing the results, he observed that the score of a given individual on a certain test T is more related to that individual's scores on other tests than to the scores of other individuals on other tests. Put differently, the scores of individuals often are interrelated to a certain degree, which can be made sense of by assuming that an individual's test score on a given test is composed of a part peculiar to that test and a part that is a function of more fundamental quantities. As this latter function of more fundamental quantities, also called the *systematic part*, enters into the scores of that individual on different tests, this is what accounts for the interrelation. In Spearman's original 'single factor theory' of human intelligence, he tried to explain the observed relationship in test scores by a single underlying ability he called 'general intelligence' (g). Not being directly observable or measurable, Spearman posed that if g

³ EFA is often confused with, for example, principal component analysis (PCA) and confirmatory factor analysis (CFA); see, e.g., Bandalos and Boehm-Kaufman (2009) or Russell (2002). EFA is used when exploring possible underlying structures, and CFA when testing the fit between hypothesized structure and data. PCA is a more general technique for reducing the dimensionality of data by creating weighted composites of the observed data. In itself, PCA is not causal, whilst factor analysis is explicitly interested in latent variables *causing* covariation. To this end, on the one hand, factor analysis analyzes the variance shared by observed variables (*communality*) and tries to account for communality among a set of items by linear relations to the latent dimensions. PCA, on the other hand, analyzes all of the variance, including what is not shared with other observed variables (*communality* + *uniqueness*).

existed, partialling it out of the observed scores by means of regression would remove all the partial covariances between the observed scores (Anderson, 2003).

Generalizing, assume we have a set of *manifest* variables that we observe (e.g., our scores to Likert-type items in a survey). Denote these by a $(p \times 1)$ vector $\mathbf{x}$. Factor analysis is based on a model that assumes the manifest variables to be related to a linear combination of *latent* variables that we do not observe, and an error unrelated to this systematic part. Denote the latent variables by a $(q \times 1)$ vector $\boldsymbol{\phi}$. The aim of factor analysis can then be stated as the approximation of the p -variate observed data, without significant loss of information, by a much smaller q -variate latent representation.

Manifest and latent variables are stochastically related, expressed by a conditional probability function $\eta(\mathbf{x}|\boldsymbol{\phi})$.⁴ If we let $f(\mathbf{x})$ be the joint distribution of the manifest variables, and $g(\boldsymbol{\phi})$ that of the latent variables, their relation can be expressed as:

$$f(\mathbf{x}) = \int_{\mathfrak{R}} \eta(\mathbf{x}|\boldsymbol{\phi}) g(\boldsymbol{\phi}) d\boldsymbol{\phi},$$

where $\mathfrak{R}$ is the range space of the latent variables. As the researcher actually observes the manifest variables, this allows inference about the latent variables. Or, more formally, we can infer about the conditional probability function of the latent variables based on what we know of the joint distribution of the manifest variables. The conditional probability function we want to infer can be expressed as follows:

$$g(\boldsymbol{\phi}|\mathbf{x}) = \frac{g(\boldsymbol{\phi})\eta(\mathbf{x}|\boldsymbol{\phi})}{f(\mathbf{x})}. \quad (1)$$

⁴ For the more common Likert-type survey data, this will be a probability mass function; for continuous data, it will be a density function.

In order to turn this inference into an applicable model, we need two crucial assumptions. The first simplifies the joint distribution of the latent variables by assuming their independence:

$$g(\boldsymbol{\phi}) = \prod_{i=1}^q g(\phi_i). \quad (2)$$

Subsequently, given that the latent variables are almost always mental constructs for understanding instead of them being directly measurable variables, we can further simplify their joint distribution. It is very uncommon to, for example, design a survey asking individuals all sorts of questions related to lifestyle, possession of luxury items and the like, and using the collected data to infer an underlying latent variable ‘wealth’. The assumed latent variable in this example is itself directly measurable, although possibly sensitive. Given that the latent variables we try to infer are almost always mental constructs, like abilities and attitudes, we can arbitrarily assume a distribution for the latent variables that suits us in constructing a model. That is, we can assume the $g(\phi_i)$ to be, for example, uniformly distributed, hereby replacing the abstract range space $\mathfrak{R}$ by well-defined integrals over the (0,1) interval for each of the latent variables.⁵

The second assumption concerns the conditional independence of $\eta(\mathbf{x}|\boldsymbol{\phi})$ – that is,

$$\eta(\mathbf{x}|\boldsymbol{\phi}) = \prod_{i=1}^p \eta_i(x_i|\boldsymbol{\phi}). \quad (3)$$

This assumption captures the crucial rationale for factor analysis, being that the observed interrelation in the manifest variables can be expressed by their dependence on the latent variables. Without the conditional independence of $\eta(\mathbf{x}|\boldsymbol{\phi})$, it would not be possible to describe the covariation in the manifest variables by a much smaller set of latent variables.

⁵ We can choose any distribution as long as any monotonic transformation of the chosen distribution would work as well.

Under assumptions (2) and (3) above, factor analysis turns into an intuitive and applicable model. As discussed, the manifest variables are partitioned into a systematic part, being a linear combination of the latent variables, and an error:

$$\mathbf{x} = \boldsymbol{\alpha} + \mathbf{\Lambda}\boldsymbol{\phi} + \mathbf{u}. \quad (4)$$

Going back to Spearman's original research, the manifest variables can be the scores on each of the p administered tests. The $(p \times 1)$ vector $\boldsymbol{\alpha}$ would then contain the average scores on each of the tests in the population. The identically shaped vector $\mathbf{u}$ is composed of a true measurement error in each of the tests, as well as a *unique factor* particular to each test. However, as we cannot distinguish between these components based on the survey data collected, we jointly refer to them as 'error' and think of the elements of $\mathbf{u}$ as the parts of the test scores not explained by the systematic part.

Finally, the systematic part itself is a linear combination of the latent variables, with the coefficients, called the *factor loadings*, being the elements of the $(p \times q)$ matrix $\mathbf{\Lambda}$. Assuming the error to be distributed independently of the latent variables, with zero mean and diagonal $(p \times p)$ covariance matrix $\boldsymbol{\Psi}$, and the latent variables to be distributed with mean zero and $(q \times q)$ covariance matrix $\boldsymbol{\Phi}$, the $(p \times p)$ covariance matrix $\boldsymbol{\Sigma}$ of the manifest variables can be written as:⁶

$$\boldsymbol{\Sigma} = \mathbb{E}(\mathbf{x} - \boldsymbol{\alpha})(\mathbf{x} - \boldsymbol{\alpha})^T = \mathbf{\Lambda}\boldsymbol{\Phi}\mathbf{\Lambda}^T + \boldsymbol{\Psi}. \quad (5)$$

Intuitively, factor analysis thus tries to approximate the observed covariance matrix of the manifest variables, $\boldsymbol{\Sigma}$, using a small set of q latent variables or *factors*. Factor analysis, in all its variants, can thus be said to operate on the covariation between survey items, trying to

⁶ More formally, $\mathbb{E}\boldsymbol{\phi}\mathbf{u}^T = 0$, $\mathbb{E}\mathbf{u} = 0$, $\mathbb{E}\mathbf{u}\mathbf{u}^T = \boldsymbol{\Psi}$, $\mathbb{E}\boldsymbol{\phi} = 0$, and $\mathbb{E}\boldsymbol{\phi}\boldsymbol{\phi}^T = \boldsymbol{\Phi}$. The structure of latent variables is then said to 'fit the q -dimensional common factor model' when there exists a diagonal matrix $\boldsymbol{\Psi}$ with nonnegative diagonal entries such that $\boldsymbol{\Sigma} - \boldsymbol{\Psi}$ is positive semidefinite of rank q . If, for whatever reason, $\mathbb{E}\boldsymbol{\phi} \neq 0$, the expected value of the manifest variables would be $\mathbf{\Lambda}\mathbb{E}\boldsymbol{\phi} + \boldsymbol{\alpha}$. It would always be possible to redefine $\boldsymbol{\alpha}$ to absorb $\mathbf{\Lambda}\mathbb{E}\boldsymbol{\phi}$ in this case. In case the latent variables are furthermore *orthogonal* with unit variance, the approximation simplifies further to $\boldsymbol{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}^T + \boldsymbol{\Psi}$.

uncover a lower dimensional structure able to reproduce the observed covariation in manifest variables with as little loss of information possible.

3 Missing data and EFA

Introducing some more notation, let $\mathbf{D}$ denote the data matrix. For a typical survey, this matrix could consist of a row for each respondent that (partly) completed the survey, and a column for each of the questions or items. In the presence of missing data, $\mathbf{D}$ has an observed and a missing part. In other words, for all cells in the data matrix, we assume values ‘exist’ (in a specific metaphysical sense), yet some are not ‘observed’.⁷ Denote this by $\mathbf{D} = \{\mathbf{D}^{obs}, \mathbf{D}^{mis}\}$, with $\mathbf{D}^{mis}$ the values assumed to exist but not observed. Also, let $\mathbf{R}$ be a matrix of identical dimensions to the data matrix containing response indicators. That is, r_{ij} is equal to 1 if d_{ij} is observed, and equal to 0 if it is missing.

The main missing data mechanisms can now be described in terms of our ability to predict the values of $\mathbf{R}$ (King et al., 2001). Missing data is *missing completely at random* (MCAR) if it cannot be predicted any better by making use of information in $\mathbf{D}$. More formally, the response matrix is independent of the data matrix: $\mathbb{P}(\mathbf{R}|\mathbf{D}) = \mathbb{P}(\mathbf{R})$. The probability of missing data on a variable is, under MCAR, unrelated to both the variable with missing data itself, as well as any of the other variables on which data is collected. This could, for example, happen if individuals make the decision to answer questions by flipping a coin, or if the missing data results from filled out surveys that went missing in the mail. In practice, missing data will

⁷ Whilst some individuals might prefer to not disclose their income, they do have an income (positive or negative). This is what is meant by ‘exist’. If a survey contains specialist questions whilst also being conducted outside of the matching population of specialists, it may, of course, be the case that a ‘don’t know’ does not signal a preference to not reveal the true existing value, but actually signals the question being inapplicable (Kroh, 2006). Assigning all individuals an opinion, even on questions that might be inapplicable to them, causes an underrepresentation of the less informed and introduces a bias.

almost never be MCAR; rather, not answering a question is often related to some other characteristic in the survey.

Missing data is *missing at random* (MAR) if the values in $\mathbf{R}$ cannot be predicted any better by making use of information in $\mathbf{D}^{mis}$, but might be by using information in $\mathbf{D}^{obs}$. That is, $\mathbb{P}(\mathbf{R}|\mathbf{D}) = \mathbb{P}(\mathbf{R}|\mathbf{D}^{obs})$. The probability of missing data on a variable is, under MAR, related to other variables on which data is collected, but not to the variable with the missing data itself. If, for example, individuals with a more right-wing political preference are less inclined to answer questions on income (resulting in relatively more missing data) than individuals with a more left-wing political preference, and this (difference in) preference can be predicted by making use of other observed data like level of education or issue positions, the missing data would be MAR.⁸ Or, even simpler, if young people do not answer questions on their political preference, the missing data in political preference is related not to political preference itself, but to another variable (age) on which data is collected. The missing data in political preference can thus be predicted by making use of data collected on other variables.

Finally, missing data is *missing not at random* (MNAR) if the response matrix depends on the data matrix, both the observed and the missing part; $\mathbb{P}(\mathbf{R}|\mathbf{D})$ does not further simplify. The probability of missing data on a variable is related, even after controlling for all other variables on which data is collected, to that variable itself. A classic example of MNAR would be high-income individuals refusing to answer questions about income while simultaneously being unable to predict which individuals have high income based on all other data collected (King et al., 2001; Little & Rubin, 2014; Rubin, 1976).

⁸ The prediction requires no causality. If the response pattern in $\mathbf{R}$ can be better predicted by making use of the information in $\mathbf{D}^{obs}$, in any way whatsoever, the missing data is MAR.

3.1 *Deletion methods*

The common practice of deleting individuals that did not answer one or more questions (listwise or pairwise) will bias results for missing data that is MAR or MNAR.⁹ Intuitively, only if data is missing completely at random, is the remaining data after deletion a random sample of the survey data originally collected. However, even in case of MCAR, deletion methods are inefficient as they disregard part of the actually observed data, and thus use only part of the information. As quantified by King et al. (2001), even under the optimistic assumption of missing data being MCAR, the efficiency loss of using deletion methods is, on average, one standard deviation. That is, the point estimates reported in many applied social science studies are, on average, about one standard deviation farther away from their true population value due to the common practice of using deletion methods in addressing missing values. Under less optimistic conditions, their inefficiency will be more severe, and their results will also be biased.

Inference based on pre-processing by imputation methods, in contrast, are more efficient under all possible types of missing data, and unbiased under MCAR or MAR. Under MNAR, both deletion and imputation methods can be biased, and additional steps must be taken.¹⁰ However, abandoning the common practice of deletion methods and moving towards imputation methods clearly improves the results of analyses in almost all circumstances, and will almost never make them worse.¹¹ Safe to say, applied social scientists should move away

⁹ Whether or not missing data is MCAR can be tested for. See, for example, Jamshidian and Jalal (2010). In practice, MCAR will almost always be rejected in favor of MAR/MNAR. Given its dependence on missing data, the presence or absence of MNAR can never be demonstrated based only on observed data.

¹⁰ Generally speaking, exploratory research is part art, part science. Existing formal tests on the type of missing data should, of course, be used. Additionally, though, first-hand knowledge about the data collection methods and the data itself will be very important in making the right choices.

¹¹ Deletion methods can outperform imputation methods, but this only happens under very unusual configurations of assumptions, methods, and analysis models. See King et al. (2001) for further details.

from their current most popular practice of addressing missing data. However, what should they do instead, as imputation comes in many different forms and shapes?

3.2 Imputation

Imputation methods fill missing data with ‘sensible’ or ‘likely’ values to create a completed data set that can be analyzed using traditional methods. Contrary to deletion methods, imputation makes use of all actually observed data and is thus more efficient, no matter the type of missing data. Imputation methods can be *single* or *multiple*. The former creates a single completed data set by filling in missing data, whereas the latter creates multiple completed data sets by filling in multiple likely values for each d_{ij} for which $r_{ij} = 0$. Whilst single imputation can lead to approximately unbiased point estimates, it always underestimates the uncertainty of imputed values (e.g., Enders, 2010). Intuitively, imputed values are treated just like observed values, whilst they are, of course, nothing but likely values with inherent uncertainty. As multiple imputation (MI) does account for the random fluctuations between each imputation, it allows for accurate inferential conclusions and is, in nearly all respects, preferred over single imputation.

MI can be applied using a wealth of methods and techniques. Generally speaking, MI consists of three stages: imputation, analysis, and pooling. In the first stage, m completed data sets are created by predicting each d_{ij} for which $r_{ij} = 0$. The observed d_{ij} , for which $r_{ij} = 1$, are the same in each of the m completed data sets. We thus end up with $\mathbf{D}_{(m)} = \{\mathbf{D}^{obs}, \mathbf{D}_{(m)}^{mis}\}$, where $\mathbf{D}_{(m)}$ are the m completed data sets, and $\mathbf{D}_{(m)}^{mis}$ are the m sets of predicted values for the missing data. The better the prediction model performs for a given imputed d_{ij} , the smaller its variation across $\mathbf{D}_{(m)}^{mis}$.

In the second stage, an analysis is performed on each of the completed data sets $\mathbf{D}_{(m)}$, resulting in m quantities of interest (like a matrix of factor loadings, a regression coefficient, or some other statistic). Let the quantity of interest be denoted by Q , and let $\hat{Q}(\cdot)$ be an estimator. Furthermore, let U denote the sampling variance of Q , with $\hat{U}(\cdot)$ being its estimator. For each completed data set $\mathbf{D}_m \in \mathbf{D}_{(m)}$, let $\hat{Q}_m = \hat{Q}(\mathbf{D}_m)$ denote the estimate of the quantity of interest, and $\hat{U}_m$ the estimate of its sampling variance.

Finally, in the third stage of MI, the final inference is arrived at through pooling the m results. The final quantity of interest, the MI point estimate $\bar{Q}$, is the average of the estimated quantities of interest on each of the m completed data sets:

$$\bar{Q} = \frac{1}{m} \sum_m \hat{Q}_m. \quad (6)$$

The variance of $\bar{Q}$ takes two sources of uncertainty into account. The first is the average of $\hat{U}_m$ over the m completed data sets, denoted by $\bar{U}$. This is the average of the sampling variances of each of the estimated quantities of interest *within* each completed data set $\mathbf{D}_m$. The second component is the excess variance due to the missing data being imputed (the variance *between* completed data sets):

$$\bar{B} = \frac{1}{m-1} \sum_m (\hat{Q}_m - \bar{Q})^2. \quad (7)$$

The total variance of the MI point estimate $\bar{Q}$ can be expressed as:

$$\bar{T} = \bar{U} + \left(1 + \frac{1}{m}\right) \bar{B}, \quad (8)$$

where the variance between completed data sets is multiplied by a factor that corrects for the fact that we have constructed a finite amount of completed data sets (Rubin, 1987).

MI is a general approach with obvious benefits. It can be used with various prediction models for the actual imputation. Although technically demanding, there are widely available (open source) software packages implementing MI. Furthermore, the technical sophistication

can indeed become quite extreme, but research has shown the exact distributional details to contribute less to the soundness of the outcome than the general approach. For example, Graham and Schafer (1999) and Schafer and Olsen (1998), amongst others, have shown MI models assuming the variables to be jointly multivariate normally distributed, an obvious approximation for survey data, to work almost as well as more sophisticated models specifically taking distributional assumptions for ordinal data into account. The improvements that can be realized versus the common practice of applied social scientists are, for the most part, related to using MI's general methodology, and only secondarily due to the sophistication of the prediction model used to create the completed data sets.

3.3 Challenges for EFA

The use of MI as a general methodology can be considered the de facto best practice standard. Purely seen from the statistical perspective of imputing missing data, few will advise against MI. And although there are numerous active discussions on the 'best' prediction model to be used within MI, which are likely to continue for years to come, applied social scientists should arguably be more focused on the use of MI as general methodology per se, and can rest assured that many mature implementations of MI exist in easy-to-use software.¹² However, notwithstanding this clear standard, not applying MI is the common practice in applied social sciences, including organizational research, which is certainly implies a suboptimal state of affairs.

There are a number of additional challenges when the intended analysis after addressing missing data is EFA. In order for the third stage of MI (pooling) to make sense, the estimated

¹² To name just two out of many: *Amelia* (Blackwell, Honaker, & King, 2017; Honaker, King, & Blackwell, 2011) and *MICE* (van Buuren & Groothuis-Oudshoorn, 2011). Both are readily available in the R ecosystem, mature, and extremely easy to use.

quantities of interest in the second stage (analysis) must be comparable. In other words, in terms of research strategy, MI seems to assume a theory being tested. The completed data sets created in the first stage should be analyzed using a single hypothesis. This hypothesis can be an assumed factor structure, like in CFA, or a regression model, or some other model, but it has to be the *same* hypothesis for all independent analyses on the m completed data sets. Only if this condition is met will the estimated parameters be comparable, and can they be pooled to obtain a single point estimate $\bar{Q}$ and its total variance $\bar{T}$.

EFA, however, has no single hypothesis. It commences without assuming any specific number of factors to be present and without having a clear-cut prior on the structure of the latent variables. For each of the $\mathbf{D}_{(m)}$, a different estimated covariance matrix $\hat{\Sigma}_{(m)}$ will result, possibly influencing the estimated quantities of interest and turning the subsequent pooling into comparing apples to oranges. More specifically, even when using the same approach to determine the number of factors that can be extracted, it cannot be guaranteed to give the same outcome for all $\hat{\Sigma}_{(m)}$. Furthermore, even if the number of factors would be the same, the order of the factors might not be. What we consider to be the first factor in the EFA performed on $\hat{\Sigma}_{(1)}$, might very well be the second factor in the EFA performed on $\hat{\Sigma}_{(2)}$, and vice versa. In pooling the factor loading matrices of these analyses, rather than properly taking the uncertainty of imputation into account, we would be averaging parameters related to different factors into a meaningless result.

4 Suggested approaches

Assume the applied social scientist has collected primary data using a survey. The sample size is just about what she needs to perform an exploratory analysis in search for a possible latent structure. However, if she would address the missing data by deletion methods,

the remaining sample would become too small for any sensible analysis. Additional data collection is not an option, so she must preserve all data collected and impute the missing values, ideally without introducing bias and correctly taking the inherent uncertainty of ‘likely’ imputed values properly into account. MI seems to be what is needed in terms of dealing with the missing data, but cannot be performed in combination with the EFA she intends to use in her analysis. How to proceed?

These problems are far from new, and many approaches have been suggested. So many even that we can only shortly mention a very few of them. A range of methods focuses on optimal estimation of the covariance matrix (on which EFA operates). Wold (1966), for example, already suggested more than five decades ago a nonlinear iterative partial least squares (NIPALS) estimation to determine the factors sequentially, even in the presence of missing data. Using the observed data, NIPALS tries to prevent singularity issues, like estimated covariance matrices not being positive-definite. Another decade-old suggestion is that of Kiers (1997), proposing an iterative principal components analysis (IPCA) method that includes a single imputation of the missing data in its operation. And full information maximum likelihood (FIML), somewhat similarly, tries to maximally incorporate information from all observed data by allowing the dimensions of the mean and covariance matrix to vary for individuals (cf. Enders & Bandalos, 2001).

Most of these methods, though, are as sensitive to the small sample sizes the applied social scientist has to work with as the intended methods for exploratory analysis. The estimates produced, even when missing data is MCAR, are unbiased only *asymptotically*,

implying that small samples still pose severe problems (cf. Schlomer et al., 2010).¹³ Furthermore, these methods are prone to overfitting – a problem becoming more severe in the context we focus on. Versions including regularization parameters (e.g., Ilin & Raiko, 2010; McNeish, 2015) can control the problem of overfitting to some degree, but require extensive tuning of the regularization parameter, with which the applied social scientist might very well be unfamiliar and/or uncomfortable.

Various other approaches suggest addressing the missing data separately at the level of the data set instead of within the covariance matrix estimation. Generally, these approaches attempt to get as close to the general methodology of MI as possible (for obvious reasons). Starting from a small sample that is also plagued by missing data, MI would first create multiple completed data sets. Performing the analysis on each of these would alleviate the issues of potential bias discussed above. Instead of iteratively estimating the covariance matrix based on the observed, small sample of data, the completed data sets have markedly higher dimension. The issues of possible singularity and small sample bias (even under MCAR) will be less acute. Furthermore, having multiple completed data sets provides the applied social scientist with a very intuitive representation of the inherent uncertainty in imputation that some of the iterative methods performing single imputation within estimation lack.

Ideally, MI incorporates the uncertainty all the way to the parameter level. That is, the effects of uncertainty are included in as many steps of the process possible, ideally only to be collapsed into a single point estimate at the stage of pooling. As this is not possible for EFA,

¹³ The same can be said for many prediction models used within MI. When using, e.g., predictive mean matching (PMM), where actually observed values from a ‘similar’ donor are imputed, small samples will result in a very small pool of donor cases that might not be that ‘similar’ to the individual for which missing data is imputed, or a single donor might be used on many occasions. Imputations using PMM for small samples may thus become less relevant and restricted in variation (e.g., Barnes, Lindborg, & Seaman, 2006; White, Royston, & Wood, 2011). For, e.g., the regressions of fully conditional specification or chained equations approaches, we might also have to cope, knowingly or unknowingly, with insignificant parameters and/or overfitting (McNeish, 2015).

due to parameter dimensions and order not necessarily being constant across analyses performed on the completed data sets, Dray and Josse (2014), for example, have suggested averaging the completed data sets. Besides the resulting data set being less intuitive (averaged values of $\mathbf{D}^{obs}$ remaining ordinal, whilst averaged values of the $\mathbf{D}_{(m)}^{mis}$ becoming real-valued), the uncertainty can be said to be collapsed almost before the MI process has started. EFA, operating on the covariation between survey items, is now presented with a single completed data set containing a mix of observed values and values that could not even have been observed. One can question whether the resulting estimated covariance matrix somehow incorporates the uncertainty in covariation between items. Intuitively, though, averaging completed data sets seems to be a suboptimal approach that does not provide a helpful representation of the uncertainty of imputation to the applied social scientist. And as intuition is important in adoption, this method does not seem to be prime candidate for widespread use within the applied social sciences.

Almost at the other end of the spectrum, Josse et al. (2011) and Lorenzo-Seva and van Ginkel (2016) have suggested an approach that tries to incorporate the uncertainty of imputation for as many stages of the MI process as possible by directly addressing the lack of comparability in parameters blocking the third stage of pooling. In their approach, the imputation and analysis stage of MI are still executed as would be in case of having a single hypothesis. However, to increase comparability of the m outcomes (like estimated factor loadings), an extra step is inserted right before the pooling stage. Using a rotation that is based on the generalized Procrustes rotation (see ten Berge (1977) and Lorenzo-Seva et al. (2002) for technical details), the matrices of factor loadings are made “as similar to one another as possible” (Lorenzo-Seva & van Ginkel, 2016, p. 599). After rotation, the parameters should be comparable enough to be pooled in the third stage of MI.

However, from the perspective of the applied social scientist, there are a number of challenges with this suggested approach. Firstly, one cannot ascertain if ‘as similar to one another as possible’ is indeed similar enough to warrant the pooling stage of MI. What is known about this rotation and its outcomes is furthermore described almost exclusively in highly technical outlets that might not be within reach of the average applied social scientist. Related to this, even the most user-friendly software implementations of the rotation require coding and some in-depth understanding of the underlying mathematics.¹⁴ Secondly, and more importantly, the actual doing of the factor analysis has to be standardized to some extent to make this approach practically applicable. As Lorenzo-Seva and van Ginkel (2016, p. 599) already mention: “It must be noted that the decision on how many r factors to extract (one factor for each latent trait) has to be the same for the K copies of data.” With precisely this number of factors that can be extracted seen as (one of) the most important decisions to be made in the doing of EFA, fixing this beforehand might be too restrictive and even disproportionate for what it makes possible in terms of addressing missing data (cf. van der Eijk & Rose, 2015; McNeish, 2017).

Furthermore, many applied social scientists will acknowledge the doing of an EFA to be ‘part art, part science.’ Oftentimes, the researcher will be, for example, comparing outcomes in terms of factors to retain using criteria ranging from scree tests and the Kaiser criterion, to parallel analysis and the domain expertise of the researcher herself. These will, most likely, come to different conclusions, and choices have to be made along the way. The same goes for the type of rotation to use (varimax, quartermax, oblimin, biquartimin, et cetera), possibly taking out cross-loading items in iterations of achieving simple structure, checking the various goodness of fit and reliability statistics whilst making choices, and more. In aiming for maximal

¹⁴ Perhaps the most easy-to-use implementation is the PCovR package within the R ecosystem (Vervloet, Kiers, van den Noortgate, & Ceuleman, 2015).

comparability, most of the doing of EFA will not be possible or desirable. This may well be the reason for Lorenzo-Seva and van Ginkel (2016) running a highly standardized and mechanical factor analysis on all of the completed data sets. All of the mechanical standardization to make the rotation and the application of MI's three stages possible might, however, be quite harmful to the final results of the analysis in and of itself.

Given the context of focus here (intended EFA to be performed by applied social scientist on small sample of primary data plagued by missing data), a number of intermediate conclusions can be reached. First of all, deletion methods should be abandoned. The fact that these are still by far the most common method of addressing missing data undermines (the quality of) applied social science. Secondly, from a missing data perspective, MI has a clear preference. It should be used as a general methodology, knowing that many mature and easy-to-use implementations using different prediction models exist. However, from an EFA perspective, MI is not an option as the quantities of interest of the analysis cannot be guaranteed to be comparable enough to pool and collapse into a single point estimate.

Thirdly, there are many different approaches to address missing data, each with their own favorable aspects and qualities. However, none seems to be an obvious candidate from the perspective of the applied social scientist. Iterative models aimed at estimating the covariance matrix whilst making optimal use of all available information are sensitive to the also typical circumstance of small sample sizes in primary data collected. The challenges of small sample size and missing data mutually aggravate each other. Single imputation or the averaging of multiple completed data sets does not properly recognize the uncertainty of imputation and/or does not incorporate it deep enough into the three-staged process of MI. Recognizing the uncertainty and keeping it incorporated into all of the stages of MI, in contrast, requires an extra step in which the quantities of interest are rotated to make them 'as similar to one another as possible.' This approach, however, means that the crucial doing of EFA,

including very important decisions on, for example, the number of factors, needs to be standardized to a degree that seems, in the end, more harmful than the method it makes possible for addressing missing data is beneficial.

5 Best of all possible worlds

Whilst the possible combination of MI and EFA arguably offers the best of both worlds for the applied social scientist, the adaptation of the former to make it applicable in combination with the latter is challenging. The representation of imputation uncertainty in the form of multiple completed data sets is very intuitive and should, ideally, be incorporated into as many stages of the MI process as possible. The second stage, of analysis, is pivotal. Staying true to the spirit of MI means EFA is performed on all m completed data sets. However, this means the EFA itself needs to be highly standardized and mechanically executed, quite contrary to the common practice of EFA in applied social science. Also, some of the most important decisions to be made in doing EFA have to be disregarded and set to some arbitrary fixed value beforehand. And finally, an additional rotation needs to be performed that might very well be counterintuitive to many applied social scientists, and may draw on technical skills they might not have at their disposal.

If, rather, one aims for the doing of EFA, with all of its subtleties and interdependent choices along the way, as the applied social scientist is used to and familiar with, this means it should be performed once on an estimated covariance matrix that incorporates and expresses the uncertainty of imputation as properly as can be. Averaging the completed data sets, however, was shown to result in a counterintuitive mix of observed ordinal values and real-valued imputations that could never have been observed. The best of all possible worlds approach we suggest is, therefore, the intermediate solution between averaging completed data

sets and averaging artificially adapted matrices of factor loadings.¹⁵ That is, for each of the completed data sets $\mathbf{D}_{(m)}$, we estimate a covariance matrix $\hat{\Sigma}_{(m)}$. Next, and in accordance with MI, we average these covariance matrices to obtain a single point estimate that properly incorporates the uncertainty of imputation:

$$\bar{\Sigma} = \frac{1}{m} \sum_m \hat{\Sigma}_m. \quad (9)$$

EFA is performed on this point estimate in the manner familiar to applied social scientists.

Our approach hereby circumvents the issues of comparability, as there is no pooling of multiple quantities of interest from separate EFAs. There also is no need for a priori artificial fixation of important decisions on the number of factors, nor any problem regarding the ordering of them. Easy-to-use and mature software implementations can be used in generating the completed data sets with a minimum of coding, all well within the quantitative operations applied social scientists are familiar with. The resulting $\mathbf{D}_{(m)}$ are all data sets containing values that could have been observed (ordinal values in accordance with Likert-type data). The EFA, as a contextual conversation between the data collected and the (domain expertise of the) researcher, proceeds as it would when deletion methods would have been used. No observed values are disregarded, no information is thrown away, no singularity issues emerge in the estimation of the covariance matrices due to the imputation of likely, and possible values, no bias is introduced that cannot be controlled (by the number of completed data sets generated), and no mechanical and artificial exploration of the data for latent structure is applied. This seems to be the best of all possible worlds.

¹⁵ This approach was already suggested by Nassiri, Lovik, Molenberghs, and Verbeke (2018). Their argument, however, is mainly focused on estimating confidence intervals for the proportion of explained variance by the first k factors, in combination with MI. The suggestion of averaging covariance matrices is offered almost as an aside and has not, as far as we are aware, been taken up by other researchers, neither in follow-up theoretical research nor in applied research.

There are, of course, downsides. Properly executing all stages of MI, without artificial interventions, like one can when performing a CFA, would be better still. Furthermore, working with a single pooled point estimate of the covariance matrix on which EFA is performed, precludes the application of pooling for precision purposes. However, given the context, the applied social scientist, at this stage of her research, is mostly interested in descriptive purposes. Possible inferences and formal tests that require precision are more common in latter stages of research, possibly after additional data collection based on the outcomes of the exploratory stage. Although these thus are, from a methodological standpoint, severe shortcomings, they might be less acute for the intended exploratory research. Put differently, the benefits of our suggested approach are not in any way ‘for free,’ but do outweigh the problems introduced and offer a way forward that is both more rigorous and more sound, and as intuitive and easy-to-use as applied social scientist are used to. How this ‘best of all worlds’ approach does look like, in very practical terms, is what we illustrate next.

6 Illustration

In order to illustrate our suggested approach, we simulate data with an easy to grasp one-dimensional latent structure. We do so in the user-friendly R programming language familiar to many applied social scientists (R Development Core Team, 2020). We have deliberately kept the coding to a minimum and gave preference to readability over coding best

practices.¹⁶ For reasons of comparability, we align our method of generating data to that of van der Eijk and Rose (2015).¹⁷

Each of the respondents is characterized by a position on the simulated underlying latent variable. More specifically, we assume the population distribution defined over the latent variable to be a normal distribution with mean 50 and standard deviation 20 (denoted by $\mathcal{N}(50, 20)$) from which we randomly draw the position of each of the respondents. In accordance with Table 1 in van der Eijk and Rose (2015), we define a set of Likert-type items (5-point ordinal scores) by their category boundaries. An item defined as (21, 38, 51, 59) would then translate a respondent's position of 40 on the latent variable into a score of 3, whilst a position of 83 would result in a score of 5. Finally, we translate these 'true scores' (in terms of classical test theory) into manifest data by adding errors in a manner that preserves the respondents' position on the latent variable, takes the smaller probability of larger errors into account, and maintains both the discreteness and the 'lumpy' character of Likert-type items.

More specifically, for each cell in $\mathbf{D}$, we independently draw a random variable $z \sim \mathcal{N}(0, 1)$. We then correct the 'true score' upwards or downwards according to the sign and the following cut-offs: no adjustment for $|z| \leq 1.2$, adjustment of one for $1.2 < |z| \leq 2.2$, adjustment of two for $2.2 < |z| \leq 3$, and an adjustment of three for $3 < |z|$ (all truncated to

¹⁶ Although, e.g., R is a *functional* programming language, we have opted for the readability of for loops over the preferred coding practice of using variants of apply and/or map whenever we deemed it to aid understanding. At other places in the script we have, e.g., opted for a `tibble` (Müller & Wickham, 2020) purely because it is generated in a more intuitive manner than a matrix, or have constructed extremely verbose lists of data transformations. We are fully aware of, and agree with, the fact that using a single paradigm in all of the code is a better practice and encourage applied social scientists to familiarize themselves with, e.g., `tidyverse` (Wickham et al., 2019). Given our intentions here, however, we have decided otherwise. The script file, made available for replication and suitable for changes to, e.g., accommodate different assumptions or simulate many data sets under different assumptions for testing, serves a purely illustrative purpose and is in no way an example of how the analyses performed could be optimally coded.

¹⁷ We use the same methodology and specification of boundaries between the categories of the Likert-type items (Table 1 in van der Eijk and Rose (2015)). However, as we have a different aim, we use less variation in terms of assumed underlying population distribution of the latent variable, number of items, etc. If desired, these elements can easily be included by combining the script file with those of van der Eijk and Rose (2015).

remain within the 5-point response categories). In this manner, slightly less than one out of four ‘true scores’ will be adjusted. The manifest response data that results is realistic in that it resembles response distributions observed in actual surveys. Furthermore, it clearly demonstrates the often-ignored fact that the distribution of manifest responses in no way indicates the distribution of the underlying latent variable(s).¹⁸

6.1 *Simulated missingness*

As various review studies (e.g., Henson & Roberts, 2006) have shown the majority of studies within our focus to have less than 200 respondents, we simulated manifest data **D** for 200 respondents on 15 items. Using the MICE package, we introduced missing data assuming the empirically common MAR condition. In addition to the standard patterns of missingness in MICE, we introduced three additional patterns that have missing values in two or more items. Furthermore, instead of the equally weighted frequency vector that determines the size of the subsets the manifest data is divided into, and thereby the relative occurrence of the patterns of missingness, we used a random vector containing draws for the (0, 1) interval summing to 1. For each of the respondents, a weighted sum score is computed that determines whether or not it receives a missing value. Four different logistic distribution functions are used in the allocation of these probabilities; we have randomized these for each of the items.¹⁹ Besides these randomizations, we used the default settings (cf. Schouten et al., 2018; Schouten & Vink, 2018).

The resulting data, including missing values, is comparable to what the applied social scientist could have collected with the intent of performing EFA. It has the approximate 50%

¹⁸ See Figures 2 and 3 in van der Eijk and Rose (2015) for a more detailed analysis and discussion.

¹⁹ See the script file and the excellent vignette of the `ampute` function of the MICE package for further details.

(98 out of 200) of respondents without any missing data, as mentioned by King et al. (2001), and about the percentage of missing elements (120 out of 3,000) observed in many review studies (e.g., Enders, 2004; Peugh & Enders, 2004). Most items have few missing values (13 out of 15 have between 2 and 9 missing values), whilst a few have many (one items with 15 and another with 30 missing values). A similar ‘lumpiness’ is observed for the respondents, with 97 out of the 102 respondents with missing values having a missing value for one item. In all, we have created a ‘typical’ dataset, with all the features the applied social scientist tends to have to deal with.

6.2 *Applied analysis*

How would the applied social scientist proceed if following our suggested approach? With the general methodology of MI being more important than the specific imputation model, we created $\mathbf{D}_{(m)}$ completed data sets using the polytomous logistic regression within MICE. We used 20 iterations and $m = 15$ (cf. van Buuren, 2007), although for many practical situations 5–10 iteration for a small number of 3–10 completed data sets suffices. Creating the $\mathbf{D}_{(m)}$ is done by a single command, after which we extract them instead of continuing our analyses within MICE (as one would in case of a single hypothesis).²⁰

Irrespective of the package and/or imputation model used, subsequent analyses are straightforwardly performed on the resulting list of completed data sets (`imps$sets`), each of dimension (200×15) in our example. Calculating $\bar{\Sigma}$, for example, is achieved by the following, intuitive expression:

```
Reduce("+", imps$sets %>% map(cov, method = "spearman")) / n_imp,
```

²⁰ Only the section “imputation” in the script is what is needed to generate and extract the completed data sets.

where the inner part of the expression applies the covariance function to each of the completed data sets, and the outer part sums the values for each cell in each of the $\mathbf{D}_{(m)}$ and divides by m . This $\bar{\Sigma}$ can be saved in any format desired to facilitate the doing of EFA in any software the applied researcher is familiar with. Alternatively, this $\bar{\Sigma}$ could serve as input for one of the many packages for EFA with R, like the mature `psych` package (Revelle, 2020), or essentially doing the same in a more visual manner within the excellent JASP environment (JASP Team, 2020).

6.3 Outcomes

As the values of ordinal variables can only be compared in terms of their ordering, covariation and correlation should be based on Spearman's rank-order.²¹ Furthermore, the family of affine-invariant Riemannian metrics called *Log-Euclidean* (LE) introduces computationally simple and fast metrics to measure distances between symmetric positive-definite matrices (for our purposes: covariance matrices).²² As the goal of EFA is to uncover a lower dimensional structure able to reproduce the observed covariation in manifest variables with as little loss of information possible, distances between covariance matrices constructed using different approaches provide us with rough indications concerning their impact.²³

²¹ Spearman's rank-order correlation can be understood as a nonparametric version of the more familiar Pearson product-moment correlation. It determines the strength and direction of monotonic relationships (not just linear). It can be used for variables measured on interval or ratio scales that violate (one of) the important but often not explicitly checked underlying assumptions of Pearson's product-moment correlation. Ordinal data by definition violates the assumed measurement on a continuous scale, meaning that association can only be measured by either Spearman's correlation or the polychoric correlation coefficient (Pearson's alternative). Although believed to be very different measures, they actually are not. The former can even be shown to be a deterministic transformation of the empirical version of the latter. As the polychoric correlation coefficient *requires* the latent variable to be continuous, Spearman's rank-order correlation often is the slightly more appropriate measure of association.

²² See, e.g., Arsigny et al. (2006) for a short and concise exposition.

²³ A smaller distance does not necessarily translate into 'uncovers the true underlying structure more precisely'.

With $\widehat{\Sigma}_{LD}$ representing the covariance matrix estimated on the collected data remaining after listwise deletion, and $\widehat{\Sigma}_M$ the covariance matrix estimated on the complete manifest data before the introduction of missing data, the distance between them is seen to be:

$$d_{LE}(\widehat{\Sigma}_M, \widehat{\Sigma}_{LD}) = \|\log \widehat{\Sigma}_M - \log \widehat{\Sigma}_{LD}\|_F = 5.43$$

This can be interpreted as an indication of the information lost even before the attempted lower dimensional approximation using EFA.²⁴ Creating completed data sets using MI as we did, estimating the covariance matrix on each of them and averaging these to obtain $\bar{\Sigma}$, results in $d_{LE}(\widehat{\Sigma}_M, \bar{\Sigma}) = 0.39$. Furthermore, averaging the completed data sets and estimating a covariance matrix on the resulting counterintuitive mix of observed ordinal values and real-valued imputations that could never have been observed, results in $d_{LE}(\widehat{\Sigma}_M, \widehat{\Sigma}_{AC}) = 0.44$.²⁵ This illustrates the importance of using the general methodology of MI over the finer details of, for example, imputation models and approach.

7 Conclusion and discussion

Multi-item scales and factor analyses are widely used standard tools in survey research. In the social sciences, including organizational research, such tools are omnipresent. Similarly, in the applied social sciences, including organizational research, missing data abound. We have to collect many of our data by approaching respondents working in practice, many of whom fail to fully deliver the data we ask for, for a variety of reasons. The typical response of the applied researcher is to adopt a simple data deletion strategy. However, this common practice

²⁴ For any matrix $\mathbf{M}$, its exponential is given by $\exp(\mathbf{M}) = \sum_{k=0}^{\infty} \frac{\mathbf{M}^k}{k!}$. The *matrix logarithm* used in d_{LE} is the inverse of this exponential. The norm used is the Frobenius norm.

²⁵ Being comparably close to a third covariance matrix does not imply closeness to each other, as $d_{LE}(\bar{\Sigma}, \widehat{\Sigma}_{AC}) = 0.22$ shows. The (proportionality of) comparisons and concluding indications remain the same when using correlation instead of covariance. Calculations included in the accompanying script file for reproduction.

is highly damaging for the quality of our work, introducing substantial loss of information, reduced power and biases in our estimates. This is particularly troublesome as we often have to work with rather small samples. Hence, imputation strategies are clearly superior to their deletion counterparts. In this paper, we focus on the case of the widely used technique of EFA, because imputation is then not straightforward due to the unknown factorial structure of the data.

The easy alternative to data deletion is to keep all respondents, but to apply imputation to missing values. We suggest an approach that combines various theoretical recommendations without becoming so technically involved that applied researchers will ignore this solution. We suggest our ‘best of all possible worlds’ approach should thus always be seen in light of both quantitative and qualitative considerations. Ease of use, intuition, a natural representation of uncertainty, and optimal use of MI in case of an intended EFA, make us suggest our approach as much as small numerical improvements do. All considering, we deem the ‘best of all possible worlds’ approach to have the greatest potential to change a common and, unfortunately, highly suboptimal practice of within applied social science for the better. To increase the likelihood that applied researchers will switch from deletion to imputation strategies, along the lines suggested above, we include all key coding, using standard and easy-to-use R packages, in the Appendix (available from the authors upon request). Starting to use this ‘best of all possible worlds’ approach will substantially increase the quality of our work, avoiding loss of information, reduced power, and biased estimates.

References

- Alexandre, J., Morin, S., Myers, N. D., & Lee, S. (2020). Modern Factor Analytic Techniques: Bifactor Models, Exploratory Structural Equation Modeling (ESEM), and Bifactor-ESEM. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of Sport Psychology* (4 ed., pp. 1044-1073). John Wiley & Sons. <https://doi.org/10.1002/9781119568124>
- Allison, P. D. (2003). Missing Data Techniques for Structural Equation Modeling. *Journal of Abnormal Psychology, 112*(4), 545-557. <https://doi.org/10.1037/0021-843X.112.4.545>
- Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis* (3 ed.). John Wiley & Sons.
- Arsigny, V., Fillard, P., Pennec, X., & Ayache, N. (2006). Log-Euclidean Metrics for Fast and Simple Calculus on Diffusion Tensors. *Magnetic Resonance in Medicine, 56*(2), 411-421. <https://doi.org/10.1002/mrm.20965>
- Barnes, S. A., Lindborg, S. R., & Seaman, J. W., Jr. (2006). Multiple Imputation Techniques in Small Sample Clinical Trials. *Statistics in Medicine, 25*(2), 233-245. <https://doi.org/10.1002/sim.2231>
- Bartholomew, D. J. (1980). Factor Analysis for Categorical Data. *Journal of the Royal Statistical Society. Series B (Methodological), 42*(3), 293-321. <https://doi.org/10.1111/j.2517-6161.1980.tb01128.x>
- ten Berge, J. M. (1977). Orthogonal Procrustes Rotation for Two or More Matrices. *Psychometrika, 42*(2), 267-276. <https://doi.org/10.1007/BF02294053>
- Blackwell, M., Honaker, J., & King, G. (2017). A Unified Approach to Measurement Error and Missing Data: Overview and Applications. *Sociological Methods & Research, 46*(3), 303-341. <https://doi.org/10.1177/0049124115585360>
- de Boeck, P., & Wilson, M. (2004). *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. Springer. <https://doi.org/10.1007/978-1-4757-3990-9>
- van Buuren, S. (2007). Multiple Imputation of Discrete and Continuous Data by Fully Conditional Specification. *Statistical Methods in Medical Research, 16*(3), 219-242. <https://doi.org/10.1177/0962280206074463>

- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate Imputation by Chained Equations. *Journal of Statistical Software*, 45(3), 1-67. <https://doi.org/10.18637/jss.v045.i03>
- Comrey, A. L., & Lee, H. B. (1992). *A First Course in Factor Analysis* (2 ed.). Lawrence Erlbaum Associates, Inc. <https://doi.org/10.4324/9781315827506>
- Conway, J. M., & Huffcutt, A. I. (2003). A Review and Evaluation of Exploratory Factor Analysis Practices in Organizational Research. *Organizational Research Methods*, 6(2), 147-168. <https://doi.org/10.1177/1094428103251541>
- Dray, S., & Josse, J. (2014). Principal Component Analysis With Missing Values: A Comparative Survey of Methods. *Plant Ecology*, 216(5), 657-667. <https://doi.org/10.1007/s11258-014-0406-z>
- van der Eijk, C., & Rose, J. (2015). Risky Business: Factor Analysis of Survey Data — Assessing the Probability of Incorrect Dimensionalisation. *PLoS ONE*, 10(3), 1-31. <https://doi.org/10.1371/journal.pone.0118900>
- Enders, C. K. (2004). The Impact of Missing Data on Sample Reliability Estimates: Implications for Reliability Reporting Practices. *Educational and Psychological Measurement*, 64(3), 419-436. <https://doi.org/10.1177/0013164403261050>
- Enders, C. K. (2010). *Applied Missing Data Analysis*. Guilford Press.
- Enders, C. K., & Bandalos, D. L. (2001). The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models. *Structural Equation Modeling*, 8(3), 430-457. https://doi.org/10.1207/S15328007SEM0803_5
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the Use of Exploratory Factor Analysis in Psychological Research. *Psychological Methods*, 4(3), 272-299. <https://doi.org/10.1037/1082-989X.4.3.272>
- Graham, J. W. (2009). Missing Data Analysis: Making It Work in the Real World. *Annual Review of Psychology*, 60, 549-576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>
- Graham, J. W., & Schafer, J. L. (1999). On the Performance of Multiple Imputation for Multivariate Data with Small Sample Size. In R. Hoyle (Ed.), *Statistical Strategies for Small Sample Research* (pp. 1-33). Sage.

- Henson, R. K., & Roberts, J. K. (2006). Use of Exploratory Factor Analysis in Published Research: Common Errors and Some Comment on Improved Practice. *Educational and Psychological Measurement*, 66(3), 393-416. <https://doi.org/10.1177/0013164405282485>
- Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A Program for Missing Data. *Journal of Statistical Software*, 45(7), 1-47. <https://doi.org/10.18637/jss.v045.i07>
- Ilin, A., & Raiko, T. (2010). Practical Approaches to Principal Component Analysis in the Presence of Missing Values. *Journal of Machine Learning Research*, 11(66), 1957-2000.
- Jamshidian, M., & Jalal, S. (2010). Tests of Homoscedasticity, Normality, and Missing Completely at Random for Incomplete Multivariate Data. *Psychometrika*, 75(4), 649-674. <https://doi.org/10.1007/S11336-010-9175-3>
- JASP Team. (2020). JASP. In (Version 0.14.1) <https://jasp-stats.org/>
- Josse, J., Pagès, J., & Husson, F. (2011). Multiple Imputation in Principal Component Analysis. *Advances in Data Analysis and Classification*, 5(3), 231-246. <https://doi.org/10.1007/s11634-011-0086-7>
- Kiers, H. A. (1997). Weighted Least Squares Fitting Using Ordinary Least Squares Algorithms. *Psychometrika*, 62(2), 251-266. <https://doi.org/10.1007/BF02295279>
- King, G., Honaker, J., Joseph, A., & Scheve, K. (2001). Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation. *American Political Science Review*, 95(1), 49-69. <https://doi.org/10.1017/S0003055401000235>
- Kroh, M. (2006). Taking 'Don't Knows' as Valid Responses: A Multiple Complete Random Imputation of Missing Data. *Quality & Quantity*, 40(2), 225-244. <https://doi.org/10.1007/s11135-005-5360-3>
- Little, J. R. (1988). Missing-Data Adjustments in Large Surveys. *Journal of Business & Economic Statistics*, 6(3), 287-296. <https://doi.org/10.1080/07350015.1988.10509663>
- Little, J. R., & Rubin, D. B. (2014). *Statistical Analysis with Missing Data* (2 ed.). John Wiley & Sons. <https://doi.org/10.1002/9781119482260>

- Lorenzo-Seva, U., & van Ginkel, J. R. (2016). Multiple Imputation of Missing Values in Exploratory Factor Analysis of Multidimensional Scales: Estimating Latent Trait Scores. *Annals of Psychology*, 32(2), 596-608. <https://doi.org/10.6018/analesps.32.2.215161>
- Lorenzo-Seva, U., Kiers, H. A., & ten Berge, J. M. (2002). Techniques for Oblique Factor Rotation of Two or More Loading Matrices to a Mixture of Simple Structure and Optimal Agreement. *British Journal of Mathematical & Statistical Psychology*, 55(2), 337-360. <https://doi.org/10.1348/000711002760554624>
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample Size in Factor Analysis. *Psychological Methods*, 4(1), 84-99. <https://doi.org/10.1037/1082-989X.4.1.84>
- McNeish, D. (2015). Using Lasso for Predictor Selection and to Assuage Overfitting: A Method Long Overlooked in Behavioral Sciences. *Multivariate Behavioral Research*, 50(5), 471-484. <https://doi.org/10.1080/00273171.2015.1036965>
- McNeish, D. (2017). Exploratory Factor Analysis With Small Samples and Missing Data. *Journal of Personality Assessment*, 99(6), 637-652. <https://doi.org/10.1080/00223891.2016.1252382>
- Müller, K., & Wickham, H. (2020). *tibble: Simple Data Frames*. In (Version 3.0.4) <https://CRAN.R-project.org/package=tibble>
- Nassiri, V., Lovik, A., Molenberghs, G., & Verbeke, G. (2018). On Using Multiple Imputation for Exploratory Factor Analysis of Incomplete Data. *Behavior Research Methods*, 50(2), 501-517. <https://doi.org/10.3758/s13428-017-1000-9>
- Peugh, J. L., & Enders, C. K. (2004). Missing Data in Educational Research: A Review of Reporting Practices and Suggestions for Improvement. *Review of Educational Research*, 74(4), 525-556. <https://doi.org/10.3102/00346543074004525>
- R Development Core Team. (2020). *R: A Language and Environment for Statistical Computing*. In <https://www.R-project.org/>
- Raghunathan, T. E., Lepkowski, J. M., van Hoewyk, J., & Solenberger, P. (2001). A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models. *Survey Methodology*, 27(1), 85-95.

- Revelle, W. (2020). *psych: Procedures for Psychological, Psychometric, and Personality Research*. In (Version 2.0.12) <https://CRAN.R-project.org/package=psych>
- Rubin, D. B. (1976). Inference and Missing Data. *Biometrika*, 63(3), 581-592. <https://doi.org/10.1093/biomet/63.3.581>
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons. <https://doi.org/10.1002/9780470316696>
- Russell, D. W. (2002). In Search of Underlying Dimensions: The Use (and Abuse) of Factor Analysis in *Personality and Social Psychology Bulletin*. *Personality and Social Psychology Bulletin*, 28(12), 1629-1646. <https://doi.org/10.1177/014616702237645>
- Schafer, J. L., & Olsen, M. K. (1998). Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst's Perspective. *Multivariate Behavioral Research*, 33(4), 545-571. https://doi.org/10.1207/s15327906mbr3304_5
- Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best Practices for Missing Data Management in Counseling Psychology. *Journal of Counseling Psychology*, 57(1), 1-10. <https://doi.org/10.1037/a0018082>
- Schouten, R. M., Lugtig, P., & Vink, G. (2018). Generating Missing Values for Simulation Purposes: A Multivariate Amputation Procedure. *Journal of Statistical Computation and Simulation*, 88(15), 2909-2930. <https://doi.org/10.1080/00949655.2018.1491577>
- Schouten, R. M., & Vink, G. (2018). The Dance of the Mechanisms: How Observed Information Influences the Validity of Missingness Assumptions. *Sociological Methods & Research*, OnlineFirst(October 9), 1-16. <https://doi.org/10.1177/0049124118799376>
- Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, 15(2), 201-292. <https://doi.org/10.2307/1412107>
- Vervloet, M., Kiers, H. A., van den Noortgate, W., & Ceuleman, E. (2015). PCovR: An R Package for Principal Covariates Regression. *Journal of Statistical Software*, 65(8), 1-14. <https://doi.org/10.18637/jss.v065.i08>
- White, I. R., Royston, P., & Wood, A. M. (2011). Multiple Imputation Using Chained Equations: Issues and Guidance for Practice. *Statistics in Medicine*, 30(4), 377-399. <https://doi.org/10.1002/sim.4067>

- Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Lin Pedersen, T., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wold, H. (1966). Estimation of Principal Components and Related Models by Iterative Least Squares. In P. R. Krishnaiah (Ed.), *Multivariate Analysis* (pp. 391-420). Academic Press.