Multiple Imputation Guided by Full Law and Target Law Identifiability

Anonymous authors
Paper under double-blind review

Abstract

The central challenges in missing data models concern the identifiability of two distributions: the target law and the full law. The target law refers to the joint distribution of the data variables, whereas the full law refers to the joint distribution of the data variables and their corresponding response indicators. However, the relationship between the identifiability of these two distributions and the feasibility of multiple imputation has not been clearly established. We present a procedure where the choice of the imputation method is guided by identifiability considerations. We show that imputations can be drawn from the correct conditional distributions for all possible missing data patterns if and only if the full law is identifiable. This result implies that standard multiple imputation methods—which keep observed values unchanged and replace missing values with imputed values—are invalid when the target law is identifiable but the full law is not. We demonstrate that alternative imputation strategies can sometimes enable the estimation of the target law in such cases. Specifically, we introduce factorizable imputation where certain observed values are also imputed and the imputed data are weighted in the analysis.

1 Introduction

Multiple imputation is a popular tool for practical data analysis with missing data (Van Buuren, 2018). A standard imputation method keeps observed values unchanged and replaces all missing values with imputed values producing many imputed versions of the full data. The theoretical basis of multiple imputation (Rubin, 1987) has been built on the missing at random (MAR) assumption (Rubin, 1976; Little & Rubin, 2019; Seaman et al., 2013). This assumption is not strictly necessary and multiple imputation has been applied also in some cases where the data are missing not at random (MNAR) (Galimard et al., 2016; Hammon & Zinn, 2020; Beesley et al., 2021; Muñoz et al., 2024).

Alongside the MAR assumption, there has been growing interest in missing data literature toward more detailed nonparametric assumptions and their implications for identifiability, also known as recoverability. These assumptions are often represented by graphical missing data models (Daniel et al., 2012; Mohan et al., 2013; Karvanen, 2015). The identifiability of the full law, i.e., the joint distribution of data variables and response indicators, and the identifiability of the target law, i.e., the joint distribution of data variables, have been among the main questions of interest (e.g., Tian, 2017; Bhattacharya et al., 2020; Mohan & Pearl, 2021; Guo et al., 2023; Tikka & Karvanen, 2025). The full law and the target law are identifiable also in some cases that are classified as MNAR. Currently, the sufficient and necessary conditions for the nonparametric identifiability are known for the full law (Nabi et al., 2020), but not for the target law.

Multiple imputation and identifiability of the full law have not been explicitly linked in the literature. Mathur et al. (2025) give a graphical condition for an imputation model trained on complete cases to be valid. However, imputation methods such as multiple imputation by chained equations (Van Buuren & Groothuis-Oudshoorn, 2011) use also partially observed cases, which extends the applicability of imputation.

In this paper, we study the consequences of the full law and the target law identifiability for multiple imputation. First we show that imputations can be drawn from the correct conditional distributions for all possible missing data patterns if and only if the full law is identifiable. This result is intuitive but has

not been presented in the literature. For the purposes of this paper, we define a conditionally complete imputation method as a procedure that produces samples from the correct conditional distributions. We consider only correctly specified imputation models and ignore the challenges related to their estimation. The link between multiple imputation and the identifiability of the full law implies that multiple imputation is applicable also in some MNAR cases. The result means that we can evaluate the potential of standard multiple imputation in a MNAR case by checking the sufficient and necessary conditions (Nabi et al., 2020) for nonparametric identifiability of the full law.

There are also cases where the full law is not identifiable, and consequently, the standard application of multiple imputation leads to biased estimates. We show that in such cases, alternative non-standard imputation strategies can sometimes be valid, for example when the target law is identifiable. When the target law factorizes into identifiable components, we show that a non-standard imputation method which we call factorizable imputation yields valid imputations. The factorization mandates two non-standard practices: replacing actual observations by imputed values under certain conditions and weighting the imputed data in the analysis.

The rest of the paper is organized as follows. In Section 2, the necessary notation is introduced and the results on identifiability and multiple imputation are presented. Section 3 presents a practical procedure for multiple imputation guided by identifiability considerations. The implications of the theoretical results are demonstrated via examples in Section 4. Section 5 concludes the paper.

2 Theoretical results

2.1 Notation and definitions

Let $\mathbf{X}$ denote the set of data variables that may be missing (called partially observed variables). Under missing data, we do not observe these variables directly. Instead, we observe a set of proxy variables $\mathbf{X}^*$ and their response indicators $\mathbf{R}$. There may also be variables that are fully observed, denoted by $\mathbf{O}$. The proxy variables $X^* \in \mathbf{X}^*$ are defined as follows:

$$X^* = \begin{cases} X & \text{if } R_X = 1, \\ \text{NA} & \text{if } R_X = 0, \end{cases} \tag{1}$$

where a missing value is indicated by NA (not available). In other words, the proxies represent the true values of the partially observed variables when they are not missing ($R_X = 1$), and the value is missing otherwise ($R_X = 0$). We generalize the notation of response indicators to sets $\mathbf{Y} \subseteq \mathbf{X}$ as follows. $R_{\mathbf{Y}} = 1$ denotes that all individual response indicators R_Y such that $Y \in \mathbf{Y}$ have the value assignment $R_Y = 1$. The notation $R_{\mathbf{Y}} = 0$ is analogous.

A set of distributions over a set of random variables $\mathbf{V}$ that can be partitioned into the sets $\mathbf{X}$, $\mathbf{X}^*$, $\mathbf{R}$, and $\mathbf{O}$ is called a missing data model. When considering a particular value assignment of $\mathbf{r}$ to $\mathbf{R}$, we further partition $\mathbf{X}$ into the sets $\mathbf{X}_1$ and $\mathbf{X}_0$, where variables $\mathbf{X}_1$ are observed and variables $\mathbf{X}_0$ are missing under this particular value assignment, respectively. In other words, when $\mathbf{r}$ is fixed, we have that $R_X = 1$ for all $X \in \mathbf{X}_1$, and $R_X = 0$ for all $X \in \mathbf{X}_0$.

In missing data models, the joint distribution of $\mathbf{V}$ factorizes as follows

$$p(\mathbf{V}) = p(\mathbf{O}, \mathbf{X}^*, \mathbf{X}, \mathbf{R}) = p(\mathbf{X}^* | \mathbf{X}, \mathbf{R}) p(\mathbf{O}, \mathbf{X}, \mathbf{R}),$$

where $p(\mathbf{O}, \mathbf{X}, \mathbf{R})$ is the *full law*, and $p(\mathbf{X}^* | \mathbf{X}, \mathbf{R})$ represents the deterministic component related to (1). The full law is characterized by the *target law* $p(\mathbf{O}, \mathbf{X})$ and the *missingness mechanism* $p(\mathbf{R} | \mathbf{O}, \mathbf{X})$. Finally, the distribution $p(\mathbf{O}, \mathbf{X}^*, \mathbf{R})$ that is available for practical analyses is called the *observed data law*.

A nonparametric missing data model $\mathcal{M}_\Omega$ is a missing data model whose members share a description Ω. The description is a set of nonparametric assumptions about the missing data model, such as (counterfactual) conditional independence restrictions for variables $\mathbf{V}$. These restrictions can be listed directly or they can be described by a graphical model, for example. Missing completely at random (MCAR) is an example of a description Ω that defines a nonparametric missing data model $\mathcal{M}_\Omega$.

We focus on graphical missing data models depicted by directed acyclic graphs (DAGs) (Daniel et al., 2012; Mohan et al., 2013; Karvanen, 2015). In other words, the description Ω encodes a missing data model such that the joint distribution factorizes according to the structure of the DAG, i.e.,

$$p(\mathbf{V}) = \prod_{V_i \in \mathbf{V}} p(V_i \mid \mathrm{pa}_{\mathcal{G}}(V_i)),$$

where $\mathcal{G}$ is a DAG and $\mathrm{pa}_{\mathcal{G}}(V_i)$ denotes the parents of V_i in $\mathcal{G}$. We also assume positivity for the full law, i.e., $p(\mathbf{O} = \mathbf{o}, \mathbf{X} = \mathbf{x}, \mathbf{R} = \mathbf{r}) > 0$ for all value assignments $(\mathbf{o}, \mathbf{x}, \mathbf{r})$ to $(\mathbf{O}, \mathbf{X}, \mathbf{R})$. This means that there are no deterministic relationships between the variables in the missing data model, except those related to the definition of proxy variables in equation (1). Conditional independence restrictions implied by the DAG can be derived using d-separation (Pearl, 1995). To simplify the presentation, we assume in the rest of the paper that all variables may have missing values, i.e., $\mathbf{O} = \emptyset$. Accordingly, $p(\mathbf{X}, \mathbf{R})$ will stand for the full law, $p(\mathbf{X})$ for the target law, and $p(\mathbf{X}^*, \mathbf{R})$ for the observed data law.

We define a general notion of identifiability that is not specific to any particular description Ω of a nonparametric missing data model.

Definition 2.1. An estimand or a probabilistic query Q is said to be *identifiable* from $(\Omega, p(\mathbf{X}^*, \mathbf{R}))$ if there exists a functional f such that $Q = f(p(\mathbf{X}^*, \mathbf{R}))$ for every element of a nonparametric missing data model $\mathcal{M}_\Omega$.

In this paper we focus on the identifiability of the target law and the full law meaning that our queries of interest are $Q = p(\mathbf{X})$ and $Q = p(\mathbf{X}, \mathbf{R})$, respectively.

An imputation method turns an incomplete dataset into a complete dataset (Van Buuren, 2018, chapter 2.3.3). We characterize this notion in the following definition.

Definition 2.2. Let $\mathcal{M}_\Omega$ be a nonparametric missing data model with observed data law $p(\mathbf{X}^*, \mathbf{R})$. An *imputation method* is a collection of stochastic kernels, called *imputation kernels*

$$\xi_{\Omega, p(\mathbf{X}^*, \mathbf{R})} = \left\{ \xi_{\mathbf{r}}(\mathbf{x}_0 \mid \mathbf{x}_1) : \mathbf{r} \in \{0,1\}^{|\mathbf{X}|} \right\},$$

where, for each missingness pattern $\mathbf{r}$, the variables are partitioned into observed variables $\mathbf{X}_1$ and missing variables $\mathbf{X}_0$. The kernel $\xi_{\mathbf{r}}(\mathbf{x}_0 \mid \mathbf{x}_1)$ is a conditional distribution over the missing variables $\mathbf{X}_0$, given the observed variables $\mathbf{X}_1 = \mathbf{x}_1$, whose form is determined only by Ω and the observed data law $p(\mathbf{X}^*, \mathbf{R})$.

In practice, imputations are drawn from the imputation kernels. Naturally, not all imputation methods correspond with the true conditional distribution of $\mathbf{X}_0$ for all partitions of $\mathbf{X}$.

2.2 Conditionally complete imputation and the full law

Following Rubin (1987, chapter 2.6), we define a conditionally complete imputation method as follows.

Definition 2.3. Let $\mathcal{M}_\Omega$ be a nonparametric missing data model with observed data law $p(\mathbf{X}^*, \mathbf{R})$. An imputation method $\xi_{\Omega, p(\mathbf{X}^*, \mathbf{R})}$ is *conditionally complete* if, for every missingness pattern $\mathbf{r}$, the corresponding imputation kernel satisfies

$$\xi_{\mathbf{r}}(\mathbf{x}_0 \mid \mathbf{x}_1) = p(\mathbf{X}_0 \mid \mathbf{X}_1 = \mathbf{x}_1, R_{\mathbf{X}_1} = 1, R_{\mathbf{X}_0} = 0), \tag{2}$$

for all possible value assignments $(\mathbf{x}_0, \mathbf{x}_1)$ to $(\mathbf{X}_0, \mathbf{X}_1)$.

In practice, many imputation methods estimate the conditional predictive distributions and then simulate values from the estimated models also known as imputation models. The imputation model may contain interactions and nonlinear relationships. In MAR settings, $\mathbf{X}_0$ in equation (2) is independent from $\mathbf{R}$ but in MNAR settings the response indicators are usually needed as predictors. We define standard multiple imputation as a procedure where the NA values are replaced with multiple imputed values but actual observations are not altered. Removing rows of the data or changing observed values to missing values are in general considered dubious operations that lead to information loss and may cause bias. This may sound

obvious, but later we will also consider non-standard imputation methods where valid imputations cannot be produced unless these principles are violated.

With the definitions above, we are ready the present the first result of the paper:

Theorem 2.4. *Let $\mathcal{M}_\Omega$ be a nonparametric missing data model. A conditionally complete imputation method $\xi_{\Omega,p(\mathbf{X}^*,\mathbf{R})}$ exists if and only if the full law $p(\mathbf{X},\mathbf{R})$ is identifiable from $(\Omega, p(\mathbf{X}^*,\mathbf{R}))$.*

Proof. Assume first that $p(\mathbf{X},\mathbf{R})$ is identifiable. Then any conditional distribution of the form of equation (2) is also identifiable and the stochastic kernels required for the conditionally complete imputation method can be constructed based on the identifying functional $f(p(\mathbf{X}^*,\mathbf{R})) = p(\mathbf{X},\mathbf{R})$.

Assume then that a conditionally complete imputation method $\xi_{\Omega,p(\mathbf{X}^*,\mathbf{R})}$ exists. The full law for the case where $\mathbf{X}_1$ is observed and $\mathbf{X}_0$ is missing can be then written as

$$p(\mathbf{X}_1, \mathbf{X}_0, R_{\mathbf{X}_1} = 1, R_{\mathbf{X}_0} = 0) = p(\mathbf{X}_0 \,|\, \mathbf{X}_1, R_{\mathbf{X}_1} = 1, R_{\mathbf{X}_0} = 0)p(\mathbf{X}_1, R_{\mathbf{X}_1} = 1, R_{\mathbf{X}_0} = 0).$$

Now, $p(\mathbf{X}_1, R_{\mathbf{X}_1} = 1, R_{\mathbf{X}_0} = 0)$ is directly identifiable from the observed data law $p(\mathbf{X}^*,\mathbf{R})$, and $p(\mathbf{X}_0 \,|\, \mathbf{X}_1, R_{\mathbf{X}_1} = 1, R_{\mathbf{X}_0} = 0)$ is obtained from the imputation kernels of the conditionally complete imputation method $\xi_{\Omega,p(\mathbf{X}^*,\mathbf{R})}$. Since the partition of $\mathbf{X}$ into $\mathbf{X}_0$ and $\mathbf{X}_1$ is arbitrary, it follows that the full law $p(\mathbf{X},\mathbf{R})$ is identifiable. □

Theorem 2.4 implies that standard multiple imputation (as defined above) is a valid approach for the estimation of the full law if and only if the full law is identifiable. As the target law $p(\mathbf{X})$ and the full law $p(\mathbf{X},\mathbf{R})$ have the same data variables and the same possible partitions of $\mathbf{X}$, standard multiple imputation is a valid approach for the estimation of the target law if and only if the full law is identifiable. In other words, one should check the identifiability of the full law under the assumed model $\mathcal{M}_\Omega$ before applying standard multiple imputation, and proceed only if the answer is positive.

2.3 Factorizable imputation and the target law

Alternative, non-standard multiple imputation approaches are sometimes valid even when the full law is not identifiable. Next, we will consider ordered variables. When an ordering is fixed and Z precedes W in the ordering, we write $Z \prec W$. We use subscripts within parentheses to denote variables in an ordering, e.g., $X_{(1)} \prec X_{(2)} \prec \cdots \prec X_{(n)}$. We also introduce the following notation for the set of predecessors:

$$\mathbf{X}_{\prec(i)} = \{X_{(j)} \in \mathbf{X} \mid X_{(j)} \prec X_{(i)}\},$$

with the convention that $\mathbf{X}_{\prec(1)} = \emptyset$. In other words, $\mathbf{X}_{\prec(i)}$ denotes the set of variables that precede $X_{(i)}$ in the ordering. Our goal is to construct an imputation method that leverages a specific ordering of the variables in $\mathbf{X}$ to draw correct imputations. The following definition describes such a method, which we call directly factorizable imputation.

Definition 2.5. Let $\mathcal{M}_\Omega$ be a nonparametric missing data model and let the variables in $\mathbf{X}$ be ordered as $X_{(1)} \prec \cdots \prec X_{(n)}$. An imputation method is *directly factorizable* with respect to the ordering if its imputation kernels are of the form

$$\xi^{\mathrm{DF}}_{\mathbf{r}}(\mathbf{t} \,|\, \mathbf{h}) = \prod_{j=m(\mathbf{r})}^{n} p(X_{(j)} \,|\, \mathbf{X}_{\prec(j)}), \tag{3}$$

where $m(\mathbf{r})$ is the index of the first variable with a missing value in the ordering with respect to the missingness pattern $\mathbf{r}$, $\mathbf{H}(\mathbf{r}) = \{X_{(j)} : j < m(\mathbf{r})\}$, $\mathbf{T}(\mathbf{r}) = \{X_{(j)} : j \geq m(\mathbf{r})\}$, and $\mathbf{h}$ denotes the values of $\mathbf{H}(\mathbf{r})$ and $\mathbf{t}$ denotes the values of $\mathbf{T}(\mathbf{r})$.

Note that a directly factorizable imputation method may impute $X_{(k)}$ even if $R_{X_{(k)}} = 1$ in cases where $R_{X_{(j)}} = 0$ for some $1 < j < k$. In practice, the observed value is then replaced by the imputed value; an operation that is not part of standard multiple imputation. In other words, the missingness pattern is

forced to be monotone. This practice may sound counter-intuitive as it leads to loss of information, but it is required to ensure that the imputations are drawn from the correct conditional distributions. However, we note that the loss of information is not as substantial as in complete case analysis, because information up to the first missing value is still available for analysis. The next theorem provides a sufficient criterion for the existence of a directly factorizable imputation method.

Theorem 2.6. *Let $\mathcal{M}_\Omega$ be a nonparametric missing data model. If there exists an ordering $X_{(1)} \prec \cdots \prec X_{(n)}$ of the variables in $\mathbf{X}$ such that $X_{(1)} \perp\!\!\!\perp R_{(1)}$ and $X_{(j)} \perp\!\!\!\perp R_{X_{(j)}}, R_{\mathbf{X}_{\prec(j)}} \,|\, \mathbf{X}_{\prec(j)}$ for all $j = 2, \ldots, n$, then a directly factorizable imputation method exists.*

Proof. It follows from the assumed conditional independence restrictions that any distribution of the form $p(X_{(j)} \,|\, \mathbf{X}_{\prec(j)})$ is identifiable from the observed data law. Thus the imputation kernels defined in equation (3) can be constructed based on the identifying functional. □

In practice, finding a suitable ordering required by Theorem 2.6 can be a challenge for large models because the number of possible orderings is the factorial of n. To alleviate this, we note that the target law is identifiable under the conditions of Theorem 2.6; a special case of identification via an admissible factorization presented by Mohan et al. (2013), and thus their results can be used to reduce the search space of possible orderings in some cases.

Corollary 2.7. *Let $\mathcal{M}_\Omega$ be a nonparametric missing data model. The target law $p(\mathbf{X})$ is identifiable under the conditions of Theorem 2.6.*

Proof. Identifiability of the target law follows directly from the assumed conditional independence restrictions as follows

$$\begin{aligned} p(\mathbf{X}) &= \prod_{j=1}^{n} p(X_{(j)} \,|\, \mathbf{X}_{\prec(j)}) \\ &= \prod_{j=1}^{n} p(X_{(j)} \,|\, \mathbf{X}_{\prec(j)}, R_{X_{(j)}} = 1, R_{\mathbf{X}_{\prec(j)}} = 1) \\ &= \prod_{j=1}^{n} p(X^*_{(j)} \,|\, \mathbf{X}^*_{\prec(j)}, R_{X_{(j)}} = 1, R_{\mathbf{X}_{\prec(j)}} = 1) \end{aligned}$$

Each term in the final right-hand side expression is directly identifiable from the observed data distribution $p(\mathbf{X}^*, \mathbf{R})$. □

Factorizable imputation can be extended to scenarios where the joint distribution of a subset of variables is identifiable by other means. We define *conditionally factorizable imputation* as follows.

Definition 2.8. Let $\mathcal{M}_\Omega$ be a nonparametric missing data model and let $(\mathbf{A}, \mathbf{B})$ be a partition of $\mathbf{X}$. Suppose the variables in $\mathbf{B}$ are ordered as $B_{(1)} \prec \cdots \prec B_{(k)}$. A *conditionally factorizable imputation method* for the partition $(\mathbf{A}, \mathbf{B})$ consists of:

1. restriction to the subpopulation with $R_{\mathbf{A}} = 1$;
2. a collection of stochastic kernels $\xi^{\mathrm{CF}}_{\mathbf{r}}(\mathbf{t_B} \,|\, \mathbf{a}, \mathbf{h_B})$ defined for missingness patterns $\mathbf{r}$ satisfying $\mathbf{r_A} = 1$, where
$$\xi^{\mathrm{CF}}_{\mathbf{r}}(\mathbf{t_B} \,|\, \mathbf{a}, \mathbf{h_B}) = \prod_{j=m_{\mathbf{B}}(\mathbf{r})}^{k} p(B_{(j)} \,|\, \mathbf{A} = \mathbf{a}, \mathbf{B}_{\prec(j)}),$$
and $m_{\mathbf{B}}(\mathbf{r})$, $\mathbf{T}_B(\mathbf{r})$, $\mathbf{H_B}$, $\mathbf{t_B}$ and $\mathbf{h_B}$ are defined analogously to Definition 2.5;
3. a weight function
$$w(\mathbf{a}) = c\, p(R_{\mathbf{A}} = 1 \,|\, \mathbf{A} = \mathbf{a})^{-1},$$
for an arbitrary normalizing constant c.

The next theorem provides sufficient criteria for the existence of a conditionally factorizable imputation method.

Theorem 2.9. *The target law is identifiable and there exists a conditionally factorizable imputation method for a partition* $(\mathbf{A}, \mathbf{B})$ *of* $\mathbf{X}$ *if the following conditions hold*

(a) $\mathbf{B} \perp\!\!\!\perp R_{\mathbf{A}} \mid \mathbf{A}$,

(b) $p(R_{\mathbf{A}} = 1 \mid \mathbf{A})$ *is identifiable from* $p(\mathbf{A}^*, R_{\mathbf{A}})$, *and*

(c) there exists an ordering $B_{(1)} \prec B_{(2)} \prec \cdots \prec B_{(k)}$ *of* $\mathbf{B}$ *for which* $B_{(1)} \perp\!\!\!\perp R_{B_{(1)}} \mid \mathbf{A}, R_{\mathbf{A}}$ *and* $B_{(j)} \perp\!\!\!\perp R_{B_{(j)}}, R_{\mathbf{B}_{\prec(j)}} \mid \mathbf{B}_{\prec(j)}, \mathbf{A}, R_{\mathbf{A}}$,

Proof. The target law factorizes as $p(\mathbf{X}) = p(\mathbf{A})p(\mathbf{B} \mid \mathbf{A})$. The identifiability of $p(\mathbf{A})$ follows directly from condition (b), which assumes the identifiability of $p(R_{\mathbf{A}} = 1 \mid \mathbf{A})$, because

$$p(\mathbf{A}) = \frac{p(R_{\mathbf{A}} = 1)p(\mathbf{A} \mid R_{\mathbf{A}} = 1)}{p(R_{\mathbf{A}} = 1 \mid \mathbf{A})}.$$

In other words, complete cases $p(\mathbf{A} \mid R_{\mathbf{A}} = 1)$ are weighted by $p(R_{\mathbf{A}} = 1)p(R_{\mathbf{A}} = 1 \mid \mathbf{A})^{-1}$. The directly factorizable imputation is applied for the data where $\mathbf{A}$ is fully observed. The validity of the imputation follows from Theorem 2.6 and condition (c) when we consider an ordering where $\mathbf{A}$ precedes variables $B_{(1)} \prec B_{(2)} \prec \cdots \prec B_{(k)}$. The imputations are from distribution $p(\mathbf{B} \mid \mathbf{A}, R_{\mathbf{A}} = 1)$ which by condition (a) equals $p(\mathbf{B} \mid \mathbf{A})$. Thus, both components of the factorization $p(\mathbf{A})p(\mathbf{B} \mid \mathbf{A})$ are identifiable. □

Theorem 2.6 is a special case of Theorem 2.9 when $\mathbf{A} = \emptyset$. Theorem 2.9 requires $p(R_{\mathbf{A}} = 1 \mid \mathbf{A})$ to be identifiable from $p(\mathbf{A}^*, R_{\mathbf{A}})$ but does not specify the way how the identification should be achieved. A practical approach will be available if $p(\mathbf{A}, R_{\mathbf{A}})$ is identifiable from $p(\mathbf{A}^*, R_{\mathbf{A}})$. Using reasoning analogous to Theorem 2.4, we may conclude that conditionally complete imputation can be applied to variables $\mathbf{A}$. Then the weights $p(R_{\mathbf{A}} = 1 \mid \mathbf{A})^{-1}$ can be estimated from the imputed data for $p(\mathbf{A}, R_{\mathbf{A}})$.

Conditionally factorizable imputation may be applicable in some longitudinal settings where $\mathbf{A}$ represents baseline measurements and $\mathbf{B}$ represents repeated measurements in later time points. The missingness of $\mathbf{B}$ may be affected by $\mathbf{A}$ and $R_{\mathbf{A}}$. However, it is natural to assume that the missingness of $\mathbf{A}$ does not depend on future measurements $\mathbf{B}$, which fulfills the condition (a) of Theorem 2.9. If in addition, $p(R_{\mathbf{A}} = 1 \mid \mathbf{A})$ can be identified from the baseline data and the repeated measurements $\mathbf{B}$ ordered in time fulfill condition (c) of Theorem 2.9, a conditionally factorizable imputation method can be applied.

3 Practical procedure

In this section, we show how the theoretical results of Section 2 can be applied the in practical data analysis. As a starting point, we have a graphical missing data model and a random sample from the distribution $p(\mathbf{X}^*, \mathbf{R})$ available. Although Theorems 2.4 and 2.9 are independent of each other, it seems reasonable to first check the identifiability of the full law. The overall procedure can be presented as follows:

1. If the full law is identifiable, apply conditionally complete multiple imputation.
2. If the conditions of Theorem 2.6 hold for some ordering, apply factorizable imputation with this ordering.
3. If the conditions of Theorem 2.9 hold for some partition and some ordering, apply factorizable imputation with this partition and this ordering.
4. Use Do-search or apply do-calculus manually to check if the target law is identifiable. If yes, estimate the target law using the obtained identifying functional.

5. If the target law is not identifiable, it can be still checked if some marginal distribution or a joint distribution of a subset of variables can be identified. Respecify the problem and start again from Step 1.

Next, we will discuss each step of the procedure. Checking the identifiability of the full law in Step 1 requires examining relatively simple graphical conditions. When unmeasured confounders are not present, the full law is identifiable if and only if the graph does not contain self-censoring edges or colluders (Nabi et al., 2020, Theorem 1). A colluder is a specific type of collider structure, where a partially observed variable and its response indicators have an outgoing edge into the response indicator of another partially observed variable, e.g., $X \to R_Y \leftarrow R_X$. A self-censoring edge is an edge from a partially observed variable into its corresponding response indicator, e.g., $X \to R_X$. With unmeasured confounders, the full law is identifiable if and only if the graph does not contain colluding paths (Nabi et al., 2020, Theorem 4).

The implementation of conditionally complete imputation requires some attention. As can be seen from equation (2), the imputation model should include the response indicators (Tompsett et al., 2018; Beesley et al., 2021). However, if the data are MAR, the response indicators are not needed. In software packages for multiple imputation, response indicators are usually not included by default and must be manually specified in the imputation model. In addition, the user has to choose the imputation method (linear model, predictive mean matching, random forest, etc.), the number of imputations and other details. Multiple imputation by chained equations also requires specifying the number of iterations needed to guarantee the convergence of the parameters of the imputation models.

In Steps 2 and 3, the key requirement is finding a suitable of ordering of the variables. While this is straightforward in a small graph, it can become challenging when dealing with a large graph. Factorizable imputation needs a preprocessing phase where the value of a variable is set as missing if any of its preceding variables has a missing value. Due to the preprocessing and the specified ordering, the covariates of the imputation models are always available and no iterations are needed. Response indicators are not used in the imputation models in factorizable imputation. Step 3 involves weights that must be estimated from the data. Standard techniques for weight / propensity score estimation are applicable here.

Step 4 emphasizes that in some cases the target law is identifiable but neither conditionally complete multiple imputation nor factorizable imputation is applicable. In such cases, the estimation method need to be tailored case by case. Finally, Step 5 remarks that even if the target law was not identified in the previous steps, it might be possible to identify the distribution of some subset of variables.

4 Examples

To illustrate the application of the theoretical results of Section 2, we consider multiple imputation when the data are MNAR. We investigate the identifiability of the full law and the target law in our example cases and use Theorems 2.4, 2.6 and 2.9 to make conclusions about the applicability of multiple imputation. These conclusions are verified by simulation studies where the aim is to compare the methods for the estimation of basic summary statistics. The codes for the simulation examples are available in the supplementary material.

The estimation methods in the comparison include three versions of multiple imputation, complete case analysis (CCA), available case analysis (ACA) and estimation using the identifying functionals. The imputation methods are multiple imputation without response indicators (MI), multiple imputation with response indicators (MIRI) and factorizable multiple imputation (FMI). In all imputation methods, we apply multiple imputation by chained equations using the R package `mice` (Van Buuren & Groothuis-Oudshoorn, 2011) with Bayesian linear regression (method "norm"). The imputation models are specified manually for MIRI and FMI. In FMI, some observations are replaced by missing values before the imputation to transform the data into the required monotonic missing data pattern.

4.1 Examples with two variables

We consider six graphs depicted in Figure 1. All graphs represent MNAR scenarios because the missingness of Y depends on X that can be missing as well. We study the identification and estimation of the full law $p(X, Y, R_X, R_Y)$ and the target law $p(X, Y)$ in these scenarios.

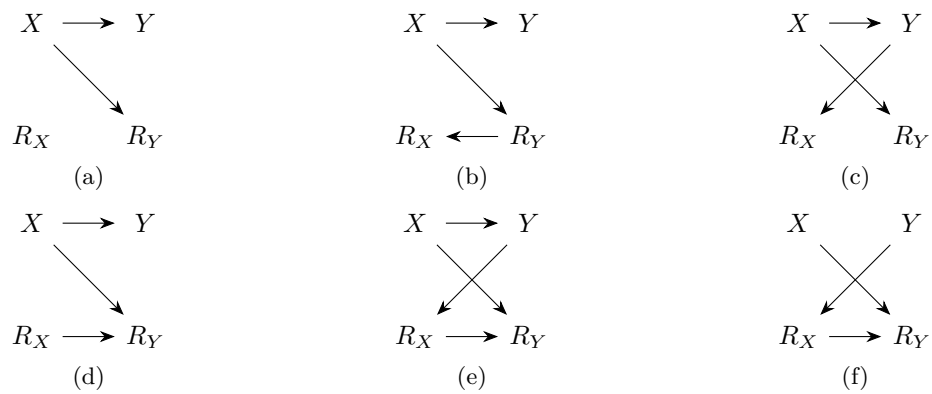

Figure 1: Graphs used for the two-variable examples. The full law and the target law are identifiable in graphs (a), (b) and (c). The target law is identifiable also in graphs (d) and (f) but not in graph (e). Proxy variables and edges related to them are not shown for simplicity.

First we consider the identifiability of the full law $p(X, Y, R_X, R_Y)$ and the target law $p(X, Y)$ graph by graph. The identifying functionals can be derived manually or applying the Do-search algorithm (Tikka et al., 2021), for example. The results are summarized in Table 1.

Table 1: Summary of identifiability and validity of imputation for the graphs of Figure 1.

	Graph					
	1a	1b	1c	1d	1e	1f
Full law identifiable	✓	✓	✓	×	×	×
Target law identifiable	✓	✓	✓	✓	×	✓
Theorem 2.4 applies	✓	✓	✓	×	×	×
Theorem 2.6 applies	✓	×	×	✓	×	×

In the graph of Figure 1a, the full law is identified by the formula

$$p(X, Y, R_X, R_Y) = p(Y^* \mid X^*, R_Y = 1, R_X = 1)p(R_Y \mid X^*, R_X = 1)p(X^* \mid R_X = 1)p(R_X).$$

thus a conditionally complete imputation method exists. The conditional distributions for imputation are identified using the full law as follows

$$\begin{aligned}
p(X \mid Y, R_X{=}0, R_Y{=}1) &= \frac{p(X, Y, R_X{=}0, R_Y{=}1)}{\sum_X p(X, Y, R_X{=}0, R_Y{=}1)}, \\
p(Y \mid X, R_X{=}1, R_Y{=}0) &= \frac{p(Y, X, R_X{=}1, R_Y{=}0)}{\sum_Y p(Y, X, R_X{=}1, R_Y{=}0)}, \\
p(X, Y \mid R_X{=}0, R_Y{=}0) &= \frac{p(Y, X, R_X{=}0, R_Y{=}0)}{\sum_{X,Y} p(Y, X, R_X{=}0, R_Y{=}0)},
\end{aligned} \tag{4}$$

where the summations should be understood as integrals when X and Y are continuous variables. The target law is identified by the formula

$$p(X, Y) = p(Y^* \mid X^*, R_X = 1, R_Y = 1)p(X^* \mid R_X = 1), \tag{5}$$

which can be seen by noting that $Y \perp\!\!\!\perp R_Y \,|\, X$ and $X \perp\!\!\!\perp R_X$. The conditions of Theorem 2.6 are thus satisfied by the ordering $X \prec Y$ which implies that a factorizable imputation method also exists. Applying Corollary 2 by Mohan & Pearl (2021), we conclude that neither CCA nor ACA is a valid estimation method for the target law. We also see that the m-backdoor criterion (Mathur et al., 2025) is not applicable in this case. These conclusions on the applicability of CCA, ACA and the m-backdoor criterion apply also to the other graphs in Figure 1.

The graph of Figure 1b has an additional edge $R_Y \to R_X$ compared to Figure 1a. The full law is still identifiable and the identifying functional is

$$p(X, Y, R_X, R_Y) = p(Y^* \,|\, X^*, R_X = 1, R_Y = 1)p(X^*, R_X = 1)\frac{p(R_Y \,|\, X^*, R_X = 1)p(R_X|R_Y)}{p(R_X = 1 \,|\, R_Y)}. \tag{6}$$

Therefore, a conditionally complete imputation method exists also in this case. The target law is identified by the formula

$$p(X, Y) = p(Y^* \,|\, X^*, R_X = 1, R_Y = 1) \sum_{R_Y, Y^*} p(R_Y, Y^*)p(X^* \,|\, R_X = 1, R_Y, Y^*). \tag{7}$$

As neither X nor Y is independent from its response indicator, the conditions of Theorem 2.6 cannot be satisfied.

In the graph of Figure 1c, the full law is identifiable via the odds ratio factorization (Yun Chen, 2007) which can be used to first identify the missingness mechanism as

$$p(R_X, R_Y \,|\, X, Y) = \frac{\alpha(R_X, R_Y, X, Y)}{\beta(X, Y)} \tag{8}$$

where

$$\begin{aligned}
&\alpha(R_X, R_Y, X, Y) = p(R_X \,|\, R_Y{=}1, X, Y)p(R_Y \,|\, R_X{=}1, X, Y)\mathrm{OR}(R_Y, R_X \,|\, X, Y),\\
&\beta(X, Y) = \sum_{r_x, r_y} \alpha(R_X = r_x, R_Y = r_y, X, Y),\\
&\mathrm{OR}(R_Y, R_X \,|\, X, Y) = \frac{p(R_Y \,|\, R_X, X, Y)}{p(R_Y = 1 \,|\, R_X, X, Y)} \frac{p(R_Y = 1 \,|\, R_X = 1, X, Y)}{p(R_Y \,|\, R_X = 1, X, Y)}.
\end{aligned}$$

All terms on the right-hand side of equation (8) are identifiable as no colluders or self-censoring edges are present. This implies the identifiability of the full law and consequently the existence of a conditionally complete imputation method. The identifying functional for the target law is

$$p(X, Y) = \frac{p(R_X = 1)p(X^* \,|\, R_X = 1)}{p(R_X = 1 \,|\, Y^*, R_Y = 1)} p(Y^* \,|\, X^*, R_X = 1, R_Y = 1). \tag{9}$$

Like the previous graph, neither X nor Y is independent from its response indicator and the conditions of Theorem 2.6 cannot be satisfied.

The graph of Figure 1d has a colluder structure $X \to R_Y \leftarrow R_X$. We conclude that the full law is not identifiable and consequently a conditionally complete imputation method does not exist. The identifying formula for the target law is

$$p(X, Y) = p(Y^* \,|\, X^*, R_X = 1, R_Y = 1)p(X^* \,|\, R_X = 1). \tag{10}$$

We have that $Y \perp\!\!\!\perp R_Y \,|\, X$ and $X \perp\!\!\!\perp R_X$ implying that a factorizable imputation method can be used under the ordering $X \prec Y$.

The graph of Figure 1e exemplifies a case where neither the full law nor the target law can be identified. The non-identifiability of the full law follows again from the presence of the colluder structure $R_X \to R_Y \leftarrow X$. The non-identifiability of the target law was demonstrated in this case by Nabi & Bhattacharya (2023). Therefore neither conditionally complete imputation nor factorizable imputation can be applied.

In the graph of Figure 1f, X and Y are independent. The full law cannot be identified because of the colluder $X \to R_Y \leftarrow R_X$ but the target law is identified by the formula

$$p(X, Y) = p(X^* \mid R_X = 1) \sum_{R_X} p(R_X)(Y^* \mid R_Y = 1, R_X). \tag{11}$$

In this case, X and R_X are independent but Y is not independent from R_X and R_Y. Thus, the conditions of Theorem 2.6 are not fulfilled.

Next, we will focus on the large-sample estimation of $p(X, Y)$. For this purpose, we simulate observations from joint distributions $p(X^*, Y^*, R_X, R_Y)$ that factorize according to the graphs of Figure 1. The simulation examples are referred to as Example 1a, Example 1b, etc. We simulate data on X and Y from a bivariate normal distribution where the expectations are 0, the variance of X is 1, the variance of Y is 2, and the correlation between X and Y is $\sqrt{2}/2 \approx 0.71$ with the exception of Example 1f where the correlation is zero. The response indicator R_X is simulated from a Bernoulli distribution where $p(R_X = 1) = 0.7$ in Examples 1a and 1d, $p(R_X = 1) = 0.4 + 0.3R_Y$ in Example 1b, and $\text{logit}(p(R_Y = 1)) = X$ in Examples 1c, 1e and 1f. The response indicator R_Y is simulated from a Bernoulli distribution where $\text{logit}(p(R_Y = 1)) = X$ in Examples 1a, 1b and 1c and from a Bernoulli distribution where $\text{logit}(p(R_Y = 1)) = (X + 0.4)(2R_X - 1)$ in Examples 1d, 1e and 1f. In the latter case, the functional form is chosen so that X and R_X have a strong interaction when the graph has the colluder $X \to R_Y \leftarrow R_X$. As we are interested in the validity of imputation, we consider only large sample properties and set the sample size to one million with no repetitions because the sample variation is negligible. The observed data are a random sample from $p(X^*, Y^*, R_X, R_Y)$.

The simulation results in Table 2 are in agreement with the theoretical considerations. MIRI provides unbiased estimates in Examples 1a, 1b and 1c as expected and suffers from some bias in other examples. MI is always biased because the data are MNAR and therefore the imputation models cannot be correctly specified without the response indicators. FMI gives unbiased estimates in Examples 1a and 1d where the ordering $X \prec Y$ fulfills the conditions of Theorem 2.6 because $X \perp\!\!\!\perp R_X$ and $Y \perp\!\!\!\perp R_Y \mid X$. As the first step of FMI, the observed values of Y are set to be missing when $R_X = 0$. FMI draws imputations in the ordering $X \prec Y$ regardless of whether Theorem 2.6 applies. CCA and CCA provide biased estimates for the target law in all examples. The use of identifying functionals leads to unbiased estimates in Examples 1a, 1b, 1d, 1c and 1f where the target law is identifiable. In Example 1e, the target law is not identifiable and an identifying functional does not exist.

4.2 Examples with four variables

Figure 2 presents four graphs with four variables. The full law $p(X, W, Z, Y, R_X, R_W, R_Z, R_Y)$ is not identifiable in these graphs due to the colluders: $Z \to R_W \leftarrow R_Z$ in graphs 2a and 2d, $W \to R_Z \leftarrow R_W$ and $Y \to R_Z \leftarrow R_Y$ in graph 2b, and $W \to R_X \leftarrow R_W$ and $Z \to R_Y \leftarrow R_Z$ in graph 2c. Using Theorem 2.4, we conclude that a conditionally complete imputation method does not exist.

The target law $p(X, W, Z, Y)$ is identifiable in all graphs of Figure 2. In the graph of Figure 2a we can order the partially observed variables as $Z \prec W \prec X \prec Y$ to that satisfies the conditional independence restrictions required in Theorem 2.6, which are

$$\begin{aligned}
&Z \perp\!\!\!\perp R_Z, \\
&W \perp\!\!\!\perp R_Z, R_W \mid Z, \\
&X \perp\!\!\!\perp R_Z, R_W, R_X \mid Z, W, \\
&Y \perp\!\!\!\perp R_Z, R_W, R_X, R_Y \mid Z, W, X.
\end{aligned}$$

Thus, a directly factorizable imputation method draws imputations in the ordering $Z \prec W \prec X \prec Y$ from the following distributions:

$$\text{If } R_Z = 0, \text{impute } Z \text{ from } p(Z \mid R_Z = 1). \tag{12}$$

$$\text{If } R_Z R_W = 0, \text{impute } W \text{ from } p(W \mid Z, R_Z R_W = 1). \tag{13}$$

$$\text{If } R_Z R_W R_X = 0, \text{impute } X \text{ from } p(X \mid Z, W, R_Z R_W R_X = 1). \tag{14}$$

$$\text{If } R_Z R_W R_X R_Y = 0, \text{impute } Y \text{ from } p(Y \mid Z, W, X, R_Z R_W R_X R_Y = 1). \tag{15}$$

Table 2: Simulation results for the examples illustrated in Figure 1. The biases are reported for multiple imputation without response indicators (MI), multiple imputation with response indicators (MIRI), factorizable multiple imputation (FMI), complete case analysis (CCA), available case analysis (ACA), and estimation using the identifying functional.

			Bias					
Example	Statistic	Truth	MI	MIRI	FMI	CCA	ACA	IF
1a	$\mathrm{E}(X)$	0.00	0.03	0.00	0.00	0.41	0.00	0.00
	$\mathrm{E}(Y)$	0.00	0.11	−0.00	−0.00	0.41	0.41	−0.00
	$\mathrm{sd}(X)$	1.00	−0.00	−0.00	−0.00	−0.09	−0.00	−0.00
	$\mathrm{sd}(Y)$	1.41	−0.02	0.00	0.00	−0.06	−0.06	0.00
	$\mathrm{Cor}(X,Y)$	0.71	−0.01	−0.00	−0.00	−0.03	−0.03	−0.00
1b	$\mathrm{E}(X)$	0.00	0.14	−0.00	0.11	0.41	0.11	−0.00
	$\mathrm{E}(Y)$	0.00	0.20	−0.00	0.11	0.41	0.41	−0.00
	$\mathrm{sd}(X)$	1.00	−0.01	0.00	−0.01	−0.09	−0.01	0.00
	$\mathrm{sd}(Y)$	1.41	−0.02	0.00	−0.00	−0.06	−0.06	0.00
	$\mathrm{Cor}(X,Y)$	0.71	−0.01	−0.00	−0.00	−0.03	−0.03	−0.00
1c	$\mathrm{E}(X)$	0.00	0.22	0.00	0.36	0.67	0.36	0.00
	$\mathrm{E}(Y)$	0.00	0.29	0.00	0.73	0.99	0.41	0.00
	$\mathrm{sd}(X)$	1.00	−0.06	−0.00	−0.07	−0.14	−0.07	−0.01
	$\mathrm{sd}(Y)$	1.41	−0.07	0.00	−0.19	−0.23	−0.06	−0.00
	$\mathrm{Cor}(X,Y)$	0.71	−0.01	−0.00	−0.05	−0.08	−0.08	−0.01
1d	$\mathrm{E}(X)$	0.00	−0.03	−0.15	−0.00	0.35	−0.00	−0.00
	$\mathrm{E}(Y)$	0.00	−0.10	−0.29	−0.00	0.35	0.15	−0.00
	$\mathrm{sd}(X)$	1.00	0.00	0.03	0.00	−0.08	0.00	0.00
	$\mathrm{sd}(Y)$	1.41	0.01	0.07	0.00	−0.06	−0.01	0.00
	$\mathrm{Cor}(X,Y)$	0.71	0.01	0.02	0.00	−0.03	−0.03	−0.00
1e	$\mathrm{E}(X)$	0.00	0.12	−0.15	0.36	0.62	0.36	NA
	$\mathrm{E}(Y)$	0.00	0.01	−0.32	0.73	0.94	0.09	NA
	$\mathrm{sd}(X)$	1.00	−0.03	0.06	−0.07	−0.13	−0.07	NA
	$\mathrm{sd}(Y)$	1.41	0.10	0.19	−0.20	−0.23	0.12	NA
	$\mathrm{Cor}(X,Y)$	0.71	0.02	0.04	−0.06	−0.08	−0.08	NA
1f	$\mathrm{E}(X)$	0.00	−0.04	−0.07	0.00	0.35	0.00	0.00
	$\mathrm{E}(Y)$	0.00	0.07	−0.00	0.72	0.72	0.12	−0.00
	$\mathrm{sd}(X)$	1.00	−0.00	0.00	−0.00	−0.09	−0.00	−0.00
	$\mathrm{sd}(Y)$	1.41	−0.00	0.00	−0.20	−0.20	−0.00	0.00
	$\mathrm{Cor}(X,Y)$	0.00	0.19	0.04	0.00	0.00	0.00	−0.00

From the graph of Figure 2a, we also obtain the conditional independence restrictions $X \perp\!\!\!\perp Z \,|\, W$ and $Y \perp\!\!\!\perp W \,|\, X, Z$ that can be used to simplify the imputation models for X and Y, respectively.

In the graph of Figure 2b, an ordering required for the directly factorizable imputation does not exist. However, we can apply Theorem 2.9 for the partition $(\{X, W\}, \{Z, Y\})$. To see this, we need to check that the conditions of Theorem 2.9 are fulfilled. From the graph we see that $Z, Y \perp\!\!\!\perp R_X, R_W \,|\, X, W$ which is the requirement of condition (a) of Theorem 2.9. The identifiability of $p(R_{XW} = 1 \,|\, X, W)$ from $p(X^*, W^*, R_X, R_W)$ can be confirmed by applying Do-search, which fulfills the condition (b) of Theorem 2.9.

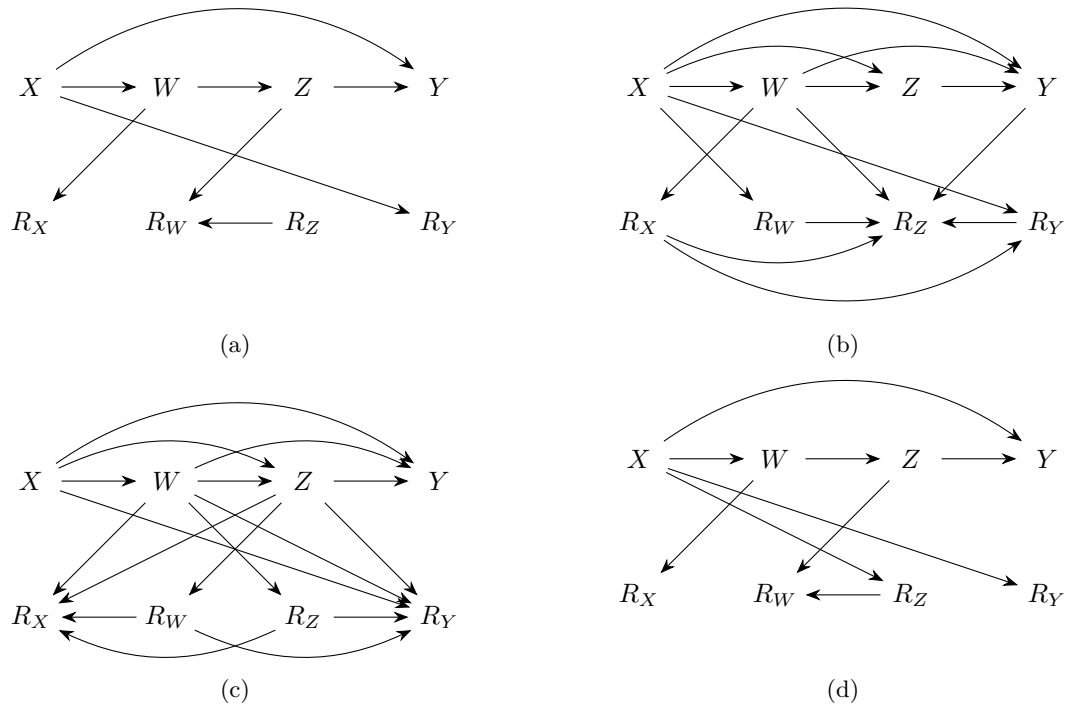

Figure 2: Graphs where the full law is not identifiable but the target law can be identified. In graph (a), a factorizable imputation method can be used in the ordering $Z \prec W \prec X \prec Y$. Weighting with imputation can be applied for the partition $(\{X, W\}, \{Z, Y\})$ in graph (b), and for the partition $(\{W, Z\}, \{X, Y\})$ in graph (c). In graph (d), the conditions of Theorem 2.6 or Theorem 2.9 are not satisfied but the target law is still identifiable. Proxy variables and edges related to them are not shown for simplicity.

This implies that we can apply conditionally complete imputation for X and W and use then the imputed data to estimate the probabilities $p(R_{XW} = 1 \,|\, X, W)$. Finally, condition (c) of Theorem 2.9 holds for the ordering $Y \prec Z$ because $Y \perp\!\!\!\perp R_Y \,|\, X, W, R_X, R_W$ and $Z \perp\!\!\!\perp R_Z, R_Y \,|\, Y, X, W, R_X, R_W$. This means that a conditionally factorizable imputation method draws imputation for the data where $R_X R_W = 1$ from the following distributions

$$\text{If } R_Y = 0, \text{impute } Y \text{ from } p(Y \,|\, X, W, R_X R_W R_Y = 1). \tag{16}$$

$$\text{If } R_Y R_Z = 0, \text{impute } Z \text{ from } p(Z \,|\, X, W, Y, R_X R_W R_Y R_Z = 1). \tag{17}$$

The reasoning is similar for Figure 2c. This time, Theorem 2.9 applies for the partition $(\{W, Z\}, \{X, Y\})$ and the ordering is $X \prec Y$. The full law $p(W, Z, R_W, R_Z)$ is identifiable from $p(W^*, Z^*, R_W, R_Z)$ licensing the use of conditionally complete imputation in the weight estimation. The conditions $X \perp\!\!\!\perp R_X \,|\, W, Z, R_W, R_Z$ and $Y \perp\!\!\!\perp R_Y, R_X \,|\, X, W, Z, R_W, R_Z$ hold permitting the factorizable imputation of X and Y given W and Z when $R_{WZ} = 1$. Thus, a conditionally factorizable imputation method draws imputation for the data where $R_W R_Z = 1$ from the following distributions

$$\text{If } R_X = 0, \text{impute } X \text{ from } p(X \,|\, W, Z, R_W R_Z R_X = 1). \tag{18}$$

$$\text{If } R_X R_Y = 0, \text{impute } Y \text{ from } p(Y \,|\, W, Z, X R_W R_Z R_X R_Y = 1). \tag{19}$$

The graph of Figure 2d exemplifies a case where neither Theorem 2.6 nor Theorem 2.9 is not applicable but the target law is still identifiable. The graph is similar to the graph of Figure 2a but there is an additional edge $X \to R_Z$. In this scenario, it is possible to derive multiple identifying functionals for the target law. We consider two of such functionals. The first functional is obtained using Do-search and it assigns weights

for the complete observations

$$p(X,W,Z,Y) = \frac{p(X,W,Z,Y,R_X=1,R_W=1,R_Z=1,R_Y=1)}{p(R_X=1|W,R_W=1)p(R_W=1|Z,R_Z=1)p(R_Z=1,R_Y=1|X,R_X=1)}. \tag{20}$$

The second functional is obtained from the first one by manual derivation

$$p(X,W,Z,Y) = \frac{p(R_X=1)p(X|R_X=1)p(Z|X,R_XR_Z=1)p(W,Y|X,Z,R_XR_WR_ZR_Y=1)}{p(R_X=1|W,R_W=1)}. \tag{21}$$

In Examples 2a – 2d, we simulate one million observations from a linear Gaussian model constructed according to the graphs of Figure 2a – 2d, respectively. The detailed description of the simulation models is presented in Appendix A. The observed data are a random sample from $p(X^*, W^*, Z^*, Y^*, R_X, R_W, R_Z, R_Y)$.

The methods in comparison are the same as in Example 1 with the addition of conditionally factorizable imputation. MI uses the partially observed variables X, W, Z and Y to impute each other. MIRI uses imputation models where these variables have interactions with the response indicators. CCA and ACA are included as well.

FMI operates differently in each example. In Example 2a, FMI uses modified data where W^* is set to be missing if $R_Z = 0$, X^* is set to be missing if $R_Z = 0$ or $R_W = 0$, and Y^* is set to be missing if $R_Z = 0$ or $R_W = 0$ or $R_X = 0$. The imputation models are specified according to equations (12), (13), (14) and (15).

In Example 2b, FMI uses first MIRI to obtain imputed data on $p(X, W, R_X, R_W)$. Then a generalized additive model is fitted to estimate probabilities $p(R_XR_W = 1 \,|\, X, W)$ whose inverses will be the analysis weights. The data are restricted to the subset where $R_XR_W = 1$ and Y and Z are imputed according to equations (16) and (17), respectively.

In Example 2c, FMI uses first MIRI to obtain imputed data on $p(W, Z, R_W, R_Z)$. Then a generalized additive model is fitted to estimate probabilities $p(R_WR_Z = 1 \,|\, W, Z)$ whose inverses will be the analysis weights. The data are restricted to the subset where $R_WR_Z = 1$ and X and Y are imputed according to equations (18) and (19), respectively.

In Example 2d, the conditions required for FMI are not met for any partition. However, for the comparison, FMI is applied in same way as in Example 2a. The identifying functionals of equations (20) and (21) are applied in estimation in Example 2d.

From Table 3, we can see that the estimates are unbiased for FMI and biased for the other methods in Examples 2a, 2b and 2c. In Example 2d, FMI also provides biased results as expected but estimation using equations (20) and (21) gives unbiased or nearly unbiased estimates. The same conclusions hold for standard deviations and correlations which are reported in Tables 4 and 5 in Appendix. The simulation results are thus in agreement with Theorems 2.4, 2.6 and 2.9.

5 Conclusion

We have studied the application of multiple imputation in graphical missing data models. First we clarified the connection between full law identifiability and standard multiple imputation. Theorem 2.4 ties the identifiability of the full law and the applicability of standard multiple imputation together. In practice, an analyst can first check if the full law is identifiable in the assumed missing data model. If this holds true, a conditionally complete imputation method exists and standard multiple imputation could be applied. If this turns out to be false, standard multiple imputation is likely to lead to biased estimates because it is theoretically not possible to devise a conditionally complete imputation method. The presented simulation examples are in agreement with Theorem 2.4. Standard multiple imputation provided unbiased estimates only in the cases where the full law is identified. However, this conclusion holds only if the imputation models include the response indicators.

We introduced factorizable imputation that can be used in some cases when the full law is not identifiable. Factorizable imputation is a non-standard multiple imputation method in the sense that certain observed

Table 3: Simulation results for the examples illustrated in Figure 2. The biases are reported for factorizable multiple imputation (FMI), standard multiple imputation without response indicators (MI), multiple imputation with response indicators (MIRI), complete case analysis (CCA) and available case analysis (ACA), and in the case of Example 2d, for estimation using the referred equations (Eq.).

		Bias				
Example	Statistic	FMI	MI	MIRI	CCA	ACA
2a	$\mathrm{E}(X) = 0$	0.00	0.09	0.00	0.72	0.29
	$\mathrm{E}(W) = 0$	0.00	−0.01	−0.06	0.78	0.11
	$\mathrm{E}(Z) = 0$	0.00	−0.01	−0.04	0.69	−0.00
	$\mathrm{E}(Y) = 0$	0.00	0.08	0.00	0.71	0.31
2b	$\mathrm{E}(X) = 0$	−0.00	−0.03	−0.11	0.69	−0.25
	$\mathrm{E}(W) = 0$	−0.00	0.04	0.03	0.45	0.25
	$\mathrm{E}(Z) = 0$	−0.00	0.05	0.02	0.57	0.42
	$\mathrm{E}(Y) = 0$	−0.00	−0.01	−0.05	0.62	0.01
2c	$\mathrm{E}(X) = 0$	0.00	0.14	−0.07	0.81	0.39
	$\mathrm{E}(W) = 0$	0.01	0.13	−0.09	0.92	0.28
	$\mathrm{E}(Z) = 0$	0.00	0.14	−0.09	0.93	0.28
	$\mathrm{E}(Y) = 0$	0.01	0.17	−0.15	0.90	0.54

Example 2d	Bias						
Statistic	FMI	MI	MIRI	CCA	ACA	Eq. (20)	Eq. (21)
$\mathrm{E}(X) = 0$	0.41	0.10	0.00	0.95	0.29	−0.01	−0.01
$\mathrm{E}(W) = 0$	0.29	−0.03	-0.09	0.92	0.03	−0.00	−0.01
$\mathrm{E}(Z) = 0$	0.21	0.00	-0.06	0.78	0.21	−0.02	−0.00
$\mathrm{E}(Y) = 0$	0.31	0.10	0.00	0.87	0.31	−0.02	−0.01

values are also imputed and weights are used in the analysis of the imputed data. Examples 1a, 2a, 2b and 2c demonstrate cases where the full law is not identified but factorizable imputation leads to unbiased estimates. These results are in agreement with Theorems 2.6 and 2.9. We studied only large-sample estimation. Investigating the small-sample properties of different estimation methods is a natural next step for future research.

There are also cases where the target law is identifiable but neither standard multiple imputation nor factorizable imputation is applicable. In these cases, the target law can be still estimated using the identifying functionals as Examples 1f and 2d demonstrate.

There is a growing understanding that the MCAR / MAR / MNAR classification could be replaced by more detailed non-parametric assumptions which are often expressed in the form of graphical models (Lee et al., 2023). Clarifying the link between multiple imputation and the identifiability of the full law and the target law helps researchers in choosing the right tool for missing data analysis in this context.

References

Lauren J Beesley, Irina Bondarenko, Michael R Elliot, Allison W Kurian, Steven J Katz, and Jeremy MG Taylor. Multiple imputation with missing data indicators. *Statistical Methods in Medical Research*, 30 (12):2685–2700, 2021.

Rohit Bhattacharya, Razieh Nabi, Ilya Shpitser, and James M. Robins. Identification in missing data models represented by directed acyclic graphs. In *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, volume 115, pp. 1149–1158, 2020.

Rhian M Daniel, Michael G Kenward, Simon N Cousens, and Bianca L De Stavola. Using causal diagrams to guide analysis in missing data problems. *Statistical Methods in Medical Research*, 21(3):243–256, 2012.

Jacques-Emmanuel Galimard, Sylvie Chevret, Camelia Protopopescu, and Matthieu Resche-Rigon. A multiple imputation approach for MNAR mechanisms compatible with Heckman's model. *Statistics in Medicine*, 35(17):2907–2920, 2016.

Anna Guo, Jiwei Zhao, and Razieh Nabi. Sufficient identification conditions and semiparametric estimation under missing not at random mechanisms. In Robin J. Evans and Ilya Shpitser (eds.), *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, volume 216 of *Proceedings of Machine Learning Research*, pp. 777–787, 2023.

Angelina Hammon and Sabine Zinn. Multiple imputation of binary multilevel missing not at random data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 69(3):547–564, 2020.

Juha Karvanen. Study design in causal models. *Scandinavian Journal of Statistics*, 42(2):361–377, 2015.

Katherine J Lee, John B Carlin, Julie A Simpson, and Margarita Moreno-Betancur. Assumptions and analysis planning in studies with missing data in multiple variables: moving beyond the MCAR/MAR/MNAR classification. *International Journal of Epidemiology*, 52(4):1268–1275, 2023.

Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*. John Wiley & Sons, 3rd edition, 2019.

Maya B Mathur, Willi Zhang, and Ilya Shpitser. Imputation without nightMARs: Graphical criteria for valid imputation of missing data, 2025. URL `http://dx.doi.org/10.31219/osf.io/zqne9`.

Karthika Mohan and Judea Pearl. Graphical models for processing missing data. *Journal of the American Statistical Association*, 116(534):1023–1037, 2021.

Karthika Mohan, Judea Pearl, and Jin Tian. Graphical models for inference with missing data. In *Advances in Neural Information Processing Systems*, volume 26, 2013.

Johanna Muñoz, Orestis Efthimiou, Vincent Audigier, Valentijn MT de Jong, and Thomas PA Debray. Multiple imputation of incomplete multilevel data using Heckman selection models. *Statistics in Medicine*, 43(3):514–533, 2024.

Razieh Nabi and Rohit Bhattacharya. On testability and goodness of fit tests in missing data models. In *Uncertainty in Artificial Intelligence*, pp. 1467–1477, 2023.

Razieh Nabi, Rohit Bhattacharya, and Ilya Shpitser. Full law identification in graphical models of missing data: Completeness results. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pp. 7153–7163, 2020.

Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

D. B. Rubin. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons Inc., New York, 1987.

Donald B Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

Shaun Seaman, John Galati, Dan Jackson, and John Carlin. What is meant by "missing at random"? *Statistical Science*, 28(2):257–268, 2013.

Jin Tian. Recovering probability distributions from missing data. In *Proceedings of the 9th Asian Conference on Machine Learning*, volume 77, pp. 574–589, 2017.

Santtu Tikka and Juha Karvanen. Monotone missing data: A blessing and a curse. *Transactions on Machine Learning Research*, 11, 2025.

Santtu Tikka, Antti Hyttinen, and Juha Karvanen. Causal effect identification from multiple incomplete data sources: A general search-based approach. *Journal of Statistical Software*, 99(5):1–40, 2021.

Daniel Mark Tompsett, Finbarr Leacy, Margarita Moreno-Betancur, Jon Heron, and Ian R. White. On the use of the not-at-random fully conditional specification (NARFCS) procedure in practice. *Statistics in Medicine*, 37(15):2338–2353, 2018.

Stef Van Buuren. *Flexible imputation of missing data*. CRC press, Boca Raton, FL, 2018.

Stef Van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45:1–67, 2011.

Hua Yun Chen. A semiparametric odds ratio model for measuring association. *Biometrics*, 63(2):413–421, 2007.

A Detailed description of simulation setup and simulation results

The data for Examples 2a–2d are simulated as follows. Variables X, W, Z and Y follow multivariate normal distribution with zero means. The standard deviations and pairwise correlation coefficients can be read from the column "Truth" in Tables 4 and 5. Response indicators without parents are simulated from a Bernoulli distribution where the probability of 1 is 0.7. Response indicators with one parent (always X, W, Z or Y in our examples) are simulated from a logistic regression model. Logistic regression is used also when a response indicator has multiple parents which may be both data variables and other response indicators. The linear predictor then contains high order interaction terms in addition to linear terms. The exact parametrization varies example by example and can be seen from the simulation code.

Table 4 presents additional simulation results for Examples 2a, 2b and 2c. It can be seen that estimated standard deviations and correlations are unbiased for factorizable imputation and biased for other approaches. Table 5 shows additional simulation results for Example 2d. The estimated standard deviations and correlations are biased for all methods except for the estimation with identifying functionals.

Table 4: Simulation results on standard deviations and correlations for the examples illustrated in Figures 2a, 2b and 2c. The biases are reported for factorizable multiple imputation (FMI), standard multiple imputation without response indicators (MI), multiple imputation with response indicators (MIRI), complete case analysis (CCA) and available case analysis (ACA).

					Bias		
Example	Statistic	Truth	FMI	MI	MIRI	CCA	ACA
2a	$\text{sd}(X)$	1.00	−0.00	−0.03	−0.00	−0.13	−0.04
	$\text{sd}(W)$	1.00	−0.00	0.00	−0.01	−0.14	−0.01
	$\text{sd}(Z)$	1.00	−0.00	0.00	−0.02	−0.12	−0.00
	$\text{sd}(Y)$	1.00	−0.00	−0.04	−0.00	−0.12	−0.05
	$\text{Cor}(X, W)$	0.71	−0.00	−0.03	−0.01	−0.08	−0.04
	$\text{Cor}(X, Z)$	0.50	−0.00	−0.03	−0.01	−0.10	−0.05
	$\text{Cor}(X, Y)$	0.75	−0.00	−0.02	−0.00	−0.06	−0.05
	$\text{Cor}(W, Z)$	0.71	−0.00	0.00	−0.01	−0.07	−0.03
	$\text{Cor}(W, Y)$	0.71	−0.00	−0.02	−0.01	−0.08	−0.04
	$\text{Cor}(Z, Y)$	0.75	−0.00	−0.01	−0.02	−0.06	−0.01
2b	$\text{sd}(X)$	1.00	−0.00	−0.00	0.02	−0.26	−0.04
	$\text{sd}(W)$	1.00	−0.00	−0.01	−0.03	−0.30	−0.04
	$\text{sd}(Z)$	1.00	0.00	−0.02	−0.02	−0.25	−0.09
	$\text{sd}(Y)$	1.00	−0.00	−0.02	−0.04	−0.31	−0.12
	$\text{Cor}(X, W)$	0.71	−0.00	−0.04	−0.07	−0.22	−0.07
	$\text{Cor}(X, Z)$	0.81	0.00	−0.02	−0.06	−0.14	−0.04
	$\text{Cor}(X, Y)$	0.86	−0.00	−0.01	−0.03	−0.11	−0.04
	$\text{Cor}(W, Z)$	0.81	−0.00	−0.01	−0.02	−0.16	−0.05
	$\text{Cor}(W, Y)$	0.86	−0.00	−0.01	−0.03	−0.15	−0.06
	$\text{Cor}(Z, Y)$	0.89	0.00	−0.01	−0.01	−0.09	−0.04
2c	$\text{sd}(X)$	1.00	−0.00	−0.03	0.09	−0.27	−0.07
	$\text{sd}(W)$	1.00	−0.00	−0.03	0.11	−0.32	−0.05
	$\text{sd}(Z)$	1.00	−0.00	−0.04	0.12	−0.35	−0.05
	$\text{sd}(Y)$	1.00	−0.00	−0.05	0.25	−0.33	−0.17
	$\text{Cor}(X, W)$	0.71	−0.00	−0.02	0.04	−0.30	−0.07
	$\text{Cor}(X, Z)$	0.81	−0.00	−0.01	0.03	−0.19	−0.04
	$\text{Cor}(X, Y)$	0.86	−0.00	−0.01	0.02	−0.14	−0.13
	$\text{Cor}(W, Z)$	0.81	−0.00	−0.01	0.03	−0.24	−0.05
	$\text{Cor}(W, Y)$	0.86	−0.00	−0.01	0.02	−0.17	−0.12
	$\text{Cor}(Z, Y)$	0.89	−0.00	−0.01	0.02	−0.14	−0.10

Table 5: Simulation results on standard deviations and correlations for the example illustrated in Figure 2d. The biases are reported for factorizable multiple imputation (FMI), standard multiple imputation without response indicators (MI), multiple imputation with response indicators (MIRI), complete case analysis (CCA) and available case analysis (ACA), and for estimation using the referred equations (Eq.).

		Bias						
Statistic	Truth	FMI	MI	MIRI	CCA	ACA	Eq. (20)	Eq. (21)
$\mathrm{sd}(X)$	1.00	-0.09	-0.03	-0.00	-0.18	-0.04	0.00	0.00
$\mathrm{sd}(W)$	1.00	-0.04	0.03	0.02	-0.16	0.06	-0.01	0.00
$\mathrm{sd}(Z)$	1.00	-0.02	-0.00	-0.03	-0.13	-0.02	0.01	-0.00
$\mathrm{sd}(Y)$	1.00	-0.05	-0.03	-0.00	-0.14	-0.05	0.02	0.00
$\mathrm{Cor}(X, W)$	0.71	-0.03	-0.00	0.01	-0.09	0.00	-0.00	0.00
$\mathrm{Cor}(X, Z)$	0.50	-0.03	0.00	0.00	-0.12	-0.08	0.02	0.00
$\mathrm{Cor}(X, Y)$	0.75	-0.03	-0.01	-0.00	-0.08	-0.05	0.01	0.00
$\mathrm{Cor}(W, Z)$	0.71	-0.02	0.02	-0.00	-0.08	-0.05	0.00	0.00
$\mathrm{Cor}(W, Y)$	0.71	-0.03	-0.01	-0.01	-0.10	-0.00	-0.00	0.00
$\mathrm{Cor}(Z, Y)$	0.75	-0.01	-0.00	-0.02	-0.07	-0.02	0.00	-0.00