Towards an MI-proper Predictive Mean Matching

Philipp Gaffert[1], Florian Meinfelder[2] and Volker Bosch[1]

[1]Division Marketing & Data Sciences, GfK SE, Nordwestring 101, 90419 Nuremberg, Germany

[2]Econometrics and Statistics, Otto-Friedrich-University, Feldkirchenstrasse 21, 96052 Bamberg, Germany

January 25, 2016

Abstract

Statistical analysis with missing data is commonly conducted using the concept of Multiple Imputation (MI) [35]. Predictive Mean Matching (PMM) has become a very popular semi-parametric method within the MI framework to impute values from the support of an incomplete variable. Moreover, it can be shown that PMM is more robust to model misspecification than purely parametric methods. However, these benefits come at the price of deviating from the Bayesian framework of MI, where imputations are based on draws from the posterior predictive distribution of the missing data. At present, several different PMM algorithms in MI software exist, but no theoretical foundation has yet been provided that shows either of the matching variants to be superior. We review all the existing software implementations and show their MI-improperness. As a consequence of the identified problems, we develop a new PMM procedure called *MIDAStouch* that largely builds on ideas by Siddique & Belin 2008 [42] and is publicly available as an *R* package [26]. This new method is finally investigated within a simulation study, where it is compared to existing software implementations. We find that the MI-improperness of the existing PMM procedures leads to an attenuation bias of the total variance estimate. The simulation study reveals that the bias is large for small sample sizes, but can be avoided by applying the proposed *MIDAStouch* procedure.

Keywords: Multiple imputation, Predictive Mean Matching, Approximate Bayesian Bootstrap, Distance-Based Donor Selection

1 Introduction

> 'Multiple imputation is a statistical technique designed to take advantage of the flexibility in modern computing to handle missing data. With it, each missing value is replaced by two or more imputed values in order to represent the uncertainty about which value to impute.'

These few first words in Rubin's seminal book on this matter [35] fully express the rationale of multiple imputation. For a more formal representation we let Y denote a multidimensional variable that can be split into an observed part Y_{obs}, and a missing part Y_{mis}, such that $Y = [Y_{obs}, Y_{mis}]$. The missing data can be imputed under a Bayesian framework by making random draws from the posterior predictive distribution

$$f(Y_{mis}|Y_{obs}) = \int_{\psi\in\Psi} f(\psi|Y_{obs}) f(Y_{mis}|\psi, Y_{obs}) d\psi, \tag{1}$$

where ψ are the imputation model parameters. Since direct draws from $f(Y_{mis}|Y_{obs})$ are usually not feasible, MI algorithms typically make random draws from the observed-data posterior distribution $f(\psi|Y_{obs})$, followed by subsequent random draws from the conditional predictive distribution of the missing data $f(Y_{mis}|\psi, Y_{obs})$ [29, p88]. The components of this typical MI procedure are referred to as the Posterior (P) step and the Imputation (I) step, respectively [21, p201]. However, in order to avoid treating imputed and observed values alike, these two steps are repeated $M \geq 2$ times, yielding M data sets which are identical for Y_{obs}, but typically not for Y_{mis}. Applying Rubin's rules leads to an additional variance component for an estimated parameter that accounts for the uncertainty induced by the missing data, thus making frequentist inference more conservative. The detailed combining rules and theoretical justification can be found in [35].

MI has developed into one of the dominating strategies on handling incomplete data, various books have been published, and algorithm are implemented in numerous statistical software packages, such as R, SAS, SPSS, and Stata.

1.1 The evolution of MI algorithms

Probably the first publicly available MI algorithm was *NORM* by Joseph Schafer in 1997 [39, p399]. The *NORM* algorithm assumes that $Y \sim N_p(\mu, \Sigma)$[1]. Random draws from the posterior distribution are composed

[1]There are extensions for e.g. categorical variables [39, p399].

of drawing $\Sigma|Y_{obs}$ from an inverse Wishart distribution, and subsequently drawing $\mu|\Sigma, Y_{obs}$ from a multivariate normal. The missing values are imputed by drawing $Y_{mis}|\mu, \Sigma, Y_{obs}$ from a multivariate normal as well. The procedure is MCMC based for non-monotone missing patterns, which, however, need not be within the scope of this paper. If Y is normally distributed, this so-called joint modeling algorithm leads to proper multiple imputation [35, p118f.][39, p105f.].

Multivariate distributions are hard to define for real data sets, though. Hence a different class of algorithms emerged that is based on conditional univariate imputations [27]. The advantage of this so-called fully conditional specification (FCS) approach[2] is the flexibility in terms of choosing an appropriate imputation model for each variable. The conditional predictive distribution of the missing data can now be denoted as $f(Y_{j,mis}|Y_{j,obs}, Y_{-j}, \psi_j)$. For non-monotone missingness patterns this procedure is akin to Gibbs sampling [10].

The *MICE* (multivariate imputations by chained equations) algorithm by Stef van Buuren and Karin Oudshoorn was published in 2000, and has laid the foundation for many similar FCS algorithms like *Stata::ice* [31], *IVEware* in *SAS* [28], or the *SPSS::MI* routines [15]. Meanwhile the FCS approach has gained the upper hand over joint modeling.[3] The work in this paper adds to the FCS algorithms, too.

1.2 The evolution of PMM algorithms

It is important to note that all the early FCS implementations were purely parametric, and could be considered proper MI algorithms, if the joint distribution of Y existed [22], and the distributional assumptions were correct. While FCS led to a considerable gain in flexibility, it did not solve the problem of potential implausible imputations. Well defined conditional predictive distributions are simply not capable of reflecting the empirical distributions of survey data, appropriately. This is where Predictive Mean Matching (PMM, [34, p92], [20]) steps in.

PMM is a semi-parametric hot-deck imputation method [41, p429] that is now not only implemented in numerous software packages (see table 1) but is even the default procedure for continuous variables in many of them. For instance, *MICE* is comprising PMM as the default for continuous variables ever since its very first version [49, p33]. The reasons for PMM's popularity within MI algorithms are multifold. Compared to a fully parametric imputation, PMM is less sensitive to model misspecification [41, p429], namely non-linear associations, heteroscedastic residuals, and deviations from normality [24, p4]. An illustrative example is given in van Buuren 2012 [47, p70]. Moreover, PMM imputations are considered plausible, because the imputed values are observed in the data set [41, p429]. However, the quality of PMM imputations largely depends upon the availability of near donors. Applying PMM in e.g. truncation settings does not make any sense [18, p38].

1.3 Linking MI and PMM

Although the concept of multiple imputation has been combined with PMM many times (see table 1), there is some skepticism around. Little & Rubin 2002 state about PMM [21, p69]

> '... properties of estimates derived from such matching procedures remain largely unexplored.'

Koller-Meinfelder 2009 notes [18, p32]

> 'The difficult part about Predictive Mean Matching is to utilize its robust properties within the Multiple Imputation framework in a way that Rubin's combination rules still yield unbiased variance estimates.'

And, recently Morris, White & Royston 2014 warned in the same context [24, p5]

> '... there is thus no guarantee that Rubin's rules will be appropriate for inference.'

This theoretical uncertainty along with the prominence of PMM in practice motivated our work. The remainder of this paper is structured as follows. The sections 2 and 3 introduce into both, the details of the previous work we refer to and the applied notation. Section 4 reveals the differences between the existing PMM algorithms and the theory of multiple imputation. In section 5 we present a new PMM procedure that we call *MIDAStouch* and that overcomes the identified differences while being largely built on the work by Siddique & Belin 2008 [42]. Section 6 presents a simulation study comparing all major PMM software implementations.

2 The univariate case

Let the data be univariate normal

$$y_{n\times 1} \sim N(\mu, \sigma_y^2) \tag{2}$$

[2] also referred to as 'switching regressions', 'sequential regressions' or most commonly 'chained equations'

[3] As an indicator we refer to R-Studio [32] reporting the number of *R*-package [26] installations from their own servers on `http://cran-logs.rstudio.com/`. In the whole year 2015 the FCS flagship package *R::mice* [48] was installed more than six times as often as the joint modeling flagship package *R::norm* [40] (64,671 times versus 10,015 times).

Software	Command /Instructions	Match types available[A]	Option to specify match type	Default value of k[B]	Option to specify k	Parameter uncertainty[C]	Prediction of donors / recipients[D]	Source of information
R::mice	mice.impute.pmm	1	-	5	donors = #	parametric	in/out	[48]
R::Hmisc	aregImpute	1,2,3	pmmtype = #	3	kclosest = #	bootstrap	in/out	[12]
R::BaBooN	BBPMM	2	-	1	-	Bayesian Bootstrap	in/in	[23]
R::mi	.pmm	2	-	1	-	parametric	in/in	[9]
SAS::proc mi	regpmm	1	-	5	K = #	parametric	in/out	[38]
SAS::MIDAS	MIDAS	2	-	n_{obs}	-	Approximate Bayesian Bootstrap	in/out	[43]
Solas	Analyze → Multiple Imputation → Predictive Mean Matching method ...	2	-	10	Select 'Use # closest cases' option in 'Donor pool' tab.	parametric	in/out	[46]
SPSS	multiple imputation /impute scalemodel = PMM	2	-	1	-	parametric	in/out	[15]
Stata	mi impute pmm	2	-	1	knn(#)	parametric	in/out	[44]
Stata	ice, match	1,2,3	matchtype(#)	3	matchpool(#)	parametric	in/out	[31]

[A] see section 4.2 [B] see section 4.3 [C] see section 2.1 [D] *in* and *out* mean in-sample and out-of-sample prediction, respectively, see section 4.1

Table 1: Summary of existing software implementations based on [24, p3]

We denote n_{mis} as the number of missing values or recipients and $n_{obs} = n - n_{mis}$ as the number of non-missing values or donors in y. We assume ignorability for the missingness as

$$P(y = missing) = \alpha_0 \neq f(y), 0 < \alpha_0 < 1 \tag{3}$$

In this simple case fully parametric multiple imputation can be carried out by repeating the following two steps $M \geq 2$ times to correctly reflect the uncertainty about the parameters of the imputation model.

1. The P-step
   (a) Draw $\tilde{\sigma}_y^2 | y_{obs} \sim \Gamma^{-1}(\frac{n_{obs}}{2}, \frac{1}{2} \cdot \sum_{i=1}^{n_{obs}} (y_i - \hat{\bar{y}}_{obs})^2)$.
   (b) Draw $\tilde{\mu} | y_{obs}, \tilde{\sigma}_y^2 \sim N(\hat{\bar{y}}_{obs}, \frac{\tilde{\sigma}_y^2}{n_{obs}})$.
2. The I-step
   (a) Draw n_{mis} times from $\tilde{y} \sim N(\tilde{\mu}, \tilde{\sigma}_y^2)$.

This procedure corresponds to method number four in Little & Rubin 2002 [21, p216].

2.1 The P-step and the bootstrap

The P-step of the above procedure is substituted by a bootstrap in a number of existing software packages (see table 1) as originally proposed in Heitjan & Little 1991 [13, p18]. Instead of drawing M times from the posterior distributions of the parameters to reflect their uncertainty, the Maximum-Likelihood (ML) estimates of M independent bootstrap samples are utilized. This procedure corresponds to method number six in Little & Rubin 2002 [21, p216].

It may be worth noting that even though the packages claim to use different bootstrap procedures they are doing very much the same thing. In the initial paper [33, p131] on the Bayesian Bootstrap (used by *R::BaBooN* [23]) Rubin 1981 already shows that its statistical properties are very similar to those of the simple bootstrap [8]. The Approximate Bayesian Bootstrap (ABB) for the P-step (used by *SAS::MIDAS* [43]) is even just a simple bootstrap [42, p85].

2.1.1 The Approximate Bayesian Bootstrap

However, the ABB for the univariate case as presented in the original paper, is a kind of shortcut-imputation [36, p368]. Rather than drawing parameters in the P-step the ABB just draws a bootstrap sample. And, rather than drawing from the conditional predictive distribution in the I-step, the ABB just draws from the empirical bootstrap sample considering the integer bootstrap weights ω. This model-free imputation procedure, however, is unique to the univariate case. Kim 2002 [16, p472] showed that the ABB delivers correct inferences for $n_{obs} \to \infty$ only. This is due to the fact that the bootstrap estimator ignores the appropriate degrees of freedom correction just like the Maximum-Likelihood estimator [6, p22]. Thus, for finite n_{obs} the total parameter variance is underestimated. Parzen et al. 2005 [25, p973] suggest the following bias-correction factor PLF:

$$PLF = \frac{\frac{n^2}{n_{obs}} + \frac{n_{mis}}{M} \cdot \left(\frac{n-1}{n_{obs}} - \frac{n}{(n_{obs})^2} \right)}{\frac{n^2}{n_{obs}} + \frac{n_{mis}}{M} \cdot \left(\frac{n-1}{n_{obs}} - \frac{n}{(n_{obs})^2} \right) - \frac{n \cdot n_{mis}}{n_{obs}} \cdot \left(\frac{3}{n} + \frac{1}{n_{obs}} \right)} \tag{4}$$

Some criticism about this correction factor stems from Demirtas et al. 2007 [7].

2.2 The I-step and Predictive Mean Matching (PMM)

PMM [34, p92] substitutes the draw from the conditional predictive distribution. This was first described by Little 1988 [20, p292]. Translating his approach to our easy example from the equations (2) and (3) gives:

1. Calculate the (predictive) mean for the n_{obs} elements of y as $\hat{\bar{y}}_{obs} = \frac{1}{n_{obs}} \cdot \sum_{i=1}^{n_{obs}} y_i$.
2. Draw the (predictive) mean for the n_{mis} elements of y as $\tilde{\mu} | y_{obs}, \tilde{\sigma}_y^2 \sim N(\hat{\bar{y}}_{obs}, \frac{\tilde{\sigma}_y^2}{n_{obs}})$.
3. Match each element of $\tilde{\mu}_{y_{mis}}$ to the its respective closest element of $\hat{\bar{y}}_{obs}$.
4. Impute the observed y_{obs} of the closest matches.

Note that in this univariate case all recipients are equally far away from all donors in terms of their predictive means. I.e., the procedure described just samples randomly with replacement from the set of donors in the sample. Hence, the total parameter variance of this procedure is even smaller than the already downward biased variance of the ABB, that samples randomly from a bootstrap sample (see section 2.1.1).

3 The multivariate case

The univariate case is especially important for understanding the ABB (see section 2.1.1). Even though PMM can be applied in a univariate setting (see section 2.2) it was developed for the multivariate case. This section is to formerly introduce PMM.

Suppose that in addition to the variable with missing values (see equation (2)) we observe a set of $p-1$ fully observed variables.

$$y_{n\times 1}, X_{n\times(p-1)} \sim N_p(\mu, \Sigma) \tag{5}$$

We still assume ignorability for the missingness, now a little bit more complex,

$$P(y = missing) = \Phi\left(X^* \cdot \alpha + \eta\right) \tag{6}$$

where X^* denotes the X matrix from above with a leading constant column, η i.i.d. normal noise and Φ the normal c.d.f.. The correct conditional imputation model is the linear model.

$$y = X^* \cdot \beta + \epsilon \tag{7}$$

with ϵ denoting i.i.d. normal noise with zero mean and variance σ^2_ϵ. Note that the missing pattern is monotone and thus no Gibbs sampler iterations are necessary [47, p104]. Fully parametric imputation can be carried out in a very similar fashion as in the univariate case.

1. The P-step
   (a) Draw $\tilde{\sigma}^2_\epsilon | y_{obs}, X_{obs} \sim \Gamma^{-1}(\frac{n_{obs}}{2}, \frac{1}{2} \cdot (\hat{\epsilon}^{ML}_{obs})' \cdot (\hat{\epsilon}^{ML}_{obs}))$
   (b) Draw $\tilde{\beta} | y_{obs}, X_{obs}, \tilde{\sigma}^2_\epsilon \sim N_p(\hat{\beta}^{ML}_{obs}, \tilde{\sigma}^2_\epsilon \cdot (X^{*'}_{obs} \cdot X^*_{obs})^{-1})$
2. The I-step
   (a) Draw independently from $\tilde{y}_{mis} \sim N(X_{mis} \cdot \tilde{\beta}, \tilde{\sigma}^2_\epsilon)$.

A detailed description is given in van Buuren 2012 [47, p58]. Again, the P-step oftentimes is substituted by a bootstrap [47, p59]. Little's 1988 PMM for the multivariate case is conducted as follows [20, p292].

1. Calculate the predictive mean for the n_{obs} elements of y as $\hat{y}_{obs} = X^*_{obs} \cdot \hat{\beta}^{ML}_{obs}$.
2. Calculate the predictive mean for the n_{mis} elements of y as $\hat{\tilde{y}}_{mis} = X^*_{mis} \cdot \tilde{\beta}$.
3. Match each element of $\hat{\tilde{y}}_{mis}$ to the its respective closest element of $\hat{y}_{obs}$.
4. Impute the observed y_{obs} of the closest matches.

4 Why PMM does it wrong

Now that we have presented the fundamentals we may turn to our criticism. We identified four imputer's degrees of freedom in the specification of PMM. We will refer to those four items in each of the upcoming sections as follows:

#	**imputer's degrees of freedom**	**introduction section**	**improvement section**	**simulation section**
1	In-sample versus out-of-sample predictions	4.1	5.1	6.2.1
2	Type-1 versus Type-2 matching	4.2	5.2	6.2.2
3	Neighbor selection	4.3	5.3	6.2.3
4	Considering the uncertainty about all parameters	4.4	5.3	6.2.4

This section shall give an introduction and point out the issues in the current implementations. It may be worth noting that the items 2 and 3 are frequently discussed in the existing literature (see e.g. [47]), whereas we are doing pioneer work in this section by shedding light on the items 1 and 4 in the PMM context.

4.1 In-sample versus out-of-sample predictions

Little's 1988 PMM (see section 3) proposes to estimate the model parameters based on all the donors and to use them for the prediction of $\hat{y}$ for both, the donors and the recipients. Hence, in this setting the donor predictions are in-sample whereas the recipient predictions are out-of-sample. Especially, for small n_{obs} the model is closer to the donors, because it is optimized in this way, than to the recipients (for a proof assuming the simplest model see appendix 8.1). Consequently, the residual variance added to the recipients will be too small.

This implementation is still the rule. We found two exceptions, though. The procedures by the packages *R::mi* [9] and *R::BaBooN* [23] estimate the parameters on the full set of observations while using previously imputed values for y_{mis}.

4.2 Type-1 versus Type-2 matching

Little's 1988 procedure of matching $\hat{y}_{obs}$ to $\hat{\tilde{y}}_{mis}$ (see section 3) has later been called Type-1 matching. In contrast, matching $\hat{\tilde{y}}_{obs}$ to $\hat{\tilde{y}}_{mis}$ is called Type-2 matching [13, p19]. Van Buuren 2012 argues that following Little's 1988 procedure but using Type-2 matching in the case where only one predictor is present the M multiple imputations are exactly alike and therefore parameter uncertainty does not propagate at all [47, p71]. Few packages offer Type-3 matching (see table 1), that draws twice from the set of parameters, once for the donors and once for the recipients and then matches $\hat{\tilde{y}}^{I}_{obs}$ to $\hat{\tilde{y}}^{II}_{mis}$.

4.3 Neighbor selection

Little's 1988 nearest neighbor approach (see section 3), the so-called deterministic hot-deck [1, p44], is a special case of the general *k-nearest-neighbor* selection [13, p16], that is most commonly applied in today's software packages. An adaptive procedure to choose the optimal k has been developed, but is hardly used in practice [41]. Van Buuren 2012 discusses this issue in detail and states that $k = 1$ led to selecting the same donor over and over again and was therefore undesirable. Choosing a large k resolved this issue but hindered the procedure to preserve the correlation structure [47, p71].

We want to add to this discussion by focusing on the point estimate for the variance of y. If the distributions of the donors and recipients are roughly comparable then a large k will increase the probability for the donors closer to the center to give their value to the recipients closer to the bounds. That inevitably decreases the variance of y (for a proof see appendix 8.2). Hence, the estimate of the variance of y based on the imputed data is downward biased for larger k.

4.4 The consideration of $\hat{\sigma}^2$'s uncertainty

Note that Little's 1988 procedure (see section 3) for all M imputations draws from a conditional predictive distribution described by $N(X^*_{mis} \cdot \tilde{\beta}, \hat{\sigma}^2_\epsilon)$ rather than from $N(X^*_{mis} \cdot \tilde{\beta}, \tilde{\sigma}^2_\epsilon)$. In other words, the uncertainty about the imputation model parameter $\hat{\sigma}^2_\epsilon$ does not at all propagate in the PMM approaches as they are implemented in a broad variety of software packages. Type-2 matching makes things better, but not well enough. Notice that in Type-2 matching the residual variance varies among the M imputations, however, only due to the variation by $\tilde{\beta}$. In other words, Type-2 matching assumes that

$$Var(\hat{\sigma}^2_\epsilon)|\beta = 0 \tag{8}$$

whereas it is 'more realistic in most cases' to assume that [11, p57]

$$Var(\hat{\sigma}^2_\epsilon)|\beta = Var(\hat{\sigma}^2_\epsilon) > 0 \tag{9}$$

The only exception is *SAS::MIDAS* [43]. We could not find any evidence on the authors (or anybody else) being aware of this issue, though. Their approach, however, is somewhat different from Little's 1988 original procedure and is therefore described below.

4.4.1 Distance-based donor selection by Siddique & Belin 2008

The distance-based donor selection uses the donor's bootstrap weights not only for the P-step, but also for the I-step. This ensures that the parameter uncertainty about $\hat{\sigma}^2$ propagates. For recipient j donor i from the full donor pool is drawn with probability

$$P(i \overset{imputes}{\rightarrow} j) = f(\omega, \hat{\tilde{y}}_{obs}, \hat{\tilde{y}}_j, \kappa) = \frac{\omega_i \cdot d_{i,j}^{-\kappa}}{\sum_{i=1}^{n_{obs}} (\omega_i \cdot d_{i,j}^{-\kappa})} \tag{10}$$

ω denotes the non-negative integer bootstrap weights of the donors, $d_{i,j}$ the scalar absolute distance between the predictive means of donor i and recipient j and κ a 'closeness' parameter adjusting the importance of the distance. For $\kappa = 0$ the procedure is equivalent to the Approximate Bayesian Bootstrap (see section 2.1.1), for $\kappa \to \infty$ the procedure becomes equivalent to the deterministic hot deck (see section 4.3). Siddique & Belin 2008 propose $\kappa = 3$, which has become the default in *SAS::MIDAS*.

5 Towards a proper I-step substitute using PMM

Having identified the issues in current PMM implementations we show now how to do it better.

5.1 In-sample versus out-of-sample predictions

We follow the common approach of estimating the imputation model parameters using the donors only [27, p94]. Hence, in-sample prediction for the recipients (see e.g. [23]) is impossible. We propose to estimate n_{obs} sets of parameters to obtain the donor predictions by the leave-one-out principle. This way, the donor predictions are out-of-sample, too, avoiding the variance underestimation described in section 4.1. The distance between the i'th donor and the j's recipient is then calculated as follows assuming Type-2 matching.

$$d_{ij} = |(x_i - x_j) \cdot \tilde{\beta}_{-i}| \tag{11}$$

x_i denotes the row-vector of X^* for the i'th donor, x_j the row-vector of X^* for the j'th recipient and $\tilde{\beta}_{-i}$ a random draw from the distribution of $\hat{\beta}_{-i}$ not conditional on the i'th donor data.

5.2 Type-1 versus Type-2 matching

The parametric I-step is conditional on one set of parameters drawn in the P-step. Type-2 matching does it likewise. Both, Type-1 and Type-3 matching, however, condition on an extra set of parameters, which is the Maximum-Likelihood estimate for the former and an additional draw for the latter. So, neither Type-1 nor Type-3 matching can substitute the I-step appropriately. We therefore argue to use Type-2 matching. Note that van Buuren's zero between-variance criticism (see section 4.2) does not apply if the uncertainty about $\hat{\sigma}^2$ is considered properly as in the *MIDAS* procedure (see section 4.4.1).

5.3 Neighbor selection and the consideration of $\hat{\sigma}^2$'s uncertainty

We generally suggest to use the distance-based donor selection by Siddique & Belin 2008 (see section 4.4.1). In addition to section 5.1 we introduce two slight modifications to the *MIDAS* procedure below. We call this touched up version of *MIDAS MIDAStouch*.

5.3.1 Estimate a value for κ from the data

Rather than using a fixed κ we recommend to set

$$\kappa(R^2_{obs}) = \left(\frac{50 \cdot R^2_{obs}}{1 + \delta - R^2_{obs}} \right)^{\frac{3}{8}} \tag{12}$$

where R^2_{obs} denotes the coefficient of determination based on the full donor set and δ a very small positive scalar number to ensure real results also for the unlikely case of $R^2_{obs} = 1$. The functional form is the inverse of Little's 2004 sales response to advertising function [19, p1845]. The key idea is that the better y can be explained by X^* the more important the $d's$, i.e. the conditionality on X^*, become in equation (10). Note that

$$\frac{\partial \kappa}{\partial R^2_{obs}} > 0 \tag{13}$$

In the extreme case, where $R^2_{obs} = 0$ the *MIDAS* procedure considering equation (12) is equivalent to the Approximate Bayesian Bootstrap (see section 2.1.1). Siddique & Belin 2008 further state that reasonable values for κ are within $[0; 10]$ [42, p88] and found in a simulation study that the ideal value for κ is 3 [42, p98] in a setting with $R^2 = .29$. Equation (12) reflects these findings as follows:

$$\kappa(R^2_{obs} = 0) = 0 \tag{14}$$

$$\kappa(R^2_{obs} = .9) \approx 10 \tag{15}$$

$$\kappa(R^2_{obs} = .29) \approx 3 \tag{16}$$

5.3.2 Apply the PLF correction to the total variance

Since the *MIDAS* concept is based on the ABB it suffers from the same shortcomings as the ABB itself, namely the underestimation of the total variance for finite n_{obs} (see section 2.1.1). We recommend to apply the PLF correction factor (see equation (4)). However, since the drawing probabilities are now heterogeneous after conditioning on the bootstrap weights a slight adaption will be necessary. We propose to substitute n_{obs} by a measure of the effective sample size n_{eff} [17, p427] and consequently substitute n by $n_{\mathit{eff}} + n_{mis}$. The expression is $n_{\mathit{eff}} = n^2_{obs} / \sum \left(\frac{w_1}{w_0} \right)^2$ [3, p5]. w_1 denotes the applied weights, that in our application stem from the *MIDAS* procedure (see equation (10)) and w_0 the starting (i.e. bootstrap) weights ω. Averaging over all recipients and the M imputed data sets gives:

$$n_{\mathit{eff}} = \frac{1}{M \cdot n_{mis}} \cdot \sum_{m=1}^{M} \sum_{j=1}^{n_{mis}} \left(\sum_{i=1}^{n_{obs}} \left(\frac{d^{-\kappa_m}_{i,j,m}}{\sum_{i=1}^{n_{obs}} (\omega_{i,m} \cdot d^{-\kappa_m}_{i,j,m})} \right)^2 \right)^{-1} \tag{17}$$

Such a variance correction has yet been developed for the mean only. A starting point for a variance correction of linear regression parameters can be found in Wu 1986 [50, p1280].

6 A simulation study

We stated above that all existing software packages suffered from the theoretical shortcomings of PMM. This section presents a simulation study in order to illustrate the magnitude of both, the identified shortcomings and the proposed improvements. To give a full picture we compare our proposed method to all PMM software packages listed by Morris et al. [24] (see table 1)[4]. Furthermore we compare to two benchmark algorithms, a fully parametric one utilizing the additional information of a normal likelihood (*mice.impute.norm*) and a fully improper PMM approach that treats the Maximum-Likelihood parameter estimates as the truth (*pmm.nob*).

6.1 Simulation setup

For simplicity we refer to the multivariate normal setting presented above (see section 3) and set all off-diagonal elements of the correlation matrix equal. To recognize different challenges in real-world applications we set up a full factorial design considering the following four binary factors.

1. Missing completely at random (*MCAR*, i.e. $\alpha_{-1} = 0$) versus missing at random (*MAR*, i.e. no restrictions on α in equation (6)) [21, p12]. We operationalize *MAR* as $P(y = missing) = \Phi\left(\frac{1}{4} \cdot (X_1 + N(0, 3))\right)$.
2. Number of covariates $p-1 = 1$ versus $p-1 = 8$. We want to address van Buuren's zero between-variance criticism about Type-2 matching in the presence of one predictor only (see section 4.2).
3. Coefficient of determination $R^2 = 0$ versus $R^2 = .75$. The former is very similar to the univariate case (see section 2) whereas the latter might be more realistic.
4. Number of donors $n_{obs} = 10$ versus $n_{obs} = 200$. Our main criticism is that the existing software packages mostly ignore $\hat{\sigma}^2$'s uncertainty (see section 4.4). This effect should become particularly obvious for a small number of donors.

Furthermore, we fix $M = 25$, $n_{mis} = 100$, all marginal means at zero, all marginal variances at one and the number of Monte Carlo simulations at $n_{sim} = 250$ for each combination.

6.2 Simulation results

We focus on the estimates of the mean of y, denoted as $\hat{\bar{y}}$, and of the regression coefficient of X_1 in the linear regression model of y on X^*, denoted as $\hat{\beta}_1$. Utilizing the concept of multiple imputation [21, p211] and the appropriate degrees of freedom [2] we construct 95% frequentist confidence intervals. For each simulation run we note whether or not the true parameter value is covered by such a confidence interval. For each cell in the results tables (see tables 2 and 3) we average the coverages over $\frac{2^4}{2^2} \cdot n_{sim} = 1,000$ simulation runs. Since we expect the number of donors n_{obs} to be the most important factor we present all its interactions.

In section 4 we have developed four imputer's degrees of freedom in the specification of PMM and have pointed out the associated issues in the existing software implementations. In the remainder of this section we review these issues in the light of the simulation results and with a focus on the relative performance of the proposed *MIDAStouch* procedure.

6.2.1 In-sample versus out-of-sample predictions

Most existing implementations mix in-sample and out-of-sample predictions (see table 1). The two exceptions, namely *R::BaBooN* and *R::mi*, perform in-sample predictions for both donors and recipients. However, the results of these two implementations seem rather in line with all others than outstanding. All *MIDAStouch* procedures perform out-of-sample predictions for donors and recipients (see section 5.1). This seems to be a significant improvement over the existing *SAS::MIDAS* implementation (see tables 2 and 3) with an overall coverage of 905‰ versus 857‰ for $n_{obs} = 10$ and 947‰ versus 939‰ for $n_{obs} = 200$.

6.2.2 Type-1 versus Type-2 matching

For the larger number of covariates ($p - 1 = 8$) the matching type does not seem influencial. However, for one covariate only ($p - 1 = 1$) Type-2 matching along with the deterministic hot-deck leads to a between variance estimate of zero (see section 4.2). This effect can be observed in the respective column of the simulation results tables (see tables 2 and 3). The coverages of the implementations *R::mi*, *SPSS* and *Stata::mi* are equal to the *R::pmm.nob* benchmark coverages. Even though all *MIDAStouch* procedures are built on Type-2 matching no such effect can be observed for them, which is what we have expected (see section 5.2).

[4] We excluded *Solas* here for technical reasons. *Solas* neither provides a batch mode nor a loop functionality to efficiently handle many incomplete data sets coming along naturally with a simulation study.

PMM procedures		Missing mechanism				Number of covariates				Coefficient of determination				overall			source
		MCAR		MAR		$p-1=1$		$p-1=8$		$R^2=0$		$R^2=.75$					
coverages of 95% confidence intervals in ‰		$\hat{\bar{y}}$	$\hat{\beta}_1$	$\hat{\bar{y}}$	$\hat{\beta}_1$	$\hat{\bar{y}}$	$\hat{\beta}_1$	$\hat{\bar{y}}$	$\hat{\beta}_1$	$\hat{\bar{y}}$	$\hat{\beta}_1$	$\hat{\bar{y}}$	$\hat{\beta}_1$	$\hat{\bar{y}}$	$\hat{\beta}_1$	both	
Proposed procedure[A]																	
R::MIDAStouch	mice.impute.MIDAStouch	967	938	905	878	948	819	924	997	976	979	896	837	936	908	922	[26],[48]
	PLF correction	991	-	955	-	972	-	974	-	988	-	958	-	973	-	-	
R::MIDAStouch	mice.impute.MIDAStouch(kappa=3)	968	914	894	843	951	757	911	1000	970	972	892	785	931	879	905	[26],[48]
	PLF correction	985	-	934	-	963	-	956	-	990	-	929	-	960	-	-	
Benchmark procedures																	
R::mice	mice.impute.norm	964	963	960	973	970	960	954	976	962	966	962	970	962	968	965	[26],[48]
R	pmm.nob[B]	479	265	286	226	396	251	369	240	314	271	451	220	383	246	314	[26]
PMM software packages listed by Morris et al. 2014 [24, p3]																	
R::mice	mice.impute.pmm	732	738	476	710	597	508	611	940	699	939	509	509	604	724	664	[26],[48]
R::Hmisc	aregImpute[C]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	[26],[12]
R::BaBooN	BBPMM	771	686	600	655	658	519	713	822	764	815	607	526	686	671	678	[26],[23]
R::mi	.pmm	704	650	443	601	396	251	751	1000	583	648	564	603	574	626	600	[26],[9]
SAS::proc mi	regpmm	616	680	358	675	435	412	539	943	557	854	417	501	487	678	582	[37],[38]
SAS::MIDAS	MIDAS	930	888	802	808	850	707	882	989	913	966	819	730	866	848	857	[37],[43]
SPSS	multiple imputation /impute scalemodel = PMM	718	626	561	584	396	251	883	959	604	622	675	588	640	605	622	[14],[15]
Stata	mi impute pmm	687	608	521	565	396	251	812	922	574	608	634	565	604	587	595	[45],[44]
Stata	ice, match	578	750	307	750	446	500	439	1000	574	1000	311	500	443	750	596	[45],[31]

[A] *R::MIDAStouch* is available at `https://cran.r-project.org/web/packages/midastouch/index.html`. It uses the open *R::mice* framework.
[B] This procedure is not implemented anywhere, it just serves as a bad benchmark.
[C] Unfortunately, for small sample sizes the procedure keeps returning errors that we are unable to resolve without making major changes inside.

Table 2: Simulation results with $n_{obs} = 10$.

PMM procedures		Missing mechanism				Number of covariates				Coefficient of determination				overall			source
		MCAR		MAR		$p-1=1$		$p-1=8$		$R^2=0$		$R^2=.75$					
coverages of 95% confidence intervals in ‰		$\hat{\bar{y}}$	$\hat{\beta}_1$	$\hat{\bar{y}}$	$\hat{\beta}_1$	$\hat{\bar{y}}$	$\hat{\beta}_1$	$\hat{\bar{y}}$	$\hat{\beta}_1$	$\hat{\bar{y}}$	$\hat{\beta}_1$	$\hat{\bar{y}}$	$\hat{\beta}_1$	$\hat{\bar{y}}$	$\hat{\beta}_1$	both	
Proposed procedure[A]																	
R::MIDAStouch	mice.impute.MIDAStouch	949	951	941	960	943	941	947	970	954	974	936	937	945	956	950	[26],[48]
	PLF correction	973	-	979	-	976	-	976	-	964	-	988	-	976	-	-	
R::MIDAStouch	mice.impute.MIDAStouch(kappa=3)	951	939	940	957	947	920	944	976	956	958	935	938	946	948	947	[26],[48]
	PLF correction	985	-	982	-	986	-	981	-	985	-	982	-	984	-	-	
Benchmark procedures																	
R::mice	mice.impute.norm	949	951	942	962	945	954	946	959	951	962	940	951	946	957	951	[26],[48]
R	pmm.nob[B]	885	866	867	855	874	851	878	871	843	855	909	866	876	861	868	[26]
PMM software packages listed by Morris et al. 2014 [24, p3]																	
R::mice	mice.impute.pmm	948	944	933	956	936	940	945	960	941	968	940	932	941	950	945	[26],[48]
R::Hmisc	aregImpute	936	940	929	947	924	939	941	948	926	958	939	929	933	944	938	[26],[12]
R::BaBooN	BBPMM	941	948	932	952	932	934	941	966	938	976	935	924	937	950	943	[26],[23]
R::mi	.pmm	908	925	906	908	874	851	940	982	894	914	920	919	907	917	912	[26],[9]
SAS::proc mi	regpmm	931	932	924	943	920	911	935	964	927	948	928	927	928	938	933	[37],[38]
SAS::MIDAS	MIDAS	941	934	933	948	928	918	946	964	939	950	935	932	937	941	939	[37],[43]
SPSS	multiple imputation /impute scalemodel = PMM	907	916	906	900	874	851	939	965	892	915	921	901	907	908	907	[14],[15]
Stata	mi impute pmm	914	913	907	900	874	851	947	962	899	912	922	901	911	907	909	[45],[44]
Stata	ice, match	943	942	927	955	932	931	938	966	936	966	934	931	935	949	942	[45],[31]

[A] *R::MIDAStouch* is available at `https://cran.r-project.org/web/packages/midastouch/index.html`. It uses the open *R::mice* framework.
[B] This procedure is not implemented anywhere, it just serves as a bad benchmark.

Table 3: Simulation results with $n_{obs} = 200$.

6.2.3 Neighbor selection

In section 4.3 we argued that the point estimate for the variance of y is downward biased for larger k. We cannot clearly see this effect in the simulation results tables that show the coverages only (see tables 2 and 3). This is why, we had a closer look. For the $n_{obs} = 10$ simulation runs the mean point estimate for the variance of y (the true variance is 1) is .848 and .773 for all implementations with $k = 1$ and $k > 1$, respectively[5]. This difference is highly significant. For the $n_{obs} = 200$ runs the differences diminish, because k is small relative to the number of donors, the mean point estimates are .995 and .992.

The mean point estimates for the variance of y of the proposed *MIDAStouch* implementation are .822 and .999 for $n_{obs} = 10$ and $n_{obs} = 200$, respectively.

6.2.4 The consideration of $\hat{\sigma}^2$'s uncertainty

SAS::MIDAS and hence the *MIDAStouch* procedures are much closer to the 95% coverage than any other existing PMM software implementation (see table 2). This result supports our criticism about ignoring $\hat{\sigma}^2$'s uncertainty in many implementation (see section 4.4).

6.2.5 The proposed *MIDAStouch* procedure

We used the open *R::mice* framework to implement the *MIDAStouch* procedure. All implementations presented in the top boxes of the tables 2 and 3 differ from the originally proposed *MIDAS* procedure by performing out-of-sample predictions only (see section 4.1). Additionally, we explicitly distinguish between a fixed closeness parameter κ as originally proposed and our suggested variable κ (see equation (12)). Also, we show separately the effect of the modified PLF correction (see equations (4) and (17)).

Especially the results from table 2 ($n_{obs} = 10$) indicate that our proposed touching up of the *MIDAS* procedure has led to an improvement. This seems to be true for all means, i.e. the out-of-sample predictions for the donors (compare *SAS::MIDAS* to *R::MIDAStouch* with $\kappa = 3$), the modified closeness parameter (compare *R::MIDAStouch* with $\kappa = 3$ to the *R::MIDAStouch* presented two lines above) and the application of the PLF correction (compare the *R::MIDAStouch* to the second lines given), the latter being available for the $\hat{\bar{y}}$ only. The results of the fully implemented *R::MIDAStouch* with the PLF correction are never below the 95% threshold and thus seem rather conservative.

7 Conclusion

We have found that the existing implementations of multiple imputation Predictive Mean Matching generally lead to overly progressive inferences and why this is so (see section 4). The propagation of parameter uncertainty, the key idea behind multiple imputation, is the main issue. From there on we have identified the *MIDAS* procedure [42] as the only one doing things right in this aspect (see section 4.4.1). It is based on the ideas of the Approximate Bayesian Bootstrap (see section 2.1.1) and Predictive Mean Matching (see section 3). In section 5 we have presented our proposed touched up version of *MIDAS* called *MIDAStouch*. The simulation study results (see section 6.2) clearly show that the *MIDAS* procedure is superior to all other implementations and that the proposed *MIDAStouch* implementation is a significant improvement towards the existing implementation in *SAS* and thus the 'golden' way of doing multiple PMM imputation.

[5] Due to the relative small number of donors the domain of X is smaller for the donors than it is for the recipients. This is a case of truncation as described in [18, p38] and the reason for the attenuation bias of the variance point estimate. This effect is amplified by deviations from the MCAR assumption.

8 Appendix

8.1 In-sample versus out-of-sample prediction

Consider the univariate case, where both, the donors and the recipients are from the same population. The mean squared deviation of the donors from the model is

$$V_{don} = \frac{1}{n_{obs}} \cdot \sum_{i=1}^{n_{obs}} (y_i - \hat{\bar{y}}_{obs})^2 \tag{18}$$

Introducing the true mean by adding $0 = \mu - \mu$ gives [5, p26]

$$\begin{aligned} V_{don} &= \frac{1}{n_{obs}} \cdot \sum_{i=1}^{n_{obs}} ((y_i - \mu) - (\hat{\bar{y}}_{obs} - \mu))^2 & (19) \\ &= \frac{1}{n_{obs}} \cdot \left(\sum_{i=1}^{n_{obs}} (y_i - \mu)^2 \right) - (\hat{\bar{y}}_{obs} - \mu)^2 & (20) \end{aligned}$$

Analogously, the mean squared deviation of the recipients from the model is

$$\begin{aligned} V_{rec} &= \frac{1}{n_{mis}} \cdot \sum_{j=1}^{n_{mis}} ((y_j - \mu) - (\hat{\bar{y}}_{obs} - \mu))^2 & (21) \\ &= \frac{1}{n_{mis}} \cdot \left(\sum_{j=1}^{n_{mis}} (y_j - \mu)^2 \right) + \hat{\bar{y}}_{obs} \cdot (\hat{\bar{y}}_{obs} - 2 \cdot \hat{\bar{y}}_{mis}) - \mu \cdot (\mu - 2 \cdot \hat{\bar{y}}_{mis}) & (22) \end{aligned}$$

Taking the difference and utilizing the homoscedasticity assumption, we get

$$\begin{aligned} V_{don} - V_{rec} &= \frac{1}{n_{obs}} \cdot \left(\sum_{i=1}^{n_{obs}} (y_i - \mu)^2 \right) - \frac{1}{n_{mis}} \cdot \left(\sum_{j=1}^{n_{mis}} (y_j - \mu)^2 \right) + 2 \cdot (\hat{\bar{y}}_{obs} - \hat{\bar{y}}_{mis}) \cdot (\mu - \hat{\bar{y}}_{obs}) & (23) \\ &= 2 \cdot (\hat{\bar{y}}_{obs} - \hat{\bar{y}}_{mis}) \cdot (\mu - \hat{\bar{y}}_{obs}) & (24) \end{aligned}$$

For a large recipient sample, $\hat{\bar{y}}_{mis} = E(\hat{\bar{y}}_{mis}) = \mu$ holds. Thus,

$$V_{don} - V_{rec} = -2 \cdot (\mu - \hat{\bar{y}}_{obs})^2 \leq 0 \tag{25}$$

I.e., as long as the model based on the donor sample differs randomly from the true population model, the residual variance for the donors is smaller than the one for the recipients. This difference diminishes for $n_{obs} \to \infty$.

8.2 Variance underestimation due to kNN

Suppose that the predictive mean π is within the bounds $\pi \in [-.5, .5]$ and that the distribution of the donors is discrete and equidistant within this range, so $\pi_{obs} = (-.5, -.5 + \frac{1}{n_{obs}-1}, \ldots, -.5 + \frac{n_{obs}-1}{n_{obs}-1})$. Further suppose that the recipients are distributed in the exact same way, so $\pi_{obs} = \pi_{mis}$. We denote $n = n_{obs} = n_{mis}$ and for simplicity allow it to be uneven only. We assume that the predictive mean π also is the characteristic of interest. This may be the case in a multivariate setting where the fully observed variables perfectly determine the variable with missing values.

π_{mis} is imputed using π_{obs} leading to π_{imp}. We want to learn about the point estimate for the variance of π_{imp} as a function of the relative size of the neighborhood that is chosen from randomly for a single recipient. We define this relative size excluding the exact nearest neighbor as $\Theta = (\frac{0}{n-1}, \frac{1}{n-1}, \ldots, \frac{n-1}{n-1})$. We decompose the variance in a between and a within component

$$T^{\pi_{imp}}(\Theta) = B^{\pi_{imp}}(\Theta) + W^{\pi_{imp}}(\Theta) \tag{26}$$

where B denotes the interrecipient variance and W the intrarecipient variance. It can easily be seen that if the exact nearest neighbor is chosen the interrecipient variance of π_{imp} will equal the variance of π_{mis}.

$$T^{\pi_{imp}}(\Theta = 0) = B^{\pi_{imp}}(\Theta = 0) = Var(\pi_{mis}) \tag{27}$$

For larger Θ the intrarecipient variance increases according to the variance formula for the discrete uniform distribution as follows [30, p372].

$$W^{\pi_{imp}}(\Theta) = -\Delta W(\Theta) = \frac{\Theta^2}{12} + \frac{\Theta}{6 \cdot (n-1)} \tag{28}$$

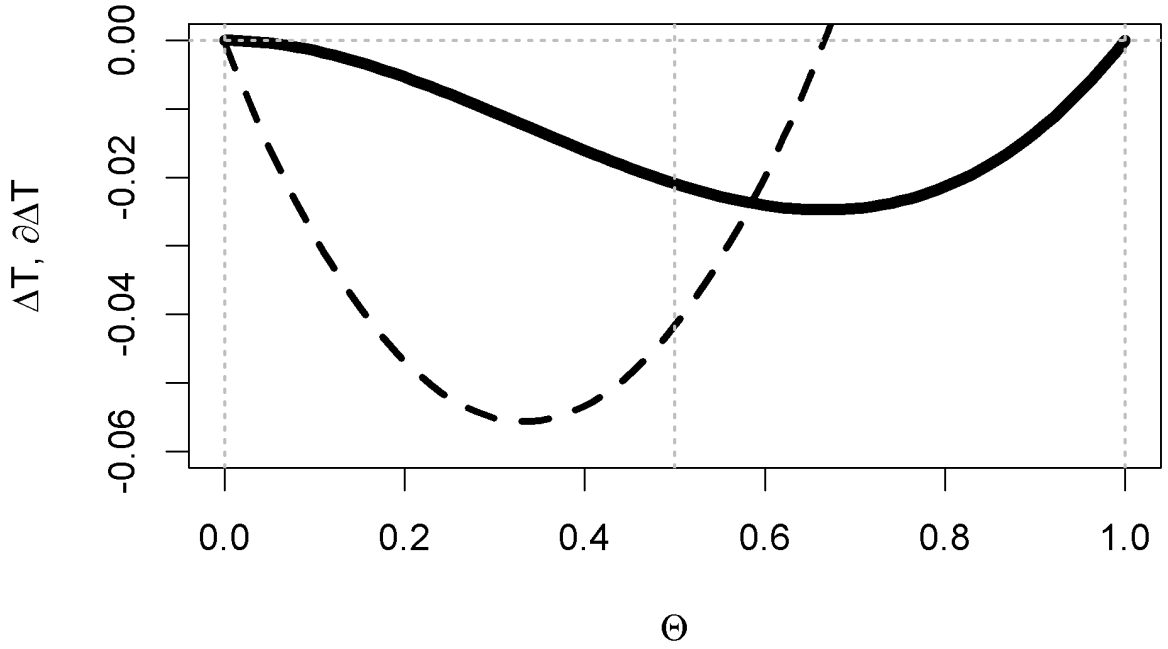

Figure 1: The solid line shows the bias of the point estimate of the variance and the dashed line its derivative, both plotted against $\Theta = \frac{k-1}{n_{obs}}$. The bias is zero for $k = 1$ or $k = n_{nobs}$ and small for small k relative to n_{obs}.

The interrecipient variance can be regarded as the variance of the expectations. The expectation of a uniform distribution is just the mean of its bounds. Since the range of π is limited to both sides, the interrecipient variance decreases with increasing Θ. More specifically we see for the left side, i.e. $\pi^i_{mis} < 0$

$$E\left(\pi^i_{imp}|\Theta, \pi^i_{mis} < \frac{\Theta - 1}{2}\right) = \frac{\Theta - 1}{2} \tag{29}$$

We assume that the mean of π is known to be zero. We can then write based on the left side

$$Var(\pi_{mis}) - B^{\pi_{imp}}(\Theta) = \Delta B(\Theta) = \frac{2}{n}\sum_{i=1}^{\frac{n-1}{2}}((\pi^i_{mis})^2 - (E(\pi^i_{imp}))^2) \tag{30}$$

We now focus just on the part of the left side for which $\pi^i_{mis} < \frac{\Theta-1}{2}$ holds. The rest can be ignored since all respective elements of the sum in equation (30) are zero. Then, using the assumption of equidistance and equation (29) we get

$$\Delta B(\Theta) = \frac{2}{n}\sum_{i=1}^{\frac{(n-1)\cdot\Theta}{2}}\left[\left(\frac{\Theta-1}{2}\right)^2 - \frac{(\Theta-1)\cdot i}{n-1} + \frac{i^2}{(n-1)^2} - \left(\frac{\Theta-1}{2}\right)^2\right] \tag{31}$$

The last term in equations (31) is also the last term in (30) and cancels out. A little bit of rewriting makes the series obvious that can be used for further simplification [4, p20]

$$\Delta B(\Theta) = \frac{2}{n}\left(\frac{1-\Theta}{n-1}\sum_{i=1}^{\frac{(n-1)\cdot\Theta}{2}} i + \frac{1}{(n-1)^2}\sum_{i=1}^{\frac{(n-1)\cdot\Theta}{2}} i^2\right) \tag{32}$$

Some more algebra leads to the third order polynomial

$$\Delta B(\Theta) = \frac{\Theta\cdot(\Theta\cdot(n-1)+2)\cdot(2\cdot\Theta\cdot(n-1)-3\cdot n+2)}{-12\cdot(n-1)\cdot n} \tag{33}$$

Adding $\Delta W(\Theta)$ gives

$$\Delta T(\Theta) = \frac{\Theta\cdot(\Theta-1)\cdot(\Theta\cdot(n-1)+2)}{6\cdot n} \tag{34}$$

with the two obvious roots at $\Theta = 0$ (see equation (27)) and $\Theta = 1$ where $B^{\pi_{imp}} = 0$. The third root does not exist given the limits for n and Θ. The first derivative is

$$\frac{\partial\Delta T(\Theta)}{\partial\Theta} = \frac{3\cdot\Theta^2\cdot(n-1) - 2\cdot\Theta\cdot(n-3) - 2}{6\cdot n} \tag{35}$$

For $n \to \infty$ $\Delta T(\Theta)$ has a minimum at $P_{min}(\Theta = \frac{2}{3}, \Delta T = -\frac{2}{81})$ and a falling inflection point at $P_{infl}(\Theta = \frac{1}{3}, \Delta T = -\frac{1}{81})$. We conclude that the point estimate for the variance of π_{imp} is downward biased for all Θ, but $\Theta = 0$ and $\Theta = 1$.

References

[1] Andridge, R. R., and Little, R. J. A review of hot deck imputation for survey non-response. *International Statistical Review 78*, 1 (2010), 40–64.

[2] Barnard, J., and Rubin, D. B. Small-sample degrees of freedom with multiple imputation. *Biometrika 86*, 4 (1999), 948–955.

[3] Bosch, V. A generalized measure of effective sample size. *Unpublished* (2005), 1–12.

[4] Bronstein, I. N., Semendjajew, K. A., and Musiol, G. *Taschenbuch der Mathematik*. Harri Deutsch, 2013.

[5] Cochran, W. G. *Sampling techniques*. John Wiley & Sons, 1977.

[6] Davison, A. C., and Hinkley, D. V. *Bootstrap methods and their application*, vol. 1. Cambridge university press, 1997.

[7] Demirtas, H., Arguelles, L. M., Chung, H., and Hedeker, D. On the performance of bias–reduction techniques for variance estimation in approximate bayesian bootstrap imputation. *Computational Statistics & Data Analysis 51*, 8 (2007), 4064–4068.

[8] Efron, B. Bootstrap methods: another look at the jackknife. *The Annals of Statistics 7*, 1 (1979), 1–26.

[9] Gelman, A., and Hill, J. Opening windows to the black box. *Journal of Statistical Software 40* (2011). R package version 1.0.

[10] Geman, S., and Geman, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 6 (1984), 721–741.

[11] Greenberg, E. *Introduction to Bayesian econometrics*. Cambridge University Press, 2013.

[12] Harrell, F. E., with contributions from Charles Dupont, and many others. *Hmisc: Harrell Miscellaneous*, 2015. R package version 3.16-0.

[13] Heitjan, D. F., and Little, R. J. Multiple imputation for the fatal accident reporting system. *Journal of the Royal Statistical Society C 40*, 1 (1991), 13–29.

[14] IBM Corp. IBM SPSS Statistics for Windows, Version 23.0, 2015. Armonk, NY: IBM Corp.

[15] IBM SPSS Statistics 23. IBM Knowledge Center Multiple Imputation. `http://www-01.ibm.com/support/knowledgecenter/SSLVMB_23.0.0/spss/mva/syn_multiple_imputation.dita?lang=en`, 2015. Retrieved 9/9/2015.

[16] Kim, J. A note on approximate bayesian bootstrap imputation. *Biometrika 89*, 2 (2002), 470–477.

[17] Kish, L. *Survey sampling*. John Wiley & Sons, 1965.

[18] Koller-Meinfelder, F. *Analysis of Incomplete Survey Data–Multiple Imputation via Bayesian Bootstrap Predictive Mean Matching*. PhD thesis, Dissertation, Otto-Friedrich-University Bamberg, 2009., 2009.

[19] Little, J. D. Models and managers: The concept of a decision calculus. *Management science 50*, 12_supplement (2004), 1841–1853.

[20] Little, R. J. Missing-data adjustments in large surveys. *Journal of Business & Economic Statistics 6*, 3 (1988), 287–296.

[21] Little, R. J., and Rubin, D. B. *Statistical analysis with missing data*. John Wiley & Sons, 2002.

[22] Liu, J., Gelman, A., Hill, J., Su, Y.-S., and Kropko, J. On the stationary distribution of iterative imputations. *Biometrika 101*, 1 (2014), 155–173.

[23] Meinfelder, F., and Schnapp, T. *BaBooN: Bayesian Bootstrap Predictive Mean Matching - Multiple and Single Imputation for Discrete Data*, 2015. R package version 0.2-0.

[24] Morris, T. P., White, I. R., and Royston, P. Tuning multiple imputation by predictive mean matching and local residual draws. *BMC medical research methodology 14*, 75 (2014), 1–13.

[25] Parzen, M., Lipsitz, S. R., and Fitzmaurice, G. M. A note on reducing the bias of the approximate bayesian bootstrap imputation variance estimator. *Biometrika 92*, 4 (2005), 971–974.

[26] R Core Team. *R 3.2.2: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015.

[27] Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., and Solenberger, P. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey methodology 27*, 1 (2001), 85–96.

[28] Raghunathan, T. E., Solenberger, P. W., and Van Hoewyk, J. Iveware: Imputation and variance estimation software. *Ann Arbor, MI: Survey Methodology Program, Survey Research Center, Institute for Social Research, University of Michigan* (2002).

[29] Rässler, S. *Statistical matching: A frequentist theory, practical applications, and alternative Bayesian approaches*, vol. 168. Springer, New York, 2002.

[30] Rinne, H. *Taschenbuch der Statistik*. Harri Deutsch, 2008.

[31] Royston, P., and White, I. R. Multiple imputation by chained equations (mice): implementation in Stata. *Journal of Statistical Software 45*, 4 (2011), 1–20.

[32] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA, 2015.

[33] Rubin, D. B. The bayesian bootstrap. *The Annals of Statistics 9*, 1 (1981), 130–134.

[34] Rubin, D. B. Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business & Economic Statistics 4*, 1 (1986), 87–94.

[35] Rubin, D. B. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, 1987.

[36] Rubin, D. B., and Schenker, N. Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association 81*, 394 (1986), 366–374.

[37] SAS Institute Inc. SAS software, version 9.4., 2015. University Edition.

[38] SAS Institute Inc. SAS/STAT(R) 13.1 user's guide: The mi procedure. `http://support.sas.com/documentation/cdl/en/statug/66859/HTML/default/viewer.htm#statug_mi_details08.htm`, 2015. Retrieved 9/9/2015.

[39] Schafer, J. L. *Analysis of incomplete multivariate data*. CRC press, 1997.

[40] Schafer, J. L., and Novo, A. A. *norm: Analysis of multivariate normal datasets with missing values*, 2013. R package version 1.0-9.5.

[41] Schenker, N., and Taylor, J. M. Partially parametric techniques for multiple imputation. *Computational Statistics & Data Analysis 22*, 4 (1996), 425–446.

[42] Siddique, J., and Belin, T. R. Multiple imputation using an iterative hot-deck with distance-based donor selection. *Statistics in medicine 27*, 1 (2008), 83–102.

[43] Siddique, J., and Harel, O. Midas: a SAS macro for multiple imputation using distance-aided selection of donors. *Journal of Statistical Software 29*, 9 (2009).

[44] StataCorp. *Stata 14 mi impute pmm Manual*, 2015. College Station, TX: Stata Press.

[45] StataCorp. *Stata Statistical Software: Release 14*, 2015. College Station, TX: StataCorp LP.

[46] Statistical Solutions Ltd. Solas version 4. `http://www.statsols.com/wp-content/uploads/2013/12/Solas-4-Manual1.pdf`, 2013. Retrieved 9/9/2015.

[47] van Buuren, S. *Flexible imputation of missing data*. Chapman & Hall/CRC, 2012.

[48] van Buuren, S., and Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software 45*, 3 (2011), 1–67. R package version 2.22.

[49] Van Buuren, S., and Oudshoorn, C. Multivariate imputation by chained equations. *MICE V1. 0 user's manual. Leiden: TNO Preventie en Gezondheid* (2000).

[50] Wu, C.-F. J. Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics* (1986), 1261–1295.