Imputation of restricted data
Tempelman, Caren

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2007

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 4

Imputation of Data Subject to Multiple Balance Restrictions

Economic data are subject to several linear restrictions. Common well-known imputation methods provide imputations that are likely not to satisfy the linear balance restrictions on economic data. In the previous chapter we discussed imputation procedures for variables that need to satisfy non-negativity constraints and one linear balance restriction. The method that was suggested, based on Dirichlet distribution, is however not capable of incorporating multiple linear balance restrictions. In this chapter we will therefore consider imputation methods for data items that need to satisfy multiple linear balance restrictions. We suggest using the singular normal distribution to model these data.

4.1 Introduction

The flexibility of the Dirichlet distribution, that was described in chapter 3, is an attractive feature for imputation purposes. Unfortunately, the Dirichlet method can only deal with one balance restriction at a time. If we are confronted with a more complex balance edit structure, where variables are present in more than one balance restriction, the Dirichlet distribution cannot be utilised.

This will be illustrated in the following example. The total operating ex-
expenses of a company can be subdivided into several items, such as the purchasing price of goods, labour costs, costs of housing, depreciation and so on. Most of these items are themselves also the total of another subdivision. Labour costs can, for example, be subdivided into gross wages, social security costs, pension charges and other social costs. Now assume we want to model a dataset consisting of the following variables: \(X_{11}, X_{12}, X_{13}, X_{14}, X_{1t}, X_{2t}, X_{3t}, X_{t}\), for which it holds that

\[X_{11} + X_{12} + X_{13} + X_{14} = X_{1t}\] and \[X_{1t} + X_{2t} + X_{3t} = X_{t}\].

The Dirichlet distribution can be employed to model the data if these two restrictions are combined into one balance restriction, e.g.

\[X_{11} + X_{12} + X_{13} + X_{14} + X_{1t} + 2X_{2t} + 2X_{3t} = 2X_{t}\].

Note, however, that in this case the fact that \(X_{11} + X_{12} + X_{13} + X_{14} = X_{1t}\) is not taken into account. This means that the imputed data may very well not satisfy this linear restriction. Consequently, the Dirichlet method cannot be applied in this manner. Another option is to use the Dirichlet distribution in conjunction with hierarchical imputation. In this instance \(X_{1t}, X_{2t}, X_{3t}\) are modelled first and subsequently \(X_{1t}\) is used to model \(X_{11}, X_{12}, X_{13}\) and \(X_{14}\). In an imputation context this procedure can, however, lead to inconsistencies as the observed values for \(X_{11}, X_{12}, X_{13}\) and \(X_{14}\) are not taken into account when imputing \(X_{1t}\). This means that the sum of the observed values may exceed the imputed total for \(X_{1t}\), which is not allowed.

So the Dirichlet distribution cannot be straightforwardly extended to the presence of multiple balance restrictions with interdependencies and therefore the need arises to develop an imputation method that can handle any type of linear balance edit structure. In this chapter we will discuss the use of a singular normal distribution for this purpose.

First the edit structure will be discussed in section 4.2. Next in section 4.3 the probability density function of the singular normal is derived, maximum likelihood estimation for completely observed data and in the presence of missing data is treated in sections 4.4 and 4.5 respectively. In section 4.6 it is shown that the maximum likelihood estimates and the imputed data concur with the balance restrictions. Deterministic and stochastic imputation methods are given in section 4.7 and in section 4.8 some results of these imputation procedures on empirical data are shown. Finally, in section 4.9 some concluding remarks will be given.
4.2 Balance edit restrictions

Consider an $n \times k$ data matrix $X$ and a $p \times k$ restriction matrix $A$, with $p$ the number of linear balance restrictions, where it holds that $AX' = 0$. For convenience we assume that every variable will be present in at least one linear balance restriction, this is not necessary however. Furthermore we assume that there are no redundant balance restrictions, which means that $A$ is of full row rank.

Note that the solution to this system of linear equations for respondent $i$ is

$$X_i = (I - A^{-1} A) Z_i,$$

where $I$ is an $k \times k$ identity matrix and $Z_i$ is an arbitrary vector of order $k \times 1$, see e.g. Rao (1973). The matrix $A^{-1}$ denotes a generalised inverse of $A$. A generalised inverse of the matrix $A$ is a matrix for which it holds that $AA^{-1} A = A$. If $A$ is square, and consequently nonsingular, there is only one solution for $X_i$: $X_i = 0$.

The multivariate singular normal distribution is useful for modelling data subject to several linear balance restrictions as the singularity, that has arisen due to the restrictions, is immediately taken into account. This method can also be applied for the case where $p = 1$, i.e. if there is only one linear balance restriction present.

Non-negativity restrictions, however, are not taken into account as the singular normal is defined on an affine subspace of $\mathbb{R}^k$. Using natural logarithmic transformations of the data items would seem to be a straightforward solution to this, as $\ln X \in \mathbb{R}$, for $X > 0$. Unfortunately, if we take logarithms or employ some other nonlinear transformation (e.g. Box-Cox transformations) the singular structure of the data will be lost as $\ln X_1 + \ldots + \ln X_{k-1} \neq \ln X_k$. This means that the imputed data will not satisfy the linear restrictions, which rules out the use of nonlinear transformations.

4.3 Multivariate singular normal distribution

Nonsingular distributions are often defined by specifying the density function of the distribution with respect to Lebesgue measure on $\mathbb{R}^k$. For a singular distribution, the rank of the covariance matrix is: \(\text{rank}(\Sigma) = q < k\), so the inversion of the covariance matrix in the probability density of the normal distribution is not possible and therefore no explicit determination of the density function with respect to Lebesgue measure in $\mathbb{R}^k$ is possible. The density function, however,
does exist on a subspace.

Assume that $X_i \sim N_k(\mu, \Sigma)$, with rank($\Sigma$) = $q$. Let $L$ be the null space (or kernel) of $\Sigma$: Null($\Sigma$) = \{$y \in \mathbb{R}^k : \Sigma y = 0$\}. The dimension of $L$ is $k - q$, which is equal to the number of linear balance restrictions $p$. The subspace $L^\perp$ is the orthogonal complement of $L$, that is the set of all vectors in $\mathbb{R}^k$ that are orthogonal to $L$. Now $X_i$ has the following probability density function (e.g. Khatri, 1968)

$$\varphi(x_i) = (2\pi)^{-k/2} |\Sigma|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(x_i - \mu)'\Sigma^+(x_i - \mu)\right), \tag{4.1}$$

for $x_i \in \mu + L^\perp$, which is an affine subspace. The pseudo-determinant $|\Sigma|_q$ is defined as $\prod_{j=1}^q \lambda_j(\Sigma)$, that is $|\Sigma|_q$ equals the product of the non-zero eigenvalues of $\Sigma$. The matrix $\Sigma^+$ denotes the Moore-Penrose inverse of $\Sigma$. The Moore-Penrose inverse $B^+$ of a matrix $B$ is a unique matrix for which it holds that

1. $BB^+B = B$
2. $B^+BB^+ = B^+$
3. $(BB^+)' = BB^+$
4. $(B^+B)' = B^+B$.

This situation is illustrated graphically in a 3-dimensional setting by Figure 4.1, where $\mu + L^\perp$ represents the affine subspace in which the $X_i$ lie. The circles represent the contour lines of the (bivariate) normal density.

The density function of $X_i$ in (4.1) can be derived as follows (see Khatri, 1968). Let $C$ be the orthogonal matrix of eigenvectors of $\Sigma$. Now partition $C$ as follows

$$C = (C_1 \ C_2),$$

where $C_1$ is the matrix of eigenvectors corresponding to the non-zero eigenvalues of $\Sigma$. This means that $C_2$ is the matrix of eigenvectors corresponding to the zero eigenvalues of $\Sigma$ and therefore $\Sigma C_2 = 0$. Note that the columns of $C_2$ span $L$, the null space of $\Sigma$, and the columns of $C_1$ span its orthogonal complement $L^\perp$. Another matrix that spans the null space is $A'$ as it is of full rank and

$$\Sigma A' = (E[X_iX_i'] - \mu\mu')A'$$
$$= E[X_i(AX_i)'] - \mu(A\mu)'$$
$$= 0,$$

as $AX_i = A\mu = 0$.

The matrix $\Sigma$ can be decomposed by means of an eigenvalue decomposition
Figure 4.1: Plot of the affine subspace in which X lies, the curves represent contour lines of the normal density.

in \( \Sigma = C\Lambda C' \), where \( \Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_q, 0, \ldots, 0\} \) is the matrix of eigenvalues. Therefore it holds that \( \Sigma = C_1\Lambda_1C_1' \), where \( \Lambda_1 = \text{diag}\{\lambda_1, \ldots, \lambda_q\} \) and the Moore-Penrose inverse of \( \Sigma \) is \( \Sigma^+ = C_1\Lambda_1^{-1}C_1' \).

Now consider the following transformation

\[
Y_i^{(1)} = C_i'X_i \\
Y_i^{(2)} = C_2'X_i.
\]

Then

\[
E[Y_i^{(2)}] = C_2'\mu \quad \text{and} \quad \text{Var}(Y_i^{(2)}) = C_2'\Sigma C_2 = 0.
\]

This means that \( Y_i^{(2)} = C_2'\mu \) with probability 1. Note that \( Y_i^{(1)} \) is normally distributed

\[
Y_i^{(1)} \sim \mathcal{N}_q(C_i'\mu, C_i'\Sigma C_i),
\]

where the covariance matrix is nonsingular as \( C_i'\Sigma C_1 = \text{diag}\{\lambda_1, \ldots, \lambda_q\} \).

Therefore \( Y_i^{(1)} \) has the probability density

\[
(2\pi)^{-\frac{q}{2}}|C_i'\Sigma C_i|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(Y_i^{(1)} - C_i'\mu)'(C_i'\Sigma C_i)^{-1}(Y_i^{(1)} - C_i'\mu)\right). \quad (4.2)
\]
So the density of $X_i$ or similarly the density of $(Y_i^{(1)}, Y_i^{(2)})$ is determined by (4.2) and $Y_i^{(2)} = C'_2 \mu$. Note that

\[
(Y_i^{(1)} - C'_1 \mu)'(C'_1 \Sigma C_1)^{-1}(Y_i^{(1)} - C'_1 \mu)
= (C'_1 X_i - C'_1 \mu)'(C'_1 \Sigma C_1)^{-1}(C'_1 X_i - C'_1 \mu)
= (X_i - \mu)'C'_1 (C'_1 \Sigma C_1)^{-1}C'_1 (X_i - \mu)
= (X_i - \mu)'C'_1 \Lambda_1^{-1}C'_1 (X_i - \mu)
= (X_i - \mu)'\Sigma^+(X_i - \mu).
\]

Hence the density in (4.2) becomes

\[
(2\pi)^{-\frac{q}{2}} \left( \prod_{j=1}^{q} \lambda_j \right)^{-\frac{1}{2}} \exp \left( -\frac{1}{2}(x_i - \mu)'\Sigma^+(x_i - \mu) \right), \quad \text{where } x_i \in \mu + L^\perp,
\]
with $C'_2 X_i = C'_2 \mu$.

### 4.4 Maximum likelihood estimation for the singular normal distribution

Khatri (1968) also shows that this singular distribution leads to the same maximum likelihood estimates as the nonsingular model. From nonsingular theory the maximum likelihood estimates of $C'_1 \mu$ and $C'_1 \Sigma C_1$ are

\[
C'_1 \mu = C'_1 \bar{x}
\]
\[
C'_1 \Sigma C_1 = C'_1 S C_1,
\]

where $S = \frac{1}{n} \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})'$. So $\hat{\mu} = \bar{x}$ and $\hat{\Sigma} = S$. We established before that $C'_2 X_i = C'_2 \mu$ with probability 1. This leads to

\[
C'_2 \mu = C'_2 \bar{x} = \frac{1}{n} \sum_{i=1}^{n} C'_2 x_i = C'_2 \mu
\]

and

\[
C'_2 \Sigma C_2 = C'_2 S C_2 = \frac{1}{n} \sum_{i=1}^{n} (C'_2 x_i - C'_2 \bar{x})(C'_2 x_i - C'_2 \bar{x})' = 0 = C'_2 \Sigma C_2.
\]

So the singular case yields the same maximum likelihood estimates as the nonsingular case.
4.4.1 Maximum likelihood estimation and linear balance restrictions

These maximum likelihood estimates concur with the linear balance restrictions on the data. First of all

\[ A\hat{\mu} = A\bar{x} = \frac{1}{n} \sum_{i=1}^{n} Ax_i = 0, \]

as the balance restrictions hold for each respondent.

Furthermore, the null space of the estimated covariance matrix needs to be equal to the null space of \( \Sigma \). Recall that \( \text{Null}(\Sigma) \) is spanned by \( A' \), as was established in section 4.3. So postmultiply \( \Sigma \) with \( A' \)

\[ \Sigma A' = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})' A' = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(Ax_i - A\bar{x})' = 0. \]

This means that \( \text{Null}(\Sigma) = \text{Null}(\Sigma) \). The maximum likelihood estimates are therefore in concurrence with the linear balance restrictions.

4.5 The EM algorithm applied to singular normal data

Now suppose we encounter missing data. The EM algorithm can be used to find maximum likelihood estimates in the presence of nonresponse. Again suppose that \( X_{ij}, i = 1, \ldots, n, \) of order \( k \times 1 \) contains the item responses for record \( i \), and the complete \( n \times k \) data matrix is \( X \), with elements \( X_{ij}, i = 1, \ldots, n, j = 1, \ldots, k \). For each record \( i \) partition the items into a missing and an observed part, thus \( X_i = (X_{i,mis}, X_{i,obs}) \) and partition \( \mu \) and \( \Sigma \) accordingly. Let \( m_i \) denote the number of missing items for respondent \( i \).

It is well known that, if \( X_{i} \) is normally distributed with mean vector \( \mu \) and covariance matrix \( \Sigma \), the following holds (see e.g. Anderson, 1984, for a proof)

\[ X_{i,mis} | X_{i,obs} = x_{i,obs} \sim N_{m_i}(\mu_{mis,obs}, \Sigma_{mis,mis,obs}), \]

where

\[ \mu_{mis,obs} = \mu_{mis} + \Sigma_{mis,obs} \Sigma_{obs,obs}^{-} (x_{i,obs} - \mu_{obs}) \]

\[ \Sigma_{mis,mis,obs} = \Sigma_{mis,mis} - \Sigma_{mis,obs} \Sigma_{obs,obs}^{-} \Sigma_{obs,mis}. \]
Note that the matrix $\Sigma_{\text{obs,obs}}$, as opposed to the Moore-Penrose inverse, is not unique. Pringle and Rayner (1971) establish the invariance of $\mu_{\text{mis,obs}}$ and $\Sigma_{\text{mis,mis,obs}}$ under any choice of $\Sigma_{\text{obs,obs}}^{-1}$.

The matrix $\Sigma_{\text{obs,obs}}$ is singular if all items in at least one linear balance restriction are observed. If $\Sigma_{\text{obs,obs}}$ is nonsingular the generalised inverse equals the regular inverse. Note that the dependence of the partitioned mean vector and the partitioned covariance matrix on $i$ is left out, this is done for ease of notation. It is, however, important to keep in mind that the missingness patterns and consequently $\mu_{\text{mis,obs}}$ and $\Sigma_{\text{mis,mis,obs}}$ vary across respondents. Also note that $\Sigma_{\text{mis,mis,obs}}$ will always be singular as all variables are incorporated in at least one linear restriction, so one item value can always be derived from the other items. A formal proof will be given in subsection 4.7.1.

For the estimation of $\mu$ and $\Sigma$ we will use the EM algorithm, which calculates the maximum likelihood estimates in the presence of missing data. For a detailed description of the EM algorithm in general see chapter 2 of this thesis. See Schafer (1997), and Little and Rubin (2002) for a description of the EM algorithm for normally distributed data.

In short, the EM algorithm for data that are distributed according to a singular normal is as follows. Since the singular normal distribution is a member of an exponential family, the E-step consists of estimating the expected sufficient statistics

$$E[T_1(X)] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$$

$$E[T_2(X)] = E[X'X] = \frac{1}{n} \sum_{i=1}^{n} E[X_iX_i'].$$ 

Recall that

$$E[X_{i,mis} | X_{i,obs} = x_{i,obs}, \mu^{(t)}, \Sigma^{(t)}] = \mu^{(t)} + \Sigma^{(t)}_{\text{mis,obs}}(\Sigma^{(t)}_{\text{obs,obs}})^{-1}(x_{i,obs} - \mu^{(t)})$$  \hspace{1cm} (4.3)$$

and

$$\text{Var}(X_{i,mis} | X_{i,obs} = x_{i,obs}, \mu^{(t)}, \Sigma^{(t)}) = \Sigma^{(t)}_{\text{mis,mis}} - \Sigma^{(t)}_{\text{mis,obs}}(\Sigma^{(t)}_{\text{obs,obs}})^{-1}\Sigma^{(t)}_{\text{obs,mis}}.$$ \hspace{1cm} (4.4)
Let $X_{i}^{*}$ denote the elements of $X_{i,mis}$ estimated by equation (4.3). Then for $i = 1, \ldots, n$, and $j, l = 1, \ldots, k$,

$$
\hat{E}[X_{ij} \mid X_{i,obs} = x_{i,obs}, \mu^{(t)}, \Sigma^{(t)}] = \begin{cases} 
    x_{ij} & \text{if } X_{ij} \text{ is observed} \\
    X_{ij}^{*} & \text{if } X_{ij} \text{ is missing}
\end{cases}
$$

and

$$
\hat{E}[X_{ij}X_{il} \mid X_{i,obs} = x_{i,obs}, \mu^{(t)}, \Sigma^{(t)}] = \begin{cases} 
    x_{ij}x_{il} & \text{if } X_{ij} \text{ and } X_{il} \text{ are both observed} \\
    X_{ij}^{*}x_{il} & \text{if } X_{ij} \text{ is missing and } X_{il} \text{ is observed} \\
    x_{ij}X_{il}^{*} & \text{if } X_{ij} \text{ is observed and } X_{il} \text{ is missing} \\
    X_{ij}^{*}X_{il}^{*} + \text{covariances} & \text{if } X_{ij}^{*} \text{ and } X_{il}^{*} \text{ are both missing}
\end{cases}
$$

where covariances $= \text{Cov}(X_{ij}^{*}, X_{il}^{*} \mid X_{i,obs} = x_{i,obs}, \mu^{(t)}, \Sigma^{(t)})$, which are the elements of the $m_{j} \times m_{l}$ covariance matrix of the missing items estimated by (4.4). The covariances $\text{Cov}(x_{ij}, x_{il} \mid x_{i,obs}, \mu^{(t)}, \Sigma^{(t)})$, $\text{Cov}(X_{ij}^{*}, x_{il} \mid x_{i,obs}, \mu^{(t)}, \Sigma^{(t)})$ and $\text{Cov}(x_{ij}, X_{il}^{*} \mid x_{i,obs}, \mu^{(t)}, \Sigma^{(t)})$ all are equal to zero since at least one of the $X$’s is observed and thus regarded as fixed. Now create a $k \times k$ matrix $V^{(t)}(X_{i,mis} \mid X_{i,obs} = x_{i,obs}, \mu^{(t)}, \Sigma^{(t)})$ with the following elements for $j, l = 1, \ldots, k$

$$
V^{(t)}(X_{i,mis} \mid X_{i,obs} = x_{i,obs}, \mu^{(t)}, \Sigma^{(t)})_{jl} = \begin{cases} 
    \text{Cov}(X_{ij}^{*}, X_{il}^{*} \mid x_{i,obs}, \mu^{(t)}, \Sigma^{(t)}) & \text{if } X_{ij}^{*} \text{ and } X_{il}^{*} \text{ are both missing} \\
    0 & \text{otherwise}
\end{cases}
$$

Now $X_{i}' = (X_{i,mis}', X_{i,obs}')$ is replaced by its expectation: $X_{i}' = (E[X_{i,mis} \mid X_{i,obs} = x_{i,obs}, \mu^{(t)}, \Sigma^{(t)}], x_{i,obs}')$ and then $\mu$ and $\Sigma$ are re-estimated by

$$
\mu^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} X_{i}
$$

$$
\Sigma^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} (X_{i} - \mu^{(t+1)})(X_{i} - \mu^{(t+1)})' + 
\frac{1}{n} \sum_{i=1}^{n} V^{(t)}(X_{i,mis} \mid X_{i,obs} = x_{i,obs}, \mu^{(t)}, \Sigma^{(t)}) 
= \frac{1}{n} \sum_{i=1}^{n} \left( X_{i}X_{i}' + V^{(t)}(X_{i,mis} \mid X_{i,obs} = x_{i,obs}, \mu^{(t)}, \Sigma^{(t)}) \right) + 
- \mu^{(t+1)}(\mu^{(t+1)})'.
$$
Next the updated parameter estimates are used to re-estimate equations (4.3) and (4.4), which in their turn are used to update the parameter estimates. This process is iterated until convergence. That is, until the change in parameter estimates is sufficiently small.

4.6 EM estimates and linear balance restrictions

It is crucial that the estimates generated by the EM algorithm concur with the linear balance restrictions on the data, in order to obtain imputations that satisfy these restrictions. This means that $A\mu^{(t)}$ should equal zero and $\text{Null}(\Sigma^{(t)})$ should equal $\text{Null}(\Sigma)$.

For $A\mu^{(t)}$ to be equal to zero it is sufficient if it holds that

$$A\mu^{(t)} = A \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{n} \sum_{i=1}^{n} AX_i = 0,$$

for $X_i$ at iteration $t$. We know that this holds for the completely observed data and need to show that this also holds for the data with missing items. Again partition $X_i$ into a missing and an observed part and partition $A$ accordingly. This means we need to show for the estimated data that

$$A_{i,\text{mis}}\hat{X}_{i,\text{mis}} = -A_{i,\text{obs}}X_{i,\text{obs}}$$

holds at every iteration $t$. Let $R$ denote the set of respondents without any item nonresponse and let $r$ denote the number of respondents without any item nonresponse.

In the EM algorithm the missing items for respondent $i$ will be replaced by their expected values:

$$E[X_{i,\text{mis}} \mid X_{i,\text{obs}}, \mu^{(t)}, \Sigma^{(t)}] = \mu^{(t)}_{\text{mis}} + \Sigma^{(t)}_{\text{mis,obs}}(\Sigma^{(t)}_{\text{obs,obs}})^{-1}(X_{i,\text{obs}} - \mu^{(t)}_{\text{obs}}).$$

First iteration

We will start at $t = 0$. Assuming there are sufficient complete cases, the complete cases mean and covariance matrix will be used as starting values, which means that $A\mu^{(0)} = 0$ and $\text{Null}(\Sigma^{(0)}) = \text{Null}(\Sigma)$ as $\Sigma^{(0)}A' = 0$ which we found in subsection 4.4.1.
4.6. EM estimates and linear balance restrictions

**Expectation step**
The updated $\mu^{(1)}$ will satisfy the linear balance restrictions if the imputed items in this iteration satisfy the balance restrictions, so it has to hold that

$$A_{i, \text{mis}} E[X_{i, \text{mis}} | X_{i, \text{obs}}, \mu^{(0)}, \Sigma^{(0)}] = -A_{i, \text{obs}} X_{i, \text{obs}}.$$ 

Now obtain

$$A_{i, \text{mis}} E[X_{i, \text{mis}} | X_{i, \text{obs}}, \mu^{(0)}, \Sigma^{(0)}] = A_{i, \text{mis}} \mu_{\text{mis}}^{(0)} + A_{i, \text{mis}} \Sigma^{(0)}_{\text{mis, obs}} (\Sigma^{(0)}_{\text{obs, obs}})^{-1} (X_{i, \text{obs}} - \mu^{(0)}_{\text{obs}}),$$

where

$$A_{i, \text{mis}} \Sigma^{(0)}_{\text{mis, obs}} = \frac{1}{p} \sum_{h \in R} (X_{h, \text{mis}} - \mu^{(0)}_{\text{mis}})(X_{h, \text{obs}} - \mu^{(0)}_{\text{obs}})'$$

$$= \frac{1}{p} \sum_{h \in R} (A_{i, \text{mis}} X_{h, \text{mis}} - A_{i, \text{mis}} \mu^{(0)}_{\text{mis}})(X_{h, \text{obs}} - \mu^{(0)}_{\text{obs}})'$$

$$= \frac{1}{p} \sum_{h \in R} (-A_{i, \text{obs}} X_{h, \text{obs}} + A_{i, \text{obs}} \mu^{(0)}_{\text{obs}})(X_{h, \text{obs}} - \mu^{(0)}_{\text{obs}})'$$

$$= -A_{i, \text{obs}} \frac{1}{p} \sum_{h \in R} (X_{h, \text{obs}} - \mu^{(0)}_{\text{obs}})(X_{h, \text{obs}} - \mu^{(0)}_{\text{obs}})'$$

$$= -A_{i, \text{obs}} \Sigma^{(0)}_{\text{obs, obs}}.$$ 

So

$$A_{i, \text{mis}} E[X_{i, \text{mis}} | X_{i, \text{obs}}, \mu^{(0)}, \Sigma^{(0)}] = A_{i, \text{mis}} \mu^{(0)}_{\text{mis}} - A_{i, \text{obs}} \Sigma^{(0)}_{\text{obs, obs}} (\Sigma^{(0)}_{\text{obs, obs}})^{-1} (X_{i, \text{obs}} - \mu^{(0)}_{\text{obs}}). \quad (4.5)$$

If $\Sigma^{(0)}_{\text{obs, obs}}$ is nonsingular, the generalised inverse becomes the regular inverse and it immediately follows that $A_{i, \text{mis}} E[X_{i, \text{mis}} | X_{i, \text{obs}}, \mu^{(1)}, \Sigma^{(1)}] = -A_{i, \text{obs}} X_{i, \text{obs}}$.

However, if $\Sigma^{(0)}_{\text{obs, obs}}$ is singular the proof is less straightforward. Decompose $\Sigma^{(0)}_{\text{obs, obs}}$ by means of an eigenvalue decomposition

$$\Sigma^{(0)}_{\text{obs, obs}} = CAC' = C \begin{pmatrix} A_1 & 0 \\ 0 & 0 \end{pmatrix} C',$$
where \( \mathbf{C} \) is the orthogonal matrix of eigenvectors, and \( \mathbf{A}_1 \) a diagonal matrix with the nonzero eigenvalues of \( \Sigma^{(0)}_{\text{obs,obs}} \) on the diagonal. Since \( \Sigma^{(0)}_{\text{obs,obs}} \) is symmetric, its generalised inverse is of the form (Pringle and Rayner, 1971)

\[
(\Sigma^{(0)}_{\text{obs,obs}})^{-} = \mathbf{C} \left( \begin{array}{ccc} A_1^{-1} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{array} \right) \mathbf{C}',
\]

where \( \mathbf{U}, \mathbf{V} \) and \( \mathbf{W} \) are arbitrary matrices. Partition the matrix \( \mathbf{C} \) into \( \mathbf{C} = (\mathbf{C}_1, \mathbf{C}_2) \), where \( \mathbf{C}_1 \) is the matrix of eigenvectors corresponding to the nonzero eigenvalues of \( \Sigma^{(0)}_{\text{obs,obs}} \). This means that \( \Sigma^{(0)}_{\text{obs,obs}} \mathbf{C}_2 = 0 \), as \( \mathbf{C}_2 \) corresponds to the zero eigenvalues of \( \Sigma^{(0)}_{\text{obs,obs}} \).

Now recall that the marginal distribution of the observed data for respondent \( i \) in this iteration is

\[
\mathbf{X}_{i,\text{obs}} \sim \mathcal{N}(\mu^{(0)}_{\text{obs}}, \Sigma^{(0)}_{\text{obs,obs}}).
\]

Transforming \( \mathbf{X}_{i,\text{obs}} \) leads to

\[
\mathbf{C}'_2(\mathbf{X}_{i,\text{obs}} - \mu^{(0)}_{\text{obs}}) \sim \mathcal{N}(0, \mathbf{C}'_2 \Sigma^{(0)}_{\text{obs,obs}} \mathbf{C}_2),
\]

where \( \mathbf{C}'_2 \mathbf{C}_2 = 0 \), since \( \Sigma^{(0)}_{\text{obs,obs}} \mathbf{C}_2 = 0 \). So \( \mathbf{C}'(\mathbf{X}_{i,\text{obs}} - \mu^{(0)}_{\text{obs}}) = 0 \) with probability one.

The right-hand side of equation (4.5) now becomes

\[
\mathbf{A}_{i,\text{mis}}^{(0)} \mu^{(0)}_{\text{mis}} - \mathbf{A}_{i,\text{obs}} \mathbf{C} \left( \begin{array}{ccc} A_1 & 0 \\ 0 & 0 \end{array} \right) \mathbf{C}' \left( \begin{array}{ccc} A_1^{-1} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{array} \right) \mathbf{C}'(\mathbf{X}_{i,\text{obs}} - \mu^{(0)}_{\text{obs}}),
\]

which reduces to

\[
\mathbf{A}_{i,\text{mis}}^{(0)} \mu^{(0)}_{\text{mis}} - \mathbf{A}_{i,\text{obs}} \mathbf{C} \left( \begin{array}{ccc} \mathbf{I}_{k-m_i-p_i} & A_1 \mathbf{U} \\ 0 & 0 \end{array} \right) \mathbf{C}'(\mathbf{X}_{i,\text{obs}} - \mu^{(0)}_{\text{obs}}),
\]

since \( \mathbf{C}' \mathbf{C} = \mathbf{I}_{k-m_i} \), where \( p_i \) denotes the number of nonredundant balance restrictions on the missing items for respondent \( i \). This can then be written as

\[
\mathbf{A}_{i,\text{mis}}^{(0)} \mu^{(0)}_{\text{mis}} - \mathbf{A}_{i,\text{obs}} (\mathbf{C}_1 \mathbf{C}_1' + \mathbf{C}_1 A_1 \mathbf{U} \mathbf{C}_2') (\mathbf{X}_{i,\text{obs}} - \mu^{(0)}_{\text{obs}}).
\]

We already established that \( \mathbf{C}'_2(\mathbf{X}_{i,\text{obs}} - \mu^{(0)}_{\text{obs}}) = 0 \) with probability one, so this equation results in

\[
\mathbf{A}_{i,\text{mis}}^{(0)} \mu^{(0)}_{\text{mis}} - \mathbf{A}_{i,\text{obs}} \mathbf{C}_1 \mathbf{C}_1' (\mathbf{X}_{i,\text{obs}} - \mu^{(0)}_{\text{obs}}). \tag{4.6}
\]
Finally, note that the following holds
\[ I_{k-m_i} = CC' = (C_1 C_2) \begin{pmatrix} C_1' \\ C_2' \end{pmatrix} = C_1 C_1' + C_2 C_2'. \]

Post-multiplying this equation with \( X_{i,obs} - \mu_{obs}^{(0)} \) leads to
\[
X_{i,obs} - \mu_{obs}^{(0)} = (C_1 C_1' + C_2 C_2')(X_{i,obs} - \mu_{obs}^{(0)})
= C_1 C_1'(X_{i,obs} - \mu_{obs}^{(0)}).
\]

Now (4.6) becomes
\[
A_{i,mis} \mu_{mis}^{(0)} - A_{i,obs} X_{i,obs} + A_{i,obs} \mu_{obs}^{(0)} = -A_{i,obs} X_{i,obs}.
\]

This means that \( A_{i,mis} E[X_{i,mis} | X_{i,obs}, \mu^{(0)}, \Sigma^{(0)}] = -A_{i,obs} X_{i,obs} \) for both singular and nonsingular \( \Sigma_{obs,obs} \).

Note that the above line of reasoning not only holds for a singular \( \Sigma_{obs,obs} \) at iteration \( t = 0 \), but for a singular \( \Sigma_{obs,obs} \) at any iteration \( t \).

**Maximisation step**

Now proceed to the M-step. The mean, \( \mu^{(1)} \) is calculated based on the observed and estimated data in the expectation step and as for each respondent in the imputed dataset it holds that \( A_{i,mis} E[X_{i,mis} | X_{i,obs}, \mu^{(0)}, \Sigma^{(0)}] = -A_{i,obs} X_{i,obs} \), it will also hold that \( A_{i,mis} \mu_{mis}^{(1)} = -A_{i,obs} \mu_{obs}^{(1)} \). Next \( \Sigma \) will be re-estimated using an updated \( X_i = (E[X_{i,mis} | X_{i,obs}, \mu^{(0)}, \Sigma^{(0)}], X_{i,obs})' \):
\[
\Sigma^{(1)} = \frac{1}{n} \sum_{i=1}^{n} \left( (X_i - \mu^{(1)})(X_i - \mu^{(1)})' + V^{(0)}(X_{i,mis} | X_{i,obs}, \mu^{(0)}, \Sigma^{(0)}) \right).
\]

For each record \( i \) it holds that \( AV^{(0)}(X_{i,mis} | X_{i,obs}, \mu^{(0)}, \Sigma^{(0)}) = 0 \). This can be established as follows. Again partition \( A \) into a missing and an observed part and partition the matrix \( V^{(0)}(X_{i,mis} | X_{i,obs}, \mu^{(0)}, \Sigma^{(0)}) \) accordingly. Now for any record \( i \)
\[
AV^{(0)}(X_{i,mis} | X_{i,obs}, \mu^{(0)}, \Sigma^{(0)})
= (A_{i,mis} A_{i,obs}) \begin{pmatrix} \text{Var}(X_{i,mis} | X_{i,obs}, \mu^{(0)}, \Sigma^{(0)}) & 0 \\ 0 & 0 \end{pmatrix}
= (A_{i,mis} \text{Var}(X_{i,mis} | X_{i,obs}, \mu^{(0)}, \Sigma^{(0)}) 0),
\]
where
\[
\text{Var}(X_{i,\text{mis}} \mid X_{i,\text{obs}}, \mu^{(0)}, \Sigma^{(0)}) = \Sigma^{(0)}_{\text{mis, mis}} - \Sigma^{(0)}_{\text{mis, obs}} (\Sigma^{(0)}_{\text{obs, obs}})^{-1} \Sigma^{(0)}_{\text{obs, mis}}.
\]

Now
\[
\text{A}_{i,\text{mis}} \text{Var}(X_{i,\text{mis}} \mid X_{i,\text{obs}}, \mu^{(0)}, \Sigma^{(0)})
= \text{A}_{i,\text{mis}} \Sigma^{(0)}_{\text{mis, mis}} - \text{A}_{i,\text{mis}} \Sigma^{(0)}_{\text{mis, obs}} (\Sigma^{(0)}_{\text{obs, obs}})^{-1} \Sigma^{(0)}_{\text{obs, mis}}
= \text{A}_{i,\text{mis}} \Sigma^{(0)}_{\text{mis, mis}} + \text{A}_{i,\text{obs}} \Sigma^{(0)}_{\text{obs, obs}} (\Sigma^{(0)}_{\text{obs, obs}})^{-1} \Sigma^{(0)}_{\text{obs, mis}}.
\]

If \(\Sigma^{(0)}_{\text{obs, obs}}\) is nonsingular this leads to
\[
\text{A}_{i,\text{mis}} \text{Var}(X_{i,\text{mis}} \mid X_{i,\text{obs}}, \mu^{(0)}, \Sigma^{(0)}) = \text{A}_{i,\text{mis}} \Sigma^{(0)}_{\text{mis, mis}} + \text{A}_{i,\text{obs}} \Sigma^{(0)}_{\text{obs, mis}}
= \text{A}_{i,\text{mis}} \Sigma^{(0)}_{\text{mis, mis}} - \text{A}_{i,\text{mis}} \Sigma^{(0)}_{\text{mis, mis}}
= 0.
\]

However, if \(\Sigma^{(0)}_{\text{obs, obs}}\) is singular, equation (4.7) becomes
\[
\text{A}_{i,\text{mis}} \text{Var}(X_{i,\text{mis}} \mid X_{i,\text{obs}}, \mu^{(0)}, \Sigma^{(0)})
= \text{A}_{i,\text{mis}} \Sigma^{(0)}_{\text{mis, mis}} + \text{A}_{i,\text{obs}} (C_1 C_1' + C_1 A_1 U C_2') \Sigma^{(0)}_{\text{obs, mis}}.
\]

Since
\[
\Sigma^{(0)}_{\text{obs, mis}} = \frac{1}{p} \sum_{h \in \mathcal{H}} (X_{h,\text{obs}} - \mu^{(0)}_{\text{obs}})(X_{h,\text{mis}} - \mu^{(0)}_{\text{mis}})',
\]
and
\[
C_2'(X_{i,\text{obs}} - \mu^{(0)}_{\text{obs}}) = 0 \quad \text{and} \quad C_1 C_1'(X_{i,\text{obs}} - \mu^{(0)}_{\text{obs}}) = (X_{i,\text{obs}} - \mu^{(0)}_{\text{obs}}),
\]
equation (4.8) reduces to
\[
\text{A}_{i,\text{mis}} \text{Var}(X_{i,\text{mis}} \mid X_{i,\text{obs}}, \mu^{(0)}, \Sigma^{(0)}) = \text{A}_{i,\text{mis}} \Sigma^{(0)}_{\text{mis, mis}} + \text{A}_{i,\text{obs}} \Sigma^{(0)}_{\text{obs, mis}}
= 0,
\]
and consequently \(\text{A} \Sigma^{(0)}(X_{i,\text{mis}}) = 0\).

The null space of \(\Sigma^{(1)}\) can now be established by postmultiplying \(\Sigma^{(1)}\) with
4.6. EM estimates and linear balance restrictions

\[ \mathbf{A} \]

\[
\Sigma^{(1)} \mathbf{A}' = \frac{1}{n} \sum_{i=1}^{n} \left( (\mathbf{X}_i - \mu^{(1)})(\mathbf{X}_i - \mu^{(1)})' + \mathbf{V}^{(0)}(\mathbf{X}_{i,\text{mis}}) \right) \mathbf{A}'
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \left( (\mathbf{X}_i - \mu^{(1)})(\mathbf{A}\mathbf{X}_i - \mathbf{A}\mu^{(1)})' + \mathbf{V}^{(0)}(\mathbf{X}_{i,\text{mis}}) \mathbf{A}' \right)
\]

\[ = 0. \]

This means that the null space of \( \Sigma^{(1)} \) is equal to the null space of \( \Sigma \). We have now established that the parameter estimates obtained in the first iteration of the EM algorithm concur with the linear balance restrictions.

**Subsequent iterations**

Next consider the case where we are at \( t = 1 \). The parameter estimates now depend not only on the observed data but on the imputed data as well.

**Expectation step**

Impute for nonrespondent \( i \) once again, but now based on the new estimates for \( \mu \) and \( \Sigma \):

\[
\mathbb{E}[\mathbf{X}_{i,\text{mis}} | \mathbf{X}_{i,\text{obs}}, \mu^{(1)}, \Sigma^{(1)}] = \mu^{(1)} + \Sigma^{(1)} \Sigma_{\text{mis},\text{obs}} (\Sigma_{\text{obs},\text{obs}})^{-1} (\mathbf{X}_{i,\text{obs}} - \mu^{(1)}).
\]

Then

\[
\mathbf{A}_{i,\text{mis}} \mathbb{E}[\mathbf{X}_{i,\text{mis}} | \mathbf{X}_{i,\text{obs}}, \mu^{(1)}, \Sigma^{(1)}] =
\]

\[
= \mathbf{A}_{i,\text{mis}} \mu^{(1)} + \mathbf{A}_{i,\text{mis}} \Sigma_{\text{mis},\text{obs}} (\Sigma_{\text{obs},\text{obs}})^{-1} (\mathbf{X}_{i,\text{obs}} - \mu^{(1)}).
\]

From \( \mathbf{A} \Sigma^{(1)} = 0 \), it follows that \( \mathbf{A}_{i,\text{mis}} \Sigma_{\text{mis},\text{obs}} = -\mathbf{A}_{i,\text{obs}} \Sigma_{\text{obs},\text{obs}} \). So

\[
\mathbf{A}_{i,\text{mis}} \mathbb{E}[\mathbf{X}_{i,\text{mis}} | \mathbf{X}_{i,\text{obs}}, \mu^{(1)}, \Sigma^{(1)}] =
\]

\[
= \mathbf{A}_{i,\text{mis}} \mu^{(1)} - \mathbf{A}_{i,\text{obs}} \Sigma_{\text{obs},\text{obs}} (\Sigma_{\text{obs},\text{obs}})^{-1} (\mathbf{X}_{i,\text{obs}} - \mu^{(1)}).
\]

This immediately leads to \( \mathbf{A}_{i,\text{mis}} \mathbb{E}[\mathbf{X}_{i,\text{mis}} | \mathbf{X}_{i,\text{obs}}, \mu^{(1)}, \Sigma^{(1)}] = -\mathbf{A}_{i,\text{obs}} \mathbf{X}_{i,\text{obs}} \) if \( \Sigma_{\text{obs},\text{obs}} \) is nonsingular. If \( \Sigma_{\text{obs},\text{obs}} \) is singular we can use the same line of reasoning as in the first iteration to find that

\[
\Sigma_{\text{obs},\text{obs}} (\Sigma_{\text{obs},\text{obs}})^{-1} (\mathbf{X}_{i,\text{obs}} - \mu^{(1)}) = \mathbf{X}_{i,\text{obs}} - \mu^{(1)}
\]
and thus: \( A_{i,mis}E[X_{i,mis} \mid X_{i,obs}, \mu^{(1)}, \Sigma^{(1)}] = -A_{i,obs}X_{i,obs} \).

**Maximisation step**

The M-step now estimates \( \mu^{(2)} \). For both the estimated and the observed data at this iteration, we found that \( A_{i,mis}E[X_{i,mis} \mid X_{i,obs}, \mu^{(1)}, \Sigma^{(1)}] = -A_{i,obs}X_{i,obs} \), so it will also hold that \( A_{i,mis} \mu^{(2)}_{mis} = -A_{i,obs} \mu^{(2)}_{obs} \).

Next the covariance matrix will be re-estimated by

\[
\Sigma^{(2)} = \frac{1}{n} \sum_{t=1}^{n} \left( (X_t - \mu^{(2)})(X_t - \mu^{(2)})' + V^{(1)}(X_{t,mis}) \right).
\]

Now

\[
AV^{(1)}(X_{i,mis}) = (A_{i,mis} \text{Var}(X_{i,mis} \mid X_{i,obs}, \mu^{(1)}, \Sigma^{(1)}) 0),
\]

where

\[
A_{i,mis} \text{Var}(X_{i,mis} \mid X_{i,obs}, \mu^{(1)}, \Sigma^{(1)}) = A_{i,mis} \Sigma^{(1)}_{mis,mis} +
A_{i,mis} \Sigma^{(1)}_{mis,obs}(\Sigma^{(1)}_{obs,obs} - \Sigma^{(1)}_{obs,mis})^{-1} \Sigma^{(1)}_{obs,mis} +
A_{i,obs} \Sigma^{(1)}_{obs,obs}(\Sigma^{(1)}_{obs,obs})^{-1} \Sigma^{(1)}_{obs,mis}.
\]

Once again the same line of reasoning as in the first iteration can be used to establish that \( \Sigma^{(1)}_{obs,obs}(\Sigma^{(1)}_{obs,obs})^{-1} \Sigma^{(1)}_{obs,mis} = \Sigma^{(1)}_{obs,mis} \) for a singular \( \Sigma^{(1)}_{obs,obs} \) and therefore that \( AV^{(1)}(X_{i,mis}) = 0 \). So \( \Sigma^{(2)}A' = 0 \), which means that \( \text{Null}(\Sigma^{(2)}) = \text{Null}(\Sigma) \).

This can be straightforwardly extended to the subsequent iterations \( t = 2, 3, \ldots \). So at each iteration in the EM algorithm the maximum likelihood estimates concur with the linear balance restrictions, that is \( A\mu^{(t)} = 0 \) and the null space of \( \Sigma^{(t)} \) is equal to the null space of \( \Sigma \) and this will result in imputed values that satisfy the balance restrictions.

### 4.6.1 Starting values

The EM algorithm requires a starting value for \( \mu \) and \( \Sigma \). The mean vector and covariance matrix can, for example, be calculated from the completely observed data or by using the available cases for each variable. Using the complete cases provides consistent estimates of the parameters if the data are MCAR and if there are at least \( k + 1 \) observations. In general the choice of the starting value
is not crucial, unless the fraction of missing data is very high (Schafer, 1997).

In our case, however, the choice of the starting value is crucial in the sense that if the starting values do not correspond with the balance restrictions, i.e. \( A_\mu^{(0)} \neq 0 \) or \( \text{Null}(\Sigma^{(0)}) \neq \text{Null}(\Sigma) \), the final estimates of the parameters will not either. This rules out the use of available cases estimates for data subject to linear balance restrictions.

### 4.7 Imputation

#### 4.7.1 The singularity of \( \Sigma^{(t)}_{mis,mis,obs} \)

Previously we mentioned that the covariance matrix \( \Sigma^{(t)}_{mis,mis,obs} \), that is used to generate imputations, is always singular if all variables are present in at least one balance restriction. We will present a formal proof here. Recall that for each respondent \( i \) at iteration \( t \) it holds that

\[
\Sigma^{(t)}_{mis,mis,obs} = \Sigma^{(t)}_{mis,mis} - \Sigma^{(t)}_{mis,obs} (\Sigma^{(t)}_{obs,obs})^{-1} \Sigma^{(t)}_{obs,mis}
\]

We leave out the dependence on \( i \) for ease of notation. Premultiply \( \Sigma^{(t)}_{mis,mis,obs} \) with \( A_{i,mis}' \), the restriction matrix on the missing items of respondent \( i \)

\[
A_{i,mis} \Sigma^{(t)}_{mis,mis,obs} = A_{i,mis} \Sigma^{(t)}_{mis,mis} - A_{i,mis} \Sigma^{(t)}_{mis,obs} (\Sigma^{(t)}_{obs,obs})^{-1} \Sigma^{(t)}_{obs,mis}
\]

since \( A \Sigma^{(t)} = 0 \), which was shown in section 4.6.

If \( \Sigma^{(t)}_{obs,obs} \) is nonsingular, it immediately follows that \( A_{i,mis} \Sigma^{(t)}_{mis,mis,obs} \) equals zero. We will use the fact that \( \Sigma^{(t)}_{obs,obs} (\Sigma^{(t)}_{obs,obs})^{-1} \Sigma^{(t)}_{obs,mis} = \Sigma^{(t)}_{obs,mis} \) (see section 4.6), if \( \Sigma^{(t)}_{obs,obs} \) is singular. Hence \( \Sigma^{(t)}_{mis,mis,obs} A'_{i,mis} = 0 \) for both singular and nonsingular \( \Sigma^{(t)}_{obs,obs} \). This means that the columns of \( A'_{i,mis} \) are in the null space of \( \Sigma^{(t)}_{mis,mis,obs} \). Since every variable is present in at least one linear balance restriction and since there are no redundant balance restrictions, the restriction matrix \( A \) is of full rank. Therefore it holds that \( \text{rank}(A_{i,mis}) \geq 1 \), from which it follows that \( \Sigma_{mis,mis,obs} \) is always singular with rank \( m_i - \text{rank}(A_{i,mis}) \).

#### 4.7.2 Imputation of missing data items

The missing data items can now be imputed using either deterministic or stochastic imputation. Deterministic imputation can be done by using the expected
missing data items that were calculated in the E-step for imputation. Using expectations, however, may lead to an underestimation of the true variance and may attenuate relationships between variables. If one is mostly interested in the estimation of means and totals this method is expected to work well. This means that the imputed values are

\[
\hat{X}_{i,\text{imp}} = E[X_{i,\text{mis}} \mid X_{i,\text{obs}} = x_{i,\text{obs}}, \mu^{(t)}, \Sigma^{(t)}],
\]

which, as we established earlier, will satisfy the balance restrictions.

If the aim of the imputation procedure is to obtain a general purpose dataset, it would be better to use a stochastic imputation method. For example by using draws from the singular normal, with parameter estimates from the EM algorithm, as imputations. In this case the true variances of the variables are better represented in the imputed dataset. The missing data items have the following distribution

\[
X_{i,\text{mis}} \mid X_{i,\text{obs}} = x_{i,\text{obs}} \sim \mathcal{N}_{m_i}(\mu_{\text{mis,obs}}, \Sigma_{\text{mis,obs}}).
\]

The \( m_i \times m_j \) matrix \( \Sigma_{\text{mis,obs}} \) is singular (see subsection 4.7.1), which means that we can decompose it by means of an eigenvalue decomposition into \( \mathbf{C}\Lambda\mathbf{C}' \), where \( \mathbf{C} \) is an orthogonal matrix of eigenvectors and \( \Lambda \) is a diagonal matrix with the eigenvalues of \( \Sigma \) on the main diagonal. So \( \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_{m_i-p_i}, 0, \ldots, 0) \), assuming that \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{m_i-p_i} > 0 \), where \( \lambda_i \) is the \( i \)-th eigenvalue, again \( m_i \) denotes the number of missing items and \( p_i \) denotes the number of nonredundant balance restrictions on the missing items for respondent \( i \). Then \( X_{i,\text{mis}} \) can be transformed as follows

\[
\mathbf{C}'(X_{i,\text{mis}} - \mu_{\text{mis,obs}}) \mid X_{i,\text{obs}} = x_{i,\text{obs}} \sim \mathcal{N}_{m_i}(0, \Lambda).
\]

Generate \( Z_j \sim \mathcal{N}(0, 1) \) for \( j = 1, \ldots, m_i-p_i \) and take \( Z_j \) equal to zero otherwise.

Then calculate \( \tilde{Z} = \Lambda^{1/2}Z \) so that \( \tilde{Z} \sim \mathcal{N}_{m_i}(0, \Lambda) \). Finally, impute the missing data items by

\[
\hat{X}_{i,\text{imp}} = \mathbf{C}\tilde{Z} + \mu_{\text{mis,obs}}.
\]

### 4.8 Imputation performance

Similarly to chapter 3 we will use this method to impute data that have been gathered by Statistics Netherlands on a part of the wholesale industry for businesses with more than 10 employees, in order to assess the performance of this method on empirical data. The effects on estimation of population parameters will be analysed.
4.8. Imputation performance

![Graphs showing 95% confidence intervals for the parameter estimates of \( \mu \).]

4.8.1 Estimation of population parameters

The effects of imputation will be investigated with respect to estimation of the population parameters \( \mu \) and \( \sigma \). As we mentioned in chapter 3, estimates of population parameters are potentially subject to sampling variance and nonresponse variance. Nonresponse variance is introduced by the nonresponse in the sample and is currently our main interest. First of all, this method is incorporated in the simulation study that was executed in chapter 3 for the datasets on labour costs and costs of third party rendering of services, which are both subject to one balance restriction. Next this method will be used for data that are subject to multiple linear balance restrictions.
Figure 4.3: 95% confidence intervals for the parameter estimates of $\sigma$.

### 4.8.2 One linear balance restriction

Clearly, the singular normal distribution can also be applied to data that are subject to only one linear balance restriction ($p = 1$). This means that we can compare the results for this imputation method with the results that were obtained in chapter 3. Recall, however, that the methods in chapter 3 also included non-negativity constraints, which is not the case for the singular normal model. Consequently in this instance the individual imputed values may violate the non-negativity constraints. For parameter estimation this will probably not pose any problems as it is highly unlikely that the parameter estimates will become negative.

Two datasets were used, the first dataset dealt with labour costs and the second with costs of third party rendering of services. Then 100 random samples of missing data were generated and next the remaining datasets containing miss-
ing data items were used for parameter estimation, either by using the incomplete data or data completed by imputation. The averaged parameter estimates and their 95% confidence intervals that are provided for the incomplete datasets using the EM algorithm for singular normal data (SN) are added to the results from chapter 3, which is presented in Figures 4.2, 4.3, 4.4 and 4.5.

For the data on labour costs (Figures 4.2 and 4.3) it is observed that the parameter estimates generated by the EM algorithm for the singular normal distribution are quite accurate point estimates for $\mu$ and $\sigma$. The nonresponse variance is, however, somewhat larger than for example the nonresponse variance produced by the Dirichlet method with expectations imputed (Dir).
Figure 4.5: 95% confidence intervals for the parameter estimates of $\sigma$.

With respect to the data on costs of third party rendering of services (Figures 4.4 and 4.5) the EM algorithm for singular normal data produces very accurate point estimates for $\mu$, again with a somewhat larger nonresponse variance than the Dirichlet method. This method is less capable of finding accurate point estimates for $\sigma$, however, and in this case the nonresponse variance is again quite substantial.

In conclusion, the EM algorithm for singular normal data appears to be a promising approach. For the mean parameters accurate point estimates are obtained, but on the other hand the dispersion parameters are less well estimated. Besides, the nonresponse variance can be substantial in some cases, meaning
that the parameter estimates depend on the realised set of missing data. Furthermore, as the EM algorithm for singular normal data converges quite slowly the procedure can become time-consuming, in which case the other imputation methods are probably preferred.

The main advantage of this method, however, is that it is able to impute data that are subject to any set of multiple linear balance restrictions. Besides, in this case we do not need to assume that the variable, which represents a certain total, is completely observed. In the next section the effects on parameter estimation for a dataset subject to multiple linear balance restrictions are analysed.

### 4.8.3 Multiple linear balance restrictions

#### 4.8.3.1 Description of the data

The dataset that will be examined concerns the cost structure of companies in part of the wholesale industry with more than 10 employees. The variables $X_{1t}$ and $X_{2t}$, representing total labour costs and total costs of third party rendering of services respectively, that were used in the previous analyses are in fact part of the total operating expenses of a business. The total operating expenses ($X_t$) is composed of

\[
\begin{align*}
X_{1t} &= \text{total labour costs} \\
X_{2t} &= \text{total costs of rendering of services} \\
X_{3t} &= \text{total costs of sales} \\
X_{4t} &= \text{total other personnel costs} \\
X_{5t} &= \text{total costs of transportation} \\
X_{6t} &= \text{total costs of energy} \\
X_{7t} &= \text{total costs of housing} \\
X_{8t} &= \text{total costs of machinery and equipment} \\
X_{9t} &= \text{total sales expenses} \\
X_{10t} &= \text{communication expenses} \\
X_{11t} &= \text{other company expenses} \\
X_{12t} &= \text{depreciation on fixed assets}.
\end{align*}
\]
The variables $X_{11}, \ldots, X_{14}$ and $X_{21}, \ldots, X_{26}$, used in chapter 3 and the previous section are also present in the dataset. The following balance restrictions hold:

\begin{align}
X_{1t} &= X_{11} + \cdots + X_{14} \quad (4.9) \\
X_{2t} &= X_{21} + \cdots + X_{26} \quad (4.10) \\
X_t &= X_{1t} + \cdots + X_{12t} \quad (4.11)
\end{align}

As these variables represent some sort of operating expense, they all are non-negative. Moreover, most of the variables $X_{3t}, \ldots, X_{12t}$ also represent the total of an underlying balance restriction, which are left out for now in order to reduce the complexity of the model.

Again records with missing items are removed from the original survey and missing values are generated in this dataset based on the MCAR assumption, using Bernoulli draws with parameter $p$ chosen such that the nonresponse rate is similar to the rate observed in the original survey. As most of these variables are aggregate values, the observed nonresponse rates for the variables $X_{1t}, \ldots, X_{12t}$, $X_t$ are much lower (about 17\%) than the nonresponse rates that were found for the variables $X_{11}, \ldots, X_{14}$ and $X_{21}, \ldots, X_{26}$ (about 31\%). After generation of the missing items deductive imputation is applied to impute the data items that can be derived with certainty, based on the edit constraints given by (4.9), (4.10), (4.11) and the non-negativity restrictions.

Parameter estimates are calculated based on the remaining dataset. The complete cases (CC), available cases (AC) and the EM algorithm for the singular normal model (SN) are used to obtain parameter estimates. Due to the fact that there are no other imputation methods that can straightforwardly yield imputed values or parameter estimates that satisfy multiple linear balance restrictions, no other imputation methods can be used for comparison purposes.

This means that we can only assess the performance of the singular normal model with respect to the complete and available cases estimates. In chapter 3 and section 4.8.2 we found that these estimates strongly depend on the realised set of missing data. As the amount of missing values is reduced in this dataset we expect that the incomplete data procedure that uses the available cases will result in accurate point estimates with lower nonresponse variance.

#### 4.8.3.2 The effects on parameter estimation

In Figures 4.6 and 4.7 the parameter estimates and their 95\% confidence intervals for respectively $\mu$ and $\sigma$ are provided. Due to the high number of variables present in this dataset, only the results for the aggregate variables, $X_{1t}, \ldots, X_{12t}$,
and $X_t$ are shown. It is clear that the method based on the available cases now produces good results with respect to accuracy as well as precision for the estimation of both $\mu$ and $\sigma$. The balance restrictions that need to hold are, however, violated. The average absolute violation relative to the variable representing the total of that restriction is 8%, 9% and 8% for restrictions (4.9), (4.10) and (4.11) respectively.

For the mean parameter, the estimates that are obtained with the EM al-
Figure 4.7: 95% confidence intervals for the parameter estimates of $\sigma$. 

algorithm for singular normal data closely resemble the available cases estimates, so both methods appear quite capable of preserving the true parameter estimates. The confidence intervals, indicating the sensitivity of the results with respect to the realised set of missing data, are quite acceptable as well for both methods. A striking result is the fact that the parameter estimates obtained with the EM algorithm for $\mu_{1t}$ and $\mu_{2t}$ are very close to the true value with a relatively small confidence interval. This is probably due to the fact that the vari-
ables $X_{1t}$ and $X_{2t}$ are aggregates of the variables $X_{11}, \ldots, X_{14}$ and $X_{21}, \ldots, X_{26}$ respectively, which are also taken into account in the imputation process. This means that, although these variables contain missing values as well, more information is available for the imputation of $X_{1t}$ and $X_{2t}$, which clearly provides better estimates. This is an encouraging result as most of the other variables $X_{3t}, \ldots, X_{12t}$ are also the aggregate of an underlying balance restriction. These variables were left out in this analysis as the dataset would have become quite substantial in that case, meaning that a simulation study would take up too much computing time. This will certainly be a topic of further research, however.

With regard to the estimation of $\sigma$, the available cases estimates appear to have a slight advantage over the estimates obtained through the EM algorithm for singular normal data. The latter shows on average less accuracy and a larger nonresponse variance. Moreover, the EM estimates appear to have a tendency to somewhat overestimate $\sigma$. The major disadvantage of the available cases method is of course that the covariance matrix produced is not singular. With respect to the variables $X_{1t}$ and $X_{2t}$ again we observe that the estimates produced by the EM algorithm are very accurate, with small confidence intervals. This means that also for the estimation of $\sigma$ it appears quite beneficial to incorporate information from underlying balance restrictions. Also note that for these variables the EM estimates do not overestimate $\sigma$. In summary, we have proposed a promising new method to take multiple linear balance restrictions into account.

4.9 Concluding remarks

In this chapter we have discussed the use of the singular normal distribution to generate imputations for data that are subject to multiple linear restrictions. We have shown that, by choosing appropriate starting values, the EM algorithm for the singular normal distribution will produce maximum likelihood estimates for the mean and dispersion parameters that concur with the balance restrictions. This means that imputations can be obtained that satisfy multiple linear balance restrictions, which, to our knowledge, is not possible with any other imputation technique developed so far.

Another advantage of this model is the fact that the balance restrictions need not be specified as they are embedded in the singularity of the covariance matrix, which is very convenient.

In section 4.8 on empirical data we found that this method is quite capable of
obtaining accurate parameter estimates for the population parameters on mean and dispersion for both data subject to one restriction as well as data subject to multiple balance restrictions. For the latter case the results indicate that the performance of the singular normal model strongly improves if data items are present in multiple balance restrictions, which is not surprising as more information for imputation is available in that instance. This area could be a topic for future research on this imputation method.

The performance of this procedure on individual level has not been investigated as there were no alternative imputation procedures available that could impute data subject to multiple restrictions, immediately satisfying these restrictions. If more procedures become available, this could also be a subject for further research.

The procedure is generally applicable to continuous data as the singular normal model can incorporate both variables that are present in a balance restriction as well as variables that are not. This means that the imputation method can be used for a large part of a business survey at once. So variables concerning company expenses, company turnover, company profits and employment can be simultaneously imputed using all available balance restrictions. A downside however, reducing its general applicability, is the fact that non-negativity restrictions cannot be incorporated. This means that the singular normal distribution will mostly be of use for the imputation of data representing aggregate variables that are far from zero.

Sometimes economic data need to satisfy other inequality restrictions as well. Clearly, the singular normal distribution does not take inequality restrictions into account. This means that there is still a need for an imputation procedure that can deal with both balance and inequality restrictions simultaneously. This will be the main topic of the next chapter.