Statistical Matching and the Likelihood Principle: Uncertainty and Logical Constraints

Marcello D'Orazio, Marco Di Zio and Mauro Scanu[1]

ISTAT, Italian National Statistical Institute, Roma, Italy
madorazi@istat.it, dizio@istat.it, scanu@istat.it

Abstract

Statistical Matching is the art of combining information from different sources. In particular it tackles the problem of variables not jointly observed. This situation leads to handle the problem of drawing conclusions when just partial knowledge of the phenomenon is available. Thus, uncertainty on conclusions arises naturally, unless strong and non-testable hypotheses have been assumed. Hence the main goal of statistical matching can be reinterpreted as the study of the key aspects of uncertainty in this context, and what conclusions can be taken. In this paper we give a formalization of the concept of uncertainty in statistical matching when the variables are categorical, and formalize the key elements to be investigated. An estimator of the elements characterising uncertainty is suggested. This analysis leads to state some inferences also in this context. Furthermore, the introduction and the effect on uncertainty of logical constraints is studied. All the analyses have been performed according to the likelihood principle. An application on real data of the proposed tools and a comparison with other approaches already defined has been carried out.

Key words: Data fusion, Synthetical matching, Constrained maximum likelihood estimation, EM algorithm.

1 Introduction

Information plays a key role for understanding phenomena. In some cases it may be obtained by combining two or more data sources: some examples are database marketing (Kamakura and Wedel, 1997 and 2003) and economic research via microsimulation (e.g. the Social Policy Simulation Database created at Statistics Canada, Singh, Mantel, Kinack and Rowe, 1993, and references therein). Another particularly suitable application field is Official Statistics because of the high number of files maintained in National Statistical Institutes (NSIs).

Integration of data from different sources can be performed by means of three different methodologies: merging, record linkage and statistical matching. While the first two aim at linking the same units from two or more different files, the third one faces the problem of integration when the files do not contain the same units. The main target of statistical matching is to give joint information on variables observed in different sources. This integration problem may be represented by the following situation. There are two different sources, A and B, two groups of variables never jointly observed, Y in A and Z in B, and one group of variables available in both data sources, X (see Fig. 1).

[1]Marcello D'Orazio is Researcher, ISTAT, via Cesare Balbo 16, 00184 Roma, Italy (e-mail: madorazi@istat.it), Marco Di Zio is Researcher, ISTAT, via Cesare Balbo 16, 00184 Roma, Italy (e-mail: dizio@istat.it), and Mauro Scanu is Researcher, ISTAT, via Cesare Balbo 16, 00184 Roma, Italy (e-mail: scanu@istat.it).

Almost all the papers on statistical matching aimed at integrating the files at the unit level in order to obtain a comprehensive database with all the variables (i.e. a synthetical dataset). These techniques have been developed since the 1970s (see references in Rässler, 2002), and may be broadly divided in two large groups. The first one contains those techniques (implicitly) based on a specific model: Y and Z are probabilistically independent conditionally on X (Conditional Independence Assumption, CIA henceforth). When this model is not adequate, the integrated synthetical dataset may be significantly different from the truth and when the usual estimators of parameters are applied they may be heavily biased (Rodgers, 1984, Paass, 1986, Barry, 1988, Goel and Ramalingam, 1989, Singh *et al.*, 1993, Renssen, 1998). The second group of techniques faces this problem using auxiliary information on (Y, Z) (e.g. Singh *et al.* 1993, and references therein). In particular, Singh *et al.* (1993) show by simulation studies how the accuracy of results of the matching procedure can be improved in this setting.

However, both the groups of techniques are dependent on assumptions that cannot be tested. Actually many distributions on (X, Y, Z) are compatible with the available partial information, i.e. many different *worlds* may have generated the observed data, and those worlds are indistinguishable. On the contrary, the two groups of techniques are constrained to just a single world: in the first we assume that the world is that described by the CIA, in the second we describe the closest world (w.r.t. the Kullback-Leibler distance, see Csizár, 1975) to that of the auxiliary information (e.g. previous year) and coherent with data currently observed. Furthermore, the latter, while an important special case, is not always feasible (Ingram, O'Hare, Scheuren and Turek, 2000) because the required external information is either on the parameters on the statistical relationships between (Y, Z) or on the (Y, Z) distribution.

In this paper we describe a different approach to statistical matching. This approach consists in assessing all *the possible worlds*, i.e. all the parameters' values consistent with the available information. At first (Section 2) we analyse the case of marginal complete information on (X, Y) and (X, Z) and discuss what we intend for uncertainty. Then (Section 3) we consider the case when marginal information on (X, Y) and (X, Z) is provided by two independent samples. In this case, uncertainty is estimated by Maximum Likelihood, i.e. all the possible worlds maximising the likelihood are equally informative and are taken into consideration. We also suggest the use of the elements characterising uncertainty in order to draw some conclusions (decisions) on parameters' values. In order to exclude some non-possible worlds, it is important to introduce logical constraints (when available), i.e. constraints characterising the phenomenon. In Section 4 we will consider structural zeros and inequality constraints between pairs of distribution parameters. Their introduction implies, as expected, a decrease of the uncertainty on the parameters.

This way of dealing with the statistical matching problem has some similarities to those in Rubin (1986), Moriarity and Scheuren (2001), and Rässler (2002), that refer to the continuous case. It is also worth to note that the concept of uncertainty has been investigated in other scientific contexts, see for instance Walley (1991) and Manski (1995). Finally in section 5 we develop a toy-example in the context of NSI to better show advantages and drawbacks of the proposed method. In the last section, concluding remarks and some directions for further research are presented.

All the considerations in the next sections have been developed when X, Y and Z are univariate variables. The extension to the multivariate context is straightforward, provided the blocks of observed data are as in Fig. 1.

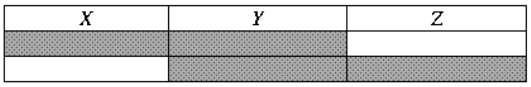

Figure 1: Typical situation for statistical matching in a unit (row) by variable (column) matrix. Spaces in grey correspond to observed data, while white spaces are missing data. The first block of data corresponds to source A, while the second to source B.

2 Uncertainty in a statistical matching context

This section underlines that the statistical matching context is inevitably characterised by uncertainty, i.e. even in the optimal case of complete knowedge on the (X, Y) and (X, Z) distributions, it is not possible to draw unique and certain conclusions on the overall distribution (X, Y, Z). Actually in a real context, just two samples from respectively (X, Y) and (X, Z) are available. The statistical analysis for this context is discussed in section 3.

Let us consider the triplet (X, Y, Z) with respectively I, J, and K categories,

$$\Delta = \{(i, j, k) : i = 1, \ldots, I, j = 1, \ldots, J, k = 1, \ldots, K\},$$

whose distribution is:

$$\theta^*_{ijk} = P(X = i, Y = j, Z = k) \qquad i, j, k \in \Delta. \tag{1}$$

If the overall distribution (1) is unknown, the distribution $\{\theta^*_{ijk}\}$ may be one in the following set:

$$\Theta = \left\{ \boldsymbol{\theta} : \theta_{ijk} \geq 0; \sum_{i,j,k} \theta_{ijk} = 1 \right\}. \tag{2}$$

Actually the distribution $\{\theta^*_{ijk}\}$ is *totally uncertain*.

Let us now assume that the marginal distributions for the two pairs (X, Y) and (X, Z) are perfectly known:

$$\theta^*_{ij.}, \qquad i, j \in \Delta; \tag{3}$$

$$\theta^*_{i.k}, \qquad i, k \in \Delta. \tag{4}$$

Information in (3) and (4) restricts the set of possible distributions Θ to the following subset:

$$\begin{cases} \sum_k \theta_{ijk} = \theta^*_{ij.} & i, j \in \Delta \\ \sum_j \theta_{ijk} = \theta^*_{i.k} & i, k \in \Delta \\ \theta_{ijk} \in \Theta. \end{cases} \tag{5}$$

Each set (2) and (5) represents uncertainty, i.e. multiplicity of plausible solutions given the available information. In particular (5) may be considered as the description of the uncertainty connected to the statistical matching problem when complete knowledge on the marginal distributions is available.

Both (2) and (5) have the following characteristics:

1. each parameter θ_{ijk} has associated an interval $\theta^L_{ijk} \leq \theta^U_{ijk}$; in particular in set (2) $\theta^L_{ijk} = 0$ and $\theta^U_{ijk} = 1$ for all (i, j, k), while in set (3) $\theta^L_{ijk} > 0$ and $\theta^U_{ijk} < 1$ possibly for some (i, j, k);

2. the true, but unknown, parameter lies in the previous interval $\theta^L_{ijk} \leq \theta^*_{ijk} \leq \theta^U_{ijk}$;

3. the frequency of all the plausible parameter values forms a polynomial distribution, $m_{ijk}(\theta)$, $\theta^L_{ijk} \leq \theta \leq \theta^U_{ijk}$. For example, when the marginal distributions are known, the polynomial distribution $m_{ijk}(\theta)$ is computed on all the distributions in (5) and the degree of the polynomial depends on the number of categories and the number of constraints (in this case the two marginal distributions (X, Y) and (X, Z)), while the coefficients depend on the marginal distributions for (X, Y) and (X, Z).

These characteristics represent the uncertainty on each single parameter. In particular a key role is assumed by the distribution $m_{ijk}(\theta)$ and its dispersion: the less it is dispersed, the less we are uncertain about the parameter value. Hence we underline that, in this context, it is not only important to assess the width of the interval of equally plausible (given the marginals) values (as it has been described in Rässler, 2002). Let us consider two parameters, θ_{ijk} and $\theta_{i'j'k'}$, and assume that the interval of equally plausible values have the same width:

$$\theta^U_{ijk} - \theta^L_{ijk} = \theta^U_{i'j'k'} - \theta^L_{i'j'k'}.$$

Let us cut a given α% of the tails of the distributions $m_{ijk}(\theta)$ and $m_{i'j'k'}(\theta)$, so that the trimmed interval is the shortest possible. If the new trimmed interval for θ_{ijk} is narrower than the one for $\theta_{i'j'k'}$, we may say that θ_{ijk} is less uncertain than $\theta_{i'j'k'}$. In fact, discarding an equal number of equally plausible distributions for both the parameters leads to a more intense reduction of the interval width for the first parameter. With an abuse of notation the percentages of the cut tails can be denominated as *error frequency*, that is the frequency of distributions that, given the available marginal constraints, may be the true but unknown one, and are not considered in the chosen interval.

Uncertainty in terms of $m_{ijk}(\theta)$ may suggest different decisions. If the decision is in terms of an interval $(a, b) \subset (\theta^L_{ijk}, \theta^U_{ijk})$, it is important to compare the error frequency

$$\int_{\theta^L_{ijk}}^{a} m_{ijk}(\theta)\, d\theta + \int_{b}^{\theta^U_{ijk}} m_{ijk}(\theta)\, d\theta \tag{6}$$

with the gain given by the reduction of the interval width

$$1 - (b - a)/(\theta^U_{ijk} - \theta^L_{ijk}). \tag{7}$$

Once α is fixed, the shortest interval (a, b) is determined by the region $\Gamma = \{\theta : m_{ijk}(\theta) \geq c\}$ where c is chosen such that $1 - \alpha = \int_\Gamma m_{ijk}(\theta)\, d\theta$.

The dispersion of $m_{ijk}(\theta)$ can suggest also a punctual approximation of the true parameter θ^*_{ijk}. For instance, a reasonable approximation for θ^*_{ijk} can be the average of $m_{ijk}(\theta)$. Let $\bar{\theta}_{ijk}$ be this average value:

$$\bar{\theta}_{ijk} = \int_{\theta^L_{ijk}}^{\theta^U_{ijk}} \theta m_{ijk}(\theta)\, d\theta. \tag{8}$$

It is easy to see that $\boldsymbol{\theta} = \{\bar{\theta}_{ijk}\}$ describes a distribution (each parameter is non-negative and their sum is 1). A useful description of the goodness of the approximation $\bar{\theta}_{ijk}$ is the dispersion of $m_{ijk}(\theta)$, for each single parameter. As a limit case, when the dispersion is null, the marginal distributions are sufficient for determining θ^*_{ijk}.

There are practical situations where logical contraints are available. In these contexts it is necessary to use these additional partial constraints in order to exclude impossible distributions in (5), thus reducing uncertainty. This aspect will be studied in Section 4.

3 The statistical model

Let us consider n i.i.d. realisations of (X, Y, Z). Dealing with discrete variables from the multinomial distribution in (1) and denoting the vector of observed frequencies with

$$\mathbf{n} = \{n_{ijk}, (i,j,k) \in \Delta\}$$

where n_{ijk} is the number of units in the sample with $(X = i, Y = j, Z = k)$, then the *complete* likelihood in the present context obviously is:

$$L(\boldsymbol{\theta}|\mathbf{n}) = \prod_{i,j,k} \theta_{ijk}^{n_{ijk}}, \qquad \boldsymbol{\theta} \in \Theta. \tag{9}$$

The statistical matching context in Fig. 1 happens when the n statistical units are divided in two subgroups A and B (two independent subsamples) of respectively n_A and n_B units, $n_A + n_B = n$. Let us assume that Z is not observed on the units in A and Y is not observed in B. According to Rässler (2002, p. 75), we suppose that this missing data mechanism is ignorable. Under this assumption, in the situation of Fig. 1, marginalisation of the complete data likelihood (9) gives the *observed* data likelihood (Little and Rubin, 1983):

$$L(\boldsymbol{\theta}|\mathbf{n}_A, \mathbf{n}_B) = \prod_{i,j} (\theta_{j|i}\theta_{i..})^{n^A_{ij.}} \prod_{i,k} (\theta_{k|i}\theta_{i..})^{n^B_{i.k}}, \qquad \boldsymbol{\theta} \in \Theta, \tag{10}$$

where $\theta_{j|i} = \theta_{ij.}/\theta_{i..}$, $\theta_{k|i} = \theta_{i.k}/\theta_{i..}$, $\mathbf{n}_A$ and $\mathbf{n}_B$ have the same meaning as $\mathbf{n}$. Note that the factorization in (10) is a straightforward application of the Factorization Lemma in Rubin (1974). Although (10) is a function of the overall distribution $\boldsymbol{\theta} \in \Theta$, the right hand side depends explicitly only on some marginal parameters. The maximum of (10) with respect to these parameters is uniquely determined by:

$$\hat{\theta}_{j|i} = \frac{n^A_{ij.}}{n^A_{i..}},\ \hat{\theta}_{k|i} = \frac{n^B_{i.k}}{n^B_{i..}},\ \hat{\theta}_{i..} = \frac{n^A_{i..} + n^B_{i..}}{n}, \quad (i,j,k) \in \Delta. \tag{11}$$

The previous statements allow us to find the final estimates for the following parameters:

$$\hat{\theta}_{ij.} = \frac{n^A_{ij.}}{n^A_{i..}} \frac{n^A_{i..} + n^B_{i..}}{n},\ \hat{\theta}_{i.k} = \frac{n^B_{i.k}}{n^B_{i..}} \frac{n^A_{i..} + n^B_{i..}}{n},\ \hat{\theta}_{i..} = \frac{n^A_{i..} + n^B_{i..}}{n},$$

$(i,j,k) \in \Delta$.
However we are interested in estimating the overall distribution $\boldsymbol{\theta}$. The maximum of the observed likelihood function (10) in θ_{ijk} is not unique. Every distribution $\boldsymbol{\theta} = \{\theta_{ijk}\}$ that satisfies the following set of equations:

$$\begin{cases} \sum_k \theta_{ijk} = \hat{\theta}_{ij.} = \frac{n^A_{ij.}}{n^A_{i..}} \left(\frac{n^A_{i..} + n^B_{i..}}{n}\right) \\ \sum_j \theta_{ijk} = \hat{\theta}_{i.k} = \frac{n^B_{i.k}}{n^B_{i..}} \left(\frac{n^A_{i..} + n^B_{i..}}{n}\right) \\ \theta_{ijk} \geq 0, \ \ \sum_{i,j,k} \theta_{ijk} = 1 \end{cases} \tag{12}$$

is an MLE. The set composed by all the MLEs forms a region called *likelihood ridge*. It is easy to see that the likelihood ridge is the MLE of the set (5), and consequently may be used for estimating the uncertainty of the statistical matching process. Given that the likelihood ridge is composed by maximum likelihood estimates, all the distributions in

the likelihood ridge are equally informative, given the data. A consequence of the properties of likelihood estimators is that uncertainty is estimated according to the likelihood principle.

One of the most important features of (12) is that it is dependent on the samples $\mathbf{n}_A$ and $\mathbf{n}_B$ through the maximum likelihood estimates $\hat{\theta}_{ij.}$ and $\hat{\theta}_{i.k}$, $(i,j,k) \in \Delta$. The sample variability of the likelihood ridge (12) decreases when n_A and n_B diverge to $+\infty$, due to the consistency of the MLEs of the marginal distributions $\hat{\theta}_{ij.}$ and $\hat{\theta}_{i.k}$, $(i,j,k) \in \Delta$. In other words, the likelihood ridge converges (almost surely) to the set of distributions in (5) that describes the uncertainty connected with the statistical matching context when *complete* knowledge on (X,Y) and (X,Z) is available. Another consequence is that we can use the MLE counterpart of θ^U_{ijk}, θ^L_{ijk} and $m_{ijk}(\theta)$, in the following $\hat{\theta}^U_{ijk}$, $\hat{\theta}^L_{ijk}$ and $\hat{m}_{ijk}(\theta)$. All these estimators are consistent, and can be usefully considered for the computation of (6), (7) and (8). Since uncertainty is a factor strongly characterising statistical matching, the most important thing is to reduce uncertainty, i.e. reduce the dispersion of the distributions $m(\theta)$. One possibility is offered by logical constraints (Section 4).

We also underline that the likelihood ridge (12) contains the solutions under some of the approaches already defined for statistical matching. For instance, under the CIA the parameters of the distribution $\boldsymbol{\theta}$ assume the form:

$$\theta_{ijk} = \frac{\theta_{ij.}\theta_{i.k}}{\theta_{i..}}. \tag{13}$$

Consequently, the (unique) maximum likelihood estimate is

$$\hat{\theta}_{ijk} = \frac{n^A_{ij.}}{n^A_{i..}} \left(\frac{n^B_{i.k}}{n^B_{i..}} \right) \left(\frac{n^A_{i..} + n^B_{i..}}{n} \right), \qquad \forall\, i,j,k,$$

and this distribution is clearly inside the likelihood ridge (12).

4 Logical constraints

There are situations when it is possible to introduce logical constraints. We intend for logical constraints those rules that make some of the distributions in Θ illogical for the investigated phenomenon. Thus their introduction is needed in order to eliminate non-possible worlds. Various are the examples of logical constraints. Two frequent cases, that we will use in the next paragraphs, are:

- *existence of some quantities*: e.g. it cannot be accepted that a unit in the population is both ten years old and married
- *inequality constraints*: e.g. a person with a degree has higher probability of being a manager rather than a worker.

For the statistical model described in section 3, they can be expressed as:

$$\theta_{ijk} = 0, \qquad \text{for some } (i,j,k) \tag{14}$$

$$\theta_{ijk} \leq \theta_{i'j'k'}, \qquad \text{for some } (i,j,k),\ (i',j',k'). \tag{15}$$

Constraint (14) is usually called *structural zero* (see, e.g., Agresti, 1990). This constraint occurs when :

1. (i, j, k) contains at least a pair of incompatible categories;

2. each pair in (i, j, k) is plausible but the triplet is incompatible.

Note that great caution should be posed on the definition of the set of logical constraints. It may happen that the constraints are not compatible each other, i.e. Θ is restricted to the empty set (see Bergsma and Rudas, 2002, for more details and references therein). From now on, we suppose that the chosen logical constraints are compatible.

The main effect of these constraints is the possible reduction of the likelihood ridge. It is clear that the size of the reduction is dependent on the amount of information introduced. In some circumstances, information carried by logical constraints can be so informative that, using them in addition to the observed marginal distributions (X, Y) and (X, Z), it is possible to reduce the likelihood ridge to a unique distribution. This happens, for instance, when $(J-1)(K-1)$ independent structural zero constraints are set for each $X = i$, for $i = 1, \ldots, I$ (i.e. maximum dependence among Y and Z conditional to X). Structural zeros are also very effective because, with the exception of limit cases, distributions $\{\theta_{ijk}\} \in \Theta$ satisfying the CIA become illogical. In fact, when θ_{ijk} is set to 0 for some (i, j, k), the CIA (i.e. parameters as in (13)) holds only when either $\hat{\theta}_{ij.} = 0$ and/or $\hat{\theta}_{i.k} = 0$, otherwise that distribution is outside the restricted parameter space, and cannot be considered in the estimation phase.

Generally speaking, let us suppose that the imposed logical constraints restrict Θ to a subspace $\Omega \subset \Theta$ which is closed and convex (any combination of structural zeros and inequality constraints leads to such restriction). The problem of the maximization of the likelihood function when constraints are imposed may be solved following two different strategies. These strategies refer to these situations: 1) Ω has a non-empty intersection with the unconstrained likelihood ridge (12); 2) Ω has an empty intersection with the unconstrained likelihood ridge (12).

In the first case the likelihood ridge reduces to the set of solutions of:

$$\begin{cases} \sum_k \theta_{ijk} = \hat{\theta}_{ij.} = \frac{n^A_{ij.}}{n^A_{i..}} \frac{n^A_{i..}+n^B_{i..}}{n} \\ \sum_j \theta_{ijk} = \hat{\theta}_{i.k} = \frac{n^B_{i.k}}{n^B_{i..}} \frac{n^A_{i..}+n^B_{i..}}{n} \\ \boldsymbol{\theta} \in \Omega. \end{cases} \tag{16}$$

In the second case it happens that the set of equations in (16) has not any solutions, i.e. Ω does not contain any derivative of the observed data likelihood (10) with respect to $\boldsymbol{\theta}$ equal to zero. In other words, the relative maximum(s) of the likelihood (10) may be only on the border of the subspace Ω. In this case we propose to use an iterative algorithm in order to find the maximum in $\boldsymbol{\theta}$ of (10) constrained to $\boldsymbol{\theta} \in \Omega$.

The likelihood maximisation problem in a proper closed and convex subset has been studied by many authors by means of many different approaches (e.g. see Judge, Griffiths, Hill and Lee, 1980, Chapter 17). We adopt a version of the "projection method" (Judge *et al.*, p. 749) described in Winkler (1993) that makes use of the EM algorithm (Dempster, Laird, and Rubin, 1979). It consists of the following steps:

1. initialize the algorithm with a $\hat{\boldsymbol{\theta}}^0 \in \Omega$

2. if at iteration t, $t \geq 1$, the EM unconstrained estimate $\hat{\boldsymbol{\theta}}^t$ does not satisfy the constraints, such solution is "projected" to the boundary of the closed and convex subspace Ω; otherwise it is left unchanged.

Table 1: Response categories for the variables considered in the example

Variables	Transformed response categories
Age (AGE)	"1"=15-17 years old; "2"=18-22; "3"=25-64; "4"=65 and more
Education Level (EDU)	"C"=None or compulsory school; "V"=Vocational school; "S"'= Secondary school; "D"=Degree
Professional Status (PRO)	"M"= Manager; "E"= Clerk; "W"= Worker

Such approach is convenient in our context because the likelihood in (9) is a mixture of multinomial distributions. In this case, a theorem by Haberman (theorem 4, 1977; see also Winkler, 1993) suggests the following: if $\hat{\boldsymbol{\theta}}^{t-1}$ and $\hat{\boldsymbol{\theta}}^{t}$, $t \geq 1$, are successive estimates, and $\hat{\boldsymbol{\theta}}^{t-1} \in \Omega$ while $\hat{\boldsymbol{\theta}}^{t} \notin \Omega$, then $\hat{\boldsymbol{\theta}}^{t}$ should be replaced by the linear combination of $\hat{\boldsymbol{\theta}}^{t-1}$ and $\hat{\boldsymbol{\theta}}^{t}$ so that $\alpha\hat{\boldsymbol{\theta}}^{t-1}+(1-\alpha)\hat{\boldsymbol{\theta}}^{t}$ lies on the boundary of Ω ($0 \leq \alpha \leq 1$). Given that Ω is closed and convex, such α exists and is unique. Additionally, the theorem by Haberman states that the likelihood of the successive M step solutions of this modified EM algorithm (Winkler calls this method: EMH) is non-decreasing. Judge *et al*. (1980) and Winkler warn that this algorithm may stuck at a solution on the boundary of Ω which is not a local maximum. Most of the times it is not difficult to determine α. In fact, structural zero constraints (14) may be easily fulfilled setting to zero the corresponding $\hat{\theta}^{0}_{ijk}$ in the initialisation step of the EM algorithm, for details see Schafer (1997, pp. 52-53). Also inequality constraints are easily fulfilled. In fact, inequality (15) is satisfied when

$$\alpha = \frac{\hat{\theta}^{t}_{i'j'k'} - \hat{\theta}^{t}_{ijk}}{\hat{\theta}^{t-1}_{ijk} - \hat{\theta}^{t-1}_{i'j'k'} - \hat{\theta}^{t}_{ijk} + \hat{\theta}^{t}_{i'j'k'}}.$$

If more than one inequality constraint is imposed, the smallest α should be considered.

We remark that the multinomial model in (9) is saturated, consequently each M step in the EM algorithm gives solutions in closed form. However, if a different loglinear model is assumed, the ECM algorithm can be adopted instead of the EM algorithm, as in Winkler (1993).

5 An example

In order to show how to introduce logical constraints and the corresponding advantages and drawbacks, we have developed a toy-example in the context of Official Statistics where logical constraints are frequently used. A subset of 2,313 employees (people at least 15 years old) has been extracted from the 2000 pilot survey of the Italian Population and Households Census. Only three variables have been analyzed: Age (AGE), Educational Level (EDU) and Professional Status (PRO). For the sake of simplicity and without loss of information for our aim, the original variables have been transformed by grouping homogeneous response categories. The results of this grouping are shown in Tab. 1.

To reproduce the situation of Fig. 1, the original file has been randomly split in two almost equal sub-sets. The variable Educational Level has been removed from the first sub-set (file A), containing 1,148 units, and the variable Professional Status has been removed from the second sub-set (file B), consisting of the remaining 1,165 observations.

Tab. 2 shows the true relative frequencies of the original dataset for each cell. Structural zeros are represented by "-". For instance a 17 years old person cannot have a

Table 2: True cell counts (n_{ijk}) and relative frequencies (θ_{ijk}), and corresponding CIA estimates ($\hat{\theta}_{ijk}$)

Cell	AGE	EDU	PRO	n_{ijk}	θ_{ijk}	$\hat{\theta}_{ijk}$
1	1	C	M	-	-	-
2	2	C	M	-	-	-
3	3	C	M	-	-	0.0540
4	4	C	M	-	-	0.0048
5	1	V	M	-	-	-
6	2	V	M	-	-	-
7	3	V	M	-	-	0.0143
8	4	V	M	-	-	-
9	1	S	M	-	-	-
10	2	S	M	-	-	-
11	3	S	M	142	0.0614	0.0649
12	4	S	M	4	0.0017	0.0013
13	1	D	M	-	-	-
14	2	D	M	-	-	-
15	3	D	M	220	0.0951	0.0220
16	4	D	M	5	0.0022	0.0009
17	1	C	E	-	-	-
18	2	C	E	-	-	0.0022
19	3	C	E	-	-	0.1336
20	4	C	E	-	-	0.0009
21	1	V	E	-	-	-
22	2	V	E	1	0.0004	0.0009
23	3	V	E	123	0.0532	0.0350
24	4	V	E	0	0	0
25	1	S	E	-	-	-
26	2	S	E	8	0.0035	0.0022
27	3	S	E	653	0.2823	0.1604
28	4	S	E	3	0.0013	0.0004
29	1	D	E	-	-	-
30	2	D	E	-	-	-
31	3	D	E	87	0.0376	0.0545
32	4	D	E	0	0	0
33	1	C	W	15	0.0065	0.0065
34	2	C	W	27	0.0117	0.0078
35	3	C	W	759	0.3281	0.1466
36	4	C	W	12	0.0052	0.0017
37	1	V	W	0	0	0
38	2	V	W	7	0.0030	0.0035
39	3	V	W	90	0.0389	0.0385
40	4	V	W	0	0	0
41	1	S	W	-	-	-
42	2	S	W	12	0.0052	0.0073
43	3	S	W	143	0.0618	0.1755
44	4	S	W	0	0	0.0004
45	1	D	W	-	-	-
46	2	D	W	-	-	-
47	3	D	W	2	0.0009	0.0597
48	4	D	W	0	0	0.0004
Tot.				2313	1.0000	1.0000

Table 3: Distribution of Professional Status vs Age in file A

	Professional Status			
Age	**M**	**E**	**W**	Tot.
1	-	-	9	9
2	-	5	17	22
3	179	443	486	1108
4	6	1	2	9
Tot.	185	449	514	1148

Table 4: Distribution of Education Level vs Age in file B

	Education Level				
Age	**C**	**V**	**S**	**D**	Tot.
1	6	0	-	-	6
2	14	6	13	-	33
3	387	102	464	158	1111
4	10	0	3	2	15
Tot.	417	108	480	160	1165

degree. Tab. 3 and 4 show respectively the distribution of Age vs Professional Status in file A, and Age vs Educational Level in file B, after the original data-set has been splitted. Note that each structural zero in a marginal table implies a set of structural zeros on the joint distribution. But the joint distribution has some additional structural zeros that cannot be inferred from the marginal tables 3 and 4 because they refer to structural zeros of the variables (PRO,EDU). This happens, for instance, in cells 3 and 4 that correspond to managers (PRO = "M") but with at maximum a compulsory school educational level (EDU ="C").

If the two files are matched by means of a technique based only on the common variable Age, without considering any auxiliary information about the relationship existing between the two variables Educational Level and Professional Status, the final output would give estimates under the CIA. In our case, results under the CIA are reported in the last column of Tab. 2. As it can be observed, the CIA produces unrealistic estimates for some cells. In particular, it gives non zero estimated probabilities for certain events that in real life cannot happen, i.e. structural zeros for (PRO,EDU). On the contrary, as expected, structural zeros are preserved when observed in the marginals tables 3 and 4, e.g. cells 9, 25 and 41 corresponding to the structural zero (AGE = "1"), (EDU = "S").

In order to explore the likelihood ridge, we have decided to run the EM algorithm with different random starting points:

S0 starting point in the full space Θ;

S1 starting point in the space Ω resctricted by structural zeros;

S2 starting point in the space Ω resctricted by structural zeros and the following inequality constraint:
P(AGE ="3", EDU ="D", PRO ="M")$\geq$P(AGE ="3", EDU ="D", PRO ="E").

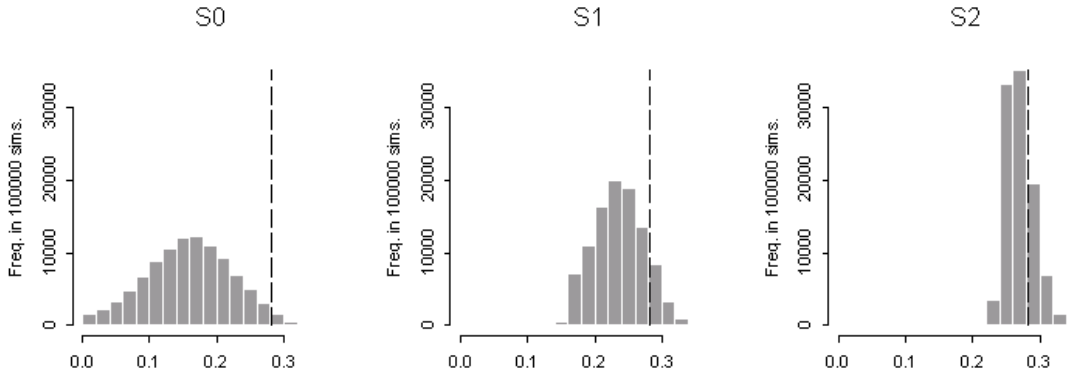

Figure 2: Likelihood ridge for cell 27 when no constraints (left), structural zeros (center) and the additional inequality constraint (right) are imposed. The vertical bar is the true probability.

The last inequality states that, a person with Age in class "3" with Educational Level in class "D" has a higher probability of being a Manager (PRO ="M") (cell 15) rather than a Worker (PRO ="E") (cell 31). In this last case we have used a modified version of EM so to satisfy both structural zeros and the inequality constraint regarding these cells.

Tab. 5 reports the simulation extremes of the likelihood ridge found by running EM 100,000 times for each of the above mentioned different starting configurations. As expected, when no restrictions were imposed on the starting point (S0), EM produces non-null estimates in correspondence of the structural zeros as under the CIA. In this case the CIA solution is always included in the interval found through EM. On the contrary, when structural zeros are introduced in the starting point (S1), EM produces zero estimated probabilities in correspondence of structural zeros. Moreover, for non-null probabilities it can be observed how the introduction of this kind of auxiliary information results in a general reduction of the ranges of estimated cell probabilities. When, in addition to the structural zeros, the inequality constraint involving cells 15 and 31 is introduced (S2) the results change quite markedly. The introduction of this inequality constraint makes the likelihood ridge shrink (see e.g. Fig. 2 and 3).

In general, in comparison with the initial situation of absence of auxiliary information about the phenomena under study (S0), an overall reduction of ranges for most of the estimated probabilities can be observed. When the final ranges (S2) of estimated probabilities are compared with those of S1 it comes out that about a half of them remains unchanged while for the others a decrease occurs. This reduction is really marked for cell 31 where the maximum for this estimated probability reduces from 0.1364 to 0.0678. On the other hand, for cell 15 the maximum remains unchanged while the lower value increases from 0 to 0.0260. For cells 33-36 it can be observed that the introduction of structural zeros is so informative (in terms of degrees of freedom) that makes the EM converge to a unique value, in all cases close to the true ones.

The width is not the only element to consider as an evaluation measure of the uncertainty on the parameters. The density of each single parameters in the likelihood ridge is another important aspect. We have approximated such density with the frequency distribution relative to the 100,000 simulations. In general, the dispersion reduces, also for those cells where the width of the interval does not change from S1 to S2.

In figures 2 and 3 we represent the evolution of this density in the three simulation context here considered for cells 15 and 27.

The joint analysis of the range and dispersion is essential to understand the uncer-

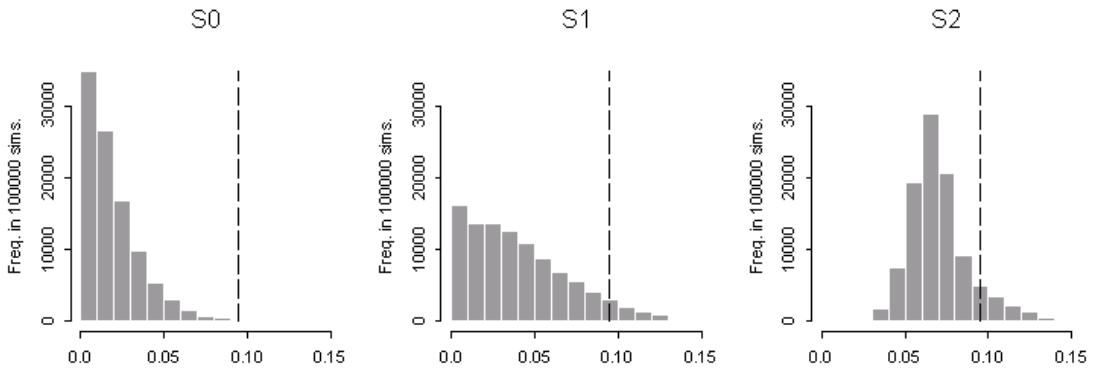

Figure 3: Likelihood ridge for cell 15 when no constraints (left), structural zeros (center) and the additional inequality constraint (right) are imposed. The vertical bar is the true probability.

tainty. Fig. 2, shows how the final distribution (S2) of the parameter in the likelihood ridge is totally different if compared to the initial one (S0), and in particular it is more concentrated. In other words, there is at the same time a shrinkage of the interval of values for the parameter and a higher concentration in its distribution. In Fig. 3, the dispersion and width of the interval again decreases from S1 to S2. This does not happen when S2 is compared with S0, although the effect of the constraints is still important. In fact, in this case the S2 distribution is more concentrated near the true value than S0.

The introduction of auxiliary information reduces uncertainty but does not eliminate it. However, as already introduced in (Section 4), a further reduction of uncertainty can be obtained by trimming a number of extreme solutions. For example, for cell 27 the cut of 5% of EM solutions in the tail of its distribution produces a reduction of about 50% of the range. Obviously we performed this operation on single parameters without caring of what happens to other cells; in fact the distributions whose parameter value for a cell is extreme does not necessarily have extreme values for other cells.

Finally, in the last column of Tab. 6 we have reported the average value ($\bar{\theta}_{ijk}$) of the 100,000 estimates respectively for S1 (that will be used in the next paragraph) and for S2, as representative parameters values among those in the ridge.

As a last remark, logical constraints do not only help in reducing uncertainty for inestimable parameters, but also may improve the MLE for estimable parameters (i.e. the ones with a unique MLE, in our example the ones for the marginal distributions (AGE, PRO) and (AGE, EDU)). In particular, the MLE for estimable parameters has the following behaviour:

- for those samples whose likelihood ridge is compatible with the constraints (i.e. some distributions in the ridge satisfy the constraints), the MLE of the estimable parameters remains unchanged in the constrained and unconstrained case;
- for those samples whose likelihood ridge is not compatible with the constraints, the MLE of the estimable parameters is forced to respect the constraints.

Hopefully, in the second case each constrained maximum likelihood estimate is moved towards the true parameter.

We have shown this last situation in our example, where some structural zeros (imposed in S1) are not compatible with the unconstrained likelihood ridge (12) when the samples in tables 3 and 4 are observed. For instance, let us consider

$$\theta_{ijk} = P(AGE = \text{“4”},\ PRO = \text{“}M\text{”},\ EDU = \text{“}C\text{”}).$$

Table 5: Range of probabilities estimates in 100,000 runs of EM with three different starting settings compared with the true counts (n_{ijk}) and frequencies (θ_{ijk})

						S0		S1		S2	
Cell	AGE	EDU	PRO	n_{ijk}	θ_{ijk}	Min	Max	Min	Max	Min	Max
1	1	C	M	-	-						
2	2	C	M	-	-						
3	3	C	M	-	-	0.0000	0.1549				
4	4	C	M	-	-	0.0035	0.0067				
5	1	V	M	-	-						
6	2	V	M	-	-						
7	3	V	M	-	-	0.0000	0.0839				
8	4	V	M	-	-						
9	1	S	M	-	-						
10	2	S	M	-	-						
11	3	S	M	142	0.061	0.0000	0.1549	0.0186	0.1550	0.0186	0.1290
12	4	S	M	4	0.002	0.0000	0.0021	0.0024	0.0031	0.0024	0.0031
13	1	D	M	-	-						
14	2	D	M	-	-						
15	3	D	M	220	0.095	0.0000	0.1260	0.0000	0.1363	0.0260	0.1364
16	4	D	M	5	0.002	0.0000	0.0014	0.0013	0.0021	0.0013	0.0021
17	1	C	E	-	-						
18	2	C	E	-	-	0.0000	0.0054				
19	3	C	E	-	-	0.0000	0.3261				
20	4	C	E	-	-	0.0000	0.0012				
21	1	V	E	-	-						
22	2	V	E	1	0.000	0.0000	0.0043	0.0000	0.0043	0.0000	0.0043
23	3	V	E	123	0.053	0.0000	0.0880	0.0014	0.0881	0.0015	0.0881
24	4	V	E	0	0	0	0	0	0	0	0
25	1	S	E	-	-						
26	2	S	E	8	0.004	0.0000	0.0054	0.0011	0.0054	0.0011	0.0054
27	3	S	E	653	0.282	0.0000	0.3776	0.1591	0.3776	0.2279	0.3780
28	4	S	E	3	0.001	0.0000	0.0012	0.0000	0.0007	0.0000	0.0007
29	1	D	E	-	-						
30	2	D	E	-	-						
31	3	D	E	87	0.038	0.0000	0.1362	0.0000	0.1364	0.0000	0.0678
32	4	D	E	0	0	0.0000	0.0011	0.0000	0.0007	0.0000	0.0007
33	1	C	W	15	0.006	0.0065	0.0065	0.0065	0.0065	0.0065	0.0065
34	2	C	W	27	0.012	0.0047	0.0101	0.0101	0.0101	0.0101	0.0101
35	3	C	W	759	0.328	0.0000	0.3278	0.3342	0.3342	0.3342	0.3342
36	4	C	W	12	0.005	0.0000	0.0023	0.0052	0.0052	0.0052	0.0052
37	1	V	W	0	0	0	0	0	0	0	0
38	2	V	W	7	0.003	0.0000	0.0043	0.0000	0.0043	0.0000	0.0043
39	3	V	W	90	0.039	0.0000	0.0880	0.0000	0.0866	0.0000	0.0865
40	4	V	W	0	0			0	0	0	0
41	1	S	W	-	-						
42	2	S	W	12	0.005	0.0040	0.0094	0.0040	0.0083	0.0040	0.0083
43	3	S	W	143	0.062	0.0000	0.3926	0.0000	0.0866	0.0000	0.0866
44	4	S	W	0	0	0.0000	0.0021	0	0	0	0
45	1	D	W	-	-						
46	2	D	W	-	-						
47	3	D	W	2	0.001	0.0000	0.1361	0.0000	0.0855	0.0000	0.0859
48	4	D	W	0	0	0.0000	0.0014	0	0	0	0

Table 6: True probabilities (θ), extremes of the likelihood ridge estimates (min and max), extremes of the 95% trimmed interval (95% lv and 95% uv), gain of the interval (95% lv, 95% uv) with respect to the interval (min, max), average values in the ridge for ($\bar{\theta}$ S2) for the S2 case. All the values are computed over the 100,000 runs of EM

Cell	ETA	EDU	PRO	θ	Min	Max	95% lv	95% uv	Gain	$\bar{\theta}$
11	3	S	M	0.0614	0.0186	0.1290	0.0446	0.1174	0.3412	0.0850
12	4	S	M	0.0017	0.0024	0.0031	0.0024	0.0031	0.0729	0.0027
15	3	D	M	0.0951	0.0260	0.1364	0.0376	0.1104	0.3412	0.0700
16	4	D	M	0.0022	0.0013	0.0021	0.0014	0.0021	0.0729	0.0018
22	2	V	E	0.0004	0.0000	0.0043	0.0000	0.0037	0.1486	0.0017
23	3	V	E	0.0532	0.0015	0.0881	0.0413	0.0881	0.4599	0.0730
24	4	V	E	0	0	0	0	0		0
26	2	S	E	0.0035	0.0011	0.0054	0.0017	0.0054	0.1486	0.0037
27	3	S	E	0.2823	0.2279	0.3780	0.2345	0.3099	0.4974	0.2698
28	4	S	E	0.0013	0.0000	0.0007	0.0001	0.0007	0.0729	0.0004
31	3	D	E	0.0376	0.0000	0.0678	0.0097	0.0640	0.1987	0.0408
32	4	D	E	0	0.0000	0.0007	0.0000	0.0007	0.0729	0.0003
33	1	C	W	0.0065	0.0065	0.0065	0.0065	0.0065		0.0065
34	2	C	W	0.0117	0.0101	0.0101	0.0101	0.0101		0.0101
35	3	C	W	0.3281	0.3342	0.3342	0.3342	0.3342		0.3342
36	4	C	W	0.0052	0.0052	0.0052	0.0052	0.0052		0.0052
37	1	V	W	0	0	0	0	0		0
38	2	V	W	0.0030	0.0000	0.0043	0.0006	0.0043	0.1486	0.0026
39	3	V	W	0.0389	0.0000	0.0865	0.0000	0.0467	0.4599	0.0150
40	4	V	W	0	0	0	0	0		0
42	2	S	W	0.0052	0.0040	0.0083	0.0040	0.0076	0.1486	0.0057
43	3	S	W	0.0618	0.0000	0.0866	0.0062	0.0813	0.1331	0.0459
44	4	S	W	0	0	0	0	0		0
47	3	D	W	0.0009	0.0000	0.0859	0.0000	0.0590	0.3134	0.0257
48	4	D	W	0	0	0	0	0		0

Table 7: Comparison among some probability estimates and the corresponding true probabilities

	True	MLE (S0)	MLE (S1)
P(PRO = M, AGE = 4)	0.0039	0.0069	0.0044
P(EDU = S, AGE = 4)	0.0030	0.0021	0.0031
P(EDU = D, AGE = 4)	0.0022	0.0014	0.0021
P(EDU = C, AGE = 4)	0.0050	0.0069	0.0052

The unconstrained MLEs are

$$\hat{\theta}_{ij.} = \frac{6}{9} \times \frac{9+15}{2313}, \qquad \hat{\theta}_{i.k} = \frac{10}{15} \times \frac{9+15}{2313}, \qquad \hat{\theta}_{i..} = \frac{9+15}{2313}.$$

From the standard inequality

$$\max\{0, \theta_{ij.} + \theta_{i.k} - \theta_{i..}\} \leq \theta_{ijk} \leq \min\{\theta_{ij.}, \theta_{i.k}\}$$

it holds that

$$\hat{\theta}_{ijk} \geq \frac{6}{9} \times \frac{9+15}{2313} + \frac{10}{15} \times \frac{9+15}{2313} - \frac{9+15}{2313} > 0,$$

which is not compatible with the constraint so far introduced that (AGE ="4", EDU = "C", PRO ="M") is a structural zero. As a consequence, this structural zero restricts Θ to a set Ω where $\theta_{ij.}$ cannot be equal to its unconstrained maximum likelihood estimate $\hat{\theta}_{ij.}$. In fact, the constrained MLE of this parameter is moved towards the true value. In Tab. 7 it is shown the effect of all the imposed structural zeros (S1) to some parameters estimates. Note that the constrained MLEs for these marginal parameters are unique and may be obtained marginalising the corresponding estimates $\bar{\theta}_{ijk}$.

6 Conclusions

In the last years, the main goal of the statistical matching procedures may be reinterpreted as the efficient use and combination of all available and relevant information. However, in the context of statistical matching the available information is in terms of partial knowledge of the phenomenon (e.g. two independent samples on some marginal distributions). Therefore, we can just make conclusions under uncertainty. Whenever it is possible to use also auxiliary information, we can draw conclusions on the parameters with a lower degree of uncertainty. Extreme cases are those when particular models can be assumed, as the CIA, or external information like that in Singh *et al*. (1983). In these cases it is possible to provide a unique conclusion on the joint distribution.

When either the CIA or external auxiliary information are not usable, uncertainty on conclusions can be rarely avoided. We propose to analyse it throughout the descprition of all the plausible solutions, i.e. all those distributions coherent with the observed data according to the likelihood principle. In this context we have focused on some aspects of uncertainty and we have proposed some statistics in order to draw conclusions on the phenomenon at different levels, for instance either regarding single parameters of the distribution, or the entire distributions.

We also describe the situation when a particular auxiliary information is available: logical constraints. Logical constraints can be very useful in order to decrease uncertainty

on parameters value. The usage of logical constraints is not immediate, and an algorithm for introducing them in the statistical matching procedure has been proposed.

Finally we remark that all the analyses of uncertainty due to the partial knowledge of the phenomenon investigated is made with respect to the likelihood principle.

Since the concept of uncertainty and partial knowledge has been deeply investigated in other contexts than stastical matching, like for instance artificial intelligence, we feel it is very important to analyse the common aspects and the solutions proposed in these contexts. The contamination with other frameworks are not only worthy for the study of uncertainty, but also for the use of logical constraints, e.g. Coletti and Scozzafava (2002) and Vantaggi (2003). Further researches will be devoted to the inspection of these other scientific contexts.

References

Agresti A. (1990), *Categorical Data Analysis*, New York: John Wiley.

Barry, J. T. (1988), "An Investigation of Statistical Matching", *Journal of Applied Statistics*, 15, 275-283.

Bergsma, W. P., and Rudas, T. (2002), "Marginal Models for Categorical Data", *The Annals of Statistics*, 30, 140-159.

Coletti, G., and Scozzafava, R. (2002), *Probabilistic logic in a coherent setting*, Trends in logic n. 15, Dordrecht: Kluwer Academic Publishers.

Csizár, I. (1975), "I-divergence Geometry of Probability Distributions and Minimization Problems", *The Annals of Probability*, 3, 146-158.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society, B*, 39, 1-38.

Goel, P. K., and Ramalingam, T. (1989), *The Matching Methodology: Some Statistical Properties*, Lecture Notes in Statistics, New York: Springer Verlag.

Haberman, S. (1977), "Product Models for Frequency Tables Involving Indirect Observation", *Annals of Statistics*, 5, 1124-1147.

Ingram, D. D., O'Hare, J., Scheuren, F., and Turek, J. (2001), "Statistical Matching: a New Validation Case Study". *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 746-751.

Judge, G. G., Griffiths, W. E., Hill, R. C., and Lee, T. C. (1980) *The Theory and Practice of Econometrics*, New York: John Wiley.

Kadane, J. B. (1978), "Some Statistical Problems in Merging Data Files", in *Compendium of Tax Research*, Department of Treasury, U.S. Government Printing Office, Washington D.C., 159-179.

Kamakura, W. A., and Wedel, M. (1997), "Statistical Data Fusion", *Journal of Marketing Research*, 34, 485-498.

Kamakura, W. A., and Wedel, M. (2003), "List Augmentation with Model Based Multiple Imputation: a Case Study Using a Mixed-Outcome Factor Model", *Statistica Neerlandica*, 57, 46-57.

Little, R. J. A., and Rubin, D. B. (1983), "On Jointly Estimating Parameters and Missing Data by Maximising the Complete-Data Likelihood", *The American Statistician*, 37, 218-220.

Manski, C. F. (1995), *Identification Problems in the Social Sciences*, Cambridge, Massachusetts: Harvard University Press.

Moriarity, C., and Scheuren, F. (2001), "Statistical Matching: a Paradigm for Assessing the Uncertainty in the Procedure", *Journal of Official Statistics*, 17, 407-422.

Paass, G. (1986), "Statistical Match: Evaluation of Existing Procedures and Improvements by Using Additional Information", in *Microanalytic Simulation Models to Support Social and Financial Policy*, eds. G. H. Orcutt, and H. Quinke, Amsterdam: Elsevier Science, pp. 401-422.

Rässler, S. (2002), *Statistical Matching: a Frequentist Theory, Practical Applications and Alternative Bayesian Approaches*, Lecture Notes in Statistics, New York: Springer Verlag.

Renssen, R. H. (1998), "Use of Statistical Matching Techniques in Calibration Estimation", *Survey Methodology*, 24, 171-183.

Rodgers, W. L. (1984), "An Evaluation of Statistical Matching", *Journal of Business and Economic Statistics*, 2, 91-102.

Rubin, D. B. (1974), "Characterizing the Estimation of Parameters in Incomplete-Data Problems", *Journal of the American Statistical Association*, 69, 467-474.

Rubin, D. B. (1986), "Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations", *Journal of Business and Economic Statistics*, 4, 87-94.

Schafer, J. L. (1997), *Analysis of Incomplete Multivariate Data*, London: Chapman & Hall.

Singh, A. C., Mantel, H., Kinack, M., and Rowe, G. (1993), "Statistical Matching: Use of Auxiliary Information as an Alternative to the Conditional Independence Assumption", *Survey Methodology*, 19, 59-79.

Vantaggi, B. (2003), "Conditional Independence Structures and Graphical Models", *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, forthcoming (Vol. 11(5), October 2003).

Walley, P. (1991), *Statistical Reasoning with Imprecise Probabilities*, London: Chapman & Hall.

Winkler, W. E. (1993), "Improved Decision Rules in the Fellegi-Sunter Model of Record Linkage", *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 274-279.