Published in Statistical Journal of the United Nations ECE 16 (1999) pp 297-309.

SMALL AREA STATISTICS FROM SURVEY AND IMPUTED DATA.[1]

Svein Nordbotten[2]
P.O.Box 309 Paradis
N-5856 Bergen
Norway

Abstract: This paper discusses improvement of statistics for small areas and groups utilizing sample survey data and data imputed by intensive use of background data from available census or administrative register sources. Reliable accuracy prediction for such statistics is important and is also discussed and investigated in the paper.

Keywords: Small area statistics, accuracy prediction, aggregates, aggregate errors, individual data imputation, imputation errors.

1. Purpose

Moving into a new millennium, our societies are rapidly becoming more dependent on detailed information about the socio-economic state and development for small areas and/or groups. Frequently, demands for more detailed statistics cannot be served by traditional data collection and processing because of associated high expenses. Details requested from the 2000 Population Censuses are typical examples to which the national statistical offices will be able to provide a fraction of the statistics wanted.

It has for a long time been usual to combine censuses with small sample surveys, to obtain statistics, which were too expensive to obtain on a complete basis in the census itself. Unfortunately, traditional estimation methods will not always give reliable results for areas or groups if the samples from these are below a certain size. The purpose of this paper is to investigate if reliable statistics for such small areas/groups can be provided by individual imputation of survey data for all non-observed units based on an imputation model and available census and administrative data.

The approach proposed is to train neural networks on survey and census data from a sample to identify relationships between the survey and the census data. The trained networks, which can be considered as a set of non-linear regression functions, are subsequently used to impute survey data for individuals not in the sample to obtain a complete set of data, some observed and some imputed, for the whole population. To investigate this approach, an empirical experiment was executed on data from the 1990 Norwegian Population Census[3].

---

[1] This paper is results from research in SIS- Statistical Information Systems, a project carried out in cooperation with several statistical agencies. In formation about SIS is available on WWW at the http://nordbotten.com/sis/ .
[2] The author is Professor Emeritus at the University of Bergen, Norway, and can be contacted by email: svein@nordbotten.com.
[3] The author is very grateful to Statistics Norway for making the experiment possible

This report belongs to a series of studies investigating the possibilities for using neural networks in statistical information systems [Nordbotten 1995]. Others have used similar approaches based on neural networks to study related problems [Roddick 1993, Teague 1996, Steen Larsen and Madsen 1999].

2. Aggregates of imputed data and their errors

Consider a population of individuals identified by $u=1,...,N$ in an area (municipality). Two sets of variables are associated with each of the individuals, one set of J variables $y_1,.., y_J$, called *survey variables*, and a second set of I variables $x_1,..., x_I,$ called *census variables*. Assume that observations of the y-variable are limited to a random sample of individuals from the population. Our approach is to impute survey values for all the individuals not in the sample and aggregate imputed and survey values as estimates for the population targets

The imputation model we use is a so-called feedforward 2-layer neural network [Rumelhart and McClelland 1986]. It can be trained to represent a set of complex relations between individual survey and census variables. The network can be considered as a model of J non-linear regression functions between individual survey variables and census variables [Bishop 1995]:

$$y_j = 1/\,(1+\exp\text{-}[w_{0j}+\Sigma_l{}^L\, w_{lj}*(1/(1+\exp\text{-}[\, w_{0l}+\ \Sigma_i{}^I\, w_{bl}*x_{ib}\,])])+r_j \qquad \text{for } \ j=1,..,J$$

The w-s are parameters characterizing the relationships while the r-s are residuals representing the effect on y_j by non-specified variables not included among the census variables. The first layer in the network is equivalent to a set of L latent variables.

Numeric values of the w-parameters are estimated from survey and census data in a sample of n individuals by means of an iterative training algorithm called backpropagation. Imputations y_j' can then be calculated by means of the trained model for any individual for which we know the census variables [Nordbotten 1996].

An individual imputation of a survey variable is related to its target value by

$$y_j\,' = y_j\ + e_j \qquad \text{for } j=1,...,\ J,$$

The e_j are *imputation errors* defined as deviations between imputed values and target values. They are assumed to be independent random variables. Their means and variances can also be estimated from the sample. Experience, however, indicates that the results easily become biased [Moody 1993]. We have therefore chosen to extend the available survey to $n+m$ units and divide the sample randomly in one subsample of n individuals for estimating the regression functions, and a second subsample of m individuals for estimating the means and variances of the residual variables. By this strategy, we may loose accuracy in estimating the w-parameters, but gain accuracy in the calculated means and variances of the residual variables.

We index the units of the population in the first sample as $u=1,\dots,n$, those in the second sample as $u=(n+1),\dots,\ (n+m)$ and the units of the population not surveyed as $u=(n+m+1),\ \dots,\ N$. Without loss of generality, we can disregard the index j for different target variables, and define an *imputation aggregate*

$$Y' = \Sigma_l{}^n\, y_u\ + \Sigma_{(n+1)}{}^{(n+m)}\, y_u\ +\ \Sigma_{(n+m+1)}{}^N\, y_u\,'$$

as an estimate of the total Y in the population P[4]. This aggregation can be rewritten as

$$Y' = \Sigma_1^N y_u + \Sigma_{(n+m+1)}^N e_u$$

The estimators used for the mean of the error e_u are

$$m = \Sigma_{(n+1)}^m \; e_u / m$$

and for the mean square error

$$mse = \Sigma_{(n+1)}^m (e_u - m_z)^2/(m-1).$$

The *aggregation error E* associated with the aggregate Y' can now be defined by

$$E = \sqrt{(N-n-m)*mse}$$

It should be noted that aggregation error is a measure of the uncertainty, which is caused by the random errors e_u while the sampling error of a traditional estimate is a measure of the uncertainty caused by the sampling process.

If N is large enough, the distribution of Y' will be approximately normal according to the Central Limit Theorem. With a probability 0.75, we can then predict that the deviation $|Y'-Y|$ will not be larger than for example 1.15* E, i.e.

$$pr(|Y'-Y| <= 1.15*E) = 0.75.$$

The aggregation error can also be used for deciding if an aggregate Y' satisfy a given accuracy requirement or not. An example can be that a user of Y'-aggregates has an application requiring that aggregates should not in general deviate with more than +/- 5 individuals from the target total. If the user can take the risk that 25% of the aggregates will be incorrectly rejected, then aggregates should be accept if their aggregation errors $E<=5/1.15 = 4.3$.

A second example can be a producer of statistics who wants to restrict publication to aggregates with a relative deviation $|Y'-Y|/Y'<=0.3$. Assume that the producer also wants to keep the risk of suppressing incorrectly an acceptable aggregate to 0.25. The producer should then suppress from publication all aggregates which do not satisfy the condition $1.15*E/Y<=0.3$. Since the target total Y is unknown, it must in this condition be substituted by the aggregate Y', i.e. aggregates with an error $E>0.26*Y'$ should not be released for publication.

Based on the discussion above, we shall investigate and discuss three propositions:

1. the imputation aggregator Y' can provide useful statistics for populations in areas/groups based on a set of real data from a small random sample and imputed data for each individual not in the sample. Survey and census data are used for estimating the imputation functions.
2. aggregation error E for an aggregate can be computed from a second, small survey sample to provide a reliable accuracy prediction for imputation aggregate.

[4] The terms *aggregator*, *aggregate* and *aggregation error* are used to avoid any confusion with the traditional concepts estimator, estimator and sampling error.

3. the trained imputation models represent a persistent structure and can be applied for imputations also in populations not represented in the survey sample.

3. Empirical investigation

3.1 Two experiments

Two experiments were performed on Norwegian population data from 1990 Population Census. Data for two municipalities, Municipality I with 17,326 individuals distributed to 56 census tracts, and Municipality II with 10,102 individuals distributed to 44 census tracts, were used in for the experiments. We shall focus our attention to imputation aggregates for these 90 small areas.

The two selected municipalities differ in several respects. The first municipality is located in the middle part of the country near a city with a mix of farming, manufacturing and transport as its main industries. The second municipality is located in the northern part of the country. Fisheries and fish processing are its main industries. The average census tract size of the Municipality I was 310 inhabitants while the average size of the tracts in Municipality II was 230 inhabitants in 1990.

For most municipalities in Norway, survey observations were collected in connection with the 1990 Population Census for a random sample of the inhabitants in addition to census data available for each individual. The two municipalities used in the study required, however, statistics based on complete counts also for the survey variables and paid Statistics Norway for the additional observations themselves.

In the experiments, we simulated the situation that a random sample survey of 2,007 individuals was taken also in the first municipality and that no survey observations were made at all in the second. The average sample size for each tract in Municipality I was 35 individuals, but because of the skew distribution a large proportion of the tracts had was represented with 20 or less individuals in the sample. Because complete survey data existed for both populations, the data represented an excellent basis for testing of the propositions made above. Imputation aggregates and accuracy predictions for the aggregates could be made from sample and compared with real target totals and evaluated.

3.2 Imputations and their mean square errors

In total 15 variables were assumed observed in the sample survey and 97 x-variables were available from census. The 15 y-variables are listed in Table 1. Two models with altogether 15 simultaneous imputation functions were used to impute 15 variable values for each inhabitant not included in the sample.

Model 1 included 9 imputation functions and provided individual imputation values for y-variables concerning *Cohabitation* for each individual, while Model 2 provided imputation values for the remaining 6 y-variables concerning *Means of transportation* to work. Both models used the individual values of the 97 x-variables census variables as independent variables. In addition, both models included 25 latent or hidden variables. The design and construction of the neural network models used for imputations and the computation of the mean square errors have been described in detail in papers referred to above.

In the first experiment, two random and mutually exclusive samples were drawn from the population in Municipality I. We assumed that the survey of *y*-variables was carried out in both samples. The first sample counted 1845 individuals and was used to estimate 5240 parameters/weights in the two imputation models. The second sample comprised only 165 individuals and was used to estimate the mean square errors and means of the imputed variables. Together these samples had approximately the relative size used in the 1990 Census for the majority of municipalities.

Table 1: Variables, estimated mean square errors and biases.

MODEL/VARIABLES	*MEAN SQUARE ERRORS*	*MEANS*
Model 1:		
1. Alone	0.062	0.002
2. Spouse	0.027	-0.024
3. Cohabitant	0.038	-0.018
4. Daughter/son	0.073	0.073
5. Mother/father	0.035	-0.002
6. Siblings	0.049	-0.020
7. Inlays	0.003	0.009
8. Grandpar./grandchildren	0.004	0.011
9. Other	0.030	-0.011
Model 2:		
1. Car	0.100	-0.092
2. Bus	0.009	0.016
3. Train	0.008	-0.000
4. Boat	0.000	-0.001
5. Bicycle	0.057	0.053
6. Other	0.002	0.007

According to the imputation aggregator the sum of the observed *y*-values for each of the 2,007 inhabitants in the two samples were used as the first component of the aggregates. The sum of the imputed values for the remaining 15,319 individuals was added as the second component to obtain the final imputation aggregates of the totals in Municipality I.

Means and mean square errors for the 15 prediction variables were computed from second sample and are also reported in Table 1.

3.3 Imputation aggregates and their predicted accuracy

Altogether 840 Y'-totals were aggregated and the corresponding real target Y totals computed for the 56 areas in Municipality I. For each aggregate, the aggregation error E was computed. To test the validity of using the E-s for predicting the accuracy of the imputation aggregates, all aggregates were classified as rejectable if $E>4.3$, if not they were classified as acceptable. Because the real totals were available in the experiment, the imputation aggregates were also classified by their real deviation from the targets. The results are given in Table 2 and 3 for the *Cohabitation* and *Means of transportation* totals, respectively.

Table 2: Predicted and real accuracy.
Cohabitation totals. Municipality I.

<table>
<tr><td></td><td></td><td colspan="3">Real |Y'Y|:</td></tr>
<tr><td></td><td></td><td><=5</td><td>>5</td><td>Sum</td></tr>
<tr><td>Predicted</td><td><=5</td><td>363</td><td>67</td><td>430</td></tr>
<tr><td>|Y'-Y|</td><td>>5</td><td>51</td><td>23</td><td>74</td></tr>
<tr><td></td><td>Sum</td><td>414</td><td>90</td><td>504</td></tr>
</table>

Table 1 shows that 414 *Cohabitation* imputation aggregates were acceptable and 90 were rejected indicating a high aggregate accuracy. Out of the 414 aggregates, 363 were correctly accepted while 51 were incorrectly classified as not acceptable. On the other hand, 67 aggregates were incorrectly accepted. Finally, the accuracy predictions correctly rejected 23 aggregates.

Table 3: Predicted and real accuracy.
Means of transportation totals. Municipality I.

<table>
<tr><td></td><td></td><td colspan="3">Real |Y'Y|:</td></tr>
<tr><td></td><td></td><td><=5</td><td>>5</td><td>Sum</td></tr>
<tr><td>Predicted</td><td><=5</td><td>235</td><td>48</td><td>283</td></tr>
<tr><td>|Y'-Y|</td><td>>5</td><td>9</td><td>44</td><td>53</td></tr>
<tr><td></td><td>Sum</td><td>244</td><td>92</td><td>336</td></tr>
</table>

For the 336 *Means of transportation* totals 244 deviated from their respective targets with 5 or less individuals. The accuracy predictions were correct for 235 out of the 244 acceptable aggregates. The accuracy predictions incorrectly accepted 48 aggregates, which were deviating with more than 5 individuals from their target totals.

For release of survey statistics, Statistics Norway used a publication rule for the 1990 Census results stating that only estimates with coefficients of variation less than 0.3 should be published in printed tables. In the present study a similar requirement has been tested. We required that imputation aggregates to be accepted must satisfy $|Y'-Y|/Y'$=< 0.3 with a confidence probability 0.75. Accordingly, aggregates with an aggregation error E<=0.26*Y' were identified and predicted acceptable. The accepted aggregates were then confronted with the real deviations and the results presented in Tables 4 and 5.

Table 4 shows that a slightly higher number of aggregates, 421 aggregates, satisfied this acceptance condition than the absolute criterion on which Table 2 was based on. As to the accuracy predictions, the figures indicate that the reliability is about the same level as above.

Table 4: Predicted and real relative accuracy.
Cohabitation. Municipality I.

		Real $\lvert Y'-\rvert/Y'$		
		<=0.3	*>0.3*	*Sum*
Predicted	*<=0.3*	365	47	412
$\lvert Y'-Y\rvert/Y'$	*>0.3*	56	36	92
	Sum	421	83	504

The aggregates for the *Means of transportation* satisfied the relative condition at about the same level as the absolute with 236 aggregates identified as acceptable. However, only 198 of these were predicted acceptable. The predictions also incorrectly accepted 79 aggregates, which should have been rejected as shown in Table 5.

Table 5: Predicted and real relative accuracy.
Means of transportation. Municipality I.

		Real $\lvert Y'-\rvert/Y'$		
		<=0.3	*>0.3*	*Sum*
Predicted	*<=0.3*	198	79	277
$\lvert Y'-Y\rvert/Y'$	*>0.3*	38	21	59
	Sum	236	100	336

In the second experiment, Municipality II was assumed not be surveyed at all, but census data-were available to the experiment for all inhabitants as a benchmark. The imputation models from the first municipality were used to impute survey variable values for all inhabitants in Municipality II. The individual imputations were aggregated to 660 imputation aggregates for the 44 census tracts of the municipality.

Table 6: Predicted and real accuracy.
Cohabitation totals. Municipality II.

		Real $\lvert Y'-Y\rvert$:		
		<=5	>5	Sum
Predicted	*<=5*	285	30	315
$\lvert Y'-Y\rvert$	*>5*	44	37	81
	Sum	329	67	396

Table 6 shows that 396 *Cohabitation* aggregates were computed. Out of these 329 aggregates had an acceptable level of accuracy. The accuracy prediction failed to accept 44 aggregates, which in fact were correct. On the other hand 30 aggregates were incorrectly accepted,

Table 7: Predicted and real accuracy.
Means of transportation totals. Municipality II.

		Real |Y'-Y|:		
		<=5	*>5*	*Sum*
Predicted	*<=5*	220	21	241
|Y'-Y|	*>5*	12	11	23
	Sum	232	32	264

The 232 aggregates of Table 7 for *Means of transportation* out of 264 were within the 5 or less individuals. The accuracy predictions classified 220 of these correctly. The table also shows that 21 were incorrectly predicted as acceptable.

Table 8: Predicted and real relative accuracy.
Cohabitation. Municipality II.

		Real |Y'-Y|/Y'		
		<=0.3	*>0.3*	*Sum*
Predicted	*<=0.3*	181	15	196
|Y'-Y|/Y'	*>0.3*	85	115	200
	Sum	266	130	396

Tables 8 and 9 display the classification of aggregates by predicted and real deviations according to the relative classification condition. These tables give about the same picture of aggregates for Municipality II as the two previous tables.

Table 9: Predicted and real relative accuracy.
Means of transportation. Municipality II.

		Real |Y'-Y|/Y'		
		<=0.3	*>0.3*	*Sum*
Predicted	*<=0.3*	78	54	132
|Y'-Y|/Y'	*>0.3*	49	83	132
	Sum	127	137	264

4. Discussion

4.1 Sizes of the census tracts

The majority of census tracts in Municipality I have from 100 to 300 inhabitants with an average of 310. On the other end, a few tracts have more than 1000 inhabitants. Because of the skew distribution, a large part of the tracts would be represented in the survey with 10-20 individuals subject to the sampling assumption we made. For these tracts, traditional estimators could not be expected to provide useful statistics.

The tracts in Municipality II have even smaller populations than tracts in the first municipality. Most census tracts of Municipality II have an average of 230 inhabitants. Out of the 44 tract have 32 less than 200 inhabitants and 13 less than 100. In our experiments, no sample survey was assumed at all for this municipality and traditional estimates could therefore not be computed at all.

4.2 Target totals and imputation aggregates

A large number of imputation aggregates were computed for each municipality, 840 for tracts in Municipality I and 660 for tracts in Municipality II. As pointed out above, real data were available for all totals and used as targets. The differences between the aggregates and the corresponding target totals were studied in detail.

For *Cohabitation* totals in Municipality I, the largest deviation between an aggregate and the corresponding target total was for the number of individuals living alone in a tract with a total population of 1,046 persons and for which Cohabitation was imputed for 926 individuals. The target total for people living alone was 248, which by the prediction aggregate, was underestimated with 26 individuals.

For *Means of transportation*, larger differences were identified. Aggregates for the number of people, who reported to use bicycle as a means of transportation to work demonstrated for example great deviations from their targets in some areas. In one area, the target total of inhabitants who used bicycles as their means of transportation to work was 165. This total was underestimated with 66 inhabitants. The explanation is most likely that the topology permitted few other means of transportation than bicycle in this tract, a fact, which was not reflected well in the imputation model.

The absolute deviations between aggregated and target totals are less than 5 individuals for nearly 82 per cent of the 504 aggregates for *Cohabitation* totals. For the *Means of transportation*, the deviations of 5 or fewer individuals count for 73 per cent of 336 aggregated totals. Keeping the sizes of the areas in mind, we can conclude that the imputation aggregates in general have high quality.

4.3 Real deviations and accuracy prediction reliability

Both producers and users of statistics wish to identify which belong to the 82 per cent, respectively 73 percent, high quality aggregates. The producers want a tool for providing quality declarations for the statistics while the users need a tool for evaluating the usability of the aggregates for their particular applications. We have used the aggregate error as an accuracy indicator for predictions, and it is crucial to determine how reliable this measure is.

Comparisons of accuracy predictions and real deviations between aggregates and targets for *Cohabitation totals* in Municipality I was summarized in Table 2. The accuracy predictions were based on the condition $E<=4.3$ which corresponds to the probability 0.75 that the aggregate deviated with 5 or less individuals from its target total. Of the 504 *Cohabitation* aggregates, 430 were predicted to satisfy this requirement. Comparisons with the real deviations show that the accuracy of 363 or 72% of the aggregates was correctly predicted.

The two types of errors known from statistical theory of testing can be used for the discussion of accuracy prediction. From the assumptions made, we would expect that the number of Type I errors, rejecting incorrectly aggregates that satisfy the requirement, would be less than 126 or 25% of the aggregates. Table 2 indicates that 51, or only 10%, were incorrectly classified not to satisfy the condition. Type II errors, accepting incorrectly aggregates that do not satisfy the condition, were 67.

If an application of aggregate4 is such that Type II errors are relatively serious or expensive, a critical limit for E less than 4.3 will reduce the number of these errors, but at the same time also increase the number of Type I errors. A balance for the relative importance of the two types of errors must be found and used to adjust the classification criterion.

For the *Means of transport* totals, Table 3 shows that 283 aggregates were predicted to deviate from the targets with 5 or less individuals. The Type II errors committed were 48 aggregates incorrectly predicted to meet the condition. The Type I errors made in predictions, were only 9 or less than 3% of the 336 aggregates. These figures indicate that accuracy predictions based on aggregate errors should be reliable for a number of applications.

Tables 4 and 5 give results from the evaluation of accuracy classification when the criterion was based on the relative deviation, i.e. $|Y'-Y|/Y' <= 0.3$, similar to the publication requirement of Statistics Norway. The number of aggregates satisfying this condition was 421 for *Cohabitation* and 236 *Means of transportation* aggregates, which is about the same general level as for the case of absolute deviations discussed above. We observe the same patterns concerning Type I and Type II errors as above. Both for *Cohabitation* as well as for *Means of transportation*, the Type I errors were about 10-11% while Type II errors were made in 24% of the *Means of transportation* accuracy classifications.

4.4 Imputation aggregates and accuracy prediction reliability for areas not sampled

The second experiment assumed that no survey data were collected in Municipality II. No imputation functions could therefore be estimated in this municipality. It was assumed that the relationships between survey and census data were shared by the municipalities and the imputations models derived from data for Municipality I was therefore used to impute all y-variable values for each individual in Municipality II. These assumptions were rather challenging.

The imputation aggregates computed for Municipality II included therefore only imputed individual values. Likewise the imputation means and mean square errors from Municipality I were used to compute the aggregation errors of the imputation aggregates and predict the accuracy of the aggregates for Municipality II. Since all survey data were available for the experiment, the accuracy of the imputation aggregates and the reliability of predictions for their quality could be studied and evaluated in detail also for this municipality.

The deviations of imputation aggregates from the real survey data total were similar in Municipality II to those for Municipality I. For the *Cohabitation* aggregates, 329, or 83% of the 396 aggregates deviated with 5 or less individuals. The distribution of aggregate deviations for *Means of transportation* totals shows even relatively higher accuracy than the corresponding distribution in Municipality I with 232, aggregates deviating with 5 or less individuals, or 87% of

264 aggregates for the municipality. Type 1 and Type II accuracy prediction classification errors are equal or smaller than the corresponding aggregate errors of Municipality I, even though the same mean square errors were used.

Taking into account that these aggregates included no observed survey data at all and that they were completely based on individually imputed values, the real accuracy and the prediction accuracy are remarkable. It is even more remarkable taking into account that the imputation functions used and the mean square errors were estimated on data from a very different municipality.

When we study the figures of Table 9 which gives classification results based on relative accuracy conditions they are as good as the results in Table 5 for aggregates in Municipality I.

5. Summary and conclusion

- The experiments support the hypothesis that reliable population statistics based on sample surveys and censuses can be derived for areas with populations too small for traditional estimation. Imputation aggregates can be computed by adding survey values for individuals in the sample and imputed values for individuals not in the sample.
- Imputation models can be trained to identify functions between individual survey and census data. Trained models can subsequently be used for imputing individual survey variable values for non-sampled individuals using census data as arguments in the model functions.
- Accuracy predictions for these aggregates can be calculated using estimated mean square errors for residuals between imputed and target values obtained from another independent sample. Comparing accuracy predictions and real deviations between aggregates and target totals indicate that the accuracy predictions can be used to distinguish accurate aggregates from inaccurate.
- Imputation models developed by data from one municipality seem to represent persistent structures for individuals also in other municipalities. If so, the models can be used for producing imputation aggregates also for small areas not covered at all by the survey sample.

The imputation aggregates computed for small areas in the experiments were surprisingly accurate. If further research supports the above findings, the approach may have a potential for providing more statistical information for areas/groups with small populations.

6. References

Bishop, C, M. (1995): Neural Networks for Pattern Recognition. Clarendon Press. Oxford.

Moody, J.E. (1993): Prediction risk and architecture selection for neural networks. In Charkassy, V., Friedman, J.H. and Wechsler, H. (eds.): From Statistics to Neural Networks. Theory and Pattern Recognition Application. Springer. Berlin.

Nordbotten, S. (1995): Editing and imputation by means of neural networks. Statistical Journal of the United Nations ECE, IOS Press. Amsterdam, pp. 110-129.

Nordbotten, S. (1996): Neural Network Imputation Applied on Norwegian 1990 Population Census Data. Journal of Official Statistics. Statistics Sweden. Stockholm.

Nordbotten, S. (1998): Estimating Population Proportions from Imputed Data Computational Statistics & Data Analysis, Vol. 27, pp. 291-309.

Roddick, H. (1993): Data Editing Using Neural Networks. Statistics Canada. Ottawa.

Rumehart D.E. and McClelland, J.L. (1986): Parallel Distributed Processing. The MIT Press. Cambridge.

Steen Larsen, B. and Madsen, B. (1999): Error Identification and Imputation with Neural Networks. UN/ECE Work Session on Statistical Data Editing in Rome. ECE. Geneva.

Teague, A. and Thomas, J. (1996): Neural Networks as a Possible Means for Imputing Missing Census Data in the 2001 British Census of Population. Survey and Statistical Computing 1996. Association for Survey Computing.