Estimating Labour Market Flows by the ADETON Method

Uwe Blien[1], Friedrich Graef[2]

Abstract: A problem encountered in many contexts is the estimation of tables from heterogeneous, incomplete and contradictory information. In this paper the ADETON method is presented which can be used for this estimation task under quite general conditions.

The ADETON method was originally developed to estimate flow matrices of regional labour markets in the Multi-Accounting System (MAS). Purpose of the MAS is to give a comprehensive survey of the stocks and flows of people in education and training, in employment, unemployment, and household activities. The flows between the stocks are represented by transition matrices. These data are used for theory-based analyses concerning the relationship between the labour market and adjacent statuses. The results are important for a rational design of regional labour market policy.

To estimate the transition matrices within MAS a new method based on the optimisation of the Chi-square function was developed. With the ADETON algorithm tables are computed which are most similar to given reference tables and satisfy a prescribed set of linear inequality and equality constraints. In addition it is possible to specify "soft constraints" which are to be observed up to a specified accuracy. The results of ADETON are approximately equal to those obtained by entropy optimization. In this respect the method is a generalisation of the well known Iterative Proportional Fitting Algorithm (used in log-linear models) or of the equivalent RAS method (used in input-output analysis). Since entropy optimization is a general estimation principle like maximum likelihood, the matrix estimated is under specified conditions "the most probable one".

ADETON is applicable for many purposes, e.g. to estimate contingency tables or input-output and other flow matrices.

1. Introduction

Flow analyses are important to understand what's going on in the labour market. For this purpose the Institute for Employment Research developed two mighty tools, the Educational Accounting System and the Labour Market Accounting System which shows stocks and flows on the labour market in its entirety. To extend these tools and to cover regional labour markets the Multi-Accounting System (MAS) was developed. Its purpose is to give a comprehensive survey of the stocks (states) and flows of people in education

[1] Institute for Employment Research, Regensburger Str. 104, 90327 Nürnberg, Germany
[2] Institute of Applied Mathematics, University of Erlangen-Nürnberg, Martensstr. 3, 91058 Erlangen, Germany

and training, in employment, unemployment, and household activities. The flows between the stocks are represented by transition matrices. These data are used for theory-based analyses concerning the relationship between the labour market and adjacent statuses. The results are important for a rational design of regional labour market policy.

In this paper a short overview of the construction principles of the Multi-Accounting System (MAS) is given and some selected results are shown. The main focus, however, is the so-called ADETON procedure which is used to estimate the transition matrices. It can be used for many purposes to estimate matrices (or tables) from heterogeneous, incomplete and contradictory information. With the ADETON algorithm tables are calculated which are most similar to given reference tables and satisfy a prescribed set of linear inequality and equality constraints (e.g. row or column sums). In addition it is possible to specify "soft constraints" which are to be observed up to a specified accuracy. If no reference matrix is available the target matrix can be obtained by using constraints only.

The results of ADETON are obtained either by applying directly the principle of entropy optimization or by minimizing a Chi-squared function that gives results approximately equivalent to entropy optimization. Since entropy optimization is a general estimation principle like maximum likelihood, the matrix estimated is "the most probable one" under specified conditions. The ADETON method is a generalization of the well known Iterative Proportional Fitting Algorithm (IPF, used in log-linear models) or of the equivalent RAS method (used in input-output analysis). Whereas with the IPF/RAS-method only equations could be used to prescribe a target matrix, in the case of ADETON also inequalities may be included.

With ADETON, the constraints can be of a soft character (similar to a "fuzzy" type) which means that they are not required to be exactly met. The target matrix should only show a structure which is similar to the one specified by the soft constraint. The degree of accuracy to which the constraint is met can be controlled by a set of weights. Therefore it is possible to include information about sampling errors or about standard errors of multivariate models. In addition – if required – rather weak expert judgements could be included in the estimation process, all with the appropriate weights.

Because of the great flexibility of the method it is possible to estimate matrices based on heterogeneous pieces of information of very different kind. Consider the case of a flow matrix in the Multi-Accounting System. It may be the case that parts of the transition matrix are known for a different year. This piece of information can be used as a hint for the estimation of a target matrix. In addition some of the row and the column sums might be known from official statistics. For them equality constraints could be specified that are required to be exactly met. In addition there might be information about flows in the target matrix coming from different survey samples. Because there is a sampling error there could be contradictions between different surveys' results. In these cases soft constraints could be specified. Then there might be relatively vague expert judgements formulating the expectation that the aggregate of some flows is between so and so. Again this could be included into the estimation procedure as a soft constraint with a low weight. Finally there might be flows that are not admissible because they are ruled out by institutional provisions. This can be modelled by a hard constraint.

To summarize, the ADETON method can be applied to many problems for other purposes outside the MAS, since it works under very general conditions. The method may be applied wherever information about the structure of the matrix to be estimated is given by a set of linear equations and/or inequalities with respect to the matrix entries. In addition, an a priori known reference matrix can be specified. The result, i.e. the structure of the estimated table, is as close as possible to this reference matrix. Typical tasks within the range of applicability of ADETON are:

- the disaggregation of data solely available in a summarized form
- the estimation of tables from heterogeneous, incompatible and incomplete data
- the calculating of transition matrices in Markov processes
- the construction of forecasts (e. g. when the joint distribution of some variables has to be estimated after computing univariate distributions in a first step, see Blien, Tassinopoulos 2001).

The ADETON procedure is designed to calculate matrices in a way which makes optimal use of the given information and avoids biases. The result can be interpreted according to criteria from statistics and from the formal information theory (Shannon, Weaver 1949, Kullback 1968, Golan, Judge, Miller 1996).

The remainder of the paper includes sections about the Multi Accounting System, the description of the ADETON method and about some results by applying the method in the MAS.

This paper is a complement to another one written by two of our colleagues (Haas, Rothe 2005b) at the Institute for Employment Research, who worked with us in the project group responsible for the development of the Multi-Accounting System. They concentrate on the MAS itself and on the results to be obtained with it. Our special concern is the ADETON method applied to generate these results.

2. The Multi Accounting System (MAS)

In the MAS the states of people located in a regional labour market are shown for defined time points, the beginning and the ending of a year. There is a number of these states which are exhaustive and exclusive: One person can have only one state. The states of the beginning of a year and those of the ending of a year are connected by transition flows (see also Table 1). These include the stayers in the main diagonal of the table who do not change their states. The states are also called “accounts” or “stocks”.

In principle the movements (flows) between individual stocks, can be analysed in two different ways: This can be done either in a *personalized approach* or in a *case-based approach*. In the *personalized approach* the state of a person at time points t_0 and t_{0+1} are compared and a transition is defined for this person. “That leads to an underestimation of flows in any case were multiple transitions have occurred, especially if a person has returned to the same labour force status he/she had the year before" (Kruppe 2001). Furthermore, the personalized approach does not allow the identification and thus the analysis of active labour market policy measures taken during the year. This means that the objective of the MAS of being a supportive analytical instrument for the development of labour market policies would be missed.

The *case-based approach* includes all transitions between the accounts occurring during the observation period (between t_0 and t_{0+1}). If necessary several changes between the defined statuses are registered since the flows are counted completely. Therefore the total flows could exceed the sum of all flows plus the stayers. The *case-based* variant is selected for the MAS, at the same time providing the stayers and the stock data at the beginning and at the end of the year.

In order to display the transitions of people in the labour market a two-dimensional matrix is suitable to represent the different states of the MAS. The source accounts are shown in the rows for time point t_0 and the target accounts in the columns, for time point t_{0+1}. The matrix entries x_{ij} quantify the transitions between the states. The main diagonal displays the number of people staying in a certain state ("stayers").

Table 1: Structure of the MAS for the labour market of a specific employment agency

	State at t_{0+1} (target account)				Σ
State at t_0 (source account)	x_{11}	x_{12}	.....	x_{1K}	b^r_1
	x_{21}	x_{22}	.....	x_{2K}	b^r_2
	x_{31}	x_{32}	.....	x_{3K}	b^r_3
		…			…
	x_{I1}	x_{I2}	.....	x_{IK}	b^r_I
Σ	b^c_1	b^c_2	.....	b^c_K	b^{rc}

In Germany there are 180 employment agencies, all with a specific area defining a regional labour market. In principle it is intended to calculate a MAS for every of these 180 units. At the moment the MAS is in an experimental stage. It is only available for a few agencies.

Data about the stocks are available with a high degree of differentiation from official statistics. Concerning the flows the situation is worse. Part of the transition matrix is also available, but many flows are known only from survey samples. In many cases the information is not available on the level of relatively small regions. A transition might be known for West Germany as a whole but not for a specific employment agency area. Therefore an estimation mention must be used for many of the flows. This estimation method should use all the available information from many different sources and collected by the project group with the task of developing the MAS.

The method used in the end was especially developed for the purpose of the MAS but could be also applied to solve many problems which are formally related.

3. The ADETON Method

ADETON is a method and a related software tool for the estimation of matrices under restrictions as described above. In this section the rationale behind this program is explained.

3.1 Matrix Estimation

In purely mathematical terms the table or matrix estimation problem as it occurs with the MAS is to recover the unknown components x_{ik} of a matrix

$$x = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1K} \\ x_{21} & x_{22} & \dots & x_{2K} \\ . & . & \dots & . \\ x_{I1} & x_{I2} & \dots & x_{IK} \end{pmatrix} \tag{1}$$

— or x = (x_{ik}) for short — from information about aggregate totals of the matrix such as row sums, column sums, weighted sums over a subset of matrix cells, differences of such sums etc. These aggregates are special cases of weighted sums

$$L(x) = \sum_{i,k} a_{ik} x_{ik} \tag{2}$$

of the matrix components with fixed known weights a_{ik}. In what follows they will be called **matrix aggregates**.

It is assumed that the unknown matrix satisfies a certain number of **restrictions** of the form

$$L_j(x) = b_j \qquad \text{(equality restriction)} \tag{3}$$

$$L_j(x) \le b_j \ \text{ or } \ L_j(x) \ge b_j \qquad \text{(inequality restriction)} \tag{4}$$

$$b_j' \le L_j(x) \le b_j'' \qquad \text{(bandwidth restriction)} \tag{5}$$

for j = 1,2,…,J where the $L_j(x)$ are matrix aggregates and b_j, b_j', b_j'' are known real numbers which will be called constraint values or **constraints** for short. In the applications described above natural restrictions are the nonnegativity $x_{ik} \ge 0$ of the components of the matrix. They may be thought of being contained in the above restriction list as inequality restrictions of type (4).

In general the matrix is not fully specified by these restrictions. The set A of **admissible matrices** satisfying the given restrictions then contains an infinity of matrices. In this case the "true" matrix cannot be recovered. It is only possible to compute an *estimate* of the true matrix according to a suitable statistical estimation principle. The classical approach is to assume some kind of *structural similarity* of the unknown matrix to a known reference matrix and to compute an estimate by minimizing a corresponding **structural distance measure** on the set of admissible matrices. This will be explained in the following section.

An additional problem arises when some of the constraints do not represent exactly bounds to the aggregates of the true matrix but only bounds that are distorted by sampling errors or some other kind of "noise". It frequently occurs in this situation that restrictions are contradictory or inconsistent such that no matrix exists which satisfies all given restrictions, i.e. the set of admissible matrices is empty. Restrictions of this type should not be tried to be met exactly. Rather the distance between matrix aggregate $L_j(x)$ and the distorted constraint b_j should be made as small as possible by minimizing a **restriction distance measure** which will be defined below. For obvious reasons in the sequel this kind of restriction resp. constraint will be called **soft restriction** resp. **soft constraint**.

In presence of soft restrictions there are two objectives to follow with matrix estimation: minimizing structural distance *and* minimizing restriction distance. But in general these two objectives are contradictory to each other. Diminishing structural distance enlarges restriction distance and vice versa. Therefore the matrix estimation problem under soft restrictions is a **bicriterial optimization problem** (Ehrgott 2005) which has no unique solution. Mathematics can only provide the set of so called *Pareto optimal* or *efficient* solutions. The researcher then has to decide which of these solutions meets his expectations best. ADETON is a software tool for the computation of Pareto optimal solutions for the matrix estimation problem. It will be described below in some detail.

3.2 Structural Distance Measures

If there is more than one admissible matrix additional information about the "true" matrix has to be sought in order to obtain a unique estimate. In case of transition matrices there often exists a known transition matrix $u = (u_{ik})$ for the same sectors be it from another time period, from a comparable region or from a sample survey and of which one can argue that it is comparable to the unknown $x = (x_{ik})$ not in absolute size but with respect to the relative sizes of the components. That means that the quotients x_{ik}/x_{rs} and u_{ik}/u_{rs} are more or less equal. We will call this kind of equivalence of two matrices **structural similarity**.

Given such a **reference matrix** u it makes sense to compute an estimate by minimizing a **structural distance measure $D_T(x,u)$** on the set of admissible matrices which has the property that its unconstrained minima are obtained by matrices x for which the above mentioned quotients coincide for all its components. The most widely used structural distance measures[3] are variants of the relative entropy

$$E(x,u) = \sum_{i,k} x_{ik} \ln\left(\frac{x_{ik}}{u_{ik}}\right) \tag{6}$$

and the Chi-square distance

$$C(x,u) = \sum_{i,k} \frac{(x_{ik} - u_{ik})^2}{u_{ik}} \tag{7}$$

[3] A general class of distance measures is discussed by Deville, Saerndal, Sautory (1993)

The choice between these objective functions is of more or less philosophical nature. Solutions obtained by (6) may be justified as following the Minimum Information Principle, while those from (7) may be interpreted statistically as Maximum Likelihood estimators (Shannon, Weaver 1949, Kullback 1968, Golan, Judge, Miller 1996). Experience shows that with a large number of well behaved restrictions the differences between the numerical results obtained from these objective functions are neglectable. Mathematically, resp. numerically, the Chi-square distance is easier to handle.

The case that there is no additional information about the matrix to be estimated may be incorporated into the above context by choosing a reference matrix with components identically set to 1.

3.3 The Restriction Distance Measure

Let $j=1,2,\ldots,J_S$ number the soft restrictions and $j=J_S+1,\ldots,J$ the **hard** ones i.e. those which have to be strictly met. Let further A_H denote the set of matrices satisfying all hard restrictions. The objective of approaching the soft constraints b_j as close as possible can be achieved by minimizing a **restriction distance function**

$$D_R(x) = \sum_{j=1}^{J_S} c_j Q_j (L_j(x) - b_j) \tag{8}$$

on the set A_H of matrices admissible with respect to the hard restrictions. Here the Q_j are suitable penalty functions measuring the deviation of the matrix aggregates $L_j(x)$ from the corresponding soft constraints b_j. In ADETON $Q_j(t) = t^2$ is used for soft equality constraints and $Q_j(t) = t^2$ for $t > 0$ and $Q_j(t) = 0$ for $t \leq 0$ resp. $Q_j(t) = 0$ for $t > 0$ and $Q_j(t) = t^2$ for $t \leq 0$ for soft inequalities. Soft bandwidth restrictions are represented by pairs of soft inequalities.

The weights c_j can be used to control the relative degree of approximation within the set of soft restrictions. A value of c_r which is large in relation to other weights forces restriction r to be more accurately met than others.

3.4 Matrix Estimation as Bicriterial Optimization

In general it is not possible to minimize simultaneously a structural distance function $D_T(x,u)$ and restriction distance function $D_R(x)$ on the set A_H of matrices satisfying the hard restrictions. Diminishing $D_T(x,u)$ enlarges $D_R(x)$ and vice versa. Under these circumstances matrix estimation is a decision problem with two conflicting objective functions. In mathematics this type of problem is dealt with by *Multicriteria Optimization*, also called *Vector Optimization* (Ehrgott 2005) and the case of only two objectives is termed *Bicriteria Optimization*.

The solution of a multicriteria decision problem is a two-step procedure. In the first — purely mathematical — step the so called *efficient* or *Pareto optimal* solutions are computed. In matrix estimation a matrix x_0 is **efficient** if no other admissible matrix exists which is "better" than x_0 with respect to both distance measures.

In the second step a decision has to be made which of the efficient solutions is to be regarded as the best one. This in general is a decision based on non-mathematical grounds and cannot be performed by mathematical algorithms.

Mathematics can only provide the set of efficient solutions. Under mild regularity conditions which are met by the above defined distance functions the set of efficient matrices is that of those matrices x(λ) which minimize the **weighted distance functions**

$$D_{\lambda}(x) = (1-\lambda)D_T(x,u) + \lambda D_R(x) \quad (9)$$

on the set A_H, where λ ranges from 0 to 1. When λ is set to 0 soft restrictions are disregarded and only the structural distance is minimized. For λ = 1 the best fit to the soft constraints is sought without respect to structural similarity to a reference matrix. For λ greater than 0 and lower than 1 the solutions x(λ) represent compromises between structural similarity and adaption to constraints. The decision maker has to choose which one out of these compromises is the best one in the underlying situation.

3.5 ADETON

ADETON is a method which realizes matrix estimation with the described properties. It is also a software tool which computes the efficient matrices x(λ) minimizing the functions (9). Besides the matrix it outputs statistics about restriction and structural deviations. Therefore by varying the parameter λ a fast search for the "best" efficient solution may be undertaken interactively.

ADETON is implemented as standalone program for Linux and Windows and as an Add-In to Microsoft Excel. Reference matrix data and restriction descriptions can be read either from simple text files or from Excel sheets. For ADETON a special, relatively simple description "language" has been developed to specify the restrictions and other aspects of the estimation process.

4. Results obtained

In the following examples of results of the application of the ADETON procedure are shown. The examples are only of an illustrative nature. Hundreds of restrictions are to be specified to estimate a complete MAS matrix. Therefore it is not possible to reproduce the complete estimation process in a short paper. The matrices were obtained by using the Chi-square distance.

The first example is taken from the Haas, Rothe (2005b) paper on the Multi-Accounting System. A characteristic of the MAS is its regional perspective. Here, the labour market of Ingolstadt, one of 180 German employment agencies, is used as an example. Since the transition matrix for every region is large — they relate to over 60 states — we focus on a special part of this matrix and look at the transitions from vocational training to various labour market statuses. In Table 2 the transition rates from vocational training into employment broken down by several industries for the area of Ingolstadt are shown. Complete matrices are available from the authors by request.

Table 2: Stocks and transitions from vocational training into dependent employment in the Ingolstadt employment agency in per cent (2002)

		Transitions into dependent employment									
		Year-beginning stock	SEC1	SEC2	SEC3	SEC4	SEC5	SEC6	SEC7	Total outflows	Year-end stock
Source account (on-the-job/off-the-job training	1. Agriculture and forestry (SEC1)	124	**30.4**	1.3	1.3	2.5	1.3	1.3	11.4	79	125
	2. Manufacturing (SEC2)	3.740	0.2	**59.0**	0.5	3.0	0.6	0.5	1.9	1.284	3.775
	3. Construction (SEC3)	1.363	0.2	1.1	**42.9**	3.5	2.0	0.7	3.5	538	1.326
	4. Trade and transport	2.553	0.2	1.5	0.4	**43.6**	0.6	1.7	2.2	982	2.601
	5. Business related services (SEC5)	997	0.2	2.0	0.4	4.6	**47.8**	1.3	2.2	456	1.007
	6. Other services (SEC6)	1.626	0.2	2.2	0.7	2.1	5.0	**36.9**	1.8	937	1.673
	Sector unknown (SEC7)	137	0.3	2.3	1.2	4.3	3.5	3.8	2.3	345	87

Source: Haas, Rothe (2005b)

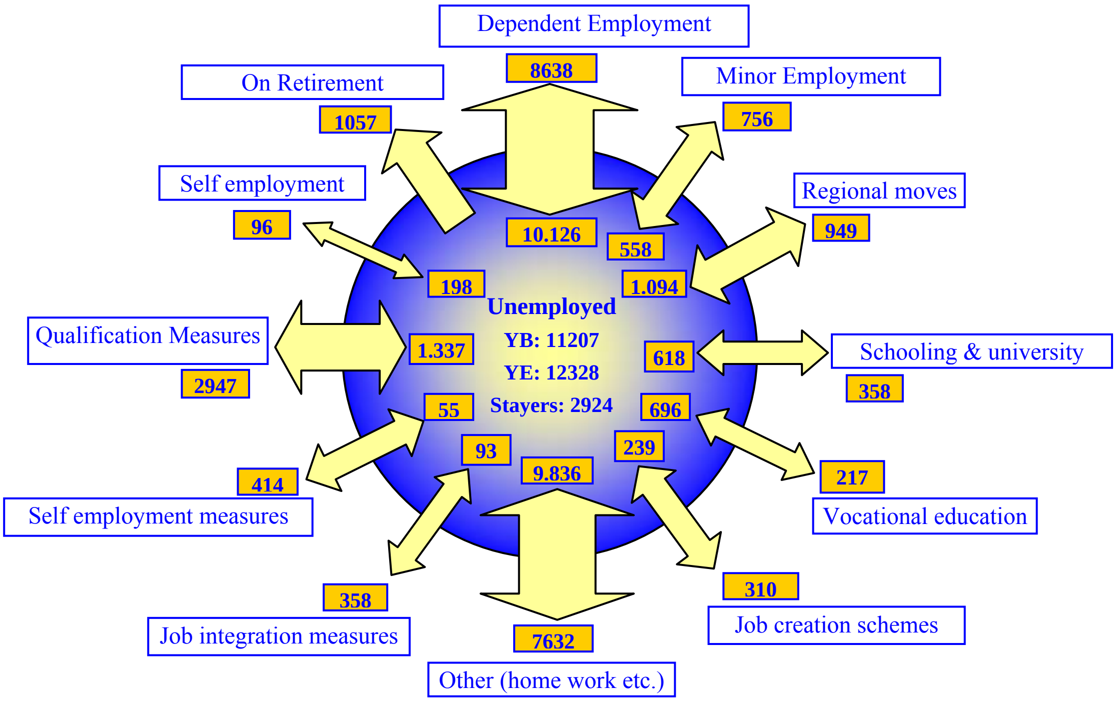
In Table 2 the transitions into employment were shown as proportions of all outflows from training in the respective economic sector. They can therefore be interpreted as transition probabilities. The probability of obtaining work in the same economic sector after vocational training was 59 percent in manufacturing. This is a figure nearly twice as high as the one for agriculture. Good prospects for further employment were also offered in business related services. In “other services” the transition rate was only 37 %. The MAS facilitates a closer look at this specific sector to find out what was going on there. The transition rate into unemployment of 11 % was below the average of all sectors (15 %). In 5 % of the cases there were transitions in additional vocational training. 7 % of the former apprentices went to a university. Finally, apprentices from this sector are relatively mobile, because 14 % of them moved from the Ingolstadt area to other regions. All these figures are important for labour market policy. With them it is possible to identify labour market problems in advance.

Figure 1 shows the flows from and into employment in the Ingolstadt area in the year 2002. The transition flows are given a graphic representation. From the figure it can be seen, that the figure for unemployment changed not much over the year, at the beginning it was 11207 (= YB) and at the end of the year it was 12328(= YE). Regarding this stability of the stock, the high in- and outflow rates indicate much dynamics in behind. There were many people coming in and out of employment, but the size of the flows relating the stock of the unemployed to the stock of people outside the labour market (and the qualification system = “other”) were especially remarkable. Many people resigned and withdrew from the labour market if they were unemployed for a long time. Others came back, e. g. after a phase of home-work.

Similar patterns can be observed with respect to employment. Politicians are often disappointed when there is an improvement in the course of the business cycle, but the rate of unemployment hardly falls. In these cases many people take over who have been not registered as unemployed but waited for a job. The sizes of all these flows show great regional variation.

Therefore, the measures of regional labour market policy (shown also in Figure 1) should be optimized with respect to maximal success, i.e. the reduction of regional unemployment and the increase of regional employment. This could be done best by using the special pattern of the flows on the regional labour market.

Figure 1: Graphical representation of the flows in and out of unemployment in the labour market of Ingolstadt



Source: Haas, Rothe (2005a)

5. Conclusion

In this paper we have argued that the ADETON procedure is an instrument that could be used for the estimation of large matrices on the basis of complex and partly contradictory information. It is possible to estimate a Multi-Accounting System describing the flows on regional labour markets. The required information is taken from many different and inconsistent sources. Other possible applications include a wide range of research problems. Regional input-output tables or regional mobility tables could also be estimated.

A special feature makes ADETON useful in particular. Since soft constraints can be taken into account the handling of inconsistent information is a lot easier than by using an algorithm which can only digest hard constraints. Applying only hard constraints invariably leads to systems with an empty solution set, which then must be corrected in a tedious search process. Apart from this, the inclusion of soft constraints is in the very nature of statistical information that is available only with a specific error. The gap between “reality” and statistical measurement is due to the limited reliability and validity of each measurement process. It is further due to sampling error and to differing measurement concepts of different data sources. The inclusion of soft constraints is a rational way to handle all these data problems, since they are not ignored or assumed away but taken into account in the estimation process.

References

Balderjahn, Ingo; Mathar, Rudolf; Schader, Martin (1998) (ed.): "Classification, Data Analysis, and Data Highways" (Proceedings of the 21st Annual Conference of the Gesellschaft für Klassifikation), Berlin etc.: Springer

Blien, Uwe; Graef, Friedrich (1998): "Entropy Optimizing Methods for the Estimation of Tables", in: Balderjahn et al. (1998): 3-15

Blien, Uwe; Tassinopoulos, Alexandros (2001): "Forecasting Regional Employment with the ENTROP Method", in: Regional Studies 35/2: 113-124

Deville, Jean-Claude; Saerndal Erik; Sautory, Olivier (1993): “Generalized Raking Procedures in Survey Sampling”, Journal of the American Statistical Association 88/423: 1013-1020

Ehrgott, Matthias (2005): “Multicriteria Optimization”, Berlin, Heidelberg: Springer

Golan, Amos; Judge, George; Miller, Douglas (1996): "Maximum Entropy Econometrics. Robust Estimation with Limited Data", New York etc.: John Wiley and Sons

Haas, Anette; Rothe, Thomas (2005a): "Dynamik am Arbeitsmarkt aus der regionalen Perspektive. Das Mehrkontenmodell als Instrument für Strategie und Steuerung", in: Zeitschrift für Arbeitsmarktforschung

Haas, Anette; Rothe, Thomas (2005b): "Labour market dynamics from a regional perspective: The multi-account system as a strategy instrument for labour market policy", IAB Discussion Paper, Nürnberg

Kruppe, Thomas (2001): “Assessing labour market dynamics. European Evidence”, International Labour Office, Employment paper 2001/15, Geneva

Kullback, Salomon (1968): "Information Theory and Statistics" (2nd ed.), New York: Dover

Shannon, Claude E.; Weaver, Warren (1949): "The Mathematical Theory of Communication", Urbana: University of Illinois Press