Research Report 0006/A

Estimating missing data within an accounting and aggregation framework

Commissioned by: Ministry of Economic Affairs

Ton Kwaak
Niels Bosma

ISBN: 90-371-0791-5

Price: NLG 20.-

Order number: H0006

This study is funded by the Ministry of Economic Affairs.

EIM / Business & Policy Research employs 170 professionals. EIM provides policy- and practice-oriented socio-economic information on and for all sectors in private enterprise and for policy-makers. EIM is established in Zoetermeer. Besides on the Netherlands, EIM also focuses on the European economy and on other continents. You may contact us for more information about EIM and its services.

Address:	Italiëlaan 33
Mailing address:	P.O. Box 7001
	2701 AA Zoetermeer
Telephone:	+ 31 79 341 36 34
Fax:	+ 31 79 341 50 24
Website:	www.eim.nl

Contents

Summary

In this study we investigate methods to estimate missing data in an accounting and aggregation framework. It is important for EIM Business & Policy Research to do this, as EIM's mission is 'to provide knowledge about the enterprise sector to governments and other parties'. Of course, the estimates obtained should be reliable. EIM has the expert knowledge to do so. The missing data may be estimated 'by hand', but this is very time-consuming. Therefore, a formal procedure, which can to a large extent be automated, is suggested.

First of all, preliminary estimates are prepared for all missing variables. These estimates need not necessarily fit exactly in the accounting/aggregation framework[1]. In the next stage, the preliminary estimates are fitted to the accounting/aggregation framework by means of an optimisation process, in which penalties are given for the differences between final and preliminary estimates. The sum of all penalties has then to be minimised. On both theoretical and empirical aspects, a weighted least-squares approach fits best to our particular problem of estimating missing data.

An empirical example shows that, when preliminary estimates are produced in a rather mechanical - though sensible - way, final estimates are close to the real values. Applying the procedure to more or less randomly assigned preliminary estimates results in considerably worse final estimates. This means that people, who are capable of producing preliminary estimates that do make sense, especially in the cases where relatively many data are missing, should implement the procedure. However, their work is relieved by the fact that they can focus on the individual observations themselves, without bothering much about the accounting/aggregation framework.

[1] Fitting these in the framework actually is the most labour-intensive part of the 'by hand' procedure.

1 Introduction

The mission of EIM Business & Policy Research is 'to provide knowledge about the enterprise sector to governments and other parties'. Correspondingly, EIM should have available among other things quantitative information about the economy. Part of this information can be derived from national statistical institutes, like CBS (Statistics Netherlands) and international organisations, such as EUROSTAT and OECD. However, since often detailed information is needed, such institutions fail to provide all information that is deemed necessary, because of for example confidentiality of data. To provide a complete picture of the enterprise sector, EIM should, therefore, enrich the information from statistical sources.

EIM is able to prepare the additional estimates needed because of its thorough knowledge of the enterprise sector. Often, this information is brought in the analysis 'by hand', which is, however, very time-consuming. Thus, productivity can be increased if a systematic method is available to perform the estimates needed. Such a method should be able to take all available information into account. This holds not only regarding the availability of data as such, but also with respect to the information which is contained in the accounting framework in which usually the data are placed. That is:

- sectors of industry should make up the macro-economic total. This is important, because at the highest level of information, often data on more variables are available. Also, within industries, size classes should add to the industry total.
- usually, variables are related to one another. For example, profits can be viewed as the difference between sales and costs.

In this paper, such a method is presented, and illustrated with an example. In chapter 2, a non-technical outline of the method is presented. Then, chapter 3 gives the formal description. Chapter 4 gives an application of (the various variants of) the method. The example starts from known data, so the example can serve to evaluate the extent to which the method is useful for its purpose. Finally, chapter 5 presents some conclusions.

2 Broad overview of the method

2.1 Setting the scene

In general, data are made available to EIM in tabular form. The enterprise sector is subdivided into sectors of industry, and within industries, size classes are distinguished. These units of observations constitute the columns of the table. Variables are presented as rows of the table. A typical data matrix made available to EIM is presented in figure 1.

figure 1 typical data delivery to EIM

	units of observation								
	industry 1			industry N			total of industries*		
variable	small firms	large firms	total	small firms	large firms	total	small firms	large firms	total
1									
2	n.a.**	n.a.**		n.a.**	n.a.**		n.a.**	n.a.**	
…									
…	n.a.**	n.a.**		n.a.**	n.a.**		n.a.**	n.a.**	
M									

* Often to be referred to as the macro level.
** On shaded entries, no data are available.

From this figure, the obvious relations between units of observations follow:

- Within each industry (and at the macro level), small and large firms[1] should add to the industry total.
- Within each size class at the macro level, industries should add to the macro-level data.

Similar rules should hold at the sub-industry level.

There are also relations between variables. These, however, depend on the subject matter of data. For example:

- In a labour-market context: the number of part-time workers and the number of full-time workers should add to total employment in each unit of observation.
- In a national-accounts context: value added equals the difference between gross production and intermediate use of goods and services.

In general, statistical institutes are not able to fill all the entries of the data matrix. In figure 1, these entries are shaded.

1 In the sequel, the terms ‘firms’ and ‘enterprises’ will be used interchangeably.

The methods proposed in this paper attempt to answer the following question: *how can (preferably statistically optimal) estimates be prepared for the missing data in a data matrix?* There are two main assumptions on which the estimation procedures are built:

- For each variable, data are available at the total macro level (or: the last column in figure 1 is filled completely).
- For each unit of observation, data are available for at least one variable (these data may or may not be present in the table to be dealt with).

2.2 Two-stage procedure

Basically, the methods consist of two steps:

- First, so-called preliminary estimates for the missing entries are constructed. These preliminary estimates are either based on expert knowledge, or they are derived in a more mechanic way by assuming constant relations between units of observations or between variables. A few examples might illustrate these preliminary estimates:
  - In many service industries, stock building is not a significant phenomenon; thus, an expert will decide to set stockbuilding to zero.
  - As a first approximation, labour costs might be assumed to be distributed over industries or size classes according to employment.
  - Cost structures tend to be rather stable over time, so if recent observations on the cost structure are available, they can be used to prepare preliminary estimates.
- The preliminary estimates prepared in the first stage do neither take into account the relations between industries and size classes, nor the relations between variables. In the second stage, therefore, the information on accounting rules and aggregation is used to adjust the preliminary estimates to fit with these relations. Here, the following methods are analysed:
  - the RAS-method of proportional adjustments of rows and columns;
  - optimisation methods: a least-squares method, which aims at minimising the differences between the preliminary estimates and the final estimates (in a least-squares sense), and an entropy method.

This is discussed in more detail in chapter 3.

3 Formal representation of methods

This chapter discusses a formal representation of the methods. First, the available data will be described; also, some remarks will be made regarding the preliminary estimates. Then, three methods aiming at fitting the preliminary data to the information enclosed in the accounting rules - between units of observation and the relations between variables - will be discussed.

3.1 Data and preliminary estimates

Data and restrictions

The available data matrix is denoted as $\mathbf{X}$, with characteristic element x_{ij}: the value of variable i in unit of observation j. Some elements of $\mathbf{X}$ are given by the statistical institute; these elements of the matrix are denoted as x_{ij}^{g}. The remaining elements - which are as yet unknown - are denoted as x_{ij}^{u}. There are (probably linear) relations between the columns. This can be represented as follows:

(I) $\Lambda_c(\mathbf{X}) = 0$, where Λ_c denotes a set of (linear) relations between columns (units of observation) of $\mathbf{X}$ (aggregation).

Similarly, a set of relations exists between rows of $\mathbf{X}$:

(II) $\Lambda_r(\mathbf{X}) = 0$, where Λ_r denotes a set of (linear) relations between rows (variables) of $\mathbf{X}$ (accounting rule).

Finally, it seems reasonable to assume the existence of a set of inequalities regarding the values of $\mathbf{X}$, since from economic theory (or common sense) values for some variables should fulfil certain restrictions concerning their sign[1]. Thus:

(III) $\mathrm{T}_r(\mathbf{X}) \geq 0$, where T_r is an operator multiplying by -1 if that variable is to be negative, and by 1 if it is to be positive.

A formal representation of the estimation method is: *Find values for* x_{ij}^{u} *, such that relations (I), (II) and (III) are fulfilled*.

[1] For example, negative turnover should be excluded as a result from the estimation procedures.

Preliminary estimates and their weights

Preliminary values for all x_{ij}^{u} are obtained using information (regarding e.g. variable totals, corresponding variable figures in related industries and values known from a previous year) and expert knowledge. Some examples are stated in section 2.2. The intention for these preliminary values is to estimate them close to the real values, which will facilitate the process of obtaining final estimates.

In the next section, we shall see that in deriving these final estimates, weights may be implemented for each x_{ij}^{u}. These weights have their interpretation as the (un)certainties of the preliminary estimates, measured as the inverse of the expected variance of the entry. Often, the preliminary estimate a_{ij} is taken as a proxy for this variance.

3.2 Final estimates

In general, final estimates are obtained by adjusting the preliminary estimates to the known totals, within the accounting and aggregation framework. We have two methods: the RAS-method and an optimisation method. For the latter, two approaches are identified: the least-squares and the entropy approach. The representations set out below are derived from Schneider and Zenios (1989).

3.2.1 RAS-method

The RAS algorithm is a scaling method for solving relations (I), (II) and (III), in which the preliminary matrix $\mathbf{A}$ is adjusted in turn for the row and the column totals. Scaling the rows induces the column sums to change, and scaling the columns changes the row sums. The iterative scaling process is repeated until the matrix is balanced. The following algorithm is applied:

Basic RAS-method

Step 0. Initialisation

Set k=0 and $\mathbf{A^0} = \mathbf{A}$ (preliminary estimates)

Step 1. Row Scaling

Define $\rho_i^k = \dfrac{u_i}{\sum_j a_{ij}^k}$ for $i = 1,2,\ldots,m$,

where u_i is the given row total, $u_i = \sum_j x_{ij}$.

Update $\mathbf{A}^k$ by

$$a_{ij}^k \leftarrow \rho_i^k a_{ij}^k, \quad \begin{array}{l} i = 1,2,\ldots,m \\ j = 1,2,\ldots,n \end{array}$$

Step 2. Column Scaling

Define $\sigma_i^k = \dfrac{v_j}{\sum_i a_{ij}^k}$ for $j = 1,2,\ldots,n$,

where v_j is the given column total, $v_j = \sum_i x_{ij}$;

and define $\mathbf{A}^{k+1}$ by

$$a_{ij}^{k+1} = a_{ij}^k \sigma_j^k, \quad \begin{array}{l} i = 1,2,\ldots,m \\ j = 1,2,\ldots,n. \end{array}$$

Step 3. Convergence?

If convergence is achieved, stop the algorithm. Otherwise, replace k with k+1 and return to step 1.

This basic method is adjusted to our current problem with respect to the following:

1. generalisation to the N-dimensional problem (straightforward)
2. take into account restriction (III)
3. set $x_{ij} = x_{ij}^g$ when variable i of observation unit j is known. This is implemented by setting the element to zero and subtracting x_{ij}^g from the corresponding row and column totals.

3.2.2 Optimisation methods

Optimisation methods use relations (I), (II) and (III) as restrictions in optimising a certain function. We consider two methods: the least-squares method and the entropy method.

Least squares

The least-squares method aims at minimising the sum of squared differences between the final estimates and the preliminary ones, subject to relations (I), (II) and (III) (see box). If all w_{ij}'s are given the value 1, we deal with a non-weighted least-squares approach. The quadratic function has a statistical interpretation as a generalised least-squares estimator. Using a quadratic function with weight terms defined by the inverse of the preliminary estimates $(w_{ij} = 1/a_{ij})$ gives a chi-square estimate[1]. Using this weighting, each entry adjustment takes place in proportion to its magnitude.

Additional constraints may be imposed, such as upper and lower bounds or non-negativity constraints on particular entries. For an extensive analysis, see Harthoorn (1994).

Basic least-squares method

Minimise $$\sum_{i,j} w_{ij}(x_{ij}^u - a_{ij})^2$$

subject to the restrictions (I), (II) and (III),

where the a_{ij} is the preliminary estimate for the unknown x_{ij}^u and w_{ij} is the corresponding weight.

Entropy

The entropy method differs from the least-squares method in the form of the function to be minimised (see box). This entropy function has a theoretical justification based on the principles of information theory. A useful survey regarding the principles and extensions of maximum entropy economics is provided by Golan et al. (1996). Disadvantage of the entropy method is that, on practical application, problems may occur concerning estimated values that are assigned the opposite signs of the preliminary estimates. This, of course, depends on the numerical algorithm used to solve to problem[2].

1 The preliminary estimates are considered as approximations to the variances of all missing entries. In this context, the resulting function converges to a chi-square distribution under the hypothesis that the final estimates fit the preliminary estimates.

2 Obviously, the entropy problem has no analytical solution, so numerical algorithms have to be applied.

Basic entropy method

Minimise $\sum_{i,j} w_{ij} x_{ij}^u \left[\ln\left(\frac{x_{ij}^u}{a_{ij}} \right) - 1 \right]$

subject to the restrictions (I), (II) and (III),

where the a_{ij} is the preliminary estimate for the unknown x_{ij}^u and w_{ij} is the corresponding weight.

Lests squares versus entropy

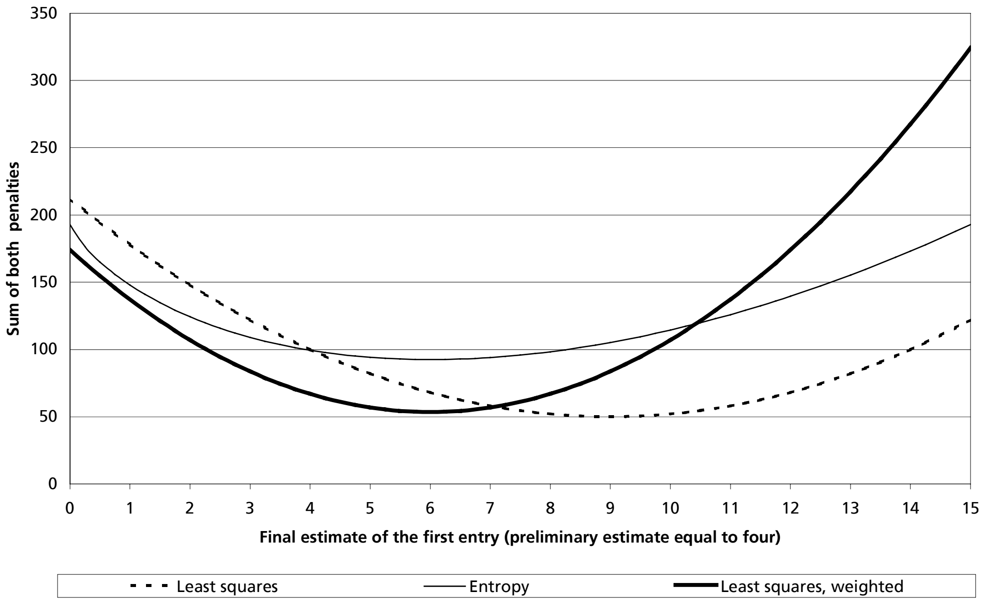
To explain the differences between the penalty functions in the two optimisation techniques set out above, we use a very simplified example. Assume there are only two observations for variables A and B. Suppose we have preliminary estimates of these variables: a_A=4 and a_B=16. These estimates may, for example, stem from known values for the previous year. Suppose further that these two variables should add up to 30 in the current year. What will be the final estimates for A and B for each method?

For each final estimate value of A, x_A^u, the final value of B is determined as 30-x_A^u. The sum of both penalties is easily calculated, and for each method, the final estimate of A will be the value that minimises this sum.

The least-squares approach relates penalties to absolute errors. This implies that the total error of 30-(4+16)=10 is equally divided between A and B, resulting in a final estimate for A of x_A^u=9. We see indeed in figure 2 that for A=9, a minimum is achieved for the sum of both penalties. The entropy method instead considers relative errors. As a result, the matching penalty function takes on its minimum near x_A^u=6. We argued earlier that relative errors can be achieved in the least-squares approach as well, by means of implementing weights. The bold line in figure 2 reflects this weighted least-squares approach. In this case, a minimum is achieved at x_A^u=6 as well.

This extremely simplified example may help to understand the differences between the two penalty functions and their impacts on the final estimates. In the next section, the merits and drawbacks of all methods are discussed.

figure 2 illustrating the differences between penalty functions (two variables, A and B; preliminary estimates: a_A= 4, a_B= 16; A + B= 30)



Discussion

The three methods are, without forcing a 'best method' decision yet, compared regarding theoretical foundations and applicability.

Theoretical foundations

As pointed out above, both the least-squares method and the entropy method have theoretical justifications. The RAS-method lacks of a theoretical meaning, though Bregman (1967) showed that if the problem is feasible (i.e. it is possible to satisfy all restrictions), the RAS algorithm converges to the balanced matrix that minimises $\sum_{i,j} x_{ij}^u \ln\left(\frac{x_{ij}^u}{a_{ij}}\right)$ subject to the given sum and row totals.

When all totals are known, this corresponds to the solution of the non-weighted entropy method.

Applicability

The RAS algorithm can only be used when all row and column totals of the balanced matrix are given. The least-squares method and the entropy method can be used in a situation of (some) unknown row and column totals. In this case, preliminary estimates with their particular weights are implemented.

Further, the least-squares and the entropy optimisation methods serve in imposing the linear relations (I), (II) and (III) more easily than the RAS-method.

Another feature of RAS is that it has the restrictive property of sign-preserving (and zero-preserving) final estimates. The entropy could give some problems when, in the optimisation process, the sign of the estimated entry value becomes the opposite of the sign of the preliminary entry value. Dependent on the application programme used, the procedure may unnecessarily be aborted. It is possible to minimise the chance of this disadvantageous occurrence by using the outcomes of some other method as a starting point. The final estimates will, however, always have the same sign as the preliminary estimates.

Conclusions

We conclude that, from a theoretical point of view, optimisation techniques are superior to the RAS-method. RAS is related to the entropy method, since the RAS solution converges to the non-weighted entropy method. We therefore do not consider the RAS technique any further as an option in estimating missing data. Various studies examining potential inaccuracies arising from the RAS-method (e.g. Lynch, 1986) justify this conclusion.

The restrictive sign-preserving property of the entropy method is not a sufficient reason to opt for the least-squares method immediately. The impact of the different optimisation functions should be investigated before a choice is made. In order to capture the merits of each method for our particular problem, we describe results from applications to an existing data set in the next chapter.

4 Illustration with an example

4.1 Data

Empirical testing of the methods set out in chapter 3 has been done using an example that serves the purpose of the procedure well. A subset of the BLISS data set[1] has been used. This subset contains values from 1995, that are already known. Now, some of these known values were *assumed to be unknown*. The resulting matrix we deal with can be found in appendix I.

4.2 Preliminary estimates

Preliminary estimates were obtained using known employment data (by industry and size class) as weights for all unknown variables to obtain estimates for the size classes. These weights were used to distribute given totals among size classes.

Applying this rule is not very accurate. It is known that in most industries export per employee is considerably less in small businesses compared to large enterprises. This means that we expect that for exports in the small businesses, the method will result in too high estimates.

Using two different matrices of preliminary estimates may point out the importance of bringing in good preliminary estimates. Therefore, we shall do the same exercise with (more or less) random preliminary estimates.

4.3 Final estimates

Final estimates were calculated from five methods:

1. LSQ non-weighted least squares
2. ENT non-weighted entropy
3. LSQW weighted least squares, using the inverse of the preliminary estimate $(w_{ij} = 1/\left|a_{ij}\right|)$ as weights
4. ENTW weighted entropy, using the inverse of the preliminary estimate $(w_{ij} = 1/a_{ij})$ as weights
5. LSQDW double weighted least-squares, using the inverse of the preliminary estimate squared $(w_{ij} = 1/a_{ij}{}^2)$ as weights.

The final estimates are presented in appendix II, together with the preliminary values and the real values. For a first comparison, we calculated the correla-

[1] BLISS (Enterprise sector information system; in Dutch: BedrijfsLeven InformatieSySteem) is the EIM database containing very detailed information on the Dutch enterprise sector.

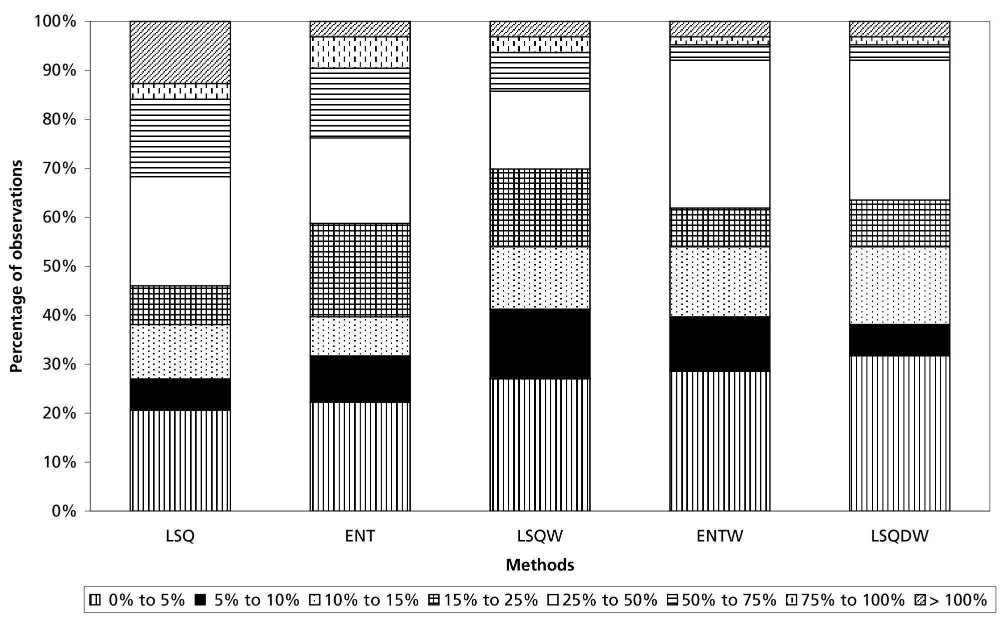
tion of all these values as printed in table 1. The correlation between the real and the preliminary estimates is equal to 0.862.

table 1 correlations between final estimates and actual values and preliminary estimates

	Correlation with actual values	Correlation with preliminary estimates
LSQ	0.980	0.879
ENT	0.993	0.874
LSQW	0.997	0.870
ENTW	0.998	0.866
LSQDW	0.998	0.866

To make for a better comparison, we calculated the percentage errors of all estimates, relative to the real values. The distribution of these errors is shown in figure 3. In this figure, we omitted the errors that were equal to zero in all cases. In the table of appendix II, the variables having this property are shaded.

figure 3 cumulative distribution of the relative errors for five methods



We observe that between 46% (LSQ) and 70% (LSQW) of the estimates have a relative error less than 25%. Errors over 100% (2 observations in four out of five methods) appear to be the result of overestimating exports in small businesses, which was expected. The consequence of overestimating these values is, of course, that large firms' exports are underestimated, as total exports by industry are given exogenously. This illustrates the importance of implementing accurate preliminary estimates.

4.3.1 Least squares

From figure 3, it appears that applying weights in the least-squares method has a significant effect on the final outcomes. The weighted least-squares method dominates the non-weighted least-squares method. Comparison between the weighted least-squares method (LSQW) and double weighted least squares (LSQDW) does not point out a favourable method: the single weighted least-squares method produces more estimates in the categories below 25%, but less in the categories below 50 and 75%. This could be the result of our particular example and may be different in other examples. As LSQW has a theoretical foundation, this method is to be preferred over LSQDW.

4.3.2 Entropy

The weighted entropy produces better estimates than the non-weighted entropy. The non-weighted entropy method (ENT) may be compared to LSQW, while weighted entropy (ENTW) may be compared to LSQDW. In both cases, the entropy method does not produce more accurate estimates than the least-squares method.

4.4 Alternative preliminary estimates

To analyse the importance of using appropriate preliminary estimates, we applied the procedure with (more or less) randomly assigned preliminary estimates[1]. We did this first using the weighted least-squares method (LSQW). Figure 4 displays the results, with LSQW2 measuring the final estimates using the alternative random primary estimates.

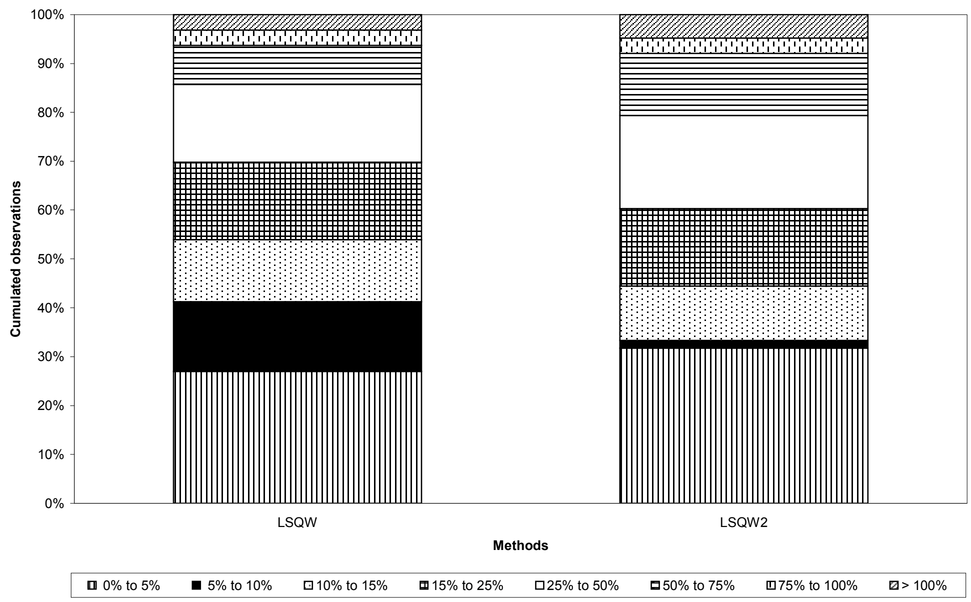
Strikingly, more final estimates fall within the 5% error level with randomly assigned preliminary estimates. For the remaining part, however, the scores are considerably less. It should be noted that for one final estimate a non-tolerable negative sign was produced as well.

Applying the entropy method (ENT) to the same randomly assigned preliminary estimates resulted in an abortion from the procedure, caused by the problem of opposite signs during the optimisation process. Final estimates could not be derived with the LSQW2 solution as starting values, because this solution had a negative estimate (as mentioned above).

Summarising, this exercise shows that the preliminary estimates should be determined very carefully. For this, expert knowledge is a necessity.

[1] The preliminary estimates were constructed as follows. The given totals for variable i were divided by three to obtain equal figures z_i for each size class (small, medium and large). The preliminary estimates were obtained using random figures from a normal distribution with mean z_i and standard deviation $z_i/2$. The correlation between these estimates and the real values is equal to 0.767 and the correlation of the two preliminary estimates used is equal to 0.883.

figure 4 comparison of the LSQW method with different preliminary estimates



4.5 Discussion

Taken all arguments together, it seems best to apply the weighted least-squares method (LSQW). This decision is based on the following arguments.

- The method has a theoretical foundation, unlike double weighted least-squares (LSQDW) and the weighted entropy method (ENTW).
- The method does not cause problems when negative values occur in the optimisation process, unlike both entropy methods.
- In the empirical example, the weighted least-squares is generally not out-performed by other methods, while it certainly gives better results than non-weighted least-squares and the non-weighted entropy method.

It is an option to use the entropy method as a check on the reliability of the estimates obtained by weighted least squares. Starting values for the entropy method algorithm should then be taken from the least-squares solution to the problem. Differences between the outcomes may serve as an indication of reliability of the estimated data.

5 Conclusion

Conclusions from this study

Considering all arguments made in this study, it is worthwhile to set up a procedure for estimating missing data. The advantage of an automated procedure is that the data are estimated, accounting for the entire economic framework. For researchers, there will be no missing values, and for all data well-documented information about the way they are estimated will be available. When some of the data would be estimated ad hoc by an expert, chances are that the estimates will be contradictory to other data that may not be considered, but that are really relevant within the accounting/aggregation framework.

The optimisation method to be used is an approach using weighted least-squares penalties for the differences between preliminary and final estimates. This method has a theoretical foundation, and empirical applications are satisfactory. It is very important, especially when relatively many data are missing, to make sure that the preliminary estimates are quite accurate already. This will pave the way for a smooth optimisation process.

How to proceed?

From the conclusions above, the following approach might be implemented at EIM Business & Policy Research: much effort will be used for obtaining adequate preliminary estimates for one and appropriate year, probably 1995. All missing data are then estimated using the weighted least-squares optimisation method. Next, the years after 1995 may be estimated using the results from 1995, following a rather mechanical procedure. Every three or four years, an extensive check has to be made on the preliminary values.

Literature

Bregman, L.M. (1967). Proof of the Convergence of Sheleikhovskii's Method for a Problem With Transportation Constraints*. USSR Computational Math. and Mathem. Phys.* **1**(1), 191-204.

Golan, A., G. Judge and D. Miller (1996). *Maximum Entropy Economics (Robust Estimation with Limited Data)*. Wiley.

Harthoorn, R. (1994). *On the Integrity of Data and Methods in the Static Open Leontief Model.* Ph. D. Thesis, Twente University, Enschede.

Lynch, R.G. (1986). An assessment of the RAS method for updating input-output tables. In: I. Sohn (Ed*.), Readings in Input-Output Analysis*, pp. 271-284. Oxford University Press, New York.

Schneider, M.H., and S.A. Zenios (1990). A Comparative study of algorithms for matrix balancing. *Operations Research* **38**, 439-55.

Appendix I: Data of the empirical example

table I query on datamart 'policy' for the Netherlands, 1995

	manufacturing				construction				trade				trade			
	large	small	medium	total	large	small	medium	total	large	small	medium	total	large	small	medium	total
turnover	269,898	29,923	86,196	386,017	31,184	20,946	40,373	92,503	162,332	168,982	228,689	560,003	6,800	17,727	8,251	32,778
purch. value of trade	27,040	8,325	10,559	45,924	573	159	320	1,053	120,282	126,708	181,036	428,026	112	182	135	429
sales	242,858	21,597	75,637	340,093	30,611	20,786	40,053	91,450	42,050	42,274	47,653	131,977	6,687	17,545	8,117	32,349
export	n.a.	n.a.	n.a.	172,654	n.a.	n.a.	n.a.	999	n.a.	n.a.	n.a.	18,174	0	0	0	0
consumption goods	n.a.	n.a.	n.a.	43,119	n.a.	n.a.	n.a.	1,679	n.a.	n.a.	n.a.	61,122	3,500	15,604	6,737	25,840
investment goods	n.a.	n.a.	n.a.	17,464	n.a.	n.a.	n.a.	51,037	n.a.	n.a.	n.a.	3,828	0	0	0	0
intermediate goods	n.a.	n.a.	n.a.	106,856	n.a.	n.a.	n.a.	37,735	n.a.	n.a.	n.a.	48,852	3,188	1,942	1,380	6,509
stockbuilding	n.a.	n.a.	n.a.	1,099	n.a.	n.a.	n.a.	0	n.a.	n.a.	n.a.	0	0	0	0	0
gross production	n.a.	n.a.	n.a.	341,191	n.a.	n.a.	n.a.	91,450	n.a.	n.a.	n.a.	131,977	6,687	17,545	8,117	32,349
use of raw materials	n.a.	n.a.	n.a.	180,252	n.a.	n.a.	n.a.	49,500	n.a.	n.a.	n.a.	6,497	1,244	4,469	1,549	7,263
use of energy	5,775	334	1,264	7,373	147	144	227	518	553	798	744	2,095	118	580	208	907
other use	30,553	3,173	9,111	42,837	3,360	2,378	4,220	9,958	15,148	13,229	15,894	44,271	1,307	3,835	1,925	7,067
total use	166,496	13,219	50,747	230,462	21,661	12,019	26,296	59,976	17,842	15,837	19,183	52,862	2,669	8,884	3,683	15,237
value added	77,061	8,379	25,290	110,729	8,950	8,767	13,756	31,474	24,208	26,437	28,470	79,115	4,018	8,661	4,433	17,113
employment	468	115	271	854	104	121	169	393	257	356	283	896	45	151	49	245

	transport				business services				non-primary private enterprises (totals)			
	large	small	medium	total	large	small	medium	total	large	small	medium	total
turnover	53,478	7,972	17,467	78,917	83,152	24,654	27,651	135,457	606,843	270,204	408,627	1,285,675
purch. value of trade	909	185	248	1,342	959	571	802	2,332	149,875	136,131	193,100	479,107
sales	52,569	7,787	17,218	77,575	82,193	24,083	26,849	133,125	456,968	134,073	215,527	806,568
export	13,504	1,456	4,538	19,498	3,458	1,007	1,481	5,946	n.a.	n.a.	n.a.	217,271
consumption goods	13,685	1,192	2,773	17,650	12,219	806	2,032	15,057	n.a.	n.a.	n.a.	164,466
investment goods	603	1	4	607	681	1,296	1,309	3,286	n.a.	n.a.	n.a.	76,223
intermediate goods	24,778	5,138	9,903	39,820	65,835	20,974	22,027	108,836	n.a.	n.a.	n.a.	348,608
stockbuilding	121	0	0	121	0	0	0	0	n.a.	n.a.	n.a.	1,220
gross production	52,690	7,787	17,218	77,696	82,193	24,083	26,849	133,125	n.a.	n.a.	n.a.	807,788
use of raw materials	2,235	103	212	2,550	11,542	4,609	7,047	23,197	n.a.	n.a.	n.a.	269,258
use of energy	2,369	713	1,501	4,582	496	294	174	965	9,457	2,862	4,119	16,439
other use	19,795	2,929	5,966	28,689	19,172	5,240	4,170	28,582	89,335	30,785	41,285	161,405
total use	24,399	3,744	7,679	35,822	31,210	10,143	11,391	52,743	264,277	63,846	118,979	447,102
value added	28,292	4,042	9,540	41,874	50,983	13,940	15,459	80,382	193,511	70,227	96,948	360,686
employment	216	49	89	354	499	194	145	838	1,589	986	1,005	3,580

Appendix II: Results of the empirical example

table II preliminary values, real values and final estimates for the five methods

Name	Preliminary value	Real Value	LSQ	ENT	LSQW	ENTW	LSQDW
export large (manuf.)	94,651	141,246	87,133	109,071	118,471	126,524	127,722
export small (manuf.)	23,227	4,017	5,639	13,163	11,393	8,950	8,304
export medium (manuf.)	54,777	27,391	79,883	50,421	42,791	37,181	36,629
consumption large (manuf.)	23,638	30,402	50,047	35,333	33,448	27,478	27,012
consumption small (manuf.)	5,801	3,684	4,131	1,731	2,940	4,141	4,465
consumption medium (manuf.)	13,680	9,032	-11,059	6,055	6,731	11,500	11,641
investment large (manuf.)	9,574	8,241	24,822	15,050	12,786	10,616	10,587
investment small (manuf.)	2,349	2,439	912	775	1,834	2,096	2,175
investment medium (manuf.)	5,541	6,783	-8,270	1,639	2,844	4,752	4,702
intermediate large (manuf.)	58,579	62,968	80,857	83,405	78,153	78,241	77,537
intermediate small (manuf.)	14,375	11,456	10,916	5,928	5,431	6,410	6,653
intermediate medium (manuf.)	33,901	32,431	15,083	17,522	23,272	22,205	22,665
stockbuilding large (manuf.)	602	699	449	248	453	477	505
stockbuilding small (manuf.)	148	0	206	220	186	186	179
stockbuilding medium (manuf.)	349	400	444	631	459	436	415
gross production large (manuf.)	187,044	243,557	243,307	243,106	243,311	243,335	243,363
gross production small (manuf.)	45,899	21,597	21,803	21,817	21,783	21,784	21,776
gross production medium (manuf.)	108,248	76,037	76,081	76,268	76,097	76,073	76,052
use of raw materials large (manuf.)	98,816	130,168	130,168	130,168	130,168	130,168	130,168
use of raw materials small (manuf.)	24,249	9,711	9,711	9,711	9,711	9,711	9,711
use of raw materials medium (manuf.)	57,187	40,372	40,372	40,372	40,372	40,372	40,372
export large (constr.)	264	553	196	199	262	264	264
export small (constr.)	306	95	227	241	306	304	304
export medium (constr.)	429	350	575	559	430	430	430
consumption large (constr.)	443	610	434	474	448	446	446
consumption small (constr.)	515	418	452	366	507	507	507
consumption medium (constr.)	721	650	793	838	724	726	725
investment large (constr.)	13,467	20,740	27,666	19,790	20,724	18,572	18,420
investment small (constr.)	15,645	8,846	10,230	11,820	9,916	11,855	12,075
investment medium (constr.)	21,925	21,451	13,141	19,427	20,397	20,610	20,542
intermediate large (constr.)	9,957	8,706	2,315	10,148	9,177	11,329	11,480
intermediate small (constr.)	11,567	11,427	9,877	8,358	10,057	8,120	7,900
intermediate medium (constr.)	16,211	17,602	25,544	19,229	18,501	18,287	18,355
stockbuilding large (constr.)	0	0	0	0	0	0	0
stockbuilding small (constr.)	0	0	0	0	0	0	0
stockbuilding medium (constr.)	0	0	0	0	0	0	0
gross production large (constr.)	24,130	30,611	30,611	30,611	30,611	30,611	30,611
gross production small (constr.)	28,033	20,786	20,786	20,786	20,786	20,786	20,786
gross production medium (constr.)	39,287	40,053	40,053	40,053	40,053	40,053	40,053
use of raw materials large (constr.)	13,061	18,154	18,154	18,154	18,154	18,154	18,154
use of raw materials small (constr.)	15,174	9,497	9,497	9,497	9,497	9,497	9,497
use of raw materials medium (constr.)	21,265	21,849	21,849	21,849	21,849	21,849	21,849

Final estimate solely determined by the accounting rules and therefore equal to real value.

Intolerable sign of final estimate.

Table II (continued)

Name	Preliminary value	Real Value	LSQ	ENT	LSQW	ENTW	LSQDW
export large (trade)	5,225	3,977	6,339	3,306	5,041	5,155	5,227
export small (trade)	7,216	5,182	1,693	6,426	6,608	7,059	6,954
export medium (trade)	5,733	9,015	10,143	8,442	6,525	5,961	5,993
consumption large (trade)	17,571	20,262	17,986	22,458	19,449	20,333	20,282
consumption small (trade)	24,269	24,271	20,630	17,570	20,615	17,947	18,019
consumption medium (trade)	19,282	16,589	22,507	21,095	21,058	22,842	22,821
investment large (trade)	1,101	1,678	2,348	1,493	1,139	1,116	1,117
investment small (trade)	1,520	762	492	1,328	1,485	1,513	1,515
investment medium (trade)	1,208	1,389	988	1,007	1,205	1,199	1,197
intermediate large (trade)	14,044	16,132	15,378	14,793	16,421	15,446	15,424
intermediate small (trade)	19,397	12,060	19,459	16,950	13,566	15,755	15,786
intermediate medium (trade)	15,411	20,660	14,015	17,110	18,865	17,651	17,642
stockbuilding large (trade)	0	0	-267	0	0	0	0
stockbuilding small (trade)	0	0	267	0	0	0	0
stockbuilding medium (trade)	0	0	0	0	0	0	0
gross production large (trade)	37,940	42,050	41,783	42,050	42,050	42,050	42,050
gross production small (trade)	52,402	42,274	42,541	42,274	42,274	42,274	42,274
gross production medium (trade)	41,635	47,653	47,652	47,653	47,653	47,653	47,653
use of raw materials large (trade)	1,868	2,141	2,141	2,141	2,141	2,141	2,141
use of raw materials small (trade)	2,580	1,810	1,810	1,810	1,810	1,810	1,810
use of raw materials medium (trade)	2,050	2,545	2,545	2,545	2,545	2,545	2,545
export large (totals)	96,463	162,739	110,629	129,537	140,736	148,904	150,175
export small (totals)	73,019	80,679	97,871	87,668	82,749	77,661	77,144
export medium (totals)	33,841	31,943	56,119	37,616	35,932	31,587	31,408
consumption large (totals)	154,774	181,608	192,350	202,147	197,551	198,816	198,242
consumption small (totals)	542	820	303	369	574	598	626
consumption medium (totals)	358,640	457,789	457,271	457,337	457,542	457,566	457,594
investment large (totals)	119,545	165,485	165,485	165,485	165,485	165,485	165,485
investment small (totals)	59,830	11,758	10,022	22,293	20,770	18,776	18,025
investment medium (totals)	45,289	45,974	42,812	37,268	41,662	40,195	40,592
intermediate large (totals)	20,989	13,344	12,931	15,221	14,532	16,762	17,061
intermediate small (totals)	95,996	62,997	68,307	59,291	57,109	58,339	58,394
intermediate medium (totals)	336	0	473	220	186	186	179
stockbuilding large (totals)	222,440	134,073	134,545	134,292	134,259	134,259	134,251
stockbuilding small (totals)	74,145	30,199	30,199	30,199	30,199	30,199	30,199
stockbuilding medium (totals)	203,047	42,774	96,620	65,440	55,765	49,591	49,071
gross production large (totals)	50,709	37,813	23,783	39,530	40,055	46,610	46,730
gross production small (totals)	20,538	30,936	7,172	23,386	25,759	27,873	27,754
gross production medium (totals)	125,666	104,004	87,952	87,171	93,948	91,453	91,973
use of raw materials large (totals)	1,292	400	444	631	459	436	415
use of raw materials small (totals)	401,252	215,927	215,971	216,158	215,987	215,963	215,942
use of raw materials medium (totals)	211,982	73,575	73,575	73,575	73,575	73,575	73,575

Final estimate solely determined by the accounting rules and therefore equal to real value.

Intolerable sign of final estimate.

Appendix III: List of Research Reports

The research report series is the successor of both the research paper and the 'researchpublicatie' series. There is a consecutive report numbering followed by /x. For /x there are five options:

/E: a report of the business unit Strategic Research, written in English;
/N: like /E, but written in Dutch;
/F: like /E, but written in French;
/A: a report of one of the other business units of EIM/Small Business Research and Consultancy;
/I: a report of the business unit Strategic Research for internal purposes; external availability on request.

9301/E The intertemporal stability of the concentration-margins relationship in Dutch and U.S. manufacturing; Yvonne Prince and Roy Thurik
9302/E Persistence of profits and competitiveness in Dutch manufacturing; Aad Kleijweg
9303/E Small store presence in Japan; Martin A. Carree, Jeroen C.A. Potjes and A. Roy Thurik
9304/I Multi-factorial risk analysis and the sensitivity concept; Erik M. Vermeulen, Jaap Spronk and Nico van der Wijst
9305/E Do small firms' price-cost margins follow those of large firms? First empirical results; Yvonne Prince and Roy Thurik
9306/A Export success of SMEs: an empirical study; Cinzia Mancini and Yvonne Prince
9307/N Het aandeel van het midden- en kleinbedrijf in de Nederlandse industrie; Kees Bakker en Roy Thurik
9308/E Multi-factorial risk analysis applied to firm evaluation; Erik M. Vermeulen, Jaap Spronk and Nico van der Wijst
9309/E Visualizing interfirm comparison; Erik M. Vermeulen, Jaap Spronk and Nico van der Wijst
9310/E Industry dynamics and small firm development in the European printing industry (Case Studies of Britain, The Netherlands and Denmark); Michael Kitson, Yvonne Prince and Mette Mönsted
9401/E Employment during the business cycle: evidence from Dutch manufacturing; Marcel H.C. Lever and Wilbert H.M. van der Hoeven
9402/N De Nederlandse industrie in internationaal perspectief: arbeidsproduktiviteit, lonen en concurrentiepositie; Aad Kleijweg en Sjaak Vollebregt
9403/E A micro-econometric analysis of interrelated factor demand; René Huigen, Aad Kleijweg, George van Leeuwen and Kees Zeelenberg

9404/E Between economies of scale and entrepreneurship; Roy Thurik
9405/F L'évolution structurelle du commerce de gros français; Luuk Klomp et Eugène Rebers
9406/I Basisinkomen: een inventarisatie van argumenten; Bob van Dijk
9407/E Interfirm performance evaluation under uncertainty, a multi-dimensional frame-work; Jaap Spronk and Erik M. Vermeulen
9408/N Indicatoren voor de dynamiek van de Nederlandse economie: een sectorale analyse; Garmt Dijksterhuis, Hendrik-Jan Heeres en Aad Kleijweg
9409/E Entry and exit in Dutch manufacturing industries; Aad Kleijweg and Marcel Lever
9410/I Labour productivity in Europe: differences in firm-size, countries and industries; Garmt Dijksterhuis
9411/N Verslag van de derde mondiale workshop Small Business Economics; Tinbergen Instituut, Rotterdam, 26-27 augustus 1994; M.A. Carree en M.H.C. Lever
9412/E Internal and external forces in sectoral wage formation: evidence from the Netherlands; Johan J. Graafland and Marcel H.C. Lever
9413/A Selectie van leveranciers: een kwestie van produkt, profijt en partnerschap?; F. Pleijster
9414/I Grafische weergave van tabellen; Garmt Dijksterhuis
9501/N Over de toepassing van de financieringstheorie in het midden- en kleinbedrijf; Erik M. Vermeulen
9502/E Insider power, market power, firm size and wages: evidence from Dutch manufacturing industries; Marcel H.C. Lever and Jolanda M. van Werkhooven
9503/E Export performance of SMEs; Yvonne M. Prince
9504/E Strategic Niches and Profitability: A First Report; David B. Audretsch, Yvonne M. Prince and A. Roy Thurik
9505/A Meer over winkelopenstellingstijden; H.J. Gianotten en H.J. Heeres
9506/I Interstratos; een onderzoek naar de mogelijkheden van de Interstratos-dataset; Jan de Kok
9507/E Union coverage and sectoral wages: evidence from the Netherlands; Marcel H.C. Lever and Wessel A. Marquering
9508/N Ontwikkeling van de grootteklassenverdeling in de Nederlandse Industrie; Sjaak Vollebregt
9509/E Firm size and employment determination in Dutch manufacturing industries; Marcel H.C. Lever
9510/N Entrepreneurship: visies en benaderingen; Bob van Dijk en Roy Thurik
9511/A De toegevoegde waarde van de detailhandel; enkele verklarende theorieën tegen de achtergrond van ontwikkelingen in distributiekolom, technologie en externe omgeving; J.T. Nienhuis en H.J. Gianotten

9512/N Haalbaarheidsonderzoek MANAGEMENT-model; onderzoek naar de mogelijkheden voor een simulatiemodel van het bedrijfsleven, gebaseerd op gedetailleerde branche- en bedrijfsgegevens; Aad Kleijweg, Sander Wennekers, Ton Kwaak en Nico van der Wijst

9513/A Chippen in binnen- en buitenland; De elektronische portemonnee in kaart gebracht; een verkenning van toepassingen, mogelijkheden en consequenties van de chipcard als elektronische portemonnee in binnen- en buitenland; drs. J. Roorda en drs. W.J.P. Vogelesang

9601/N Omzetprognoses voor de detailhandel; Pieter Fris, Aad Kleijweg en Jan de Kok

9602/N Flexibiliteit in de Nederlandse Industrie; N.J. Reincke

9603/E The Decision between Internal and External R&D; David B. Audretsch, Albert J. Menkveld and A. Roy Thurik

9604/E Job creation by size class: measurement and empirical investigation; Aad Kleijweg and Henry Nieuwenhuijsen

9605/N Het effect van een beursnotering; drs. K.R. Jonkheer

9606/N Een Micro-werkgelegenheidsmodel voor de Detailhandel; drs. P. Fris

9607/E Demand for and wages of high- and low-skilled labour in the Netherlands; M.H.C. Lever and A.S.R. van der Linden

9701/N Arbeidsomstandigheden en bedrijfsgrootte. Een verkenning met de LISREL-methode; drs. L.H.M. Bosch en drs. J.M.P. de Kok

9702/E The impact of competition on prices and wages in Dutch manufacturing industries; Marcel H.C. Lever

9703/A FAMOS, een financieringsmodel naar grootteklassen; drs. W.H.J. Verhoeven

9704/N Banencreatie door MKB en GB; Pieter Fris, Henry Nieuwenhuijsen en Sjaak Vollebregt

9705/N Naar een bedrijfstypenmodel van het Nederlandse bedrijfsleven; drs. W.H.M. van der Hoeven, drs. J.M.P. de Kok en drs. A. Kwaak

9801/E The Knowledge Society, Entrepreneurship and Unemployment; David B. Audretsch and A. Roy Thurik

9802/A Firm Failure and Industrial Dynamics in the Netherlands; David B. Audretsch, Patrick Houweling and A. Roy Thurik

9803/E The determinants of employment in Europe, the USA and Japan; André van Stel

9804/E PRISMA'98: Policy Research Instrument for Size-aspects in Macro-economic Analysis; Ton Kwaak

9805/N Banencreatie bij het Klein-, Midden- en Grootbedrijf; Henry Nieuwenhuijsen, Ben van der Eijken en Ron van Dijk

9806/A Milieumodel; drs. K.L. Bangma

9807/A Barriers for hiring personnel; Jacques Niehof

9808/A Methodiek kosten en baten Arbowetgeving; drs. K.M.P. Brouwers, dr. B.I. van der Burg, drs. A.F.M. Nijsen en ir. H.C. Visee

9809/E Business Ownership and Economic Growth; An Empirical Investigation; Martin Carree, André van Stel, Roy Thurik and Sander Wennekers

9810/E The Degree of Collusion in Construction; M.H.C. Lever, H.R. Nieuwenhuijsen and A.J. van Stel

9811/E Self-employment in 23 OECD countries; Ralph E. Wildeman, Geert Hofstede, Niels G. Noorderhaven, A. Roy Thurik, Wim H.J. Verhoeven and Alexander R.M. Wennekers

9812/E SICLASS: Forecasting the European enterprise sector by industry and size class; Niels Bosma and Ton Kwaak

9901/E Scanning the Future of Entrepreneurship; drs. N.S. Bosma, drs. A.R.M. Wennekers and drs. W.S. Zwinkels

9902/E Are Small Firms Really Sub-optimal?; Compensating Factor Differentials in Small Dutch Manufacturing Firms; David B. Audretsch, George van Leeuwen, Bert Menkveld and Roy Thurik

9903/E FAMOS; A size-class based financial analysis model; W.H.J. Verhoeven and E.A. van Noort

9904/E Conduct and Performance in Dutch Manufacturing; An Application of Appelbaum 1982 with a Plausibility-Check; Frank A. Hindriks, Henry R. Nieuwenhuijsen and Adriaan J. van Stel

9905/E Non-competitive Rents in Dutch Manufacturing; Conduct and Performance in the New Empirical Industrial Organization; Frank A. Hindriks

9906/E A human-resource-based theory of the small firm; Charlotte Koch and Jan de Kok

9907/N Van werknemer naar ondernemer; Een hybride of directe start?; ir. H.C. Visee en drs. W.S. Zwinkels

9908/E Modelling returns to R&D: an application on size effects; Peter Brouwer and Henry Nieuwenhuijsen

9909/E Turbulence and productivity in the Netherlands; Niels Bosma and Henry Nieuwenhuijsen

9910/E Start-up capital: Differences between male and female entrepreneurs. ‘Does gender matter?’; Ingrid Verheul and Roy Thurik

9911/E Modelling Business Ownership in the Netherlands; Niels Bosma, Sander Wennekers, Gerrit de Wit and Wim Zwinkels

9912/A Measuring innovative intensity: Scale construction; J.P.J. de Jong

9913/A Determinants of firm size; Y. Bernardt and R. Muller

0001/E Strategies, uncertainty and performance of small business startups; Marco van Gelderen, Michael Frese and Roy Thurik

0002/E Determinants of Successful Entrepreneurship; Niels Bosma, Mirjam van Praag and Gerrit de Wit

0003/E Comparative Advantages in Estimating Markups; Frank A. Hindriks, Henry R. Nieuwenhuijsen and Gerrit de Wit

0004/A The ARKO labour-cost model: Characteristics and application; G.Th. Elsendoorn and A.H. Nieuwland

0005/E The impact of contestability on prices in manufacturing industries; Frank Bosman