WP. 42
ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

Joint UNECE/Eurostat work session on statistical data confidentiality
(Geneva, Switzerland, 9-11 November 2005)

Topic (vi): Software for statistical disclosure control

IMPROVING CONFIDENTIALITY WITH t-ARGUS BY FOCUSSING ON CLEVER USAGE OF MICRODATA

Supporting Paper

Submitted by Statistics Netherlands[1]

[1] Prepared by Roland van der Meijden (rmdn@cbs.nl).

IMPROVING CONFIDENTIALITY WITH τ-ARGUS BY FOCUSSING ON CLEVER USAGE OF MICRODATA

Roland van der Meijden MSc, Statistics Netherlands

Abstract: Users of tabular data want more and more detailed information, which has huge consequences for confidentiality. In order to meet the users' needs without losing a lot of information (caused by primary and secondary suppressions) one should focus on clever usage of microdata and available tools. Clever usage of microdata can be divided into four different areas. First, reconstructing hierarchies of classifications from narrow to wide hierarchies will diminish primary suppressions as well as secondary suppressions on higher hierarchical levels when making publications on lower hierarchical levels. Second, improving small area estimates will improve turnover estimates and distribution of turnover, which results in less dominance and thus less primary suppressions. Third, deliberately adjusting microdata in extraordinary situations can counteract a lot of primary and secondary suppressions. Fourth and last, bringing the different publication obligations into accordance with each other will diminish the existence of unnecessary suppressions. Clever usage of available tools can be divided into two areas. First, (mis)using the history file in τ-Argus in order to direct the confidentiality pattern. Second, using different information loss weights for improving secondary suppressions. When using all these clever adjustments together, considerably less information will be lost due to confidentiality reasons.

I. INTRODUCTION

1. At Statistics Netherlands the software tool τ-Argus is used for statistical disclosure control (SDC) of tabular data for (most) business statistics. τ-Argus uses cell suppression as SDC technique in order to make it impossible to exactly or approximately recalculate sensitive cells in published tables. The cell suppression technique comprises two steps. First, suppressing the primarily sensitive cells and second suppressing a number of cells in order prevent disclosure due to the additive relationship between the cells of the table (the so-called secondarily sensitive cells). Although the algorithms used are rigid, the outcome of cell suppression (and subsequent information loss) can be directed. τ-Argus offers a couple of features that can be (mis)used for directing cell suppression and it can also be beneficial to make changes to the base material used for constructing the tables.

2. This paper will give an overview of the possibilities in directing cell suppression and improving confidentiality by changing base material and using some specific features of τ-Argus. This paper will also shortly mention what can be done in order to improve confidentiality without the need for tooling or changing base material. Hopefully this paper can give enough practical guidance to diminish the occurrence of overprotected tabular data.

3. This paper starts in Section II with an explanation of why software tools are important when working on confidentiality. This section also mentions the different confidentiality rules that τ-Argus offers. In Section III the different tuning possibilities of τ-Argus are extensively described. This section also descibes how tuning is done and shows several results of the different tuning techniques. The effects that changes in base material have on confidentiality will be discussed in Section IV. In Section V the coordination of publication obligations is mentioned.

II. WHY ONE SHOULD WORK ON CONFIDENTIALITY

IIa. GENERAL REASONS

4. So many papers have been written on confidentiality that everybody already knows that statistical offices do not want to jeopardize the privacy of the individuals that have supplied the information used. If individuals think that there is a considerable chance that their privacy could be violated, they will not be willing to supply their private information anymore. So far nothing new!
5. However more and more users of tabular data want more detailed information. This is a problem, since the availability of more detail in tables increases the chance of extracting information about individuals. The more detail is given, the more important confidentiality protection becomes. For more information on confidentiality protection in general we refer to Willenborg and De Waal (2001).
6. If time and energy is put in elaborate preparation before statistical disclosure control is applied, more detailed information can be published without compromising confidentiality (as will be shown in this paper).

IIb. WHY USE A TOOL LIKE t-ARGUS?

7. Protecting tabular data against disclosure is an inevitable part of statistics. It is not easy to protect tabular data against disclosure, because of the several dimensions in the table that should be considered. Since the development of automated cell suppression software, the design of a useful suppression pattern is no longer a time consuming (and error prone) activity. It is also possible to calculate alternative suppression patterns.
8. To show the complexity of protecting tabular data and the necessity of automated cell suppression software a few examples will be given successively. See also Van der Meijden and Schalen (2004).

Table 1: Turnover of business sector X in a region is completely produced by one enterprise.

Region	State “Groningen” (NUTS 2 level)	State “Friesland” (NUTS 2 level)	State “Drente” (NUTS 2 level)	North-Netherlands (NUTS 1 level)
Turnover	40 (enterprise A)	60 (enterprise B, C en D)	-	100

9. According to τ-Argus the cells in “Groningen” and “Friesland” (Table 1) are primarily confidential because both cells have less than four contributors to the turnover (see also Section IIc). τ-Argus can also “see” that the enterprise in “Groningen” is able to recalculate the turnover in “Friesland” based on the total turnover in “North Netherlands” and the enterprise’s own turnover. Since the turnover in “Friesland” must stay confidential, the enterprise in “Groningen” should not be able to disclose this information. Therefore τ-Argus will also make “North Netherlands” secondarily confidential. See also Salazar-González (2004), where the mathematical model underlying τ-Argus is explained.

10. τ-Argus must also protect against recalculation of “North Netherlands” based on other dimensions (Table 2). See also Giessing (2001) and Hundepool (2001). When taking into account the other dimensions of a table, more secondary suppressions will occur than expected at first sight.
11. τ-Argus takes two steps in order to prevent disclosure of “North Netherlands”:

1- Another region in the same business sector will be made confidential.
2- Another business sector in the same region will be made confidential.

Table 2: The secondarily confidential region “North Netherlands” can be disclosed if t- Argus *does not make another region as well as another business sector (or totals) secondarily confidential.*

	North	East	South	West	Total
Business sector X	Secondarily confidential	300	100	400	1000
Business sector Y	200	…	…	…	…
Business sector Z	500	…	…	…	…
Total	800	…	…	…	…

12. When all the different dimensions (for example NACE, size class and region) in the table are made confidential sequentially, the chance exists that the resulting table is still not completely safe. Therefore, τ-Argus calculates the confidentiality effects on all dimensions simultaneously. See also Van der Meijden and Schalen (2004) and De Wolf (2002).
13. From the simple example mentioned above it can be concluded that it is not easy to protect more complex sets of tabular data. It can also be concluded that a cell with one contributor (singleton) has effects on confidentiality in other dimensions that cannot always (easily) be overseen. Especially when working on large tables with three or more dimensions, statistical disclosure control becomes difficult. See also Feuvrier and Faes-Cannito (2003). A tool like τ-Argus does oversee all the effects and eases the task of protecting the confidentiality of respondents.

IIc. CONFIDENTIALITY RULES IN t-ARGUS

14. τ-Argus offers four confidentiality rules, of which three are “dominance” rules and one is a minimum frequency rule. The “dominance” and frequency rules work independent of each other and determine together how the primary suppressions are selected. The frequency rule states how many contributors a cell must have in order to be safe. A frequency unsafe cell thus has fewer contributors than what is stated as safe. The “dominance” rules are:

1. (*n*,*k*)-rule (also called the dominance rule) – this rule states that a cell is primarily unsafe if a number of *n* contributors is responsible for more than *k* percent of the total value of that cell.
2. *p*%-rule (*p* percent rule) – This rule states that a cell is primary unsafe if an individual contribution can be recalculated within p percent of the actual value.
3. *p*-*q*-rule (prior-posterior-rule) – this rule takes into consideration the prior knowledge of the magnitude of the different contributions. The assumption is made that a contribution is a priori known with a margin of *q* percent. A cell is primary unsafe if an individual contribution can be recalculated within *p* percent of the actual value (*q* is therefore larger than *p*!).

15. For more information about the confidentiality rules in τ-Argus see Loeve (2001).

III. TUNING POSSIBILITIES FOR τ-ARGUS

IIIa. GENERAL

16. The statistical disclosure control provided by τ-Argus is based on input from four different domains. Tuning of cell suppression patterns is therefore only possible by making changes to these four different domains. See Van der Meijden et al (2004).

17. The four different domains that τ-Argus uses as input are:

1. Hierarchies: The way hierarchies are built is of influence on how secondary suppressions are applied.
2. History file: A preference can be given for which cells may or must be secondarily confidential.
3. Information loss weights: Information will be lost when applying secondary suppressions. The way τ-Argus calculates this information loss can be adjusted.
4. Base material: The way the microdata and preferred output are composed is of influence on the way secondary suppressions are applied.

18. The order of the above mentioned domains indicates at which domains generally the best results are expected when making changes to the four domains. The four different domains will be discussed successively in the next chapters.

19. For more information about τ-Argus one should read the user guide of Hundepool et al (2004).

IIIb. HIERARCHIES

20. τ-Argus needs to know how the hierarchies of classifications (for example NACE, size class and region) in the base material are constructed in order to choose counterparts for primary suppressions, that is secondary suppressions.

21. It is easier to determine secondary suppressions within a category of a hierarchy (and not influence the higher levels of the classification) if there are more subcategories within the category. Furthermore “singletons” (cells with one contributor) have less influence within a category if τ-Argus can find enough counterparts in that category. In Figure 1 is shown what is meant by more subcategories. Especially the hierarchical parts “A-0”, “C-4” and “D-5/6” in the “narrow” classification may give problems (that is more primary and secondary suppressions at higher hierarchical levels) because there are no or not enough subcategories that can act as counterpart for primary suppressions.

22. It is difficult to find enough “counterbalance” for the primary suppressions when using a “narrow” hierarchy (a break down in only one or two subcategories within a category of a hierarchy). As a result τ-Argus will apply secondary suppressions at a higher level within the hierarchy.

23. In order to keep the secondary suppressions at the same level where the primary suppressions occur, the hierarchy has to be “widened”. This is only possible in ordinal classifications like NACE and size class. A change in an ordinal classification can be carried out without influencing the existing obligations for publication.

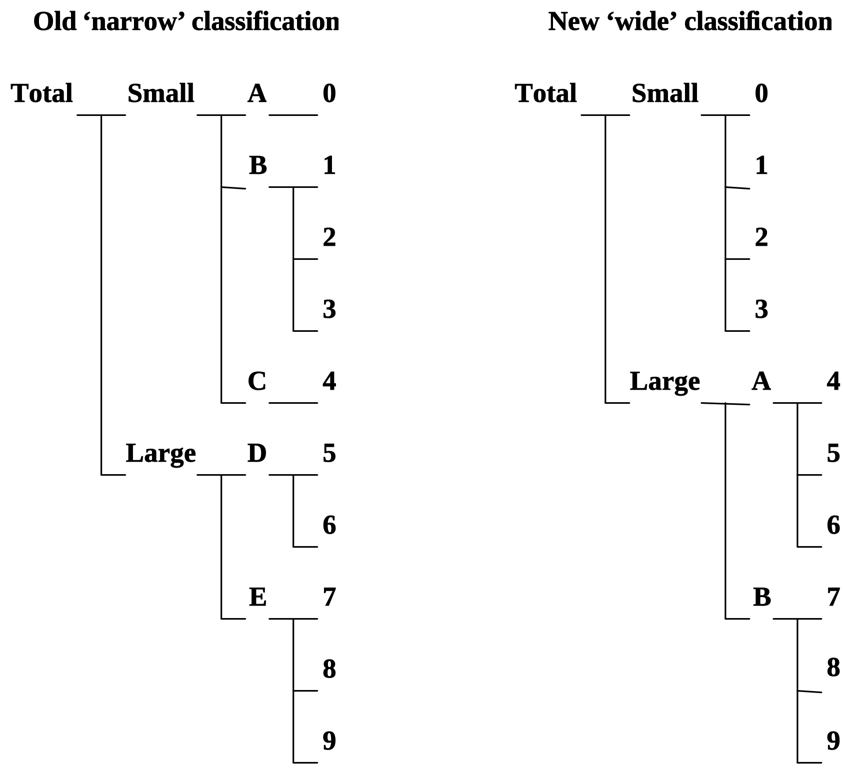

Figure 1: A rearrangement of subcategories within a size class classification.

24. If it is possible to make changes to an ordinal classification, a “widening” of the classification can supply for enough counterbalance for primary suppressions within a category of a hierarchy. If and how a classification can be “widened” depends (of course) on the obligations for publication and the type of classification. See also Westlake (2003).

25. The results (and differences) of using a “narrow” or “wide” classification (Figure 1) can be seen in Table 3. In this table we used the narrow and wide hierarchies from Figure 1 in order to calculate the number of primary and secondary sensitive cells for NACE “Industry”. Notice that the percentage of frequency unsafe and secondary unsafe cells has been reduced in the higher levels of the hierarchy when using the “wide” hierarchy. Also notice that the percentage of safe cells in the higher levels of the hierarchy has increased. In general a “wide” hierarchy also requires less secondary suppressions than a “narrow” hierarchy.

26. Because the NACE classification is an international standard, it is not desirable and possible to make changes to this classification. However, it is possible to harmonise the lowest level of the NACE classification that is used with the publication obligations in order to minimize the number of suppressed cells.

27. When more detail is needed, a lower hierarchical level within a classification is used. However, because the lower hierarchical level will contain smaller cells the chance becomes higher that the number of enterprises in a cell will be below the threshold of the minimum number of enterprises in a cell (for frequency sensitivity). Therefore more detail in the lower hierarchical level can result in more primary (and indirectly in more secondary) suppressions.

28. When more detail in a lower hierarchical level is combined with a “narrow” classification it may result in secondary suppressions on a higher hierarchical level than the level used. The result is exactly contrary to the desire to show more detail.

Table 3: The number and percentage of primary and secondary confidential cells as well as safe cells when using different (narrow or wide) hierarchies of the size class classification. The numbers in bold italics are the most interesting changes when using a wide hierarchy.

Status	**size class total**		**size class S - L**		**size class A - E**		**size class 1 - 9**		**total**	
narrow size class	**# cells**	**%**	**# cells**	**%**	**# cells**	**%**	**# cells**	**%**	**# cells**	**%**

A frequency unsafe	3863	32,8	8383	42,0	16951	52,9	33005	61,4	62202	52,9
B dominance unsafe	577	4,9	295	1,5	164	0,5	182	0,3	1218	1,0
C history file	11	0,1	0	0,0	0	0,0	0	0,0	11	0,0
D secondary unsafe	3202	27,2	6284	31,5	8035	25,1	9635	17,9	27156	23,1
V safe	4130	35,1	5018	25,1	6896	21,5	10948	20,4	26992	23,0
Total	11783	100	19980	100	32046	100	53770	100	117579	100

Status wide size class	**size class total**		**size class S - L**		**size class A - B**		**size class 1 - 9**		**total**	
	# cells	**%**	**# cells**	**%**	**# cells**	**%**	**# cells**	**%**	**# cells**	**%**
A frequency unsafe	3863	32,8	8325	***41,1***	12859	***49,3***	33005	61,4	58052	***51,9***
B dominance unsafe	577	4,9	333	1,6	182	0,7	182	0,3	1274	1,1
C history file	11	0,1	0	0,0	0	0,0	0	0,0	11	0,0
D secondary unsafe	3128	***26,5***	6369	31,5	7410	28,4	9735	18,1	26642	23,8
V safe	4204	***35,7***	5217	***25,8***	5627	***21,6***	10848	20,2	25896	***23,1***
Total	11783	100	20244	100	26078	100	53770	100	111875	100

29. Sometimes it is better to use a higher hierarchical level within a classification because it can result in showing more information/detail than when using a lower hierarchical level. The reason for this is less (or even no!) occurrence of primary and secondary suppressions at both the same and higher hierarchical levels as used. See also De Wolf and Mulder (2002).

30. Consider, for example, the Transport statistics (NACE 634); this is a very “narrow” classification. NACE 634 only consists of NACE 6340 which in its turn consists of NACE 63401 and 63402 (see Figure 2). For the annual Transport statistics at Statistics Netherlands one GK-NACE combination at the fifth digit NACE-level was primarily confidential, while there was not enough counterbalance at the fifth digit NACE-level, which in combination with the “narrow” classification resulted in secondary suppressions at a higher level; in this case even up to the third digit NACE. If NACE 63402 is primarily confidential, while NACE 63401 cannot provide for enough counterbalance, NACE 6340 and NACE 634 will become secondarily confidential. Publishing on the fourth digit NACE, provided that it is not generally known that most of the turnover is in NACE 63402, will result in safe cells.

31. Publishing at the fourth digit NACE instead of the fifth digit does not lead to disclosure of any sensitive information. Thus it can be beneficial to publish on a higher hierarchical level because of less primary and secondary suppressions in the higher hierarchical levels when using τ-Argus.

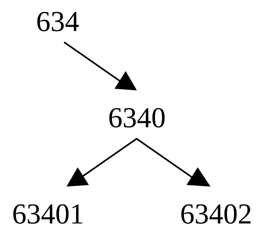

Figure 2: A part of the NACE for Transport.

32. It can be concluded that it is beneficial to tune the hierarchies of classifications. When classifications are "widened", the effect of "singletons" will be diminished. Therefore secondary suppressions at higher hierarchical levels will not exist anymore. If classifications cannot be "widened", publishing on a higher hierarchical level can result in less suppressions in the higher hierarchical levels and thus possibly more information is available.

IIIc. HISTORY FILE

33. The statistician can adjust the status of confidentiality of cells by hand in the history file. Originally the history file was meant for τ-Argus to take into account the suppression pattern of former years (hence the name history file). See also De Wolf (2003). However, the history file can also be misused for directing the confidentiality pattern towards the wishes of the statistician.

Table 4: The possible statuses that can be assigned to a cell.

Status	Meaning of the status
G	Confidential, so suppress primarily.
P	Publishable, so do not suppress primarily.
L	Preferably do not suppress secondarily.
R	Preferably do suppress secondarily.

34. The history file can be used for both giving a preference for suppressing and not suppressing a cell (Table 4). Through the history file both the primary and secondary suppressions can be directed. τ-Argus tries to grant the settings in the history file as much as possible.

35. However, there are some rules when assigning status codes:

1) Cells without any contributors cannot get a status by hand.
2) Codes 'G', 'L' and 'R' can only be applied on cells that are primarily safe.
3) Code 'P' can only be applied on cells that are primarily unsafe.

36. The retail trade for example has used the history file in an optimal way. The "department stores" were primarily confidential and because of their minor total turnover the "retail sale of antiques" could not be a proper counterpart for secondary suppression. Therefore instead of the not so important "retail sale of antiques" some important cells became secondarily confidential. With the help of a subject matter expert in retail trade the history file was complemented with a list of cells (NACE/size class combinations) that could be used for secondary suppressions in order to get enough turnover as counterpart for the primarily confidential cells. Some important cells were also given a somewhat "protected" status by giving them the status "L". Because of the use of the history file the important cells stayed publishable.

37. Although the history file's primary functionality is to take into account the confidentiality pattern of former years, it can also be used to direct τ-Argus in choosing both primary and secondary suppressions.

IIId. INFORMATION LOSS

38. Secondary suppressions are inevitable in order to make it (almost) impossible to retrieve the information from cells that are primary confidential. The results of the method used for determining the secondary suppressions depend on how information loss is measured.

39. τ-Argus will try to keep the information loss at a minimum while finding the cells for the secondary suppressions. In order to keep the information loss at a minimum τ-Argus must know how the information loss should be measured. For that purpose every cell in the table must be assigned a value which reflects the amount of information in that cell.

40. τ-Argus offers the following standards for measuring information loss:

<u>Cell value</u>: the amount of information in a cell equals the cell value. The consequence of minimizing this information loss is that cells with a higher value will be used less frequently as secondarily confidential cells than cells with a lower value.

<u>Frequency</u>: the amount of information in a cell equals the number of contributors to that cell. Minimizing this kind of information loss results in secondary suppressions applied on cells with fewer contributors.

<u>Equal</u>: the amount of information in a cell is exactly the same for each cell. Minimizing this kind of information loss results in producing the smallest number of cells that are

suppressed secondarily. The consequence of this method is that important cells are suppressed as easily as unimportant cells.
Distance: the amount of information in a cell is related to the distance of that cell to a primarily confidential cell. Minimising this kind of information loss results in a clustering of secondarily confidential cells around the primarily confidential cells.

Table 5: The number and percentage of primary and secondary confidential cells as well as safe cells when using different methods for determining information loss. The numbers in bold italics are the most interesting differences when using another method.

	Methods for determining information loss							
	Cell value				**Frequency**			
Status	2nd digit NACE	3rd digit NACE	4th digit NACE	5th digit NACE	2nd digit NACE	3rd digit NACE	4th digit NACE	5th digit NACE
A frequency unsafe	0	195	2686	7995	0	195	2686	7995
B dominance unsafe	0	26	216	641	0	26	216	641
D secondary unsafe	4	321	2837	6806	4	***298***	***2638***	***6423***
V safe	285	1382	4813	8351	285	***1405***	***5007***	***8724***

41. It should be clear that the different information loss measures for obtaining secondary suppressions result in different suppression patterns. In Table 5 the results of the "cell value" and "frequency" methods for determining information loss are shown for different NACE levels. The results are for the wholesale business. The experience is that it is very useful to try the different information loss measures in order to get the best possible way of secondary suppression.

IV. TUNING BASE MATERIAL

IVa. SMALL AREA ESTIMATION

42. Statistics Netherlands has worked on improving their method for small area estimation of business statistics. See Witvliet and Heerschop (2005). This improved method especially improves the small area estimates for enterprises in size classes 3 – 5. This new methodology improves the estimates of the turnover of enterprises that have not been surveyed, which results in more null cells as well as a better distribution of the turnover. The better turnover estimates and better distribution of turnover result in less sensitive cells according to the dominance rule. The better distribution also results in more counterbalance for primary suppressions. Results and differences of using the old and new small area estimation method on confidentiality can be seen in Table 6.

Table 6: The number and percentage of primary and secondary confidential cells as well as safe cells when using different small area estimation methods for the wholesale business. The numbers in bold italics are the most interesting changes when using the new small area estimation method.

	Region					
Status	**NL total**		**Country parts** NUTS level 1		**Provinces** NUTS level 2	
	#cells	**%**	**#cells**	**%**	**#cells**	**%**
	Old method					
A frequency unsafe	1054	23,7	4589	31,3	14950	43,0
B dominance unsafe	44	1,0	171	1,2	471	1,4
D secondary unsafe	1055	23,8	4390	30,0	9530	27,4
V safe	2287	51,5	5495	37,5	9853	28,3
	New method					

A frequency unsafe	555	***12,5***	3758	***23,7***	14558	***36,4***
B dominance unsafe	46	1,0	362	2,3	1116	2,8
D secondary unsafe	639	***14,4***	4508	***28,4***	11627	29,0
V safe	3195	***72,0***	7256	***45,7***	12731	***31,8***

IVb. DELIBERATELY ADJUSTING MICRODATA

43. First of all it should be clear that adjusting microdata is not the normal way how statistics are made. However, extraordinary situations can occur in which adjustment of microdata could be justifiable.
44. We give an example of a situation in which adjustment of microdata could be acceptable. Suppose an enterprise of a certain NACE just moved from size class 3 to size class 4 and crossed a “crucial” border (Figure 3). Also suppose that there are no more enterprises present within size classes 4 – 9 for that specific NACE.

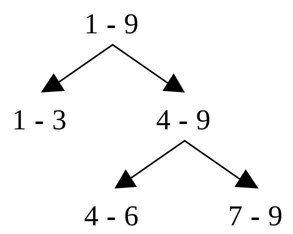

Figure 3: An example of a size class classification hierarchy.

45. Because only one enterprise is present in size class 4 – 6, this size class will become primarily confidential. Since size class 7 – 9 is completely empty, size class 4 – 9 is also primarily confidential. There are two possible options for the secondary suppression: size class 1 – 3 or size class 1 – 9 must become secondarily confidential. What option is used by τ-Argus depends on what the effects are on the other dimensions that are present.
46. To prevent the primary and secondary suppressions and subsequent effects on other dimensions, a recode of that single enterprise from size class 4 to size class 3 could be applied.

V. COORDINATION OF PUBLICATION OBLIGATIONS

47. Although this section is not about tuning base material or using specific features in τ-Argus, it is about improving confidentiality by clever usage of the microdata and preferred output. Statistical offices are working so hard on developing new methodologies and software tools for disclosure control that one almost forgets that thinking about confidentiality starts way before a publication is made.
48. Thinking about confidentiality starts when making agreements with customers about what information to deliver. What are the publication obligations? What obligations are enforced (for example by Eurostat and government)? And what obligations are agreed on with specific customers? Obligations that are agreed on can be changed in order to improve the total amount of information released!
49. How often does it happen that agreements with customers about publishing certain groupings of, for example, size class do not correspond with the groupings that are enforced? The more agreements are made about different groupings of variables, classifications, etcetera, the bigger the confidentiality problem becomes. See also De Wolf et al. (2002).
50. The Industry statistics at Statistics Netherlands, for example, are made for different customers (Eurostat and EIM – Economical Institute for small to Medium-sized enterprises). For both customers another size class classification is used. Because EIM is especially interested in SME (small to medium-sized enterprises), the size class categories are a little different from that of Eurostat. The result is that the different size class classifications are partly overlapping, which is

especially limiting the number of cells that can be published without disclosing information. Although the customer is happy to get the exact classification that they wanted; they also get a lot of unwanted disclosed information. The question is whether the customer is not better off with the other classification but with less disclosed information!

51. Using different classifications or variables for different customers leads to inevitably much unnecessary confidentiality for all customers, thus not only for the customer who wants something else than other customers. Because the publications for EIM are agreed upon, they should be brought into accordance with the enforced obligation of Eurostat in order to minimize the confidentiality and improve the amount of information released for all customers.

52. If there is more cooperation between those responsible for statistical disclosure control and those who make the agreements with customers about specific tables that are produced, a lot of measures to avoid disclosure can be limited or will even not be needed. See also Samuelson (2001).

VI. REFERENCES

De Wolf, P.-P. (2002), *HiTaS: A heuristic approach to cell suppression in hierarchical tables*, Inference Control in Statistical Databases, Springer-Verlag, Berlin Heidelberg, pp. 74 – 82.

De Wolf, P.-P. (2003), *Guideline t-Argus interface for business statistics* (internal note), Statistics Netherlands, Voorburg.

De Wolf, P.-P., Hundepool, A., Loeve, A. and De Waal, T. (2002), *Securing tabular data, now and in the future* (internal note), Statistics Netherlands, Voorburg.

De Wolf, P.-P. and Mulder, A. (2002), *Comparison of different table protection tools* (internal note), Statistics Netherlands, Voorburg.

Feuvrier, P. and Faes-Cannito, F. (2003), *Cell suppression in Eurostat on structural business statistics – an example of statistical disclosure control on tabular data*, paper presented at the joint ECE / Eurostat work session on statistical data confidentiality, Luxembourg, 2003.

Giessing, S. (2001), *New tools for cell suppression in tau-Argus: one piece of the casc project work draft*, Federal Statistical Office of Germany, paper presented at the joint ECE / Eurostat work session on Statistical Data Confidentiality, Skopje, 2001.

Hundepool, A. (2001), *The CASC project*, Statistics Netherlands, paper presented at the joint ECE / Eurostat work session on Statistical Data Confidentiality, Skopje, 2001.

Hundepool, A., Van de Wetering, A., Ramaswamy, R., De Wolf, P.-P., Giessing, S., Fischetti, S., Salazar, J.J., Castro, J. and Lowthian, P. (2004), *t-Argus user's manual version 3.0.*

Loeve, J.A. (2001), *Dominance rule versus P-Q rule* (internal note), Statistics Netherlands, Voorburg.

Salazar-González, J.-J. (2004), *Mathematical models for applying cell suppression methodology in statistical data protection,* European Journal of Operations Research 154, pp. 740-754.

Samuelson P. (2001), *Statistics – a balance between official and confidential data*, Ministery of Justice, Sweden.

Van der Meijden, R., Schalen, J., De Wolf, P.-P. and Hundepool, A. (2004), *Study results and guidelines for tuning t-Argus* (internal note), Statistics Netherlands, Voorburg.

Van der Meijden, R. and Schalen, J. (2004), *Confidentiality issues for t-Argus* (internal note), Statistics Netherlands, Voorburg.

Westlake, A. (2003), *Security and disclosure for statistical information*, Survey & statistical computing, London.

Willenborg, L. and De Waal, T. (2001), *Elements of Statistical Disclosure Control*, Lecture Notes in Statistics 155, Springer-Verlag, New York.

Witvliet, H. and Heerschop, M.J. (2005), *small area estimations in ESB* (internal note), Statistics Netherlands, Voorburg.

VI. ACKNOWLEDGEMENT

I would like to thank John Schalen for answering al my questions and for helping me with producing the data for the tables that I wanted in this paper.