JRC Scientific and Technical Reports

The assessment and revision of the public procurement Indicator 3: Total value of published tenders.

Athina Karvounaraki
With contributions from Martin Hradisky, William Castaings and Stefano Tarantola

EUR 23695 EN - 2008

JRC
EUROPEAN COMMISSION

ipsc
Institute for the Protection and Security of the Citizen

The Institute for the Protection and Security of the Citizen provides research-based, systems-oriented support to EU policies so as to protect the citizen against economic and technological risk. The Institute maintains and develops its expertise and networks in information, communication, space and engineering technologies in support of its mission. The strong cross-fertilisation between its nuclear and non-nuclear activities strengthens the expertise it can bring to the benefit of customers in both domains.

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen

Contact information
Address: Via E. Fermi,2749
E-mail: athina.karvounaraki@jrc.it
Tel.: +39 0332 786190
Fax: +39 0332 785733

http://ipsc.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC 49565

EUR 23695 EN
ISSN 1018-5593

Luxembourg: Office for Official Publications of the European Communities

Printed in Italy

LIST OF CONTENTS

Executive summary

The main objective of this study is to assess and revise the existing Indicator 3 that estimates the total value of published tenders. This indicator is currently produced by DG MARKT annually. The work performed by DG JRC focuses on the analysis of the quality of the available data, and the assessment of the methodology used for the calculation of the indicator.

Detailed exploration of the data recorded during year 2005, highlighted the lack of consistency in the way some information is recorded and the existence of non informative variables, which should be eliminated.

The analytical break down of the database revealed that less than 24% of the information stored is used for the calculations of indicator 3. During this exercise, we examined how Contract Award Notices (CANs) are distributed among Member States in terms of number and value involved and we noticed a significant difference in the reporting rate.

The current methodology for the calculation of the total value of published tenders for each Member State and year of interest consists of three main components:

1. The average value of all CANs below 100 million euros for each sector (works, supplies, services)
2. The number of published calls (CALLs) in each sector
3. The total value of the large CANs , (above 100 million euros) addressed as outliers.

By looking at each of these components, we identified the limitations of the current method and we proposed some alternatives for the improvement of this indicator. Our efforts focused on three Member States, Great Britain, Spain and Finland, and the results are presented in this document. Two main issues were raised: the lack of correspondence between the CANs and the CALLs of the given year and the appropriateness of the arithmetic mean as a measure of location due to the skewness of the distribution of the CANs towards higher valued contracts.

Reclassification of the contracts according to their product code and the use of trimmed means, were the techniques applied so as to increase the homogeneity of the samples and to reduce the bias.

In order to understand to what extent the mismatch between CANs and CALLs influences the estimated amount of published tenders we developed an approach that looks back through the data in time according to which the calculation of indicator 3 for a given year is based on the CANs directly linked to the CALLs of that year. The CANs may have been published up to three years after the publication of the CALL and for this reason we had to use all available data from previous years.

1. Data description and quality assessment

We initially explored the 2005 data on Public Procurements. The dataset contains 243065 records and 73 variables of which 29 were numeric, one hyperlink imported as text, and the rest text. Variables containing free text, such as ActPXLine and EstPXLine haven't been systematically analyzed. Our findings can be summarized as follows.

- There are six completely empty text variables.

Obs	VARIABLE	DESCRIPTION
1	AB	Abstract
2	COCY	Country of Successful Contractor
3	MS	Member State offering Contract
4	OT	Original Text
5	OX	
6	PG	Page of official Journal

- Often there is not one to one correspondence between the numeric codes and the description of the codes. See for example Variables AA (Description of type of Awarding Authority) and AACODE (Numeric code for the type of Awarding Authority). The same situation appears between the variables: NC and NCCODE, PR and PRCODE, RP and RPCODE, TD and TDCODE.

	AACode									
	1	3	4	5	6	8	9	N	R	Total
AA										
Body governed by public law	1	0	0	1	47383	0	0	0	0	47385
Central government	26870	0	0	0	0	0	0	0	0	26870
EC institutions	0	0	0	3817	0	0	0	0	0	3817
Local authorities	0	85828	0	0	0	0	0	0	0	85828
National or federal Agency/Office	0	0	0	0	0	0	0	52	0	52
Not applicable	0	0	0	0	0	0	111	0	0	111
Other	1	0	0	1	0	30569	0	0	0	30571
Regional or local Agency/Office	0	0	0	0	0	0	0	0	67	67
Regional or local authority	0	18838	0	0	0	0	0	0	0	18838
Utilities	0	0	23165	0	0	0	0	0	0	23165
Missing	5784			575						6361
Total	32656	104666	23165	4392	47383	30569	111	52	67	243063

- The same currency or the same official language is represented by two categories most probably due to typing errors. For example, EUR and EURO are two different categories in the variable ESTPXCURRENCY.

<ESTPXCURRENCY>								
ALPHABETICALLY SORTED				SORTED BY FREQUENCY				
NUM	CATEGORY	#	%	CATEGORY	#	%	CUM #	CUM %
1		207026	85.17		207026	85.17	207026	85.17
2	CYP	1	0.00	EUR	24382	10.03	231408	95.20
3	CZK	5	0.00	PLN	4966	2.04	236374	97.25
4	DKK	370	0.15	GBP	2589	1.07	238963	98.31
5	DKR	2	0.00	LTL	1420	0.58	240383	98.90
6	EEK	20	0.01	SEK	764	0.31	241147	99.21
7	EUR	24382	10.03	DKK	370	0.15	241517	99.36
8	EURO	12	0.00	HUF	364	0.15	241881	99.51
9	GBP	2589	1.07	åõñþ	312	0.13	242193	99.64
10	HUF	364	0.15	SKK	307	0.13	242500	99.77
11	LTL	1420	0.58	SIT	291	0.12	242791	99.89
12	LVL	201	0.08	LVL	201	0.08	242992	99.97
13	MTL	28	0.01	MTL	28	0.01	243020	99.98
14	PLN	4966	2.04	EEK	20	0.01	243040	99.99
15	ROL	1	0.00	EURO	12	0.00	243052	99.99
16	SEK	764	0.31	CZK	5	0.00	243057	100.00
17	SIT	291	0.12	STERLING	3	0.00	243060	100.00
18	SKK	307	0.13	DKR	2	0.00	243062	100.00
19	STERLING	3	0.00	CYP	1	0.00	243063	100.00
20	TRY	1	0.00	ROL	1	0.00	243064	100.00
21	åõñþ	312	0.13	TRY	1	0.00	243065	100.00

- The data set contains four date variables i.e. PD, DS, DR and DT all of which were imported as text variables due to problems in their formatting. In fact, we have noticed that dates are recorded in one of the two following formats: YYYY-MM-DD and DD/MM/YYYY. In all 4 variables the first format is more frequent covering over 80% of the cases.

VARIABLE	FORMAT 1 YYYY-MM-DD		FORMAT 2 DD/MM/YYYY		TOTAL	MISSING
	Freq	%	Freq	%		
PD	204168	84.0	38897	16.0	243065	0
DS	204167	84.0	38897	16.0	243064	1

DR	204167	84.0	38897	16.0	243064	1
DT	127318	86.5	19795	13.5	147113	95952

- By looking the frequency table of the existing categories we have noticed that there are Awarding Authorities with very similar names. In reality, it is the same Authority recorded in a slightly different way. For example:

  "Dimosia Epicherisi Ilektrismoy AE" with frequency 95 and

  "Dimosia Epicherisi Ilektrismoy (DEI) AE with frequency 86.

- The same situation has been seen in the variable TW (Town of Awarding Authority). For example ***"Milan" appears in 1345 records and "Milano" in 7.***
- In total there are 29 numeric variables of which nine are Control variables, ('0', '1' type e.g. SW_TPX) and five are Code Variables(e.g. ABCode1) containing alphanumeric codes of 7 or 8 digits. All of them could have been recorded as text variables similar to PC (Product Code). Moreover, there are two non informative variables (StartingLineOrigin, empty and Class, only zeros)
- In the majority of numeric variables zero value represents missing information. There are only two exceptions: variables ActPX and EUPX that identify missing values as such. Nine of the numeric variables contain prices

Variable	**Missing**		**Zeros**		**Positive**		**Total**
	Freq	***%***	***Freq***	***%***	***Freq***	***%***	
EstPX	0	0	205952	85	37113	15	243065
EstEUPX	0	0	207209	85	35856	15	243065
EstLowPX	0	0	240361	99	2704	1	243065
EstHiPX	0	0	240244	99	2821	1	243065
ActPX	193022	79	769	0	49274	20	243065
ActEUPX	0	0	210915	87	32150	13	243065
LowPX	0	0	235260	97	7805	3	243065
HiPX	0	0	235241	97	7824	3	243065
EUPX	167554	69	16887	7	58624	24	243065

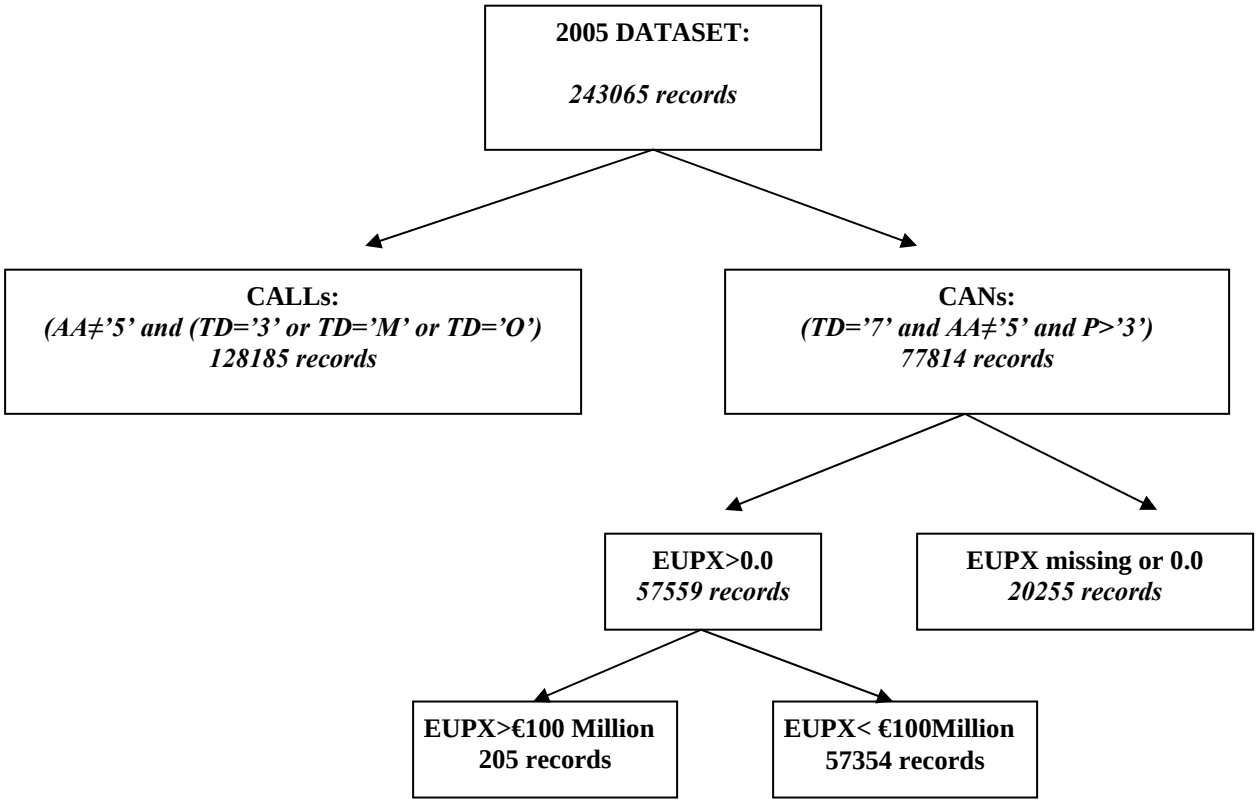
- There are cases where the Estimated (EstPX) or the Actual Price of Contract, (ActPX) is positive (3.4% and 35% respectively) and the Estimated or the Actual Price Converted to Euros (EstEUPX, ActEUPX) is zero.

- There are cases (approx. 3%) where there is an Estimated or an Actual Price of the Contract but the Currency of the Contract is missing. The price line from the original document shows that in most of these cases the amounts recorded are not in Euros and therefore, conversion is required.
- When we tried to calculate simple statistics for the numeric variables expressed in Euros (e.g. mean, standard deviation, range) we realized that there are errors in the reported prices that affect seriously such calculations. See for example the following two cases of the variable EUPX.

ND	CY	NCCode	RPCode	EstEUPX	EstPXLine	ActEUPX	ActPXLine	EUPX
100314-2005	BE	4	3	4	Estimated total value (excluding VAT): For the 4-year contract term: EUR 3 200 000 (contract for EUR 800 000 per year renewable annually).	0	V.1.2) Information on value of contract or on highest and lowest tenders taken into consideration (excluding VAT): Value: EUR 1 515 121,60.	4
101313-2005	GB	4	3	4	Estimated total value: The estimated contract value for both lots 1 and 2 over a 4-year period is EUR 186 953,08.	0		4

2. Data exploration

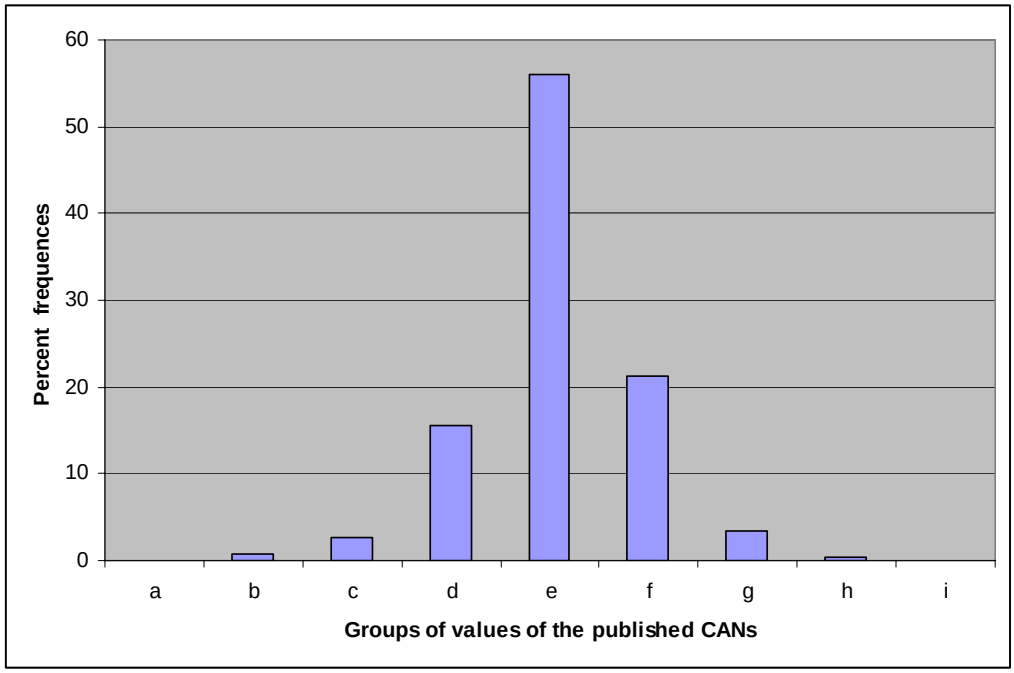
The structure of the 2005 dataset is presented in the chart below:



- It can be seen that 32% of the database refers to Contract Award Notices and only 74% of these contain the value which means that our calculations will be based on less than 24% of the entire database. Similarly, the published Calls for Tenders represent almost 53% of the database.
- There are only 205 contracts (≈0.4%) above the threshold of €100 Million. The sum of their amounts corresponds to 36.5% of the total amount of the contracts awarded in 2005.
- There are 1197 small contracts, of a price less than 5000 Euros (2.1% of the total number of contracts with a reported amount). Over 50% of these small contracts come from Lithuania (LT). Only 3 of the 449 contracts (value <1000 Euro) concern Works.

The distribution of the contracts according to their value can be seen in the table and the graph below.

classest	Frequency	Percent
(10, 1E+2]	5	0.01
(1E+2,1E+3]	444	0.77
(1E+3, 1E+4]	1536	2.67
(1E+4, 1E+5]	8922	15.50
(1E+5, 1E+6]	32241	56.01
(1E+6, 1E+7]	12189	21.18
(1E+7, 1E+8]	2019	3.51
(1E+8, 1E+9]	200	0.35
>1E+9	3	0.01
Total	57559	100.00



a =10 -100 €
b =100 -1000 €
c =1000 -10000 €
d =10 -100 K€
e =100-1000 K€
f = 1000 -10000 K€
g = 10 -100 M€
h = 100 -1000 M€
i = >1000 M€

- Over 60% of the Finish contracts do not report prices.
- Contracts for Works are the fewest, only 17% but the most expensive as they cover almost 40% of the total Public Procurements.

Contracts Awards				
NCCode	Num	%	Amount in million €	%
1 = works	10025	17.4	63214.7	39.8
2 = Supplies	23053	40.1	36540.5	23.0
3 = combined	2	0.0	1.4	0.0
4 = services	24479	42.5	58906.9	37.1
TOTAL	57559		158663.5	

- The following table shows that 28% of the published contracts come from France. In terms of amounts, Spain comes first with total public procurements of approximately 30 billion Euros. Great Britain comes second and France third in the list.

CY	Num	%	Amount in million €	%
AT	884	1.5	1930.9	1.2
BE	685	1.2	2107.3	1.3
BG	4	0.0	2.0	0.0
CY	78	0.1	227.0	0.1
CZ	1369	2.4	1547.9	1.0
DE	6590	11.4	10550.4	6.6
DK	528	0.9	2024.3	1.3
EE	234	0.4	520.3	0.3
ES	5051	8.8	30013.7	18.9
FI	342	0.6	1620.8	1.0
FR	16110	28.0	22228.1	14.0
GB	3907	6.8	28183.7	17.8
GR	450	0.8	3340.3	2.1
HU	1052	1.8	4112.2	2.6
IE	331	0.6	818.6	0.5
IT	3996	6.9	22071.8	13.9
LT	1747	3.0	706.7	0.4
LU	126	0.2	85.4	0.1
LV	555	1.0	914.9	0.6
MT	47	0.1	24.7	0.0
NL	1322	2.3	3208.0	2.0
PL	9881	17.2	15130.4	9.5
PT	206	0.4	813.5	0.5
RO	1	0.0	9.7	0.0
SE	1195	2.1	3942.2	2.5
SI	350	0.6	450.9	0.3
SK	517	0.9	2077.8	1.3
TR	1	0.0	0.4	0.0
Total	57559	100	158663.6	100.0

- The Call for Tenders notices of each country are usually more than the contract awards. There are three exceptions: Slovakia (SK), Lithuania (LT) and Estonia (EE).
- Over 30% of the contracts refer to products under Chapters 45 and 74. In terms of amounts, these two chapters account for 50% of the total public procurements.

Note: For the evaluation of indicator 3 we have used the data of three more years, 2003, 2004 and 2006 of only three Member States, GB, ES and FI, without carrying out a similar analysis on the structure and the quality. Each of these MSs has a different level of quality in the reporting of the data. In particular, only 40% of the Finish CANs report a price, 64% of the British and 91% of the Spanish CANS report a price in the 2005 dataset.

3. Indicator 3: Total amount of published Tenders

3.1 Description of the current methodology

Indicator 3 provides for a given year and Member State an estimation of the total amount of the CALLs published in the official journal during that year. Given that no value is reported in the CALLs, indicator 3 has to be estimated using other information such as the values of the CANs.

The methodology used so far for the estimation of indicator 3 can be summarised as follows:
Usually, the estimation is based on the Contract Award Notices (CAN) published during the given year. In the case of smaller countries the calculations are based on all available values, including those of previous years.
So far we have seen that not all Contract Award Notices report the value. For example, for the year 2005, 58624 of the 79488 CANs (73.5%) contain a value.
When the Awarding Authority is one of the EC Institutions the contract is excluded from the calculations. Moreover, when the procurement is based on a Regulation concerning External Aid and European Development Fund or Community Institutions and International Organizations or TACIS and countries of Central and Eastern Europe, the contract is also excluded from the calculations. In 2005 exercise, due to these constrains the percentage of useable contracts dropped down to 72%. Finally, very large contracts are treated separately and they are considered as outliers.
For each Member State and Type of Contract (i.e. Works, Supplies, Services) the formula used for the calculation of indicator 3 is the following:

$$V_{ij} = \frac{Total\ CAN\ value - Total\ Outlier\ value}{Num\ of\ CANs - Num\ of\ Duplicate\ RN} * Tenders + Total\ Outlier\ value \qquad (1)$$

Where CAN= Contract Award Notices for which a price exists
Outlier value= Total value of the CANs above the threshold of 100 millionEuro
Duplicate RN= CANs with the same Reference Number.

The same method can be written in a more mathematical form as follows:
For each member state i and type of contract j we calculate the average $\overline{X}_{ij}$ and the and the indicator $V_{ij} = \overline{X}_{ij} * T_{ij} + U_{ij}$ where $i = 1..27$ and $j = 1..3$
The total value of Published Tenders for each member state according the above formula will be $V_i = \sum_{j=1}^{3} (\overline{X}_{ij} * T_{ij} + U_{ij})$ and for the whole EU will be $V = \sum_{i=1}^{27} \sum_{j=1}^{3} (\overline{X}_{ij} * T_{ij} + U_{ij})$

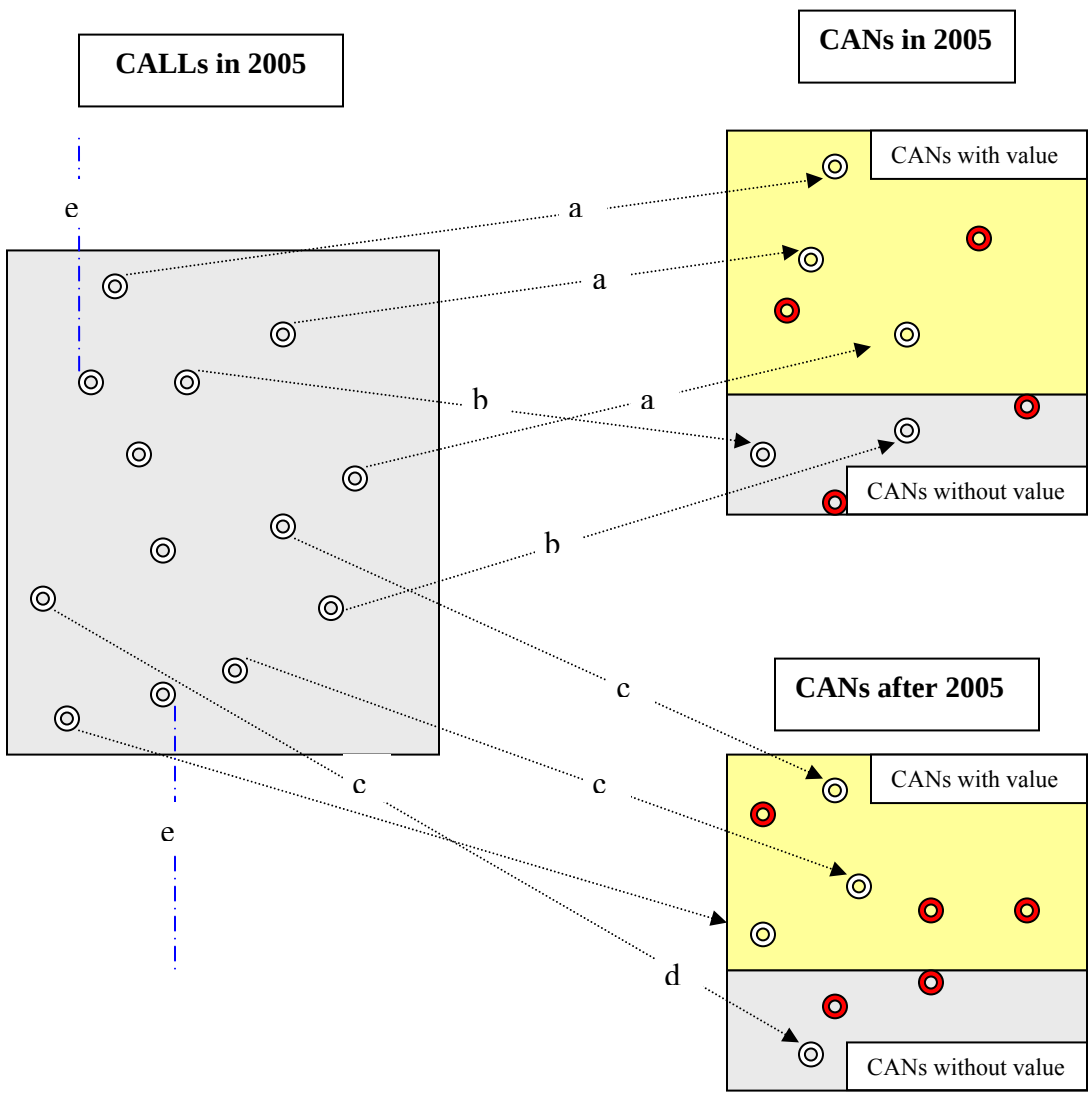
V_{ij} = Value of Published Tenders for each Member State and Type of Contract

X_{ij} = Average value of the CANs below 100 million Euro after duplicate RNs have been removed

T_{ij} = Number of Published Tenders for each Member State and Type of Contract

U_{ij} = Value of contracts above the threshold for each Member State and Type of Contract

One serious drawback of this method is the lack of correspondence between the CALLs and the CANs published a given year. The structure of the database and the relationship between the CALLs and the CANs is presented graphically below.



It can be seen that a CALL in 2005 may lead to a Contract Awarded in 2005 for which the value may (a)or may not be reported (b). A CALL published in 2005 may also lead to

a Contract not the same year but even 5 years after the initial publication (c and d). Finally, a CALL published in 2005 may not lead to any Contract, or may lead to a contract that has never and will never be published (e).
In a similar way the set of the CANs available for 2005 consists of contracts the CALLs of which have been published either in 2005 or previous years (in the graph these cases are represented by the red circles).
As it will be seen, this lack of correspondence between CALLs and CANs of the same year affects the estimation of Indicator 3.

3.2 Assessment of the current method

In the calculation of indicator 3 there are three components:

- The average value of all CANs below 100 million euros for each sector (works, supplies, services)
- The number of published calls (CALLs) in each sector
- The total value of the large CANs, (above 100 million euros) addressed as outliers.

Below we shall see how the indicator is affected by changes in these components.

We have seen that a CALL of a specific year may or may not lead to a contract the same year. In fact, our analysis has shown that depending on the country and the amount involved, less than half of the published CALLs correspond to contracts awarded the same year of the publication. Moreover, it is logical to assume that expensive contracts require more negotiations, more paper work and therefore more time to be completed. In the current methodology, when we add the total amount of the large CANs, which may have been published one to even five years before the current date, we assume that the total value of the large contracts remains more or less stable at MS level.
Both examples presented below demonstrate that this assumption does not hold. In the first case, Finland, the amount involved in large contracts increases between 2003 and 2005 and decreases significantly in 2006. In Spain there is a continues and progressive increase in the amounts involved in large contracts.

Finland

Year	**Works**		**Services**		**Total**	
	#	Amount	#	Amount	#	Amount
2003	0		1	120000000	1	120000000
2005	1	638000000	0		1	638000000
2006	1	103040000	0		1	103040000

Spain

Year	Works		Supplies		Combined		Services		Total	
	#	Amount	#	Amount	#	Amount	#	Amount	#	Amount
2003	10	2932629155	2	753227554	3	578682159	1	470655832	16	4735194700
2004	17	4476698591	3	688141417	0		2	949366283	22	6114206291
2005	26	5272138377	5	1177147448	0		3	7809404570	34	14258690395
2006	21	7916864140	17	30787861160	0		15	29517267156	53	68221992456

Due to the improvement in the reporting there is an increase in the numbers of CALLs published every year. This general tendency follows different patterns over the years depending on the Member State. Finland, for example, shows the highest increase between 2003 and 2004 whereas Spain has published significantly more CALLs in 2006. (See tables below)

Finland

Table of YR by NCCode					
YR(YR)	**NCCode(NCCode)**				**Total**
Frequency	**Works**	**Supplies**	**Combined**	**Services**	
2003	141	803	4	499	1447
2004	132	891	0	622	1645
2005	139	888	0	638	1665
2006	173	939	0	730	1842
Total	585	3521	4	2489	6599

Spain

Table of YR by NCCode					
YR(YR)	**NCCode(NCCode)**				
Frequency	**Works**	**Supplies**	**Combined**	**Services**	**Total**
2003	736	2191	70	2943	5940
2004	786	2187	0	3176	6149
2005	1113	2447	0	3714	7274
2006	1478	3215	0	4503	9196
Total	4113	10040	70	14336	28559

3.3 Alternative methodologies

3.3.1 Change in the classifications

In order to increase the ***homogeneity*** in each category and therefore to get more representative and robust averages we tried to classify both CALLs and CANs in chapters of products and in groups of products.
For the definition of the chapters we used the first two digits of the eight digit product code given in the variable PC.
Examples:

Product Code (variable PC)	Chapter	Description
34 100000	34	Motor vehicles, trailers and vehicle parts.
45 240000	45	Construction work.
74 700000	74	Consultancy Services:Architectural, Construction, Legal, Accounting and Business.

Groups 1, 2 and 3 were formed by putting chapters of products together. In particular, all products falling in the first 44 chapters formed group 1 which in principal should cover Supplies. Product codes of chapter 45 formed the second group which refers exclusively to Works and the rest chapters formed group 3 which should cover Services.

Cross-tabulation between the frequencies of the three groups, as defined above, and the three sectors, as appear in variable NCCODE, shows that the correspondence is not one to one but it holds for the majority of the cases.

Below we present the situation in three Member States: Great Britain, Finland and Spain. The correspondence between Works and group 2 varies from 84% to 93% of the cases depending on the Member State. Similarly, correspondence between Supplies and group 1 varies from 94% to 97% and between Services and group 3 from 80% to 96%.

Contract Award Notices for GB in 2005

Table of group by NCCode				
group	**NCCode**			
	1 = Works	**2 = Supplies**	**4 = Services**	**Total**
1= chap<45	27	1630	244	1901
2 = chap=45	223	19	75	317
3 = chap>45	16	86	1587	1689
Total	266	1735	1906	3907

Contract Award Notices for FI in 2005

Table of group by NCCode				
group	**NCCode(NCCode)**			
Frequency	**1 = Works**	**2 = Supplies**	**4 = Services**	**Total**
1 = chap<45	3	154	6	163
2 = chap=45	31	1	23	55
3 = chap>45	0	6	118	124
Total	34	161	147	342

Contract Award Notices for ES in 2005

Table of group by NCCode				
group	**NCCode(NCCode)**			
Frequency	**1 = Works**	**2 = Supplies**	**4 = Services**	**Total**
1 = chap<45	29	1546	59	1634
2 = chap=45	668	13	55	736
3 = chap>45	19	38	2624	2681
Total	716	1597	2738	5051

The methodology used for the calculation of indicator 3 remains the same. For each chapter we calculated the average amount involved, after excluding all large contracts, and we multiplied it by the number of Calls reported in 2005 under that chapter. To get the final figure we summed up the amounts of each chapter and we added the total amount of the large contracts. Two problems were encountered when using this approach:

- For some chapters only contract award notices were available and no Calls or vice versa.
- For some chapters the number of available contracts was so small that the calculation of an average didn't make any sense.

To overcome these two problems we have used information of two consecutive years that is 2004 and 2005 together. Averages calculated in this way were based on more cases and therefore were more informative and reliable. Below, we give as an example the calculation of indicator 3 for Spain in 2005 based on chapter averages as explained previously.

Example: Indicator 3 for Spain based on 2004 and 2005 chapter averages.

chap	num_ES	mean_ES	sum_ES	CALLs	indicator 3
1	15	334019.41	5010291.18	24	8016465.84
2	3	955022.31	2865066.92	1	955022.31
5	3	582343	1747028.99	7	4076401
11	8	3475377.37	27803018.94	10	34753773.7
14	22	1668159.19	36699502.16	20	33363183.8
15	86	731715.38	62927522.93	70	51220076.6
17	50	537294.53	26864726.51	33	17730719.49
18	78	750314.93	58524564.66	66	49520785.38
19	6	518187.09	3109122.55	2	1036374.18
20	10	446489.05	4464890.45	6	2678934.3
21	59	473262.61	27922493.75	41	19403767.01
22	115	669326.14	76972505.64	70	46852829.8
23	67	1959605.29	131293554.6	52	101899475.1
24	507	1483158.01	751961109.7	265	393036872.7
25	82	1038833.74	85184366.31	48	49864019.52
26	7	520169.46	3641186.2	5	2600847.3
27	13	946202.52	12300632.71	12	11354430.24
28	112	821558.46	92014547.83	96	78869612.16
29	318	1968510.66	625986390	241	474411069.1
30	286	1690153.23	483383822.7	206	348171565.4
31	104	3063999.4	318655937.1	84	257375949.6
32	116	1091454.56	126608728.4	98	106962546.9
33	978	731625.43	715529667	686	501895045
34	162	2538672.11	411264882.3	170	431574258.7
35	130	6244627.7	811801600.5	38	237295852.6
36	128	886942.35	113528620.9	128	113528620.8
40	52	2643322.81	137452786.1	59	155956045.8
45	1325	13136545.67	17405923014	1121	14726067696
50	497	1848163.48	918537250.6	364	672731506.7
52	2	209240	418480	2	418480
55	45	3396027.32	152821229.4	26	88296710.32
60	72	1059581.17	76289843.95	75	79468587.75
62	19	3521770.41	66913637.72	10	35217704.1
63	70	1735539.12	121487738.2	52	90248034.24
64	60	4848833.05	290929982.9	41	198802155.1
65	11	2121189.25	23333081.73	13	27575460.25
66	53	2459500.32	130353517	71	174624522.7
67	6	953781.14	5722686.84	7	6676467.98
70	12	5147104.71	61765256.46	14	72059465.94
72	348	1248993.18	434649627	251	313497288.2
73	12	518946.13	6227353.51	10	5189461.3
74	3370	1182786.54	3985990625	2383	2818580325
75	24	1613644	38727456	18	29045592
77	53	1092958.11	57926779.78	67	73228193.37
78	46	424248.56	19515433.84	25	10606214

80	22	741359.52	16309909.46	4	2965438.08
85	27	1429132.44	38586575.97	20	28582648.8
90	153	2342407.8	358388393.7	107	250637634.6
92	41	776378.53	31831519.59	22	17080327.66
93	44	861405.94	37901861.49	31	26703584.14
Total	9829				23282708042
Total amount from contracts larger than 100000000 Euro					14258690395
Indicator 3, total amount of PP for 2005					37541398437

Note: Two Calls were excluded from the calculations as they were not corresponding to CANs. If we had used only 2005 data four CANs and two CALLs would have been excluded.

Below we present the results from the calculation of indicator 3 for GB, ES and FI for the year 2005 using the classification in groups of chapters as explained previously.

Indicator 3: Great Britain 2005

Statistics on the amounts of the contracts with EUPX < 100000000 Euro

Obs	group	num_GB	mean_GB	sum_GB	max_GB	median_GB	min_GB	ratio_GB
1	1	1887	2499031.89	4715673183.0	96519450	619698.74	146	4.03266
2	2	295	11081218.36	3268959415.7	95057034	5530536.70	26323	2.00364
3	3	1676	3039534.91	5094260512.8	87744955	584966.36	101	5.19608

Distribution of the large contracts among the three groups

group	Frequency	Cumulative Frequency
1	14	14
2	22	36
3	13	49

Total amount of the large contracts for 2005 is 15104846629 Euro that is approximately 15.1 billion Euro

Estimation of the indicator 3 for 2005

group	mean_GB	num of calls	Indicator 3
1 = chap<45	2499031.9	5879	14691808481
2 = chap=45	11081218	1350	14959644786
3 = chap>45	3039534.9	5442	16541148980
Total of small contracts			46192602248
Total of big contracts			15104846629
Total estimation			61297448877

Indicator 3: Finland 2005

Statistics on the amounts of the contracts with EUPX < 100000000 Euro

group	snum_FI	smean_FI	ssum_FI	smax_FI	smedian_FI	smin_FI	sratio_FI
1 = chap<45	163	2477176.40	403779753.57	82000000	600000.00	7090.00	4.12863
2 = chap=45	54	6089459.67	328830822.02	16899950	6570551.55	36000.00	0.92678
3 = chap>45	124	2017684.81	250192916.26	82809401	433845.48	583.69	4.65070

There is only one contract above the threshold of 100000000 Euro with a total amount of 638000000 Euro

Estimation of the indicator 3 for 2005

group	mean_FI	num of calls	Indicator 3[1]
1 = chap<45	2477176.4	895	2217072878
2 = chap=45	6089459.67	163	992581926
3 = chap>45	2017684.81	607	1224734680
Total of small contracts			4434389484
Total of big contracts			638000000
Total estimation			5072389484

The averages calculated over the three sectors on procurements of all available years lead to slightly different results.

Method	Estimate
Average over NCCode of procurements 2005, Duplicate RNs included	5336801174
Average over NCCode of procurements 2005, Duplicate RNs excluded	5598147618
Multiple years average over NCCode, Duplicate RNs included	5179077373

[1] The 13 duplicate RN CAN's were not excluded. The value of the estimator when calculated over the sectors, without excluding the duplicate RN CANs is 5336801174

Indicator 3: Spain 2005

Statistics on the amounts of the contracts with EUPX < 100000000 Euro

group	snum_ES	smean_ES	ssum_ES	smax_ES	smedian_ES	smin_ES	sratio_ES
1 = chap<45	1627	1594354.84	2594015328.8	83623216	507056.21	484.18	3.14434
2 = chap=45	711	13385167.37	9516853999.0	96304662	7823968.83	5971.07	1.71079
3 = chap>45	2679	1360296.75	3644234989.8	95198399	527155.17	411.26	2.58045

Distribution of the large contracts among the three groups

group	Frequency	Cumulative Frequency
1	7	7
2	25	32
3	2	34

Total amount of the large contracts for 2005 is 14258690395 Euro that is approximately 14.3 billion Euro

Estimation of the indicator 3 for 2005

group	mean_ES	num of calls	Indicator 3[2]
1 = chap<45	1594354.84	2539	4048066939
2 = chap=45	13385167.37	1121	15004772622
3 = chap>45	1360296.75	3614	4916112455
Total of small contracts			23968952015
Total of big contracts			14258690395
Total estimation			38227642410

[2] The 96 duplicate RN CAN's were not excluded. The value of the estimator when calculated over the sectors, without excluding the duplicate RN CANs is 38,521,126,930

Comparison of the three approaches.

Was any significant improvement in the homogeneity and therefore the robustness of the averages used for the calculations?

Great Britain

The calculation of indicator 3 for 2005 under different classifications was straight forward for Great Britain since, the reporting rate and quality was sufficient, the number of CANs published was large enough to avoid aggregation over a series of years and no cases of duplicate RNs were found. The estimation we obtained using the classifications over groups or chapters did not vary significant. The original method based on sectors provided a rather higher value resulting in a difference of 4 billion Euros.

Method	**Overall estimation of Indicator 3 in million €**
By sectors	64.206
By chapters	61.102
By groups	61.298

Finland

Finland is an example of a small country with low reporting rate, where averages have to be calculated taking into consideration all available data of a series of years. Moreover, cases with double RNs have been identified and excluded from the original calculations.

Method	**Overall estimation of Indicator 3 in million €**
By sectors (duplicate RNs excluded)	5.173
By sectors (duplicate RNs included)	5.179
By groups (duplicate RNs included)	5.072
By chapters (duplicate RNs included)	5.541

Due to the above reasons it is not easy to compare the different methods applied to this dataset. For example, the groups were formed only on 2005 data, and therefore the averages calculated in this way are not exactly comparable to the averages provided after aggregating all available data. Similarly, for the calculation of the indicator based on chapter classification data of two consecutive years 2004 and 2005 have been used. In any case, as it can be seen from the above table, the estimations do not vary significantly with only one exception: the classification by chapter, which surprisingly gives a higher value.

Spain

Spain is one of the best examples of reporting rate and quality. The number of CANs and CALLs is sufficient to calculate representative averages without using data from previous years. However, cases with duplicate RNs have been identified and excluded from the original calculations. To make comparisons more meaningful, we recalculated indicator 3

without excluding the duplicate RNs and we obtained a slightly lower estimation as expected. When we applied the classification over groups the indicator decreased further and it reached its lowest value when we used the classification over chapters. The overall difference between the original value and that of chapter classification is about 2.5 billion Euros. If duplicate RNs have been excluded, this figure would have been somehow less.

Method	**Overall estimation of Indicator 3 in million €**
By sectors (duplicate RNs excluded)	39.103
By sectors (duplicate RNs included)	38.521
By groups (duplicate RNs included)	38.228
By chapters (duplicate RNs included)	37.541

3.3.2 The use of robust estimators for the averages values

Although, very large contracts are excluded from the calculations of the averages, some of the remaining contracts are still too high in comparison to the rest of the sample. The choice of the threshold 100million Euro is arbitrary from a statistical point of view. We have seen that the distributions of the CANs in the MS are very skewed even after the exclusion of the large contracts. Therefore, we have to replace the mean with another robust measure of location such as the median or the trimmed mean. Both of them are relatively insensitive to the outlying values. The trimmed mean is calculated after the k smallest and k largest observations are deleted from the sample.

Instead of having an arbitrarily chosen absolute threshold above which trimming is applied, is better to use relative thresholds, that is, to eliminate a certain percentage of the observations at each extreme of the distribution. As we increase the percentage of trimmed data, the value of the trimmed mean falls and therefore the question we must answer is: What percentage of trimmed data should be used? The change of the shape of the distribution, and the lack of sensitivity in the mean, are two possible criteria.
For the current exercise considering the length of the tails we started with a less drastic approach. We used log transformations of the data and we trimmed until all outliers were eliminated.

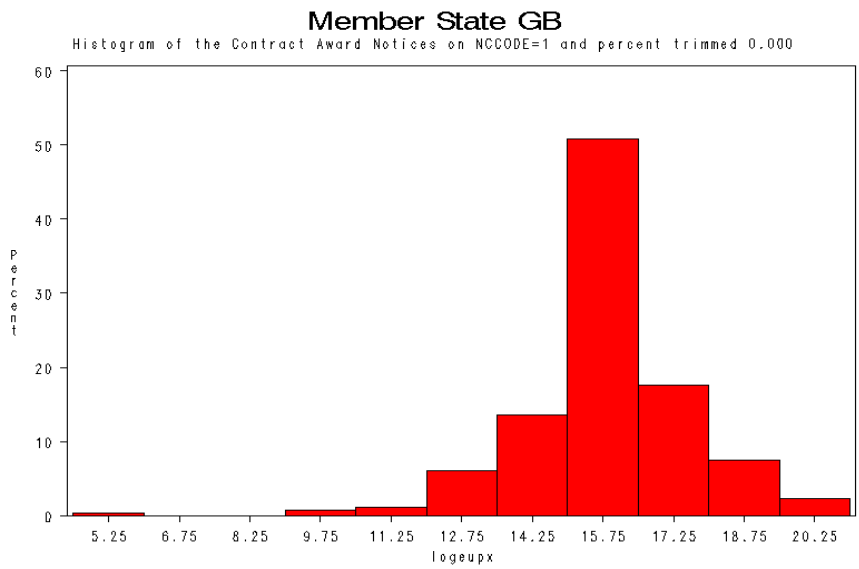
The logarithm function tends to squeeze together the larger values in the data set and stretches out the smaller values**.** In fact, the bigger the data value, the more the squeezing and the smaller the value the more the stretching. This squeezing and stretching can correct the following problems:

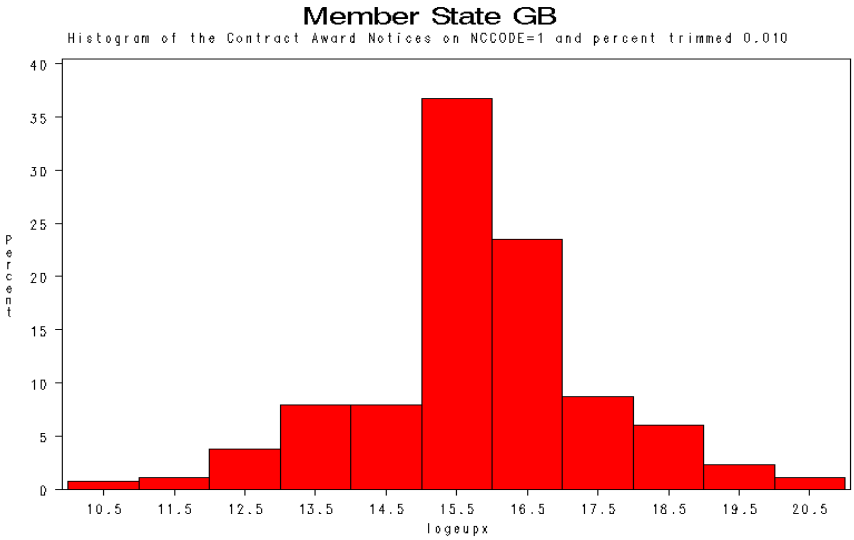
- Skewed data
- Outliers

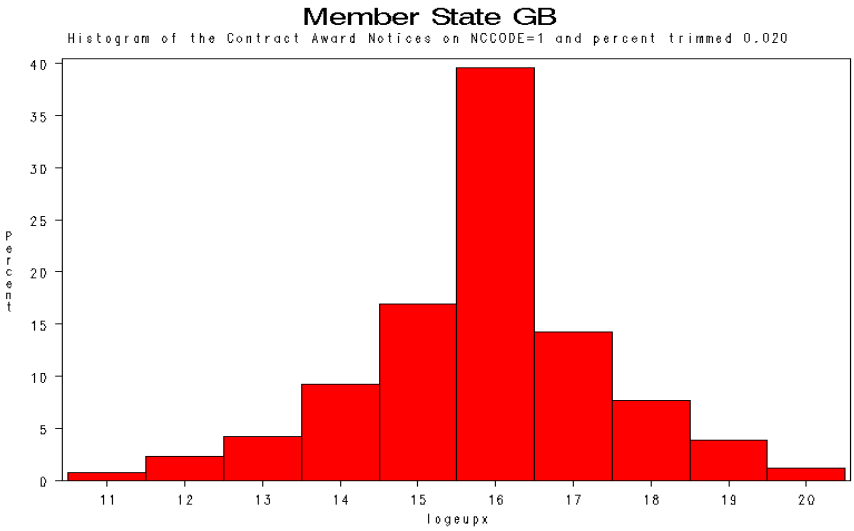
If the data are skewed to the right, a log transformation can sometimes produce a data set that is closer to symmetric**.** If the data are symmetric or skewed to the left, a log

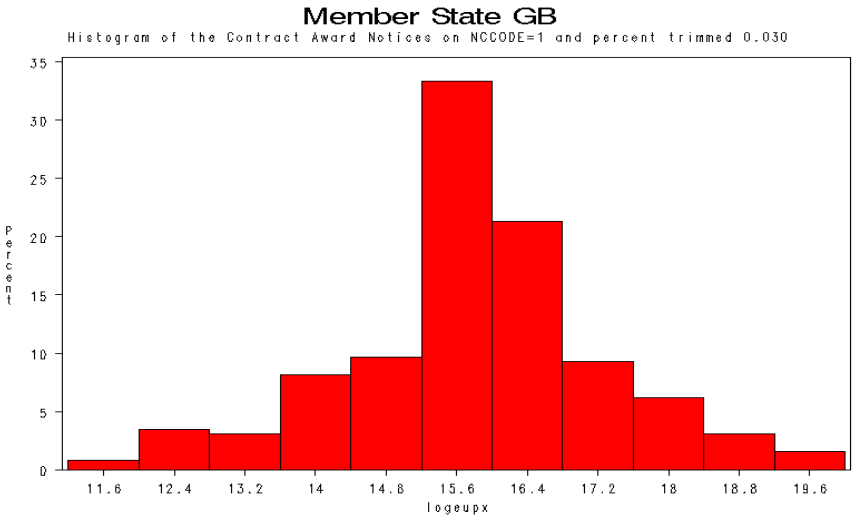
transformation could actually make things worse**.** If the dataset has outliers on the high end, as in our case, a log transformation can help since the squeezing of large values might pull that outlier back in closer to the rest of the data.

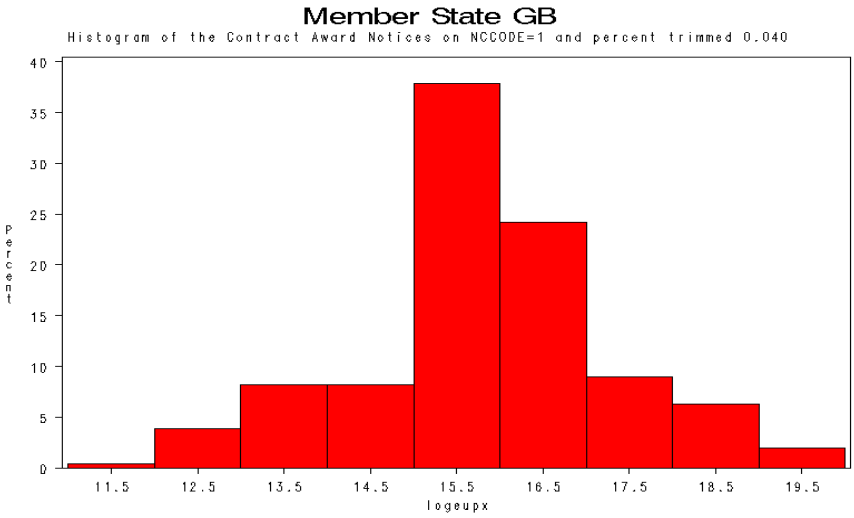
As example we will demonstrate what happened to the distribution of the contracts awarded in GB when we applied different levels of trimming and to the average values calculated for the three sectors: works, supplies and services. The histograms presented below, show how the distribution of the log transformed data changes after trimming 0% to 15% of the larger contracts of the sector works. The box-plots that follow the histograms show what is the effect of trimming on the outliers across the three sectors.
A box-and-whisker plot (sometimes called simply a box plot) is a histogram-like method of displaying data, invented by J. Tukey. The ends of the box (hinges) are at the quartiles, so that the length of the box is the interquartile range IQR. The median is marked by a line within the box and the average with a cross. The two vertical lines (called *whiskers*) are drawn from the hinges of the box to the largest value within the upper fence, located at 1.5 * IQR above the 75th percentile and the smallest value within the lower fence, located at 1.5 * IQR below the 25th percentile value. All values outside the upper and lower fences are considered as outliers and are identified as stars

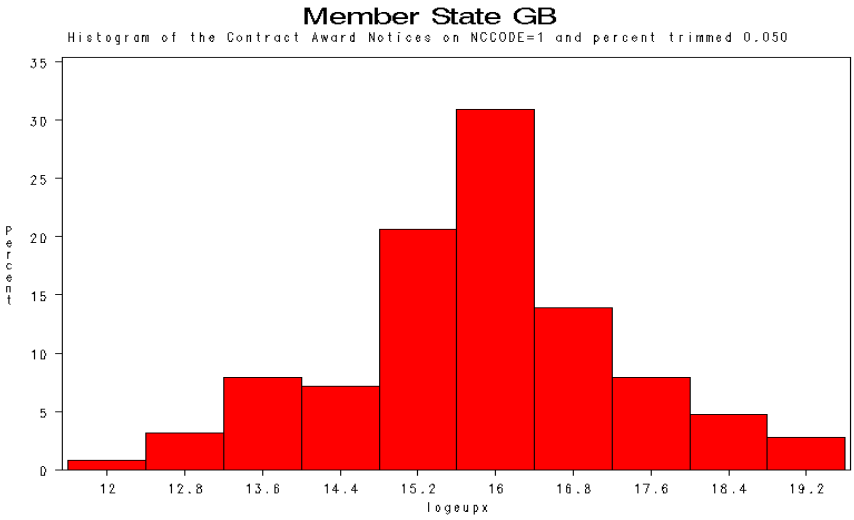


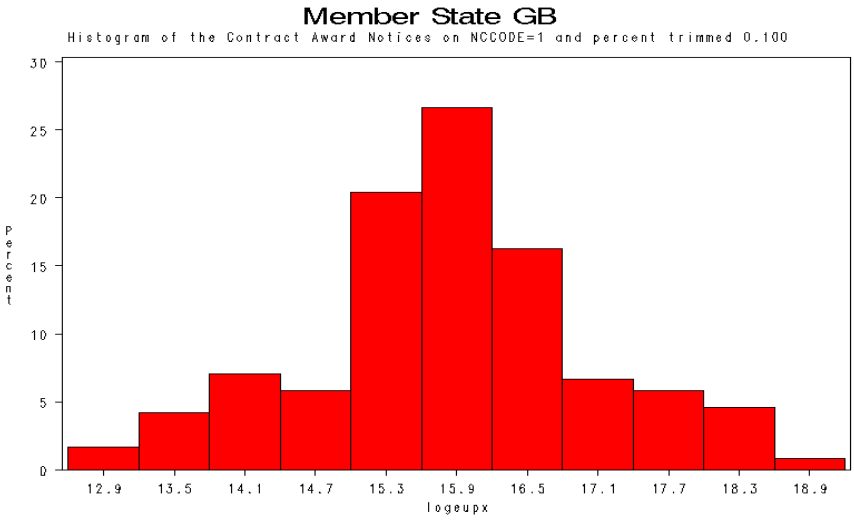

Member State GB
Histogram of the Contract Award Notices on NCCODE=1 and percent trimmed 0.010
Percent
0
5
10
15
20
25
30
35
40
10.5
11.5
12.5
13.5
14.5
15.5
16.5
17.5
18.5
19.5
20.5
logeupx

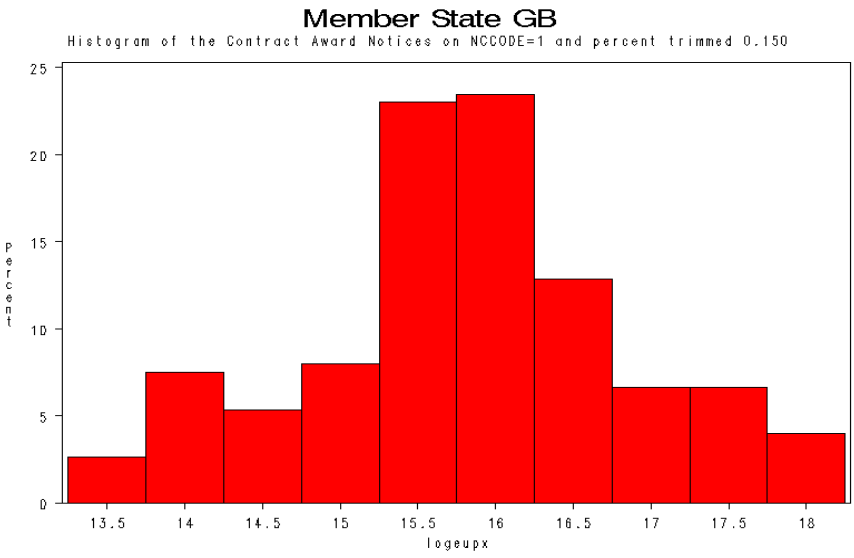

Member State GB
Histogram of the Contract Award Notices on NCCODE=1 and percent trimmed 0.020
Percent
0
5
10
15
20
25
30
35
40
11
12
13
14
15
16
17
18
19
20
logeupx


Member State GB
Histogram of the Contract Award Notices on NCCODE=1 and percent trimmed 0.030
Percent
0
5
10
15
20
25
30
35
11.6
12.4
13.2
14
14.8
15.6
16.4
17.2
18
18.8
19.6
logeupx

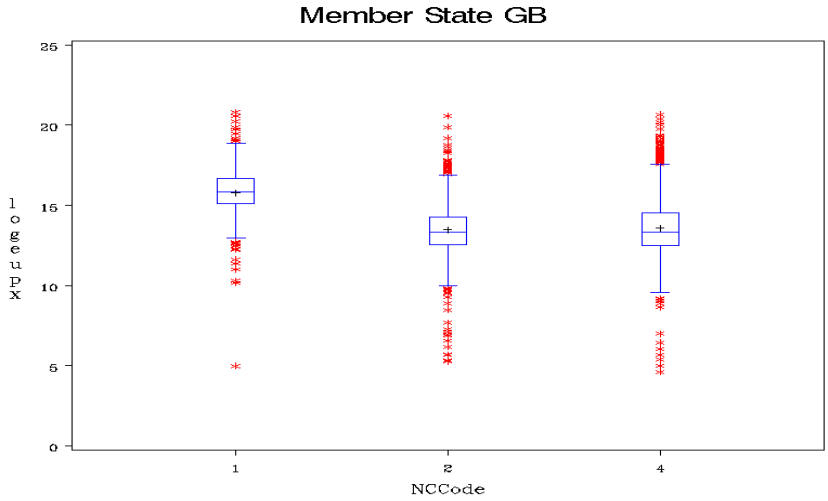

Member State GB
Histogram of the Contract Award Notices on NCCODE=1 and percent trimmed 0.040
Percent
0
5
10
15
20
25
30
35
40
11.5
12.5
13.5
14.5
15.5
16.5
17.5
18.5
19.5
logeupx


Member State GB
Histogram of the Contract Award Notices on NCCODE=1 and percent trimmed 0.050
Percent
0
5
10
15
20
25
30
35
12
12.8
13.6
14.4
15.2
16
16.8
17.6
18.4
19.2
logeupx

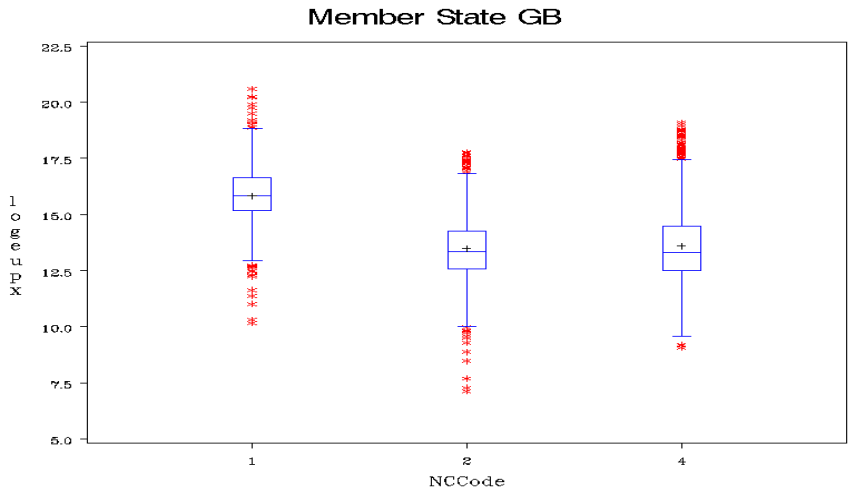

Member State GB
Histogram of the Contract Award Notices on NCCODE=1 and percent trimmed 0.100
Percent
0
5
10
15
20
25
30
12.9
13.5
14.1
14.7
15.3
15.9
16.5
17.1
17.7
18.3
18.9
logeupx


Member State GB
Histogram of the Contract Award Notices on NCCODE=1 and percent trimmed 0.150
percent
25
20
15
10
5
0
13.5
14
14.5
15
15.5
16
16.5
17
17.5
18
logeupx

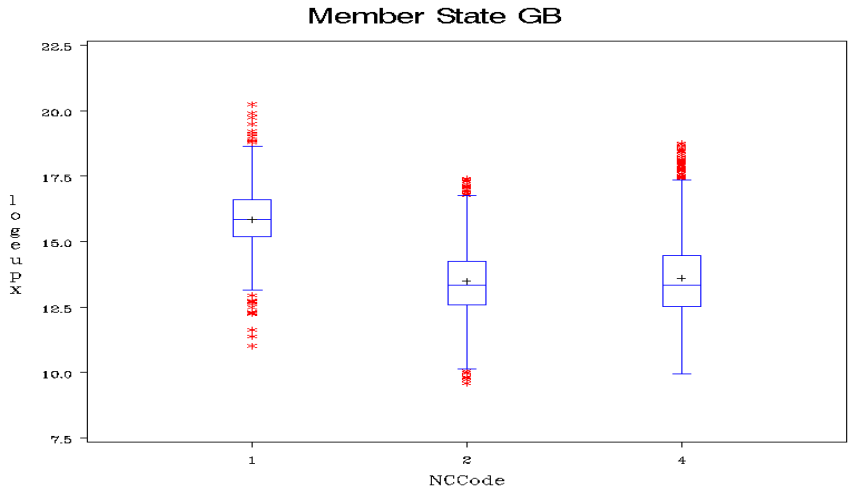
Trimmed 0.00



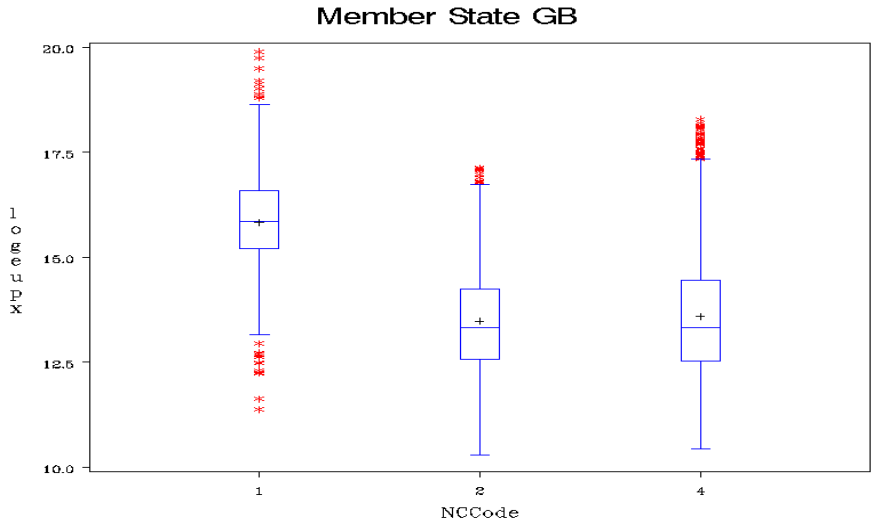
Trimmed 0.01



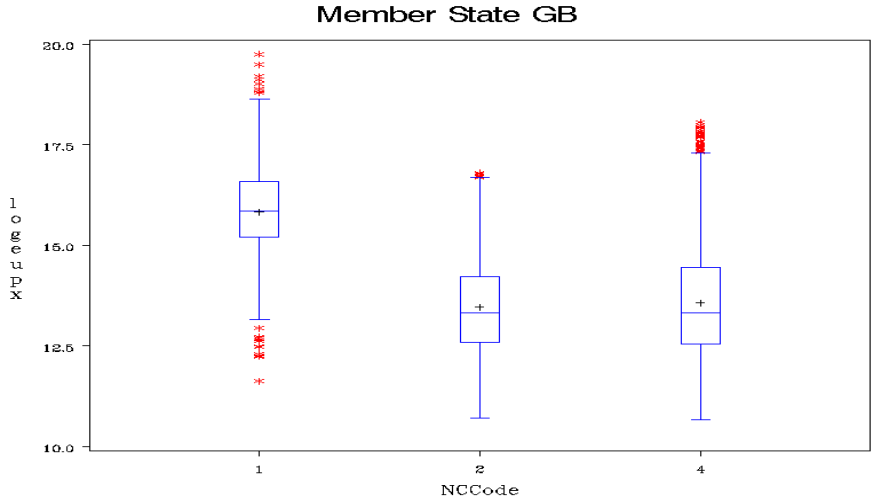
Trimmed 0.02



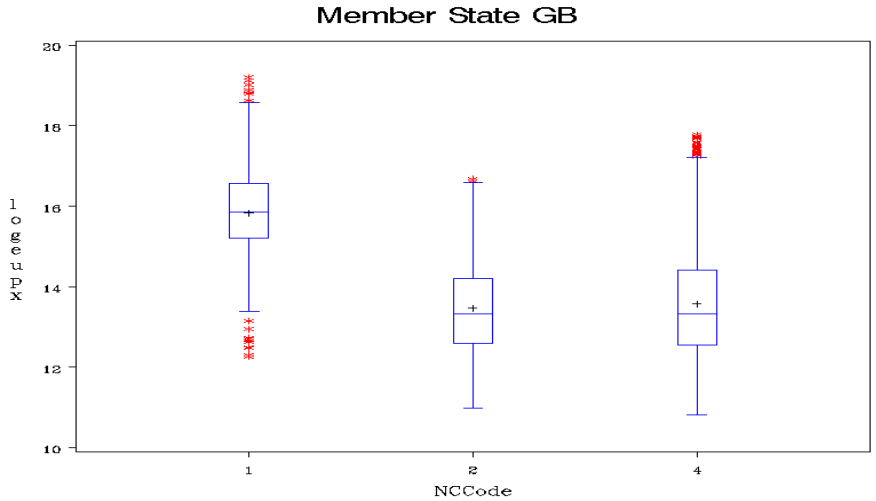
Trimmed 0.03



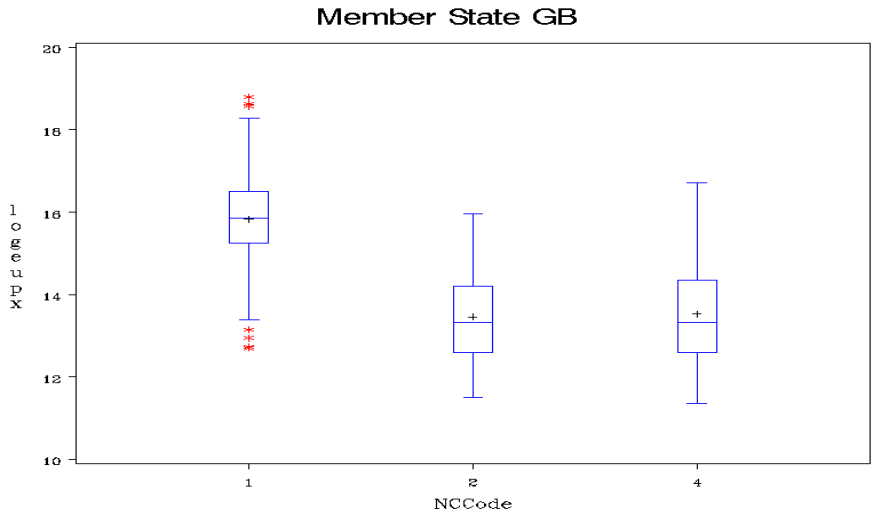
Trimmed 0.04



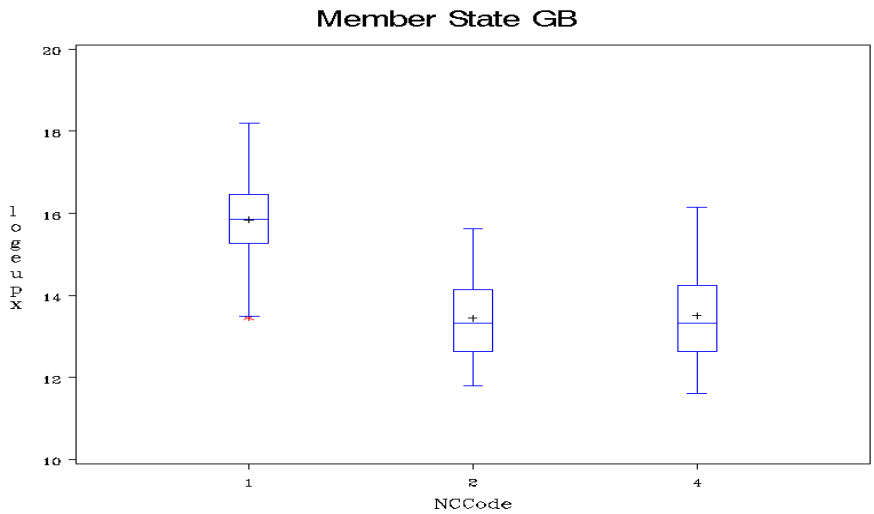
Trimmed 0.05



Trimmed 0.10



Trimmed 0.15



The following table presents summary statistics of the trimmed sets and detailed calculations of indicator 3 for each level of trimming. In particular:

Num = number of CANs after trimming.

CALLs = Number of CALLs registered for each of the sectors

Mean, Max, Q3, Median, Q1 and Min correspond to the Average, Maximum, 3rd Quartile, Median, 1st Quartile and Minimum value in the set of CANs for each sector after trimming

Totaln = Number of CANs for each sector before any trimming has been applied

NCCode = Sectors: 1 for Works, 2 for Supplies and 4 for Services

Ptrim = Percent of trimming. Varies between 0 and 15.

Ratio = Mean over Median. Gives an idea of the skewness.

Sum Outliers = Total amount of the contracts excluded due to trimming. It is added in the final calculation of the indicator.

Indicator 3 = The value of the indicator calculated for the optimum percent of trimming for each sector.

Num	Mean	CALLs	Max	Q3	Median	Q1	Min	Totaln	NCCode	Ptrim	Ratio	Sum Outliers		Indicator 3
266	33952943	1258	1096811933	17186113	7641123	3715095	146	266	1	0	4.443449		42712801797	
264	30055571	1258	877449547	16831338	7641123	3821163	26323	266	1	0.01	3.933397	1096811933	38906719822	
260	24729955	1258	614214683	16248627	7641123	3959403	60707	266	1	0.02	3.236429	2601637906	33711921393	
258	22540748	1258	438724773	16020690	7641123	3991575	87394	266	1	0.03	2.949926	3215852589	31572113309	
256	21002737	1258	379145949	15966619	7641123	4006609	112791	266	1	0.04	2.748645	3654577362	30076020955	
252	18669644	1258	219362387	15720402	7641123	4021644	212306	266	1	0.05	2.443311	4326206493	27812618518	
240	15104266	1258	146241591	14769768	7641123	4202933	329897	266	1	0.1	1.976708	5404346300	24405512456	
226	12811342	1258	79811348	14109060	7641123	4271695	698945	266	1	0.15	1.676631	6130420153	22247087975	22247087975
1735	3262645	5126	877449547	1608658	614215	281808	197	1735	2	0	5.311897		16724320277	
1717	2039404	5126	52646973	1579409	614215	285171	1250	1735	2	0.01	3.320343	2159028957	12613012115	
1701	1850275	5126	36267915	1542849	614215	288315	14624	1735	2	0.02	3.012424	2513310853	11997821807	
1683	1697610	5126	27639661	1536873	614215	289849	29248	1735	2	0.03	2.763871	2803365196	11505315732	
1665	1580764	5126	19888856	1513444	614215	292483	44477	1735	2	0.04	2.573635	3028143585	11131141242	
1649	1506165	5126	17548991	1474115	614215	292483	58497	1735	2	0.05	2.45218	3176053906	10896654092	10896654092
1561	1262766	5126	8623989	1462416	614215	295445	99444	1735	2	0.1	2.055904	3685031971	10157971878	
1475	1121178	5126	6142147	1389558	614215	307107	131617	1735	2	0.15	1.825385	3997432081	9744591461	
1906	7078472	6287	961068314	2036256	608557	265428	101	1906	4	0	11.63158		44502352215	
1886	4550673	6287	195963732	1974261	608557	272009	8687	1906	4	0.01	7.477814	4908973326	33519055652	
1868	3760852	6287	139144778	1925766	608557	276428	21132	1906	4	0.02	6.179954	6466140890	30110614777	
1848	3199573	6287	87744955	1901141	608557	278371	34267	1906	4	0.03	5.257642	7578324253	27694037406	
1830	2853840	6287	70195964	1901141	608557	280784	42539	1906	4	0.04	4.689523	8268273437	26210367651	
1810	2544174	6287	52646973	1833429	608557	283606	49898	1906	4	0.05	4.180669	8885380783	24880602231	
1716	1752017	6287	18280199	1694951	608557	292483	84820	1906	4	0.1	2.878971	10480595506	21495525346	21495525346
1620	1444987	6287	10212050	1526310	608557	307107	109681	1906	4	0.15	2.37445	11141462142	20226096887	
Overall estimation based on the optimum percent of trimming applied in each sector														54639267414

We have used the original methodology for the calculation of the indicator, as summarized in equation (1). The level of trimming was decided by looking at the box-plots and the number of outliers left after each iteration. For example, for the first sector, Works, we eliminate all outliers if we exclude from the calculations 15% of the data that is 7.5% each side. For the other two sectors the percentages are 5 and 10 respectively. The overall estimation we obtain in this way, approx.54.6 billion Euros, is significantly lower than any other figure produced by the methods discussed so far.

Currently, the method as such, is rather manual and intuitive. Automatic criteria for the determination of the percent of trimming to be applied each time are necessary. This requires significant amount of programming and time. Nevertheless, the strongest limitation of the method is related to the lack of adequately large samples on which calculations can be made.

3.3.3 Reverse Chronological "Backward" Estimation of the Indicator 3

Introduction

As we have mentioned previously, one of the drawbacks of the method currently used for the calculation of indicator 3, is the lack of correspondence between CALLs and CANs due to the time lag between the date the CALL was published and the date the contract was awarded. As a result, when we use the CANs of a specific year to estimate the CALLs of that year, in practice we use information of previous years assuming that the situation they describe has not change significantly. In order to avoid this mixing of information and given that there is available data for a series of years, we tried to create a sample of contracts awarded in different years which all correspond to calls for tenders published for a given year *t* and then estimate indicator 3 based on this sample.

We are aware that this type of backward estimates cannot be used for forecasting purposes since they can only be available for past years. However, they can be very useful for checking and verifying the results obtained by the standard methods. Moreover, depending on the quality of the data, the samples constructed in this backward approach can be used for creating empirical confidence bounds for the estimates, as we will explain later.

The backward methodology has been applied to three countries: Spain (highly complying member state), Great Britain (one of the largest EU economies) and Finland (poorly complying member state). Its use in practice is illustrated on the Spanish data. The results corresponding for the two remaining member states can be found in Annex I and II.

Application of the backward method to the Spanish data

Before explaining the backward methodology, a quick look at the TED database shows (Table 1) that the total number of entries increased steadily between 2003 and 2006. The calls for tenders[3] form the

[3] Call for tenders are defined as entries in the TED database for which "TDcode" is equal to 3, M or O and for which "AAcode" is different from 5.

largest category, covering about half of the total entries and they show a slight downward trend over the sample period. The contract awards notices[4] account for almost 40 percent.

The basic idea behind the backward methodology is, first, to create a sample of contracts awarded in different years which all correspond to calls for tenders published for a given year *t* (we will refer henceforth to this sample as the *reference sample*) and then estimate indicator 3 based on this sample assuming that CALLs and CANs come from the same distribution.

Table 1: Overview of main categories of the TED database

	2003	2004	2005	2006
Entries	11872	12632	15130	19236
Calls for Tenders	5940	6149	7274	9196
Contract Awards	4802	5015	5572	7543
As Percentages of Total Entries				
Calls for Tenders	50.0%	48.7%	48.1%	47.8%
Contract Awards	40.4%	39.7%	36.8%	39.2%

The first step in the "backward approach" that is to create the "reference" sample of CANs, is achieved by linking the "RN" field[5] of the CANs with the "ND Num" field of the CALLs. Though not unique, as several contracts may contain the same number for "RN" field, it is the only possible solution to establish such a link given information currently available in the TED database. However, when such a "double" entry is present, it concerns, in all cases, contracts corresponding to calls of the same government body, and in most cases, for the same category of products (identical "PN" field – field which describes what is being bought). In any case, the proportion of such "double" entries is rather limited[6].

In order to create this reference sample, we start by selecting contract awards for which price and RN Number are available (Table 2). The number of the latter has been growing steadily, both in absolute and relative terms. In particular, they accounted for around 76 (in the table we read 75.7%) percent of total contract awards recorded in 2006. This clearly indicates an improvement in the quality of the data stored in the database. Similarly, the number of large contracts (above 100 million euros) showed an upward trend between 2003 and 2006 both in absolute or relative terms. Nevertheless, they still represent less than 1 percent of total contract awards.

Table 2: Classification of contract awards

	2003	2004	2005	2006
Contract Awards	4802	5015	5572	7543
- with Price	2653	4839	5051	6918
- with RN	2595	3253	4255	6226
- with Price and RN	1237	3165	3899	5712
- Large Contracts	16	22	34	53
- out of which with Price and RN	14	13	28	44
As Percentages of Contract Awards				
Contract Awards				
- with Price	55.2%	96.5%	90.6%	91.7%
- with RN	54.0%	64.9%	76.4%	82.5%
- with Price and RN	25.8%	63.1%	70.0%	75.7%
- Large Contracts	0.3%	0.4%	0.6%	0.7%
- out of which with Price and RN	0.3%	0.3%	0.5%	0.6%

Source: TED database and own calculus.

[4] Entities with "TDcode" equal to 7, "AAcode" different from 5 and "RPcode" larger than 3 are considered as contract awards.

[5] In fact, "RN" field contains two numbers in the following format: "15872-2006". In the case of contract awards, the former is used to establish a link, if possible, with the "NDNum" field corresponding to calls published during the year given by the latter.

[6] To be precise, by "double" entries, we mean at least two contracts which have the same "RN number". However, they may be three or more. For example, in the case of contracts awarded in 2006 and corresponding to calls published during the same year, there are 180 such "double" entries out of 3429 available observations, *i.e.* slightly more than 5 percent of the sample.

Next we proceed with the decomposition of the contract awards according to the year of their call (Table 3). Indeed, as already explained, the "RN" field contains two numbers, one of which indicates the year of call and the other is used to link this particular contract to its CALL. In the "Call-Contract" table the rows give the number of contracts corresponding to the given year of call. For example, from the 5940 CALLs published in 2003 we trace 2007 that lead to contracts for which the price was available, that is 33.8%. Of those, 22.5% (452/2007) were awarded the same year the call was published, 73.4% one year after, 3.6% two years after and 0.4% three years after. The columns show the distribution of the contracts of a given year according to the year of their call. It can be seen that the percentage of contracts awarded in a given year corresponding to calls published during the same year has steadily increased to 60 percent in 2006 from 36.5 percent in 2003. Note that, the figures read on row "Total 1" in Table 3, are identical to those found under heading "Contract Awards with Price and RN Number" in Table 2.

Table 3: Call-Contract table[1]

	Year of Contract Award				
Year of Call	2003	2004	2005	2006	Total 2
2000	4	4	1	1	-
2001	24	12	2	0	-
2002	757	35	15	3	-
2003	452	1474	72	9	2007
2004	-	1636	1578	124	3338
2005	-	-	2231	2146	4377
2006	-	-	-	3429	3429
Total 1	1237	3161[2]	3899	5712	-
As Percentages of Total 1					
2000	0.3%	0.1%	0.0%	0.0%	-
2001	1.9%	0.4%	0.1%	0.0%	-
2002	61.2%	1.1%	0.4%	0.1%	-
2003	36.5%	46.6%	1.8%	0.2%	-
2004	-	51.8%	40.5%	2.2%	-
2005	-	-	57.2%	37.6%	-
2006	-	-	-	60.0%	-
Total	100.0%	100.0%	100.0%	100.0%	-

Note: 1) Only contract awards with non-zero price and RN Number are taken into consideration.
2) The RN Number does not allow classifying correctly 4 contract awards in 2004 (as the second part of the code contains implausible year reference - twice 1900, 2098 and 2119).

In comparison to Table 3, Table 4 below focuses only on those contracts for which the exact link with a CALL can be made. This leads to an extra reduction of the sample and explains the slight difference in the figures between the column Total 2 in Table 3 and the row Total in Table 4. The sample formed in this way from now on will be referred as the “reference sample” and it will be used for the calculations. Compared to Table 3, rows and columns are now reversed for the ease of exposition. Indeed, such a modified presentation allows for a more readable decomposition of contracts according to their size (*see* the second part of the Table 4).

Table 4: Contract-Call table for identifiable pairs only[1]

Year of Contract Award	Year of Call for Tenders 2003	2004	2005	2006
2003	442	-	-	-
2004	1432	1596	-	-
2005	68	1542	2169	-
2006	7	116	2048	3363
Total	1949	3254	4217	3363
p.m. Total Calls for Tenders	*5940*	*6149*	*7274*	*9196*
As percentages of Total Calls for Tenders				
2003	7.4%	-	-	-
2004	24.1%	26.0%	-	-
2005	1.1%	25.1%	29.8%	-
2006	0.1%	1.9%	28.2%	36.6%
Total	32.8%	52.9%	58.0%	36.6%
Non-Large Contract Awards				
2003	438	-	-	-
2004	1429	1589	-	-
2005	64	1535	2158	-
2006	4	114	2037	3343
Total	1935	3238	4195	3343
Large Contract Awards[2]				
2003	4	-	-	-
2004	3	7	-	-
2005	4	7	11	-
2006	3	2	11	20
Total	14	16	22	20

Note: 1) Out of contract awards with non-zero price and "RN" field, only those are taken into account for which a link to corresponding call can be established (reference sample).
2) Those above 100 million of euros.

It turns out that by pooling all contract awards from 2003 onwards corresponding to calls from 2003, we obtain a sample which covers about 1/3 of the calls published in 2003[7]. However, as we pointed before, that ratio rises to more than half in 2004 and 2005 despite the decrease in the time span of the related samples. But more importantly, the distinction between large and non-large contracts seems to indicate that a three-year window should be large enough to obtain a good sample which can provide reliable backward estimates of indicator 3. Definitely, the proportion of non-large contracts awarded in time t+3 is rather negligible for calls published in 2003. Moreover, it seems unlikely that this pattern will be reversed afterwards given the evolution of distributions for subsequent years. This claim holds also, though to a lesser extent, for large contracts.

The next step in the backward method is to compute the average value of the reference sample of the Non-Large contracts (CANs from time *t* onwards corresponding to CALLs from time *t. see* Table 5). It is important to **note that no classification in sectors or chapters or groups has been applied to the reference sample.** Assuming that CALLs and CANs have the same distribution in terms of amounts, we simply multiply this average by the number of CALLs published in time *t*. We then add the sum of the large contracts and we obtain the final estimate of indicator 3. For example, suppose we want to estimate indicator 3 for Spain in year 2003 (see Table 5). Our reference sample contains 1935 non-large CANs with average value 2947801 euros and 14 large CANs with a total value of 4827714010 euros. If we multiply the average value by the total number of CALLs published in 2003, that is 5940

[7] That is, 1949 divided by 5940.

and to this product we add the sum of the large contracts we will get an estimate of indicator 3 based on backward calculations. For Spain this amount is approximately 22.3 and 22.9 billion euros for 2003 and 2004 respectively.

Table 5: Estimation of the value of public procurements
(in euros)

Year of Contract Award	Year of Call for Tenders 2003	2004	2005	2006
Average Price for Identifiable Non-Large Contract Awards				
2003	3,669,052	-	-	-
2004	2,504,606	3,240,433	-	-
2005	7,988,769	2,962,833	3,238,011	-
2006	1,646,633	4,238,371	3,408,898	2,960,269
Weigted Averge	2,947,801	3,143,969	3,320,990	2,960,269
Sum of Prices for Identifiable Large Contract Awards				
2003	788,613,846	-	-	-
2004	1,370,354,849	1,294,504,874	-	-
2005	1,037,810,554	1,691,415,418	9,281,762,436	-
2006	1,630,934,761	609,786,544	28,177,368,198	10,226,088,164
Total	4,827,714,010	3,595,706,836	37,459,130,634	10,226,088,164

Estimation of the value of Public Procurements based on Identifiable Calls for Tenders

	Year of Call for Tenders 2003	2004	2005	2006
Non-Large Calls	17,509,937,398	19,332,264,050		
Large Calls	4,827,714,010	3,595,706,836		
Total (1)	22,337,651,409	22,927,970,886		
p.m. DG-Markt Estimates (2)	26,710,000,000	25,060,000,000		
Difference between (2) and (1) (in percentage points)	19.6%	9.3%		

Source: TED database, DG-Markt note MARKT/RW D(2007) and own calculus.

How our results compare to those of the DG-MARKT? Table 5 shows that there is a difference in the estimations which are significantly lower when the backward method is applied. Such behaviour is expected and justified up to a point, as the samples on which calculations were based are not the same and no classification of the contracts has been considered in the backward approach. However, the magnitude of the difference was unknown and it varies between the Member States depending on the quantity and quality of the available data. For Spain this difference is almost 20% (less than the original estimation) for 2003 and 10% for 2004.

In the case of Great Britain, presented in the Annex I the difference between the two estimations is enormous. We have reasons to believe that this is not normal. In fact, looking at the figures produced for 2003 by DG MARKT using the original method we notice that the total amount of published CALLs in GB in 2003 is much higher than every other Member State and almost double than the previous year. A mistake in the calculations, poor quality of the data, a change or adoption of a specific policy, and/or an external unforeseen factor may be the reasons for this behaviour. On the other hand, due to the small size of the reference sample used in the backward approach, less than 20% of the CALLs published that year, we can not ignore the possibility that some of the very large contracts have been excluded from the calculations.

In the case of Finland presented in the Annex II, the results indicate an underestimation of the total amount of published CALLs as calculated by the original method. The difference accounts for 20% to

25% for 2003 and 2004 respectively. For this particular Member State we should not draw conclusions without considering the low reporting and quality rate of the data provided every year. However, Finland due to the small number of large contracts reported each year, gives a good example of how the lack of correspondence between the CALLs and the CANs for the year the indicator is calculated can affect the estimation. In Table 5 we see that in 2004 there has been one CALL for a large contract: total amount 638000000 Euro. This amount has not been used in the calculations of indicator 3 for 2004 but for 2005 since the contract was awarded that year.

Comments on the usefulness and the limitations of the method

All in all, we can say that the backward method increases the homogeneity of the sample as it considers only the contracts for which there is one to one correspondence with the published CALL eliminating the time lag. Moreover, large contracts that affect the estimate refer to the real year of publication. Unfortunately, it is a method that we can only use retrospectively; ideally 3 to 5 years after the CALLs were published and therefore, it is not useful for predictions. Obviously, usefulness of any method depends on the objectives in mind. If timely estimates of the value of public procurements are of utmost importance to policymakers, the backward methodology is clearly not the most appropriate tool. Nevertheless, even in these circumstances, it may appear as a useful benchmark from a cross-checking perspective.

We have also noticed that the method is based on an assumption that may not be always true. In particular, it assumes that all large contracts published in the year of interest have been included in the reference sample. If the reference sample is relatively large, as for example in the case of Spain, the assumption is not unrealistic. If however, the reference sample covers only a small fraction of the CALLs published the year of interest, such assumption may not hold. The case of Great Britain for 2003 is an example of this situation. In general, if the reference sample is large enough, not only we can obtain good estimates of indicator 3 but we can also use re-sampling techniques in order to construct empirical confidence bounds for the indicator. Comparing the value of the indicator, as calculated at the time using the original method, with these confidence bounds we may conclude with a certain assurance whether the published figures overestimate or underestimate the real total amount of published CALLs.

4. General Conclusions - Suggestions

The main conclusions of this exercise are summarized in this section, together with some suggestions for future improvement of the indicator.

Based on our analysis we can conclude that the original method developed by DG MARKT most probably overestimates the total amount published but we can not say at what extent. It is clear that both the quality and quantity of the data needs further improvement. The structure of the database must also be revised so as to eliminate non informative variables and to insure that the same information is recorded in a stable format every year. Moreover, modification of the information required to be registered in the TED database could drastically improve the current situation. For example, the indication of a range of values to be expected in a CALL, or an upper limit in the price of the offers, are information that permit to obtain in time estimations of the total amount of published Tenders, and possibly, to develop upper and lower bounds for these estimates.

Due to the time lag between the publication of the CALL and the awarding of the contract the current method suffers from a lack of correspondence between the CALLs and the CANs of the reference year. In particular, this mismatch affects the total amount of the large contracts used in the estimation of the indicator 3, as their distribution varies significantly from year to year.

The mean values used in the calculations are predominantly influenced, by the long right tails of the distribution of the CANs. The use of a trimmed mean instead of the arithmetic mean showed that, depending on the level of trimming, the results change since the mean values become significantly lower. Therefore, it is necessary to apply automatic criteria to determine the percent of trimming taking into consideration the amount of programming and time involved. Nevertheless, even if we improve this specific part in the estimation of the indicator the lack of correspondence between CANs and CALLs and the poor reporting still remain as the major drawbacks.

Two new types of classification of the data have been applied. Both were based on the product codes involved and each of them gave different results. In theory, classification in chapters (first 2 digits of the product code) should reduce the variability of the samples leading to better estimations of the means. In practice we realised that no sufficient data was available for some of the Member States. However, when we applied these classifications to adequately large samples of good quality data we verified our initial hypothesis that the indicator based on groups (of chapters) will be lower than the indicator based on NCCodes and higher than the indicator based on chapters (cases of ES and GB).

The reverse chronological (backward) method we developed is based on a sample that contains only those CANs directly linked to the CALLs of the reference year. Assuming that this “reference” sample is large enough, the method works well and provides reliable results. Unfortunately, it can only be used retrospectively for verification and cross checking purposes.

Annex I

Results from the application of the backward method to GB

Table 1: Overview of main categories of the TED database

	2003	2004	2005	2006
Entries	20851	21573	21977	22472
Calls for Tenders	11927	12366	12671	12676
Contract Awards	6282	6236	6103	6888
As Percentages of Total Entries				
Calls for Tenders	57.2%	57.3%	57.7%	56.4%
Contract Awards	30.1%	28.9%	27.8%	30.7%

Table 2: Classification of contract awards

	2003	2004	2005	2006
Contract Awards	6282	6236	6103	6888
- with Price	2291	3751	3907	5254
- with RN	4333	4644	4101	3567
- with Price and RN	1601	2761	2578	2608
- Large Contracts	38	48	49	91
- out of which with Price and RN	29	41	34	67
As Percentages of Contract Awards				
Contract Awards				
- with Price	36.5%	60.2%	64.0%	76.3%
- with RN	69.0%	74.5%	67.2%	51.8%
- with Price and RN	25.5%	44.3%	42.2%	37.9%
- Large Contracts	0.6%	0.8%	0.8%	1.3%

Table 3: Call-Contract table

Year of Call	Year of Contract Award 2003	2004	2005	2006	Total 2
1998	8	-	-	-	-
1999	5	1	-	-	-
2000	15	6	1	-	-
2001	146	18	4	1	-
2002	1023	133	27	11	-
2003	404	1581	178	23	2186
2004	-	983	1542	149	2674
2005	-	-	826	1378	2204
2006	-	-	-	1046	1046
Total 1	1601	2722	2578	2608	-
As Percentages of Total 1					
1998	0.5%				-
1999	0.3%	0.0%			-
2000	0.9%	0.2%	0.0%		-
2001	9.1%	0.7%	0.2%	0.0%	-
2002	63.9%	4.9%	1.0%	0.4%	-
2003	25.2%	58.1%	6.9%	0.9%	-
2004	-	36.1%	59.8%	5.7%	-
2005	-	-	32.0%	52.8%	-
2006	-	-	-	40.1%	-
Total	100.0%	100.0%	100.0%	100.0%	-

Table 4: Contract-Call table for identifiable pairs only

Distribution of Identifiable Calls for Tenders

Year of Contract Award	Year of Call for Tenders 2003	2004	2005	2006
2003	396	-	-	-
2004	1517	959	-	-
2005	174	1504	806	-
2006	21	145	1341	1037
Total	2108	2608	2147	1037
p.m. Total Calls for Tenders	11927	12366	12671	12676
As percentages of Total Calls for Tenders	17.7%	21.1%	16.9%	8.2%
Non-Large Contract Awards				
2003	393	-	-	-
2004	1494	956	-	-
2005	161	1492	802	-
2006	16	132	1310	1026
Total	2064	2580	2112	1026
Large Contract Awards				
2003	3	-	-	-
2004	23	3	-	-
2005	13	12	4	-
2006	5	13	31	11
Total	44	28	35	11

Table 5: Estimation of the value of public procurements
(in €)

Year of Contract Award	Year of Call for Tenders 2003	2004	2005	2006
Average Price for Identifiable Non-Large Contract Awards				
2003	3,757,390	-	-	-
2004	2,498,439	1,970,039	-	-
2005	7,928,704	3,615,155	2,146,734	-
2006	24,628,325	12,681,692	4,156,776	2,644,827
Weigted Averge	3,333,283	3,469,439	3,393,493	2,644,827
Sum of Prices for Identifiable Large Contract Awards				
2003	664,078,599	-	-	-
2004	14,403,677,836	1,076,923,738	-	-
2005	4,140,792,968	3,923,211,466	943,473,529	-
2006	1,739,003,711	2,310,356,694	10,037,834,955	3,366,562,661
Total	20,947,553,114	7,310,491,898	10,981,308,484	3,366,562,661

Estimation of the value of Public Procurements based on Identifiable Calls for Tenders

	Year of Call for Tenders 2003	2004	2005	2006
Non-Large Calls	39,756,070,019	42,903,082,690		
Large Calls	20,947,553,114	7,310,491,898		
Total (1)	60,703,623,133	50,213,574,588		
p.m. DG-Markt Estimates (2)	118,290,000,000	80,760,000,000		
Difference between (2) and (1) (in percentage points)	94.9%	60.8%		

Annex II

Results from the application of the backward method to FI

Table 1: Overview of main categories of the TED database

	2003	2004	2005	2006
Entries	2444	2757	2921	3348
Calls for Tenders	1447	1645	1665	1843
Contract Awards	764	819	861	1052
As Percentages of Total Entries				
Calls for Tenders	59.2%	59.7%	57.0%	55.0%
Contract Awards	31.3%	29.7%	29.5%	31.4%

Table 2: Classification of contract awards

	2003	2004	2005	2006
Contract Awards	764	819	861	1052
- with Price	363	487	342	281
- with RN	312	393	367	675
- with Price and RN	139	235	170	195
- Large Contracts	1	0	1	1
- out of which with Price and RN	0	0	1	1
As Percentages of Contract Awards				
Contract Awards				
- with Price	47.5%	59.5%	39.7%	26.7%
- with RN	40.8%	48.0%	42.6%	64.2%
- with Price and RN	18.2%	28.7%	19.7%	18.5%
- Large Contracts	0.1%	0.0%	0.1%	0.1%

Table 3: Call-Contract table

	Year of Contract Award				
Year of Call	2003	2004	2005	2006	Total 2
2000	-	-	1	-	-
2001	2	1	0	-	-
2002	98	0	0	-	-
2003	39	117	8	3	167
2004	-	117	92	6	215
2005	-	-	69	66	135
2006	-	-	-	120	120
Total 1	139	235	170	195	-
As Percentages of Total 1					
2000	-	-	0.6%	-	-
2001	1.4%	0.4%	0.0%	-	-
2002	70.5%	0.0%	0.0%	-	-
2003	28.1%	49.8%	4.7%	1.5%	-
2004	-	49.8%	54.1%	3.1%	-
2005	-	-	40.6%	33.8%	-
2006	-	-	-	61.5%	-
Total	100.0%	100.0%	100.0%	100.0%	-

Table 4: Contract-Call table for identifiable pairs only

Year of Contract Award	Year of Call for Tenders 2003	2004	2005	2006
2003	37	-	-	-
2004	97	100	-	-
2005	7	76	54	-
2006	3	3	46	102
Total	144	179	100	102
p.m. Total Calls for Tenders	1447	1645	1665	1843
As percentages of Total Calls for Tenders	10.0%	10.9%	6.0%	5.5%
Non-Large Contract Awards				
2003	37	-	-	-
2004	97	100	-	-
2005	7	75	54	-
2006	3	3	46	102
Total	144	178	100	102
Large Contract Awards				
2003	0	-	-	-
2004	0	0	-	-
2005	0	1	0	-
2006	0	0	0	0
Total	0	1	0	0

Table 5: Estimation of the value of public procurements (in €)

Year of Contract Award	Year of Call for Tenders 2003	2004	2005	2006
Average Price for Identifiable Non-Large Contract Awards				
2003	2,822,884	-	-	-
2004	3,364,670	2,658,409	-	-
2005	1,217,149	2,943,012	3,004,159	-
2006	7,142,267	27,324,275	3,419,065	2,539,605
Weigted Averge	3,199,768	3,194,043	3,195,015	2,539,605
Sum of Prices for Identifiable Large Contract Awards				
2003	0	-	-	-
2004	0	0	-	-
2005	0	638,000,000	0	-
2006	0	0	0	0
Total	0	638,000,000	0	0

Estimation of the value of Public Procurements based on Identifiable Calls for Tenders

	Year of Call for Tenders 2003	2004	2005	2006
Non-Large Calls	4,630,064,086	5,254,200,686		
Large Calls	0	638,000,000		
Total (1)	4,630,064,086	5,892,200,686		
p.m. DG-Markt Estimates (2)	3,670,000,000	4,410,000,000		
Difference between (2) and (1) (in percentage points)	-20.7%	-25.2%		

European Commission

EUR 23695 EN – Joint Research Centre – Institute for the Protection and Security of the Citizen
Title: The assessment and revision of the public procurement Indicator 3: Total value of published tenders
Author: Athina Karvounaraki
Luxembourg: Office for Official Publications of the European Communities
2008 – 46 pp.
EUR – Scientific and Technical Research series – ISSN 1018-5593

Abstract
The main objective of this study is to assess and revise the existing Indicator 3 that estimates the total value of published tenders. This indicator is currently produced by DG MARKT annually. The work performed by DG JRC focuses on the analysis of the quality of the available data, and the assessment of the methodology used for the calculation of the indicator.

How to obtain EU publications

Our priced publications are available from EU Bookshop (http://bookshop.europa.eu), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

JRC
EUROPEAN COMMISSION

Publications Office
Publications.eu.int