



# EDICOM REPORT

**2005.742**

**Intrastat as a Negative Priority**

**March 2007**

Ellen Jones, Andy Watson and Romesh Paul  
Statistics and Analysis of Trade Unit  
HM Revenue & Customs  
3rd Floor Central  
Alexander House  
21 Victoria Avenue  
Southend-on-Sea SS99 1AA  
United Kingdom



# Contents

1	Background .....	5
2	Implementation timetable .....	7
3	Objectives .....	9
4	Human resources used .....	10
5	Equipment and software applications used .....	12
6	Description of the operation .....	13
6.1	Objectives achieved? .....	13
6.2	Results – Objective 1: Sampling .....	14
6.2.1	Aggregate yearly total matches .....	14
6.2.2	Aggregate correlations .....	15
6.2.3	Number of monthly aggregates within five per cent of actual totals .....	19
6.2.4	Percentage comcode-country combinations represented .....	20
6.2.5	Percentage HS4-country combinations represented .....	21
6.2.6	Percentage of comcode-country combinations within five per cent of actual totals .....	22
6.2.7	Percentage of HS4-country combinations within five per cent of actual totals .....	23
6.3	Work carried out to achieve the results – Objective 1 .....	24
6.3.1	SRS by Commodity and Country using Intrastat data (CC).....	25
6.3.2	SRS by Trade Class and Country using SIC and ESL data (SESL).....	26
6.3.3	SRS by Trade Value (VAL).....	27
6.3.4	Simple Random Sampling (SIM) .....	28
6.4	Results – Objective 2: Workshop .....	29
6.5	Work carried out to achieve the results – Objective 2: Workshop .....	31
6.6	Developments planned for the coming years .....	31
7	Summary of objectives and main results achieved .....	32

## Annexes:

A.	Contract with Consultant .....	33
B.	Breakdown of Workshop and Seminar travel and subsistence reimbursements .....	36
C.	Workshop and Seminar Agenda .....	37
D.	‘Intrastat simplification: Past, present and future’, Aleš Čapek .....	41
E.	‘Implications of Single Flow for data processing’, Jan Planovsky .....	45
F.	‘Prerequisites for a Single Flow system in the EU’, Peter Ottosen.....	46
G.	‘Administrative Burden in the Netherlands’, Marjolijn Jaarsma .....	48
H.	‘Asymmetries’, Søren Rich .....	51
I.	‘Asymmetries with the UK’, Andrew Kochen .....	54

J. 'Implications of 'Simple' and 'Qualified' Single Flow systems', Walter Seiringer.....	57
K. 'The Stratified Random Sampling option', Ellen Jones .....	60
L. 'Consequences of reducing nomenclature from CN8 to HS6', Lars Malmborg.....	64
M. 'The Threshold-raising option', Frank Weideskog .....	67
N. 'Simplification of the Threshold', Lien van Driessche .....	72
O. 'Implications of raising the UK Threshold', Rafael Mastrangelo .....	75
P. 'Declarant Burden', Mark Kelly .....	78
Q. 'UK national needs and uses of aggregate trade data', Stuart Brown.....	81
R. 'UK national needs and uses of detailed trade data', Henry Bottomley .....	83
S. 'Preferred simplification option', Sandra Tudor .....	87
T. 'Intrastat Simplification Workshop and Seminar Report', Guillaume Talon.....	89

## **Abstract**

Intrastat is the system used for collecting data on trade-in-goods within the European Union (EU). By simplifying Intrastat the burdens on business could be reduced, but this must be balanced by the continuing need for the trade statistics generated from Intrastat data. One possible solution is to use Stratified Random Sampling (SRS) to reduce the number of traders required to submit declarations, while preserving coverage of the various trade sectors as far as possible.

The SRS method was simulated and evaluated in this project. It was concluded that SRS is not in fact a promising or feasible option for the simplification of Intrastat, due to the complexity of the sampling system and the inconsistency of the figures returned.

In October 2006 an international Intrastat Simplification Workshop and Seminar were held. Delegates attended from many EU Member States (MS) and research was shared and various options discussed. It was concluded that the most promising simplification options were raising the threshold and 'single flow'. Raising the threshold would mean reducing the number of traders required to submit declarations by relieving those with the lowest annual intra-EU trade value; this option would be relatively straightforward to implement. However, it would cause an increase in asymmetries as MS with large trade volumes would lose from their figures almost all trade with MS with smaller trade volumes. The 'single flow' option would entail all MS collecting only their outgoing trade (dispatch) data, and reconstructing their incoming trade (arrivals) from other MS' dispatches. This option also promises a large reduction in burden on business, but is not feasible for introduction in the short-term as there are many issues which must first be addressed.

A Eurostat Working Group has been set up to take forward the work on Intrastat simplification. It is due to report to the Council of European Finance Ministers (ECOFIN) in October 2007.

# 1 Background

Trade-in-goods statistics provide vital economic information that is needed to monitor the micro and macro economic climate, for trade policy and monitoring trade agreements, for the System of National Accounts (SNA) and Balance of Payments statistics (BoP).

Due to the increasing globalisation of national economies, the importance of international trade statistics as a decision-making tool in international organizations continues to grow. The European Union, the European Central Bank, the World Trade Organization, the World Bank, the International Monetary Fund, and the United Nations, with its various subsidiary organizations responsible for economy and development, all depend on detailed knowledge of global trade flows to be able to carry out their work effectively.

Within the EU, trade statistics are important short-term indicators for monitoring economic trends and the trade cycle. Detailed figures on trade statistics help to track the development of home and foreign markets and in evaluating the competitive situation. Companies and trade associations also use this information to compare the sales strategy of different producers.

Intrastat was introduced in 1993 with the establishment of the European Single Market as a simplified declaration system for intra-EU trade in goods. The drive towards simplification has remained strong and further simplifications have followed regularly. Compilation of UK-EU trade data currently involves collection of detailed declarations from about 33,000 traders, who are selected using a threshold in order to capture 97 per cent of the trade value. Similar systems are in force throughout the EU. In the UK the submission process requires an average of 3.25 hours a month, representing an annual cost to business of approximately £37m<sup>1</sup>.

Motivated by potential reductions in burdens on business, Intrastat simplification is now again on the political agenda. Careful consideration and innovative solutions are needed to balance this with the enduring demand for high quality detailed intra-EU trade statistics.

This project involved the exploration of the possibility of using SRS of the trader population to reduce the number of businesses required to submit declarations while maintaining acceptable accuracy at detailed level.

SRS involves stratification of the respondent population (sampling frame) and random sampling. In this project the sampling frame consisted of traders in goods between the UK and the rest of the EU who were above the Intrastat threshold for 2004<sup>2</sup>. Stratification enhances random sampling by ensuring coverage of all sectors of the population. It involves grouping together traders with similar trade so that the variation within strata is reduced and estimates are therefore more reliable. Possible stratification criteria include:

- Trade value
- Commodity (product type)
- Partner country

---

<sup>1</sup> Intrastat Compliance Costs 2004-2005, Statistics and Analysis of Trade Unit, HRMC

<sup>2</sup> Traders with annual EU imports or exports of value greater than £221,000 in 2003, or whose cumulative EU imports or exports cross the threshold (also £221,000) during 2004.

The best criterion by which to stratify the trader register in order to minimise variation is the trade value. This was investigated in a study in Germany in 1999<sup>3</sup>. It obtained a reasonable match at partner country level but a poor match at commodity level. With the aim of improving the accuracy at detail level the major stratification criteria investigated in this study were commodity and partner country. Stratification by value and simple random sampling were tested in parallel as comparisons. Sample sizes used were a half, a third and a quarter of those traders currently on the Intrastat register. It was considered that including below-threshold traders<sup>4</sup> in a sampling system would represent a small return in terms of trade value collected, and a significant increase in burden on business and demand for HMRC educational resources.

Also prescribed in this project was the holding of an international workshop and discussion forum for issues relating to the simplification of Intrastat. Many MS have conducted and are in the process of conducting research into Intrastat simplification possibilities. It is important that the various parties are informed of each other's activities so that previous knowledge can be built upon and duplicate studies are not carried out.

---

<sup>3</sup> Durchführbarkeit der Intrahandelsstatistik auf Stichprobenbasis, Statistisches Bundesamt, Wirtschaft und Statistik 8/1999, Wolf Bihler

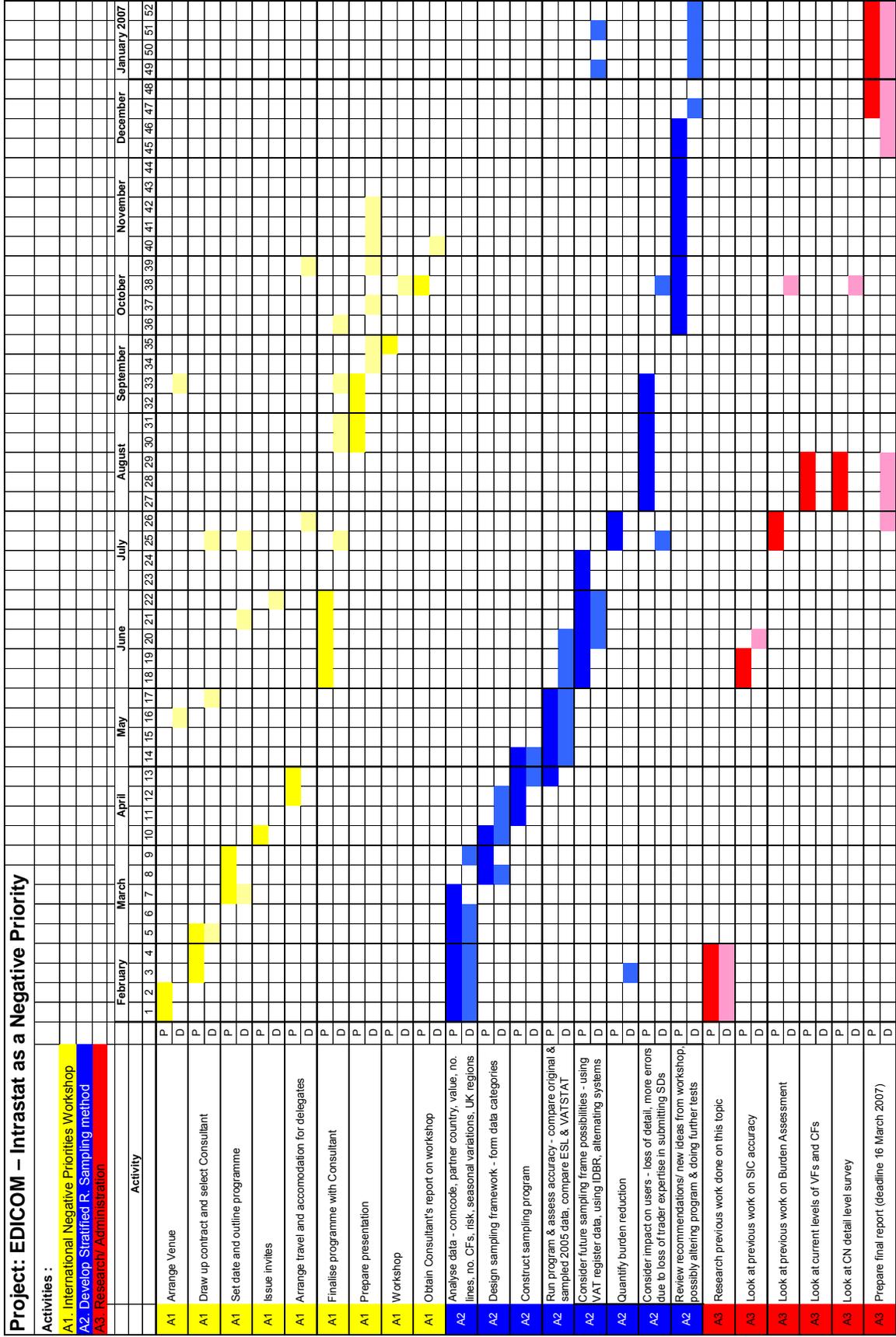
<sup>4</sup> VAT registered traders who trade in goods with other Member States but are below the Intrastat threshold.

## **2 Implementation timetable**

Figure one shows the planned timetable and actual completion record for actions relating to this project. Items to note include:

- workshop presentation preparation continued past the workshop date – this refers to preparation and distribution of CD-ROMs of workshop presentations to delegates; and
- it had been hoped that some extra research work into the impact of a sampling system on the levels Credibility and Validity Failures could be carried out, but this was not feasible within the constraints of the project.

Figure 1: Project timetable and completion record



### **3 Objectives**

#### **1. Stratified Random Sampling of Intrastat traders**

Fully explore and comment on the method of performing a stratified random sample, by which it might be possible to reduce the burden on business of the present Intrastat system. Where possible the reduction of burden on business will be quantified, along with comments regarding the suitability, accuracy and feasibility of such an action, and the ramifications thereafter.

#### **2. International Intrastat Simplification Workshop**

A workshop, in the UK, will take place for invited delegates of Member States (MS) who have already expressed an interest in “Intrastat as a Negative Priority”. This workshop will be lead by a contractor with expertise in macro and micro economic indicators and knowledge of customer needs. Opportunity will be provided for delegates to demonstrate their actions in relation to this subject and provide comment on work completed and in progress in the UK and other MS.

## 4 Human resources used

The time spent on this project by HMRC staff of various grades can be seen in part A, section a of the financial summary shown in figure two.

A Consultant was also recruited to work on the project (see financial summary part A, section e) via a restricted invitation to tender. A copy of the Consultant's contract can be found in annex A. Table one shows a summary of the time spent, work done and remuneration received by the Consultant.

**Table 1: Tasks undertaken and payment received by Consultant**

<b>Task</b>	<b>Time /days</b>	<b>Payment /€</b>
Quality assurance of UK's SRS project	6	2.690
Quality assurance of other contributions	11	4.930
Workshop agenda preparation	2	900
Workshop leading	3	1.340
Preparation of the workshop report	3	1.340
	<b>Total</b>	<b>11.200</b>

**Figure 2: Project financial summary**

Summary Financial Statement for Grant Agreement 53102.2005.001-2005.742

Period of operation 1st February 2006 until 31st January 2007

Part A Expenditure/eligible Costs	£	Exchange Rate	Euros	Part B Financing plan	£	Exchange Rate	Euros
1. cost of staff assigned to the operation				a. direct revenue expected from the operation			
Senior Statistician 5.7 Days @ £395.06	2251.84						
Statistician 0.9 days @ £340.5	306.45						
Statistician 10.8 days @ £284.00	3067.20						
Senior Statistical Officer 0.1 days @ £236.11	23.61						
Senior Statistical Officer 9.10 days @ £225.22	2049.502						
Senior Statistical Officer 0.1 days @ £225.91	22.59						
Higher Statistical Officer 8.20 days @ £170.76	1400.232						
Higher Statistical Officer 0.1 days @ £166.55	16.66						
Higher Statistical Officer 1.10 days @ £151.41	166.55						
Higher Statistical Officer 0.2 days @ £148.73	29.75						
Higher Statistical Officer 121.3 days @ £126.8	15290.81						
Senior Officer 1.4 days @£224.72	314.61						
Business Manager 1.4 days @ 225.91	316.27						
Operational Manager 1.4 days @ 298.66	418.12						
2. travel and subsistence expenses for staff involved in the operation	1630.73	0.6613	38823.83	b. contribution by the applicant @ 10%	3714.13	0.6613	5616.41
3. costs of equipment, land and immovable property (rent, purchase)	0.00	0.6613	0.00	c. contribution by other organisations (please specify)	0.00	0.6613	0.00
4. cost of consumables and supplies	0.00	0.6613	0.00	d. contribution requested from the Commission	33427.16	0.6613	50547.65
5. Subcontracting	7406.56	0.6613	11200.00				
6. other direct costs	0	0.6613	0.00	e. estimated bank interest generated by the grant requested			
Contingency Reserve @ 5%					0.00	0.6613	0.00
7. general costs charged to the operation - indirect costs of 7% of eligible costs per grant agreement.	2429.80	0.6613	3674.28	f. other contributions by the Commission for the same operation (please specify)	0.00	0.6613	0.00
<b>Total Eligible Costs</b>	<b>37141.29</b>		<b>56164.06</b>	<b>Total</b>	<b>37141.29</b>		<b>56164.06</b>
Estimate of Contribution in kind for the operation	0.00	0.6613	0.00	Estimate of the contributions in kind, where applicable	0.00	0.6613	0.00
<b>Total</b>	<b>37141.29</b>		<b>56164.06</b>	<b>Total</b>	<b>37141.29</b>		<b>56164.06</b>

## **5 Equipment and software applications used**

Microsoft Office software was used throughout this project. For research Adobe Acrobat software was also used.

For completion of objective one, SAS Statistical Analysis Software version nine was used on a Solaris Unix box.

For completion of objective two, use was made of the Churchill Conference Room in the Government Offices Great George Street building in Westminster, London. Use of this facility incurred no additional cost to HMRC. Refreshments were provided during the events by local catering company Charlton House Catering; the cost of this is included in part A, section f of the financial summary on the previous page. Delegates attending from countries which had already expressed an interest in Intrastat simplification (Denmark, the Netherlands, Portugal and Sweden) received the offer of reimbursement of the travel and subsistence costs. This resulted in a cost to HMRC of £1,630.73, which appears in part A, section b of the financial summary (please see annex B for a breakdown of this figure). Subsequent to the workshop and seminar, CD-ROMs were produced containing all presentations and conclusions from the events. These were produced using NERO software and were distributed to all delegates and other parties who had expressed an interest.

## **6 Description of the operation**

### **6.1 Objectives achieved?**

#### **1. Stratified Random Sampling of Intrastat traders**

Yes: the method of using SRS to select traders for Intrastat declaration was fully explored and comments made on the suitability, accuracy and feasibility of such an action, the ramifications thereafter and the burden reduction possible.

Four sampling methods were investigated:

- SRS by commodity and country using Intrastat data – (CC);
- SRS by trade class and country using Standard Industrial Classification (SIC) and EC Sales List (ECSL) data – (SESL);
- SRS by trade value – (VAL); and
- simple random sampling – (SIM).

Each sampling method was trialled with sample sizes of a half, a third and a quarter of traders currently on the Intrastat register. This represented burden reductions of a half, two-thirds and three-quarters respectively.

The VAL method was the most successful at producing accurate results at both aggregate and detailed levels. The quality was sufficient to justify further study, similar to results produced by modelling an increase in the assimilation threshold. However, the threshold raising method would be technically and practically much simpler than introducing an SRS system. Therefore it was not recommended that further research be carried out into the use of SRS in Intrastat.

#### **2. International Intrastat Simplification Workshop**

Yes: an Intrastat Simplification Workshop and Seminar took place in London on 16-18 October 2006. It was attended by delegates from 15 Member States, Eurostat and various UK government departments. Presentations and discussions covered burden assessment, simplification options such as Single Flow, Threshold Raising, Stratified Sampling and Nomenclature Simplification, and the impact of simplification on users of trade statistics.

## 6.2 Results – Objective 1: Sampling

### 6.2.1 Aggregate yearly total matches

Tables two and three show the percentages of the actual 2005 trade total which would have been obtained with each sampling method at the three sample sizes.

Overall, the CC and VAL stratifications produced the closest aggregate yearly totals, with average discrepancies of 4.3 per cent and 4.8 per cent respectively. The VAL stratification method exhibited a much greater degree of stability over the different sample sizes. The SIM method produced a slightly higher average discrepancy of 5.5 per cent and showed more variation, with totals being higher for dispatches than arrivals, and for smaller sample sizes in both suites. The SESL stratification produced a consistent underestimation, with an average magnitude of 7.8 per cent. The totals were lowest for the 33 per cent sample in both suites.

**Table 2: Aggregate yearly totals - Arrivals (percentage of actual)**

Sampling method	Sample size		
	25 per cent	33 per cent	50 per cent
CC	98	96	96
SESL	91	86	97
SIM	97	95	89
VAL	95	95	96

Source: HM Revenue & Customs Overseas Trade Statistics

**Table 3: Aggregate yearly totals - Dispatches (percentage of actual)**

Sampling method	Sample size		
	25 per cent	33 per cent	50 per cent
CC	106	94	96
SESL	95	89	95
SIM	108	103	97
VAL	95	95	95

Source: HM Revenue & Customs Overseas Trade Statistics

## 6.2.2 Aggregate correlations

Tables four and five show the correlations between the actual monthly aggregate totals and those obtained under the various sampling methods and sample sizes. Charts one to six on the following pages show this in graphical form.

The VAL and SESL stratification methods obtained the best correlations with averages of 0.94 and 0.93 respectively. Both sets of results were slightly better for arrivals than dispatches, and the VAL results again varied very little with sample size. The SIM method followed with an average of 0.84, improving with sample size and again better for arrivals than dispatches. The CC stratification method performed very poorly, with an average correlation of only 0.30. No relation could be observed with sample size, but on average the results were again slightly better for arrivals than dispatches.

**Table 4: Correlations of aggregate monthly totals with actuals - Arrivals**

Sampling method	Sample size		
	25 per cent	33 per cent	50 per cent
CC	0.33	0.13	0.55
SESL	0.92	0.96	0.95
SIM	0.81	0.85	0.93
VAL	0.95	0.95	0.95

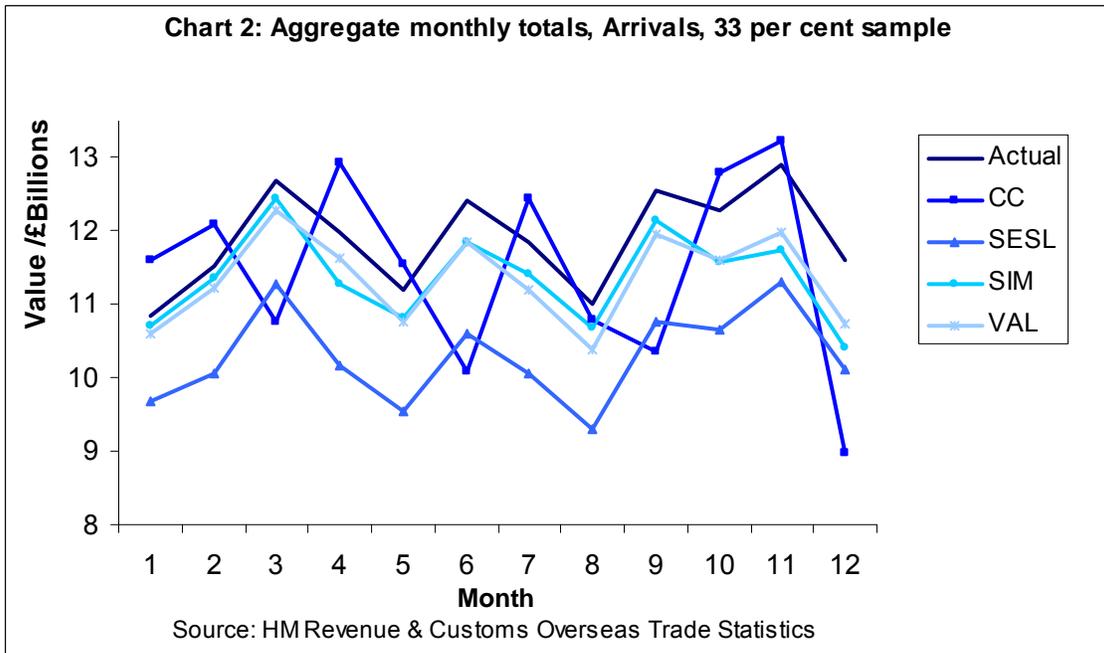
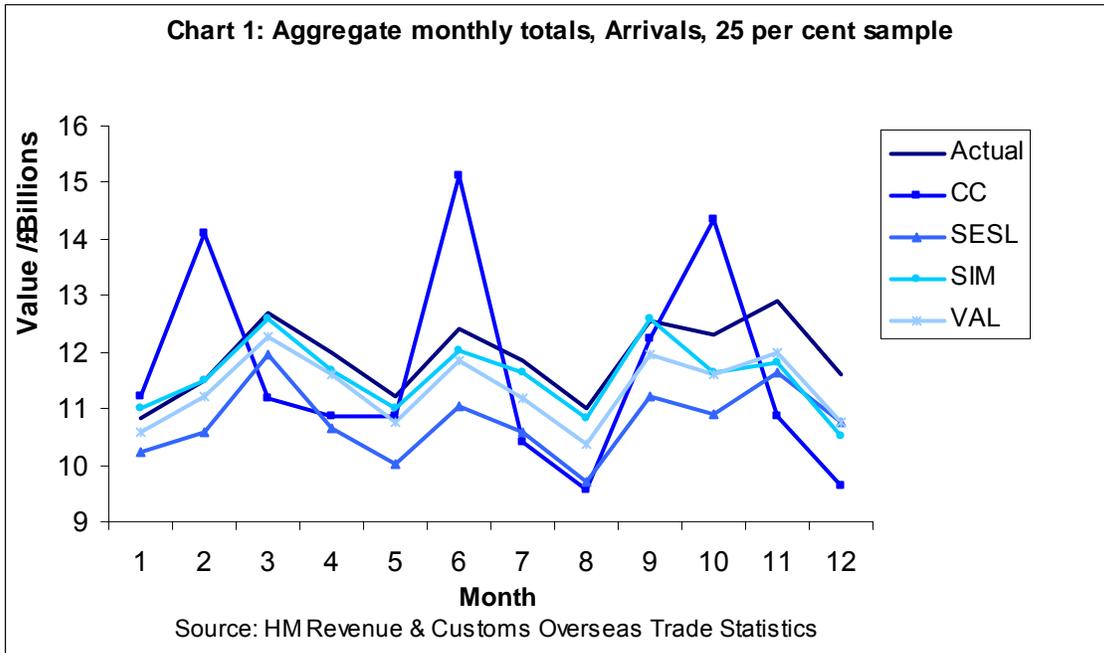
Source: HM Revenue & Customs Overseas Trade Statistics

**Table 5: Correlations of aggregate monthly totals with actuals - Dispatches**

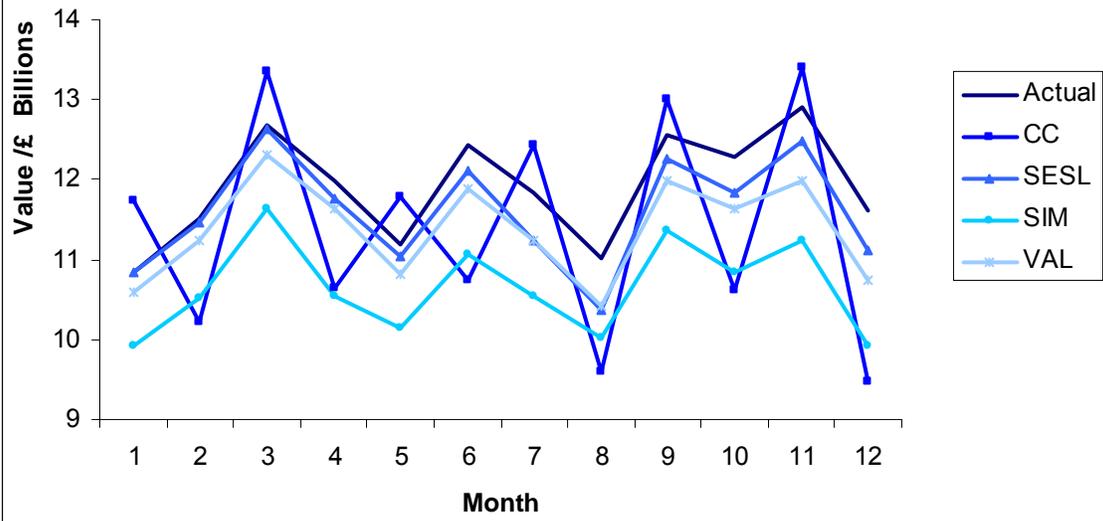
Sampling method	Sample size		
	25 per cent	33 per cent	50 per cent
CC	0.23	0.30	0.23
SESL	0.91	0.95	0.91
SIM	0.78	0.78	0.87
VAL	0.93	0.93	0.94

Source: HM Revenue & Customs Overseas Trade Statistics

Charts one to six show that the VAL stratification results matched the actual totals most closely overall. The results from the SESL stratification show good correlations but are commonly affected by a general underestimation (most noticeable in charts two and four, relating to a 33 per cent sample size). The SIM results were very good in charts one, two and six, but a general underestimation affected the values in chart three and anomalies in August and November lowered the correlations in charts four and five. The CC stratification produced the most erratic results in all cases.

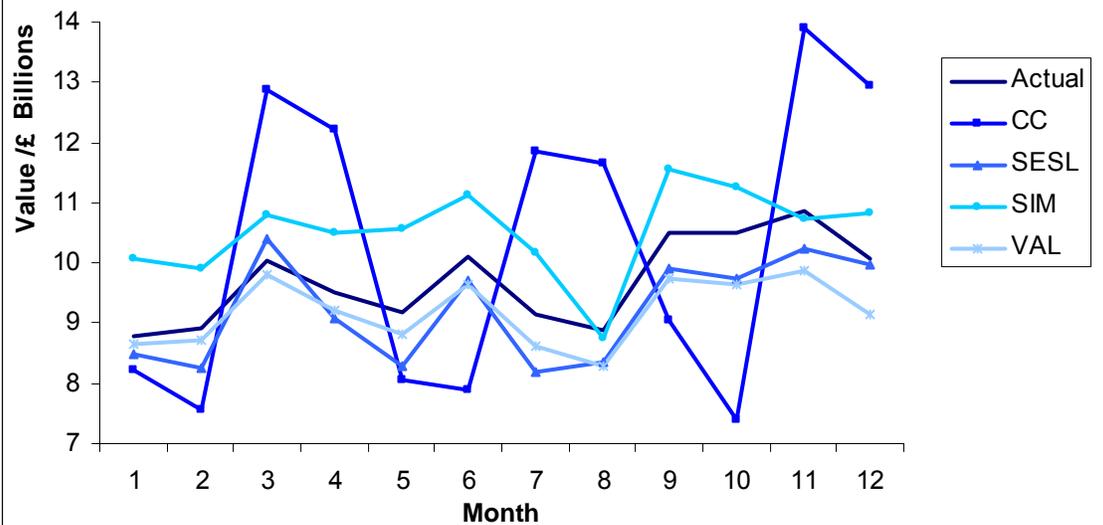


**Chart 3: Aggregate monthly totals, Arrivals, 50 per cent sample**



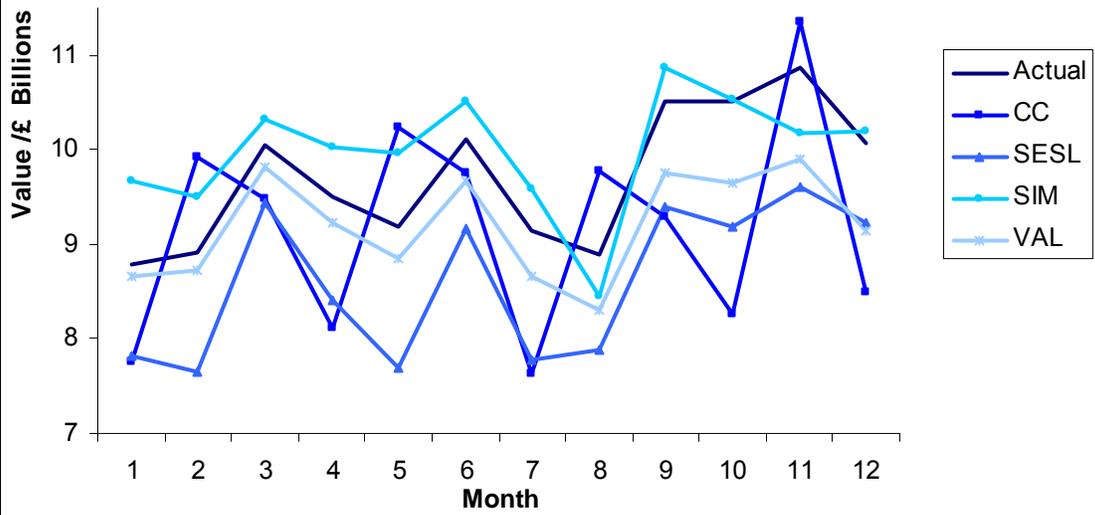
Source: HM Revenue & Customs Overseas Trade Statistics

**Chart 4: Aggregate monthly totals, Dispatches, 25 per cent sample**



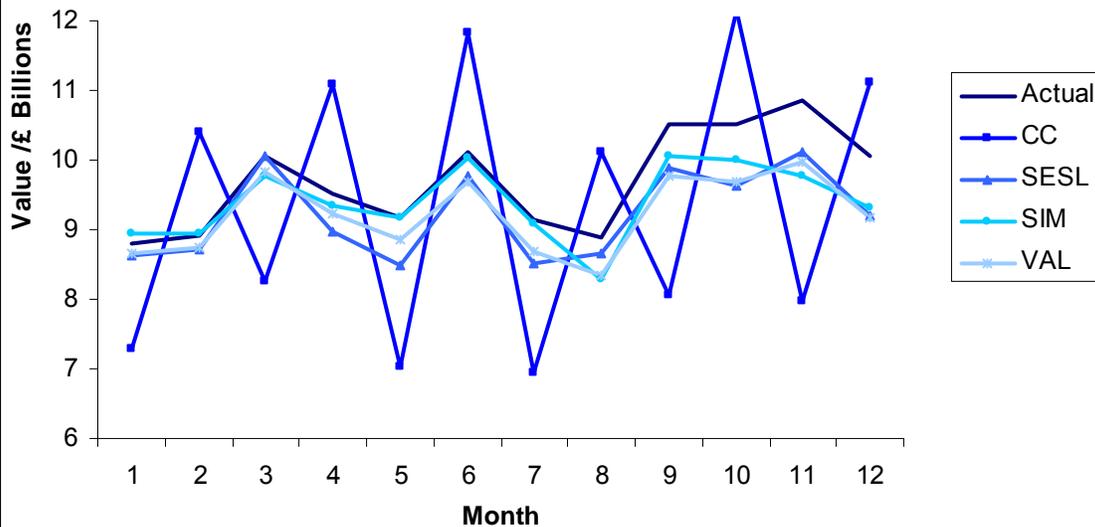
Source: HM Revenue & Customs Overseas Trade Statistics

**Chart 5: Aggregate monthly totals, Dispatches, 33 per cent sample**



Source: HM Revenue & Customs Overseas Trade Statistics

**Chart 6: Aggregate monthly totals, Dispatches, 50 per cent sample**



Source: HM Revenue & Customs Overseas Trade Statistics

### 6.2.3 Number of monthly aggregates within five per cent of actual totals

Tables six and seven show the numbers of monthly aggregates which were within five per cent of the actual totals, under each of the sampling methods and at each sample size.

The VAL stratification method produced the best results. The totals were within five per cent of the actual totals for seven months for arrivals and six months for dispatches, independent of sample size. The second best results were obtained from the SIM method, which twice achieved nine monthly totals within five per cent of actual totals. However, this method also experienced some noticeable drops for arrivals at 50 per cent and dispatches at 25 per cent. This was due to a general underestimation seen in results for arrivals at 50 per cent and an overestimation affecting dispatches at 25 per cent (these can be seen in charts three and four). The CC stratification method totals matched well with the actual totals for an average of four months for arrivals, but only one month for dispatches. Results were higher for the 33 per cent sample size in both suites. The SESL stratification method produced very variable results. The cases for which the SESL stratification results were not within five per cent of the actual totals in any month were affected by a general underestimation (this can be seen in charts one, two and five).

**Table 6: Monthly aggregate totals within five per cent of actuals - Arrivals**

Sampling method	Sample size		
	25 per cent	33 per cent	50 per cent
CC	3	5	3
SESL	0	0	10
SIM	9	8	0
VAL	7	7	7

Source: HM Revenue & Customs Overseas Trade Statistics

**Table 7: Monthly aggregate totals within five per cent of actuals - Dispatches**

Sampling method	Sample size		
	25 per cent	33 per cent	50 per cent
CC	0	2	0
SESL	5	0	5
SIM	2	7	9
VAL	6	6	6

Source: HM Revenue & Customs Overseas Trade Statistics

#### 6.2.4 Percentage comcode-country combinations represented

This is a measure of the coverage at detailed level afforded by each of the sampling methods. Tables eight and nine show the percentages of all actual comcode-country combinations which were populated with trade under each of the sampling methods and sample sizes.

The VAL stratification produced the best results, with an average of 85.3 per cent of combinations represented, improving with increasing sample size and better for dispatches than arrivals. The CC stratification method followed closely with an average of 82.8 per cent of combinations represented, again improving with sample size and better for dispatches. The SIM and SESL stratification methods produced similar results, with averages of 65.7 per cent and 63.8 per cent of combinations respectively. These results also improved with sample size.

**Table 8: Comcode-country combinations containing trade - Arrivals  
(percentage of actual number of combinations)**

Sampling method	Sample size		
	25 per cent	33 per cent	50 per cent
CC	75	82	88
SESL	55	62	74
SIM	57	64	76
VAL	78	83	89

Source: HM Revenue & Customs Overseas Trade Statistics

**Table 9: Comcode-country combinations containing trade - Dispatches  
(percentage of actual number of combinations)**

Sampling method	Sample size		
	25 per cent	33 per cent	50 per cent
CC	78	84	90
SESL	53	63	76
SIM	57	64	76
VAL	83	87	92

Source: HM Revenue & Customs Overseas Trade Statistics

## 6.2.5 Percentage HS4-country combinations represented

Tables ten and 11 are similar to tables six and seven, but for a lower level of detail: they show the percentages of all actual HS4-country combinations which were populated with trade under each of the sampling methods and sample sizes.

Again the VAL stratification produced the best results. On average 90.0 per cent of combinations were represented, closely followed by the CC stratification at 89.0 per cent of combinations on average. These figures represent improvements of about six per cent on the percentages of comcode-country combinations represented. The SIM and SESL stratification methods again produced similar results, with averages of 77.3 per cent and 76.5 per cent respectively. These were improvements of about 12 per cent on the percentages of comcode-country combinations represented. All results improved with increasing sample size and were slightly better for dispatches than arrivals.

**Table 10: HS4-country combinations containing trade - Arrivals  
(percentage of actual number of combinations)**

Sampling method	Sample size		
	25 per cent	33 per cent	50 per cent
CC	81	87	91
SESL	68	73	82
SIM	69	75	83
VAL	83	87	92

Source: HM Revenue & Customs Overseas Trade Statistics

**Table 11: HS4-country combinations containing trade - Dispatches  
(percentage of actual number of combinations)**

Sampling method	Sample size		
	25 per cent	33 per cent	50 per cent
CC	88	92	95
SESL	71	78	87
SIM	72	79	86
VAL	90	92	96

Source: HM Revenue & Customs Overseas Trade Statistics

## 6.2.6 Percentage of comcode-country combinations within five per cent of actual totals

Tables 12 and 13 show the percentages of all actual comcode-country combinations which contained trade values within five per cent of the actual totals, under each of the sampling methods and at each sample size.

Results of this analysis were poor for all cases except the VAL stratification at a 50 per cent sample size. This gave about 70 per cent of comcode-country totals within five per cent of the actual values for both suites, which was by far the best obtained. The second best results were given by the CC stratification at a 50 per cent sample and the VAL stratification at a 33 per cent sample, both of which gave matches of about 15 per cent before deteriorating for lower sample sizes. Both the SESL stratification and the SIM methods gave very low results at all sample sizes.

**Table 12: Comcode-country combination totals within five per cent of actuals - Arrivals (percentage of combinations containing trade)**

Sampling method	Sample size		
	25 per cent	33 per cent	50 per cent
CC	7	12	16
SESL	2	2	3
SIM	3	3	3
VAL	10	13	68

Source: HM Revenue & Customs Overseas Trade Statistics

**Table 13: Comcode-country combination totals within five per cent of actuals - Dispatches (percentage of combinations containing trade)**

Sampling method	Sample size		
	25 per cent	33 per cent	50 per cent
CC	6	10	14
SESL	2	2	3
SIM	2	2	3
VAL	11	16	70

Source: HM Revenue & Customs Overseas Trade Statistics

## 6.2.7 Percentage of HS4-country combinations within five per cent of actual totals

Tables 14 and 15 show the percentages of all actual HS4-country combinations which contained trade values within five per cent of the actual totals, under each of the sampling methods and at each sample size.

Again the best results were produced by the VAL stratification at a 50 per cent sample size. This gave about 60 per cent of HS4-country totals within five per cent of the actual values for both suites, ten per cent lower than its percentage of comcode-country values within five per cent. The results for the other sample sizes and sampling methods were higher than for the comcode-country comparison, with the second best percentages being about 20 per cent, again from the CC stratification at a 50 per cent sample and the VAL stratification at a 33 per cent sample. Results for both these then deteriorated for the lower sample sizes, and the SESL stratification and the SIM methods again gave very low results at all sample sizes.

**Table 14: HS4-country combination totals within five per cent of actuals - Arrivals (percentage of combinations containing trade)**

Sampling method	Sample size		
	25 per cent	33 per cent	50 per cent
CC	8	16	22
SESL	3	3	4
SIM	3	3	4
VAL	12	17	58

Source: HM Revenue & Customs Overseas Trade Statistics

**Table 15: HS4-country combination totals within five per cent of actuals - Dispatches (percentage of combinations containing trade)**

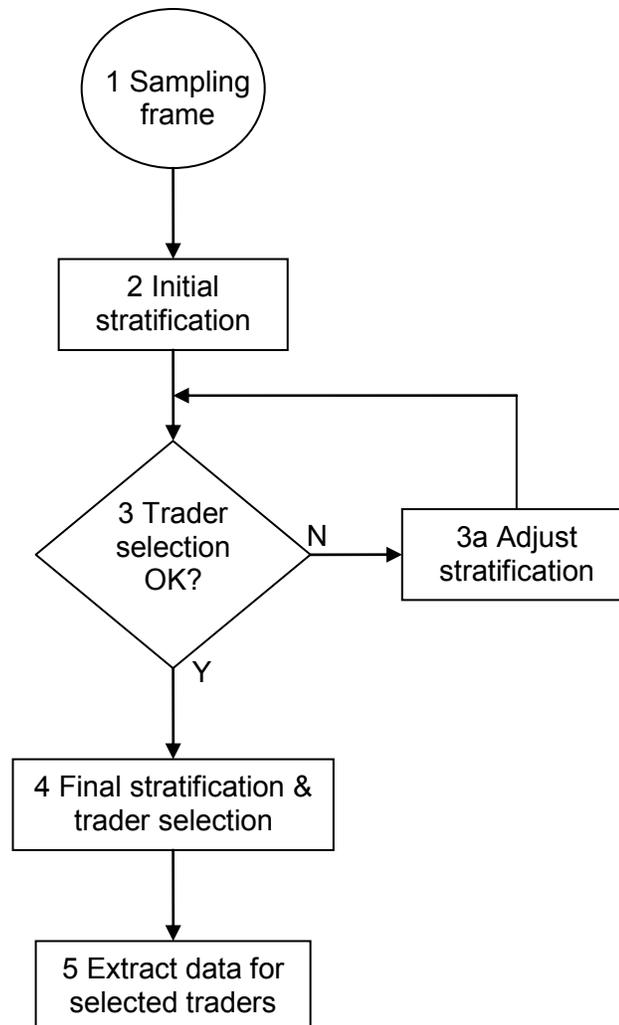
Sampling method	Sample size		
	25 per cent	33 per cent	50 per cent
CC	8	15	20
SESL	3	3	4
SIM	3	3	4
VAL	14	21	60

Source: HM Revenue & Customs Overseas Trade Statistics

### 6.3 Work carried out to achieve the results – Objective 1

The sampling processes used for each method (CC, SESL, VAL and SIM) followed a similar design. This is summarised in figure 3 and then further details are given for each method in the following sections.

**Figure 3: Process for developing stratifications and sampling traders**



### 6.3.1 SRS by Commodity and Country using Intrastat data (CC)

#### Stage 1: Sampling frame

The sampling frame used for this method was constructed from the previous year's Intrastat data, which was the option with the greatest level of detail available. This required specialised sampling arrangements in order to enable sampling frames to continue to be developed subsequent to an SRS system being introduced. These arrangements entailed traders being sampled for certain months only. This would mean that all above-threshold traders could continue to be sampled, while still providing a burden reduction due to the decreased frequency of submission.

#### Stage 2: Initial Stratification

Initially the sampling frame dataset was stratified into HS4 (four digit commodity code)-country categories. There were 16,455 arrivals categories and 22,532 for dispatches.

#### Stage 3: Check Trader Selection

The reliability of an estimate based on a sample drawn from a certain category of the population is variable. It is increased if that category contains a large number of respondents, and if the variation between the respondents' values is low. Initially, the CC stratification produced a very large number of categories, many of which contained only a few traders. Due to the volatile nature of the trader population it was particularly important that all categories contained a substantial number of traders, in order for the samples drawn to be representative. Therefore adjustment of the categories was necessary.

#### Stage 3a: Adjust Stratification

The categories were adjusted by aggregating them to a lower level of detail. This increased the number of traders per category, and joining only similar categories together preserved the variation-limiting effect of the stratification. The variation was assessed using the Neyman coefficient<sup>5</sup>. This takes into account the number of traders included and the extent of variation in their trade values and gives the optimal sample size for a each category. Categories were selected for aggregation if:

- the number of traders was less than 60; or
- the Neyman coefficient was greater than 0.5.

The Neyman coefficient limit of 0.5 represents a 50 per cent sample, which was the largest sample size to be investigated. The minimum number of traders was set at 60 as this provided a reasonable sample size at all sampling percentages and resulted in a large but manageable number of final categories, representing a thorough stratification. Categories fulfilling these criteria were aggregated together, to chapter (two digit

---

<sup>5</sup> Neyman coefficient  $N_i = \frac{\sigma_i \sum_{i=1}^k n_i}{\sum_{i=1}^k \sigma_i n_i}$  for category i of k categories

where  $\sigma_i$ =standard deviation of trader trade values in category i

$n_i$ =number of traders in category i

commodity code)-country level. The criteria were then applied again, and categories still fulfilling them were aggregated to chapter level.

#### **Stage 4: Final Stratification**

After the stratification adjustments had been carried out there were 2,505 arrivals categories, consisting of 1,778 at HS4-country level, 630 at chapter-country level and 97 at chapter level. There were 3,206 categories for dispatches, consisting of 2,128 at HS4-country level, 986 at chapter-country level and 92 at chapter level.

#### **Stage 5: Data Extraction**

Trade was extracted from the 2005 dataset for the selected traders. This was then summarised by comcode and partner country and scaled up using a simple scaling factor corresponding to the number of traders sampled<sup>6</sup>.

### **6.3.2 SRS by Trade Class and Country using SIC and ESL data (SESL)**

#### **Stage 1: Sampling frame**

The sampling frame used for this method was constructed from the traders' SIC codes and the previous year's ECSL data. The SIC data are created when traders apply to be registered for VAT, and for this project were extracted from the HMRC administrative VAT register. They consist of five digit codes representing the industrial sector in which the traders primarily trade. There are currently some concerns about the reliability of the SIC data from this source; work is ongoing to address this.

ECSL data are collected monthly or quarterly from traders making dispatches to the EU, recording each trader's trade values with each of their partner businesses. These data are exchanged among EU member states each quarter, allowing the construction of UK arrivals data from partner countries' dispatches.

#### **Stage 2: Initial Stratification**

Initially the traders were stratified into SIC-country categories: This resulted in 635 categories for arrivals and 3,542 for dispatches. There were many more categories for dispatches than arrivals because the ECSL data were much more complete for dispatches.

#### **Stage 3: Check Trader Selection**

Initially many categories contained only a few traders; therefore it was necessary to adjust the stratification.

#### **Stage 3a: Adjust Stratification**

A similar procedure was followed to that used to adjust the CC stratification: Categories were again adjusted by aggregating them to a lower level of detail. The combination criteria used were slightly different, with the lower limit of the number of traders being set

---

<sup>6</sup> Scaling factor = 100/sample percentage

at 24 rather than 60. This was done because the categories in general contained fewer traders and it was desired not to reduce the number of final categories too greatly.

The first wave of aggregations consisted of aggregating the categories to SIC five digit level. The second wave involved aggregations to SIC three digit level. Once these had been completed it was noted that a large number of categories still contained fewer than 24 traders; hence these categories were combined into a single 'low number of traders' category.

#### **Stage 4: Final Stratification**

After the stratification adjustments had been carried out there were 188 arrivals categories, consisting of 100 at SIC five digit level, 87 at SIC three digit level and the one 'low number of traders' category. There were 277 categories for dispatches, consisting of 74 at SIC-country level, 124 at SIC five digit level, 78 at SIC three digit level and the one 'low number of traders' category.

#### **Stage 5: Data Extraction**

Data were extracted and scaled up in the same way as for the CC sampling method.

### **6.3.3 SRS by Trade Value (VAL)**

#### **Stage 1: Sampling frame**

The sampling frame used for this method was constructed from above threshold traders' total trade values for the previous year. This information was taken from traders' 2004 Intrastat declarations for this project. Were the current Intrastat survey to be replaced by a sampling system, these data can also be obtained from the traders' VAT declarations, as these contain total values for EU arrivals and dispatches.

#### **Stage 2: Initial Stratification**

Traders were initially stratified into 100 categories according to their trade values, such that each category contained an equal number of traders.

#### **Stage 3: Check Trader Selection**

The number of traders to be sampled from each category was decided using the Neyman sampling allocation. This involved using the Neyman coefficients to define the ratio by which the sampled traders were split between the categories. All traders were sampled from the higher value categories, and then the sampled proportion decreased as the value decreased.

#### **Stage 5: Data Extraction**

Data were extracted and scaled up in the same way as for the previous sampling methods.

#### **6.3.4 Simple Random Sampling (SIM)**

No sampling frame or stratification was needed for this method; just the above threshold trader population. This was taken from the 2004 Intrastat register for this project, but can also be obtained from the traders' VAT declarations. The trader population was randomly sorted and traders selected from the top of the list. Data extraction and scaling then proceeded in the usual way.

## 6.4 Results – Objective 2: Workshop

The following presentations were made at the Intrastat Simplification Workshop and Seminar. A copy of the agenda, along with delegate details, can be found in annex C.

**Table 16: Locations of presentation slides from Workshop and Seminar**

<b>Subject</b>	<b>Presented by</b>	<b>Located in annex</b>
Intrastat Simplification: Past, present and future	Ales Capek EU	D
Implications of Single Flow for data processing	Jan Planovsky EU	E
Prerequisites for a Single Flow system in the EU	Peter Ottosen DK	F
Administrative Burden in the Netherlands	Marjolijn Jaarsma NL	G
Asymmetries	Søren Rich DK	H
Asymmetries with the UK	Andrew Kochen UK	I
Implications of ‘Simple’ and ‘Qualified’ Single Flow systems	Walter Seiringer AT	J
The Stratified Random Sampling option	Ellen Jones UK	K
Consequences of reducing nomenclature from CN8 to HS6	Lars Malmberg SE	L
The Threshold-raising option	Frank Weideskog SE	M
Simplification of the Threshold	Lien van Driessche BE	N
Implications of raising the UK Threshold	Rafael Mastrangelo UK	O
Declarant Burden	Mark Kelly UK	P
UK national needs and uses of aggregate trade data	Stuart Brown UK	Q
UK national needs and uses of detailed trade data	Henry Bottomley UK	R
Preferred simplification option	Sandra Tudor UK	S

## **Workshop conclusions as summarised by the Consultant** (full report in annex T)

### **General**

- Perhaps there is a wrong perception of the Intrastat burden - the European Commission will develop methodology for assessing the total administrative and statistical burdens
- Eurostat supports the creation of a task force on simplification

### **Single Flow**

- Issues which would need to be addressed before implementation:
  - Validate the choice of dispatches
  - Harmonise the methodology
  - Reduce asymmetries
  - Identify the set of data to be collected
  - Build adequate threshold system (lower ?)
  - Exchange of confidential aggregate data should be made possible
  - VAT number of the partner company should be collected in all MS
  - Exchange of data at company (national and partner) level should be made possible (for cross-checking with VAT, regional statistics)
  - Estimate the burden for companies declaring under Single Flow
  - Estimate the impact of Single Flow on other statistics (eg BoP, NA)
  - Eurostat should play a central role for data exchange and dissemination (centralised database)
  - Clarify the responsibilities between Eurostat and the MS (e.g. data validation)
  - All MS provide timely data with adequate coverage to avoid large revisions
  - Legal changes
- Expected time plan: Single Flow is not feasible in the short term

### **Threshold Raising**

- Raising of Intrastat thresholds could lead to large reduction of declaration burden
- A significant increase could lead to a larger decrease of the burden than Single Flow, but increases in asymmetries
- Raising the thresholds could cause difficulty in implementing Single Flow at a later stage - any short term option should be consistent with long term options

### **Stratified Random Sampling**

- Not a good option - all large companies would have to be included anyway to ensure sufficient accuracy, there would be an increased risk of non-response and lower quality at detailed level

### **Nomenclature Simplification**

- Decreasing the level of detail (CN8 to HS6) would reduce the number of codes, but not the burden or asymmetries significantly

## **6.5 Work carried out to achieve the results – Objective 2: Workshop**

The following tasks were undertaken in the course of organising the Intrastat Simplification Workshop and Seminar:

- Consultant's contract drawn up
- Consultant recruited
- Invitations sent to delegates
- Possible venues researched
- Agenda constructed (see annex C)
- Venue booked
- Catering arranged
- Recommended hotel list compiled
- Presentation prepared
- Events hosted
- CD-ROMs produced and distributed

## **6.6 Developments planned for the coming years**

The Eurostat Working Group on Intrastat Simplification will take forward the investigation into options: HMRC will participate for the UK in the Quality sub-group. The Working Group is scheduled to report to the ECOFIN council in October 2007.

## **7 Summary of objectives and main results achieved**

### **1. Objective 1: Sampling investigation**

The possibility of using Stratified Random Sampling to reduce the number of traders required to submit Intrastat declarations was explored. Stratifications by commodity and partner country (CC) and by trade value were investigated, along with simple random sampling as a control method.

It had been hypothesised that the CC stratification would produce the best results, preserving the highest degree of coverage at detailed level. This was reasoned because this stratification followed the supposed natural patterns of variation in the trade figures, i.e. according to the type of goods and the partner country. However, the most accurate results were in fact achieved by the stratification by trade value. This produced the most accurate value figures and equivalent coverage at detailed level. This indicates that good coverage of the various commodity-country categories can be achieved without the need to stratify the trader population according to these criteria. Also, as expected the best basis on which to predict a trader's future trade value is his past trade value.

The results produced by the stratification by value, along with the burden reductions possible, indicate that this could be a promising option for simplification of Intrastat. However, the method would also entail considerable added complication in the trader selection procedures carried out by EU National Statistical Institutes, and confusion on the part of traders who are required to cease or begin Intrastat submissions in different years. HMRC SATU has also been carrying out research into the possible effects of raising the Intrastat threshold in order to reduce the number of traders required to submit declarations. This method produces similar results to those obtained here by the value stratification (it is in effect a crude stratification by value) and has the advantage of not introducing the added complications of a stratified sampling system. Therefore SATU do not recommend that any more consideration to the SRS option.

### **2. Objective 2: Intrastat Simplification Workshop**

A three day Intrastat Simplification Workshop and Seminar was held in London in October 2006. It was attended by representatives from 15 Member States, Eurostat and various UK government departments. Research was shared and various options discussed, and it was concluded that the most promising simplification options were raising the threshold and 'single flow'. Raising the threshold would mean reducing the number of traders required to submit declarations by relieving those with the lowest annual intra-EU trade value; this option would be relatively straightforward to implement. However, it would cause an increase in asymmetries as MS with large trade volumes would lose from their figures almost all trade with MS with smaller trade volumes. The 'single flow' option would entail all MS collecting only their outgoing trade (dispatch) data, and reconstructing their incoming trade (arrivals) from other MS' dispatches. This option also promises a large reduction in burden on business, but is not feasible for introduction in the short-term as there are many issues which must first be addressed.

A Eurostat Working Group has been set up to take forward the work on Intrastat simplification. It is due to report to the Council of European Finance Ministers (ECOFIN) in October 2007.