Bakgrundsfakta

Statistiska centralbyrån Statistics Sweden

The Growth Rate Metod for Production of Rapid Estimates of the Swedish Foreign Trade

– Development, Evaluation and Preparations for the Implementation

2005:9

Ekonomisk statistik

The series Background facts presents background material for statistics produced by the Department of Economic Statistics at Statistics Sweden. Product descriptions, methodology reports and various statistic compilations are exampels of background material that give an overview and facilitate the use of statistics.

Publications in the series Background facts on Economic Statistics

2001:1 Offentlig och privat verksamhet – statistik om anordnare av välfärdstjänster 1995, 1997 och 1999

2002:1 Forskar kvinnor mer än män? Resultat från en arbetstidsundersökning riktad till forskande och undervisande personal vid universitet och högskolor år 2000

2002:2 Forskning och utveckling (FoU) i företag med färre än 50 anställda år 2000

2002:3 Företagsenheten i den ekonomiska statistiken

2002:4 Statistik om privatiseringen av välfärdstjänster 1995–2001. En sammanställning från SCB:s statistikkällor

2003:1 Effekter av minskad detaljeringsgrad i varunomenklaturen i Intrastat – från KN8 till KN6

2003:2 Consequences of reduced grade in detail in the nomenclature in Intrastat – from CN8 to CN6

2003:3 SAMU. The system for co-ordination of frame populations and samples from the Business Register at Statistics Sweden

2003:4 Projekt med anknytning till projektet "Statistik om den nya ekonomin". En kartläggning av utvecklingsprojekt och uppdrag

2003:5 Development of Alternative Methods to Produce Early Estimates of the Swedish Foreign Trade Statistics

2003:6 Övergång från SNI 92 till SNI 2002: Underlag för att bedöma effekter av tidsseriebrott

2003:7 Sveriges industriproduktionsindex 1913–2002 – Tidsserieanalys
The Swedish Industrial Production Index 1913–2002 – Time Series Analysis

2003:8 Cross-country comparison of prices for durable consumer goods: Pilot study – washing machines

2003:9 Monthly leading indicators using the leading information in the monthly Business Tendency Survey

2003:10 Privat drift av offentligt finansierade välfärdstjänster. En sammanställning av statistik

2003:11 Säsongrensning av Nationalräkenskaperna – Översikt

2003:12 En tillämpning av TRAMO/SEATS: Den svenska utrikeshandeln 1914–2003

2003:13 A note on improving imputations using time series forecasts

2003:14 Definitions of goods and services in external trade statistics

Continued on inside of the back cover!

These publications and others can be ordered from:
Statistics Sweden, Publication Services, SE 701 89 ÖREBRO, Sweden
phone +46 19 17 68 00 or fax +46 19 17 64 44.

You can also purchase our publications at our Statistics Shop:
Karlavägen 100, Stockholm, Sweden

Background Facts

The Growth Rate Method for Production of Rapid Estimates of the Swedish Foreign Trade

– Development, Evaluation and Preparations for the Implementation

Economic Statistics 2005:9

Statistics Sweden
2005

Economic Statistics

The Growth Rate Method for Production of Rapid Estimates of the Swedish Foreign Trade
– Development, Evaluation and Preparations for the Implementation

Statistics Sweden
2005

Tidigare publicering Publicerad årligen sedan 2001
Previous publication Previous publication has been made since 2001

Producent SCB, Avdelningen för ekonomisk statistik
Producer Box 24 300
104 51 Stockholm

Förfrågningar Anders Jäder, tfn: +46 8 506 947 28
Inquiries anders.jader@scb.se

Omslag: Ateljén, SCB

ISSN 1653-316x Bakgrundsfakta
ISSN 1653-3178 Ekonomisk statistik

Printed in Sweden
SCB-tryck, Örebro 2005:02

Preface

This report concludes the project Growth rate method for production of rapid estimates of the Swedish foreign trade and presents development work, evaluation and preparations for the implementation of the method. The project was carried out during September 2004 – May 2005. The implementation phase of the new method for production of rapid estimates of the total monthly value of the Swedish foreign trade now continues.

The project has been financed by the EU with the funding from the EDICOM 2004 budget.

This report has been produced by Anders Jäder at the unit for Foreign Trade and Anders Holmberg at the unit for Central Methodology.

Statistics Sweden, August 2005

Kaisa Ben Daher
Head of the Unit for Foreign Trade

Contents

Summary

This report deals with the growth rate method developed for the production of rapid estimates of the total monthly value of the Swedish foreign trade. The report starts with a renewed description of the method. Further, it is investigated what data is available in time for the production of rapid estimates. It is found that already 19 days after the reference month a large proportion of the value has already been reported. Estimates of the total trade are then produced and compared to the estimates from the sample based method and the first detailed statistics. It is found that for the second rapid estimate the growth rate method produces better results but for the first rapid estimate the sample based method seems to be somewhat better. It is also found that the growth rate method seems to be less plagued by negative bias than the sample based method. The reasons for the discrepancies between the estimates from the growth rate method and the first detailed statistics are then investigated. Many of the identified reasons were found to have large impact on the estimates. The work to make the implementation of the new method possible is also described along with solutions to some of the practical problems. The report ends with an investigation to find out whether Extrastat data can be used in order to improve the estimates. The investigation was not fruitful.

For the reference months April, May and June 2005 estimates from the growth rate method will be produced using the developed dunning procedures, editing techniques and other procedures used in real production. The production will be performed in real time, i.e. between the 19:th and the 25:th day after the reference month. At the time this report was finalised the results from this test were not ready.

1 Introduction

1.1 Background

In our earlier EDICOM 2 budget for 2002 an action on development of an alternative method for advanced estimates was conducted and compared with the current method. Comparisons suggested that the new method could produce unbiased rapid estimates with fewer resources than the current method and on speed as for the current method. The new method was promising but the accuracy of the estimates from the method was slightly worse than of the current method. We made a decision to wait with the implementation of the new method and to continue with further development, analysis and improvements of the drawbacks of the method.

1.2 Implementation timetable for the operation

The action was started 15 September 2004 and was finished 14 May 2005.

1.3 Objectives

The overall objective of this action is to provide good quality rapid estimates of foreign trade statistics according to requirements of European Union and national users. The estimates should be produced quickly and have high quality. The objective of this project is to further develop the alternative, new method in order to improve the accuracy and to implement the method in the production of rapid estimates as an integrated part of Intrastat system.

List of tasks:

- Carry out comparisons of monthly rapid estimates between the current method, the alternative method and data from the first publishing of detailed results at least for 9 months
- Analyse discrepancies between these figures
- Improve the checking of data in the alternative method
- Investigate the possibilities to use Extrastat data in the alternative method
- Investigate what changes in Intrastat system are needed for the implementation of the alternative method
- Carry out necessary changes in order to be able to integrate rapid estimates in the Intrastat system
- Create new practical routines for the work with advanced estimates e.g. for reminder actions and data checking

1.4 Human resources used

The work was carried out in Statistics Sweden by statisticians Anders Holmberg and Anders Jäder, IT-expert Pekka Koski and economist and future user of the system Martin Fors. No sub-contracting was used.

1.5 Equipment and software

For the tests and evaluations the software SAS has been used. The production system for implementation of the rapid estimates was developed mainly in dbms SYBASE using Visual Basic but the editing and determination of important PSIs was developed in SAS.

1.6 Description of the operation

All the objectives of the projects were met. We have:

- Carried out comparisons of monthly rapid estimates between the current method, the alternative method and data from the first publishing of detailed results for 29 months
- Analysed the discrepancies between these figures theoretically and empirically
- Improved as much as possible the checking in the alternative method
- Investigated the possibilities to use Extrastat data for estimates
- Investigated what changes in the Intrastat system were needed for the implementation of the alternative method.
- Carried out necessary changes in order to be able to integrate rapid estimates in the Intrastat system
- Created new practical routines for the work with advanced estimates e.g. for reminder actions and data checking

This operation consisted of the continuation of the creation of a time series data base containing Intrastat reports extracted from the ordinary Intrastat production environment at a specific time during the month. Estimates based on the extracted data were made and these were compared to the total value published in the first detailed statistics. Reasons for discrepancies between the two were identified theoretically and investigated empirically. When the decision was taken to implement the new method the work begun to investigate what changes were necessary in order to be able to implement the new method. This was done through collaboration between the operating personnel, IT-staff and statisticians. Methods and routines for editing and reminder actions were developed. A part of the action has also been devoted to the investigation whether Extrastat data can be utilized to improve the method.

The results of the evaluations were the conclusion that the growth rate method seems to produces better estimates that the sample based method for the second rapid estimate. However, for the first rapid estimate 25 days after the reference month the sample based method may produce better results. It was also found that for all estimates the growth rate method is less affected by negative bias than the sample based method. The work to develop the computer system and the SAS applications necessary to produce rapid estimates according to the alternative method is more or less finalised, but measures to improve the system might be needed.

Currently we are performing a more thorough test were the new method is run in parallel with the sample based method. The test is performed to test if the computer system and the practical routines are working well. The test will also answer the question whether the new method performs well

in a real production environment. Results from the test, which will be performed for the reference months April 2005 to June 2005, are not ready yet.

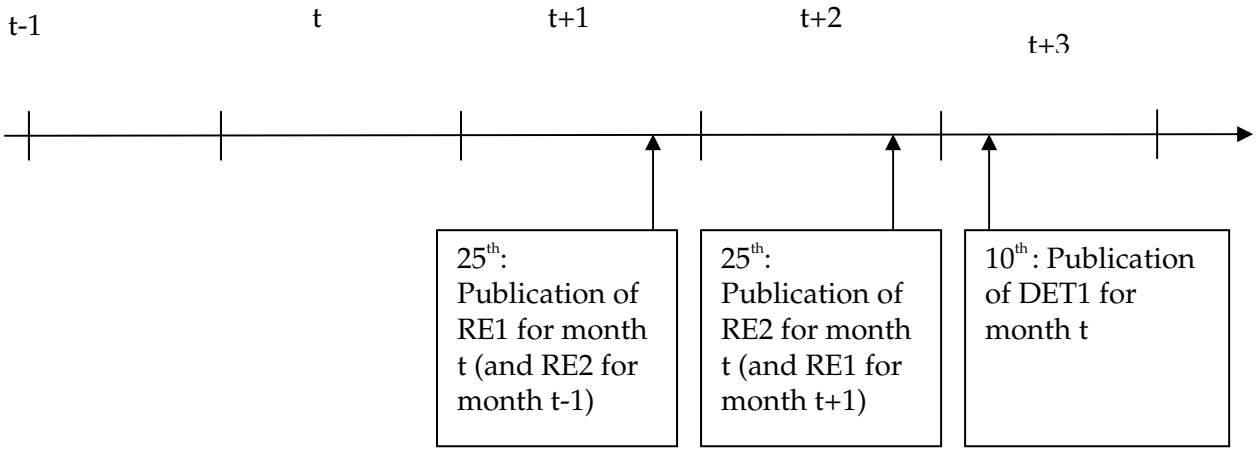
The report starts in the next chapter with a revision of the theory behind the proposed estimators. In chapter 3, the quality of the data in our time series data base is evaluated and an evaluation of the performance of the estimators is made. In this chapter the reasons for discrepancies between the estimates and the first detailed statistics is also identified theoretically and tested empirically. In chapter 4 the work performed in order to implement the new method is described. Chapter 5 is devoted to an investigation to find out whether Extrastat data can be used to improve the estimates.

2 A revision of the estimators

During the testing and evaluation of the growth rate method we have found some shortcomings of the estimators proposed in the previous EDICOM report. This is why a new description of the estimation procedure is given in this section.

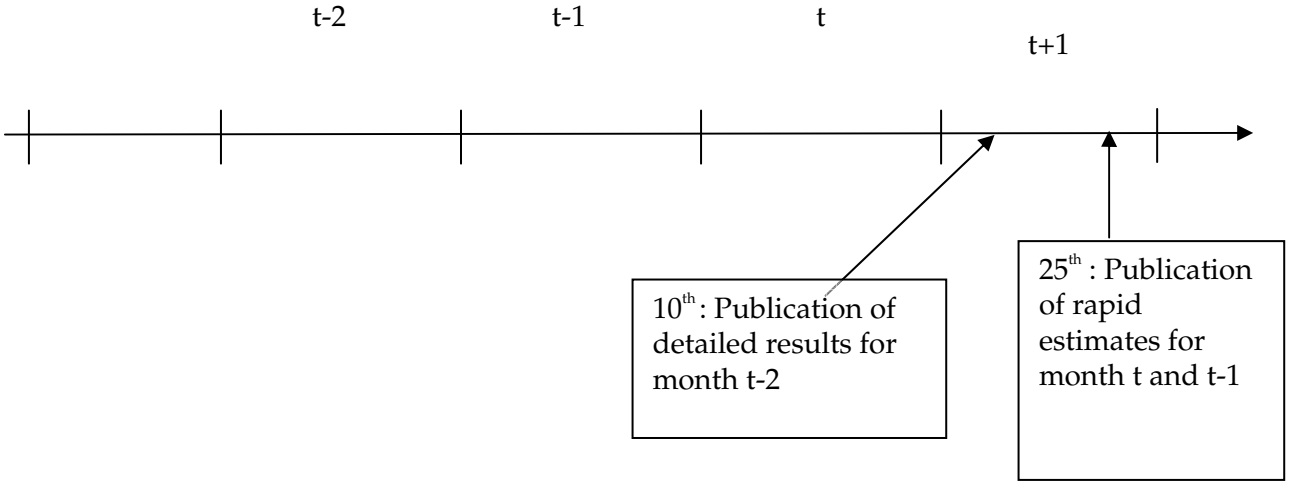
At the Foreign trade unit at Statistics Sweden the first results are published approximately 25 days after the reference month (denoted t). These results are called rapid estimates and include only the total value of trade broken down by EU and non-EU trade. Two rapid estimates, henceforth abbreviated RE1 and RE2, are produced before the first detailed results are published, henceforth abbreviated DET1 or only DET. The detailed results are published about 70 days after the reference month and consist of a complete breakdown on countries and commodities. The timetable for production of foreign trade statistics at Statistics Sweden is illustrated in Figure 1 below. The introduction of the growth rate method instead of the sample survey method will not change the timetable of publication.

Figure 1
Timetable for the production of foreign trade statistics for a reference month at Statistics Sweden



The implication of this timetable is that when the first rapid estimate for a reference month t is to be produced the detailed results for month t-2 are already published. This is illustrated in figure 2 below. This means that the total value from the detailed statistics can be used in the estimation of aggregated results for month t and t-1.

Figure 2
Illustration of the fact that the detailed statistics for month t-2 are published before the publication of rapid estimates for month t and t-1



The basic idea behind our new proposed method is to use the published total value of the EU-trade for month t-2 as a base for estimates. To produce an estimate of the total trade of month t-1 this base is multiplied by an estimate of the growth of the trade from month t-2 to month t-1 and to produce an estimate of the total trade of month t the base is multiplied by an estimate of the growth of the trade from month t-2 to month t.

The following sets can be defined:

D_t^t: The set of companies which have reported data for month 't' at the time of publication of the first rapid estimates for month t, i.e. by the 25th of month t+1.

D_{t-1}^t: The set of companies which have reported data for month 't-1' at the time of publication of the first rapid estimates for month t, i.e. by the 25th of month t+1.

D_{t-2}^t: The set of companies which have reported data for month 't-2' at the time of publication of the detailed results for month t-2, i.e. by the 10th of month t+1.[1]

Let $U_t = D_t^t \cup \overline{D}_t^t$ denote the population of PSIs at month t. Our objective is to estimate the total trade of the population for the two months t and t-1. Hence, our main parameters are:

$$T_{t-1} = \sum_{i \in U_{t-1}} y_{i,t-1} \qquad \text{Total trade for reference month t-1} \qquad (1)$$

$$T_t = \sum_{i \in U_t} y_{i,t} \qquad \text{Total trade for reference month t} \qquad (2)$$

[1] This definition is slightly different from the definition in the previous EDICOM report.

where $y_{i,t}$ and $y_{i,t-1}$ are the arrival or dispatch values of the i:th PSI for month t and t-1 respectively.

The values $y_{i,t}$ and $y_{i,t-1}$ could be considered the "true" trade values of the PSI. However, in practice the individual PSIs trade values may change over time depending on when they are measured. It may be reasonable to use the first detailed estimate as the target parameter. Another approach could be to have the final detail figure, i.e. the figure published after the final revision of a month, as the target parameter. We have chosen the first approach.

Estimation procedures

When estimating the growth rate of trade from one period to another, the ratio of the sum of the trade of the last month and the sum of the trade of the first month is calculated. It is important to calculate the sums over the same set of companies for the two periods. Otherwise the growth rates can be affected by the fact that, in some months most reports are sent in early and, in other months, most reports are delayed e.g. due to holidays. Thus, the following sets of companies are also important:

$$A^t = D^t_{t-1} \bigcap D^t_{t-2}, \quad B^t = D^t_t \bigcap D^t_{t-2} \tag{3}$$

The first set, A^t, is used to estimate the growth rate from t-2 to t-1 and the second set, B^t, is used to estimate the growth rate from t-2 to t.[2] The estimators of the respective growth rates are:

$$\hat{g}_1 = \frac{\sum_{i \in A^t} y^t_{i,t-1}}{\sum_{i \in A^t} y^t_{i,t-2}} \qquad \hat{g}_2 = \frac{\sum_{i \in B^t} y^t_{i,t}}{\sum_{i \in B^t} y^t_{i,t-2}} \tag{4}$$

Two relatively simple estimators are used to estimate the totals of equations (1) and (2). To estimate (1), i.e. to produce an RE2-estimate, we use the following expression:

$$E1 = \hat{T}^t_{t-1} = DET_{t-2} \cdot \frac{\sum_{i \in A^t} y^t_{i,t-1}}{\sum_{i \in A^t} y^t_{i,t-2}} = DET_{t-2} \cdot \hat{g}_1 \tag{5}$$

This estimator uses the estimate of the growth rate $\hat{g}_1$ and the first detailed estimate for month t-2.[3]

[2] In our previous report we also considered a third set C^t which was used to estimate the growth rate from t-1 to t. Since the use of this set did not result in distinctly better estimates we have decided to disregard this estimator for the moment. Further studies of this estimator could be done in the future.

[3] This estimator is the same as the non-response adjusted estimator in the previous report (see equation 13).

To estimate (2) we have used the estimator E2, an estimator similar to E1, using $\hat{g}_2$ instead of $\hat{g}_1$:

$$E2 = \hat{T}_t^t = DET_{t-2} \cdot \frac{\sum_{i \in B^t} y_{i,t}^t}{\sum_{i \in B^t} y_{i,t-2}^t} = DET_{t-2} \cdot \hat{g}_2 \qquad (6)$$

The separate estimators used in the previous report are not studied in this report. Neither is the estimation for countries and commodities studied. Further studies of these estimators can be done in the future.

3 Further evaluation of the method

In the report under the EDICOM 2 budget for 2002 estimates for the sample survey method as well as for the new growth rate method were presented and comparisons were made with the first published detailed statistics. The estimates presented covered the period September 2002 to March 2003. Based on these estimates and comparisons we did not feel confident enough to implement the method in production. Further evaluations and tests were deemed necessary. These will be presented in this report.

First we investigate what kind of data we have at our disposal to make the estimates. Then estimates are calculated and finally the reasons for possible under and over estimations are investigated.

3.1 Data quality

To be able to produce good estimates as early as 25 days after the reference month we must have access to sufficient amount of data. This means that companies representing a large part of the total value of the trade must have sent in their reports.

As explained in the previous report we constructed a time series data base with Intrastat data. This data was used to evaluate the amount of data received and the quality thereof. The data was also used to make estimates of the total trade and these were compared to the estimates from the sample survey method and from the first detailed statistics. We have continued to enlarge this data base by extracting reports from the Intrastat data base every month. In the first report our estimates was based on data extracted from the Intrastat data base on the 25:th of each month. Even if this was sufficient as a first evaluation of the method, it was however somewhat unrealistic. If the estimates are to be produced on the 25:th of the month we also have to allow for a couple of days for editing, reminder actions and publication work. That is why we, for the reference month October 2003 and onward, decided to extract data already after 19 days.

The decision to extract data already the 19:th day after the reference month made the evaluation more realistic. In real production there would also be a dunning procedure in which the most important PSIs would be reminded if they hadn't reported by the 19:th day. The most important PSIs are reminded in the sample based method which is used in real production. In order to mimic an imagined real production we complemented our extracted data with the largest missing PSIs reminded in the sample based method. This procedure was also started for the reference month October 2003.

The estimates are based on sets of reporting units. In our first report these sets of reporting units were formed by subsidiary reporting numbers. A subsidiary reporting number can e.g. be one of several plants belonging to a company. Starting for the reference month October 2003 we decided that the estimates should be calculated on sets of VAT-numbers instead of on sets of subsidiary reporting numbers. In that way changes in the way an enterprise is divided into reporting numbers don't affect the estimates. On

the other hand it becomes more important that all of the subsidiary reporting numbers do come in with their reports. Otherwise the trade of the enterprise is underestimated.

Let us first look at the number of PSIs in sets A and B for the different months. These figures must be set in relation to the number of obliged PSIs. The number of obliged PSIs decreased at the start of the year 2005. This can be seen in Table 1. The decrease was due to the increase in threshold values which took effect at the turn of the year. Before May 2004 the number of obliged PSIs was somewhat lower than what is shown in Table 1 since by then the ten new member states hadn't entered the European Union.

Table 1
The number of obliged PSIs before and after the increases in thresholds

	2004	2005
Arrivals	13 000	11 000
Dispatches	7 400	4 900

In Appendix 1 figures describing the size of the sets A and B can be seen. The sets A naturally contain a larger proportion of the PSIs then the sets B. For both arrivals and dispatches the sets A includes about 70-80 per cent of the PSIs. For arrivals the sets B contain from 30 per cent to 60 per cent of the PSIs. The proportion for dispatches is slightly larger. The decrease in the number of PSIs in the beginning of 2005 had a larger proportional effect on dispatches than on arrivals and the effect seems to be larger on the sets A than on the sets B. In fact the sets B for arrivals seem to have increased in size.

The decision to extract data already 19 days after the reference month seems to have had no effect on the sizes of the A sets but the B sets have decreased in size. The change in definition of reporting units from subsidiary reporting numbers to whole companies should not affect the sets A and B by more than a maximum of 200 units for arrivals and a maximum of 100 units for dispatches.

In Appendix 2 the proportion of the total trade value that is included in the sets A and B for the different months can be seen. For arrivals the sets A normally include around 90 per cent of the value and for dispatches the proportion is even higher. This gives good possibilities for high precision in the estimates of the total trade for month t-1.

For arrivals the sets B normally includes 60-70 per cent of the total value of the trade. For dispatches the proportion is a bit higher, 70-80 per cent. This means that the conditions for high precision in the estimates are less favourable for the reference month t than for the reference month t-1. Furthermore, the conditions for good estimates are better for dispatches than for arrivals. This is true for both reference month t and reference month t-1.

The decision to extract data already 19 days after the reference month seems to no have impact neither on the proportion of the value included in sets A, nor on the proportion of the value included in the sets B. This implies that no large companies are lost by extracting data already after 19

days. This might be due to the fact that we complement the database with the largest companies from the sample survey method.

In the next section estimation results based on this data are presented.

3.2 Estimation results

In appendix 3 graphs are presented illustrating the discrepancies between the first detailed statistics and the respective estimates of the sample based method and the growth rate method.[4] The discrepancies are measured in hundred million SEKs. As a comparison it can be noted that on the average the value of the Swedish EU-trade is approximately 40 000 million SEK for arrivals as well as for dispatches.[5]

In table 2 can be seen the number of months were the growth rate method produced at least as good estimates as the sample based method. It can be seen that the estimates of the growth rate method are at least as good more than 50 per cent of the time for the estimates of t-1. For the estimates of reference month t however, the estimates of the growth rate method are inferior more than 50 per cent of the time.

Table 2
Performance of the growth rate method compared to the sample based method

	Arrivals		Dispatches	
	Number of estimates were the growth rate method are at least as good	Total number of month estimated	Number of estimates were the growth rate method are at least as good	Total number of month estimated
Estimates for t-1	17	29	16	29
Estimates for t	11	29	13	29

The mean discrepancy is shown in table 3 for the sample based method as well as for the growth rate method. All the mean discrepancies are negative. That indicates that the rapid estimators seem to underestimate the total trade. However, in all cases the growth rate method has smaller mean discrepancy than the sample based method. This indicates that the negative bias is smaller for the growth rate method.

[4] The discrepancies for the sample based method for the RE1 for May 2004 and the RE2 for April 2004 is not shown due to issues concerning the EU-enlargement.

[5] 1 SEK ≈ 0.11 EUR

Table 3
Mean discrepancies for the sample based method and the growth rate method

	Arrivals		Dispatches	
	Sample based method	Growth rate method	Sample based method	Growth rate method
Estimates for t-1	-0.38	-0.02	-0.24	-0.09
Estimates for t	-0.67	-0.58	-0.48	-0.24

Table 4 shows the standard error of the discrepancies. The growth rate method has a smaller variation in the discrepancies when it comes to the estimation of reference moth t-1. However, for the estimation of reference month t the growth rate method has larger variation in the discrepancies.

Table 4
Standard deviation for the discrepancies for the sample based method and the growth rate method

	Arrivals		Dispatches	
	Sample based method	Growth rate method	Sample based method	Growth rate method
Estimates for t-1	0.78	0.51	0.60	0.48
Estimates for t	0.72	0.95	0.64	0.77

To summarise, for the estimation of reference month t-1 the growth rate method seems to be superior. In more than half of the months it gives equally good or better estimates than the sample based method. The bias of the estimator and the variation in the errors seem to be smaller.

For the estimation of reference month t the sample based method might be more favourable. In more than half of the times it produces better estimates than the growth rate method. The variation in the errors is also smaller but on the other hand the negative bias seems to be larger for the sample based method than for the growth rate method.

For the reference months April, May and June 2005 the growth rate method will be tested in a real production environment. Estimates from the growth rate method will be produced using the developed dunning procedures, editing techniques and other procedures used in real production. The production will be performed in real time, i.e. between the 19:th and the 25:th day after the reference month. At the time this report was finalised the results from this test was not ready. Our hope however, is that the estimates from the test will be somewhat better than our earlier estimates using the growth rate method.

In the next section the possible reasons for the discrepancies between the rapid estimates and the detailed statistics will be examined.

3.3 Reasons for discrepancies

There can be several reasons why the estimates differ from the target, i.e. from the first detailed statistics. We have identified 5 major reasons:

1) The reported data is not representative of the whole trade
2) Measurement errors
3) The compensation for non-response and below threshold trade in the detailed statistics
4) Presence of vessels and aircraft
5) Intra-SAD transactions

They will all be discussed in this section.

3.3.1 The reported data is not representative of the whole trade

The PSIs that send in their data early might be different from the rest of the total population of PSIs. PSIs having had little trade one month may require little time for the completion of their report. Because of that the may be the ones sending in their reports early. PSIs having had unusually large trade one month may on the other hand require more time to complete their report and their report will be sent in later. This would cause an underestimation of the trade. The growth rate method has some problem with negative bias for the estimation of reference month t. It is hard to see how to avoid this in the growth rate method.

Apart from selection bias, the fact that not all PSIs have reported their data at the time of publication will also cause variance. This variance can be estimated. In appendix 4 it is showed that our growth rate estimator can be seen as a calibration estimator. It is also argued that since all PSIs above the cut-off threshold are obliged to report, the sample design used to estimate the growth rates can be regarded as a census. At least by the time rapid estimates are produced there will however be some non-response. Thus, since the sampling variance is zero, the whole variance of our estimator can be attributed to variance from non-response. In appendix 4 it is showed that the variance of our estimator can then be estimated by the following expression:

$$\hat{V}(\hat{T}_t) = \sum_{i \in r} d_i^2 v_i (v_i - 1) e_i^2 \qquad (7)$$

where, $d_i = 1$, $v_i = DET_{t-2} \Big/ \sum_B y_{i,t-2}$ and

$$e_i = y_{it} - \frac{\sum_B y_{it}}{\sum_B y_{i,t-2}} y_{i,t-2} = y_{it} - \hat{g}_2 y_{i,t-2}$$

This variance estimator is theoretically justified and it is easy to compute. Most of its components are derived in other parts of the production process, e.g. in principle the v_i is the inverse of the proportion of the value included in set B seen in tables 2 and 4 of Appendix 2.

Calculations from our database (on the months January 2005 and February 2005) suggest that the square root of the above variance estimator (the estimated standard error of $\hat{T}_t$) varies between 674 – 717 MSEK for arrivals and 671 – 751 MSEK for dispatches. These numbers are in the range of the

margin of error we expect of the growth rate method and they indicate that the variance estimator appear to work reasonably well.

3.3.2 Measurement errors

First we investigate how measurement errors affect the estimates. For this purpose we need some notation.

The estimation formula for the estimation of reference month t can be written as below:

$$\hat{T}_t^* = DET_{t-2} \cdot \frac{\sum_B y_t^*}{\sum_B y_{t-2}} \tag{8}$$

where $\hat{T}_t^*$ is the estimate when the measurements are affected by measurement errors. The values for t-2, y_{t-2}, are edited since they have been published as a part of the detailed statistics. The values for reference month t, on the other hand, are affected by measurement errors. In the formula above y_t^* is the reported value including measurement errors. Examples of reasons for measurement errors are e.g. the following:

- One or more subsidiary reporting numbers are missing
- The value of an individual observation has been incorrectly recorded
- Mistakes have been made by the PSI during the completion of the questionnaire (electronic questionnaire or paper questionnaire.)

Let's asume that y_t^* can be divided into two parts; the correct value y_t, and a measurement error ε_t. I.e.:

$$y_t^* = y_t + \varepsilon_t \tag{9}$$

The estimation formula then becomes:

$$\hat{T}_t^* = DET_{t-2} \cdot \frac{\sum_B (y_t + \varepsilon_t)}{\sum_B y_{t-2}} = DET_{t-2} \cdot \frac{\sum_B y_t + \sum_B \varepsilon_t}{\sum_B y_{t-2}} =$$

$$DET_{t-2} \cdot \frac{\sum_B y_t}{\sum_B y_{t-2}} + DET_{t-2} \cdot \frac{\sum_B \varepsilon_t}{\sum_B y_{t-2}} = \hat{T}_t + \frac{DET_{t-2}}{\sum_B y_{t-2}} \cdot \sum_B \varepsilon_t \tag{10}$$

where $\hat{T}_t$ is the estimate that would result if no measurement errors where present. Thus, the effect of the measurement errors on the estimate can be divided into to parts: One factor determined by the size of DET_{t-2} in relation to $\sum_B y_{t-2}$ and one factor consisting of the sum of all measurement errors in the set B. The first factor is determined mainly by the size of the set B measured in value. This factor will decrease if the set B is increased. The set B is increased if more reports are sent in for month t. Thus an efficient dunning procedure should be set in place.

In addition to an efficient dunning procedure one should also make sure that the sum of the measurement errors, $\sum_B \varepsilon_t$, is as small as possible. It is obvious that negative and positive measurement errors will cancel each

other out. In addition, it can be concluded that it is the size of the absolute error measured in value that determines the impact of the measurement error on the estimate, not the measurement error as a fraction of the value t-2 or anything similar.

How large the sum of the measurement errors has been historically can be evaluated from our time series data base. As previously stated, for arrivals the sum $\sum_B y_{t-2}$ is normally 60-70 per cent of DET_{t-2}. During the period September 2002 to June 2004 the fraction varied between 57 per cent and 74 per cent with a mean of 67 per cent. That would entail a factor between 1.75 and 1.35 with a mean of 1.49. For dispatches the fraction varied between 66 per cent and 86 per cent with a mean of 77 per cent. This leads to a factor between 1.52 and 1.16 with a mean of 1.30. In the figures below can be seen the measurement error part of (10) depending on the number of incorrect total values by PSI that is corrected. The graphs are constructed under the assumption that the errors are corrected in order of size. The maximal factors have been used in order to get an upper bound for the effect of the measurement errors.

Figure 3a
The effect of the measurement errors on the estimate depending on the number of corrected errors. Values in million SEK. Arrival data for August 2004

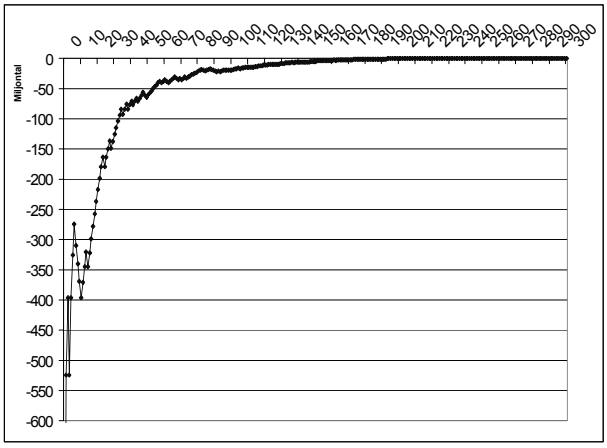
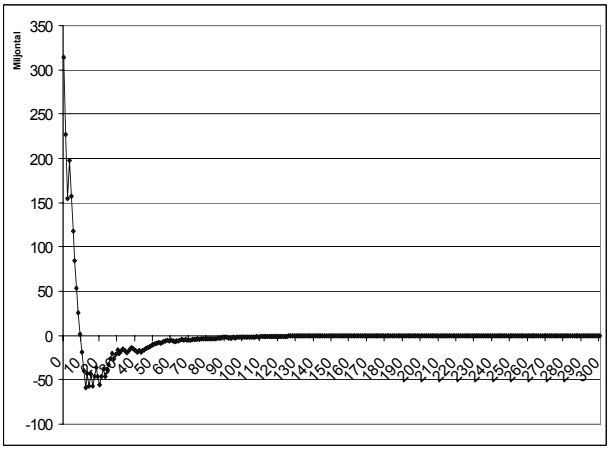


Figure 3b
The effect of the measurement errors on the estimate depending on the number of corrected errors. Values in million SEK. Dispatch data for August 2004



The published total trade values are given in thousands of millions of SEK with a precision of one decimal. This means that if the measurement errors are to have no noticeable effect on the estimates the total measurement error must be smaller than 50 million SEK. This is accomplished after about 50 corrections for arrivals and after about 20 corrections for dispatches.

In figure 3 the individual measurement errors, ε_t, are recorded by subsidiary reporting number. This implies that the effect of a missing report from one out of several subsidiary reporting numbers cannot be seen here.

To investigate whether August 2004 was a special month data for October 2004 are presented below.

Figure 4a
The effect of the measurement errors on the estimate depending on the number of corrected errors. Values in million SEK. Arrival data for October 2004

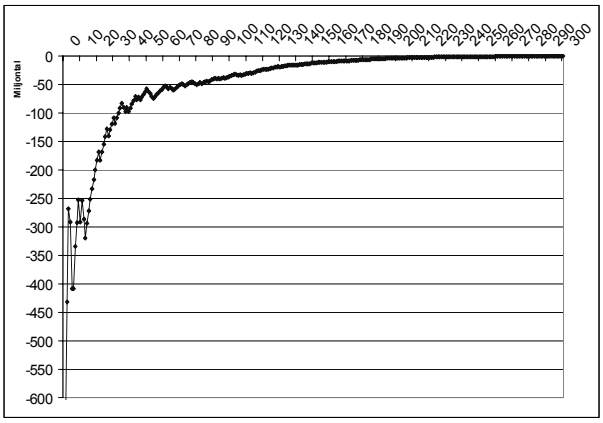
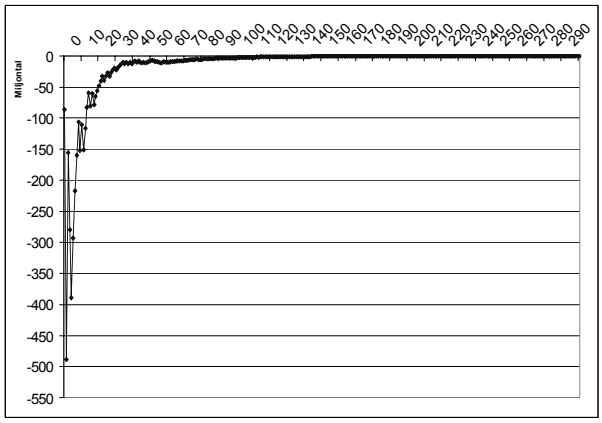


Figure 4b
The effect of the measurement errors on the estimate depending on the number of corrected errors. Values in million SEK. Dispatch data for October 2004



For the reference month October about 70 corrections are required for arrivals and 20 corrections for dispatches. This means that approximately the same number of corrections are required for both August and October.

As can be seen from the graphs the measurement errors $\varepsilon_t = y_t^* - y_t$ are most often negative. This means that for most of the PSIs the unedited value is smaller than the edited value.[6] This is true even though the evaluation is performed on the level of the subsidiary reporting number. The increases in invoiced values are thus not a result of extra subsidiary reporting numbers coming in with their report. Instead the individual subsidiary reporting number has made a supplement to their report or has written a too small invoiced value on their report. The individual terms in the sum $\sum_B \varepsilon_t$ don't seem to cancel each other out. Most of them have the same sign.

The smallest correction that needs to be found in order for the total measurement error to be below 50 million SEK is very small. In the graphs below can be seen the magnitude of the smallest correction needed. We have demonstrated that in August 70 corrections were necessary for arrivals and 20 corrections were necessary for dispatches. By locating these respective numbers on the horizontal axes it can be seen in the graphs that for arrivals corrections as small as 1 million SEK are required in order for the total measurement error to be below 50 million SEK. For dispatches all errors larger than 6.2 million SEK needs to be corrected. The results for October are similar.

[6] Previous experiences from the unit price checking has showed that when it comes to corrections for individual observations (lines) the editing most often makes the invoiced values higher. On the other hand the corrections that make the invoiced value smaller are often very large corrections.

Figure 5a
The sizes of the individual measurement errors. Arrival data in millions of SEK for August

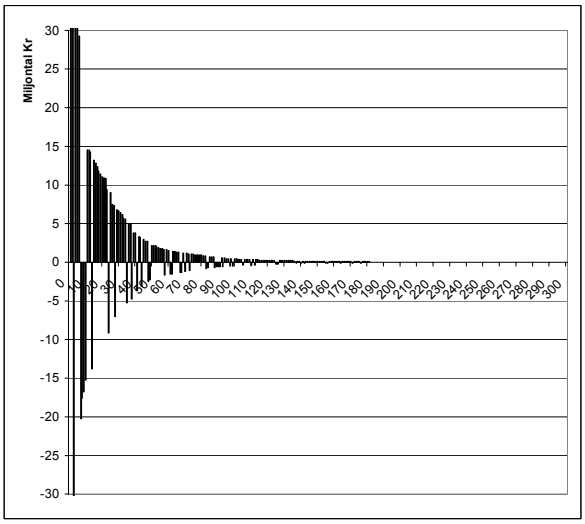
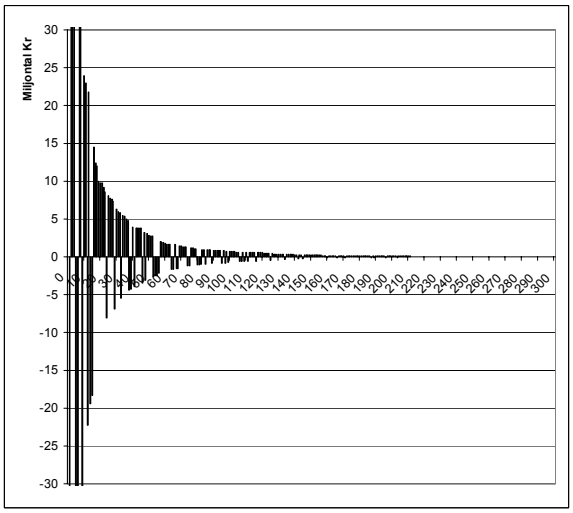


Figure 5b
The sizes of the individual measurement errors. Dispatch data In millions of SEK for August



From figure 3 and figure 4 it is obvious that there exist many small measurement errors that together give a large measurement error. Instead of chasing all these small measurement errors it might be better to have a general compensation for them. The estimation formula would then be as follows:

$$\hat{T}_t^{**} = DET_{t-2} \cdot \frac{\sum_B y_t^*}{\sum_B y_{t-2}} + K \qquad (11)$$

where K can be e.g. 100 million SEK for arrivals and e.g. 50 million SEK for dispatches. Whether this compensation is beneficial must be investigated further before it can be implemented.

Now that we have investigated what kind of errors should be corrected in order to get good results we turn the attention to the question whether we found these errors in our evaluation. When we produced estimates for the evaluation of the growth rate method we edited the data according to a method that will be described later in this report. The question is whether we found the important errors by using this method.

The question might seem easy to respond to. The problem is that the Intra-stat database is constantly changing. Even after the publication of the first detailed results a lot of changes occur. A report might be transferred to another PSI due to a merger or an organizational change. This actually means that some PSIs in the set B will no longer be included in the set. Another problem is that supplementary reports will be sent in after the publication of the first detailed results. It may be difficult afterwards to know the reason for the changed total value and it may even be difficult to know when the change occurred. In our time series data base we have not included the data by PSI that is used for the detailed publication. If this data were included in the time series data base the evaluation would be facilitated.

After all it can be found that for the first rapid estimate for arrivals for August 2004 the editing missed a 36 million SEK error occurring since one of the subsidiary reporting numbers was missing for a PSI. Furthermore a 25 million error was missed since one PSI that had reported 92 million had their report changed to 67 million. The reason for this change is now unclear. Several errors of almost the same magnitude exist. Most remaining errors seem to be negative, i.e. the edited value is lager than the unedited. For dispatches for august 2004 e.g. a nil-report was changed to 36 million SEK. A couple of almost as high errors exist and a lot of small negative errors. It is thus clear that measurement errors will have a fully noticeable effect on the estimates.

When looking at the measurement errors for October it can be found that one PSI supplemented its report by 94 million SEK during the time between the first rapid estimate and the first detailed statistics. The value before the supplementary report was sent in was 560 million SEK and the value after was 654 million SEK. This change was probably difficult to anticipate. According to what has been said previously it can be estimated that this single error would affect the estimate by approximately 140 million SEK. A couple of errors of around 20 million SEK was not corrected either. For dispatches a couple of large errors exist as well but the reason for them is unclear.

To conclude, measurement errors can have a significant effect on the estimates. But due to the constantly changing Intrastat database the exact effect of the errors can be difficult to assess however. If the time series data base were expanded to include the data used at the exact moment of publication the possibilities for evaluation would be improved.

3.3.3 The compensation for non-response and below threshold trade in the detailed statistics

As previously mentioned the total sum of trade from the detailed statistics for reference month t-2 is used as a basis for the rapid estimates. This sum is not simply the sum of received reports. A couple of percent of the total trade value consists of estimated trade values due to non responding enterprises and trade of the enterprises below the threshold. If the growth rate of the trade of these enterprises differ from the growth rate of the rest of the enterprises the result will be biased estimates. Another reason for bias might be that the trade of these enterprises is incorrectly estimated in the detailed statistics. These issues will not be investigated further in this report.

3.3.4 Presence of vessels and aircraft

Now and then a large vessel or a large aircraft is recorded in the statistics. These are often of a value of a couple of hundred million SEK. Thus, these transactions have a considerable impact on the total value of trade. These transactions are collected by a special survey and they are treated somewhat differently from the other Intrastat transactions. Even if this is not always the case, these transactions are typically reported by PSIs that normally has no other trade. That means that these PSIs are seldom included in the set A or set B. Thus the vessels and aircraft will have no influence on the growth rate. Sometimes the vessels and aircraft are recorded in time for

the publication of the first detailed statistics. That means that they are included in the base which is then multiplied by the growth rate. This is however not satisfactory since there may be no aircraft or vessels for the reference months t and t-1.

The preferred estimation formula for reference month t is the following:

$$\hat{T}_t = \left(D_{t-2} - VA_{t-2}\right) \cdot \frac{\sum_B y_t}{\sum_B y_{t-2}} + VA_t \qquad (12)$$

where VA denotes the value of vessels and aircraft. The trade values y_t and y_{t-2} should then be measured without the vessels and aircraft. The corresponding formula for reference month t-1 is:

$$\hat{T}_{t-1} = \left(D_{t-2} - VA_{t-2}\right) \cdot \frac{\sum_B y_{t-1}}{\sum_B y_{t-2}} + VA_{t-1} \qquad (13)$$

If VA_t and VA_{t-1} is not known by the time of publication it might be possible to make an informed estimate of the value.

The number of vessel and aircraft transactions over 30 million SEK was three for the year 2002, 21 for 2003, ten for 2004 and during 2005 five transactions has been reported so far. For the reference months September 2004 to February 2005 the above formula was used for the estimation. For this time period there were 2 vessels or aircraft for arrivals and 1 for dispatches. For most months the estimates get better by the use of the above formula but for some months the estimates actually get worse. Theoretically though, the formula should produce better results. The deterioration of some estimates is probably due to other problems which should be dealt with by use of other measures. It can be concluded after all that the aircraft and vessels have a large impact on the estimates.

3.3.5 Intra-SAD

Some transactions concerning the EU-trade are not reported to Statistics Sweden but to the Swedish Customs. These transactions are then added to the ordinary Intrastat transactions at the time of publication of the detailed statistics. The transactions are also added to the estimates from the sample survey method. They are thus not included in the trade values of the individual PSIs used in the estimation, i.e. y_t, y_{t-1} and y_{t-2} are all measured without the Intra-SAD transactions. That means that the transactions do not affect the estimates of the growth rates.

We think the Intra-SAD transactions are reported very soon after the reference month, which at least is the case for all the other transactions reported to the Customs, and that these values are not revised to a large extent. Therefore there is no need to estimate this part of the trade and it can be considered a known constant. Thus we can use the following estimation formulae:

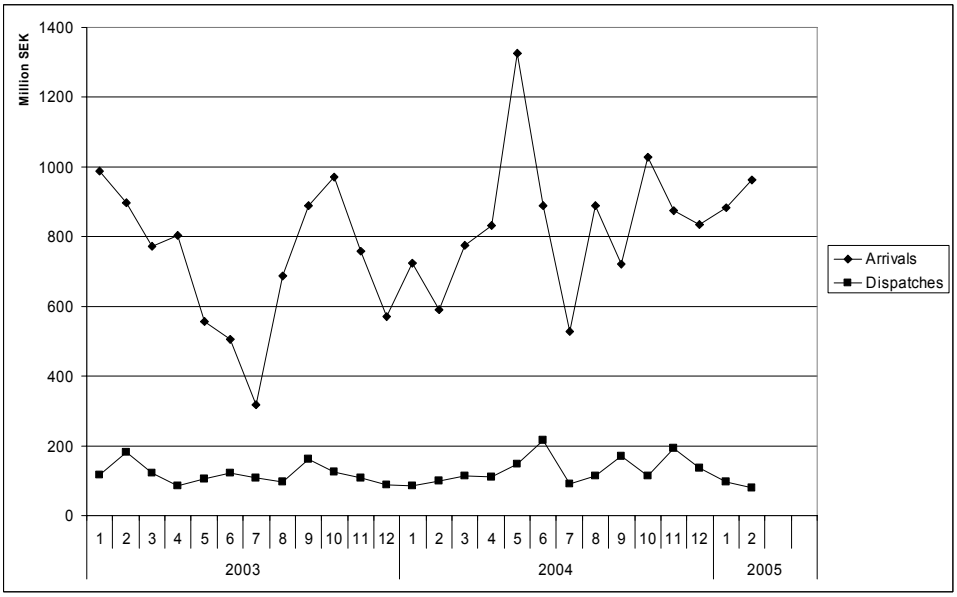
$$\hat{T}_t = \left(D_{t-2} - SAD_{t-2}\right) \cdot \frac{\sum_B y_t}{\sum_B y_{t-2}} + SAD_t \tag{14}$$

$$\hat{T}_{t-1} = \left(D_{t-2} - SAD_{t-2}\right) \cdot \frac{\sum_B y_{t-1}}{\sum_B y_{t-2}} + SAD_{t-1} \tag{15}$$

where SAD represents the sum of the value of the Intra-SAD transactions for the respective months.

In the graph below can be seen the value of the Intra-SAD transactions over time.[7] The dispatches seem to be quite stable over time while the arrivals seem to vary a lot. That indicates that the estimates could be improved for arrivals by the use of the formulae above. In our estimates for September 2004 to February 2005 we have used the above formulae. We have found that for arrivals, the use of the formulae has an impact on the estimates but for dispatches the impact is negligible.

Figure 6
The value of the Intra-SAD transactions over time



[7] Information about Intra-SAD transactions for the year 2002 is not available.

4 The implementation

In September 2004, based on the results of the estimates until June 2004, it was decided to implement the growth rate method in production. This chapter will describe the work that has been done to make that possible.

4.1 The data model

Several demands were put forward on the design of the new data system. These were made by the statistician, the computer specialist and the personnel that will be the users of the system. The following are examples of demands on the new system. It should be…

1) Integrated into the ordinary Intrastat computer system
2) Easy to work with
3) Designed so that evaluations and continuous quality improvement is possible
4) Designed for the production of best possible estimates
5) Designed not to be overly complicated to implement

It was realized after some time that some of these demands were somewhat in conflict with each other.

It was realized that Intrastat reports by PSI had to be extracted from the ordinary Intrastat production tables to separate tables used only for the production of rapid estimates. To allow for editing and other steps of production the extraction of reports had to be done a couple of days before publication, normally around the 19th day after the reference months. However it was found that from the 19th day until the day of publication (the 25th day) large changes occurred in the ordinary Intrastat production system. The values of the reports already sent in changed due to supplements and editing. Our aim was at first to in some way take these changes into account in order to get better estimates. This was supposed to be done by using a second extraction of reports after about 23 days. This turned out to be very complicated logically and computer technically since changes would also have been made in the already extracted data. In that sense point 4 above was in conflict with point 5 and probably also point 2. It was decided that the second extraction of data around the 23rd day after the reference month would only be used to complement with PSIs that had not yet sent in their reports.

The demand that the system should be designed so that evaluations and continuous quality improvement is possible also complicated the implementation somewhat and some compromises had to be made. However, the following attributes have been implemented to the new system: (i) It is possible to follow up what reports where sent in by the 19th day as well as the values of those reports. (ii) It is possible to see what values were changed or added during the production as well as the reason for the changes and additions, e.g. changes due to editing or additions due to the dunning procedure. (iii) It is possible to see what values were added during the second extraction (by day 23). Thus the possibilities for ex post

evaluations are quite good. It requires however some extra effort by the operating personnel so it is somewhat in conflict with point 2 above.

The advantage of the integration of the system for rapid estimates into the ordinary Intrastat system is that the same register of PSIs can be used in the two sub-systems. If there are any disadvantages to this - e.g. that large changes in the register might be made between the 19:th and the 25:th day after the reference month - is not clear.

4.2 The important PSIs and the dunning procedure

In the sample survey method for production of rapid estimates it is obvious which PSIs need to be part of the dunning procedure; it is the PSIs in the sample which has not reported data for the reference months. In the growth rate method however we are potentially using all the PSIs and thus it becomes impossible to remind all of them. Criteria have to be developed for the selection of the most important PSIs that should be included in the dunning procedure. These PSIs should be the ones having the largest potential impact on the estimate of the growth rate. PSIs having large impact on the estimates of the growth rate are PSIs with large trade but they should also have a variation in the trade that is different from the bulk of the trade value.

A SAS application has been developed that estimates the total trade for each of the previous 12 months using the growth rate method. Each month is estimated several times. Each time one of the PSIs is excluded from the estimation and the effect of the exclusion is measured. In that way it is possible to see how much each PSI historically has affected the estimates of the growth rates. The PSIs can actually by ranked according to the impact they have had historically on the estimates. We have decided to apply the dunning procedure to those of the 500 PSIs with the highest rank that has not reported. Arrivals and dispatches are treated separately. Thus we make sure to receive reports from 500 most important arrival-PSI and 500 most important dispatch-PSI.

It is interesting to evaluate if our approach differ from a simple use of the 500 largest PSIs in terms of yearly trade value. For the reference month May 2005 it was found that for arrivals 344 PSIs are among the 500 most important according to both methods. For dispatches 355 PSIs are among the 500 most important PSIs according to both methods. Thus, the two methods yield similar results but some differences exit.

4.3 The editing

An efficient editing system has proven very difficult to construct. The total values of the PSIs vary very much between the months. It is thus very difficult to forecast the total value of trade of the PSI. Our experience is that a trade value that seems completely unreasonable compared to the previous trade of the PSI may often, or even most of the time, turn out to be correct. On the other hand a trade value that seems to be in perfect agreement with the previous reporting of the PSI may turn out to be in error. It may be recommended that personal contact should be made each month with reporters from the largest PSI to discuss the report. This should be made regardless of whether the total value seems unreasonable or not.

An editing system was however constructed in order to detect the largest and most unreasonable errors. The system tries to estimate the total value of each PSI by the mean of the total values for the previous four months. The discrepancy in SEK between the reported value and the estimated value is then computed along with the discrepancy in percent from the estimated value. These two measures can be considered to be the impact of the potential error on the estimate and the degree of suspicion respectively. All PSI with a potential error larger than 50 million SEK is then flagged for investigation. In addition, PSIs with potential errors smaller than 50 million SEK is flagged for investigation if the value is highly suspicious. The smaller the potential impact is, the larger is the suspicion required for the PSI to be flagged for investigation.

Including the two reference months and both arrivals and dispatches, about 300 to 400 PSIs are flagged for investigation each production month. Unfortunately this system yields a hit-rate of only a couple of percent. We have elaborated with different editing systems but no distinctly better results are achieved. Our hope of finding a better system is small considering the variation that exits in the data.

5 Possibilities to use Extrastat data to improve estimation

Since more or less complete Extrastat (non EU trade) data is available by the time of the publication of RE1, we decided to look at the possibilities to use Extrastat data to improve the Intrastat estimates. The simple idea why this might work for the growth rate method is that the Extrastat and Intrastat trade should have a similar development pattern over time. In other words, we could be able to adjust the growth rate estimates in Intrastat with the aid of information about the observed growth rate in Extrastat. Unfortunately, this idea has not yet been fruitful. Nevertheless, to explain why, we report some of our findings below.

A study has been done of the connection between the Intrastat and the Extrastat. Two time series for both flows have been established, one for the first published detailed statistic, both EU and non EU trade, and one for the first rapid estimate of the alternative method for both EU and non EU-trade. The time series cover the period September 2002 to February 2005.[8] To compare the Extrastat trade with the development in Intrastat trade, we calculated the growth rate (t-2 to t) of the two, for the detailed statistics ('true growth rates') as well as for first rapid estimates. The correlations between Intrastat and Extrastat growth rates are given in the table below.

Table 5
The Correlation between Extrastat and Intrastat growth rates in the detailed and rapid estimates

Flow	Data series	Correlation of Extrastat and Intrastat growth rates
Import	Detailed statistics	0.81
Export	Detailed statistics	0.92
Import	First rapid estimate	0.64
Export	First rapid estimate	0.76

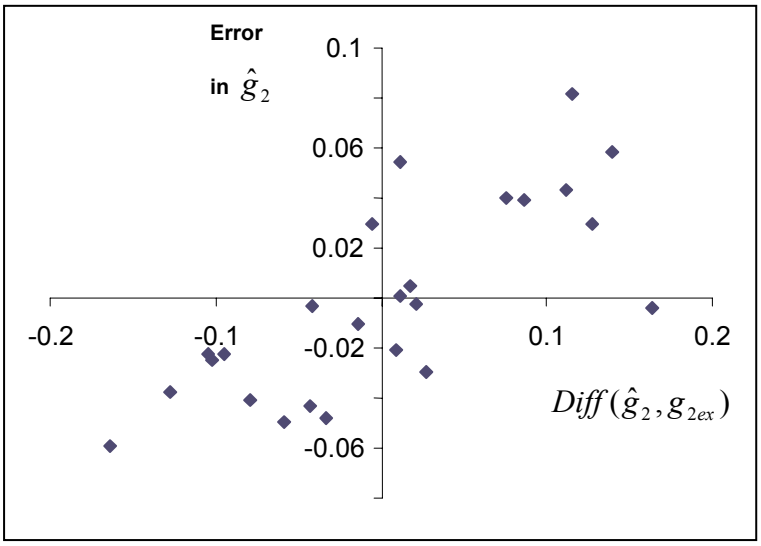
These results show that the correlation of growth rate factors between EU trade and non EU trade is quite high, especially the correlation on the export side. However, the variability in the growth rate estimates (especially for imports) seems a bit too high to be really useful in the sense of improving the growth rate estimates. This can be observed in the following graphs. (For sake of space and since the point is made anyway, we choose only to show the graph on export. The variability in the import data series is even worse than it is for the export.)

A smaller deviation between true growth rates (denoted G_2) and the estimates $\hat{g}_2$ is an improvement of the growth rate estimates in RE1. If we plot the difference between growth rates in Intrastat and Extrastat (denoted

[8] The values for May and June 2004 have been deleted due to the EU-enlargement in May 2004, which caused a structural change in the content of the Extrastat and Intrastat series.

$Diff(\hat{g}_2, g_{2ex})$) against this deviation, we get an illustration of the possibilities to use models of Extrastat data to make the deviation of $\hat{g}_2$ and G_2 smaller. For exports this is given in the scatterplot in Figures 7 below.

Figure 7
Plot of the error in growth rate estimates $\hat{g}_2$ against the difference between Extrastat and Intrastat growth rates $Diff(\hat{g}_2, g_{2ex})$, (Nov 02 – Feb 05), Exports



Although the plots reveal that the two variables covary, and have a seemingly linear relationship they also reveal that the variability is large. Our present interpretation is that the relationships between Intrastat and Extrastat growth rates are not strong enough to make valuable contributions to our estimates. At present, the Extrastat growth rates will only be used as an supplementary comparison check before final publishing is made.

References

Edicom Report 2002 Statistics Sweden, *Development of Alternative Methods to Produce Early Estimates of the Swedish Foreign Trade Statistics*

Estevao, Victor M. and Särndal Carl-Erik, (2000) *A Functional form approach to Calibration*, Journal of Official Statistics vol 16 no 4.

Lundström, Sixten and Särndal Carl-Erik, (2001) CBM *Estimation in the presence of Nonresponse and Frame Imperfections*, Statistics Sweden.

Appendix 1

Table 1
The number of PSIs in Set A. Arrivals

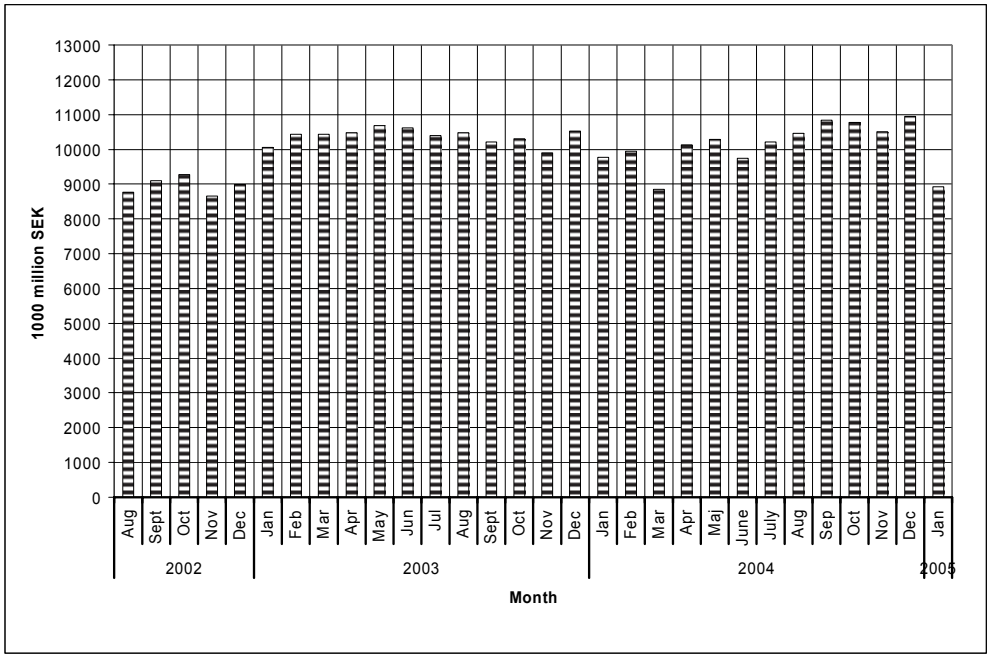


Table 2
The number of PSIs in Set B. Arrivals

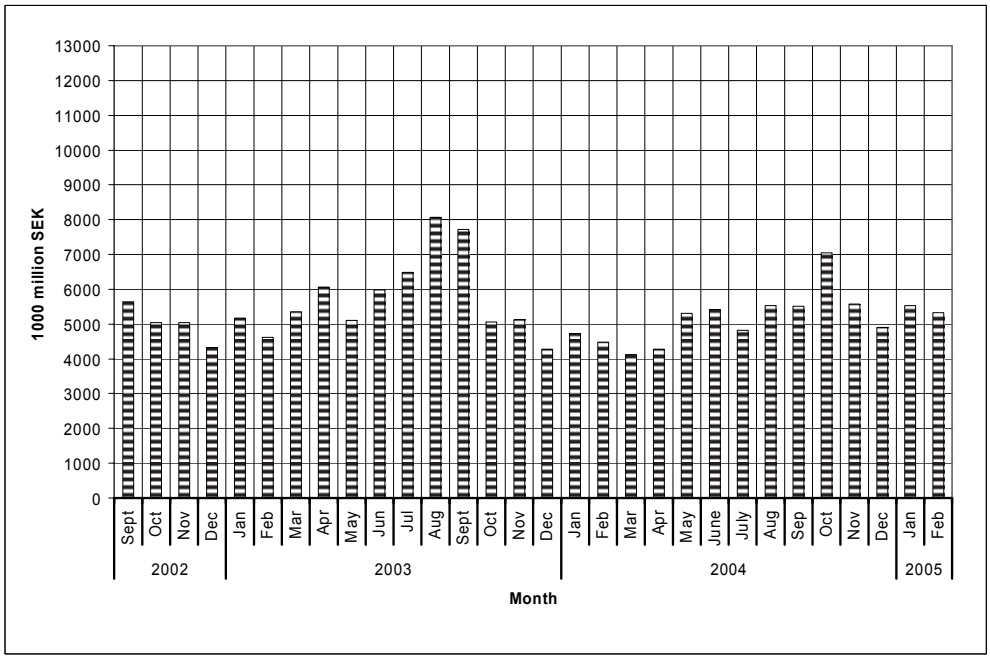


Table 3
The number of PSIs in Set A. Dispatches

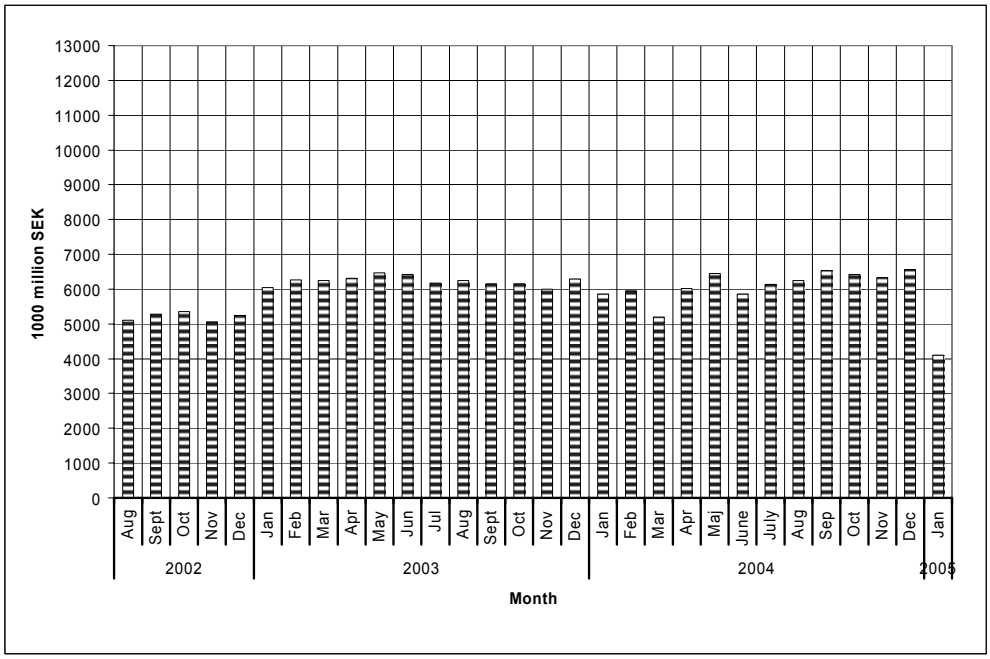
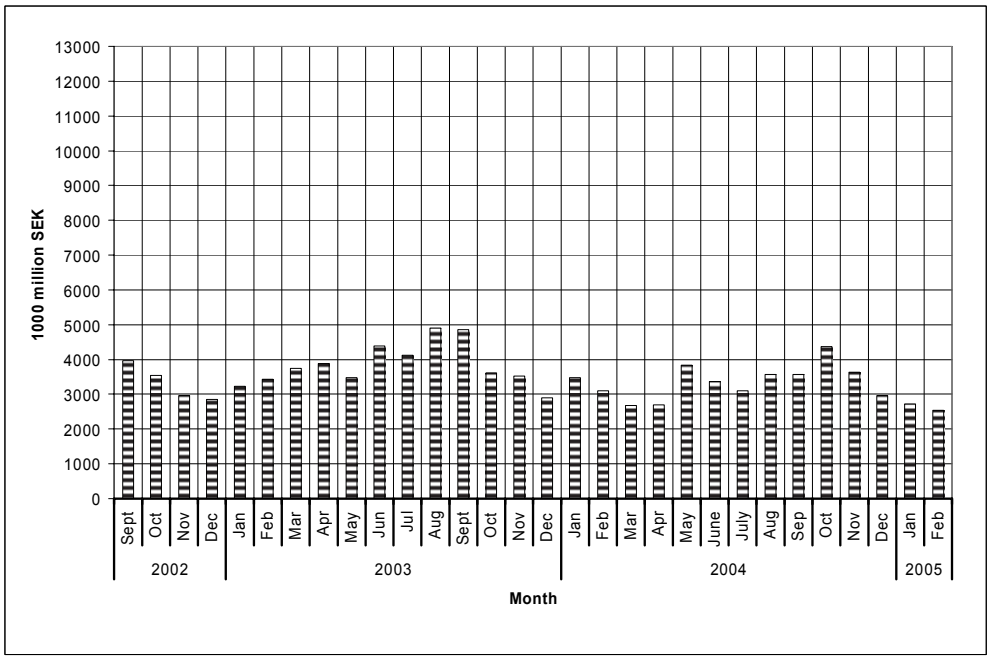


Table 4
The number of PSIs in Set B. Dispatches



Appendix 2

Table 1
The proportion of the value included in set A. Arrivals

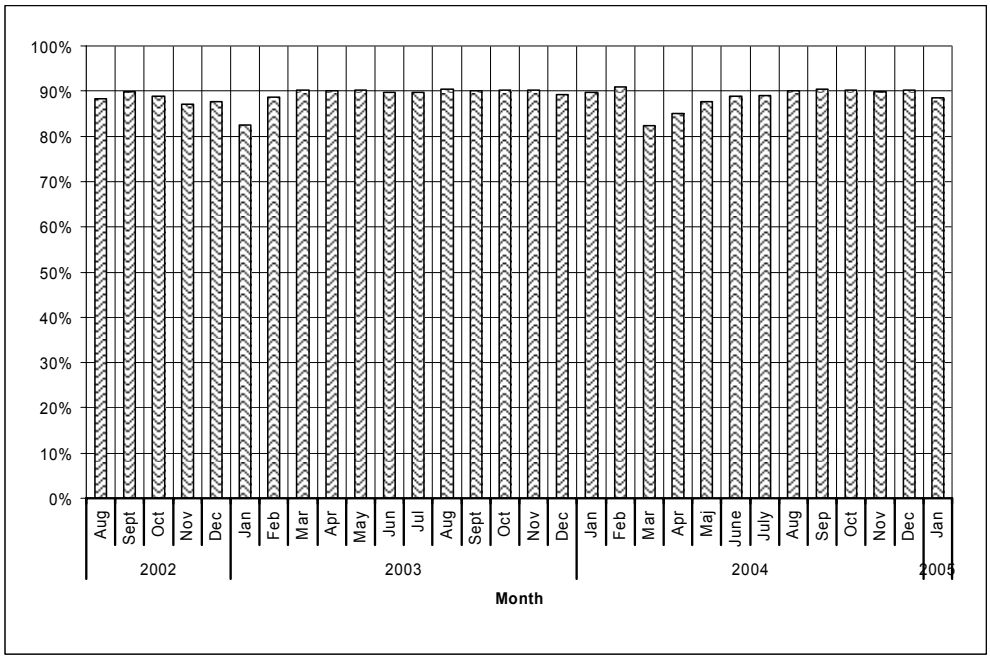


Table 2
The proportion of the value included in set B. Arrivals

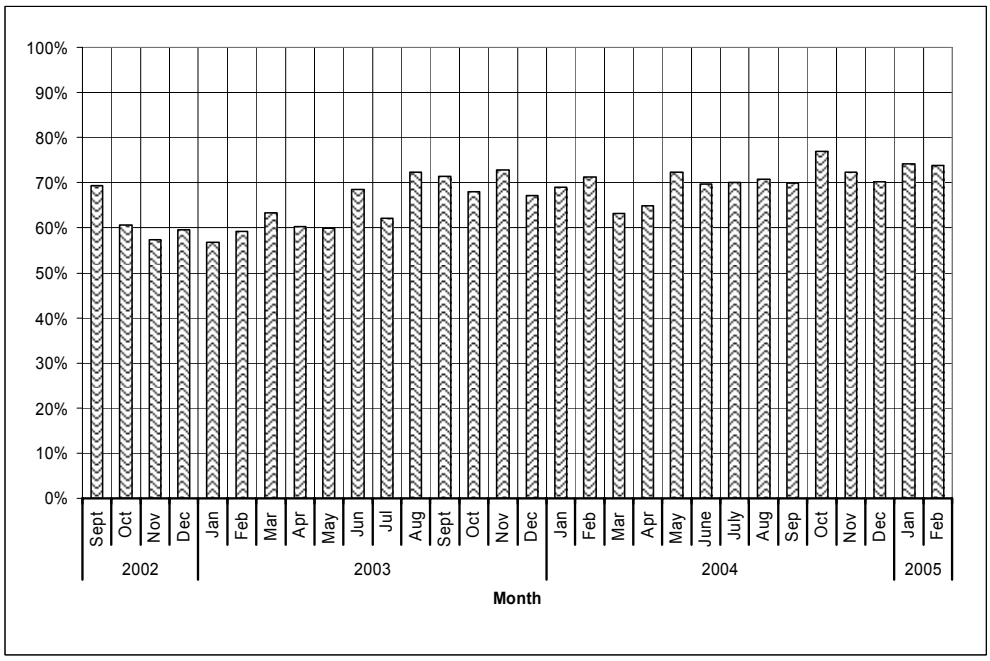


Table 3
The proportion of the value included in set A. Dispatches

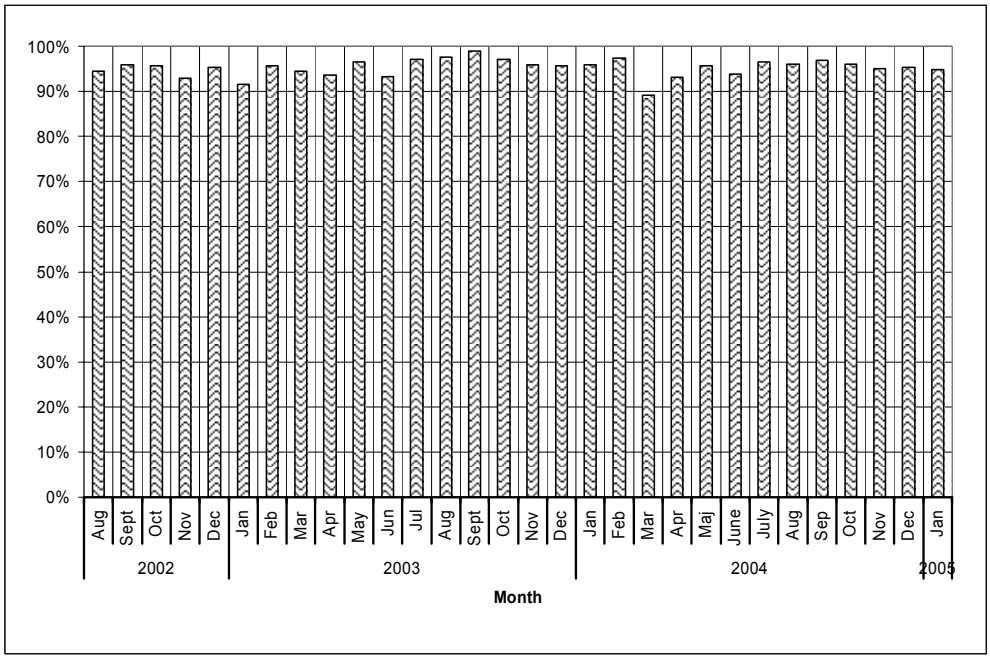
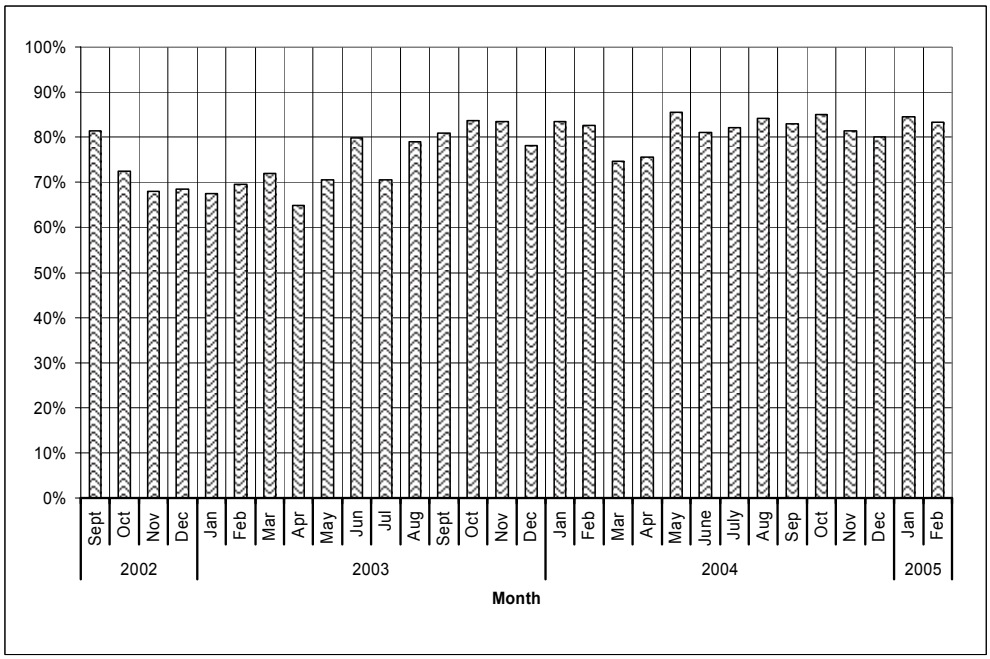


Table 4
The proportion of the value included in set B. Dispatches



Appendix 3

Table 1
Reference month t. Arrivals. Discrepancy between the first detailed estimate and the respective estimates from the sample based method and the growth rate method

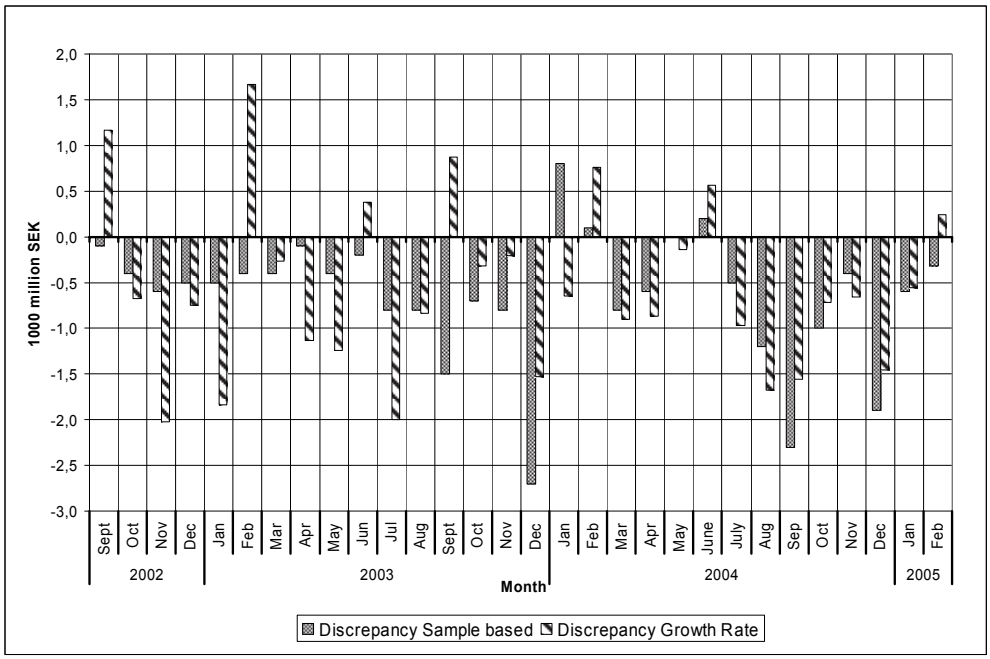


Table 2
Reference month t. Dispatches. Discrepancy between the first detailed estimate and the respective estimates from the sample based method and the growth rate method

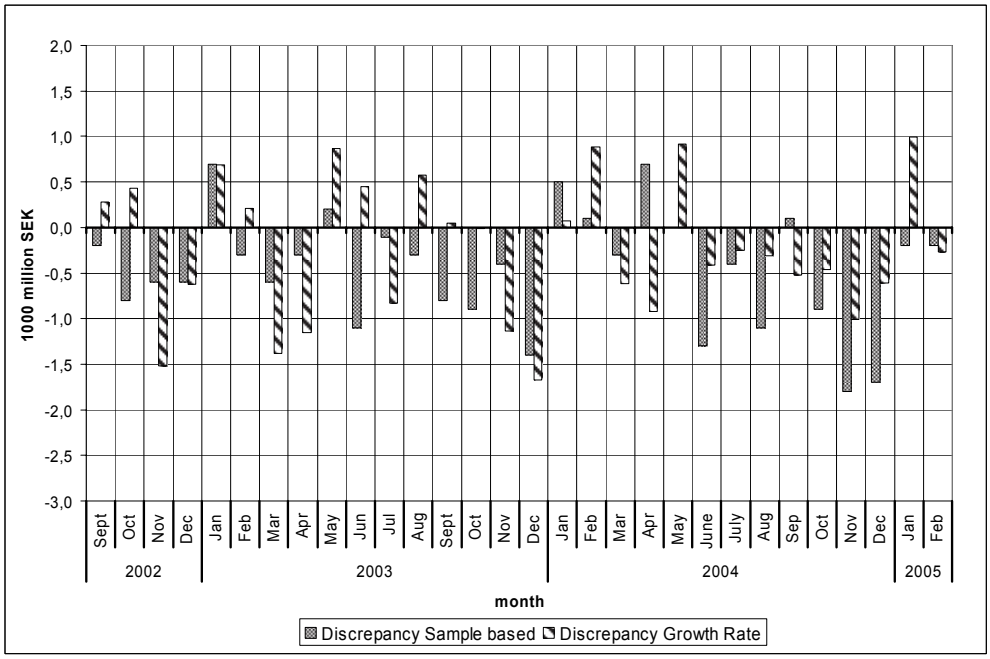


Table 3
Reference month t-1. Arrivals. Discrepancy between the first detailed estimate and the respective estimates from the sample based method and the growth rate method

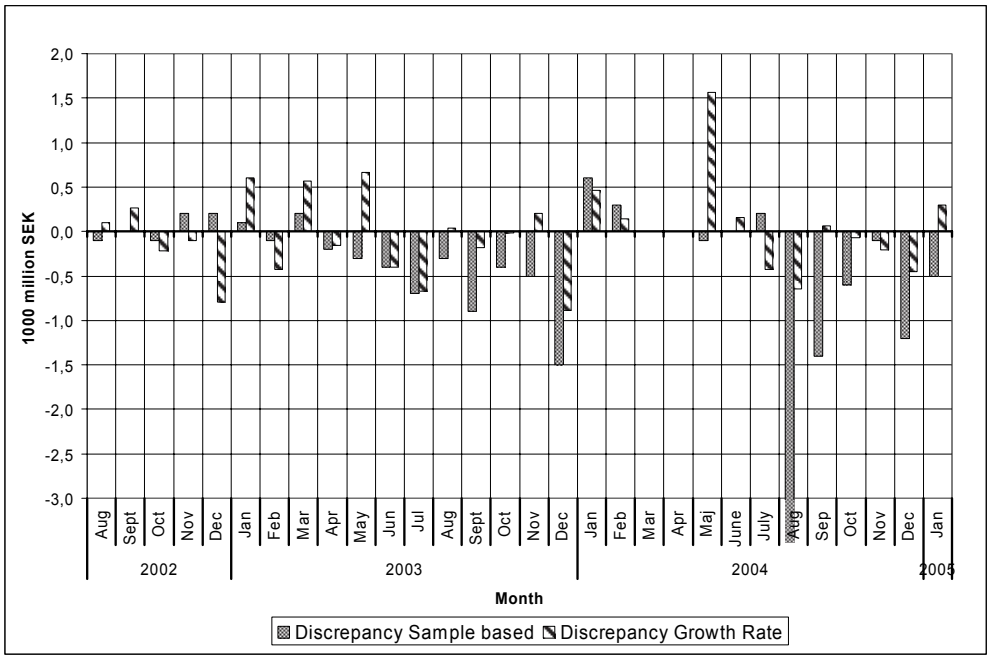
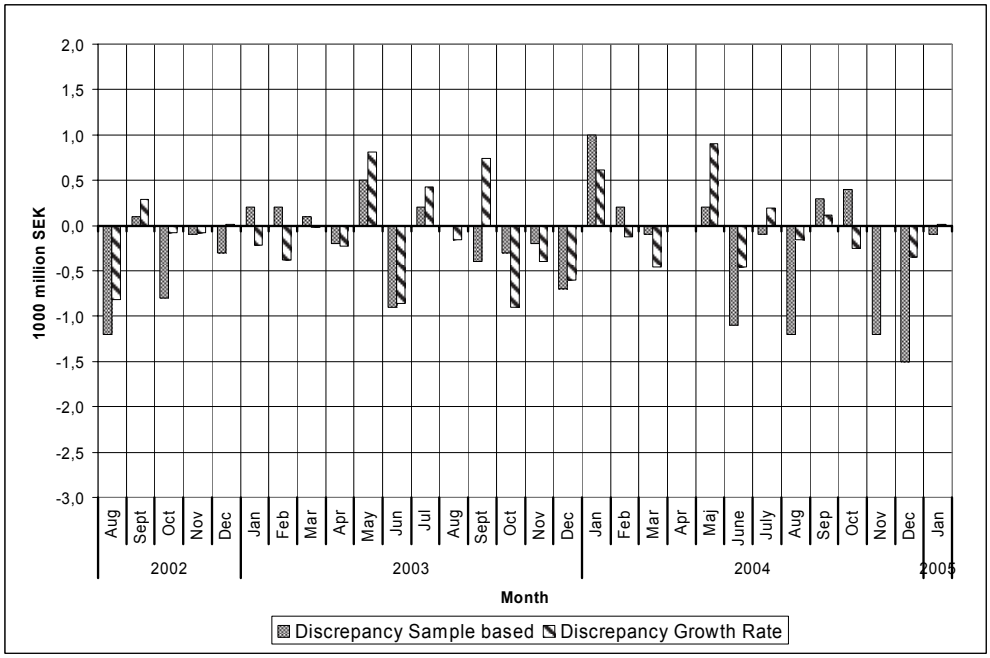


Table 4
Reference month t-1. Dispatches. Discrepancy between the first detailed estimate and the respective estimates from the sample based method and the growth rate method



Appendix 4 – Estimation of the uncertainty in the rapid estimates

As shown in section 3.3, there are a number of possible sources of the discrepancy between the detailed estimates and the rapid estimates. In section 3.3.1 we briefly mentioned selection bias. This selection bias in the rapid estimates of the growth rate method is due to nonresponse at the time of production of RE1 and RE2. In this appendix we discuss a way in which we can monitor and estimate the monthly variance due to the non-response. The framework below is also applicable to other estimators than E1 and E2 (e.g. in breakdowns or in connection with estimators using other auxiliary data). It is intended to be implemented as a quality indicator both during production and to the users. To provide a background we first briefly present some general theory on calibration.

The Calibration Estimator
In sampling applications there are two main ways of handling non-response, reweighting and imputation. Use of calibration and calibration estimators is a popular and efficient reweighting method often used in official statistics. It is done in one computational step and has nice statistical properties such as, a possible reduction of nonresponse bias, a small estimator variance and consistency.

The calibration approach and a calibration estimator can be described as follows: Let $T_t = \sum_{i \in U} y_{it}$ denote the unknown target parameter of a population U, i.e. the total of the study variable. Moreover, suppose we have a vector of auxiliary variables $\mathbf{x}_i$ with known values for all units $i \in U$, i.e. we know the vector total $\boldsymbol{T}_{xt} = \sum_{i \in U} x_{it}$. Generally, the calibration estimator of the total T_t is defined by

$$\hat{T}_{cal,t} = \sum_{i \in r} w_{it} y_{it}$$

where r is the set of responding units i in a sample s and $w_{it} = d_{it} v_{it}$ are calibration weights of every unit $i \in r$. The weights w_i are computed to satisfy the calibration equation

$$\sum_{i \in r} d_i v_i \mathbf{x}_i = \sum_{i \in U} \mathbf{x}_i = \mathbf{T}_x$$

with d_i being the inverse of the sample inclusion probability of unit i and

$$v_i = 1 + c_i \left(\sum_{i \in U} \mathbf{x}_i - \sum_{i \in r} d_i \mathbf{x}_i \right)' \left(\sum_{i \in r} d_i c_i \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \mathbf{x}_i \qquad \text{for } i \in r .$$

The calibrated weights $w_i = d_i v_i$ are derived by minimizing a distance function between d_i and w_i, details can be found in e.g. Särndal and Estevao (2000). (The factor c_i is specified by the statistician and is included

here for generality reasons.) The weights w_i will ensure that estimates computed on the response set r are consistent with the known x-totals in the population, $\mathbf{T}_x$.

The growth rate estimator as a calibration estimator

For both flows (arrivals and dispatches) the monthly Intrastat data collection is designed as a complete enumeration (census) of all obliged PSIs above the cut-off threshold. Hence, all PSIs above the threshold are obliged to report their Intrastat trade. However, at the time when early estimates are computed all PSIs have note yet reported. To estimate the monthly total trade value T_t, a suggested estimator of RE1 is (this is equivalent to expression 6 in section 2).

$$\hat{T}_t = \sum_{U_{t-2}} y_{i,t-2} \frac{\sum_{i \in B^t} y_{i,t}}{\sum_{i \in B^t} y_{i,t-2}} = T_{t-2} \frac{\sum_r y_{i,t}}{\sum_r y_{i,t-2}}$$

Basically, this is a ratio estimator, where the estimated total trade value from reference month t-2 is adjusted with a growth rate factor. The set $B^t = r$ contains the PSIs responding to both reference month t as well as t-2.

In the estimator above the y-values at t-2 play the role of the auxiliary variable, i.e. at time t these values $y_{i,t-2}$ are (more or less) known. Hence, by inserting $y_{i,t-2}$ for $\mathbf{x}_i$, $c_i = 1 / y_{i,t-2}$ and $d_i = 1$ (since all units are obliged to report) in the expression for v_i we get,

$$\begin{aligned} v_i &= 1 + \frac{1}{y_{i,t-2}} \left(\sum_{i \in U_{t-2}} y_{i,t-2} - \sum_{i \in r} y_{i,t-2} \right) \left(\sum_{i \in r} y_{i,t-2} \right)^{-1} y_{i,t-2} \\ &= \sum_{i \in U_{t-2}} y_{i,t-2} \Big/ \sum_{i \in r} y_{i,t-2} \end{aligned} \qquad \text{for } i \in r .$$

Rewriting our estimator as $\hat{T}_t = \sum_{i \in r} d_i v_i y_{it} = \sum_{i \in r} w_{it} y_{it} = \hat{T}_{cal,t}$, we see that it is a calibration estimator with the weights $w_{it} = \sum_{i \in U_{t-2}} y_{i,t-2} \Big/ \sum_{i \in r} y_{i,t-2}$ satisfying the calibration equation $\sum_{i \in r} d_i v_i y_{i,t-2} = \sum_{i \in U_{t-2}} y_{i,t-2} = T_{t-2}$.

It is well known that if the calibration estimator is to have a small non-response bias and a small variance, it depends on the strength of the auxiliary variable. In our case with the growth rate method, $y_{t-2,i}$ must be considered strong, having a high correlation with y_{ti}. Other alternative estimators can also be written as special cases of the calibration estimator. And another use of this framework is that it enables us to apply the results of the calibration estimator to derive an estimator of the error. Although not necessary for E1 variance estimates of the calibration estimator can be computed using specialized variance estimation software e.g. CLAN.

To estimate the uncertainty of the growth rate estimators we may in a given month use expressions given in the next section.

A variance estimator of the growth rate estimators

To describe the technical details how to estimate the variance of the calibration estimator would be to go too far in this report. However, we will give a short introduction to the general idea, which is based on the presentation in Lundström and Särndal (2001) section 6.4.

Generally the variance estimator suggested consists of two components, a sampling component and a nonresponse component $\hat{V}(\hat{T}_{cal,t}) = \hat{V}_{SAM} + \hat{V}_{NR}$. In our case, with a complete enumeration of U_t, only $\hat{V}_{NR}$ matters since $\hat{V}_{SAM}$ will be zero. The proposed variance estimator is then

$$\hat{V}(\hat{T}_t) = \sum_{i \in r} d_i^2 v_i (v_i - 1) e_i^2$$

where, $d_i = 1$, $v_i = \sum_{i \in U_{t-2}} y_{i,t-2} \Big/ \sum_{i \in r} y_{i,t-2}$ and

$$e_i = y_{it} - \frac{\sum_{i \in r} y_{it}}{\sum_{i \in r} y_{i,t-2}} y_{i,t-2} = y_{it} - \hat{g}_2 y_{i,t-2}$$

How well this variance estimator works in our practice is yet to be fully investigated. It is however theoretically justified and it is easy to compute. Most of its components are derived in other parts of the production process, e.g. in principle the v_i is the inverse of the proportion of the value included in set B seen in tables 2 and 4 of Appendix 2.

2004:01 Hjälpverksamhet. Avrapportering av projektet Systematisk hantering av hjälpverksamhet

2004:02 Report from the Swedish Task Force on Time Series Analysis

2004:03 Minskad detaljeringsgrad i Sveriges officiella utrikeshandelsstatistik

2004:04 Finansiellt sparande i den svenska ekonomin. Utredning av skillnaderna i finansiellt sparande Nationalräkenskaper, NR – Finansräkenskaper, FiR Bakgrund – jämförelser – analys

2004:05 Designutredning för KPI: Effektiv allokering av urvalet för prismätningarna i butiker och tjänsteställen. Examensarbete inom Matematisk statistik utfört på Statistiska centralbyrån i Stockholm

2004:06 Tidsserieanalys av svenska BNP-revideringar 1980–1999

2004:07 Labor Quality and Productivity: Does Talent Make Capital Dance?

2004:08 Slutrapport från projektet Uppsnabbning av den ekonomiska korttidsstatistiken

2004:09 Bilagor till slutrapporten från projektet Uppsnabbning av den ekonomiska korttidsstatistiken

2004:10 Förbättring av bortfallsprocessen i Intrastat

2004:11 PLÖS. Samordning av produktion, löner och sysselsättning

2004:12 Net lending in the Swedish economy. Analysis of differences in net lending National accounts (NA) – Financial accounts (FA). Background – comparisons - analysis

2004:13 Testing for Normality and ARCH. An Empirical Study of Swedish GDP Revisions 1980–1999

2004:14 Combining leading indicators and a flash estimate

2004:15 Comparing welfare of nations

2004:16 ES-avdelningens utvecklingsplan 2004

2004:17 Den svenska konsumentprisindexserien (KPI), 1955–2004. En empirisk studie av säsongsmönstret. En tillämpning av TRAMO/SEATS

2004:18 Skola, vård och omsorg i privat regi. En sammanställning av statistik

2005:01 An ignorance measure of macroeconomic variables

2005:02 Svenska hälsoräkenskaper. Ett system framtaget inom ramen för de svenska nationalräkenskaperna

2005:03 The sample project. An evaluation of pps sampling for the producer and import price index

2005:04 Finansiellt sparande i den svenska ekonomin. Utredning av skillnaderna i finansiellt sparande. Nationalräkenskaper, NR – Finansräkenskaper, FiR. Slutrapport

2005:05 Net lending in the Swedish economy. Analysis of differences in net lending National accounts, NA – Financial accounts, FA. Final report

2005:06 Varför får NR motstridiga uppgifter? Statistikens samanvändbarhet studerad inom ramen för Nationalräkenskaperna

2005:07 Evaluation of the possibility of producing statistics on production in the construction industry on a monthly basis

2005:08 Increased Automation of the Validation and Correction Processes in the Swedish Intrastat Production

ISSN 1653-316X Bakgrundsfakta
ISSN 1653-3178 Ekonomisk statistik

Statistikpublikationer kan beställas från SCB, Publikationstjänsten, 701 89 ÖREBRO, e-post: publ@scb.se, telefon: 019-17 68 00, fax: 019-17 64 44. De kan också köpas genom bokhandeln eller direkt hos SCB, Karlavägen 100 i Stockholm. Aktuell publicering redovisas på vår webbplats (www.scb.se). Ytterligare hjälp ges av Bibliotek och information, e-post: information@scb.se, telefon: 08-506 948 01, fax: 08-506 948 99.

Statistical publications can be ordered from Statistics Sweden, Publication Services, SE-701 89 ÖREBRO, Sweden (phone: +46 19 17 68 00, fax: +46 19 17 64 44, e-mail: publ@scb.se). If you do not find the data you need in the publications, please contact Statistics Sweden, Library and Information, Box 24300, SE-104 51 STOCKHOLM, Sweden (e-mail: information@scb.se, phone: +46 8 506 948 01, fax: +46 8 506 948 99).