For Official Use **STD/STESEG(2003)10**

Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development

10-Jun-2003

English - Or. English

STATISTICS DIRECTORATE

STD/STESEG(2003)10
For Official Use

OECD Short-Term Economic Statistics Expert Group

STESEG TASK FORCE ON TIMELINESS AND BENCHMARKING

MORE RAPID RETAIL TRADE STATISTICS IN SWEDEN

Paris, 26-27 June 2003

Prepared by: Anita Ullberg
Statistics Sweden
Department of Economic Statistics

Contact: Richard McKenzie (richard.mckenzie@oecd.org), Statistics Directorate, OECD

JT00145787

English - Or. English

MORE RAPID RETAIL TRADE STATISTICS IN SWEDEN

1. Introduction

Retail trade is an important short-term indicator and timeliness of the retail trade statistics has been in focus in recent years.

In 2001 the EU initiated a project on the possibility of reporting turnover data to Eurostat far earlier than at present. Statistics Sweden was one of the participants in the Task Force on Retail Trade. The aim of the work has been a substantial increase of timeliness. The ambition was that presentation of data on retail trade on the EU level should be possible within 13 days or, if this would not be possible, at least within 30 days after the end of the reference period. The use of a country stratified European sample for retail trade has been studied.

Statistics Sweden has been deeply involved in the implementation of new methods proposed and tested in the task Force on Retail Trade. The purpose of the paper is to describe the implementation of new processes to increase timeliness in Retail trade statistics in Sweden.

2. Description of the system for retail trade statistics

2.1 Introduction

Statistics Sweden conducts a monthly survey on Retail Trade. The survey is mainly used by the National Accounts to compile the Gross Domestic Product but there are several other users as well. There is only one variable collected in the survey, turnover, and the presented estimates are total turnover for retail trade (NACE 52) and a number of levels within NACE 52. This survey is designed to produce reliable estimates on a national level and cannot be used for regional breakdowns.

2.2 Definition of the population

The population consists of all active enterprises in the Swedish Business Register (FDB) classified into the economic activity NACE[1] 52 with an annual turnover exceeding 200 000 SEK. The information on annual turnover is collected from the Value Added Tax (VAT) – register.

[1] *Enterprises in FDB are classified into economic activities according to the international classification ISIC and the European counterpart NACE and at a more detailed level (5-digit) to the Swedish classification system, SNI, which corresponds to NACE at the 4-digit level and to ISIC at the 2-digit level*

An additional population is also constructed among the active enterprises in FDB classified into the economic activity NACE 52 with annual turnover less than 200 000 SEK. A sample is drawn from this population and questionnaires are sent out to these small enterprises. But there is a cut off limit introduced for enterprises with annual turnover less than 200 000 SEK and no employees. Further on in this report this additional survey on very small enterprises will be omitted, as their impact on the estimates is very small.

2.3 Stratification of the survey

2.3.1 Economic activity

The population in this survey is first divided into strata based on economic activity, so called activity strata. Principally the five-digit level is used to create these strata. The stratification is done in order to fit the domains of study as much as possible, see table 1.

Table 1 Distribution of number of enterprises over activity strata in the survey referring to the year 2000.

Activity strata	Number in population	Number in sample	Activity strata	Number in population	Number in sample
52111	5	5	52470	829	110
52112	3 096	284	52481	804	103
52120	72	25	52482	460	53
52250	3	2	52483	287	89
52260	1 710	71	52484	601	93
522XX	3 166	142	52485	2 019	188
52310	1	1	52486	381	71
5232A	504	74	52487	1 772	101
52410	1 012	50	52488	484	31
52420	3 954	371	52493	693	105
52430	866	96	52494	138	49
5244X	1 229	159	52495	117	34
5244Y	1 382	99	5249X	424	26
52451	459	52	5249Y	1 061	75
52452	1 039	143	52500	534	23
52453	309	44	52610	227	52
52454	241	32	5262A	777	46
52460	1 831	173	52700	1 816	77

522XX = 5221-5224+5227, 5232A = 5232-5233, 5244X = 52441-52442, 5244Y = 52443-52444, 5249X = 52491-52492, 5249Y = 52496-52499, 5262A = 5262-5263

2.3.2 Size strata

Each activity strata in this survey is divided into six size strata. The limits for the different size groups are based on annual turnover. Information on turnover is available for the whole population and collected from the VAT-register, but there is of course some time delay. The information from the VAT-register is about six months old every time a new sample is to be drawn.

Four different sets of size limits are used in the survey because the amount of turnover varies a lot between the activity strata.

Table 2 Distribution of number of enterprises over size groups in the survey referring to the year 2000.

Size group	Number in population	Number in sample
1	19 701	669
2	8 568	633
3	4 995	985
4	916	739
5	97	97
6	26	26

2.4 Allocation and estimation

2.4.1 Allocation

Neyman allocation is used to decide sample sizes in each stratum except for size group five and six. Enterprises in these size groups are completely enumerated in each activity strata because their impact on the estimates is very large. The variable used for allocation is annual turnover, the same as for grouping into size groups. This variable is obtainable for the whole population and correlated with the study variable.

The main purpose with the survey is to be used by the National Accounts to compile the Gross Domestic Product. This means that the most important issue with the survey is to estimate total Retail Trade with high accuracy.

The required precision, in terms of relative standard errors, is therefore set for total retail trade. The consequence is less accuracy in small and not so important domains of study.

2.4.2 Estimation

Turnover is estimated using a ratio estimator for each classification group – see page 7, *Estimation for final results*. The turnover data used for the stratification also serve as auxiliary data.

2.4.3 Sampling error

The sampling error calculated on a yearly basis for Sweden is today approximately 0.20-0.30 % in terms of relative standard error. The estimated sampling error for one month is approximately 0.40-0.50 % in terms of relative standard error.

2.5 Retail Trade Survey in practice

Once a year, in November/December, a new population is constructed and a new stratified simple random sample is drawn from the population. The population constructed and the sample drawn in November/December year (t-1) is then used in year (t). The population constructed in November/ December 1999 included 34 303 enterprises and the sample drawn from this population included 3 149 enterprises.

The majority of the selected enterprises also serve as observation units in this survey. But there are a small number of enterprises that consists of two or more legal units. In this case the legal unit could serve as observation unit, depending on what level (enterprise or legal unit) the information is obtainable for.

2.5.1 Data collection

One day before the surveyed month (T) is ended a questionnaire is sent to the enterprises financial manager. Data are collected by touchtone data entry (TDE), facsimile transmission (telefax), post or telephone. Sixty percent of the collected data is coming in by TDE. One week before the last answering day a pre-remind mail card is sent to the enterprise that haven´t been answering. One working day after last answering day telephone reminding is started. Telephone reminding is done for one week to approximately 900 enterprises each month.

Firstly the collected data is checked when data is registered or in the telephone when the enterprise leave data by TDE. Secondly the data is checked through error-lists and by a graphical check of the data.

It is very important that the large enterprises in size group five and six answer the questionnaire. But if it is impossible to get an answer from one of these enterprises the imputation is individual. The information used to impute is VAT-data, returned questionnaires from an earlier period etc.

Enterprises in size group one to four have a smaller impact on the estimates and therefore they are imputed by the average among the respondents in the respective stratum.

Today three persons are working with the retail trade statistics.

2.5.2 Time distribution of incoming answers (the situation before conforming to the EU requirements)

Time and response rates given are calculated for the reference month of October 2000.

	Event	Time	Response rate (measured as number of enterprises)
1	Sending of qusestionnaires	T-1	-
2	First answer	T+2	-
3	-	T+5	16%
4	-	T+10	35%
5	-	T+15	43%
6	Sending of pre remind mail-cards	T+19	47%
7	Last answering day	T+25	59%
8	Start telephone reminders	T+27	61%
9	End of telephone reminders	T+31	70%
10	Imputation and estimation start – Start of checking the figures	T+35	72%
11	End of checking the figures-Last estimation	T+41	76%
12	Writing press release etc.	T+42	
13	Press release	T+43	78%

For a typical month the weighted respond rate (measured with how much turnover that has been answering) is 91-93 %.

3. Description of the system conforming to the EU requirements

3.1 Introduction

The main condition for improving timeliness in the retail trade statistics was that enterprises should be able to respond earlier to the questionnaires.

In 2002 Statistics Sweden made a small survey for hundred enterprises where 81% of the enterprises answered that they could provide there figures before day T+18. For the big and important enterprises the same figure was 94 %.

Statistics Sweden decided to join the European sub-sample, but with the whole sample for the retail trade statistics (approximately 3100 enterprises). The sampling error for the Swedish retail trade statistics is approximately 0.20-0.30%, which is much less than the accepted relative standard error of 3.4 % mentioned in the final report of the Country stratified European sample for the retail trade index.

The new survey was introduced in January 2003. A new time table for different processes was implemented. The response rate was supposed to be higher than before at an earlier time but not to be as high as it was at T+43 in the old system. However a lower response rate would introduce bias in the estimation. Tests at time T+28 has shown that the estimated turnover is 1.3 - 2% lower than the estimated turnover at T+43. Statistics Sweden has introduced methods to correct this bias (non-sampling error).

3.2 Estimation

3.2.1 Estimation for preliminary results

The total turnover in each domain of study, that is activity strata, is estimated with a combined ratio estimator which is corrected for over-coverage. The auxiliary information used is annual turnover from the VAT-register. The estimator looks as follow:

The estimator for the preliminary results looks as follows:

$$\hat{t}_{y_{ra}(D)} = t_{x(D)} * \frac{\sum_{h \in D}\left[\left(\frac{N_h}{m_{\ddot{a}_h}(t) + m_{\ddot{o}_h}(t)}\right) * \sum_{k \in h} y'_k\right]}{\sum_{h \in D}\left[\left(\frac{N_h}{m_{\ddot{a}_h}(t) + m_{\ddot{o}_h}(t)}\right) * \sum_{k \in h} x_k\right]}$$

where:

$\hat{t}_{y_{ra}}$ total turnover per business

t_x total turnover from the population turnover from VAT year t-2 per business

VAT, $t_x = \sum_U x_k$

h is stratum
D is a business
N_h number of enterprise in the frame for stratum h
$m_{ä}$ number of answer exclusive over-coverage
$m_{ö}$ is number of enterprise which is identified as over-coverage
x_k is turnover for enterprise k in VAT year t-2

$$y'_k = y_k * \frac{\left(1+\gamma_h{}^p(t)\right)}{\left(1+\gamma_h{}^d(t-1)\right)}$$

where
y_k is left turnover for enterprise k.

Expression $\frac{\left(1+\gamma^p{}_h(t)\right)}{\left(1+\gamma_h{}^d(t-1)\right)}$ are over-coverage correction for preliminary results.

$\gamma_h{}^p(t)$ is estimated per stratum from preliminary figures month (t):

$$\gamma_h{}^p(t) = \frac{m_{ö_h}(t)}{m_{ä_h}(t)}$$

$\gamma_h{}^d(t-1)$ is estimated per stratum from definite figures month (t):

$$\gamma_h{}^d(t-1) = \frac{m_{ö_h}(t-1)}{m_{ä_h}(t-1)}$$

3.2.2 Estimation for final results

The total turnover in each domain of study is estimated with a combined ratio estimator. The auxiliary information used is annual turnover from the VAT-register. The estimator looks as follows:

$$\hat{Y}^{Lop}_{b,k,c} = \sum_{h=1}^{H} \frac{N_h}{m_h} \sum_{i=1}^{m} y_i * \frac{X_b}{\sum_{h=1}^{H} \frac{N_h}{m_h} \sum_{i=1}^{m_s} x_i}$$

where

y_i = Turnover for enterprise i
N_h = Number of enterprises in the population for stratum h.
H = Number of stratum.
m_h = Number of respondents in stratum h.
b = Domain (Industrial activity).
k = Month.
Lop = Current prices.
c = Year.

X_b = Total turnover for a business group (from the VAT- register year t-2).

x_i = Turnover for enterprise i in the value added tax register year t-2.

3.3 Impact on national legislation

The legal condition that says that the enterprises have to answer the questionnaires before T+25 has been changed to T+15 and a new survey based direction SCB-FS 2002:23 is set in action from January 1, 2003.

3.4 Time table for data collection

The time table for data collection has been changed as a consequence of

- The change in the legal condition for responding to the questionnaires
- Improvements that have been implemented regarding data collection and reminding.

The time table for data collection is:

No	Event	Time	Response rate
1	Sending of questionnaires	T-1	0
2	Pre remind mail-cards	T+10	40%
3	Last answering day	T+15	57%
4	Start of telephone reminding	T+16	60%
5	End of telephone reminding.	T+18	67%
6	Imputation, estimation and checking start.	T+19	67%
7	Telephone reminding to important enterprises	T+19	
8	Graphical check of figures	T+19	
9	Press release preliminary figures	T+27	75%
10	End of checking the final figures – Last estimation	T+54	78%
11	Press release final figures	T+57	78%

The weighted response rate for the more timely retail trade statistic is 90%.

The telephone reminding in the earlier system was done by two persons during five days. In the new time table four or five persons are telephone reminding under three days because more intensive reminding is necessary.

The preliminary (T+27) and final results (T+57) for the month before are presented in the same press release.

4. Risks, success factors and interdependencies

Risks

- Bias in the estimation because the responding rate is lower in the preliminary results than in the final results.
- Possible problems with the enterprises because they get less time to answer. Contact with enterprises can be worse.
- Checking of the data material has to be done during a shorter time period which can lead to the wrong figures beings accepted.
- It can be problems to get timely results for months that are calculated during vacations, holidays, first month after sample etc. Example for “problem” months is (January, June, November).

Success factors

- Timeliness is better.
- The figures are going to be more demanded by the users.
- Even if the T+27 results would be less accurate than the final figures the preliminary result has an information value for the users.

5. Impact on NSI publications

Statistics Sweden now releases figures in two publications:

- One press release after T+27 for the preliminary results. At the same time the preliminary results are sent to Eurostat.
- One press release after T+57 days for the final results. At the same time the final results are sent to Eurostat.