

SPECIAL ANALYSIS

Reduced sample size in the Business Tendency Survey

In a sample survey, sample size is an important consideration. Before the sample for the Swedish Business Tendency Survey was redrawn in May 2019, the National Institute of Economic Research (NIER) conducted an analysis of the possibility of reducing the sample size for the survey. The following provides a summary of this analysis.

A large sample reduces the risk of random errors but increases the cost of conducting the survey. At a time when the number of surveys in society is rising, it is also desirable to limit the reporting burden on firms. It is important to test the potential effects of reduced sample sizes before making such a change, because the surveys cannot be repeated.

The existing sampling methodology was studied to identify strata (see definitions) where sample sizes could potentially be reduced without appreciably affecting the results of the survey. The reductions in these strata were then used as input values for simulations on data from the NIER's actual surveys for the period from May 2018 to March 2019. For each of these 11 months, 10,000 simulations were performed for each industry, randomly excluding different firms – but the same number – in each simulation. The results of the simulations were then compared with the published results based on the actual sample.

The differences between the simulated results with a limited sample and the published results were generally small. This led to a decision to reduce the size of the sample. The number of firms in the sample was cut by approximately 20 per cent from around 6,800 to around 5,500 when the sample was redrawn in May 2019.

Reasons to reduce the sample size

The main argument in favour of reducing the sample size in the NIER's business survey is to reduce firms' reporting burden. A survey should not gather more information than necessary, and it is an explicit goal in Sweden for the authorities to work on

Definitions

Cut-off

The minimum number of employees a firm must have to be included in the sampling frame.

Target population

Firms with operations in Sweden in the industries covered by the Business Tendency Survey.

Strata

Non-overlapping groups of firms that together make up the sampling frame. Each stratum is defined in such a way that the firms in it have similar characteristics. A stratified sample is used so that the sample is representative of the target population. In the Business Tendency Survey, firms are stratified according to the number of employees they have and the industry to which they belong.

Coverage

How well the sampling frame covers the target population.

Sample industry

An industry defined on the basis of the Swedish Standard Industrial Classification (SNI) at a 2-4 digit level and used to stratify the sample. The sample industries are determined to some extent by which industries the NIER itself chooses to publish results for, and by the industries on which information is to be reported to the EU.

Sampling frame

Firms in the target population that have more employees than the cut-off for the sample industry to which they belong.

decreasing the reporting burden on firms.¹ A smaller sample also has other benefits. For the NIER, a smaller sample means lower costs, for example through less work on the survey's database of firms, fewer queries from respondents, and less work when re-drawing the sample. It also cuts the cost of mailings and hence reduces the environmental impact.

The greatest drawback of a smaller sample is an increased risk of random errors. If the sample is reduced by raising the cut-off for the number of employees, the sampling frame will also be reduced, and coverage will decrease. Low coverage can increase the risk of the results not accurately reflecting the target population. A reduction in the sample therefore risks having unwanted consequences, such as a break in the time series.

Existing sampling methodology

The Business Tendency Survey is a panel survey divided into four different sectors: manufacturing, services, retail and construction. Firms with 100 or more employees are automatically selected. A proportion of firms with fewer than 100 employees are also included. The sample is redrawn annually.² All firms included in the sample are questioned each month until the sample is redrawn. This analysis was performed on the sample that began to be used in May 2018.

To produce the sampling frame, we use Statistics Sweden's (SCB) system for co-ordinating frame populations and samples (SAMU) for March.³ SAMU provides a snapshot of SCB's business register (FDB), which includes all firms in Sweden. The use of SAMU means that the sampling frame consists of all firms with more employees than the cut-off.

The sample is stratified by industry and size in terms of employees. The sample industries are defined on the basis of the Swedish Standard Industrial Classification (SNI). The size categories are based on the number of employees at each firm. The cut-off varies between five and 50 employees depending on sector and

¹ Official Statistics Ordinance (SFS 2001:100).

² When the sample is redrawn, there are naturally changes in the sample industries, as the composition of firms of different sizes changes, and firms may be re-assigned to another industry.

³ SAMU is updated four times a year. The March update is used because of the high number of reorganisations at the start of the calendar year. These will have been captured in the FDB business register by the beginning of February.

industry. In total, samples are drawn in 96 sample industries and six size categories.⁴ Within each sector, samples are drawn in between three and 40 sample industries.

SCB produces a sampling frame to the NIER's specifications and provides the data needed by the NIER to determine the sample sizes in different industries and size categories. The NIER itself then determines how many firms to include in the survey in each stratum. Finally, SCB draws the firms from SAMU and delivers the sample for the survey.

Table 1 Sample, May 2018, number of firms

Sector	Sampling frame	Sample size	Percentage of frame
Manufacturing	3 511	1 704	49%
Construction	5 306	478	9%
Retail	8 486	1 477	17%
Services	7 860	3 129	38%
Total	24 901	6 526	26%

Table 2 Sample, May 2018, number of employees

Sector	Sampling frame	Sample size	Percentage of frame
Manufacturing	437 252	367 423	84%
Construction	200 954	91 009	45%
Retail	387 316	282 404	73%
Services	828 612	681 136	82%
Total	1 854 914	1 421 972	76%

Strategy for reducing the sample

The sampling methodology presented above was analysed to identify strata where sample sizes could potentially be reduced without appreciably affecting the results of the survey at an aggregated level.

The strategy that the NIER adopted for reducing the sample size consists of four different approaches. This section describes

⁴ The size categories used are 5-9, 10-19, 20-49, 50-99, 100-199 and 200 or more employees.

all four approaches before presenting their combined effect. For each approach, we present the changes that were proposed.

EXCLUDE STRATA WITH LOW WEIGHTS

Historically, the NIER has tried as far as possible to apply general rules for the cut-offs for the industries in each sector.⁵ This makes it easier to understand how the survey is constructed, and ordering the sample is more straightforward. The drawback of general cut-offs is that the sample then includes firms that have very little influence on the results in individual industries. Due to the industry structure, more general cut-offs also mean that some strata are so small that it is hard to rotate the sample. When a stratum is excluded, the sampling frame is affected, and the industry's weight in terms of the number of employees decreases.

Table 3 shows how the proposed new cut-offs affect the sample size. For example, the number of firms in the retail sector is reduced by almost 15 per cent.

Table 3 Proposed reduced sample sizes in strata with low weights

Sector	Number of firms in sample	Number of firms excluded	Percentage of firms excluded
Manufacturing	1 704	126	7.4%
Construction	478	-	-
Retail	1 477	218	14.8%
Services	3 129	55	1.9%
Total	6 526	399	6.1%

Table 4 shows the percentage of the sampling frame that is excluded as a result of higher cut-offs. In the case of retail, almost 5 per cent of the sector's weight is excluded.

⁵ For example, the cut-off for all firms in the manufacturing sector has been 20 employees. In the retail sector, the cut-off has differed between the three sub-sectors: 50 employees for wholesalers, 20 for the motor trade, and five for the rest of the sector.

Table 4 Change in sampling frame as a result of higher cut-offs

Sector	Number of employees in sampling frame	Number of employees excluded	Percentage of employees excluded
Manufacturing	437 252	9 735	2.2%
Construction	200 954	-	-
Retail	387 316	18 981	4.9%
Services	828 612	12 798	1.6%
Total	1 854 914	41 514	2.3%

REDUCE SAMPLE SIZES IN INDUSTRIES WITH LOW WEIGHTS

Some industries for which separate results are reported have comparatively little impact on one or more industry aggregates. The smaller size categories in these cases can still have a comparatively large effect on the results for the individual industry. Therefore, we do not raise the cut-off for the industry, but instead we reduce the size of the sample in the smaller size categories. In this and the following approaches, the sampling frame is unaffected.⁶ Table 5 shows how many firms are excluded as a result of this approach.

Table 4 Proposed reduced sample sizes in industries with low weights

Sector	Number of firms in sample	Number of firms excluded	Percentage of firms excluded
Manufacturing	1 704	60	3.5%
Construction	478	-	-
Retail	1 477	12	0.8%
Services	3 129	249	8.7%
Total	6 526	321	4.9%

REDUCE THE SAMPLE IN STRATA WHERE ADDITIONAL FIRMS ARE DRAWN

There are considerable differences in the number of firms in different strata in the Business Tendency Survey. In the strata with

⁶ This is because the weights of the remaining firms in the size category are adjusted up in line with the weights of the firms excluded. Thus, the industry retains its relative weight in the industry aggregates in which it is included.

the most firms, we have previously had larger sample sizes than needed in order to be able to comment on the individual industry. This was to get better estimates for industry aggregates. However, it may be possible to reduce the sample in these strata without any major negative consequences. The proposed reductions are presented in Table 6.

Table 5 Proposed reduced sample sizes in strata where additional firms are drawn

Sector	Number of firms in sample	Number of firms excluded	Percentage of firms excluded
Manufacturing	1 704	99	5.8%
Construction	4786	6	1.3%
Retail	1 477	156	10.6%
Services	3 129	169	5.9%
Total	6 526	430	6.6%

REDUCE SAMPLE IN "OPERATING OF OWN REAL ESTATE" INDUSTRY

In the "Operating of own real estate" industry, firms' revenue is used as a proxy for the number of employees.⁷ The NIER has previously drawn a disproportionately large number of firms in this industry in relation to its size.⁸

Table 6 Proposed reduced sample sizes for "Operating of own real estate"

Industry	Number of firms in sample	Number of firms excluded	Percentage of firms excluded
Operating of own real estate	262	177	67.5%

COMBINED EFFECT ON THE SAMPLE

Applying all four approaches above decreases the number of firms in the sample by approximately 20 per cent (see Table 8).

⁷ Many firms in this industry have few employees despite high levels of revenue. Revenue is therefore considered a better reflection of a firm's size than the number of employees.

⁸ The previous method meant that all firms above the cut-off were included in the sample. The proposal was to reduce the sample in this industry by randomly selecting firms with between 20 and 99 "employees" according to the proxy variable.

In terms of the number of employees, however, the reduction is just 2.5 per cent (see Table 9).

Table 7 Proposed reductions, number of firms

Sector	Actual 2018		Simulated 2018		Percentage of firms excluded
	Sampling frame	Sample size	Sampling frame	Sample size	
Manufacturing	3 511	1 704	3 297	1 419	16.7%
Construction	5 306	478	5 306	472	1.3%
Retail	8 486	1 477	6 254	1 091	26.1%
Services	7 860	3 129	7 561	2 479	20.8%
Total	24 901	6 526	22 156	5 376	19.5%

Table 8 Proposed reductions, number of employees

Sector	Actual 2018		Simulated 2018		Percentage of employees excluded
	Sampling frame	Sample size	Sampling frame	Sample size	
Manufacturing	437 252	367 423	427 517	353 195	3.3%
Construction	200 954	91 009	200 954	90 704	0.2%
Retail	387 316	282 404	368 335	273 561	2.3%
Services	828 612	681 136	815 814	637 799	2.9%
Total	1 854 914	1 421 972	1 812 620	1 355 259	2.5%

LIMITATIONS

Another way of reducing the sample size would be to depart from the principle of automatic selection in the size category of firms with 100-199 employees. In some industries, the weight of this size category in terms of the number of employees is comparatively small in relation to the other size categories. We have chosen not to analyse such a change at this stage, partly to limit the scope of this project and partly to limit the number of changes to the sampling methodology at one time.

Calculation of net balances

A majority of the questions in the Business Tendency Survey are reported in the form of net balances.⁹ The net balance is the

⁹ Some questions have more response options or are answered numerically.

difference between, for example, the number of respondents replying “increased” to a question and the number replying “decreased”. In this analysis, net balances are used to present the impact of the proposed sample reduction.

Example of the calculation of a net balance: If 40 per cent of firms reply that their production volumes have increased, 10 per cent that they have decreased and 50 per cent that they have remained unchanged, the net balance will be $40 - 10 = 30$. The net balance is thus always between -100 (all respondents responding negatively) and +100 (all respondents responding positively).¹⁰

Simulations

To investigate how great an effect a reduced sample would have had on previously published results, we carried out simulations based on the strategy presented above. We used a time period where the survey had an unchanged sample: May 2018 to March 2019. For each of these 11 months, we performed 10,000 simulations per sample industry, randomly excluding firms in each simulation so that the final sample size was in line with Table 8. We then merged the results for industries where new results had been simulated and industries where no changes to sample sizes were proposed, in order to see the effect at sector level as well.

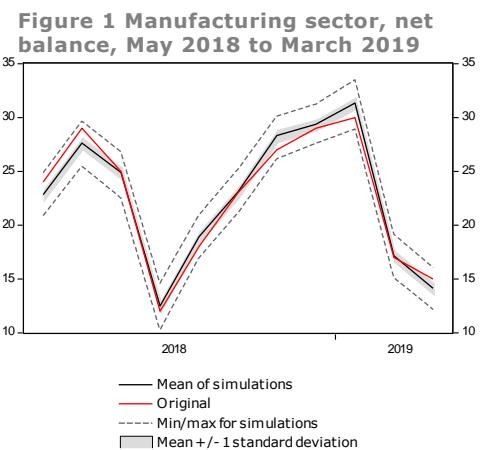
RESULTS

In this section, we report the results of the simulations with reduced sample sizes on non-seasonally-adjusted net balances for one question per sector. For the other questions that were simulated, but are not reported here, similar conclusions were drawn. The effects of the sample reductions in the construction sector were so small that the results are not presented here. From the results of the simulations, we report the mean, standard deviation and min/max values, along with the published results.

Figure 1 compares the published original series for the question on production volumes in the manufacturing sector¹¹ with a series showing the mean of the simulations. The results of the simulations show only minor differences from the original series. The standard deviation for the simulations of the net balance for

Key to diagrams

The charts below compare the published non-seasonally-adjusted original series with a mean of the 10,000 simulations. The min and max values are the lowest and highest values respectively of all the simulations. The standard deviation for the simulated values is used to show the variation in the simulations. The grey band is the interval for the mean plus/minus one standard deviation.



¹⁰ See the appendix for a more detailed description of the calculation of net balances.

¹¹ How have the firm's production volumes developed over the past three months? [increased/remained unchanged/decreased]

a period is approximately 0.5 points and much the same for all months. Half a point is comparatively little. It can be compared with the published net balance for the question having moved by an average of 5.3 points from one month to the next during the period we simulated. The min and max values from the simulations show that, even for the most extreme cases, the results are close to the original series.

Figure 2 presents the question on the business situation for firms in the service sector.¹² In this sector too, the results show only minor differences from the original series. The standard deviation again averages 0.5 points.

Finally, we report the results for the question on retailers' sales volumes¹³ in Diagram 3. The standard deviation in the simulated results is greater in the retail sector than for manufacturing and services, averaging 1.5 points over the whole period, but this variation is still low in relation to how the time series normally moves.

SENSITIVITY ANALYSIS

To examine what would happen if we instead reduce the sample sizes more drastically in a specific industry, rather than making reductions in different types of industries, we performed a test on the "Non-specialised retail sale" industry, which has a relatively high weight of small firms.

Figure 2 Service sector, net balance, May 2018 to March 2019

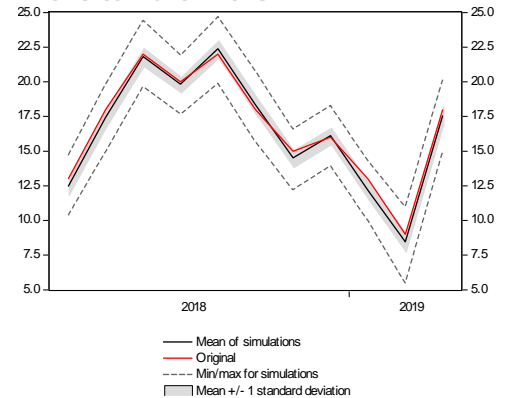
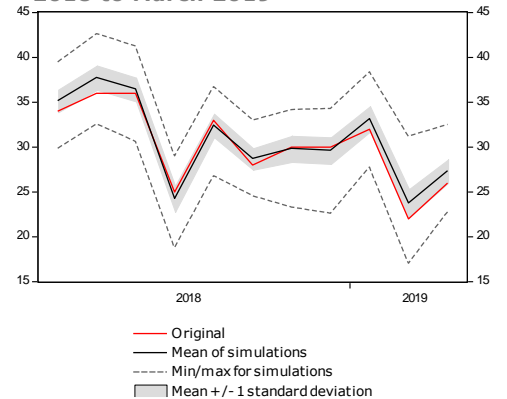


Figure 3 Retail sector, net balance, May 2018 to March 2019



¹² How has the firm's business developed over the past three months? [improved/remained unchanged/deteriorated]

¹³ How have the firm's sales volumes developed over the past three months? [increased/remained unchanged/decreased]

Table 9 Test simulation, “Non-specialised retail sale”

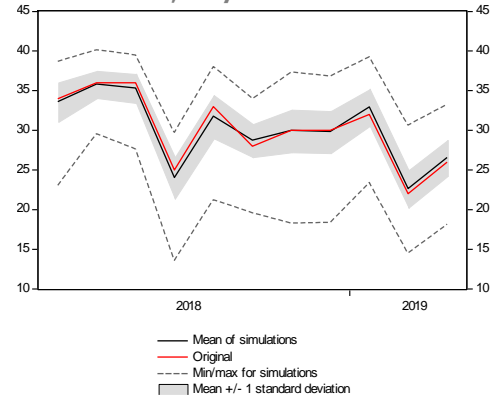
Number of employees	Number of firms in sampling frame	Number of firms in sample	Number of firms in simulation	Reduction in sample	Number of employees
5-9	592	30	10	20	3 991
10-19	634	30	10	20	8 804
20-49	465	39	10	29	14 052
50-99	160	30	10	20	11 062
100-199	64	64	64	0	8 482
>=200	22	22	22	0	34 167
Total	1 937	215	126	89	80 558

In this test, the sample was reduced by 89 firms (see Table 10), which can be compared with the simulation above where we excluded a total of 386 firms from different parts of the retail sector.

As in the first simulation for the retail sector, the mean is close to the original series (see Figure 4). However, the standard deviation in the test simulation is higher, averaging 2.5 points over the whole period. In other words, we excluded only a quarter as many firms in this test as before, but almost doubled the standard deviation. This means that the uncertainty in the estimates would be greater for the retail sector as a whole if firms were to be excluded as shown in Table 10, even though fewer firms are excluded in that simulation.

Conclusions

Overall, the differences between the simulations and the published results were generally small. This led to a decision to reduce the size of the sample in line with the proposed strategy. The number of firms in the sample was cut by approximately 20 per cent from around 6,800 to around 5,500 when the sample was redrawn in May 2019.

Figure 4 Test simulation, retail sector, net balance, May 2018 to March 2019

Appendix

To compute the net balance for a question, we begin by calculating a total weight, V_b^a , for each response option (a) and sample industry (u). This is done in two steps. First, we adjust up each firm's weight by two factors: a population adjustment and a non-response adjustment.

As the probability of firms in different strata being drawn varies, we need to extrapolate to the level the responses would have been had all firms in the population in a stratum been asked to participate in the survey. Small sampling units will have their weight adjusted up, as they represent other small units that were not included in the sample. Here, we use the inverse of each stratum's sampling fraction, i.e. the inverse probability of each sampling unit being selected ($\frac{N_j}{n_j}$, where N_j is the total number of firms in size category j in the sampling frame, and n_j is the total number of firms in the sample in size category j).

The weights of the firms choosing each response option are summed for each stratum (size category and sample industry). These values are then extrapolated to the level they would have been had all firms in the stratum in the sample answered the question ($\frac{\sum_i W_{ij}}{\sum_i W_{ij}^r}$, where $\sum_i W_{ij}$ is the total weight of all firms i in size category j , and $\sum_i W_{ij}^r$ is the total weight of all firms i in size category j that answered the question). This means that if a question is answered by firms with a combined weight of 90 per cent (90/100) of the weight of the entire sample (for a stratum), the firms' weighted responses are multiplied by a factor of 100/90.

$$W_{ij}^* = W_{ij} \sum_{j=1}^S \frac{N_j}{n_j} \frac{\sum_i W_{ij}}{\sum_i W_{ij}^r} \quad (1)$$

Where:

j is size category 1 to S , where S is the number of size categories in the industry

W_{ij} is the weight of firm i in size category j

W_{ij}^* is the upwardly adjusted weight of firm i in size category j

We next calculate the total weight for each response option in the sample industry:

$$V_u^a = \sum_{j=1}^s \sum_{i=1}^n W_{ij}^{*a} \quad (2)$$

Where:

$\sum_i W_{ij}^{*a}$ is the total weight of all firms i in size category j for response option a

To combine these total weights into an industry aggregate (b), we sum all the total weights for the relevant sample industries (u) for each response option (a):

$$V_b^a = \sum_u V_u^a \quad (3)$$

We then compute gross balances, i.e. the share of firms choosing each response option, for the different industries and industry aggregates:

$$B^a = \frac{V^a}{V^+ + V^= + V^-} \quad (4)$$

Finally, we calculate the net balance for the question:

$$N = B^+ - B^- \quad (5)$$