**Improving efficiency of the sample design and reducing survey costs in the Finnish horticultural survey**

Anna-Kaisa Jaakkonen, Natural Resources Institute Finland, Finland, anna-kaisa.jaakkonen@luke.fi

Johanna Laiho-Kauranne, Natural Resources Institute Finland, Finland, johanna.laiho-kauranne@luke.fi

Mika Kuoppa-aho, Natural Resources Institute Finland, Finland, mika.kuoppa-aho@luke.fi

**Abstract**

*The horticultural statistics are collected for the European agricultural policies based on the regulations on horticultural crops and crop statistics. The Finnish horticultural survey has been conducted annually as a total survey with a threshold on standard economic output (SO) of the horticultural enterprises. The standard output of an agricultural product (SO), is defined as the average monetary value of the agricultural output at farm-gate price, in euro per hectare or per head of livestock. Traditionally the threshold has been relatively low in the horticultural survey in comparison to the average wages of the household for example. Therefore, the impact of the small horticultural enterprises for the final survey estimates is studied including their impact on the survey costs and on the quality of the survey data.*

*In this paper, we investigate the impact of increasing the threshold both on the quality and coverage of the final estimates as well as the impacts on the survey cost component. With the sensitivity analysis using previous survey data, we can present the detection of the optimal threshold on the standard economic output to balance the survey costs and the quality criteria of the survey defined in the EU regulation survey for horticultural crops. We also present the method of deriving and monitoring the development of the standard economic output for the horticultural enterprises in the sampling frame.*

*The horticultural survey uses both the auxiliary data from statistical farm and horticultural register and collects directly information about the production of horticultural products, as well as on the production and the use of energy at the horticultural enterprises. To reduce the survey costs, and to improve the quality of the survey data, the sampling design of the survey is reviewed in detail. The data collection of the horticultural survey uses mixed-mode approach; using register data, web-survey, telephone interviews and accepts also paper questionnaires. With the increase of the thresholds we can also analyse the increase in the web-survey response rate. Thus it is expected that larger horticultural enterprises tend to respond through the web-survey more likely; while those who are interviewed tend to be on average smaller enterprises.*

**Keywords:** thresholds, cut-off sampling, response burden, survey costs

**1. Introduction**

The improvement of the sampling designs is increasingly topical as there are extensively new information needs, and the statistical offices must balance between the statistical and response burden. The driver for our study is the urgency to reduce overall survey costs. We investigate the impact of increasing the survey threshold on the survey quality, coverage of the final estimates and on the survey costs. We have compared the possibilities to increase the thresholds to reduce the sample size in contrast to reduce the frequency of the surveys or drastically cut down the information contents. However, the latter two scenarios impose the risk that a new survey for national data needs would be required, possibly even increasing the current response burden. Thus we are studying the scenario of increasing the cut-off sampling threshold to reduce costs.

Our survey costs consist of mailing costs of advance letter to the respondents, fixed costs for development of web questionnaire and varying costs of telephone interviews per interviewee who has not responded to the web questionnaire, and the salary costs of the researchers and statisticians. Thus the number of respondents in the sample is directly connected with the survey costs. With the sensitivity analysis using previous survey data, we can detect the optimal threshold on the standard economic output to balance the survey costs and the quality criteria of the survey defined in the EU regulation survey for horticultural crops. We also present the method of deriving and monitoring the development of the standard economic output for the horticultural enterprises in the sampling frame.

The horticultural survey uses both the auxiliary data from statistical farm and horticultural register and collects directly information about the production of horticultural products, as well as on the production and the use of energy at the horticultural enterprises. To reduce the survey costs, and to improve the quality of the survey data, the sampling design of the survey is reviewed in detail. The data collection of the horticultural survey uses mixed-mode approach; using register data, web-survey, telephone interviews and accepts also paper questionnaires. With the increase of the thresholds we can also analyse the increase in the web-survey response rate. Thus it is expected that larger horticultural enterprises tend to respond through the web-survey more likely; while those who are interviewed tend to be on average smaller enterprises.

**2. Horticultural survey in the nutshell and the legal framework**

The data is required from each EU-MS to monitor development of European agriculture for the purposes of CAP – Common Agricultural Policy, which makes 40 % of the total budget of the European Union and 40 % of land use in the EU. The vegetable sector is a key sector in EU agriculture, weighting 14 % of EU agricultural output in 2016. Tomatoes, carrots and onions were the most important vegetables in 2016. In Finland, the horticultural statistics illustrate both the structure of commercial horticultural enterprises, and their outdoor and greenhouse production as well as mushroom cultivation. The legal framework for the survey is in the EC Regulation 543/2009 on Crop Statistics and the 1337/2011 on horticultural crops in ESS.

**Figure 1. Output of the agricultural industry, EU-28, 2016 (% of total output)**



Source: Agriculture, forestry and fishery statistics 2017 edition

The intensity and market value orientation of horticultural farming in contrast to other areas of agriculture is shown in the economic efficiency of input and output analysis. In Finland the horticultural production is conducted in 3 700 enterprises, and the
total utilised area of agriculture is 19 000 ha, of which greenhouses cover 390 ha. The share of horticultural farms out of all agricultural enterprises is 5%, and the share of horticultural production of utilised agricultural area is only 1 % while the share of actual market value of economic output of the agriculture is 12%.

The horticultural survey has been conducted as total annual sample survey of horticultural enterprises since 1984. Introduction of new sampling methods has not been active as both the level of detail required by the EU regulation, and the heterogeneity of the enterprises has been very high due to details of horticultural crops required by density and age categories, location, large and small enterprises. Thus the horticultural survey contains a large variety of plants and crops.

The survey design has been developed to be mixed mode of mailed advance letter followed by web survey, and complemented by telephone interviews. The web response rate has been above 50% of all respondents. The total response rate achieved in the survey has been close to 98% of horticultural farms. And the survey covers of almost 100% of the land in horticultural production. The survey contains also every third year an additional component on the use of energy in the horticultural farms, with the special focus on the use of energy in the greenhouses in Finland. The pre-conditions and constraints for the change in the survey design are that the key variables to be estimated are totals, as the units and the areas of production are known from agricultural registers.

.

**3. The urgency to modernise and improve the efficiency of the survey and reduce survey costs**

*3.1. Strategic modernisation requirements from the ESS*

As modernising the survey design, we need to take into account both the EU and national information needs as well as the reliability of the energy consumption of horticultural farms statistics produced simultaneously. In addition, we need to comply with the requirements of the volume of production for the European farm structure survey in relation to horticultural farms. We have recognised that at national level the importance of the economic value has risen in connection with the monitoring the volumes of production at EU context of agricultural statistics produced for the information needs of the Common Agricultural Policy. The preconditions for modernization follow the strategic development needs identified in the ESS Statistical Program.

The changes in the survey design must result in reduction of the response burden of the horticultural enterprises and reduce the survey costs without sacrificing too much the accuracy and quality of the survey information. In the EU-Regulation on horticultural crops it is specifically mentioned that the statisticians must avoid placing an unnecessary burden on farms and administrations, and that thresholds should be established that exclude non-relevant entities from the basic entities in respect of which statistics on horticultural crops are to be collected.

We have also evaluated other means of introducing savings and analysed their impact to survey costs. Changing the annual survey into semi-annual survey using estimation would increase salary costs of the researchers while the survey costs would be reduced. In other words, saving the direct survey costs would lead into increase in salary costs and development costs. We have also been concerned on the trade-offs between abandoning mailing of advance letters, which would reduce web-response rates. Increase web-response by all possible means could also lead to reduction in interviewer costs, which we aim to promote by intensified reminders on responding to the web by using text messages. We have abandoned the option to accept only web-responses, as currently in Finland this would both reduce the response rates and lead to inadequote coverage of the survey.

Statistics to be provided on horticultural crops referred to under points (a) to (l) of Article 1(1) shall be representative of at least 95% of the total planted area producing entirely or mainly for the market of each horticultural crop referred to in each Member State.

*3.2. Choosing the survey strategy – cut-off sampling based on standardised economy output*

Standard Output (SO) is the estimated average economic yield for farm products per hectar or per production animal excluding agricultural benefits for production. In the beginning of our experiment the threshold had been set to 2 000 euros in the frame population in statistical register of farms and horticultural enterprises which is based on IACS. In the Table 1 we show how the SO per hectare varies across different production branches. We also show the results of our sensitivity analysis how the impact of increasing the threshold would increase the average number of hectares in production per SO unit.

**Table 1. The impact of standard output (SO) cut-off threshold in the survey population in relation to average hectares (ha) in production**

|  |  |  |
| --- | --- | --- |
| **Crop production**  | **SO eur/ha** | **hectares per SO**  |
| **2 000**  | **10 000** | **15 000** | **20 000** |
| Vegetables in greenhouse | 797 180 | 0,003 | 0,013 | 0,019 | 0,025 |
| Ornamentals in greenhouse | 622 090 | 0,003 | 0,016 | 0,024 | 0,032 |
| Open field vegetables and strawberry | 12 400 | 0,2 | 0,8 | 1,2 | 1,6 |
| Open field berries | 6 750 | 0,3 | 1,5 | 2,2 | 3,0 |
| Potatoes | 3 570 | 3,9 | 19,6 | 29,4 | 39,2 |
| Oats | 510 | 0,6 | 2,8 | 4,2 | 5,6 |

*3.2. Sensitivity analysis on impact of increasing the cut-off threshold*

We demonstrate the estimated a-prior impact of increasing the cut-off value initially by sensitivity analysis conducted prior the change in the survey design. Subsequently, we then show the final post impact of the change in the survey design. Our analysis indicates that the impact of increasing the threshold is large in production sectors with small economic value and thus the impact would be non-existent for the greenhouse production, which are larger professional enterprises and which is the leading value creator for the total output value for the horticultural sector (Table 2).

The agricultural sector is in our analysis seen as part of enterprise statistics, and also the trend in the structural change is towards to the large production entities that enters to the market, and not in small kitchen farms and the focus in horticultural statistics is less in social aspects. Thus the interest of the statistics is not in the smallest units and their economic value.

**Table 2.The impact of increasing the cut-off threshold on estimated production volumes and coverage of farms and utilised agricultural area**

|  |  |  |
| --- | --- | --- |
| **Crop production** | **Difference: SO 2 000 eur / 10 000 eur** | **% Difference: 2 000 eur / 10 000 eur** |
| **# Farms** | **ha** | **1 000 kg** | **# Farms** | **ha** | **1 000 kg** |
| Strawberries | -267 | -86 | -161 | -24 | -3 | -1 |
| Highbush blueberry | -53 | -25 | -33 | -36 | -35 | -28 |
| **All berries** | **-543** | **-352** | **-320** | **-34** | **-6** | **-2** |
| Carrots | -52 | -7 | -110 | -14 | 0 | 0 |
| Garlic | -23 | -5 | -5 | -34 | -26 | -10 |
| **All open land vegetables** | **-251** | **-84** | **-307** | **-18** | **-1** | **0** |
| Tomatoes | -2 | 0 | -1 | -1 | 0 | 0 |
| Butter-head lettuce | 0 | 0 | 0 | 0 | 0 | 0 |
| **All greenhouse vegetables** | **-3** | **0** | **-1** | **-1** | **0** | **0** |
| **Bedding plants (# 1 000)** | **-2** | **0** | **-2** | **0** | **-** | **0** |

Based on the results in the table 2, we focus on comparing the increase of the SO threshold from 2 000 euros upto 10 000 euros. We examine the expected impact of increasing the threshold from 2 000 euros upto 10 000 euros for horticultural farms producing mainly berries across NUTS 3 regions across all production branches. There is most impact for the berry farming, which has most heterogeneity in the distribution. The increase of the threshold from 2 000 euros to 10 000 euros cuts off 34% of berry farms, 6% of production land but only reduces 2% of the total crop coverage. With special production such as with blueberry farming the cut off of the farming land is 35-38% and in crops 28%.

In production branches where the size of the horticultural farms is larger in economic terms, the impact of the cut-off is smaller. Berry farms are often small in relation to the value of production thus the cut off affects them more severely. They are also more concentrated to the Eastern Finland while the main production branches are concentrated to the Western Finland (Cartogram 1). We estimated that by introducing a higher cut-off threshold of 10 000 eur SO per ha we can impose significant reduction in response burden with minimum impact on the reduction of survey coverage. We still can utilise register information for the number, production branch and utilised area of agriculture also for the smallest farms. Thus the savings in the survey costs outweigh are balanced with the quality of survey estimates.

**Cartogram 1. Regional variation and concentration of horticultural farms**

**Greenhouse vegetables** **Outdoor vegetables Berries**

Horticulture, gross return total 656 million euros, of which:
greenhouse vegetables 37%, outdoor vegetables 28% and berries 14%.

**4. Methods for for bias correction under investigation**

In the near future we will examine whether we can introduce even higher cut-off sample by incresing the existing 10 000 eur SO per hectare upto 15 000 eur per hectar. Then for the final estimates we would neef to introduce plausible bias correction methods for the smallest farms we can estimate from based on the available auxiliary information. These methods would include studying the usability of donor imputation of the nearest neighbourhood. The limitation of this approach is that the rare production branches and with small farms the imputation can be highly demanding. We have therefore plans to develop estimation methods based on register information on farms, growth region, local rainfall and weather statistics at more aggregate level, but for the rare production branches, we plan to study mixed method of expert opinion based on donor imputation.

**5. Conclusions**

We have shown that the survey costs can be reduced in unefficient sample designs of horticultural survey by increasing the survey threshold. By increasing the cut-off threshold from 2 000 euros of Standard Output up to 10 000 euros, we have been able to reduce the number of farms in the sample by 26% leading to the reduction of 756 enterprises from the survey population. The agricultural area for horticultural production was reduced only by 2% i.e. 441 hectares. The reduction was highest amongst open field horticultural farms in which the economic value per hectare is lower than in greenhouses. We estimate that the direct survey costs have been reduced in total by some 15 000 euros per survey. The outcome of the savings in survey costs we have gained, are higher as we originally anticipated: the reduction or survey costs between 4 000 euros and 8 000 euros and reduction with response burden by 23%. Thus we have been able to balance between the reduction of survey costs and maintaining reasonable survey quality and required quality according to the statistical regulations.

**6. References**

Eurostat (2015), Statistics explained. Glossary: Standard output (SO).
http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Standard\_output\_(SO)

Natural Resources Institute Finland (2015), Horticultural Statistics 2014 (final statistics) and Energy Consumption in Greenhouse Enterprises.

http://stat.luke.fi/en/horticultural-statistics

Regulation (EC) No 543/2009 of the European Parliament and of the Council of 18 June 2009 concerning crop statistics and repealing Council Regulations (EEC) No 837/90 and (EEC) No 959/93 (1)

Regulation (EU) No 1337/2011 of the European Parliament and of the Council of 13 December 2011 concerning European statistics on permanent crops and repealing Council Regulation (EEC) No 357/79 and Directive 2001/109/EC of the European Parliament and of the Council.

Eurostat (2017), Agriculture, forestry and fishery statistics 2017 edition
<http://ec.europa.eu/eurostat/documents/3217494/8538823/KS-FK-17-001-EN-N.pdf/c7957b31-be5c-4260-8f61-988b9c7f2316>

Eurostat (2018), Annual crop statistics Handbook, 2018 Edition, <http://ec.europa.eu/eurostat/cache/metadata/Annexes/apro_cp_esms_an1.pdf>

Luke Statistical WebServices on Agricultral, Forestry , Fishery and Bioeconomy statistics: <http://stat.luke.fi/en/>