Sample of Dutch FADN 2007

Design principles and quality of the sample of agricultural and horticultural holdings

LEI
WAGENINGEN UR

Sample of Dutch FADN 2007

Design principles and quality of the sample of agricultural and horticultural holdings

H.C.J. Vrolijk
H.B. van der Veen
J.P.M. van Dijk

Report 2009-067
August 2009
Project number 31560
LEI Wageningen UR, The Hague

LEI Wageningen UR conducts research in the following areas:

- International policy
- Development issues
- Consumers and supply chains
- Sectors and enterprises
- Environment, nature and landscape
- Rural economy and use of space

This report is part of the research area Sector and enterprises.

This publication has been produced on the basis of activities included within the statutory research tasks. The execution of these tasks is coordinated by the Centrum voor Economische Informatievoorziening (Centre for Economic Information, CEI). The CEI monitors the independent character of the execution and the quality assurance.

Photo's: Hans Vrolijk and Shutterstock

Sample of Dutch FADN 2007; Design principles and quality of the sample of agricultural and horticultural holdings

H.C.J. Vrolijk, H.B. van der Veen and J.P.M. van Dijk

Report 2009-067
ISBN/EAN: 978-90-8615-355-8
Price € 18,50 (including 6% VAT); 73 p., fig., tab., app.

The EU Farm Accountancy Data Network (FADN) requires the Netherlands to yearly send bookkeeping data of 1,500 farms to Brussels. This task is carried out by LEI and CEI. The data sent to Brussels mainly involves technical and financial economic information. For national policy purposes additional data is collected, such as pesticide use, manure production, nature management, non-farm income and rural development. This report explains the background of the farm sample for the year 2007. The report mainly focuses on the Dutch contribution to the European Farm Accountancy Data Network. All phases from the determination of the selection plan, the recruitment of farms to the quality control of the final sample are described in this report.

Mede voor de Europese Unie organiseren het CEI en het LEI jaarlijks de verzameling van technische en financieel-economische gegevens van circa 1.500 bedrijven in de akkerbouw, tuinbouw en veehouderij. Voor nationaal beleidsgericht onderzoek wordt die informatie aangevuld met gegevens over bijvoorbeeld milieubelasting, natuurbeheer en plattelandsontwikkeling. Alle gegevens worden vastgelegd in het Bedrijven-Informatienet. In dit rapport wordt verantwoording afgelegd over de steekproef 2007, toegespitst op de Nederlandse bijdrage aan het Farm Accountancy Data Network van de Europese Unie. De diverse fasen, van het opstellen van het selectieplan, het werven van de bedrijven tot het beoordelen van de kwaliteit van de resulterende steekproef worden beschreven.

Orders

+31.70-3358330/publicatie.lei@wur.nl

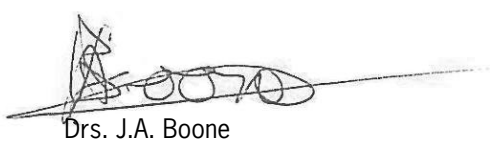
LEI is ISO 9000 certified.

Contents

Preface

The EU Farm Accountancy Data Network (FADN) requires the Netherlands to yearly send bookkeeping data of 1,500 farms to Brussels. This task is carried out by LEI and CEI. This report explains the background of the sample for the year 2007. All phases from the determination of the selection plan, the recruitment of farms to the quality control of the final sample are described in this report. This report provides essential background information for the European Commission the Dutch Ministry and researchers of LEI Wageningen UR and other organisations to fully understand the statistical aspects of the Dutch FADN sample.

Prof Dr R.B.M. Huirne
Director General LEI Wageningen UR



Drs. J.A. Boone
Head CEI

Summary

Introduction
The EU Farm Accountancy Data Network (FADN) requires the Netherlands to yearly send bookkeeping data for 1,500 farms to Brussels. This task is carried out by the Agricultural Economics Research Institute (LEI) and the Center for Economic Information (CEI). The legislation of the FADN demands that the member states prepare a selection plan and a report on the results of the selection. This report fulfils this obligation. Furthermore, the report gives an analysis of the quality of the sample.

Population and Selection plan 2007
The population (field of survey) of the FADN is defined as all farms above the threshold of 16 European Size Units (ESU). In the Netherlands farms between 16 and 2,000 ESU are included in the population (Table 3.1). A stratified random sample is drawn, in which economic farm size and type of farming are used as stratification variables. The scheme for the types of farming is based on a Dutch version of the Common Agricultural Typology that is also used by EUROSTAT. The total agricultural population contains 76,741 farms according to the agricultural census. The field of survey contains 58,787 farms. These farms cover an important part (92.8%) of the production capacity (Table 3.1). In the selection plan, LEI planned to select 1,500 farms for the 2007 accounting year.

Result of recruitment and quality of the 2007 sample
For 2007, 1,510 farms were included in the sample and were delivered to Brussels (Table 5.6). Chapter 6 gives a quantitative evaluation of the resulting sample. A comparison of the field of survey with the total agricultural population shows that 23% of the farms are below the lower threshold. These farms are responsible for a small percentage of production only. The sample results in a coverage of 90% of the production for most of the agricultural activities. In horticulture, part of the production is not covered because it takes place on farms above the upper threshold. Therefore the upper threshold has been increased to 2,000 ESU. This increase has been introduced as a trial in 2006 and has been integrated in the selection plan starting from the year 2007. Despite this increase there are still 166 firms larger than 2,000 ESU. Table 6.2 gives a description of the coverage of a large number of activities. Table 6.3 shows the relationship between types of farming and agricultural activities. The numbers

show that only a limited percentage of pigs is produced on specialised pig farms, while at the other extreme almost all mushrooms are produced on specialised mushroom farms. Two important aspects of a sample, the representativeness of the sample and the reliability of estimates, are evaluated in section 6.3.3 and 6.3.4. Table 6.4 evaluates for many variables whether there is a difference between the agricultural census and the estimate based on the FADN sample. These tables provide useful information for specific research projects enabling the researcher to determine whether the sample is representative for his or her topic.

In summary, with the delivery of 1,510 farm accounts to the European Commission for the accounting year 2007, the EU requirement has been fulfilled. This sample of farms provides a useful dataset for policy analysis of the Common Agricultural Policy (CAP) and the monitoring of the economic developments in the farming sector.

Samenvatting

Steekproef van het Nederlandse FADN 2007; Ontwerpprincipes en kwaliteit van de land- en tuinbouwbedrijven in de steekproef

Inleiding

Mede voor de Europese Unie organiseren het CEI en het LEI jaarlijks de verzameling van technische en financieel-economische gegevens van circa 1.500 bedrijven in de akkerbouw, tuinbouw en veehouderij. Voor nationaal beleidsgericht onderzoek wordt die informatie aangevuld met gegevens over bijvoorbeeld milieubelasting, natuurbeheer en plattelandsontwikkeling. Alle gegevens worden vastgelegd in het Bedrijven-Informatienet. In dit rapport wordt verantwoording afgelegd over de steekproef 2007, toegespitst op de Nederlandse bijdrage aan het Farm Accountancy Data Network van de Europese Unie. De diverse fasen, van het opstellen van het selectieplan en het werven van de bedrijven tot het beoordelen van de kwaliteit van de resulterende steekproef, worden beschreven.

Populatie en selectieplan 2007

De onderzoekspopulatie van het Bedrijven-Informatienet is gedefinieerd als alle bedrijven groter dan 16 Europese grootte-eenheden (ege) en kleiner dan 2.000 ege (tabel 3.1). Uit het steekproefkader (alle bedrijven in de landbouwtelling tussen 16 en 2.000 ege) wordt een gestratificeerde random steekproef getrokken. Economische omvang en het type bedrijf worden gebruikt als stratificatievariabelen. Voor het jaar 2007 omvat de totale agrarische populatie 76.741 bedrijven (opgenomen in de landbouwtelling). Het steekproefkader omvat 58.787 bedrijven. Deze bedrijven zijn verantwoordelijk voor 92,8% van de totale productiecapaciteit (tabel 3.1). Het selectieplan 2007 is op een aantal punten gewijzigd ten opzichte van de jaren daarvoor als gevolg van het verhogen van de bovengrens. Het selectieplan geeft aan dat er 1.500 bedrijven in administratie dienen te worden genomen.

Resultaat van de werving en kwaliteit van de resulterende steekproef 2007

Voor het jaar 2007 zijn 1.510 bedrijven uitgewerkt en aangeleverd aan Brussel (tabel 5.8). Hoofdstuk 6 geeft een kwantitatieve evaluatie van de resulterende steekproef. Een vergelijking tussen de onderzoekspopulatie en de totale agrari-

sche populatie zoals beschreven in de landbouwtelling laat zien dat 23% van de bedrijven zich beneden de benedengrens bevinden. Deze bedrijven zijn echter verantwoordelijk voor slechts een klein deel van de totale productie. De onderzoekspopulatie dekt circa 90% van de productie van de meeste agrarische activiteiten. In de tuinbouw ligt het probleem bij de grotere bedrijven. Om dit probleem voor de toekomst te verminderen is de bovengrens van de steekproef opgetrokken naar 2.000 ege. Deze verhoging is in 2006 op proefbasis ingevoerd en in 2007 definitief doorgevoerd in de steekproefopzet. Er zijn nog 166 bedrijven die boven deze grens vallen. Tabel 6.2 geeft een nadere uitwerking van de dekking voor een groot aantal activiteiten. Tabel 6.3 geeft de samenhang weer tussen typen en agrarische activiteiten. Uit de tabel blijken grote verschillen in de mate van specialisatie van activiteiten. Slechts een beperkt percentage van alle vleesvarkens wordt geproduceerd op gespecialiseerde vleesvarkensbedrijven. Aan de andere kant geldt dat bijna alle paddenstoelen worden geproduceerd door gespecialiseerde paddenstoelbedrijven. Twee belangrijke aspecten van steekproeven, de representativiteit en de betrouwbaarheid van schattingen worden geëvalueerd in paragraaf 6.3.3 en 6.3.4. Tabel 6.4 geeft voor een groot aantal variabelen een vergelijking tussen de waarde volgens de landbouwtelling en de schatting op basis van het Bedrijven-Informatienet. Deze informatie stelt de onderzoeker in staat om te beoordelen in hoeverre de steekproef representatief is voor zijn of haar specifieke onderzoeksproject.

Samenvattend, met de levering van 1.510 bedrijven aan de Europese Commissie voor het boekhoudjaar 2007 heeft Nederland voldaan aan de EU-verplichting. Deze dataset is van grote waarde voor de evaluatie van het Gemeenschappelijke Landbouwbeleid (GLB) en het monitoren van de financieel-economische situatie in de agrarische sector.

1 Introduction

1.1 Objective of the report

In 1965 the European Commission adopted a regulation (no. 79/65/EEG) in which member states were obliged to set up a network for the collection of accountancy data on the incomes and business operation of agricultural holdings in the European Economic Community. The purpose of the data network is defined as the annual determination of incomes on agricultural holdings, and a business analysis of agricultural holdings. The Netherlands were required to provide financial economic information on 1,500 farms to Brussels.

For the management of the system, the EU requires information on the selection of farms that are included in the national FADN systems. In particular the regulation prescribes the provision of data on the establishment of a selection plan and the recruitment of farms.

With respect to the selection plan the regulation EEG 1859/82 prescribes (article 6):

'Each Member State shall appoint a liaison agency whose duties shall be: […] to draw up and submit to the National Committee for its approval, and thereafter to forward to the Commission:

- the plan for the selection of returning holdings, which plan shall be drawn up on the basis of the most recent statistical data, presented in accordance with the Community typology of agricultural holdings.'

This report provides all the relevant background information on the population, the selection plan, implementation of the selection plan and quality of the sample of data that has to be provided to Brussels and which forms the basis for a wide range of national and international research projects.

1.2 Structure of the report

Chapter 2 gives a description of the background of the Dutch FADN system. Chapter 3 describes the agricultural population in the year 2007. This chapter will also consider the demarcation of the population as used in the Dutch FADN. Also the design of the sample of the Dutch FADN system is described. Chapter 4 reports on the selection plan 2007. Chapter 5 provides information on the implementation of the selection plan and the recruitment of new farms. Chapter 6 provides a qualitative and quantitative evaluation of the 2007 sample.

2 Statistical background of the Dutch FADN sample

2.1 Introduction

In the Dutch FADN detailed records on 1,500 agricultural and horticultural farms are kept. Besides financial-economic information, a broad set of technical-economic, socio-economic and environmental-economic data is collected. One of the reasons for the Dutch FADN system is the legal obligation to provide information on the financial-economic situation of farms to Brussels. However, an even more important use of the data can be found at the national level. Data from the FADN system is used for many national policy evaluations and research projects.

Based on a sample of farms estimations are made for the whole population. This might raise the question how conclusions can be drawn for the whole population if only a limited number of farms are observed. The answer to this question can be found in the selection of farms that are included in the sample. A cook also doesn't eat all the soup to judge the quality of the soup. It is important to stir well before tasting; the spoon of soup should reflect all flavours in the pan of soup. The spoon of soup should be representative on the whole pan of soup. The same is true for the FADN sample. The farms that are included in the FADN should be representative of the whole population. In this way a sample can provide better information than a census (in which all units are observed). With a fixed budget it is much easier to collect good data on a limited number of farms instead of collecting information on all farms. With a limited number of farms and thus a limited number of data collectors, it is easier to ensure good procedures and good training to collect reliable data.

An important issue is how to ensure that the farms that are included in the FADN sample are representative of the whole population. Use is made of a disproportional stratified random sample. A stratified sample implies that the population is divided into a number of groups. Subsequently farms are selected from each of the groups. The variables on which the groups are defined should be relevant variables to make sure that the farms that are included in one group are similar (at least with respect to the important aspects). Using this stratification, and selecting farms from each group, ensures that farms from all groups and consequently with different characteristics are included in the sample.

Disproportional means that not all farms have the same chance of being included in the sample. Groups which are relatively homogeneous, i.e. farms which show large similarities, have a lower chance of being included in the sample. After all, if all the farms are very similar, a limited number of observations is enough to draw reliable conclusions (in the extreme case that all farms are exactly identical, it would be enough to have only one observation). In case of less homogeneous groups it is important to have a larger number of observations to make reliable estimates. The choice of the stratification variables has therefore an important impact on the quality of the sample.

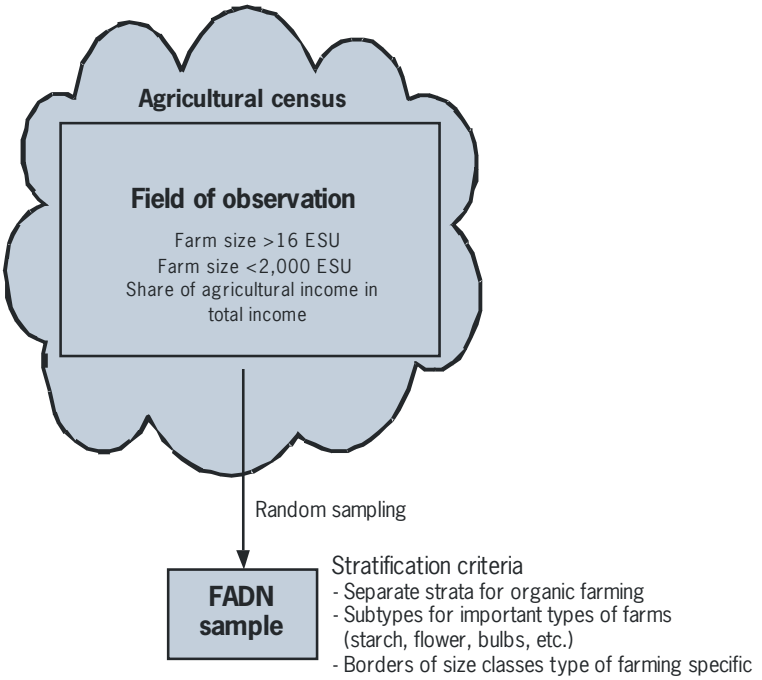
This way of selecting farms makes it possible to make unbiased estimates for the whole population of farms. Stratification assures that farms are selected from all groups, thereby allowing estimations for all groups. All groups together make up the whole population. In the FADN this is achieved by assigning a weight to each sample farm. The weight is calculated by dividing the number of population farms in a group by the number of sample farms in the same group.

Stratification also improves the representativeness in case of non-response. If a farm which is asked to join the FADN system refuses, another farm in the same size class and of the same type of farming can be selected. If there is a difference between the selection plan and the actual implementation, stratification helps to improve the representativeness by taking into account the real sampling fraction.

Finally, stratification makes the maintenance of the sample easier. Due to attrition and changes in the population it is sometimes necessary to supplement certain groups. Stratification makes a more focused replacement possible.

The relationship between the agricultural population and the FADN sample is presented in Figure 2.1. The agricultural census provides an almost complete description of the agricultural population. Part of this census or part of this population is defined as the field of observation in the FADN. In the definition of the field of observation a lower threshold and an upper threshold are applied. Furthermore, an additional criterion on the share of agricultural income in total income is used. These criteria will be further discussed.

Figure 2.1 **Agricultural population and the FADN sample**



Lower threshold

The lower threshold of 16 ESU has been used for a long period of time. It is specified in the legislation underlying the FADN. The historical background was to distinguish small farms which were only held as a hobby or as side activity from real commercial farms producing for the market. Although the number of farms excluded from the field of survey is quite substantial, the percentage of production value which is not covered because this threshold is very limited.

Upper threshold

The upper threshold was introduced to exclude some non-agricultural organisations from the field of observation. The agricultural census contains some organisations with a lot of land but which are not considered as agricultural holdings (examples are airports, nature organisations and in earlier days organisations which managed the reclamation of land from water bodies). Furthermore the inclusion of these very large farms would result in a substantial decrease in

the reliability of estimates due to the large heterogeneity of these farms. Another practical reason to exclude the large farms is the complexity and size of the bookkeeping and therefore the large demand for limited human resources.

In order not to judge each individual holding, an upper threshold was introduced to exclude these from the field of survey. Due to the growth in size of farming in especially horticulture it was decided to increase the upper threshold to fulfil the requirement to cover at least 90% of the agricultural productivity.

At the current moment a project is being undertaken to assess whether farms above the threshold can be included in the sample in the future. Issues to be addressed will be: are large farms willing to cooperate, how can they be motivated, is the farm comparison report useful for them, how much resources will it take to administer these farms etc. Based on the results of this project a decision will be made whether the upper limit will be maintained in the future.

Other income sources

For practical and methodological reasons a limitation on 'other income of the holding' is used. In earlier times the rules were not clearly specified. Firms with a high share of other income sources were excluded from the sample because of practical reasons such as the impossibility to allocate costs and revenues to different activities, firms would refuse to participate anyway because they cannot be motivated to participate etc. Recently clear rules have been specified whether a firm belongs to the field of observation or not. A firm should have at least 16 ESU from primary agricultural activities, at least 25% of the turnover should come from primary agricultural activities and agricultural activities - in the broadest sense, so as to include other gainful activities - should be the largest share of turnover of the holding.

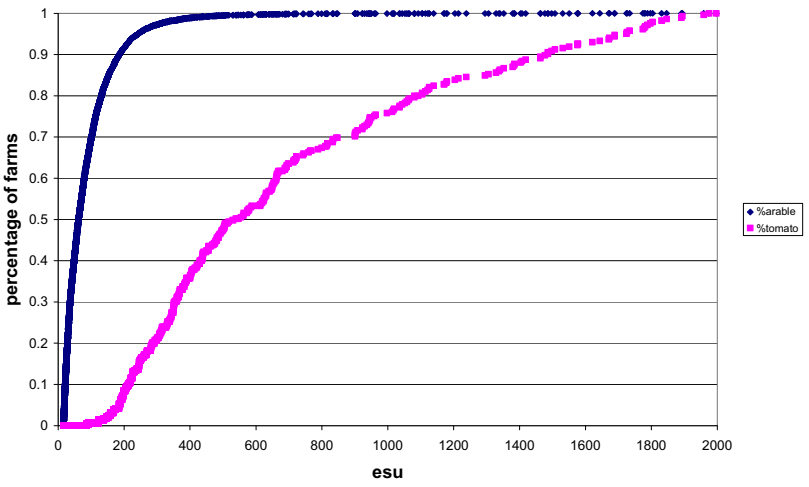
Stratification criteria

Given these three criteria the field of observation of the FADN system is defined. Within this field of observation a stratification scheme is used. The stratification of the Dutch FADN is based on size of farming and type of farming. Although these criteria are similar to those used by the commission, a more detailed look reveals substantial differences with the EU stratification. Differences are for example the use of separate strata for organic farming, and in several types of farming more detailed subtypes of farming are specified which are relevant for Dutch Agriculture (for example starch potato farms, flower bulb farms, horticultural farms by type of production).

The Dutch situation is somewhat more complicated due to the fact that the size classes are different within different types of farming. The size distribution

of, for example, horticultural farms is completely different than the size distribution of arable farms. This is illustrated in Figure 2.2. This figure illustrates that almost all arable farms are smaller than 400 ESU, almost 70% of the tomato growers are larger than 400 ESU. To take these differences into account the borders of the size classes have been established for each type of farming separately. Despite this complication the strata are still a cross section between types of farming and size-classes. In total 98 strata have been defined.

Figure 2.2 **Distribution of farms (arable and tomatoes)**

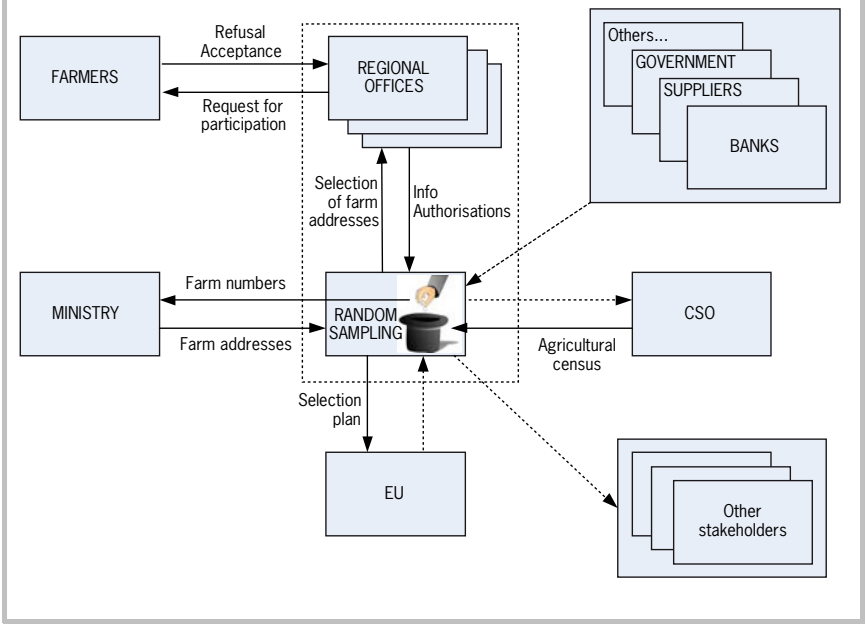


2.2 Sampling and recruitment processes

Figure 2.3 presents an overview of the sampling and recruitment processes. The agricultural census from Statistics Netherlands (CBS) is the starting point for the random sampling of farms. The random sampling takes place based on the selection plan as submitted to the European Commission. The selection plan will be further described in chapter 4. Based on the selection plan farms from the agricultural census are randomly drawn. This census (as available to researchers) does not contain addresses but only farm identifiers. These farm identifiers are sent to the ministry and the ministry then returns the addresses.

These addresses are forwarded to the regional offices who are responsible for contacting farmers to request their participation. The farmers either refuse or accept the request to participate; this recruitment process and the non-response will be described in chapter 5. The regional offices collect the authorisations and forward them to the central office in The Hague. These authorisations are used to receive electronically available information from banks, suppliers, governmental institutions and others. The information on the acceptance and refusal of farmers is also used to verify the quality of the sample (see chapter 6).

Figure 2.3 **Sampling and recruitment processes**



3 2007 population

3.1 Introduction

This chapter will describe the population or, more precisely, the field of observation as covered by the FADN sample. A lower threshold is used to define the field of observation. This threshold and the consequences of this threshold will be described in section 3.2. Section 3.3 describes the strata which are used to divide the population. Section 3.4 reports the number of farms in each of the strata.

3.2 Defining the field of observation

Collecting detailed information at farm level requires considerable time and money. To assure an efficient and effective allocation of the available budget, the sample design focuses on certain groups in the population. Given the limited capacity it is important to apply a sampling procedure that optimises the reliability of the sample estimates (through stratification).

Regulation 1859/82 of the EU Commission (adapted by regulation EEG no. 3548/85) defines the population (field of observation) for the Dutch FADN as those farms with a size of more than 16 European size units (ESU). Until 2001 this threshold was translated into 16 Dutch size units (DSU), which is roughly similar to 18.7 ESU. For the statistical use of the data and the comparability of results it was considered advisable to apply the ESU threshold. Therefore the lower limit of the Dutch FADN system has been 16 ESU since the year 2001.

In addition to a lower threshold there is also an upper threshold. This upper threshold has been adjusted every few years to take into account the growth of the average size of farms. Until 2001 the upper threshold was 800 DSU. In 2001 the upper threshold was raised to 1,200 ESU. The percentage of farms and the agricultural output excluded due to this upper threshold has been growing since 2001. For this reason the upper threshold has been increased again to 2,000 ESU.

Table 3.1	Number of farms and their relative economic importance (measured in European size units - ESU) in the 2007 agricultural census	
	Number of farms	**Percentage ESU**
All farms in the agricultural census (a)	76,741	100
minus farms smaller than 16 ESU	17,788	1.83
minus farms larger than 2,000 ESU	166	6.37
Total of non-covered farms (b)	17,954-	8.21-
Total of covered farms (a) - (b)	58,787	92.79

The increase has been introduced on a trial basis in 2006 and has been integrated in the sample and weighting scheme starting from the year 2007. In this report all analyses presented are based on the upper threshold of 2,000 ESU. In 2007, 166 farms were excluded from the field of observation because of the upper threshold of 2,000 ESU. These farms were responsible for 6.37% of the total production. Due to the lower threshold 17,788 farms were not covered by the FADN sample. Although this is a large number of farms, they are only responsible for 1.83% of the total production capacity. The number of farms and the share of economic production of these small farms show a continuous decrease compared to 2005 and 2006. The population (field of observation) of the Dutch contribution to the EU FADN system is displayed in Table 3.1.

3.3 Design of the stratification scheme

Farms are allocated to strata according to the following stratification variables: type of farming and size class. In the past, a more detailed stratification scheme was used, but this resulted in numerous practical problems due to empty or nearly empty cells. Combining cells can easily lead to a distortion in the calculated results (a bias). Farms of a certain type of farming are divided into 3 or 4 size classes. In the last years 3 size classes were used. The increase from 3 to 4 size classes is caused by the increase of the upper threshold. Increasing the upper threshold without replacing the whole sample in the largest size class results in distorted samples.

In total 29 types of farming are distinguished (see Table 3.2). For a number of types of farming a distinction is made between organic farming and non-organic farming. A compromise was found to fulfil the increasing demand for research on organic farms. Random selection of organic farms from the total

population would result in a very low number of observations because of the low proportion of organic farms. The definition of separate strata would result in many practical problems. The number of strata would double. The problem of empty or nearly empty strata would increase seriously. In line with the existing stratification, a number of types of farming were selected where organic farming is especially relevant. The types that were originally selected were: field crop farms, dairy farms, field vegetables and combined crop farms (Vrolijk and Lodder, 2002). The growth in the organic sector was however lower than expected and aimed for by policy makers. This resulted in practical problems in the recruitment of organic farms, for example due to the fact that the number of farms according to the selection plan was close to or even higher than the actual number of farms in the population. To deal with this problem a number of organic strata have been combined. 'Organic field crops farms', 'field vegetables' and 'combined crop farms' have been integrated in one stratum 'organic crop farms' (Vrolijk, 2006).

The breakdown in subtypes is as follows: 'field crop farms' have been itemised in 'starch potato farms', 'organic crops' and all 'other field crop farms'. The 'vegetables under glass' farms have been broken down in 'paprika', 'cucumber', 'tomato' and 'other'. 'Cut flowers under glass' are divided into 'roses', 'chrysanthemums' and 'other cut flowers'. The dairy farms are split into organic and non-organic dairy farms. Within 'field vegetables' and the 'combined crop farms' the organic farms have been separated. These are subsequently combined with the organic field crop farms.

The final stratification and the size thresholds for each of the strata are displayed in Table 3.2. The size classes 1 and 2 have remained the same, size class 3 has been slightly redefined due to the introduction of the 4th size class in certain types of farming. This redefinition has been based on practical criteria. A more thorough analysis of the definition of the strata will be made in the transition of standard gross margins to standard outputs as a base for the definition of farm size and types of farming.

Table 3.2	Stratification of the Dutch FADN sample			
Type of farming	**Size class**			
	1	**2**	**3**	**4**
Field crop farms				
- Starch potatoes	16-66	66-140	140-2,000	
- Organic crops	16-45	45-90	90-2,000	
- Other field crop farms	16-73	73-178	178-2,000	
Horticulture				
Vegetables under glass				
- Paprika	16-245	245-480	480-900	900-2,000
- Cucumber	16-201	201-393	393-800	800-2,000
- Tomato	16-269	269-518	518-1,100	1,100-2,000
- Other	16-106	106-336	336-600	600-2,000
Cut flowers under glass				
- Rose	16-260	260-495	495-900	900-2,000
- Chrysanthemum	16-194	194-373	373-750	750-2,000
- Other	16-142	142-342	342-550	550-2,000
Plants	16-186	186-464	464-850	850-2,000
Other glass	16-108	108-292	292-500	500-2,000
Field vegetables	16-86	86-257	257-2,000	
Fruit	16-64	64-139	139-2,000	
Nurseries	16-85	85-251	251-2,000	
Mushroom	16-188	188-445	445-900	900-2,000
Bulbs	16-185	185-477	477-900	900-2,000
Other open air	16-116	116-356	356-2,000	
Grazing livestock				
Dairy				
- Organic	16-86	86-128	128-2,000	
- Non-organic	16-89	89-159	159-2,000	
Calf fattening	16-64	64-150	150-2,000	
Other grazing livestock	16-47	47-146	146-2,000	
Intensive livestock				
Breeding pigs	16-116	116-263	263-2,000	
Fattening pigs	16-60	60-161	161-2,000	
Integrated pig farms	16-129	129 -253	253-2,000	
Laying hens	16-138	138-345	345-2,000	

Table 3.2	Stratification of the Dutch FADN sample (continued)			
Type of farming	**Size class**			
	1	**2**	**3**	**4**
Poultry	16-100	100-203	203-2,000	
Other intensive livestock	16-113	113-261	261-2,000	
Combined	16-81	81-206	206-2,000	

3.4 Number of farms in the 2007 population

Table 3.3 presents the number of farms in the population (2007 agricultural census). In this table the stratification according to size class and type of farming is applied as described in Table 3.2.

Table 3.3	Number of farms per stratum according to the 2007 agricultural census				
Type of farming	**Size class**				
	1	**2**	**3**	**4**	**total**
Field crop farms					
- Starch potatoes	439	385	186		1,010
- Organic crops	65	70	96		231
- Other field crop farms	3,779	2,134	734		6,647
Horticulture					
Vegetables under glass					
- Paprika	69	121	89	51	330
- Cucumber	52	105	67	17	241
- Tomato	51	90	87	57	285
- Other	399	249	57	40	745
Cut flowers under glass					
- Rose	74	97	90	53	314
- Chrysanthemum	64	55	66	19	204
- Other	788	648	219	154	1,809
Plants	451	321	174	118	1,064
Other glass	266	179	79	64	588
Field vegetables	430	284	123		837

Table 3.3	Number of farms per stratum according to the 2007 agricultural census (continued)				
Type of farming	**Size class**				
	1	**2**	**3**	**4**	**total**
Fruit	620	617	262		1,499
Nurseries	962	713	362		2,037
Mushroom	121	58	28	18	225
Bulbs	405	276	124	64	869
Other open air	687	398	116		1,201
Grazing livestock					
Dairy					
- Organic	125	97	75		297
- Non-organic	6,275	8,902	3,272		18,449
Calf fattening	383	514	235		1,132
Other grazing livestock	5,485	2,133	282		7,900
Intensive livestock					
Breeding pigs	764	517	135		1,416
Fattening pigs	875	497	155		1,527
Integrated pig farms	428	414	153		995
Laying hens	632	282	44		958
Poultry	197	172	65		434
Other intensive livestock	94	53	19		166
Combined	2,874	1,765	738		5,377
Total	27,854	22,146	8,132	655	58,787

This table shows that 58,787 farms fall within the field of observation. Dairy farms are clearly the largest group of farms. Almost one in every three farms is classified as a dairy farm.

4 2007 selection plan

4.1 Introduction

The allocation of the total capacity of sample farms is based on the relative importance and the heterogeneity of the different types of farming (see Dijk et al., 1995a and Vrolijk and Lodder, 2002). Within each type of farming an optimal stratification (determination of thresholds of size classes) and optimal allocation has been applied (distribution of sample capacity over the different size classes). In the introduction of the 4th size class practical considerations such as the availability of sampling farms have been considered.

4.2 2007 Selection plan

The design principles of the sample of the FADN system facilitate an efficient alignment with the goals of the system (see chapter 2). A summary of the 2007 selection plan is provided in Table 4.1. Given the goals of the FADN system the numbers provided in the table are the required number of observations per type of farming.

For the sample of 2007 the changes are related to the increase of the upper threshold. To anticipate this development in 2006 a start has been made to increase the number of sample farms in those farm types that have at least more than a few farms above 1,200 ESU. This concerns the glasshouses (vegetables as well as flowers), plant growers, mushroom growers and bulb growers. These are the types of farms where the share of production above the upper limit increased substantially during the last years.

Table 4.1	Desired sampling size per type of farming (selection plan) 2007			
Type of farming	**Code**	**Number of farms**		
		main type	**Type**	**sub type**
Field crop farms	*1*	*210*		
- Starch potatoes			30	
- Organic crops			30	
- Other field crop farms			150	
Horticulture	*2 + 3*	*538*		
Vegetables under glass	2012		134	
- Paprika				34
- Cucumber				33
- Tomato				34
- Other				33
Cut flowers under glass	2022		116	
- Rose				30
- Chrysanthemum				30
- Other				56
Plants	2022		44	
Other glass	other 2022 and 2013, 2023, 2039, 349 (>50% glass)		30	
Field vegetables	2011		30	
Fruit	3210		40	
Nurseries	3480		40	
Mushroom	2033		32	
Bulbs	2021		42	
Other open air	other 2022 and 2013, 2023, 2039, 349 (<50% glass)		30	
Grazing livestock		*410*		
Dairy	4110, 4120, 4370		330	
- Non-organic				300
- Organic				30
Calf fattening	4380		30	
Other grazing livestock	4410, 4420, 4430		50	
Intensive livestock	*5*	222		
Breeding pigs	5011		50	

Table 4.1	Desired sampling size per type of farming (selection plan) 2007 (continued)			
Type of farming	**Code**	**Number of farms**		
		main type	**Type**	**sub type**
Fattening pigs	5012		50	
Integrated pig farms	5013		40	
Laying hens	5021		34	
Poultry	5022		30	
Other intensive livestock	other 5		18	
Combined	*6,7 and 8*	*120*		
Total		*1,500*		

5 Recruitment of farms in 2007

5.1 Basic principles for 2007

The recruitment for 2007 took at the start of the year. In February an assessment was made of the farms available for the FADN system (considering farms dropping out of the system). From April till June the farms were recruited. This selection of farms was based on the 2006 agricultural census. At the end of 2007 a recruitment took place for the year 2008. Some of these farms were also used for the sample of 2007.

5.2 Elaboration of selection plan

Table 5.1 gives a more detailed description of the selection plan as presented in Table 4.1.

Table 5.1 Detailed 2007 selection plan per stratum

Type of farming	ESU size class				
	1	**2**	**3**	**4**	**total**
Field crop farms					
- Starch potatoes	10	10	10		30
- Organic crops	10	10	10		30
- Other field crop farms	45	51	54		150
Horticulture					
Vegetables under glass					
- Paprika	9	13	8	4	34
- Cucumber	9	13	6	5	33
- Tomato	9	9	8	8	34
- Other	10	10	8	5	33
Cut flowers under glass					
- Rose	10	8	8	4	30

Table 5.1	Detailed 2007 selection plan per stratum (continued)				
Type of farming	**ESU size class**				
	1	**2**	**3**	**4**	**total**
- Chrysanthemum	10	8	8	4	30
- Other	17	18	13	8	56
Plants	12	13	13	6	44
Other glass	10	10	6	4	30
Field vegetables	10	10	10		30
Fruit	12	14	14		40
Nurseries	13	13	14		40
Mushroom	10	10	8	4	32
Bulbs	13	13	11	5	42
Other open air	10	10	10		30
Grazing livestock					
Dairy					
- Organic	10	10	10		30
- Non-organic	100	100	100		300
Calf fattening	10	10	10		30
Other grazing livestock	17	16	17		50
Intensive livestock					
Breeding pigs	20	16	14		50
Fattening pigs	16	16	18		50
Integrated pig farms	14	12	14		40
Laying hens	12	12	10		34
Poultry	10	10	10		30
Other intensive livestock	6	6	6		18
Combined	37	41	42		120
Total					1,500

5.3 Recruitment of farms

Based on the available number of farms in the FADN sample and the expected number of farms ending their participation before or during 2007 an estimate was made of the number of farms to be recruited. Furthermore, the variant of bookkeeping has been explicitly considered. A distinction is made between CSP observations (corporate social performance) and the total number of observa-

tions. Poppe (2004) describes that the introduction of a new bookkeeping system and budget cuts resulted in a large pressure on available capacity. To deal with this pressure, a flexible data collection system has been introduced with two main variants in the data collection: the EU variant and the CSP variant. In the EU farm-income variant the most essential financial economic information is collected. This is the information that each member state is obliged to provide to Brussels. The information covered in this variant mainly focuses on family farm income, the balance sheet, a limited number of technical data (cropping pattern, livestock) and information on the EU subsidies. In the second variant, the CSP variant, a wide range of data is collected for EU and national purposes. It covers all the topics that are nowadays considered relevant in a report on the corporate social performance of a company or a farm. Therefore, besides the financial-economic information as collected in the EU variant, a wide range of data is collected such as environmental data, other farm incomes, off-farm income, animal welfare, animal health and the level of innovation of firms.

An evaluation has been made of the policy and research relevance of sectors and based on this importance a decision has been made whether a type of farming is assigned to the EU variant, the CSP variant or a combination of both. This implied that some farms had to be switched to the other variant. Based on the number of farms to be recruited, farms were randomly selected from the 2006 agricultural census. The random draw of farms took place per stratum. The number of farms drawn per stratum was 7 times higher than the required number of farms to ensure enough addresses, even with a high non-response rate in specific types of farming. The addresses were requested from an agency (Dienst Regelingen) of the Ministry of Agriculture. The farm identifiers of the randomly selected farms were sent to the Ministry which sent back the addresses of these farms (under the strict condition that this information was only used for the recruitment of farms for the FADN). Using these addresses farms were contacted and asked to participate in the FADN.

Farms are asked to participate in the system in order to compensate for attrition and to take structural changes in agriculture into account. Some of the farms approached during the recruitment phase refused to participate. These refusals do not cause problems if these farms do not differ from farms that participate in their place. In the case where farms that refuse to participate systematically differ from the participating farms, this could result in a bias. If for example older farmers are less inclined to participate, this will result in a different age distribution in the sample compared to the population. The representativeness of the data with respect to age will be called into question - whether this is a problem or not depends on the research goals and the extent to which the important variables cor-

relate with age. The representativeness is analysed in chapter 6. Table 5.2 describes the response rate in the different types of farming. This table only includes those farms which were asked to participate in the CSP variant.

Table 5.2 Response rate in different types of farming, recruitment for CSP variant

	Refusals	Recruited	Unsuitable	Total	Unsuitable %	Response %
Field crop farms						
- Starch potatoes	24	1	4	29	14	4
- Organic crops	0	0	0	0		
- Other field crop farms	42	5	9	56	16	11
Horticulture						
Vegetables under glass						
- Paprika	8	3	9	20	45	27
- Cucumber	5	3	2	10	20	38
- Tomato	9	5	8	22	36	36
- Other	19	2	12	33	36	10
Cut flowers under glass						
- Rose	18	4	8	30	27	18
- Chrysanthemum	10	2	3	15	20	17
- Other	11	2	7	20	35	15
Plants	36	12	11	59	19	25
Other glass	3	0	0			
Nurseries	0	0	1			
Mushroom	0	0	0			
Bulbs	0	1	0			
Other open air	0	0	0	1		
Grazing livestock						
Dairy						
- Organic	8	8	0	16	0	50
- Non-organic	22	14	1	37	3	39
Calf fattening	0	0	0	0		
Other grazing livestock	2	0	0	2		

Table 5.2	Response rate in different types of farming, recruitment for CSP variant (continued)					
	Refusals	**Recruited**	**Unsuitable**	**Total**	**Unsuitable %**	**Response %**
Intensive livestock						
Breeding pigs	17	7	3	27	11	29
Fattening pigs	58	9	11	78	14	13
Integrated pig farms	4	0	0	4		
Laying hens	0	0	0	0		
Poultry	1	0	0	1		
Other intensive livestock	0	0	0	0		
Combined	6	3	1	10	10	33
Total	303	81	90	474		

To develop a better understanding of the reasons for non-response a number of questions were asked to all farmers approached. Table 5.3 shows the results for the questions asked. The farmer had to indicate to which extent he or she agrees with a statement about his knowledge or his attitude. The table shows a clear difference between those farmers who are willing to cooperate and those who are not. Those who are willing to participate are more informed about the activities of LEI and the use of FADN data. Providing data and the FADN system is considered more useful by those who are willing to participate. The opinion about LEI with respect to objectivity and carefulness is better among the participants. The last question shows that non-participants have a significantly lower trust in the government.

Table 5.3	Attitude of farmers (-2 = disagree, 2 = agree)				
	Non participant		Participant		
	average	SE	average	SE	
1 Informed about the LEI	1.11	0.08	1.60	0.07	s
2 Informed about the FADN system	0.42	0.09	0.92	0.16	s
3 Informed about the use of FADN data	0.15	0.08	0.65	0.16	s
4 Usefulness of FADN system	0.19	0.07	1.08	0.13	s
5 Usefulness of providing data	0.13	0.08	1.27	0.13	s
6 Carefulness of LEI	0.45	0.06	1.23	0.13	s
7 Objectivity of LEI	0.39	0.06	1.21	0.11	s
8 Trust in the government	-0.22	0.06	0.12	0.10	s
SE - standard error; s - significant difference, ns - non-significant difference.					

Using these same variables discriminant analysis was applied to find the factors that are most discriminating between farmers who are willing to participate and farmers who refuse to participate. The analyses of the attitude of farmers show that 'usefulness of providing data' is the most important factor in predicting the participation of an individual farmer. The next important factors are 'Objectivity of the LEI'. 'Usefulness of FADN system' and 'Carefulness of LEI'. These results are in line with the previous recruitment (Vrolijk et al., 2008).

Table 5.4 describes the number of farms where accounts were completed for the first time for the bookkeeping year 2007. Due to several factors this is not exactly the same as the number of farms recruited. First, farms can drop out during the first year of participation. Second, some farms were already recruited during a previous year, but due to capacity problems their bookkeeping was not completed for that year.

Table 5.4	Number of farms with 2007 as first year of completion of bookkeeping, recruited for EU or CSP			
Type of farming	ESU size class			
	1	2	3	4
Field crop farms				
- Starch potatoes		1		
- Organic crops		1	1	
- Other field crop farms	7	8	15	

Table 5.4	Number of farms with 2007 as first year of completion of bookkeeping, recruited for EU or CSP (continued)			
Type of farming	**ESU size class**			
	1	**2**	**3**	**4**
Vegetables under glass				
- Paprika	2			
- Cucumber	1	1		
- Tomato	1	1	1	1
- Other	1	1	1	1
Cut flowers under glass				
- Rose	1		1	1
- Chrysanthemum	3			
- Other		1		
Plants	1	5	3	2
Other glass		1	1	
Field vegetables			2	
Fruit				
Nurseries	1	1	3	
Mushroom	1	6		
Bulbs	1			
Other open air		5	1	
Grazing livestock				
Dairy				
- Organic	2		1	
- Non-organic	3	2	6	
Calf fattening	1		1	
Other grazing livestock				
Intensive livestock				
Breeding pigs		4	3	
Fattening pigs	4	2	5	
Integrated pig farms				
Laying hens				
Poultry				
Other intensive livestock		1		
Combined	5	2	3	
Total	35	43	48	5

Tabel 5.5	Comparison of the field of observation (population) and the sample available for research purposes in 2007 (2007 agricultural census)			
Type of farming	**Code**	**Number of farms**		
		population	**total**	**CSP**
Field crop farms	*1*			
- Starch potatoes		1,010	29	26
- Organic crops		231	28	22
- Other field crop farms		6,647	155	139
Horticulture	*2+3*			
Vegetables under glass	2012			
- Paprika		330	32	31
- Cucumber		241	33	30
- Tomato		285	35	31
- Other		745	34	29
Cut flowers under glass	2022			
- Rose		314	29	26
- Chrysanthemum		204	28	28
- Other		1,809	72	54
Plants	2022	1,064	41	38
Other glass		588	29	11
Field vegetables	2011	837	30	6
Fruit	3210	1,499	39	30
Nurseries	3480	2,037	30	1
Mushroom	2033	225	21	0
Bulbs	2021	869	40	27
Other open air		1,201	30	8
Grazing livestock	*4*			
Dairy	4110+4120+4370			
- Organic		297	32	31
- Non-organic		18,449	309	239
Calf fattening	4380	1,132	28	13
Other grazing livestock	4410+4420+4430	7,900	49	33
Intensive livestock	*5*			
Breeding pigs	5011	1,416	52	45
Fattening pigs	5012	1,527	45	40

Tabel 5.5	Comparison of the field of observation (population) and the sample available for research purposes in 2007 (2007 agricultural census) (continued)			
Type of farming	**Code**	**Number of farms**		
		population	**total**	**CSP**
Integrated pig farms	5013	995	43	37
Laying hens	5021	958	39	35
Poultry	5022	434	26	21
Other intensive livestock	other 5	166	14	4
Combined	*6-8*	5,377	113	62
Total		58,787	1,485	1,097

5.4 Supply of 2007 farm results to the European Commission

The final delivery of 2007 data to the EU has taken place in December 2008. Data of 1,510 farms have been provided to Brussels (Table 5.6). This is the highest number of farms since many years and it fulfils the obligation of 1,500 farms.

Table 5.6	Comparison between the number of farms supplied to the EU and those available for research		
Bookkeeping year	**Provided to the European Commission**	**Weighted farms available for research**	**Other available farms a)**
1990/91	1,587	1,576	12
1991/92	1,505	1,547	8
1992/93	1,513	1,516	7
1993/94	1,525	1,520	7
1994/95	1,546	1,534	13
1995/96	1,536	1,530	6
1996/97	1,551	1,545	6
1997/98	1,529	1,522	7
1998/99	1,368	1,363	5
1999/00	1,341	1,334	7
2000 b)	N/A	N/A	N/A
2001	1,330	1,310	20
2002	1,358	1,344	14
2003	1,437	1,399	38
2004	1,420	1,392	28
2005	1,458	1,406	52
2006	1,506	1,472	34
2007	1,510	1,485	25

a) Other available farms are farms that are also available but without a weight. Reasons for not having a weight are: a farm is outside of the defined field of observation because a farm is too large or to small according to the information in the agricultural census. In alternative weighting systems (based on the characteristics of the farm these farms do get a weight; b) Bookkeeping year 1999/00 ended for arable farms and husbandry at 30 April 2000. Due to capacity problems related to IT problems, farm data for the period from 30 April 2000 to 31 December 2000 (respectively 1 January 2000 to 31 December 2000) are not processed but estimated based on data of 1999/00 and 2000/01.

6 Evaluation of 2007 sample

6.1 Introduction

In this chapter the FADN sample for the year 2007 is evaluated in a qualitative and quantitative way. Section 6.2 provides an evaluation of the methodology of stratification and weighting. A crucial element is the calculation of weights. Section 6.3 provides the quantitative evaluation of the year 2007. This section focuses on the quality of the estimations that can be made based on the sample. This chapter is based on the standard approach of making estimations based on weights assigned to farms. In Appendix 1 an alternative approach to make estimations is described. This alternative approach can be used to improve the quality of estimates.

6.2 Evaluation of stratification and weighting

6.2.1 Introduction

This section deals with some practical problems related to the estimation process. Weights of individual farms are used to make estimations of frequencies, totals and averages of groups of farms (aggregated results) based on the data from the agricultural census and the FADN data.

The method to calculate the weights of individual farms is crucial. The goal is to achieve unbiased estimates with a minimal variance. This enables the estimation of the confidence interval of the real population value and the minimisation of the total error. This is true for direct estimators. In the case of a ratio estimator this is not necessarily true, but ratio estimators are outside the scope of this publication (see Vrolijk et al., 2001) for a more extensive description of ratio estimators and other estimators, see also appendix 1).

6.2.2 Method of calculation of weights

The objective of the Dutch FADN system is to give a representative view of the total population. The question is therefore how to draw conclusions on totals, averages and frequencies that are valid for the whole population based on individual farm data. For example, how much is the average family farm income of

all farms in agriculture and horticulture? The practical solution is found in weighting: the individual farm data are raised to the population level (for some variables the estimated values can be compared to the data that is available for the whole population, i.e. data which are included in the yearly agricultural census). A weight is assigned to every observed farm in the FADN system. The weight is defined as the ratio between the number of farms in a stratum according the agricultural census and the number of farms in the sample (in the FADN system). For the assignment of farms in the FADN system to strata the information from the year 2007 is used. These data can be different from the data when the farm was chosen in the system for the first time. This implies some kind of post-stratification. Weights can be calculated as soon as a substantial number of farms have been completed. During the year, when additional farms are completed, the weights are recalculated. The weights of the farms are recalculated until the accounts of all farms are completed and the final set of weights can be established. For preliminary estimations based on for example 50% of the farms, one should be aware of the fact that this 50% is not necessary representative for the whole population.

The (post) stratification of the farms is based on the 2007 agricultural census. The population in a specific stratum is continuously changing. Therefore the farms that belong to a stratum in 2006 are not exactly the same as the farms that belong to that stratum in 2007. Due to these changes farms included in one stratum could have had different inclusion probabilities at the time of recruitment. In theory, to achieve unbiased estimators these differences in inclusion probabilities should be taken into account in the estimation process. However, the consequence of this would be a very complicated system with many different substrata with different inclusion probabilities. Therefore this complicated procedure is not applied. As a result, the theoretical assumption of a strict random sample can not be validated.

Although the calculation method applied in practice can lead to systematic distortions between estimated values and real values, the assumption of a random sample is made. This leads to several attractive consequences. The method to calculate weights is relatively easy, involving a limited set of homogenous strata and resulting in a more effective use of data.

Because of the applied sampling procedure (see section 2.1) the different strata have different sampling fractions. Strata with relatively homogenous units have a lower sampling fraction than very heterogeneous strata. This also implies that farms have very diverging weights. Farms from a homogeneous cluster will have a larger weight (in principal the reciprocal of the sampling fraction) and therefore represent a larger number of farms. The differences in sampling frac-

tions are shown in Table 6.1. These percentages are calculated by dividing the required number of farms in the selection plan (Table 5.1) by the number of population units (Table 3.3).

Table 6.1 Sampling fractions in different strata (2007 sample)

Type of farming	ESU size class			
	1	2	3	4
Field crop farms				
Starch potatoes	0.02	0.03	0.05	
- Organic crops	0.15	0.14	0.10	
- Other field crop farms	0.01	0.02	0.07	
Horticulture				
Vegetables under glass				
- Paprika	0.13	0.11	0.09	0.08
- Cucumber	0.17	0.12	0.09	0.29
- Tomato	0.18	0.10	0.09	0.14
- Other	0.03	0.04	0.14	0.13
Cut flowers under glass				
- Rose	0.14	0.08	0.09	0.08
- Chrysanthemum	0.16	0.15	0.12	0.21
- Other	0.02	0.03	0.06	0.05
Plants	0.03	0.04	0.07	0.05
Other glass	0.04	0.06	0.08	0.06
Field vegetables	0.02	0.04	0.08	
Fruit	0.02	0.02	0.05	
Nurseries	0.01	0.02	0.04	
Mushroom	0.08	0.17	0.29	0.22
Bulbs	0.03	0.05	0.09	0.08
Other open air	0.01	0.03	0.09	
Grazing livestock				
Dairy				
- Organic	0.08	0.10	0.13	
- Non-organic	0.02	0.01	0.03	
Calf fattening	0.03	0.02	0.04	

Table 6.1	Sampling fractions in different strata (2007 sample) (continued)			
Type of farming	**ESU size class**			
	1	**2**	**3**	**4**
Other grazing livestock	0.00	0.01	0.06	
Intensive livestock				
Breeding pigs	0.03	0.03	0.10	
Fattening pigs	0.02	0.03	0.12	
Integrated pig farms	0.03	0.03	0.09	
Laying hens	0.02	0.04	0.23	
Poultry	0.05	0.06	0.15	
Other intensive livestock	0.06	0.11	0.32	
Combined	0.01	0.02	0.06	

6.2.3 Remarks on the weights of 2007

In the report on farm results for 2007 the research population is defined as all farms in the 2007 agricultural census (between the lower and upper threshold). The weight per farm is calculated as the ratio between the number of farms in the census and the number of farms in the sample.

In the calculation of aggregated results (averages, frequencies and totals) for the year 2007 the 2007 agricultural census is the starting point. Because of the complete registration of farms in the population (almost all farms are registered in the agricultural census) the aggregated numbers of farms are exactly the same as the number of farms in the census. However, in using these numbers in the calculation of weights for estimations for 2007 two remarks should be made.

Every year all horticultural and agricultural farms are registered in the agricultural census, but this registration only represents the situation at a certain moment during the year. Therefore it is possible that farms are missing from this registration. Furthermore, the number of farms tend to fall significantly (this trend is stronger for certain types of farms and less strong for others). As a consequence estimations for the year 2007 might be overestimations of reality.

Distortions in the number of farms in the census can therefore cause incorrect estimations of aggregates.

Furthermore, the typology of farms according to the agricultural census might differ from the typology according to the FADN data. The census reflects the situation at a certain point in time, while the FADN system describes the

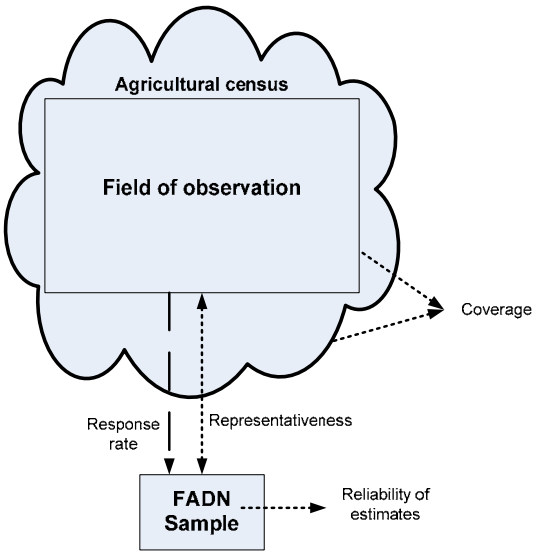
farm during a whole year. In order to take these differences into account two weighting methodologies are available in the Dutch FADN system. From a theoretical point of view weighting based on the characteristics of the farm in the census is more correct. The census is used as the sampling frame, the weights should reflect information from this sampling process. Furthermore, if there are substantial differences, then the variables type and size of farming in the agricultural census are different from the variables size and type of farming in the FADN. In a weighting procedure based on the population numbers in the census and the characteristics in the FADN these variables are considered to be the same.

6.3 Quantitative evaluation of 2007

6.3.1 Introduction

This section focuses on the quality of the estimations based on the 2007 FADN sample. Figure 6.1 shows the same structure as displayed in Figure 2.1, but it adds the quality aspects. Section 6.3.2 provides information on the coverage of the sample; the coverage compares the total population as described by the census and the field of observation of the FADN sample. Section 6.3.3 analyses the extent to which distortions might occur between the sample and the population due to over or under representation of farms with specific characteristics; it compares the characteristics of the field of observation and the actual FADN sample. Section 6.3.4 provides information on the reliability of estimates based on the FADN sample. The last quality aspect listed in Figure 6.1, the response rate and the non-response, has already been described in the previous chapter.

Figure 6.1 **Quality aspects of the Dutch FADN**



6.3.2 Coverage

It is desirable to have a sample that represents the population as well as possible. A clear distinction should be made between the coverage and the representativeness. This section describes the coverage, section 6.3.3 deals with the representativeness. To get an idea about the extent to which the total population is covered by the sample it is relevant to distinguish several aspects. Farms that are too small or are not registered in time are not part of the agricultural census (b). The sampling frame (c) is the basis for the choice of sample farms and consists of farms registered in the agricultural census and have a size of more then 16 ESU and less than 2,000 ESU. From this sampling frame the sample is drawn (d).

Figure 6.2 **Relationship between FADN sample and all farms**

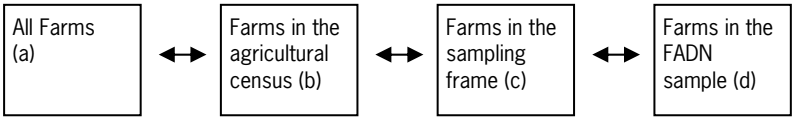


Table 6.2 gives an indication to what extent the FADN sample covers the whole population. A comparison is made between the farms in the sampling framework (all the farms that have a chance of being included in the FADN sample) (c) and the total population as described by the agricultural census (b). Direct comparison with all farms (a) would be better but the unregistered farms are unknown, and the practical difference is very limited. The sampling framework covers the population to a large extent. For example with respect to the production, almost 92% is covered by the sample. Small farms are excluded from the sampling framework, this means that a substantial number of the farms and to a lesser extent also of labour are outside of the sampling frame. With respect to agricultural activities, the table shows that some activities are not well covered by the sample. This mainly concerns the activities that are commonly found on very small (ewes) or on very large specialised farms (tomatoes).

Table 6.2 **Coverage of the sample compared to 2007 agricultural census**

Variable-agricultural census	Number according to census	Not covered in sample (%)		Percentage covered by sample
		of which <16 ESU	of which >2,000 ESU	
Farms	76,741	23.2	0.2	76.6
Dutch size units	7,215,044	1.8	6.4	91.8
Farm managers	61,889	13.7	0.3	86.0
Family labour	104,664	11.0	0.2	88.8
Paid labour	46,313	2.6	9.0	88.4
Total labour	150,978	8.4	2.9	88.7
Size in hectares				
Agricultural area	1,914,330	5.2	0.6	94.1
Arable	997,445	4.3	0.6	95.1

Table 6.2	Coverage of the sample compared to 2007 agricultural census (continued)			
Variable-agricultural census	**Number according to census**	**Not covered in sample (%)**		**Percentage covered - by sample**
		of which <16 ESU	**of which >2,000 ESU**	
Grassland	820,703	6.8	0.0	93.2
Horticulture under glass	10,374	0.2	13.5	86.3
Vegetables in the open air	85,809	1.0	5.5	93.6
Number of animals				
Dairy cows	1,413,166	0.1	0.0	99.9
Fattening calves	859,872	0.9	0.1	98.9
Ewes	678,644	20.2	0.0	79.8
Fattening pigs	5,558,828	1.0	1.0	98.0
Breeding pigs	1,266,471	0.1	1.3	98.5
Laying hens	41,226,541	0.4	0.0	99.6
Poultry	43,351,729	0.2	0.7	99.2
Size in hectares				
Winter cereal	124,429	4.4	0.5	95.2
Seed potatoes	36,729	0.2	0.5	99.4
Consumption potatoes	72,464	1.2	0.6	98.2
Starch potatoes	47,980	1.0	1.7	97.3
Sugar beets	82,026	2.6	0.7	96.8
Peas for canning	6,027	2.0	3.9	94.1
Seed onions	20,148	0.4	0.3	99.3
Grass seed	20,097	3.4	0.5	96.2
Green maize	221,554	7.7	0.1	92.2
Brussel sprouts	3,352	0.5	0.5	99.1
Cabbage all types	2,633	0.7	4.9	94.4
Asparagus	2,383	2.3	0.5	97.2
Tulips	10,740	0.1	5.6	94.2
Hedges	2,904	1.9	1.7	96.5

Table 6.2	Coverage of the sample compared to 2007 agricultural census (continued)			
Variable-agricultural census	**Number according to census**	**Not covered in sample (%)**		**Percentage covered - by sample**
		of which <16 ESU	**of which >2,000 ESU**	
Trees	5,643	0.7	9.8	89.5
Apples	9,380	1.7	0.0	98.3
Pears	7,296	1.5	0.0	98.5
Tomatoes under glass	1,545	0.0	38.2	61.7
Cucumbers under glass	617	0.0	3.6	96.4
Paprika under glass	1,188	0.0	12.0	88.0
Roses	652	0.0	8.0	92.0
Chrysanthemum	566	0.1	0.0	99.9
Fresia	155	0.0	0.0	100.0
Ornamentals leave	552	0.1	10.8	89.1
Ornamentals flower	845	0.0	21.2	78.7
Mushrooms	70	0.1	16.8	83.1

In policy analysis and research it is essential to distinguish between farming types (for example specialised pig fattening farms) and agricultural activities (pig fattening). In the report on the redesign of the FADN sample it was illustrated that types of farming should not be the only focus of research (Vrolijk and Lodder, 2002). Agricultural activities are important in many research projects.

To give a complete picture of a certain agricultural activity it is therefore important to look at the activities on all farm types. For example, not only pig fattening farms will create added value from pig fattening, also other types of farms can be involved in this activity (although it is not their main business). The next table describes to which extent a certain activity can be found on certain types of farming. For example, 82.8% of the agricultural activity fattening pigs can be found on the intensive livestock farms. This means that 17.2% of this activity can be found on farms that belong to other types of farming, for example arable farms.

Table 6.3	Relationship between types of farming and agricultural activities - share of ESU 2007								
Type of farming	**Dairy**	**Cattle**	**Sheep**	**Goat**	**Grass-land**	**Fattening pig**	**Other pig**	**Laying hen**	**Poultry**
Field crop farms									
- Starch potatoes	0.01	0.37	0.18	0.01	0.20	0.00	0.25	0.12	0.67
- Organic crops	0.00	0.17	0.08	0.00	0.26	0.00	0.05	0.07	0.00
- Other field crop farms	0.06	2.15	3.37	0.20	4.48	0.15	0.82	0.67	2.40
Horticulture									
Vegetables under glass									
	0.00	0.00	0.02	0.00	0.01	0.00	0.00	0.00	0.00
- Cucumber	0.00	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.00
- Tomato	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
- Other	0.00	0.00	0.07	0.01	0.06	0.00	0.03	0.00	0.00
Cut flowers under glass									
- Rose	0.00	0.00	0.04	0.01	0.02	0.00	0.00	0.00	0.00
- Chrysanthemum	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
- Other	0.00	0.00	0.08	0.01	0.06	0.00	0.00	0.17	0.00
Plants	0.00	0.00	0.05	0.00	0.05	0.00	0.00	0.00	0.00
Other glass	0.00	0.00	0.02	0.01	0.03	0.00	0.00	0.00	0.00
Field vegetables	0.01	0.27	0.07	0.01	0.15	0.03	0.12	0.04	0.10
Fruit	0.01	0.17	0.21	0.01	0.28	0.01	0.07	0.18	0.00
Nurseries	0.02	0.23	0.18	0.01	0.34	0.25	0.23	0.07	0.00
Mushroom	0.00	0.06	0.02	0.00	0.05	0.00	0.00	0.00	0.00
Bulbs	0.04	0.12	0.15	0.00	0.16	0.07	0.20	0.00	0.26
Other open air	0.04	0.28	0.18	0.01	0.30	0.03	0.27	0.05	0.30
Grazing livestock									
Dairy									
- Organic	92.58	40.08	21.67	1.29	3.03	2.05	8.18	1.09	1.01
- Non-organic	1.32	0.75	0.49	0.05	0.46	0.05	0.07	0.24	0.00
Calf fattening	0.02	0.71	0.55	0.09	0.07	0.00	0.22	0.15	0.14

Table 6.3	Relationship between types of farming and agricultural activities - share of ESU 2007 (continued)								
Type of farming	**Dairy**	**Cattle**	**Sheep**	**Goat**	**Grass-land**	**Fattening pig**	**Other pig**	**Laying hen**	**Poultry**
Other grazing livestock	1.59	36.89	60.99	91.23	73.56	0.26	1.22	0.50	0.09
Intensive livestock									
Fattening pigs	0.03	0.24	0.93	0.11	1.46	56.04	3.92	0.19	0.11
Breeding pigs	0.01	0.38	0.62	0.03	1.05	0.26	36.88	0.03	0.05
Integrated pig farms	0.02	0.53	0.52	0.02	0.79	25.26	25.37	0.01	0.37
Laying hens	0.01	0.15	0.58	0.01	0.81	0.14	0.33	80.02	0.11
Poultry	0.02	0.03	0.23	0.00	0.28	0.03	0.15	0.04	70.42
Other intensive livestock	0.02	0.15	0.15	0.01	0.17	1.11	1.56	4.80	3.83
Mixed	4.19	16.26	8.50	6.88	11.84	14.26	20.06	11.54	20.13
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Field crop farms									
- Starch potatoes	6.45	14.51	0.24	0.04	0.00	0.00	0.00	0.00	0.00
- Organic crops	1.64	0.95	2.62	0.48	0.07	0.00	0.07	0.03	0.00
- Other field crop farms	55.92	60.69	2.93	0.80	0.09	0.00	0.73	0.00	0.01

Table 6.3	Relationship between types of farming and agricultural activities - share of ESU 2007 (continued)								
Type of farming	**Wheat**	**Root crops**	**Vegetable open air**	**Fruit**	**Tree**	**Mushroom**	**Bulbs**	**Vegetables glass**	**Cut flowers glass**
Vegetables under glass									
- Paprika	0.01	0.00	0.04	0.02	0.01	0.00	0.00	27.00	0.04
- Cucumber	0.02	0.00	0.05	0.00	0.00	0.00	0.00	10.76	0.03
- Tomato	0.01	0.00	0.02	0.00	0.01	0.00	0.00	36.94	0.02
- Other	0.17	0.02	2.17	0.18	0.04	0.00	0.00	21.70	0.12
Cut flowers under glass									
- Rose	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.03	29.29
- Chrysan-themum	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.01	10.01
- Other	0.09	0.03	0.16	0.06	0.13	0.00	1.14	0.15	48.75
Plants	0.02	0.00	0.03	0.02	0.19	0.00	0.00	0.11	0.92
Other glass	0.07	0.02	1.50	0.25	3.02	0.00	3.32	1.81	5.45
Field vegetables	0.64	0.54	59.97	0.29	0.06	0.00	0.01	0.41	0.02
Fruit	0.42	0.20	0.22	85.15	0.10	0.00	0.00	0.02	0.00
Nurseries	0.55	0.22	0.32	0.42	84.59	0.00	0.02	0.01	0.08
Mushroom	0.02	0.00	0.00	0.11	0.00	99.34	0.00	0.00	0.00
Bulbs	0.83	1.03	0.54	0.00	0.03	0.00	76.98	0.01	1.21
Other open air	0.75	0.75	7.56	2.14	3.50	0.02	9.54	0.66	3.42
Grazing livestock									
Dairy									
- Organic	3.02	3.26	0.80	0.42	0.27	0.00	0.35	0.00	0.00
- Non-organic	0.18	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Calf fattening	0.28	0.10	0.09	0.02	0.02	0.00	0.00	0.00	0.00

Table 6.3	Relationship between types of farming and agricultural activities - share of ESU 2007 (continued)								
Type of farming	**Wheat**	**Root crops**	**Vegetable open air**	**Fruit**	**Tree**	**Mushroom**	**Bulbs**	**Vegetables glass**	**Cut flowers glass**
Other grazing livestock	4.27	0.60	0.32	0.24	0.09	0.00	0.04	0.00	0.00
Intensive livestock									
Fattening pigs	1.99	0.34	0.23	0.06	0.01	0.00	0.02	0.00	0.00
Breeding pigs	1.33	0.20	0.07	0.01	0.04	0.00	0.00	0.00	0.00
Integrated pig farms	2.09	0.58	0.31	0.04	0.09	0.00	0.01	0.00	0.00
Laying hens	0.62	0.17	0.26	0.04	0.05	0.00	0.00	0.00	0.00
Poultry	0.38	0.07	0.10	0.01	0.00	0.00	0.00	0.00	0.00
Other intensive livestock	0.22	0.05	0.01	0.00	0.01	0.00	0.00	0.00	0.00
Mixed	18.00	15.65	19.43	9.19	7.49	0.64	7.77	0.35	0.62
Total	100.0	100.0	100.	100.0	100.0	100.0	100.0	100.0	100.0

6.3.3 Representativeness

Because of the stratification scheme the sample will provide a good representation of the population on the main characteristics (stratification variables) at the beginning of a year. During the year farms might drop out of the sample and changes might occur in the population. Despite these changes the representativeness is maintained by applying post-stratification on the resulting sample and the changed population. Representativeness with respect to the stratification variables does not necessary imply that the sample is representative for all variables. Such a full representativeness is impossible unless the sample size approximates the whole population or highly correlated with the stratification variables. Table 6.4 shows to what extent the sample is representative for a number of variables in the agricultural census.

The following guideline can help in the interpretation of the table: a relative difference which is close to the relative standard error cannot be regarded as proof of systematic differences between the sample and the population. If the relative difference is more than two times the relative standard error then it is less likely that these differences can be explained by sampling errors. It is very unlikely that the difference is caused by coincidence if the relative difference is more than 3 times the relative standard error.

An example can illustrate how the table should be interpreted. The average number of DSU (Dutch size units) of pigs as measured in the 2007 agricultural census is 8.61 (i.e. the average of all farms within the field of observation). If the same variable is estimated based on the FADN sample an average of 8.77 is calculated. It might seem that the number of pigs is slightly overestimated in the sample. However, the relative standard error of the estimate is 3.2%. When this standard error is compared to the relative difference between both values (1.8%), then the conclusion that there is a significant difference, cannot be supported.

Table 6.4 **Comparison of farms in the agricultural census (16-2000 ESU) and farms in the Dutch FADN (2007 agricultural census)**

Variable	Average calculated based on		Relative standard error (FADN)	Ratio Census and FADN		
				all farms	farms with value >0	
	census (1)	FADN (2)		average (1/2)	number	average
dsu	112.66	115.63	0.91	97.4	100	97.4
Activities (dsu)						
Field crops	13.48	15	2.52	89.9	90.79	99
Grassland	2.17	2.61	13.02	83.1	96.15	86.5
Horticulture in the open	17.29	16.93	3.29	102.1	100.68	101.5
Horticulture under glass	25.38	24.54	1.91	103.4	100.97	102.4
Cattle	36.08	36.53	1.62	98.8	99.31	99.5
Dairy cows	28.92	29.38	1.69	98.4	99.87	98.6
Fattening cattle	0.82	0.98	18.58	83.8	92.83	90.2
Veal	2.03	1.98	8.32	102.3	99.12	103.2
Horses	2.05	1.04	21.24	197.2	132.78	148.5
Sheep	0.48	0.48	17.62	100.2	97.24	103

Table 6.4	Comparison of farms in the agricultural census (16-2000 ESU) and farms in the Dutch FADN (2007 agricultural census) (continued)					
Variable	**Average calculated based on**		**Relative standard error (FADN)**	**Ratio Census and FADN**		
				all farms	**farms with value >0**	
	census (1)	**FADN (2)**		**average (1/2)**	**number**	**average**
Goats	0.36	1.09	30.99	32.8	68.33	48
Pigs	8.61	8.77	3.24	98.2	101.88	96.4
Fattening pigs	4.04	3.93	4.4	102.8	104.39	98.5
Breeding pigs	4.55	4.84	4.42	94	95.19	98.7
Poultry	3.42	3.68	5.81	92.9	84.38	110
Fattening peepers	0.98	1.01	12.36	97.3	85.91	113.2
Laying hen	1.62	2.2	8.46	73.6	71.68	102.6
Turkey	0.12	0.12	36.28	100.3	142.53	70.4
Rabbits	0.05	0.2	57.41	26.3	50.37	52.2
Fur animals	0.67	1.18	36.6	57	65.39	87.2
Sizes (ha)						
UAA	30.66	32.76	1.68	93.6	99.53	94
Field crops	16.13	17.87	2.67	90.3	93.55	96.5
Horticulture open air	1.37	1.52	4.57	90	100.68	89.4
Horticulture glass	0.15	0.15	2.32	100.3	100.97	99.4
Permanent grass	12.58	12.94	3.88	97.2	99.26	98
Acreages field crops						
Grains	3.52	3.9	5.29	90.2	91.79	98.3
Leguminous plants	0.04	0.07	32.99	60.8	81.42	74.6
Commercial crops	0.17	0.26	17.36	65.9	72.58	90.7
Seeds	0.34	0.51	13.11	67.3	65.59	102.6
Tuberous and carrots	3.98	4.52	3.48	88	87.36	100.8
Green fodder	6.8	7.34	4.85	92.6	91.39	101.3
Green fertiliser	0.27	0.31	13.98	88	95.39	92.2

Table 6.4	Comparison of farms in the agricultural census (16-2000 ESU) and farms in the Dutch FADN (2007 agricultural census) (continued)					
Variable	**Average calculated based on**		**Relative standard error (FADN)**	**Ratio Census and FADN**		
				all farms	**farms with value >0**	
	census (1)	**FADN (2)**		**average (1/2)**	**number**	**average**
Horticulture in the open air						
Vegetables (market garden)	0.39	0.44	10.76	87.2	85.79	101.7
Stone fruit	0.29	0.33	6.98	86.3	108.5	79.5
Small fruits	0.03	0.03	58.02	82.7	121.05	68.3
Flower nursery	0.04	0.04	21.41	117.3	104.95	111.7
Tree nursery	0.22	0.2	16.01	112.4	112.42	100
Flower bulbs	0.37	0.46	6.54	81.2	82.87	97.9
Glass houses						
Vegetables	0.06	0.06	3.5	97	97.36	99.7
Tomatoes	0.02	0.02	4.51	103.3	108.37	95.3
Cucumbers	0.01	0.01	5.28	93.2	104.12	89.5
Paprika	0.02	0.02	3.51	95.4	102.86	92.7
Fruit	0	0	47.95	57.5	68.32	84.2
Cut flowers	0.05	0.05	4.42	98.3	98.58	99.7
Roses	0.01	0.01	4.62	109	111.46	97.8
Chrysanthemum	0.01	0.01	6.9	102.1	104.35	97.8
Plants	0.03	0.03	5.5	94.4	101.14	93.3
Tree nursery	0.01	0	27.22	141.9	99.6	142.4
Standing glass	0.15	0.15	2.3	100.8	101.65	99.2
Mushrooms						
Cell	0.03	0.03	7.56	91.9	71.04	129.3
Size (are)	0	0	11.12	86.7	71.04	122.1
Chicory						
Size (are)	0.03	0.05	45.25	58.5	55.82	104.7
Bulbs						
Tulips (pieces)	21.86	35.84	21.72	61	106.18	57.5
Narcissus (kg)	0.05	0.01	64.44	405	221.24	183.1

Table 6.4	Comparison of farms in the agricultural census (16-2000 ESU) and farms in the Dutch FADN (2007 agricultural census) (continued)					
Variable	**Average calculated based on**		**Relative standard error (FADN)**	**Ratio Census and FADN**		
				all farms	**farms with value >0**	
	census (1)	**FADN (2)**		**average (1/2)**	**number**	**average**
Substrate growing (are)						
Vegetable	0.05	0.05	3.68	93.2	96.38	96.7
Flowers	0.02	0.02	11.22	90.8	86.52	104.9
Stable capacity (number of animals)						
fattening calves	17.65	20.68	14.7	85.4	98.81	86.4
fattening pigs	115.91	111.82	4.35	103.7	108.81	95.3
peepers	850.37	994.63	15.24	85.5	74.36	115
laying hen	637.69	962.19	8.48	66.3	57.66	114.9
Characteristics firm and entrepreneur						
Age	51.01	49.69	0.77	102.7	100	102.7
Labour						
Total	3.23	3.42	4.41	94.2	99.77	94.4
Male	2.05	2.11	3.58	97.1	99.31	97.7
Female	1.18	1.31	6.18	89.7	93.9	95.5
Paid labour	1.04	1.06	13.8	97.9	91.63	106.9

The information in Table 6.4 gives an indication for which variables and consequently for which research projects it might be wise to perform post-stratification or use alternative estimation techniques to take into account the differences between the sample and the population. For example, in studies in which the age of the farmer plays an important role it might be useful to apply alternative estimation techniques. Table 6.4 gives a description for the whole population. In case of research projects on specific types of farming, similar tables could be generated for only farms of that type of farming.

The last two columns of Table 6.4 provide more detailed information on the difference between the population and the sample. These differences can be

explained on one hand by differences in the number of farms on which a certain activity occurs (a value larger than zero) and on the other by the average of this activity on farms which are in this activity. For example: the number of dsu dairy cows in the FADN is higher than in the agricultural census (no significant difference). This difference is partly explained by a marginally higher estimation of the number of farms with dairy cows and partly by a higher estimation of DSU of dairy cows on farms with dairy cows (98.4 = 99.87% * 98.6).

A comparison between the sample and the population as registered in the agricultural census does not fully answer the question whether estimations of financial, economic and technical characteristics are bias free. It is for example possible that farms with relatively good or bad management skills and therefore performance are over represented in the sample.

6.3.3 Reliability

The previous subsection provides some indicators whether there are systematic differences between the sample and the population (representativeness of sample). This section focuses on the reliability of the estimates.

The calculation of averages of groups based on sampling units implies that there can be differences between the estimated value and true population value. These differences can occur due to the random selection of units to be included in the sample. Table 6.5 provides an indication of the level of precision of the estimates for a set of important goal variables.

The precision of estimates can be measured by the standard error of the estimate of a variable. The standard error is used to calculate the confidence interval. This confidence interval describes the range in which the true population value will be given a certain level of certainty. The confidence interval ranges from the calculated average minus two times the standard error to the calculated average plus two times the standard error. The calculated averages of two groups are significantly different (with a 95% certainty) if the difference is larger than two times the square root of the sum of squares of the standard errors of the two group averages.

This section provides the reliability of estimates for a number of important goal variables for different types of farming. This calculation is based on the available CSP observations (see Section 5.3).

Table 6.5	Reliability of estimates of important goal variables per type of farming, based on FADN sample (2007)					
Type of farming	**Goal variable**					
	family farm income	**total revenues**	**return a)**	**savings**	**income farm**	**net farm result**
Field crop farms						
- Starch potatoes	17,932	123,333	3.1	14,192	17,756	14,591
- Organic crops	17,769	34,523	5.2	19,950	24,311	14,840
- Other field crop farms	10,139	31,613	2.3	11,163	10,769	7,779
Horticulture						
Vegetables under glass						
- Paprika	44,832	89,600	2.4	42,938	45,179	37,341
- Cucumber	50,682	104,891	1.6	52,317	46,650	31,151
- Tomato	48,449	185,549	1.9	57,841	48,183	43,884
- Other	13,755	39,890	4.6	11,786	13,541	9,878
Cut flowers under glass						
- Rose	47,329	212,388	2.9	34,592	46,433	38,945
- Chrysanthe-mum	32,997	110,700	2.8	32,584	33,185	25,152
- Other	20,335	48,010	2.3	18,548	21,371	14,988
Plants	29,043	132,582	2.7	25,060	28,244	27,379
Other glass	21,810	45,183	14.6	9,608	13,749	14,125
Field vegetables	*	*	*	*	*	*
Fruit	14,380	27,989	3.6	13,074	14,564	13,290
Nurseries	*	*	*	*	*	*
Mushroom	*	*	*	*	*	*
Bulbs	31,292	71,544	2.7	31,493	30,999	27,667
Other open air	90,485	690,912	10.4	97,652	92,313	93,015
Grazing livestock						
Dairy						
- Organic	6,748	16,214	2.2	15,546	6,232	6,460
- Non-organic	3,461	6,609	0.9	5,407	3,467	2,999
Calf fattening	8,437	24,563	5.0	5,558	5,579	10,731

Table 6.5	Reliability of estimates of important goal variables per type of farming, based on FADN sample (2007) (continued)					
Type of farming	**Goal variable**					
	family farm income	**total revenues**	**return a)**	**savings**	**income farm**	**net farm result**
Other grazing livestock	8,399	18,628	3.7	6,540	12,165	10,914
Intensive livestock						
Breeding pigs	12,822	42,946	2.1	15,916	13,200	12,880
Fattening pigs	12,514	34,638	2.1	13,334	12,767	9,578
Integrated pig farms	13,356	41,997	1.3	15,576	16,752	11,984
Laying hens	55,313	58,610	2.9	64,527	63,501	16,932
Poultry	14,138	82,437	1.4	12,833	12,121	11,562
Other intensive livestock	*	*	*	*	*	*
Mixed	14,995	54,977	3.1	14,240	14,979	14,185

a) Revenues per 100 euro costs; * Insufficient number of observation in CSP variant.

Table 6.6	Reliability of estimates of important goal variables per main type of farming, based on FADN sample (2007)					
Type of farming	**Goal variable**					
	family farm income	**total revenues**	**return**	**savings**	**income farm**	**net farm result**
Field crops	8,863	30,985	2.0	9,599	9,382	6,830
Vegetables under glass	16,092	45,010	2.2	16,627	15,789	12,782
Cut flowers under glass	17,293	48,047	1.8	15,422	17,992	12,971
Pigs	7,496	23,053	1.2	8,659	8,059	6,664
Poultry	38,322	47,829	2.1	44,588	43,866	12,198
Grazing livestock	3,333	6,955	1.2	4,054	4,163	3,715
All farms	3,555	17,020	0.8	3,840	3,846	3,362

There are clear differences in the significance of estimates between different types of farming. The estimates for the dairy sector are the most reliable because of the large number of farms included in the sample, which reflects the

importance of the dairy sector in Dutch agriculture. The decision on the number of farms is described in Vrolijk and Lodder (2002).

Tables 6.7 and 6.8 describe the relative standard error (coefficient of variance). This is the standard error divided by the group average. A higher relative standard error implies less reliable estimates, but the value is strongly affected by the absolute value of the average. If the average value approaches zero, the relative standard error can become very large. A meaningful evaluation of the standard error requires a simultaneous use of Tables 6.5 and 6.6 on one hand and Tables 6.7 and 6.8 on the other.

Table 6.7 **Coefficient of variation of estimates of important goal variables per main type of farming, based on FADN sample (2007)**

Type of farming	Goal variable					
	family farm income	total revenues	return	savings	income farm	net farm result
Field crops	0.122	0.100	0.023	0.332	0.117	-0.412
Vegetables under glass	0.236	0.043	0.025	-1.503	0.213	-0.411
Cut flowers under glass	0.265	0.055	0.020	-0.755	0.271	-0.317
Pigs	-0.124	0.050	0.016	-0.088	-0.188	-0.049
Poultry	0.357	0.066	0.023	0.561	0.348	-2.196
Grazing livestock	0.050	0.028	0.015	0.112	0.050	-0.114

Table 6.8 **Coefficient of variation of estimates of important goal variables per type of farming, based on FADN sample (2007)**

Type of farming	Goal variable					
	family farm income	total revenues	return	savings	income farm	net farm result
Field crop farms						
- Starch potatoes	0.22	0.36	0.03	0.29	0.18	7.33
- Organic crops	0.27	0.13	0.06	0.53	0.28	-1.48
- Other field crop farms	0.14	0.10	0.03	0.43	0.14	-0.40
Horticulture						

Table 6.8	Coefficient of variation of estimates of important goal variables per type of farming, based on FADN sample (2007) (continued)					
Type of farming	**Goal variable**					
	family farm income	**total revenues**	**return**	**savings**	**income farm**	**net farm result**
Vegetables under glass						
- Paprika	0.23	0.06	0.02	0.50	0.23	0.56
- Cucumber	-2.10	0.08	0.02	-0.38	-4.69	-0.33
- Tomato	-0.61	0.11	0.02	-0.40	-0.67	-0.26
- Other	0.14	0.08	0.06	0.31	0.13	-5.71
Cut flowers under glass						
- Rose	1.70	0.12	0.03	-0.52	1.43	-0.98
- Chrysanthemum	0.40	0.07	0.03	-1.93	0.40	-0.63
- Other	0.29	0.07	0.03	-1.44	0.30	-0.36
Plants	0.35	0.12	0.03	2.22	0.31	-2.83
Other glass	0.48	0.16	0.22	0.42	0.24	-0.34
Field vegetables	*	*	*	*	*	*
Fruit	0.31	0.10	0.04	-2.40	0.27	-0.37
Nurseries	*	*	*	*	*	*
Mushroom	*	*	*	*	*	*
Bulbs	0.45	0.10	0.03	-4.56	0.42	-0.57
Other open air	-2.64	0.77	0.14	-1.22	-2.98	-0.65
Grazing livestock						
Dairy						
- Organic	0.09	0.06	0.03	0.35	0.07	-0.20
- Non-organic	0.04	0.02	0.01	0.08	0.03	-0.21
Calf fattening	0.29	0.14	0.06	-0.40	0.17	-0.24
Other grazing livestock	-3.22	0.15	0.07	-0.27	0.48	-0.15

Table 6.8	Coefficient of variation of estimates of important goal variables per type of farming, based on FADN sample (2007) (continued)					
Type of farming	**Goal variable**					
	family farm income	**total revenues**	**return**	**savings**	**income farm**	**net farm result**
Intensive livestock						
Breeding pigs	-0.14	0.09	0.03	-0.11	-0.16	-0.07
Fattening pigs	-0.59	0.10	0.03	-0.34	-40.06	-0.13
Integrated pig farms	-0.18	0.07	0.02	-0.13	-0.32	-0.07
Laying hens	0.41	0.09	0.03	0.59	0.42	7.14
Poultry	0.33	0.09	0.01	1.04	0.18	-0.50
Other intensive livestock	*	*	*	*	*	*
Mixed	0.27	0.14	0.04	1.74	0.22	-0.39

The previous tables give an indication of the reliability of estimates for certain types of farming. These tables are used to evaluate the allocation of sampling capacity to the different types of farming. Also in research projects the tables give an indication of the reliability of estimates and should therefore be considered before drawing statistical conclusions.

The tables also give an indication of the dispersion of observations. A large dispersion makes it more difficult to make precise estimates of group characteristics. Dispersion is however also on the main advantages of the FADN systems. The micro economic information at farm level makes it possible to show and analyse differences between farms.

References

Bont, C.J.A.M. de, W.H. van Everdingen and B. Koole, *Standard Gross Margins in the Netherlands.* LEI, Den Haag, 2003.

Dijk, J., *De steekproef gewogen.* Onderzoeksverslag 53. Landbouw-Economisch Instituut, Den Haag, 1989.

Dijk, J., B. Kortekaas, K. Lodder en J. Luijt, *Netjes over de drempel; voorstel tot harmonisatie van de steekproeven voor de boekhoudnetten land-, tuin- en bosbouw*. Interne Nota 422. Landbouw-Economisch Instituut (LEI-DLO), Den Haag, 1994.

Dijk, J., K. Lodder, J. Luyt en H.C. Pruis, *Voorstel voor de indeling van de populatie land- en tuinbouwbedrijven in groepen en bijbehorend keuzeplan*. Interne Nota 437. Landbouw-Economisch Instituut (LEI-DLO), Den Haag, 1995a.

Dijk, J., J. Schering en K. Lodder, *Improvement of the weighting system of the FADN.* Landbouw-Economisch Instituut (LEI-DLO), Den Haag, 1995b.

Dijk, J., *Nonresponse in the Dutch Farm Account Data Network*. Paper presented at the Essex 1996 Fourth International Conference on Social Science Methodology, 1-5 July 1996. Landbouw-Economisch Instituut (LEI-DLO), Den Haag, 1996.

Dijk van, J.P.M., B.E. Douma en A.L.J. van Vliet*, Bedrijfsuitkomsten in de landbouw (BUL). Boekjaren 1992/93 t/m 1995/96.* Periodieke Rapportage 11-95/96. Landbouw-Economisch Instituut (LEI-DLO), Den Haag, 1997a.

Dijk van, J.P.M., B.E. Douma en A.L.J. van Vliet, *De financiële positie van de landbouw (FIP) Boekjaar 1995/96 en vergelijkingen met voorgaande jaren*. Periodieke Rapportage 12-95/96. Landbouw-Economisch Instituut (LEI-DLO), Den Haag, 1997b.

Dijk, van J.P.M., J.J.P. Groot, K. Lodder, H.C.J. Vrolijk, *De steekproef voor het Bedrijven-Informatienet van het LEI. Bedrijfskeuze 1999 en selectieplan 2000.* Rapport 6.00.94. LEI, Den Haag, 2000.

Dijk, van J.P.M., J.J.P. Groot, K. Lodder, L.C. van Staalduinen, H.C.J. Vrolijk, *De steekproef voor het Bedrijven-Informatienet van het LEI. Bedrijfskeuze 1998 en selectieplan 1999*. Rapport 6.99.94. LEI, Den Haag, 1999.

Dijk, J.P.M. van, B.E. Douma en A.L.J. van Vliet, *BUL: Bedrijfsuitkomsten in de landbouw (bedrijfsresultaten (verlies- en winstrekeningen) voor akkerbouw- en veehouderijbedrijven*. PR 11. Landbouw-Economisch Instituut (LEI-DLO), Den Haag, 1996.

Dijk, van J.P.M., K. Lodder, H.C.J. Vrolijk, *De steekproef voor het Bedrijven-Informatienet van het LEI. Bedrijfskeuze 2000 en selectieplan 2001.* Report 1.01.02, 2002. LEI, Den Haag,

Poppe, K.J., *Het Bedrijven-Informatienet van A tot Z*. Rapport 1.03.06. LEI, Den Haag, 2004.

Vrolijk, H.C.J., *Working Procedures for the selection of farms in the FADN*, in: G. Beers en K. Poppe (eds.), PACIOLI 9. Innovations in the FADN. Report 8.02.02. Agricultural Economics Research Institute (LEI), The Hague, 2002.

Vrolijk, H.C.J. en K. Lodder, *Voorstel tot vernieuwing van het steekproefplan voor het Bedrijven-Informatienet*. Rapport 1.02.02. LEI, Den Haag, 2002.

Vrolijk, H.C.J., W. Dol en G. Cotteleer, *Schatten van kenmerken van kleine deelgebieden.* LEI, Den Haag, 2002.

Vrolijk H.C.J. en G.C. Cotteleer, *Non-respons en rotatie in het Bedrijven-Informatienet; Kwantitatieve en kwalitatieve analyse van de effecten*, Report 1.05.01. LEI, Den Haag, 2005.

Vrolijk H.C.J., *Non-response in the Dutch FADN. Qualitative reasons and quantitative impacts*. In: K.J. Poppe, PACIOLI 13. Micro economic Data on Farm Diversification. Rural Businesses and the Intra-generational Transfer. Report 8.05.04. pp. 130 - 139. LEI, The Hague, 2005.

Vrolijk, H.C.J., H.B. van der Veen and J.P.M. van Dijk, Sample of Dutch FADN 2004. Design principles and quality of the sample of agricultural and horticultural holdings. Report 1.06.03.2006. LEI, The Hague, 2006

Vrolijk, H.C.J., *Sampling of organic farms in the Dutch FADN: lessons learned*. In: M. Rippin, H. Willer, N. Lampkin, A. Vaughan, Towards a European Framework for Organic Market Information. pp. 87-90, 2006.

Vrolijk, H.C.J., H.B. van der Veen and J.P.M. van Dijk, *Sample of Dutch FADN 2005. Design principles and quality of the sample of agricultural and horticultural holdings.* Report 1.08.01. LEI, The Hague, 2008.

Vrolijk, H.C.J., H.B. van der Veen and J.P.M. van Dijk, *Sample of Dutch FADN 2006. Design principles and quality of the sample of agricultural and horticultural holdings.* Report 2008-81, LEI, The Hague, 2009.

Welten, J.P.P.J., *Berekening en toepassing van Nederlandse grootte-eenheden en standaardbedrijfseenheden (nge 1992 en sbe 1993)*. Periodieke Rapportage 63-92. Landbouw-Economisch Instituut (LEI-DLO), Den Haag, 1996.

Appendix 1

Statistical matching based on FADN and agricultural census

1. Introduction

Regional results become increasingly important in agricultural research and policy making. The increased importance of rural development, cross compliance and directives like the water framework directive change the type of data needed for sound agricultural policy analysis. Furthermore policy makers are more and more interested in not only the average result of a group of farms, but especially in the distribution. For research on for example poverty in agriculture and risk management especially the tails of the distribution are of importance. Agricultural statistics on a macro level, such as supply balance sheets and the economic accounts of agriculture do not provide this type of information.

In agriculture, data from the Farm Accountancy Data Network (FADN) provides the most detailed and harmonised data on the economic performance of farms and are often used to estimate population characteristics. The use of FADN data in regional studies is often problematic due to the low number of observations. Several methods have been developed to use additional information to increase the reliability of estimates. (Dol, 1991; Baker et al. 1994). Additional information that can be used is for example the agricultural census. The agricultural census gives a complete list of the population of farms. The amount of information in this census is however limited. In this paper we will describe an option to make use of additional information from the census to make more reliable estimates in regional studies. The procedure has been implemented in the software tool Stars.

In section 2, the principles of data imputation will be explained. The selection of imputation variables is the weak link in data imputation. Therefore, in section 3 we propose a step wise procedure for the selection of imputation variables. Section 4 illustrates this approach by making estimates for dairy farmers in a small part of the Netherlands. The validity of the approach is discussed in section 5. The paper ends with some conclusions.

2. Description of data imputation

In a specific research project attention focuses on farms of a certain region, farms that belong to a certain type or a combination of both. We will call this group the population of interest or population in short. In the imputation procedure, for each farm in the population, a farm in the FADN sample is selected which resembles the farm as closely as possible. The researcher selects the variables, which are used to decide whether a farm resembles a sample farm. These variables are called the imputation variables. The imputation variables should be known for all farms in the sample and the population. Based on these variables the distance is calculated. Different methods are available to establish this distance. The sample farm with the smallest distance is regarded as the farm that resembles the population farm as closely as possible. For each farm in the population, 5 or 10 most similar farms are selected from the sample. These best fits are recorded together with the distance measures.

Based on these best fits, estimates can be made for a set of goal variables, which are known in the sample, but unknown for all population farms. In making estimations for the population of interest, a choice can be made between simple and multiple imputation. Simple imputation has the disadvantage that the variance of the estimator is underestimated. The estimated (e.g. imputed) value is treated as the real value, although there is a degree of uncertainty about this value. To overcome this problem multiple imputation can be used. In this option, the user can define how many of the best-fit farms will be used to make estimates about the population.

The approach is illustrated in Figure A1.1 and Figure A1.2. Figure A1.1 describes the traditional approach (see for example Cochran, 1977). The census describes the whole population (N units). Based on the population a stratified sample is drawn. Given the number of farms in the population and the sample, weighting factors per sample farm are calculated. A weighted average of the sample observations gives a good estimation of the population.

Figure A1.2 describes the data imputation approach. The same sample as in Figure A1 is the starting point. To make estimates of the population of interest (e.g. specific region), sample farms are matched to population farms based on the imputation variables. The sample farm that is most similar to a population farm is used to impute goal variables. The basic assumption is that if the farm is similar on the imputation characteristics, then it is likely that the farm is also similar on the goal variables. To assure that this is a valid assumption, the imputation variables have to be selected in a careful way (see section 3).

Figure A1.1 Direct estimation using weight of sample units

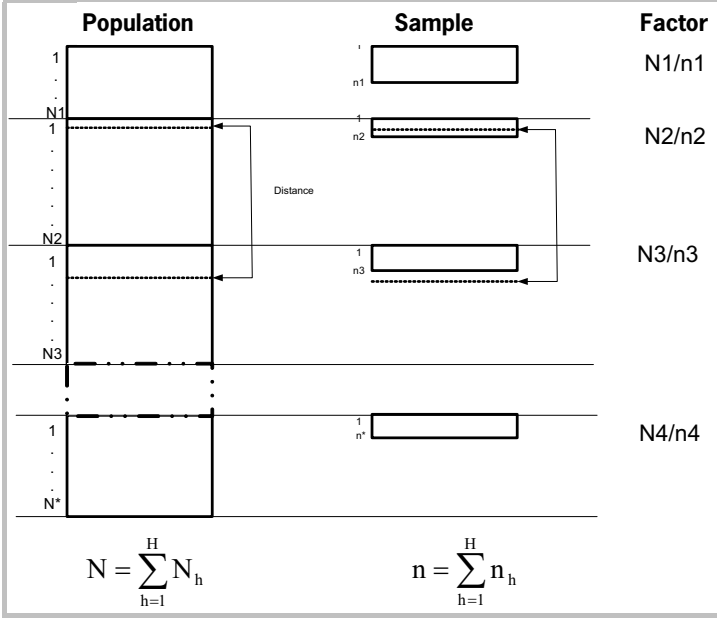


Figure A1.2 Data fusion

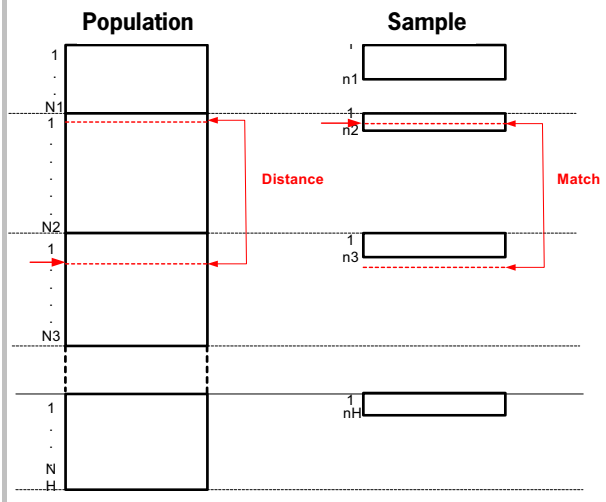


Figure A1.3 and Figure A1.4 show an illustration of both approaches. The relevance of these figures is not the exact value (different types of red), but only in the presence or absence of an estimate (green versus red). The color green indicates that it is not possible to make an estimation for that region, because no observations are available. A red color indicates that an estimation can be made (darker red indicates a higher value). The first figure shows the results of a direct estimate based on FADN data. In Figure A1.4 the same variable is estimated based on the data fusion procedure. The first figure clearly shows that there are no results for many regions because of the absence of sample farms in the regions. Obeying the requirement of publishing results based on at least 15 FADN farms would result in only a small number of regions for which an estimate could be made. Figure A1. 4 shows that after imputation an estimation can be made for all regions.

Figure A1.3 **Estimation of mean revenues based on direct estimation**

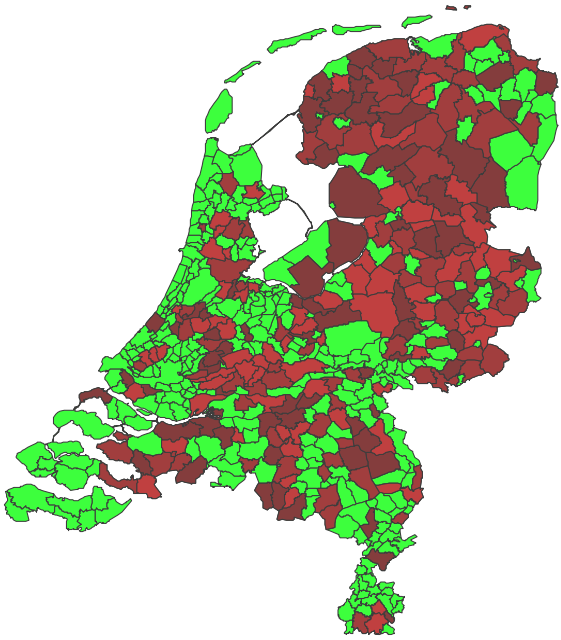
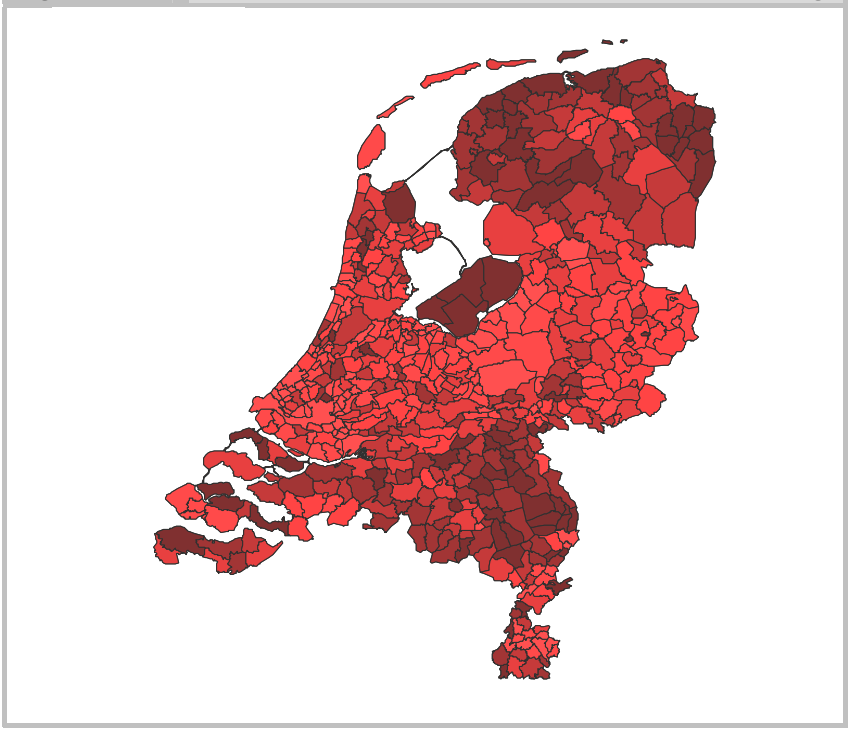


Figure A1.4 **Estimation of mean revenues based on statistical matching**



3. Stepwise selection of imputation variables

The selection of variables to calculate the distance between the population farm and the sample farm is an essential step. Simkin et al. (2004) describe the use of genetic algorithms to select the imputation variables. A clear disadvantage of this approach is the long calculation time and the possibility of local optima in the selection of the imputation variables. This section describes an alternative approach which is based on a stepwise selection procedure, similar to stepwise regression. The advantage of this method is the simplicity of the method and the speed of calculation. The steps to be performed are:

1 Select set of potential variables

2 Calculate measure of fit of solution based on each variable separately

3 Rank variables based on measure of fit

4 Start with highest ranked variable, calculate measure of fit

5 Add variable and calculate measure of fit and the improvement of fit

6 If measure of fit improves more than a certain threshold include variable in the set of imputation variables, go to step 5

7 End of selection of variables

This results in the set of imputation variables that can be used to apply the estimation process.

4. Estimating regional results of dairy farmers: an example

In this example we explore the opportunities to make estimations for dairy farms in a municipality in the northern part of the Netherlands (black area in Figure A1.5). In this example an estimate is made for the variables: total revenues, total costs, net farm result, labour income entrepreneur and number of entrepreneurs (these are the goal variables). Based on the number of observations in the FADN, it is difficult to make direct estimations. However, this municipality is part of a larger grassland area with similar production circumstances. This area, 'Noordelijk Weidegebied' (Northern Grassland Area), is one of the agricultural areas of the Netherlands (see grey area Figure A1.5). With data imputation it is possible to use the extra information from dairy farms in the larger region to make an estimation of the results of dairy farms in the specific municipality. In the FADN, 70 dairy farms from this region are included in the sample.

Figure A1.5 **Municipality of interest (black) in Northern Grassland Area (grey)**

In the estimation procedure a number of imputation variables is used:

- Age;
- Hectares grass;
- Hectares fodder crop;
- Number of dairy cows;
- Economic size.

In Table A1.1 the results of the imputation process are described. In this example a single imputation is applied. For each farm in the population in the municipality the most similar farm in the FADN sample in the Northern Grass Area is selected. The similarity is based on the 5 imputation variables as described above (to take into account the different units of measurement the variables are standardised before calculating the distance). Subsequently the average of the imputed values for all farms in the municipality are calculated, assuming that the values of the most similar farms in the Northern Grass Area provide a good approximation of the value of that specific farm.

Table A1.1	Results of imputation process (single imputation)	
Variable	**Mean**	**Standard error**
Revenues	415,020	15,028
Costs	506,479	15,103
Net farm result	-80,069	4,581
Labour income per entrepreneur	58,066	5,010
Number of entrepreneurs	1.47	0.05

Single imputation has the disadvantage of underestimating the variance. The imputed values for a specific farm are considered as the true values, although there is a certain uncertainty about these values. In Table A1.2 the results are displayed for a multiple imputation process. The 3 most similar farms are used to make an estimation for the municipality. In this multiple imputation process 100 independent replications are applied. In each replication one of the 3 nearest neighbours is randomly selected. The values of that neighbour are used to impute the values and make estimations for the region. Comparing Tables A1.1 and A1.2 shows that the estimations of the means are not very different. It also shows that the variance of the estimator increases due to the multiple imputation process. This increase is caused by the addition of between replication variance. The columns Min and Max show that the estimation of the average total revenues varies between 405 and 431 thousand. This variance is added to the variance as a consequence of differences between farms within a replication (within variance). The variance increases by 10% for the different goal variables.

Table A1.2	Results of imputation process (multiple imputation)			
Variable	**Mean**	**Standard error**	**Min**	**Max**
Revenues	417,203	16,723	405,002	431,081
Costs	505,405	16,354	492,738	521,129
Net farm result	-76,984	5,502	-85,138	-69,606
Labour income per entrepreneur	63,899	6,459	56,126	75,055
Number of entrepreneurs	1.49	0.05	1.4	1.6

5. Validity of results

In the previous section the quality of the imputation process was not explicitly considered. In this section a validation procedure is described. The quality can be judged by using the same approach for imputing values in the sample (which are known) under the restriction that the farm itself cannot be used to impute values. In this way the values of a sample farm are estimated by imputing values from one or more other sample farms that are very similar. Subsequently a statistical test can be conducted to check whether significant differences exist between the real values and the imputed values.

Table A1.3	**potential imputation variables**
Age	Percentage other grazing livestock
Hectare	Percentage breeding pigs
Hectare grass	Percentage fattening pigs
Hectare fodder crops	Percentage poultry
Dairy cows	Percentage fodder crops
Dairy cows per hectare	Percentage grains
Total added value	Percentage tuberous plants
Added value pigs	Percentage other arable farming
Percentage dairy cows	Percentage horticulture open air

Table A1.3 lists all the variables that could be used as imputation variables. The inclusion of variables as imputation variables is only useful when there is some kind of logical relationship between this variable and the goal variables. Unlike regression analysis no assumption has to be made about the shape of the relation. In Table A1.4 a naïve approach has been applied in which all potential imputation variables have been used. This table shows that the values estimated by the imputation procedure are close to the real values. No significant differences can be shown by looking at the averages and the standard errors.

Table A1.4	**Comparison of real and estimated values**		
Variable	**Real value**	**Estimated value**	**Standard error**
Revenues	476,902	493,360	32,869
Costs	569,488	573,109	33,472
Net farm result	-79,303	-66,473	9,536
Labour income per entrepreneur	67,817	80,157	11,858
Number of entrepreneurs	1.53	1.49	0.09

An important question is whether all imputation variables are relevant in the imputation process. A balance has to be found between the correctness of the model and the simplicity of the model. In Table A1.5 an extreme variant is applied in which the distance is only based on the age of the farmer and the hectares of grassland. This table shows large and significant differences between the estimated and real values. Based on this analysis the conclusion can be drawn that data imputation based on only these two variables result in a low quality.

Table A1.5 Imputation based on age and hectares of grassland

Variable	Real value	Estimated value	Standard error
Revenues	476,902	355,033	21,028
Costs	569,488	459,701	14,797
Net farm result	-79,303	-91,233	9,601
Labour income per entrepreneur	67,817	12,530	10,507
Number of entrepreneurs	1.53	1	0

In Table A1.6 the results for an imputation procedure based on 5 imputation variables is described. This table shows that the results are equally good or even better compared to an imputation procedure based on all imputation variables.

Table A1.6 Imputation based on age, hectares of grass, hectares of fodder crops, number of dairy cows and economic size

Variable	Real value	Estimated value	Standard error
Revenues	476,902	470,917	34,330
Costs	569,488	560,114	33,836
Net farm result	-79,303	-76,492	9,182
Labour income per entrepreneur	67,817	68,500	11,297
Number of entrepreneurs	1.53	1.53	0.09

This approach provides the advantage that the basic assumption of the imputation process can be tested. Besides theoretical reasons, a quantitative analysis can provide support for the choice of the imputation variables.

6. Summary and discussion

Using existing survey data in regional studies leads to several problems. The survey was often not designed for that type of research. A practical problem is that the number of observations is often to limited to make reliable estimations.

Different methods have been developed to use additional information that facilitates more reliable estimates. One of these methods is data imputation. In this paper the use of data imputation is described in which information from the agricultural census is used besides FADN data. An example is described in which more reliable estimates are made of the economic performance of dairy farms in a small region in the Netherlands.

The selection of imputation variables is of crucial importance to the quality of the end result. A theoretical model about the impact of these variables on the goal variables should be the basis for the selection. This implies that a general set of imputation variables won't work; these variables should be selected based on the characteristics of an individual research project. Quantitative analysis based on the available sample data should be performed to test the quality of the imputation process.

References

Baker, Ken, Paul Harris en John O'Brien. 'Data Fusion: An Appraisal and Experimental Evaluation.' In *Journal of The Market Research Society*, 31 (2),152-212, 1994.

Cochran, W.G.C., *Sampling Techniques*, Wiley, New York, 1977.

Dol, W., *Small area estimation, a synthesis between sampling theory and econometrics*, Wolters Noordhoff, Groningen, 1991

Ford, B., 'An Overview of Hot-Deck Procedures.' In: *Incomplete Data In Sample Surveys*, Academic Press, Volume II, Part 2, 185-207, 1983.

Simkin, S., T. Verwaart and H.C.J. Vrolijk, 'Application of a Genetic Algorithm to Nearest Neighbour Classification.' In: Moonis Ali, Floriana Esposito, *Lecture Notes in Artificial Intelligence*, Vol. 3533/2005, Springer Verlag, pp. 544-546.

Vrolijk, H.C.J. and M. Wedel (1996), 'Een Datafusie-procedure voor het maken van kruistabellen.' In: *Jaarboek van de Vereniging voor Marktonderzoek en Informatiemanagement*, Uitgeverij de Vrieseborch, Haarlem, pp. 95-106, 2005.

LEI Wageningen UR develops economic expertise for government bodies and industry in the field of food, agriculture and the natural environment. By means of independent research, LEI Wageningen UR offers its customers a solid basis for socially and strategically justifiable policy choices.

LEI is part of Wageningen University and Research Centre, forming the Social Sciences Group with the department of Social Sciences.

More information: www.lei.wur.nl

LEI report 2009-067