

Business Statistics and National Accounts – What Strategy and Methods should we adopt?

Anders Wallgren, Britt Wallgren

Formerly of the Department of Research and Development at Statistics Sweden

ba.statistik@telia.com

Abstract

In our view, sample surveys of populations of persons dominate survey theory and methodological thinking. At Statistics Sweden, statistics on persons follows high methodological standards and both register surveys and sample surveys produce high quality statistics. The microdata from these surveys are consistent and the estimates are coherent with some exceptions. Many qualified methodologists have contributed to our Labour Force Survey.

This is in sharp contrast with economic statistics. Coverage errors and other systematic errors cause that those responsible for the National Accounts get estimates that are biased and that estimates from different surveys are inconsistent. Many qualified methodologists prefer to do research on sample surveys of persons and almost no research has been done regarding statistical methods for National Accounts. Thus, we have both methodological problems and problems with attitudes regarding economic statistics.

In the following table, we compare social statistics (or statistics on persons) with economic statistics (or business statistics) – the table defines the main methodological problems we must solve if we want to improve economic statistics.

Chart 1. Methodological problems in social statistics on persons and in economic statistics

	Social statistics	Economic statistics
1. Main variables	Stock variables that exist for points in time, i.e. age, sex	Flow variables that exist only for periods (e.g. turnover, value added)
2. Definition of target population	All units that existed at a specific point in time	All units that existed during any part of a specific period (year, quarter, month)
3. Permanence of population units	Persons are the same units during their lifetime, only their variable values change	Units can merge with other units or split into new units during the period
4. Equality of population units	All persons are of equal interest	A few big units are of main interest, many small units are sometimes disregarded (cut-off sampling)
5. Main classifications used in many surveys	Very simple: Sex, age, residence, country of birth	Very complicated: Industry, institutional sector
6. Combining sample surveys and registers	Frame populations and register populations are very similar	Frame populations and register populations are created at different points in time and differ considerably in coverage
7. Geo-referencing of units	Requires population register with residential addresses	Requires legislation, so that all employers with more than one local unit give information on where each employee works.
8. Developing countries	Information on migration can be missing	Informal sector requires combinations of registers and area samples

In our view, these methodological problems for economic statistics are not handled in a sufficiently good way at Statistics Sweden and most countries have similar problems. In this paper, we will discuss the issues above for economic statistics and give our suggestions on how to tackle these problems. Our focus is on administrative registers and the combination of sample surveys and statistical registers or register surveys. How should the Business Register be designed and used?

1. The National Accounts Survey

The main purpose of all business surveys is to give aggregated data to the National Accounts. Based on all these estimates from business surveys and some other surveys, for instance the Labour Force Survey, the National Accounts Survey produces estimates of the parameters that constitute the national accounting system. The business surveys can be censuses, sample surveys or register surveys describing different parts of the non-financial, financial and public sectors.

When we discuss methods for creating and maintaining the statistical Business Register, everything we say has direct implications for the National Accounts. This fact is highlighted in Chart 2, which has been used by managers at Statistics Sweden to explain the importance of the Business Register and to explain that all economic surveys must be coordinated to support the National Accounts. However, the coordination has not been successful, the system in Chart 2 is not an integrated system of surveys: frame populations have coverage errors that vary over time, the economic register surveys have register populations that have not been created with the Business Register and they are therefore not consistent with the sample surveys. In Chart 3 some of these inconsistencies are illustrated for surveys that are relevant for the employment part of the National Accounts.

Chart 2. The system of business statistics (economic statistics)

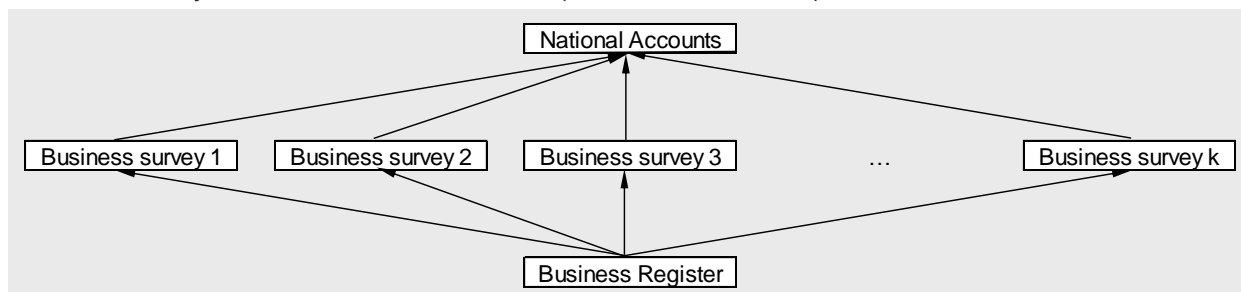


Chart 3. Employees by industry, November 2004, thousands¹

Economic activity	Business Register		Employment Register	Labour Force Survey	Error margin
	Enterprises	Establishments			
Agriculture, forestry and fishing	35	37	37	26	5
Mining, quarrying and manufacturing	688	636	717	640	23
Electricity, gas and water	21	22	28	29	5
Construction	197	209	215	199	14
Wholesale and retail trade	456	453	484	456	20
Hotels and restaurants	89	93	99	106	10
Transport, communication	240	242	243	236	15
Financial intermediation	83	77	85	78	9
Real estate, business activities	457	524	457	470	20
Government	139	215	239	230	15
Education	382	408	431	462	20
Health and social work	836	684	675	675	24
Other service activities	142	163	175	168	13
Unknown activity	0	0	38	4	
Total	3 763	3 763	3 924	3 778	43

How does National Accounts use business statistics?

Chart 3 illustrates the dilemma that those responsible for the National Accounts face. How to choose between sources that give conflicting estimates for the same parameters? The problem is usually “solved” by choosing the best source for each parameter: the source that is considered to be the best on agriculture is used for the first estimate, the best source for manufacturing is used for the second estimate, and so on. This sub-optimisation method leads to inconsistent subpopulations – the Business Register is not used as it should be used. Chart 4 illustrates this method and its consequences. The estimates for sectors 1 + 2 + 6, 3, 4 and 7 must be adjusted by “smearing” the residual on these four estimates so that they sum to 1 042 011. When all business surveys deliver their macrodata to the National Accounts, also all errors and problems are handed

¹ Based on Chart 1.2 in Wallgren and Wallgren (2014)

over and those working with the National Accounts are responsible for handling all these problems, as a rule without support from methodologists.

Chart 4. Inconsistent sources with wage sums in the National Accounts 2004²
Wage sums according to the definitions of the National Accounts, SEK millions

Wage sums 2004, total sum for the whole economy according to the best administrative source		1 042 011
<i>Wage sums from different sources:</i>		
	<i>Sector</i>	<i>Wage sum</i>
Enterprises and sole traders	1 + 2 + 6	697 361
Central government	3	69 826
Local government, municipalities	4	231 897
Non-profit organisations	7	24 034
Sum of these sources:		1 023 118
Residual = 1 042 011 – 1 023 118 =		18 893

Those who work with the National Accounts are aware of that the estimates they get from the economic surveys have systematic errors. Many systematic errors are generated by coverage errors including cut-off sampling. A common approach is to disregard the levels of the estimates from business surveys and only use the relative change between the last estimate and the previous estimate. These estimates of change in percent are used to update old levels defined by those who worked with the National Accounts many years earlier. This way of using microdata from business surveys is not an optimal way of using available statistics. In Section 2-3 we discuss how the Business Register and the survey methods can be improved to reduce these inconsistencies.

The system of social statistics or statistics on persons at Statistics Sweden is quite different. Within this system, populations are almost perfectly consistent and all variables that are based on the Population Register and other registers are consistent at the micro level. To obtain this, the work has been organised in a good way and good methods have been developed. Chart 5 illustrates this – those who work with National Accounts deserve to get macrodata of similar quality.

Chart 4. The system of social statistics, statistics on persons

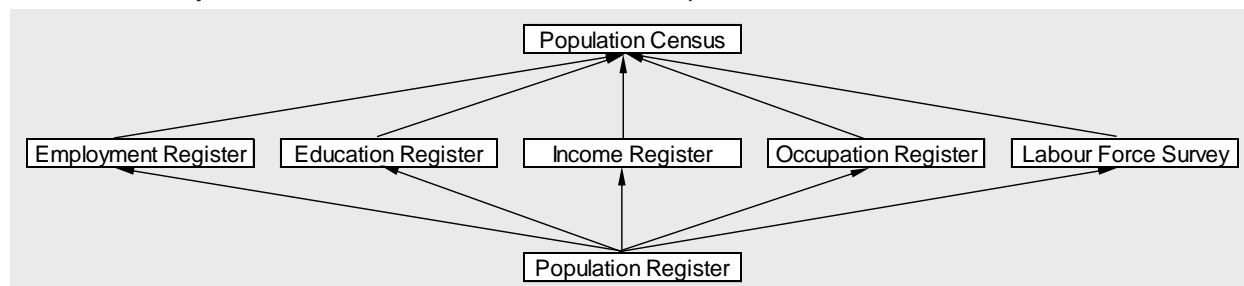


Chart 5. Estimates of number of persons from four statistical registers for one municipality 2010³

Population Register		Employment Register		Education Register					Income Register Yearly earned, SEK thousands			
Age	Number	Employed	Not employed	Com-pulsory	Upper secondary	Post-secondary	Post-graduate	Not known	0	1–139	140–279	280–
0–15	1 358	-	-	-	-	-	-	-	-	-	-	-
16–19	384	84	300	297	67	0	0	20	120	259	5	0
20–24	328	197	131	55	233	31	0	9	38	153	105	32
25–34	719	622	97	88	423	201	1	6	15	141	320	243
35–44	962	846	116	83	554	319	2	4	19	106	381	456
45–54	910	776	134	171	500	226	4	9	24	89	363	434
55–64	1 071	791	280	295	491	271	7	7	21	145	468	437
65+	1 402	138	1 264	758	438	194	8	4	1	508	719	174

² Based on Statistics Sweden (2007)

³ Based on Chart 4.11 in Wallgren and Wallgren (2014)

The inconsistencies and errors described above are caused by problems that we have in economic statistics but not in statistics on persons. In Statistics Sweden (2001), (2005) and (2007) we report issues that result in inconsistent economic statistics:

- People do not understand the importance of coordination and cooperation. “My hotel register is much better than the hotel part of the Business Register”. Instead of improving the Business Register with this information, inconsistencies are created.
- A decentralised statistical system does not support cooperation and consistence. The Swedish Energy Agency is responsible for energy statistics and one unit at Statistics Sweden works on commission with energy surveys for this agency. The unit had their own register with energy enterprises that was not coordinated with Statistics Sweden’s Business Register. “We work for the Energy Agency and do as they tell us to do.”
- The Board of Agriculture is responsible for all agricultural surveys in Sweden. The Board of Agriculture and Statistics Sweden work together and compare the farm populations in Statistics Sweden’s Business Register with the register used by the Board of Agriculture. Inconsistencies are found, but estimates are not coordinated. The Board’s register is used for the Farm Structure Survey (FSS), the Business Register is used for the Structural Business Statistics survey (SBS). The National Accounts use FSS for agriculture and SBS for other industries. Therefore, some enterprises within ISIC 01 in the Business Register are not included in the GDP estimates. Statistics Canada has been able to create a Business Register that includes the Farm Register. It will then be easier to obtain consistency.
- Many economic surveys cover different parts of the same calendar year – the time lag between the earliest survey and the last is almost two years. The frame errors in the early sample surveys could be analysed with information available when all surveys have been done and revised estimates could be produced, but this is not done. There is time for such revisions – the work with definite accounts for year t starts during spring year $t+2$.

These examples show that the National Accounts receive conflicting information due to wrong attitudes and lack of understanding among the staff. But also, new estimation methods are required. Today we use microdata for each survey alone and produce estimates that often are not coherent. Instead we must combine microdata from different surveys (e.g. the FSS and SBS mentioned above) and produce estimates based on the combined set of data. Our conclusion is: if we want to improve the National Accounts Survey we must improve the Business Register, the economic surveys and the estimation methods that are used. All these improvements must be considered together – the whole survey system must be improved.

The basic method that is used for the yearly National Accounts is to create supply and use tables for about 400 products. For one product, the sources are illustrated in Chart 6. We found that eleven surveys ($S_1 - S_{11}$) and six models ($M_1 - M_6$) were used. Thirteen parameters should be estimated, but four of these are assumed = 0 for this specific product. So, in all we have about $400 \cdot 13 = 5\ 200$ parameters that are estimated with these supply and use tables. The preliminary estimates are inconsistent and the 400 residuals Supply–Use give important information regarding the uncertainty of National Accounts. Other important quality measures are the amounts that are estimated with models instead of survey data – these amounts show the lack of information.

Chart 6. Supply and use table for one product

Supply								
Values in:	Production	Import	Customs	Trade margin	Product tax	Product subsidies	VAT	Total supply in purchasers' prices
Current prices	$S_1 S_2 S_3$	$S_4 S_5 S_6 S_7 M_1$	0	M_2	0	0	M_3	
Use								
Values in:	Intermediate consumption	Public consumption	Private consumption	Gross fixed investment	Changes in inventories	Export	Residual Supply-Use	Total use
Current prices	$S_2 S_8 M_4 M_5$	0	$S_9 S_{10}$	$M_6 S_2 S_{11}$	S_3	$S_4 S_5 S_6 M_1$		

2. The Business Register, the base register for economic statistics

The statistical Business Register should be the base register for all business surveys that are used for economic statistics and National Accounts. As a base register, it should have the following characteristics and be used and supported by all those who work with economic statistics:

- *Standardised statistical units:*
The Business Register should define important statistical units and include identity numbers that are links between different kinds of statistical units in the Business Register and related registers.
- *Standardised populations:*
The Business Register should be the master sampling frame used for all economic sample surveys and should be used to create all register populations in all economic register surveys. The Business Register should also be responsible for creating the standardised populations used for the National Accounts.
- *Standardised variables:*
The Business Register should be responsible for creating and maintaining the basic classification variables used for economic statistics:
 - Industry
 - Institutional sector
 - Geographical codes for establishments

The national statistical institute must use administrative registers as sources to be able to create and maintain the statistical Business Register. The structure of the administrative systems that generate these administrative sources must be understood by the statisticians. These systems give us both opportunities and quality problems. We give here a short description of the Swedish administrative systems that are used for the statistical Business Register.

Once a business has been registered with the Swedish Companies Registration Office, the business must be registered with the Swedish Tax Agency for taxation purposes. If a business is liable for value added tax (VAT), it is required to register for VAT and if a business has employees, it is required to register as an employer for Pay As You Earn (PAYE) and deliver monthly payroll reports and end of year reports with income statements.

An income statement is an income report for individuals issued by anyone who pays e.g. salary, pension, interest or dividends. Income statements play an essential role in Swedish society and in the Swedish tax system as they are the very basis on which the decision of the income tax each citizen should pay is made. Employers operating in more than one establishment should specify establishment number in the income statements. The data on establishment numbers are retrieved for Statistics Sweden and is used for statistics on employment and payrolls.

All systems for tax registration and tax payments generate data for *legal units*. Both the systems for *tax registration* and the systems for *tax reports and payments* can be used for the statistical Business Register. The coverage will differ: registration means that a legal unit *may* be active, a tax return with non-zero values means that a legal unit *has* been active. The system with income statements is the only administrative source that can be used to support a statistical register with *establishments*.

The national statistical institutes in the Nordic countries have spent long-term efforts on improving the national administrative systems. This requires cooperation with politicians, government agencies and municipalities. It has required parliament decisions and other improvements to replace traditional censuses with statistical registers. New legislation was required so that employers operating in more than one establishment specify establishment numbers in the income statements. Without this legislation, it would not be possible to produce regional employment statistics or to develop and maintain the establishment part of the Business Register.

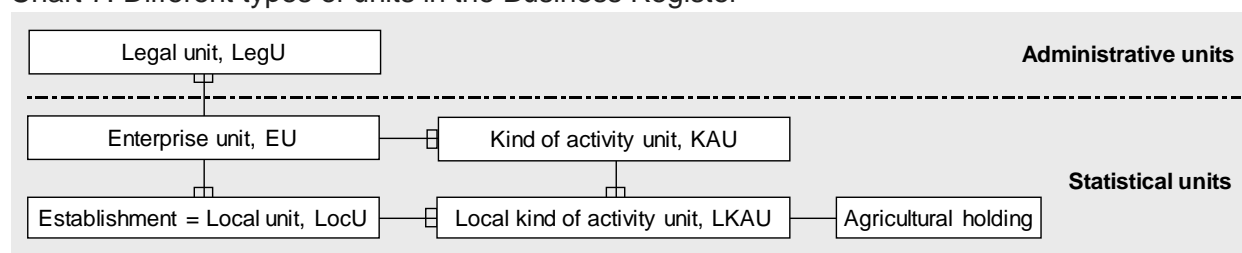
2.1 The units in the Business Register

Can we live with a single type of unit in business statistics?

According to our opinion we need the five types of units that are shown in Chart 7. Statistics Sweden has create five versions of the Business Register, one for each type of unit.

- The *legal units* (LegU) we must have, as we must use the important data-rich administrative sources with business data, and these sources contain data regarding legal units. Statistics Sweden’s Business Register consists of about 1 200 000 legal units, of these about 620 000 are sole traders and about 580 000 are enterprises.
- If we want to produce regional employment statistics and regional economic statistics, we need an *establishment* version. Statistics Sweden’s Business Register consists of about 1 300 000 local units or establishments (LocU).
- If we want to produce economic statistics by industry, we must also create a register version with *kind of activity units* (KAU). The variable industry should be grouped according to the requirements of the yearly National Accounts. About 170 KAUs have been created by splitting enterprise units (EU) into homogenous parts by industry.
- Finally, if we want regional statistics by industry we need a register version with *local kind of activity units* (LKAU). About 45 establishments have been split into 120 LKAUs by industry. The Farm Register should be an integrated part of the Business Register and the Farm Register should consist of the local kind of activity units with ISIC = 01. These units correspond to the *holdings* in agricultural statistics.

Chart 7. Different types of units in the Business Register



How is the unit problem handled when linking data sources?

When different administrative sources with economic data are linked, we find that legal units are not suitable as statistical units because the combined data set has records with conflicting variable values. Many legal units are related to other legal units through ownership patterns and these relations have a great impact on tax reporting. In Chart 8, three sources have been matched:

BIN = Business identity number of each legal unit, LegU

SBS = Turnover according to Statistics Sweden’s questionnaire

YIT = Turnover according to the yearly income tax returns

VAT = Turnover according to 12 monthly VAT returns

Chart 8. Yearly turnover for the same legal units in three sources, USD million

8A. Incomplete composite enterprise units				8B. Complete composite enterprise units			
BIN	SBS	YIT	VAT	BIN	SBS	YIT	VAT
160001	2 301	0	0	Enterprise unit 1	160001	2 301	0
160002	2 211	0	2 239	Enterprise unit 1	170001	-	2 301
160003	5 514	2 888	2 895	Enterprise unit 2	160002	2 211	0
				Enterprise unit 2	170002	-	2 211
				Enterprise unit 3	160003	5 514	2 888
				Enterprise unit 3	170003	-	2 626
							2 895
							2 622

This problem is handled by creating Enterprise units (EU), in this case three EUs are created from six LegUs.

As a rule, the unit problem is not handled at all when data sources are linked, only about 40 complex enterprise units have been created by Statistics Sweden.

Behind each legal unit in Chart 8A there is a ‘family’ of related legal units. The person at the legal unit 160001 received a questionnaire (SBS) from Statistics Sweden and responded for the whole ‘family’. But the tax reporting (YIT and VAT) within the ‘family’ was handled by another legal unit (170001), which is clear from Chart 8B.⁴

Behind the legal units shown in Chart 8, there are three different populations consisting of different kinds of units that use the same system of identity numbers, BIN:

- Those who received the questionnaire – Statistics Sweden’s *reporting units*
- Those who report yearly income tax returns – *income tax return units*
- Those who report VAT – *VAT return units*

When we match these sources, we obtain correct matches of legal units, but statistically erroneous matches with inconsistent variable values. This is a difficult issue in economic statistics – how to achieve consistent estimates from different surveys based on data for legal units? Chart 9 illustrates that the VAT-system and the yearly income declarations have different populations of legal units. Almost the same “population” of turnover-money is allocated between different sets of legal units.

The unit problems we can see in charts 8 and 9 are not handled in a systematic way. One person produces statistics with the VAT source, another person produces statistics with the income declarations. When they produce estimates by industry, the estimates will be inconsistent as the legal units in the two populations can differ regarding industry, ISIC. The inconsistent estimates are handed over to the National Accounts.

Chart 9. Each administrative system has its own object set⁵

LegU-id	Turnover, SEK million	
	VAT	Income declaration
LegU1	31	2 161
LegU2	870	2 990
LegU3	2 084	0
LegU4	2 043	0
LegU5	0	2 040
LegU6	2 036	0
LegU7	3 998	5 939
LegU8	0	1 934
LegU9	2 558	642

The large differences between the two turnover values for these legal units should not be interpreted as measurement errors or tax evasion.

When we see a zero value in one source and many millions in the other source, this should also not be interpreted as a missing value in the first source. We should not impute values in such cases. Instead we should understand that this is typical for economic data from administrative units.

Income declaration	VAT-report	Number of legal units in each category	Turnover SEK million for each category	
			Income declaration	VAT-report
1. Turnover = 0	Turnover not 0	110 826	0	674
2. Turnover not 0	Turnover not 0	696 205	6 486	6 240
3. Turnover not 0	Turnover = 0	116 770	242	0
Total turnover:			6 728	6 914

Legal units belonging to big businesses are almost always included in this kind of enterprise ‘families’, or to use the correct term, they almost always belong to *composite enterprise units* together with other legal units. All legal units belonging to the same family should be combined into one enterprise unit and the variable values are aggregates of the values for the legal units. In Statistic Sweden’s Business Register about 1 300 legal units are used to create about 40 composite enterprise units, EU. A register with enterprise groups is available to support this profiling work. These 40 enterprise units are then split into kind of activity units (KAU) by economic activities.

Problems and opportunities related to the units in the Business Register

The *legal units* give us the opportunity to use all available *administrative data* related to business enterprises and employers. It is also possible to obtain *good coverage* of the business population by using available information in the administrative systems for registration and tax reporting. In developed countries like Sweden, where everyone must produce tax returns, these systems have

⁴ Based on Section 6.2.3 in Wallgren and Wallgren (2014)

⁵ Based on Section 7.4.2 (op. cit.)

very good coverage of the population of legal units we want to have in the Business Register. In developing countries, there can be an informal sector that is not covered by taxation systems. For such countries, we recommend that administrative registers for the formal sector are combined with area samples of households that cover both the formal and informal sectors. To avoid double-counting, it is necessary to ask for identity numbers in the household surveys and use administrative registers from social security systems with employees in the formal sector.⁶

To maintain the register with *establishments*, administrative data with income statements are used. Every year, about 8 000 enterprises that have reported that they have more than one establishment get a questionnaire with the intention to update and maintain the quality of the establishment register. The quality of the establishment register is also improved by careful editing of the income statements. For each employee, the geographical codes for the person's home are compared with the codes for the person's place of work, according to the income statement. Enterprises with groups of employees with strange patterns are contacted. However, Statistics Sweden suspects that there is still some undercoverage in the establishment register.

To create the 40 *establishment units*, the 170 *kind of activity units* and the 120 *local kind of activity units* mentioned above, requires work done by Statistics Sweden's profiling group. This means that the registers with these kinds of units are expensive to maintain, and economic statistics based on these units requires data collection and a substantial response burden for the companies involved. The quality problems that these units should solve are only partially solved – there are many more EUs, KAUs and LKAUs hidden in the Business Register that generate inconsistencies. In Section 2.3 we discuss the possibilities to create a Business Register with *derived* KAUs and LKAUs to improve the quality of estimates by industry.

2.2 The Business Register – frames and register populations

The Business Register plays a quite different role in the Swedish statistical system than the Population Register. The statistical registers in the system of social statistics in charts 4 and 5 have been created in close cooperation with the Population Register. They are all advanced registers – they are based on many sources and require complex register processing. The Population Register functions as a base register for the statistical registers used for statistics on persons and the consistence is perfect.

Parallel to this, the Population Register also functions as the master sampling frame for all sample surveys of persons and these frames are consistent with the register populations, the only difference is that the frames and register populations are created at different points in time. This difference can be small – if a frame is created with the current stock of the Population Register at time t , after about one month the Population Register has been updated so that the population at time t is known. The sample surveys of persons use advanced estimation methods where register variables are used for GREG-estimation and calibration of weights.

The Business Register is used as a master sampling frame for Statistics Sweden's business sample surveys. But it is not used as a base register for statistical registers with economic statistics. No advanced registers with economic data have been created, instead the administrative sources after editing are used one by one as they are. The sample surveys of enterprises do not use register variables to improve estimation or to adjust sampling weights, except for stratification⁷.

The administrative register populations and the frame populations for business statistics also differ considerably regarding coverage. There are three reasons behind these differences:

⁶ Our experiences from Latin America are described in Wallgren and Wallgren (2015), (2016a) and (2016b).

⁷ Why is social statistics at Statistics Sweden advanced and of high quality as compared with business surveys? We think that the development of social statistics in the Nordic countries has been stimulated by a grand vision "*We must get rid of the traditional census!*" This long-term vision led to the development of the advanced registers in Chart 4 and these registers stimulated the development of advanced estimation methods for sample surveys. There has been no similar grand vision for economic statistics and that is why economic statistics is lagging behind.

- The sampling frames are created with the current stock of the Business Register at a specific *point* in time. The register populations in the administrative sources describe all legal units that have been active reporting for a specific *period* (year, quarter or month).
- The Business Register at Statistics Sweden is updated with the information regarding *registration* of legal units, these units *may* be active. The administrative sources contain data regarding units that *have* been active.
- The long time lag between the frames and the administrative sources for the same periods. The first sampling frame for year t is created during November year $t-1$, the last administrative source for year t is delivered during November year $t+1$.

Chart 10 illustrates the time lag between sampling frames and deliveries of administrative data for the calendar year 2004. In November 2004, the current stock of enterprises *registered* as active during November 2004 defined the sampling frame that should be used for yearly statistics regarding 2004. This sampling frame was also used for the SBS 2004 which was a combination of a questionnaire to some big enterprises and administrative data with income declarations for all other enterprises in the frame, plus a sample survey used to collect additional variables on investments etc. The deliveries of administrative data are shown under the time axis, the last delivery for 2004 came during the end of 2005. If the administrative sources for 2004 are compared with the November frame, over- and undercoverage in the frame will be found in all comparisons.

Early 2006, we created the calendar year register for 2004 with all legal units that *had been active* with at least one non-zero tax report in at least one administrative system for 2004. All available administrative sources were used for this calendar year register.

Chart 10. The November frame and the Calendar Year Register 2004⁸

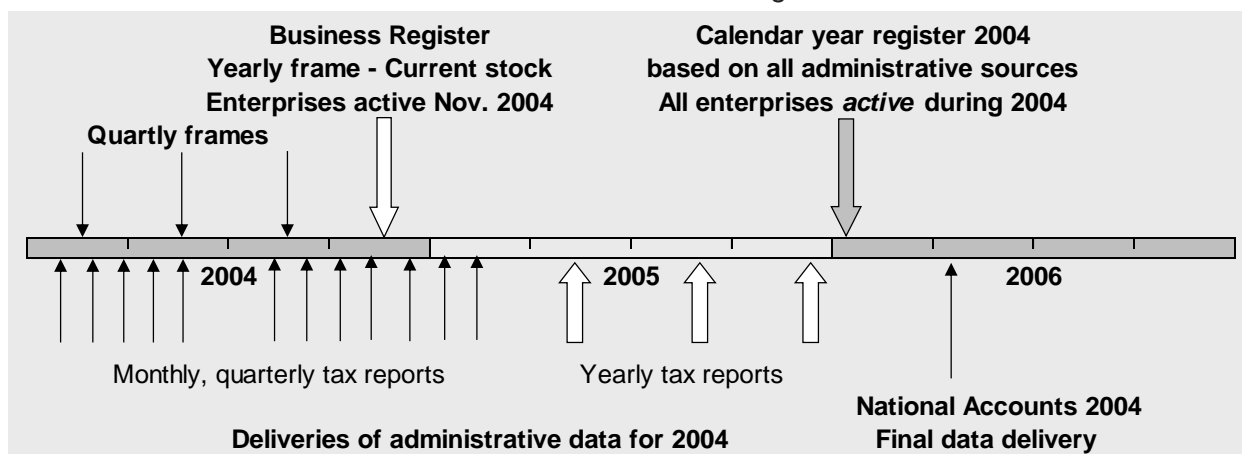


Chart 11. Differences between a sampling frame and a calendar year register

Sampling frame, November 2004	Calendar year register for 2004
Created during November 2004	Created during January 2006
Completely based on registration data: Legal units registered as active by the Tax Agency	Completely based on activity data: Tax reports for legal units combined with real tax payments
All legal units that were registered as active during the week in November 2004 when the frame was created. Units active only during January-October or December 2004 were excluded	All legal units that had been active with at least one kind of tax report and payment regarding any part of 2004. Monthly, quarterly and yearly reports are used
Suitable as sampling frame to be used during January-March 2005. Not suitable as register population due to undercoverage compared with all administrative sources for 2004	Suitable as register population, coverage suitable for all administrative sources. Not suitable as sampling frame because it is created too late

⁸ Charts 10 and 12 are based on charts 7.10 and 7.11 in Wallgren and Wallgren (2014)

Chart 12 illustrates the coverage errors in the sampling frame for November 2004 and in the Structural Business Statistics survey (SBS) for 2004. The chart highlights the following problems:

How should we create the statistical Business Register? What kinds of register versions should be created and how should we use the statistical Business Register?

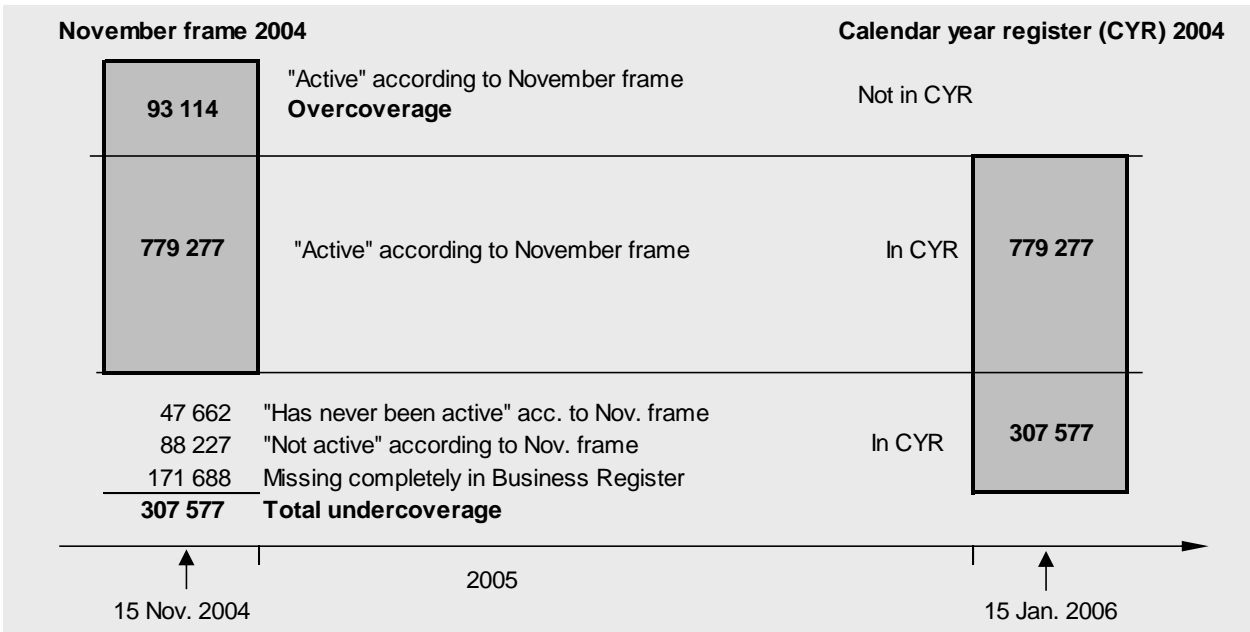
The first thing to be noted is the low quality of registration status in the Tax Agency’s register. As many as 93 114 legal units are classified as “active”, but they did not appear in any administrative register for 2004: no non-zero VAT return, payroll, export/import, income statements or income declaration had been delivered to the Tax Agency for these units. There are also 307 577 legal units that had been active during 2004 with tax reports and payments, but they are not classified as “active” in the November frame.⁹

The next thing to note is that there is plenty of time to create a register with good coverage that can be used by the National Accounts as standardised population for the yearly accounts of 2004. The work with these accounts started during spring 2006. The SBS is mainly based on the register of yearly income declarations for legal units. This register is delivered to Statistics Sweden during the autumn after the calendar year, in our case the autumn of 2005. The SBS used the sampling frame for November 2004 as register population, which means that the register population was created almost one year before that the administrative register was delivered. We think that this is an example where a sampling paradigm has dominated. Instead, if the register population had been created after the delivery of the administrative register it would have been possible to create a much better register population as there is more and better information available at that time.

How should we create the statistical Business Register? This example shows that we should give priority to activity data before registration data. This will reduce the coverage errors.

What kinds of register versions should be created and how should we use the statistical Business Register? Sample surveys need sampling frames. As these frames are created early before data collection, it may be necessary to use registration data. But also here, the quality of the frames can be improved by using short-term administrative registers as monthly pay-rolls and VAT. Register surveys need register populations created later and activity data should be used for these. The yearly National Accounts should be based on a calendar year register that should be used as a standardised population for all sample surveys and register surveys used by the National Accounts.

Chart 12. Overcoverage and undercoverage as number of legal units in the November frame



⁹ The size of the undercoverage is about 2% of total turnover and up to 15% for some industries

2.3 The Business Register – classification variables: industry

In this section, we discuss the variable *industry* – probably the variable in the national statistical system that generates most trouble of all¹⁰. Biases, inconsistencies, time series level shifts and sometimes panic¹¹ are related with this very important variable. We discuss first how *industry* is handled by Statistics Sweden and then we explain an alternative that we think is better.

The *Industry database table* in Chart 13 is used to determine 1st ISIC. In the KAU database some legal units are split into kind of activity units, KAU. For these KAUs, it is necessary to collect data as there is no administrative data available. The estimates from the register or census are biased. *Forestry* (ISIC 02) and *Cheese production* (ISIC 10) are missing completely, estimates = 0.

Chart 13. The variable *industry* in Statistics Sweden’s registers and surveys

Business Register				Register survey or census of KAUs				Estimates		
1. Legal units database table		3. KAU database table								
LegU idno	1 st ISIC	KAU idno	1 st ISIC	LegU idno	KAU idno	1 st ISIC	Payroll	ISIC	Payroll	
1	01	11	01	1	11	01	423	01	423	
2	16	21	16	2	21	16	375	16	375	
3	20	31	20	3	31	20	6 600	20	6 600	
4	24	32	24	3	32	24	6 184	24	9 640	
		41	24	4	41	24	3 456	Total	17 038	
2. Industry database table				Sample survey of KAUs						
Combination objects = LegU • Industry										
LegU idno	ISIC	Rank	Per cent	KAU idno	1 st ISIC	Payroll	weights	P • w	ISIC	Payroll
1	01	1	50	21	16	375	2.5	938	16	938
1	02	2	30	41	24	3 456	2.5	8 640	24	8 640
1	10	3	20						Total	9 578
2	16	1	60							
2	02	2	40							
3	20	1	51							
3	24	2	49							
4	24	1	90							
4	10	2	10							

In Chart 14, the variable *industry* is used in a different way. In Chapter 14 in Wallgren and Wallgren (2007) we argue that multi-valued variables must be treated with estimation methods using weights for the different categories of industry where each legal unit is active as producer. Even if the ISIC-weights are not perfect, the estimates are less biased as compared with Chart 13. The estimate 277 for *Forestry* can have an error, but is better than the estimate 0 in Chart 13.

Chart 14. The variable *industry* treated as a multi-valued variable

Business Register				Register survey or census of legal units					Estimates		
1. Legal units database table											
LegU idno				LegU idno	ISIC	Industry w	Payroll	P • w	ISIC	Payroll	
1				1	01	0.50	423	212	01	212	
2				1	02	0.30	423	127	02	277	
3				1	10	0.20	423	85	10	430	
4				2	16	0.60	375	225	16	225	
				2	02	0.40	375	150	20	6 520	
2. Industry database table				could be collected							
Combination objects = LegU • Industry											
(Derived KAU database table)											
(KAU idno)	LegU idno	ISIC	Per cent	3	20	0.51	12 784	6 520	24	9 375	
11	1	01	50	3	24	0.49	12 784	6 264	Total	17 038	
12	1	02	30	4	24	0.90	3 456	3 110			
13	1	10	20	4	10	0.10	3 456	346			
21	2	16	60	Sample survey of legal units							
22	2	02	40	LegU idno	ISIC	Payroll	Industry w	Sampling w	P • lw • Sw	ISIC	Payroll
31	3	20	51	2	16	375	0.60	2.5	563	02	375
32	3	24	49	2	02	375	0.40	2.5	375	10	864
41	4	24	90	4	24	3 456	0.90	2.5	7 776	16	563
42	4	10	10	4	10	3 456	0.10	2.5	864	24	7 776
										Total	9 578

¹⁰ Comment by Professor Svein Nordbotten, during the discussion of the Swedish version of our book on register-based statistics in Statistics Sweden’s scientific council, 2 November 2004

¹¹ When one big company suddenly changed 1st industry, we have seen the National Accounts panic – “Don’t change now, we must soon publish!”

The traditional method, illustrated in Chart 13, means that all activities in an enterprise with several industrial classifications are assigned to the largest classification. This gives biased estimates and estimates using weights for combination units, illustrated in Chart 14, will always be better.

Why is the method using 1st ISIC so commonly used? We think that this depends on deeply rooted attitudes. Many do not understand the difference between quality of microdata and quality of estimates: “Microdata in the register survey in Chart 13 is correct, we think we have correct classification of 1st ISIC. Microdata in the register survey in Chart 14 can be wrong, we do not know if the values 212, 127, etc. are correct.” Another problem is that people are not accustomed to use weights for register data. As a rule, only simple summations of values for the records in the register are done to produce the estimates.

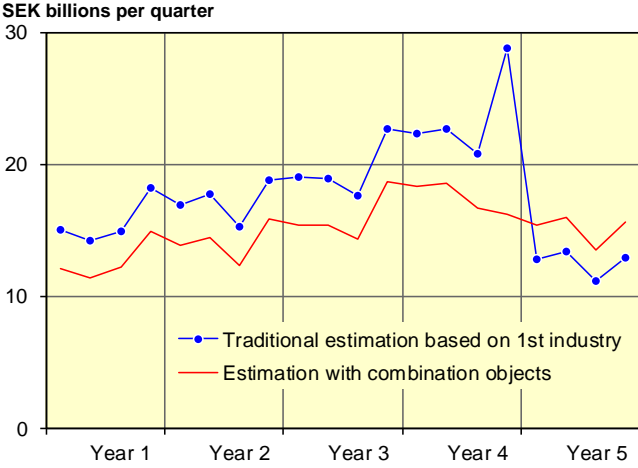
Estimates based on data for legal units will also be inconsistent with estimates based on data for establishments, as 1st industry will often differ between the legal unit and the unit’s establishments. In Chart 15, the first two columns of Chart 3 are displayed. We used the weights in Statistics Sweden’s Business Register and computed estimates for the legal unit version of the register with legal unit weights and then we used data for the establishment version with establishment weights. We arrived at the same results that are shown in column (3) for combination objects. This shows that estimation with combination objects reduces bias and makes estimates based on different kinds of units consistent.

Chart 15. Employees by industry November 2004. Two versions of the Business Register¹²

Industry	Thousands of employees	Legal units 1 st ISIC (1)	Local units 1 st ISIC (2)	Combination objects (3)	Legal-Comb. (1) - (3)	Local-Comb. (2) - (3)
Agriculture and forestry, fishing		35	37	38	2	1
Mining, quarrying, manufacturing		688	636	621	66	15
Electricity, gas and water		21	22	22	1	0
Construction		197	209	207	11	2
Wholesale and retail trade		456	453	468	12	15
Hotels and restaurants		89	93	95	6	1
Transport, communication		240	242	241	1	0
Financial intermediation		83	77	78	5	1
Real estate, business activities		457	524	526	69	2
Government		139	215	223	84	8
Education		382	408	404	22	5
Health and social work		836	684	674	162	10
Other service activities		142	163	166	24	3
Total		3 763	3 763	3 763	466 (12%)	62 (2%)
In all, 466 000 persons were allocated to the wrong industry as measured by 1 st ISIC of the employer’s enterprise (legal unit). The error is 12% of the total number of employees.				2-digit level	14%	2%
				3-digit level	18%	3%
				4-digit level	25%	7%
This kind of error increases if industry is measured at the 3-, 4- or 5-digit level.				5-digit level	38%	11%

Chart 16 illustrates that the estimation method based on 1st ISIC will also give bad time series.

Chart 16. Turnover in industry R, two estimates



We can call the industry in Chart 16 for industry R. During the 4th quarter of year 4, one legal unit with 1st industry = R merged with a legal unit with no production within industry R. During the 4th quarter of year 4 the new merged unit was classified as industry R, but from year 5 the new unit was classified as not belonging to R, as the new part belonged to another industry that from now on became the 1st industry for the whole enterprise. Nothing has happened in the real economy (red curve), Statistics Sweden’s estimates (blue dotted curve) are misleading.

¹² Chart 15 and 16 are from Chapter 14 in Wallgren and Wallgren (2007)

3. Business surveys – how to combine registers and sample surveys?

In this section, we discuss the following three topics:

- How can the calendar year version of the Business Register be used to coordinate and improve economic statistics and the National Accounts?
- How can the calendar year version of the Business Register be used when revised estimates from business sample surveys are produced?
- How can the present method of choosing the best source for National Accounts estimates be improved?

At the end of 2008, the calendar year register in Chart 17 was created. This register gives a rather accurate picture of the target population of the National Accounts for 2007. These accounts are based on several business surveys, most of them done by Statistics Sweden, but some are done by other Swedish statistical institutes. The problem of coordination and consistence is a difficult problem as the Swedish economy is not described by one single survey with consistent estimates but in a piecemeal way with many surveys for different parts of the economy. This is in sharp contrast to social statistics where this piecemeal approach is not used, for instance the Labour Force Survey describes the whole labour market with consistent estimates.

The Structural Business Statistics survey (SBS) gives consistent estimates for the non-financial enterprises and sole traders, but the National Accounts prefer to use better sources for some subpopulations as agriculture, energy and real estate. Also, the supply and use-tables that are used by the National Accounts make that the National Accounts are based on a very large number of sources, which means that the coordination issue is a serious problem.

Assume, that we want to estimate GDP for Sweden 2007. If we compute value-added for all units in Chart 17, then we want that *all* units are included in our GDP-estimate, that no units have been overlooked and that no units have been included more than once. If we compare the frame populations and the register populations for all surveys with the calendar year population, we will find the gaps and overlaps in the system of surveys that are used by the yearly National Accounts. We have done such comparisons and we found many gaps and overlaps.

One way of using the calendar year version of the Business Register is to start to work with comparisons of frame and register populations of surveys used by the National Accounts. This will be a learning process that will take many years. The aim of the work is to coordinate future surveys so that gaps and overlaps gradually are reduced.

A second way of using the calendar register is to revise the estimates of existing business surveys. The coverage errors in all sample surveys describing 2007 can be reduced if the frames that have been created early are compared with the final calendar year register. After calibration of sampling weights, new revised estimates can be used for the work with the yearly National Accounts.

Chart 17. Calendar year register for 2007, legal units by institutional sector and industry

Industry:	Sector:	Non-financial enterprises	Financial enterprises	Central & local government	Sole traders	Non-profit organisations	Total
Agriculture, forestry, fishing		11 354	0	13	236 467	546	248 380
Manufacturing, mining, energy		33 743	1	13	23 717	139	57 613
Construction		44 611	0	0	49 161	62	93 834
Trade and transport		96 626	1	5	61 606	246	158 484
Hotels and restaurants		18 598	2	0	10 966	255	29 821
Information, communication		29 010	1	1	25 807	318	55 137
Financial intermediation		10 852	2 060	10	683	1 116	14 721
Real estate, business activities		157 163	15	49	112 719	10 914	280 860
Government		70	0	298	61	247	676
Education		8 738	0	120	14 277	985	24 120
Health and social work		14 196	0	256	17 847	979	33 278
Personal and cultural services		21 837	1	94	80 281	25 949	128 162
Total		446 798	2 081	859	633 592	41 756	1 125 086

How can the present method of choosing the best source for National Accounts estimates be improved? Today, the Swedish National Accounts uses macrodata from many surveys. When two surveys give estimates for the same parameters, the source that is considered to be the best is used. Agriculture is an example where the National Accounts uses the FSS from the Board of Agriculture instead of the SBS from Statistics Sweden. The SBS uses 1st ISIC = 01 to define agriculture, but the FSS includes only the agricultural part of all establishments above defined thresholds.

These two surveys have populations that are very inconsistent. Selander (2008) describes the undercoverage in the Business Register regarding agricultural enterprises. Statistics Sweden (2007) describes the corresponding overcoverage. In Chart 18 the population in the Farm Register (FR) has been matched against the units in the Business Register (BR).

Chart 18. Coverage errors in the Business Register as compared with the Farm Register¹³

Data from 2004 and 2005	Number of units in		Activity	Coverage	SEK millions	
	FR	BR	In ISIC 01	errors in SBS		
1. ISIC 01 in BR not active		6 787	Not active at all	Overcoverage production	1 110	1.8%
2. ISIC 01 in BR not in FR		64 498	Yes, but not in 01			
3. In FR and BR	67 112	67 112	Yes, and in 01			
4. In FR not in BR	8 696		Yes, and in 01	Undercoverage turnover	4 425	6.8%
Total number of units:	75 808	138 397				

The method used today gives rise to biases and inconsistencies due to gaps and overlaps of population units. Instead, we suggest a new method that is based on *integrated microdata*, in this case microdata from the SBS and the FSS should be combined. The combined dataset should then be used to produce the estimates that are suitable for the National Accounts. In Chart 19 we outline the present estimation method and in Chart 20 the new method is illustrated.

When the National Accounts replace the agricultural part of the SBS, legal units 1 and 5 are excluded. The agricultural part of legal unit 6 is counted twice, it is included in the estimate 1 532 but also in the estimate of value added of forestry, 984. The estimate of total GDP is biased downwards due to excluded units, and biased upwards due to counting some units twice. As the National Accounts use aggregated estimates, it is not possible to see and understand the errors.

Chart 19. Statistics Sweden's present way of using business surveys

Business Register				Statistics Sweden's Structural Business Statistics survey (SBS)				Estimates from the SBS		
1. Legal units database table		3. KAU database table		LegU idno	KAU idno	1 st ISIC	Value added	ISIC	Value added	
LegU idno	1 st ISIC	KAU idno	1 st ISIC	1	11	01	423	01	1 899	
2	16	21	16	2	21	16	375	02	984	
3	20	31	20	3	31	20	6 600	16	375	
4	24	32	24	3	32	24	6 184	20	6 600	
5	01	41	24	4	41	24	3 456	24	9 640	
6	02	51	01	5	51	01	1 476	Total	19 498	
		61	02	6	61	02	984			
				Total				19 498		
2. Industry database table				Excluded by National Accounts				Estimates from the FSS		
Combination objects = LegU • Industry				Board of Agriculture				Farm Structure Survey, FSS		
LegU idno	ISIC	Rank	Per cent	LegU idno	ISIC	Value added	ISIC	Value added		
1	01	1	50			within ISIC 01	01	1 532		
1	02	2	30	1	01	282	Total	1 532		
1	10	3	20	5	01	750				
2	16	1	60	6	01	500				
2	02	2	40	Total				1 532		
3	20	1	51	Counted twice by National Accounts						
3	24	2	49							
4	24	1	90							
4	10	2	10							
5	01	1	60							
5	02	2	40							
6	02	1	55							
6	01	2	45							

¹³ Based on Table 2.2 in Wallgren and Wallgren (2010)

When microdata from the SBS and FSS in Chart 19 are combined, it will be possible to produce unbiased estimates. In a real application, the record linkage will not be as easy as in Chart 20, there will be some many-to-one links, for instance husband and wife can be two sole traders in the Business Register but one holding in the Farm Register. But after careful editing, where we search for unit errors, such problems can be handled.

When information from the SBS and FSS are combined, it is possible to improve the weights for industries that is the last column in the Industry database table. To get the final estimates we used these improved weights. Instead of getting conflicting estimates from two inconsistent surveys, the National Accounts now will get consistent estimates based all available sources.

Chart 20. A new estimation method based on integrated microdata from different surveys

Business Register				Register survey of legal units SBS and FSS integrated microdata					Estimates from the combined SBS and FSS can be used directly by the National Accounts		
1. Legal units database table				LegU idno	ISIC	Industry w	Value added	VA • w	ISIC	Value added	
1				1	01	1.00	282	282	01	1 532	
2				1	02	1.00	85	85	02	1 445	
3				1	10	1.00	56	56	10	402	
4				2	16	0.60	375	225	16	225	
5				2	02	0.40	375	150	20	6 520	
6				3	20	0.51	12 784	6 520	24	9 375	
				3	24	0.49	12 784	6 264	Total	19 498	
				4	24	0.90	3 456	3 110			
				4	10	0.10	3 456	346			
				5	01	1.00	750	750			
				5	02	1.00	726	726			
				6	02	1.00	484	484			
				6	01	1.00	500	500			
								Total	19 498		
								From FSS			
2. Industry database table Combination objects = LegU • Industry (Derived KAU database table)				(KAU idno)	LegU idno	ISIC	Per cent				
11	1	01	67								
12	1	02	20								
13	1	10	13								
21	2	16	60								
22	2	02	40								
31	3	20	51								
32	3	24	49								
41	4	24	90								
42	4	10	10								
51	5	01	51								
52	5	02	49								
61	6	02	49								
62	6	01	51								
				revised weights							
				Board of Agriculture Farm Structure Survey, FSS							
				LegU idno	ISIC	Value added within ISIC 01					
				1	01	282					
				5	01	750					
				6	01	500					
				Total		1 532					

4. Conclusions

Administrative registers are the most important sources to economic statistics in the Nordic countries and will be so in other countries also. In countries like Sweden, the Structural Business Statistics survey is mainly based on tax return data from legal units and the Structural Business Statistics survey and register-based payroll statistics are two very important sources for the National Accounts. To be able to create and maintain a Business Register with good coverage, administrative registration and taxation systems with information on legal units are indispensable.

However, the methods that are used must be improved. In this paper, we give the following proposals:

- The calendar year register should be used as standardised population for the yearly National Accounts. All surveys used for the yearly National Accounts should be consistent towards the calendar register.
- The classification variable *industry* should not be used as it is used today when all activities of a legal unit or an establishment are assigned to 1st ISIC. Instead *industry* should be treated as a multi-valued variable and weights should be used to estimate all activities.
- The National Accounts estimation method should be improved. Instead of using aggregated macrodata, consistent estimates based on integrated microdata from several business surveys should be developed.

References

- Selander, R. (2008) *Comparisons between the Business Register and Farm Register in Sweden*. Paper presented at the First International Workshop on Technology and Policy for Accessing Spectrum, TAPAS 2006.
- Statistics Sweden (2001): *The future development of the Swedish register system – Final report of the Register Project*. R&D Report 2001:1.
- Statistics Sweden (2005): *Why do the National Accounts receive Conflicting Information? Analysis of the coherence between sources used by the National Accounts*. Background Facts, Economic Statistics 2005:6 (in Swedish).
- Statistics Sweden (2007): *Register-based economic statistics based on a standardised register population – A calendar year version of the Business Register with consistent microdata designed for the yearly National Accounts*. Background Facts, Economic Statistics 2007:6 (in Swedish).
- Wallgren, A. and Wallgren, B. (2010) *Using administrative registers for agricultural statistics*. In R. Benedetti, M. Bee, G. Espa and F. Piersimoni (eds), *Agricultural Survey Methods*, pp. 27–44. Chichester: John Wiley & Sons, Ltd.
- Wallgren A., Wallgren B. (2014): *Register-based Statistics – Statistical Methods for Administrative Data*. Second edition, John Wiley & Sons, Ltd, Chichester, UK.
- Wallgren, A., Wallgren, B. (2015) *Towards an integrated statistical system based on registers*. Paper presented at the Inter-American Bank’s conference *Driving the Data Revolution: Statistical Use of Administrative Registers*, Washington 16-18 September 2015.
- Wallgren, A., Wallgren, B. (2016a) *Estimating frame coverage issues in agricultural surveys*. GS Training and Outreach Workshop on Agricultural Surveys, FAO October 2016.
- Wallgren A., Wallgren B. (2016b) *Administrative data and Agricultural Statistics – What Strategy and Methods should we adopt?* Paper presented at ICAS VII, Rome October 2016.