ENBES
Citykonferensen, Stockholm

21 January 2014

Register-based statistics production

Anders Wallgren and Britt Wallgren

Formerly of the Department of Research and Development
Statistics Sweden

ba.statistik@telia.com

Program

...

11:40 1. Response burden

12:05 1. Response burden

13:30 2. Sampling and estimation
in business surveys
Microdata, estimators & quality

14:00 Register-based statistics production

14:30 3. Turnover survey in England

15:20 4. Methodological challenges
in national accounts
Coverage & consistency

...

11:40 1. Response burden

12:05 1. Response burden

13:30 2. Sampling and estimation
in business surveys
Microdata, estimators & quality

14:00 Register-based statistics production:

14:30 3. Turnover survey in England

15:20 4. Methodological challenges
in national accounts
Coverage & consistency

1. Use administrative data!
Reduce respondents' costs
Reduce producer's costs
Improve quality

11:40 1. Response burden

12:05 1. Response burden

13:30 2. Sampling and estimation ⟷ in business (sample) surveys
Microdata, estimators & quality

14:00 Register-based statistics production:

14:30 3. Turnover survey in England

15:20 4. Methodological challenges in national accounts
Coverage & consistency

1. Use administrative data!
   Reduce respondents' costs
   Reduce producer's costs
   Improve quality

2a. Also sample surveys use registers, this saves costs and improves consistency!

11:40 1. Response burden

12:05 1. Response burden

13:30 2. Sampling and estimation in business (sample) surveys
Microdata, estimators & quality

14:00 Register-based statistics production:

14:30 3. Turnover survey in England

15:20 4. Methodological challenges in national accounts
Coverage & consistency

1. Use administrative data!
   Reduce respondents' costs
   Reduce producer's costs
   Improve quality

2b. Estimation in register surveys
Microdata, estimators & quality

11:40 1. Response burden

12:05 1. Response burden

13:30 2. Sampling and estimation
in business (sample) surveys
Microdata, estimators & quality

14:00 Register-based statistics production:

14:30 3. Turnover survey in England ⟷

15:20 4. Methodological challenges
in national accounts
Coverage & consistency

1. Use administrative data!
Reduce respondents' costs
Reduce producer's costs
Improve quality

2b. Estimation in register surveys
Microdata,
estimators & quality

3. Administrative systems,
quality, methodological issues
may differ between countries

11:40 1. Response burden

12:05 1. Response burden

13:30 2. Sampling and estimation in business (sample) surveys
Microdata, estimators & quality

14:00 Register-based statistics production:

14:30 3. Turnover survey in England

15:20 4. Methodological challenges in national accounts ⟷
Coverage & consistency

1. Use administrative data!
Reduce respondents' costs
Reduce producer's costs
Improve quality

2b. Estimation in register surveys
Microdata, estimators & quality

3. Administrative systems, may differ between countries

4. Can National Accounts obtain consistent macrodata?

Register-based statistics production:

1. Use administrative data! **How?**

   Reduce respondents' and producer's costs

   Improve quality

How?	A systems approach, combine data:
Quality?	Better coverage, can find errors

Total turnover
Sample surveys have never better accuracy
Sample surveys may have better timeliness

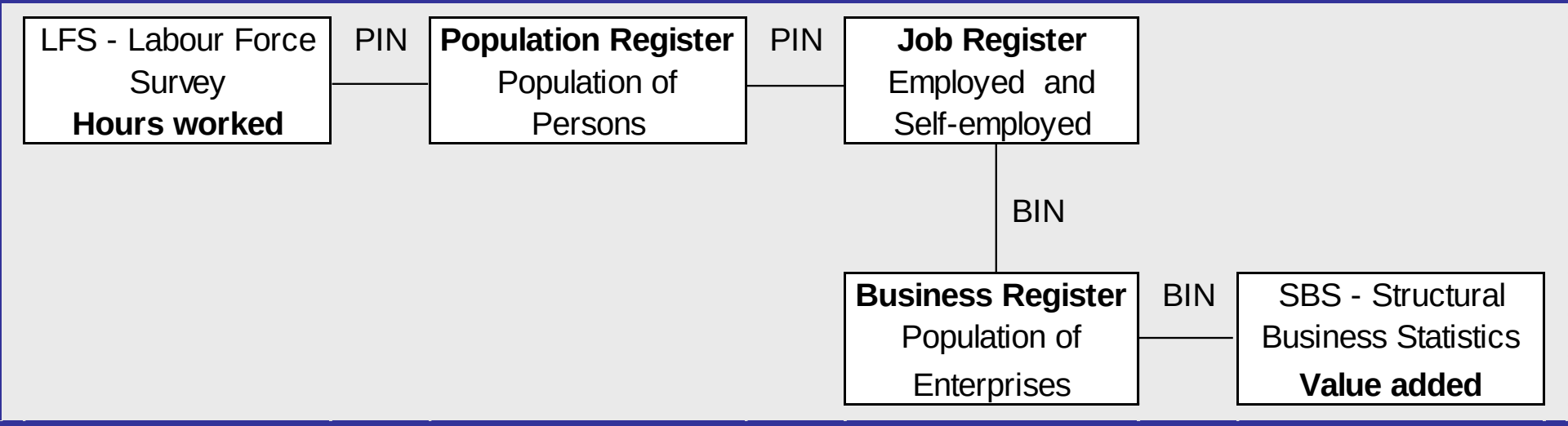
Turnover by product group
Sample survey the only possibility,
but heavy response burden

Register-based statistics production:

2b. Estimation in register surveys
Microdata, estimators & quality

Methodology: A systems approach, combine microdata

Assume that we want to estimate productivity by economic activity. The systems approach means here:



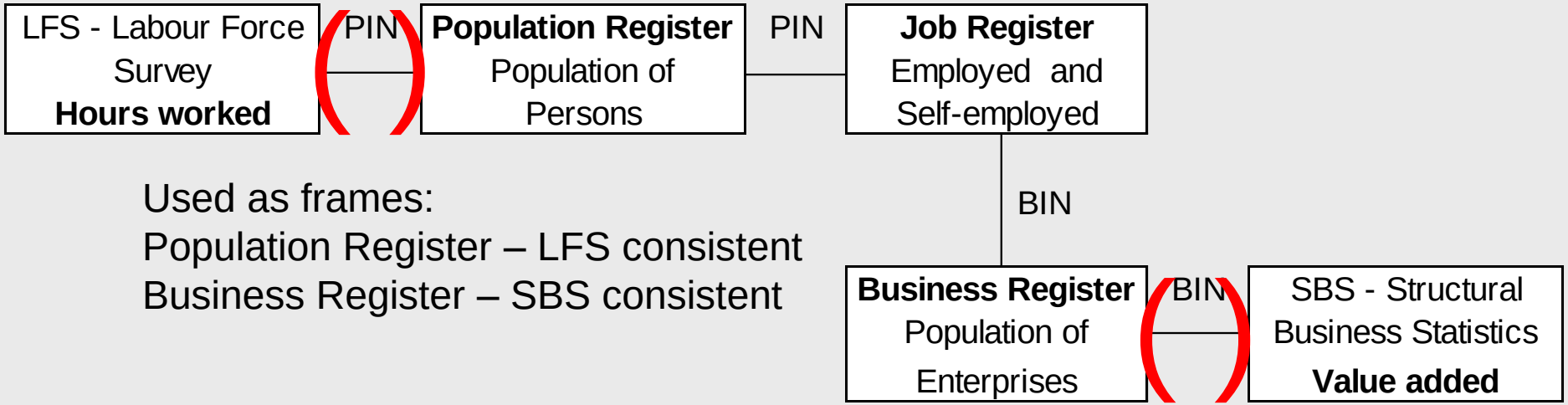
We combine microdata from one sample survey (LFS) and three registers and one combined register and sample survey (SBS)

Register-based statistics production:

2b. Estimation in register surveys
Microdata, estimators & quality

Methodology: A systems approach, combine microdata

Assume that we want to estimate productivity by economic activity. The systems approach means here:



Used as frames:
Population Register – LFS consistent
Business Register – SBS consistent

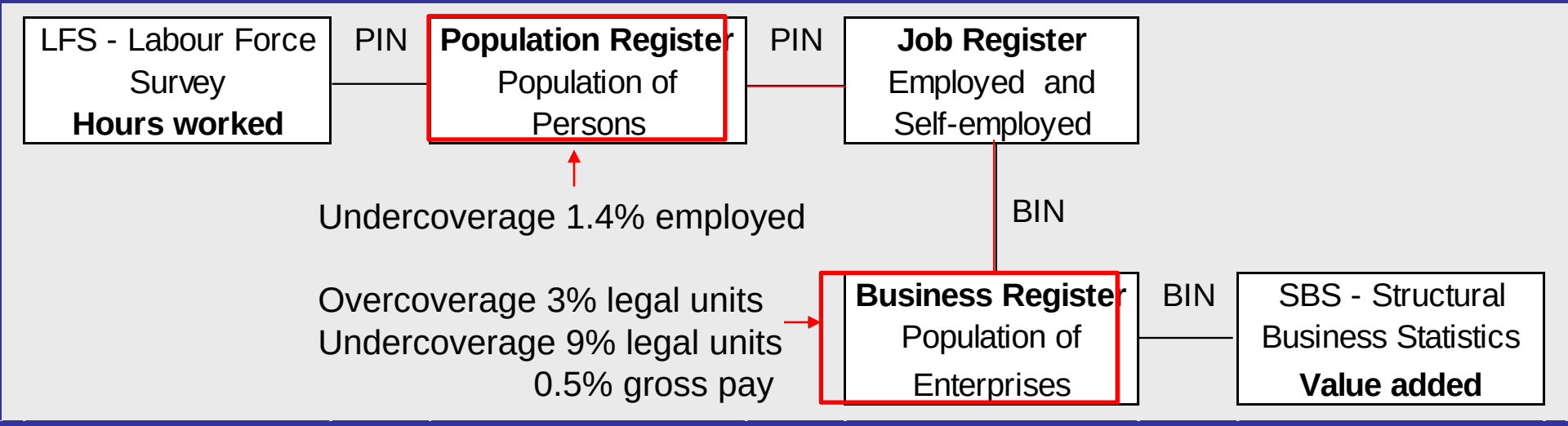
We combine microdata from one sample survey (LFS) and three registers and one combined register and sample survey (SBS)

Register-based statistics production:

2b. Estimation in register surveys
Microdata, estimators & quality

Methodology: A systems approach, combine microdata

Assume that we want to estimate productivity by economic activity. The systems approach means here:



When we correct the Population and Business Register for these coverage errors, ALL surveys using these as frames are improved

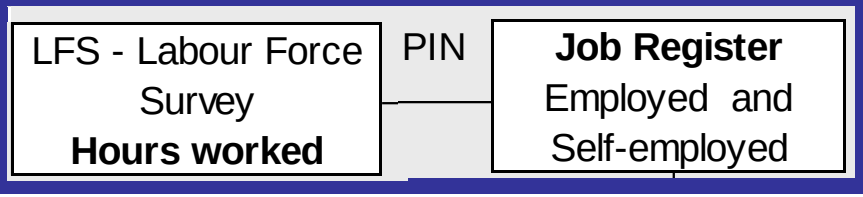


Chart 3. Example of integrated microdata from the LFS and the Job Register

LFS PIN (1)	LFS Hours worked (2)	LFS Hours usually worked (3)	LFS Sector (4)	LFS ISIC (5)	LFS Weight (6)	Job Register PIN (7)	Job Register ISIC (8)	Job Register Sector (9)
PIN1	12	20	6	56100	32.2	PIN1	56100	110
PIN1	16	20	6	56100	28.8	PIN1	56100	110
PIN1	0	20	6	56100	27.9	PIN1	56100	110
PIN1	20	20	6	56100	33.1	PIN1	56100	110
PIN2	40	40	6	56100	32.4	*	*	*
PIN2	40	40	6	56100	31.5	*	*	*
PIN2	40	40	6	56100	33.2	*	*	*
PIN3	40	40	1	01110	32.1	PIN3	81300	320
PIN4	10	10	6	01110	51.5	PIN4	43320	611
PIN5	45	40	6	01131	40.4	PIN5	01500	611
PIN6	30	30	6	01191	43.1	PIN6	*	*
PIN7	5	8	6	01191	45.7	PIN7	01134	110
PIN8	40	40	6	01199	48.1	PIN8	01430	110
PIN9	60	40	6	64190	47.1	PIN9	55102	212
PIN9	60	40	6	64190	44.7	PIN9	55102	212

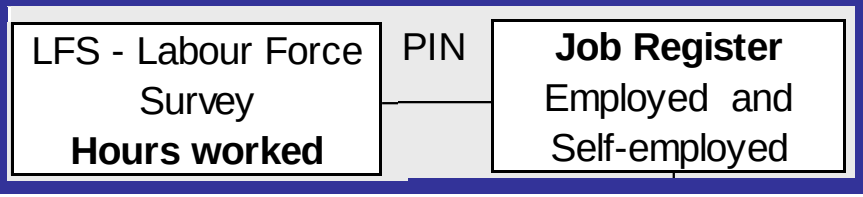


Chart 3. Example of integrated microdata from the LFS and the Job Register

LFS PIN (1)	LFS Hours worked (2)	LFS Hours usually worked (3)	LFS Sector (4)	LFS ISIC (5)	LFS Weight (6)	Job Register PIN (7)	Job Register ISIC (8)	Job Register Sector (9)
PIN1	12	20	6	56100	32.2	PIN1	56100	110
PIN1	16	20	6	56100	28.8	PIN1	56100	110
PIN1	0	20	6	56100	27.9	PIN1	56100	110
PIN1	20	20	6	56100	33.1	PIN1	56100	110
PIN2	40	40	6	56100	32.4	*	*	*
PIN2	40	40	6	56100	31.5	*	*	*
PIN2	40	40	6	56100	33.2	*	*	*
PIN3	40	40	1	01110	32.1	PIN3	81300	320
PIN4	10	10	6	01110	51.5	PIN4	43320	611
PIN5	45	40	6	01131	40.4	PIN5	01500	611
PIN6	30	30	6	01191	43.1	PIN6	*	*
PIN7	5	8	6	01191	45.7	PIN7	01134	110
PIN8	40	40	6	01199	48.1	PIN8	01430	110
PIN9	60	40	6	64190	47.1	PIN9	55102	212
PIN9	60	40	6	64190	44.7	PIN9	55102	212

Perfect consistency

LFS - Labour Force Survey
Hours worked

PIN

Job Register
Employed and Self-employed

Chart 3. Example of integrated microdata from the LFS and the Job Register

LFS PIN (1)	LFS Hours worked (2)	LFS Hours usually worked (3)	LFS Sector (4)	LFS ISIC (5)	LFS Weight (6)	Job Register PIN (7)	Job Register ISIC (8)	Job Register Sector (9)
PIN1	12	20	6	56100	32.2	PIN1	56100	110
PIN1	16	20	6	56100	28.8	PIN1	56100	110
PIN1	0	20	6	56100	27.9	PIN1	56100	110
PIN1	20	20	6	56100	33.1	PIN1	56100	110
PIN2	40	40	6	56100	32.4	*	*	*
PIN2	40	40	6	56100	31.5	*	*	*
PIN2	40	40	6	56100	33.2	*	*	*
PIN3	40	40	1	01110	32.1	PIN3	81300	320
PIN4	10	10	6	01110	51.5	PIN4	43320	611
PIN5	45	40	6	01131	40.4	PIN5	01500	611
PIN6	30	30	6	01191	43.1	PIN6	*	*
PIN7	5	8	6	01191	45.7	PIN7	01134	110
PIN8	40	40	6	01199	48.1	PIN8	01430	110
PIN9	60	40	6	64190	47.1	PIN9	55102	212
PIN9	60	40	6	64190	44.7	PIN9	55102	212

Black sector

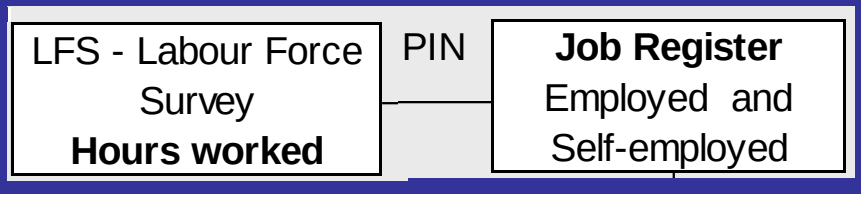


Chart 3. Example of integrated microdata from the LFS and the Job Register

LFS PIN (1)	LFS Hours worked (2)	LFS Hours usually worked (3)	LFS Sector (4)	LFS ISIC (5)	LFS Weight (6)	Job Register PIN (7)	Job Register ISIC (8)	Job Register Sector (9)
PIN1	12	20	6	56100	32.2	PIN1	56100	110
PIN1	16	20	6	56100	28.8	PIN1	56100	110
PIN1	0	20	6	56100	27.9	PIN1	56100	110
PIN1	20	20	6	56100	33.1	PIN1	56100	110
PIN2	40	40	6	56100	32.4	*	*	*
PIN2	40	40	6	56100	31.5	*	*	*
PIN2	40	40	6	56100	33.2	*	*	*
PIN3	40	40	1	01110	32.1	PIN3	81300	320
PIN4	10	10	6	01110	51.5	PIN4	43320	611
PIN5	45	40	6	01131	40.4	PIN5	01500	611
PIN6	30	30	6	01191	43.1	PIN6	*	*
PIN7	5	8	6	01191	45.7	PIN7	01134	110
PIN8	40	40	6	01199	48.1	PIN8	01430	110
PIN9	60	40	6	64190	47.1	PIN9	55102	212
PIN9	60	40	6	64190	44.7	PIN9	55102	212

Inconsistencies! **Different Sector** **Different ISIC**

Register-based statistics production:

4. Can (yearly) National Accounts get consistent macrodata?

The previous examples show that by combining microdata from different sources we can find and correct many errors in economic statistics.

There is another way of improving consistence:

By combining many sources, coverage can be improved and the populations used in different surveys can be harmonised:

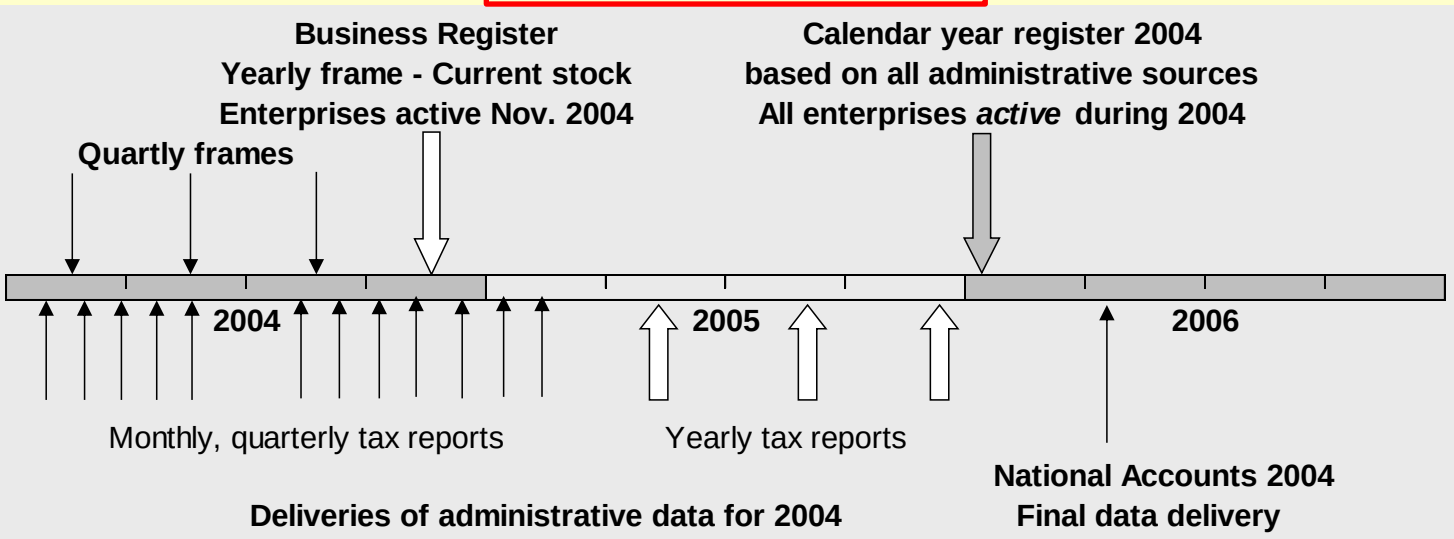
The production of ALL enterprises should be measured and there should be no double counting!

All enterprises active in at least on administrative system regarding 2004 = The calendar year population 2004

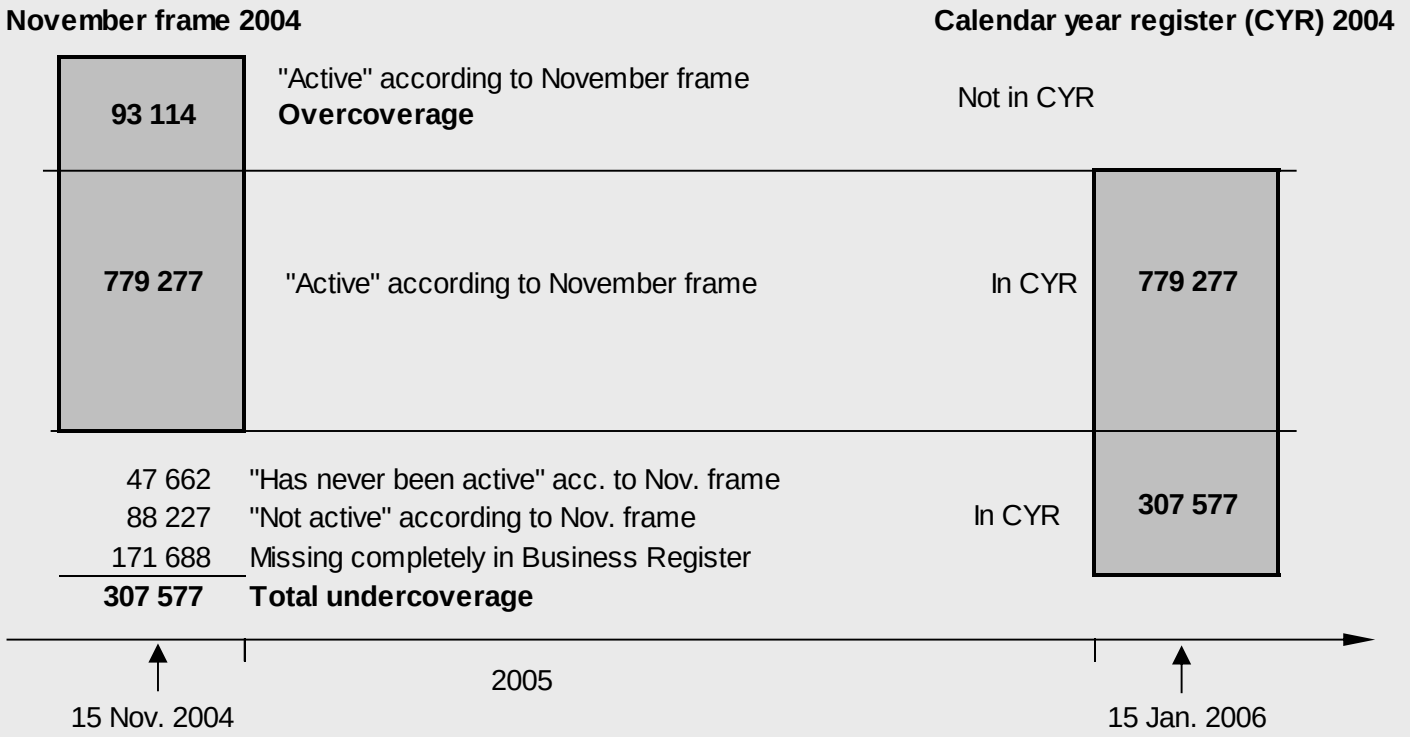
Wait until **all** administrative sources regarding 2004 are available

Coverage much better!

The November frame and the Calendar Year Register 2004



Over and undercoverage as number of legal units in the November frame



Yearly GDP is not measured in one survey, it is measured in a large number of surveys done by different statistical offices in Sweden, monthly, quarterly and yearly surveys, sample surveys and register surveys:

Serious inconsistencies regarding populations!

Legal units by institutional sector and economic activity Calendar year population

Institutional sector: / Economic activity:	Non-financial enterprises	Financial enterprises	Govern-ment	Sole traders	Non-profit organisations
Agriculture, forestry, fishing	11 354	0	13	236 467	546
Manufacturing, mining, energy	33 743	1	13	23 717	139
Construction	44 611	0	0	49 161	62
Trade and transport	96 626	1	5	61 606	246
Hotels and restaurants	18 598	2	0	10 966	255
Information, communication	29 010	1	1	25 807	318
Financial intermediation	*10 852*	2 060	10	683	1116
Real estate, business activities	157 163	15	49	112 719	10 914
Government	*70*	0	298	*61*	247
Education	8 738	0	120	14 277	985
Health and social work	14 196	0	256	17 847	979
Personal and cultural services	21 837	1	94	80 281	25 949

Questionnaire data and administrative data on enterprises

BIN = Business identity number of each legal unit/entity
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
Distance = |SBS-YIT| + |SBS-VAT| + |YIT-VAT|

Chart 2.3 Yearly turnover for the same enterprises in three sources, USD million (transformed microdata)

BIN	SBS	YIT	VAT	Distance
160001	7 179	11 941	8 089	3 175
160002	2 954	0	0	1 969
160003	843	3 561	918	1 812
160004	5 514	2 888	2 895	1 751
160005	26	538	2 536	1 673
160006	2 301	0	0	1 534
160007	2 211	0	2 239	1 493
160008	1 316	1 316	0	877
160009	638	638	0	425
160010	456	0	435	304
160011	141	141	0	94

Questionnaire data and administrative data on enterprises

BIN = Busine… …tity
SBS = Turnov… …stionnair
YIT = Turnov… …eturns
VAT = Turnover according to 12 monthly VAT returns
Distance = |SBS-YIT| + |SBS-VAT| + |YIT-VAT|

2 strata in SBS:
- Big ones get questionnaires
- The rest: Use administrative data

Chart 2.3 Yearly turnover for the same enterprises in three sources, USD million (transformed microdata)

Complete groups of enterprises

BIN	SBS	YIT	VAT
160006	2 301	0	0
170006		2 301	2 301
160007	2 211	0	2 239
170007		2 211	0

160006	2 301	0	0	1 534
160007	2 211	0	2 239	1 493
160008	1 316	1 316	0	877
160009	638	638	0	425
160010	456	0	435	304
160011	141	141	0	94

Administrative and statistical units give rise to difficult problems in economic statistics

One enterprise may consist of many legal units

Tax reporting may be orgainsed in different ways:

VAT in one way
YIT in another way

Here, SBS may get measurement errors

Employees by economic activity November 2004 thousands

Economic activity	Legal units 1st ISIC (1)	Local units 1st ISIC (2)
Agriculture and forestry, fishing	35	37
Mining, quarrying, manufacturing	688	636
Electricity, gas and water	21	22
Construction	197	209
Wholesale and retail trade	456	453
Hotels and restaurants	89	93
Transport, communication	240	242
Financial intermediation	83	77
Real estate, business activities	457	524
Government	139	215
Education	382	408
Health and social work	836	684
Other service activities	142	163
Total	3 763	3 763

Different units have consequences for economic statistics:

Inconsistencies!

We need better estimation methods!

You have seen substantial errors and inconsistencies in Statistics Sweden' economic data

Other countries have the same errors but it is not so easy to find the errors if you do not have a system of statistical registers

Errors and inconsistencies in economic data is a general problem

By using administrative data in a systematic way many sources can be combined, inconsistencies and errors will be found and can be reduced