

SOME REMARKS ON THE USE OF AUXILIARY INFORMATION IN THE FINNISH LFS

Riku Salonen, Statistics Finland

e-mail: riku.salonen@stat.fi

INTRODUCTION

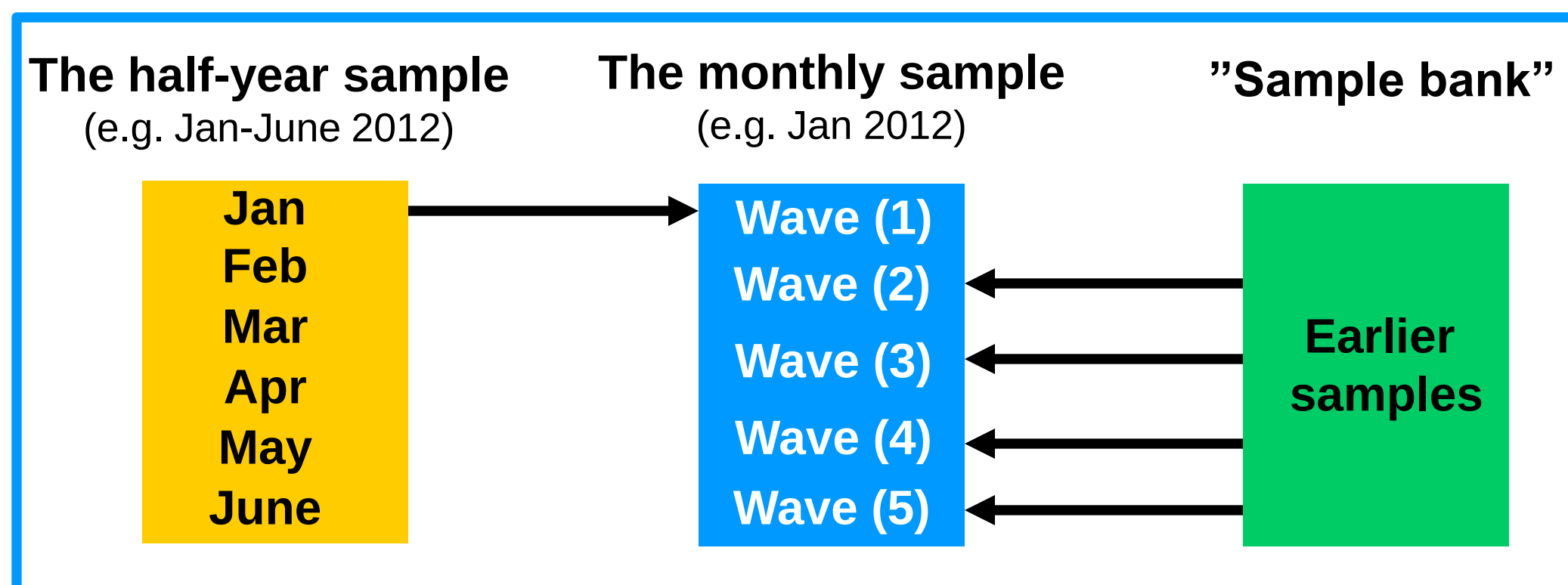
This poster examines the use of auxiliary information both at the sampling stage and at the estimation stage in a complex rotating panel design. Empirical results are based on real data from the Finnish Labour Force Survey (LFS). The motivation for this study comes from the auditing process conducted in 2011 (auditing of the Finnish LFS production system) and recent articles. Many articles evaluate the use of registers as auxiliary information in survey-based statistics and the selection of "best auxiliary variables" (see e.g. Särndal & Lundström, 2010). The use of several auxiliary information (e.g. administrative registers and published official statistics) both at the sampling stage and at the estimation stage in official statistics production are also discussed by Lehtonen & Djerf (2008) and Fuller (2009). Furthermore Särndal (2007) considers successful application of the use of auxiliary information at the estimation stage in a complex rotating panel design, which part of the information coming from the survey results in previous wave.

THE FINNISH LFS

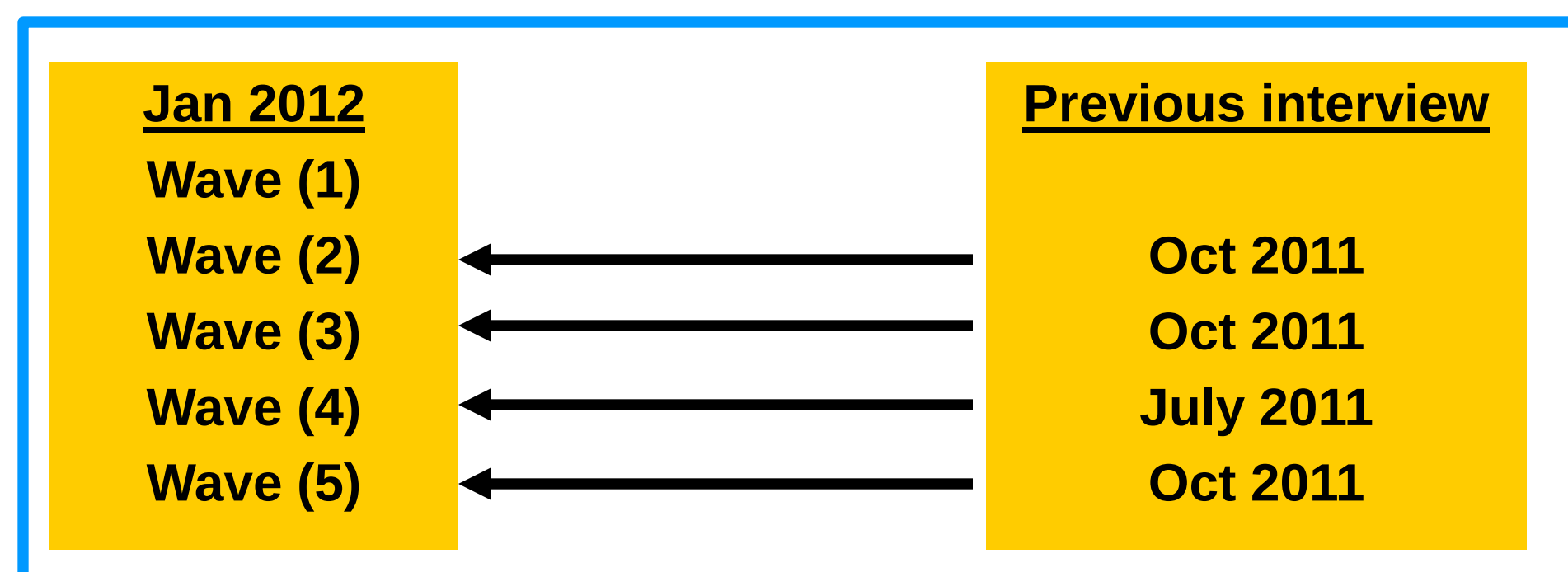
- Continuous monthly survey on individuals
- It provides monthly, quarterly and annual results
- The target population are persons aged 15 to 74

The half-year and monthly samples

- The half-year sample is drawn two times a year
- The sample size (14,400) is sufficient to meet the need for individuals in the monthly samples for the next six months
- It is allocated into six equal part - one for each month



- The monthly sample is divided into five waves
 - Wave (1) come from the half-year sample
 - Waves (2) to (5) come from "sample bank"
- The sample size is approximately 12,000
- The survey is repeated over time with partially overlapping samples (e.g. January 2012)



- Each sample person is in sample five times during 15 months
- The monthly rotation pattern: 1-(2)-1-(2)-1-(5)-1-(2)-1
 - No month to month overlap
 - 60% quarter to quarter theoretical overlap
 - 40% year to year theoretical overlap

SAMPLING STAGE

The sampling design used in LFS is stratified systematic sampling of elements. The population of individuals is divided into strata. The strata are formed according to NUTS-1 regions (Mainland Finland and The Autonomous Territory of the Åland Islands). In each stratum systematic random selection is applied to the frame sorted according to the domicile code which yields implicit geographic stratification.

ESTIMATION STAGE

The Finnish LFS uses generalized regression (GREG) estimation and it utilises auxiliary information in the estimation stage (e.g. control totals for GREG used in CLAN program).

var	N	mar1	mar2	...	mar21
region	21	282 759	345 671	...	786 285
sex	2	1 995 190	1 994 188		
age group	12	330 875	328 105	...	
job-seeker status	8	68 429	115 171	...	

The use of registers as auxiliary information

Finland belongs to the so-called register countries and the Finnish system of registers is quite up-to date, especially in register data on individual persons.

The database of the total population

- Maintained by Statistics Finland
- The primary source for information on the population of Finland
- It provides the basic information for LFS in both the sampling and estimation stage (**sex, age and region**)
- It is based on the Population Information System of The Population Register Centre
- Matching key: PIN

Job seeker register

- Maintained by The Ministry of Employment and the Economy's
- The "job-seeker status" are classified in different categories (e.g. the duration of unemployment)
- Matching key: PIN

Some audit findings and conclusions

- In the Finnish LFS it is not possible to obtain unbiased estimates on unemployment without using proper auxiliary information
- The use of job seeker register data as auxiliary information reduced the gap between the two official unemployment figures and improved the precision of unemployment estimates

FUTURE WORK

Is there more potential register-based auxiliary information available both at the sampling stage (e.g. other stratification variables) and at the estimation stage (e.g. Register-Based Employment Statistics)?

How to select the best and powerful auxiliary information?

Utilisation of information collected in previous waves (e.g. the regression composite (RC) estimator in the Canadian LFS)

REFERENCES

Fuller, W.A. (2009). Sampling Statistics, Wiley.

Lehtonen, R. & Djerf, K. (2008). Survey sampling reference guidelines: introduction to sample design and estimation techniques, Eurostat methodologies and working papers, Office for Official Publications of the European Communities, Luxembourg

Särndal, C-E. (2007). The calibration approach in survey theory and practice. Survey Methodology, 33, 99-119.

Särndal, C-E. & Lundström, S. (2010). Design for estimation: Identifying auxiliary vectors to reduce nonresponse bias. Survey Methodology, 36, 131-144.