



Consistent weighting of the LFS - monthly, quarterly, annual and longitudinal data

Martijn Souren

summary *This paper describes the challenges that come with pursuing internal consistency for (un)employment statistics on different periods. Solutions for these challenges are presented, which enable Statistics Netherlands to publish consistent statistics on monthly, quarterly, annual and longitudinal data for the most prominent (un)employment estimates.*

keywords *LFS weighting, Rotation Group Bias, consistency*

1. Introduction

The Dutch Labour Force Survey (LFS) is based on a rotating panel design. Each month a sample of addresses is drawn and data are collected by means of computer assisted interviewing of the residing households. The sampled households are re-interviewed four times with quarterly intervals. The estimation procedure of this survey for quarterly and yearly estimates is based on the generalized regression (GREG) estimator, developed by Särndal e.a. (1992).

The rotating panel design of the Dutch LFS makes it possible to compose monthly, quarterly, annual and longitudinal statistics. However, there are also consistency problems with the use of a rotating panel design and subsampling for monthly, quarterly, annual and longitudinal LFS data. The problem addressed in this paper is the way to achieve consistency between different publications. The way Statistics Netherlands deals with these problems, using the GREG estimator, is presented in this paper. Also the implications of using monthly estimators via a structural time series model (STM) are described.

2. Weighting and internal consistency

The weighting scheme of the Dutch LFS is based on a combination of different socio-demographic categorical variables. The current rotation scheme implies that the sample data obtained with the five waves in three successive months all represent unique households from independent samples. The advantage of a GREG estimator is that the whole dataset can be analysed with one set of weights. This leads to consistent statistics on all variables on a quarterly basis. There is however a problem with substantial systematic differences between the subsequent waves of the panel due to mode- and panel effects. This is a well-known problem for rotating panel designs, and in the literature is referred to as rotation group bias (RGB), see Bailer (1975). In the Dutch LFS, the level of the unemployment in the subsequent waves is substantially smaller compared to the first wave.

2.1 RGB and model based monthly estimates

A big disadvantage for Statistics Netherlands is that the weighting procedure can hardly correct for this RGB. Statistics on (un)employment have been based on computer assisted personal interviewing (CAPI) for many years. The introduction of the mixed mode rotation scheme, and a redesign of it thereafter, would therefore have introduced breaks in the time series which are not easily corrected for.

Another problem with the Dutch LFS is that the monthly sample size is too small to rely on the GREG estimator to produce official statistics about the monthly employment and unemployment. GREG estimators have a relatively large design variance in the case of small sample sizes and with the alternative of a moving average structural changes in the (un)employment appear delayed in the series of the published figures.

To deal with both problems, a structural time series model (STM) is applied as an alternative. This approach was presented in the LFS workshop of 2009 and the 2013 workshop on monthly estimates (see van den Brakel, 2009 and 2013). The STM takes advantages from sample information observed in the past to improve the precision of the estimated monthly (un)employment level. It also accounts for the systematic differences between the five monthly GREG estimates to avoid that the estimated monthly (un)employment level is biased due the RGB. Finally, the STM accounts for discontinuities in the time series due to redesigns and it can be further improved with the use of auxiliary series.

2.2 From monthly to quarterly estimates

The monthly estimates in the Netherlands fulfil a more prominent role in the public discussion about labour market trends than the quarterly estimates. In our philosophy it thus important to publish the best monthly estimates as possible and, under the condition that an average of these estimates is at least as good as the quarterly estimates, from thereon build quarterly

(and yearly and longitudinal) LFS-statistics. In fact, quarterly estimates can also benefit from the improvements of explicitly modelling RGB, auxiliary information, time series information and discontinuities. This is especially important in relation to the rotating panel design where it is not straightforward to have GREG estimators deal with for example RGB and discontinuities that will gradually grow into the figures due to the panel structure.

For these reasons, and because Statistics Netherlands does not prefer to revise already published monthly estimates, it was decided to incorporate the monthly estimates into the quarterly weighting scheme. A table is added with an average of the three model-based monthly (un)employment estimates crossed by age and sex. Another advantage of incorporating monthly estimates into the quarterly weighting scheme is that it can be made possible to also analyse on monthly weighted data for a better interpretation of the monthly (un)employment estimates. This can be done by adding a table with each monthly total (un)employment estimate, divided by three (see figure 1). It is then possible to retrieve monthly estimates from the quarterly weighted data by multiplying the quarterly weights by three in order to create monthly weights.

2.3 Consistent quarterly and annual statistics

The annual estimates can be calculated by adding all of the quarterly data and dividing the quarterly weights by four. This automatically means that monthly, quarterly and annual estimates on (un)employment (crossed by sex and age) are consistent by performing just one weighting procedure each quarter.

However, for variables that are not surveyed every wave, a different set of weights has to be calculated. The same tables with monthly (un)employment estimates (crossed by age and sex) can be added to this weighting scheme. However, all other estimates that are not part of the weighting scheme, will not be consistent for these data subsets. For example, the number of higher educated employees from the regular yearly estimates is different (1% higher) compared to the estimates based on just one wave. The differences are generally small and do not lead to many questions from users.

Finally, although the sample size is mostly too small to publish on a quarterly basis if not all waves can be used. The weighting is done quarterly as well (see figure 1), which has again important advantages for analytical purposes.

Figure 1. Schematic overview of the consistent weighting framework for cross-sectional estimates.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Monthly estimates												
W ₁												
W _{2..}												
W ₅												
Quarterly estimates												
W ₁	Jan/3 + Feb/3 + Mar/3											
W _{2..}	and/or (Jan + Feb + Mar)/3											
W ₅												
Yearly estimates = (Q ₁ + Q ₂ + Q ₃ + Q ₄)/4												
W ₁	Jan/3 + Feb/3 + Mar/3			Apr/3 + May/3 + Jun/3				
W _{2..}	and/or (Jan + Feb + Mar)/3			and/or (Jan + Feb + Mar)/3				
W ₅												
Yearly estimates (one wave) = (Q ₁ + Q ₂ + Q ₃ + Q ₄)/4												
W ₁												
W _{2..}	(Jan + Feb + Mar)/3			(Jan + Feb + Mar)/3				
W ₅												

2.4 Consistent quarterly, annual and longitudinal statistics

Next to the monthly, quarterly and annual estimates, the panel structure makes it possible to produce flow statistics. This approach was presented in the LFS workshop of 2009, see Bierings e.a.(2009). Examples are quarterly flows in and out of (un)employment. This is done as follows: for each quarter all respondents that also responded in the subsequent quarter (or the same quarter in the year after) are selected. For this subsection of respondents a similar weighting scheme is used as for the “regular” quarterly estimates. In addition there are some specific weighting terms that are related to labour market mobility, like movings and changes in unemployment registration. Also again, a table is added to the weighting scheme in order to achieve consistency with the quarterly estimates: for each period (beginning and ending quarter) the labour market status based on the monthly estimates is added to the weighting scheme. For the beginning period these are exactly the already published quarterly estimates. However, the total population is of course different for the ending period. Therefore the estimates for the (un)employed can be made consistent, but the difference in population is discounted in the non-active population. This way for the most important variables consistency is achieved (see figure 2).

Figure 2. Schematic overview of the consistent weighting framework for flow estimates

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
W ₁												
W _{2..}												
W ₅												
Quarterly flow estimates												
W ₁												
W _{2..}	S: (Oct + Nov + Dec)/3			S: (Jan + Feb + Mar)/3				
W ₅	E: (Jan + Feb + Mar)/3			E: (Apr + May + Jun)/3				
Yearly flow estimates												
W ₁												
W _{2..}												
W ₅	S: (J+F+M+A+M+J+J+A+S+O+N+D: t-1)/12 E: (J+F+M+A+M+J+J+A+S+O+N+D: t)/12											

3. Discussion

In this paper a system of weighting for the LFS is described that aims at achieving a maximum amount of consistency in estimates. The most prominent estimates for (un)employment are made consistent for monthly, quarterly, annual and longitudinal statistics. This is necessary because in all estimates of the LFS the RGB has to be taken into account. It is still an issue how far one wants to go to with this consistency procedure. Should everything that you calculate monthly also be retrievable from the quarterly files? What if we can calculate flow statistics on a monthly basis?

References

- Bailar, B.A. (1975). The Effects of Rotation Group Bias on Estimates from Panel Surveys. *Journal of the American Statistical Association*, 70, pp. 23-30.
- Brakel, J.A. van den, B. Janssen, S. Krieg and M. Souren, (2009). Estimation of the Monthly Unemployment Level through Structural Time Series Modelling in a Rotating Panel Design. Discussion paper for LFS workshop 2009. Statistics Netherlands, Heerlen.
- Brakel, J.A. van den and S. Krieg (2013) Dealing with small sample sizes, rotation group bias and discontinuities in rotating panel designs. Discussion paper for LFS workshop on monthly unemployment 2013. Statistics Netherlands, Heerlen.
- Bierings, H. H-J. Dirven, J. van der Valk and R. de Vries (2009). Labour market flows in the Netherlands. Discussion paper for LFS workshop 2009, Statistics Netherlands, Heerlen.
- Särndal, C-E., B. Swensson and J. Wretman (1992). *Model Assisted Survey Sampling*. New York: Springer Verlag.