Discussion Paper

Estimating a time series of temporary employment using a combination of survey and register data

Nino Mushkudiani
Jeroen Pannekoek

November 2019

Contents

1 Introduction 4

2 Data sources and targets of inference 5
- 2.1 The used data sets 6
- 2.2 Parameters of interest: distributions per time-point and transitions 8

3 Reconciliation methods 9
- 3.1 Objective function 10
- 3.2 Specific models 11

4 Comparison with results from the measurement error model of Pankowska et al. 16
- 4.1 Differences in combining the groups 17
- 4.2 The latent Markov model approach versus macro-integration approach 17
- 4.3 Numerical results 19

5 Discussion 20

A Solutions to the optimisation problems 24
- A.1 General solution 24
- A.2 Quadratic-loss, additive adjustments 25
- A.3 KL-loss, multiplicative adjustments 26
- A.4 Algorithm 27

B Model outcomes for pooled data 29

The goal of this paper was to investigate how macro-integration methods can be applied to the reconciliation process of labor force statistics from two sources: a survey and an administrative source. In particular, the aim was to arrive at a single estimate of the time series of temporary employment that efficiently combines the information from both sources. By varying the specifications of the objective function and constraints, four different macro-integration models were defined for this purpose. The most plausible results were obtained for a model that treats none of te sources as fixed and uses multiplicative adjustments. The results were compared with those of Pankowska et. al. 2018, who estimated the same time series on the same data with a latent Markov model. This individual level model-based approach does not lead to very different estimates of the time-series of temporary (or permanent) employment contracts but does result in smaller estimates of the proportion of ”movers”, those that change contract status from temporary to permanent or the other way around. The model-based approach also provides estimates of the measurement errors in each of the sources. On the other hand, the macro-integration approach is less restrictive in the sense that it does not impose a Markov property of the integrated times series of proportions and it is more easy to implement.

Keywords: **Macro-integration, Labor force statistics, time series, multiple source statistics.**

1 Introduction

As the information era is booming, statistical agencies have to deal with an increasing number of all kinds of data sources, including administrative registers, census and big data or sample surveys. Combining all these different sources to make official statistics figures is essential in view of the necessity to improve the quality, deal with the decreasing survey response burden and the rising costs Bakker (2011), Zhang (2012), Fosen and Zhang (2011), Guarnera and Varriale (2015). “The 20th century witnessed the birth and maturing of sample surveys; the 21st century will be the age of data integration” Zhang (2012).

This naturally involves research on methods for multi-source statistics. The major issue is to combine data from different sources that are typically collected by different institutions for their own specific purposes and are not readily together usable in their original form for statistical purposes. Furthermore, all these sources describe different dimensions of the same phenomenon. de Waal et al. (2019) give an overview of important problems that can occur when working with multi-source statistics and methods that can be used to overcome these problems are described. Many of these methods are developed quite recently and have some drawbacks, and therefore the authors suggest that more research should be carried out in this direction.

In this paper we present an experiment on integration of data from a sample survey (the Labour Force Survey, LFS) and a register (the Employment Register, ER). Both data sources contain information on labour market variables with different definitions, different population coverage and different frequency. Currently, Statistics Netherlands publishes statistics based on these sources separately. For social and economic studies of the job market, where the number of temporary employed persons is an important figure, there is a need for a reliable and undisputed estimate of this figure and especially its development in time. It is therefore desirable to have a single estimate of the time series of temporary employment that efficiently combines the available information in both sources. The goal of our paper is to investigate the possibility of obtaining reconciled figures of labour statistics using aggregated figures from both sources.

The variable of interest is the employment contract type, consisting of three categories: Permanent, Temporary and Other. In particular, we are interested in the development in time of the proportion of temporary employed persons. The proportions obtained from the two different data sources are the aggregated figures of interest. They form time series with different observation frequencies; from ER we have monthly values and from LFS quarterly. We see discrepancies between the two sources in observed proportions at the same time points and we will consider several reconciliation models to deal with these discrepancies. These models are based on macro-integration techniques. Macro-integration methods are applied at different production processes at Statistics Netherlands (see e.g. Mushkudiani et al. (2018)). In this paper we investigate a slightly different model, but we do follow the general principle of macro-integration.

Using the same data set, Pavlopoulos and Vermunt (2015) and Pankowska et al. (2018) proposed another approach based on a latent Markov model for obtaining integrated results. These models view the values obtained for the same variables in the different sources as fallible measurements of the same underlying true (or latent) variables that differ from the true values (and each other) due to (differences in) measurement errors. Estimates of proportions based on the true values will provide single estimates of the corresponding population values that are corrected for measurement errors.

Other National Statistics Institutes (NSIs) also face similar problems and investigate integration methods for labour force statistics. Goni et al. (2019) describe how the Basque Statistics Office integrated administrative information in the Labour Force Survey using a micro calibration method. This was done in order to improve the quality of survey estimations and gain coherency between internal and external results (registered employed and unemployed populations). One of the many methodological changes is that survey respondents are confronted with the administrative information that is linked beforehand. The administrative data is used as auxiliary information also during the weighting and calibration procedure. Statistics Norway also investigated micro-integration techniques for combining employment data, see Fosen (2011), Fosen and Zhang (2011). When differences are observed in the variable of interest between two data sources, additional information, for example on wages, is used to define the most probable value for the variable for each unit.

This report is organized as follows: In Section 2 we describe our data set. We also derive tables of proportions and transition proportions for the variable 'Contract type'; In Section 3 we define objective functions for macro-integration models. Next we introduce four different models, starting with the simplest additive model for fixed survey figures adjusting only register variables and making it more complex using multiplicative adjustments and adjusting both survey and register figures; In Section 4 we compare our results with the latent class analysis approach given in Pankowska et al. (2018); Finally in the discussion we share our thoughts on the way the proposed macro-integration method could be implemented in the production process. More effort should be made to implement the method we propose here. We also discuss advantages and disadvantages of the macro-integration methods versus the latent class micro-integration method.

2 Data sources and targets of inference

The data sources are a register and a survey. Information on individuals that were in both data sources is linked. Both sources contain information on the employment status of each person, which is measured by the variable 'Contract type'. The targets of inference are the population proportions in each of the categories of 'Contract type' and, in particular, the development in time of these proportions. Apart from the univariate distributions of 'Contract type' at each time point, this development can also, and more detailed, be described by the transition proportions between adjacent time points. That is, the proportions of individuals moving from one category at time-point t to another category at time point $t + 1$, for each of the 3×3 category combinations. Below we describe the data sources and target parameters in more detail and introduce some notation.

2.1 The used data sets

The data set that is used in this paper is the linked data set obtained from the two data sources; the Labor Force Survey (LFS) [1] and the Employment Register data (ER).

The LFS is a rotating panel survey consisting of five waves, collecting information about labour of households and individuals. For our study we consider only individuals and not households. The rotating panel design entails that after a respondent joins the survey, he/she will receive follow-up questionnaires for four more times with three months intervals. Each month the sample is supplemented with new respondents to compensate for those that have finished five waves. As a result of this setup there are five waves in each month.

The ER is an administrative data set that combines information from different administrative sources, mainly from Tax authorities but also from the Centre for Work and Income (CWI) and the institute for employees insurances (in Dutch Uitkeringsinstituut Werknemers Verzekeringen (UWV)). The ER consists of administrative information on persons, households, jobs, benefits and pensions. It covers the entire Dutch population, including persons living abroad but working in the Netherlands or receiving a benefit or pension from a Dutch institution, see Cobben (2009).

An important difference between the sources is the population coverage. For example, the LFS includes self-employed persons and excludes institutional residents and the ER includes all persons living abroad but working in the Netherlands and excludes self-employed persons.

The micro-data from the two sources are linked using the personal identity number and some additional background information is obtained from other linked registers. Differences on the micro-level between the two sources in the measures of contract type in the corresponding months can be expected due to differences in definitions,[2] differences in time of measurement and measurement errors.

From the linked data set, we used the data of 8886 individuals of age 25-55, observed during 15 months, from Januari 2007 until March 2008. During this period we have 5 quarterly observations of the employment status from the LFS (Jan-07, Apr-07, Jul-07, Oct-07, Jan-08) and 13 monthly measurements of employment status are available from the ER. Our data includes three cohorts or groups of LFS for this period of 15 months. The first group starts in January 2009, the second one in February 2009 and the third one in March 2009. After the first measurement at the start of each group, the group is followed-up four times at three-month intervals. For each group the monthly ER data are linked, see Figure 2.1.

Both data sources include the variable of interest: ‘Contract type’. It is redefined having three categories: Permanent (1); Temporary (2); and Other (3). Here the category “Other” includes unemployment, self-employment and education. After linkage we have two variables ‘Contract Type LFS’ and ‘Contract Type ER’. We have obtained the proportions of respondents in the three categories of both these variables. These proportions are presented in Tables 2.1 and 2.2. For the ‘Contract Type ER’ table we have monthly figures for each group. There are noticeable but

[1] https://www.cbs.nl/en-gb/our-services/methods/surveys/korte-onderzoeksbeschrijvingen/dutch-labour-force-survey–lfs–

[2] One example of differences in definitions is an unemployed person. In the LFS an unemployed person is a person that works less then 4 hours a week and is actively looking for a job. This could be different from an ER unemployed person. In the ER an unemployed persons are persons that receive unemployment benefits.

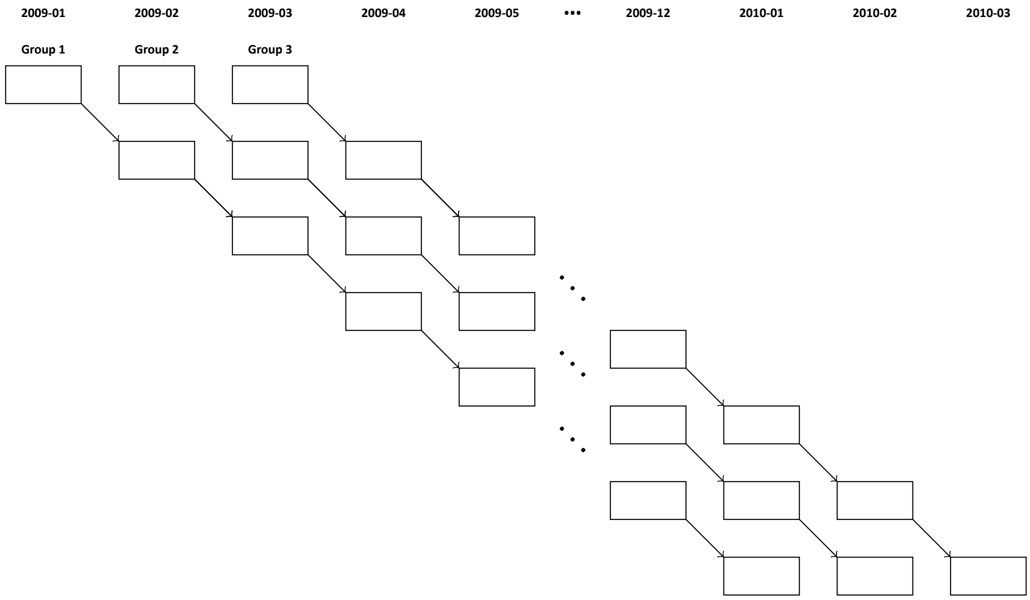

Figure 2.1 The ER monthly data for three groups.

small differences between groups. The proportions Temporary contracts are larger for group 2 than for the other two groups and group 3 has higher proportions of Permanent contracts than group 1 and 2.

Table 2.1 Proportions of contract type in the ER for the three groups.

N	Date	Group1			Group 2			Group 3		
		Perm.	Temp.	Oth.	Perm.	Temp.	Oth.	Perm.	Temp.	Oth.
1	2009 - 01	0.5858	0.1537	0.2605	-	-	-	-	-	-
2	2009 - 02	0.5901	0.1552	0.2547	0.5756	0.1629	0.2615	-	-	-
3	2009 - 03	0.5878	0.1562	0.2560	0.5752	0.1648	0.2600	0.5962	0.1545	0.2493
4	2009 - 04	0.5916	0.1508	0.2576	0.5741	0.1637	0.2621	0.6025	0.1512	0.2463
5	2009 - 05	0.5888	0.1549	0.2563	0.5727	0.1644	0.2628	0.6003	0.1501	0.2496
6	2009 - 06	0.5895	0.1517	0.2587	0.5712	0.1659	0.2629	0.5993	0.1498	0.2509
7	2009 - 07	0.5864	0.1473	0.2663	0.5683	0.1636	0.2682	0.5968	0.1468	0.2564
8	2009 - 08	0.5836	0.1489	0.2675	0.5697	0.1605	0.2698	0.5961	0.1472	0.2568
9	2009 - 09	0.5844	0.1474	0.2683	0.5673	0.1633	0.2694	0.5955	0.1472	0.2572
10	2009 - 10	0.5847	0.1461	0.2692	0.5703	0.1606	0.2691	0.5946	0.1457	0.2598
11	2009 - 11	0.5844	0.1492	0.2664	0.5676	0.1654	0.2669	0.5933	0.1439	0.2628
12	2009 - 12	0.5814	0.1479	0.2707	0.5693	0.1607	0.2700	0.5965	0.1402	0.2633
13	2010 - 01	0.5876	0.1415	0.2710	0.5763	0.1533	0.2704	0.5956	0.1414	0.2630
14	2010 - 02	-	-	-	0.5754	0.1483	0.2763	0.5967	0.1358	0.2675
15	2010 - 03	-	-	-	-	-	-	0.5981	0.1396	0.2624

In Table 2.2 we have the proportions of the 5 quarterly measurements of ‘Contract Type LFS’ for each group. We see relatively big differences between the ER and LFS. In particular, the proportions of permanent contracts are much higher in the LFS than in the ER while the number of temporary contracts is much lower.

The Tables 2.1 and 2.2 show that the different sources lead to different proportions for the three contract types for time points where both sources apply. We would like to obtain a single integrated estimator of proportions for ‘Contract type’ that combines the information of both sources.

Table 2.2 Proportions of contract type in the LFS for the three groups.

N	Date	Group1			Group 2			Group 3		
		Perm.	Temp.	Oth.	Perm.	Temp.	Oth.	Perm.	Temp.	Oth.
1	2009 - 01	0.6568	0.1117	0.2314	-	-	-	-	-	-
2	2009 - 02	-	-	-	0.6541	0.1159	0.2300	-	-	-
3	2009 - 03	-	-	-	-	-	-	0.6587	0.1121	0.2291
4	2009 - 04	0.6643	0.1050	0.2307	-	-	-	-	-	-
5	2009 - 05	-	-	-	0.6703	0.1059	0.2238	-	-	-
6	2009 - 06	-	-	-	-	-	-	0.6736	0.1034	0.2230
7	2009 - 07	0.6597	0.1070	0.2333	-	-	-	-	-	-
8	2009 - 08	-	-	-	0.6639	0.1066	0.2296	-	-	-
9	2009 - 09	-	-	-	-	-	-	0.6735	0.1030	0.2235
10	2009 - 10	0.6624	0.1039	0.2336	-	-	-	-	-	-
11	2009 - 11	-	-	-	0.6613	0.1168	0.2218	-	-	-
12	2009 - 12	-	-	-	-	-	-	0.6734	0.1044	0.2223
13	2010 - 01	0.6611	0.1008	0.2381	-	-	-	-	-	-
14	2010 - 02	-	-	-	0.6562	0.1148	0.2290	-	-	-
15	2010 - 03	-	-	-	-	-	-	0.6724	0.1009	0.2268

2.2 Parameters of interest: distributions per time-point and transitions

To describe the observed proportions of 'Contract type' for both sources and the integrated estimates of the corresponding population proportions, we introduce the following notation.

The measurements for a respondent i from group 1 at time-point t will be denoted by

R_i^t : register value of 'Contract type' for individual *i* at time-point t, for $t = 1, \dots, 13$,

S_i^t : survey value of 'Contract type' for individual *i* at time-point t, for $t \in \{1, 4, 7, 10, 13\}$,

with R_i^t and $S_i^t \in$ {Permanent, Temporary, Other}.

For group 2 we have $t = 2, \dots, 14$ for R^t and $t \in \{2, 5, 8, 11, 14\}$ for S^t. Similarly for group 3 we have $t = 3, \dots, 15$ for R^t and $t \in \{3, 6, 9, 12, 15\}$ for S^t. To keep it simple, from here on we will only deal with group 1. The results below hold similarly for groups 2 and 3.

For group 1 we denote the proportions for the 'Contract type' by:

$\mathbf{p}^{R.t} = (p_1^{R.t}, p_2^{R.t}, p_3^{R.t})$ for $t = 1, \dots, 13$: vectors with proportions from register data,

$\mathbf{p}^{S.t} = (p_1^{S.t}, p_2^{S.t}, p_3^{S.t})$ for $t \in \{1, 4, 7, 10, 13\}$: vectors with proportions from sample data.

Here subscripts define the contract types: $1 =$ Permanent, $2 =$ Temporary and $3 =$ Other. The primary parameters of interest are the population distributions over the three contract types for each month. Let us denote these unknown population proportion by:

$\pi^t = (\pi_1^t, \pi_2^t, \pi_3^t)$ for $t = 1, \dots, 13$.

The observed information from one group to estimate these population proportions is depicted in Table 2.3. Here we see that for time-points t and $t + 3$ both sources provide information to estimate the vector with population proportions, whereas for the times in between, no direct information from the survey is available.

Besides the marginal distribution π^t over the three contract types for each month, the changes between these categories over time are also of interest. These can be described by the transition

Table 2.3 Observed proportions from one group and population proportions for four consecutive time-points

...	$\mathbf{p}^{S.t}$			$\rightarrow$			$\mathbf{p}^{S.t+3}$	...
	$\downarrow$						$\downarrow$	
...	π^{t}	$\rightarrow$	π^{t+1}	$\rightarrow$	π^{t+2}	$\rightarrow$	π^{t+3}	...
	$\uparrow$		$\uparrow$		$\uparrow$		$\uparrow$	
...	$\mathbf{p}^{R.t}$	$\rightarrow$	$\mathbf{p}^{R.t+1}$	$\rightarrow$	$\mathbf{p}^{R.t+2}$	$\rightarrow$	$\mathbf{p}^{R.t+3}$	...

proportions between adjacent time points. That is, the proportions of individuals moving from one category at time-point t to another category at time point $t+1$, for each of the 3×3 category combinations. To define these transition proportions, we first consider the matrix with bivariate proportions corresponding to two variables, 'Contract type' at time-point t and 'Contract type' at time-point $t+1$. The matrix with population proportions for the cross-classification of these two variables will be denoted by $\pi^{t,t+1}$ with cell-proportions $\pi^{t,t+1}_{c_t,c_{t+1}}$ for $c_t, c_{t+1} \in$ {Permanent, Temporary, Other}.

In general, we will denote the proportions of units for all combinations of any number of variables by repeated use of superscripts and subscripts, i.e. $\pi^{t,t+1,t+2,\ldots}_{c_t,c_{t+1},c_{t+2},\ldots}$, and the corresponding matrices will be denoted by $\pi^{t,t+1,t+2,\ldots}$.

From each data source we can calculate the proportions of respondents for all possible combinations of 'Contract type' from month t till $t+3$ ($t \in \{1, 4, 7, 10, 13\}$). From the register data we observe also the intermediate months. These could be, for example, the proportion of persons that has "Permanent"→"Permanent"→"Permanent"→"Permanent" 'Contract type' during the months 1 till 4, or any other sequence of categories. We define these proportions by $\mathbf{p}^{R.t,t+1,t+2,t+3}_{c_1c_2c_3c_4}$ and the corresponding matrices by $p^{R.t,t+1,t+2,t+3}$. Observe that

$$\mathbf{p}^{R.t,t+3}_{c_1c_4} = \sum_{c_2,c_3} \mathbf{p}^{R.t,t+1,t+2,t+3}_{c_1c_2c_3c_4},$$

where $\mathbf{p}^{R.t,t+3}_{c_1c_4}$ stands for the proportions of register units that have respectively 'Contract type' c_1 and c_4 in month t and $t+3$. For the survey data we observe $\mathbf{p}^{S.t,t+3}_{c_1c_4}$ ($t \in \{1, 4, 7, 10, 13\}$) but do not have proportions for the intermediate months.

The observed values $\mathbf{p}^{R.t,t+3}_{c_1c_4}$ and $\mathbf{p}^{S.t,t+3}_{c_1c_4}$ are not consistent with each other. In the next section we define several reconciliation models that derive a single estimate $\pi^{t,t+3}_{c_tc_{t+3}}$ for these two proportions and for any other univariate or bivariate table of proportions.

3 Reconciliation methods

In this section several reconciliation methods will be presented that can be used to obtain univocal estimates of both marginal proportions and transitions. In general, reconciliation is applied when there is a need for comparing different data sets and when these data sets measure the same or related phenomenon and obtained estimates of aggregated quantities are different. Data reconciliation targets the correction of sampling or measurement errors that cause these differtences. When large systematic differences occur, with known causes, these need to be detected and corrected first before applying a reconciliation method, see e.g. Dagum and Cholette (2006).

Our aim is to have a single estimate for the variable ‘Contract type’ that combines available information from the monthly and quarterly data of ‘Contract type’ from ER and LFS and transition probabilities.

3.1 Objective function

Consider the four-way probability table corresponding to the four time-points t to $t+3$ as depicted in Table 2.3, with cell-probabilities $\pi^{t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}}$. All parameters of interest can be obtained as univariate and bivariate marginal tables with probabilities obtained from this four-way table. A reconciliation strategy is now to estimate this table $\tilde{\pi}^{t,t+1,t+2,t+3}$ using all the information from both sources, so that it is as close as possible to the register four-way proportions, $\mathbf{p}^{R.t,t+1,t+2,t+3}_{c_1 c_2 c_3 c_4}$ and has marginal probabilities $\tilde{\pi}^{t,t+3}$ which are as close as possible to the corresponding survey proportions, $\mathbf{p}^{S.t,t+3}_{c_1 c_4}$. From this table we can also obtain all required univariate and bivariate marginal probabilities.

This reconciliation problem is defined as constrained minimization problem of an objective function, which can be formulated in general as:

$$\begin{aligned} D = {} & D^R(\mathbf{p}^{R.t,t+1,t+2,t+3}, \tilde{\pi}^{t,t+1,t+2,t+3}) \\ & + D^S(\mathbf{p}^{S.t,t+3}, \tilde{\pi}^{t,t+3}) \text{ for } t \in (1,4,7,10) \end{aligned} \quad (1)$$

with D^R a measure of the discrepancy between the reconciled proportions $\tilde{\pi}^{t,t+1,t+2,t+3}$ and the observed register proportions $\mathbf{p}^{R.t,t+1,t+2,t+3}$ and D^S a measure of the discrepancy between the reconciled proportions $\tilde{\pi}^{t,t+3}$ and the observed survey proportions $\mathbf{p}^{S.t,t+3}$. Note that in D we have separate components for the reconciled estimates of the four dimensional table $\tilde{\pi}^{t,t+1,t+2,t+3}$ and its two-way margin $\tilde{\pi}^{t,t+3}$. To enforce that this two-way table is indeed a margin of the four-way table, we must have in any case the constraints:

$$\sum_{c_{t+1}, c_{t+2}} \tilde{\pi}^{t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}} = \tilde{\pi}^{t,t+3}_{c_t c_{t+3}}.$$

Different reconciliation models can be defined by different specifications of the objective function and constraints. Below we define a number of such models.

Constraints As mentioned above, since we have a linked data set on a micro level we mainly have to deal with measurement errors. These errors can occur in both ER and LFS variables for different reasons. ER variable ‘Contract type ER’ could have errors due to delayed updates, if for e.g. a person had a temporary contract at time t and got a permanent contract at time $t+1$. It could take some time to change his/her status in the register data. On the other hand if a person has a temporary contract with the intention of getting a permanent contract, he/she might fill in “permanent” in the LFS questionnaire.

For some of the reconciliation models we choose to adjust the register data and let the survey data remain unchanged. These models will have fixed survey-proportions and in (1) the component D^S vanishes. The constraints on the reconciled proportions $\pi^{t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}}$ for these models can be written as:

$$\sum_{c_{t+1}, c_{t+2}} \tilde{\pi}^{t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}} = p^{S.t,t+3}_{c_t c_{t+3}}, \quad t \in (1,4,7,10). \tag{2}$$

These constraints ensure that the bivariate marginal proportions for time points $t, t+3$ after reconciliation are equal to the observed survey-counterparts.

For models where both the register and the survey proportions can be adjusted, we have the following equality constraints:

$$\sum_{c_t c_{t+1} c_{t+2} c_{t+3}} \tilde{\pi}^{t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}} = 1,\ t \in (1,4,7,10), \tag{3a}$$

$$\sum_{c_{t+1}, c_{t+2}} \tilde{\pi}^{t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}} = \tilde{\pi}^{t,t+3}_{c_t c_{t+3}},\ t \in (1,4,7,10), \tag{3b}$$

$$\sum_{c_1} \tilde{\pi}^{t=1,t=4}_{c_1,c_4} = \sum_{c_7} \tilde{\pi}^{t=4,t=7}_{c_4 c_7}, \tag{3c}$$

$$\sum_{c_4} \tilde{\pi}^{t=4,t=7}_{c_4,c_7} = \sum_{c_{10}} \tilde{\pi}^{t=7,t=10}_{c_7 c_{10}}, \tag{3d}$$

$$\sum_{c_7} \tilde{\pi}^{t=7,t=10}_{c_7,c_{10}} = \sum_{c_{13}} \tilde{\pi}^{t=10,t=13}_{c_{10} c_{13}}. \tag{3e}$$

Constraint (3a) ensures that the probabilities $\tilde{\pi}$ can be interpreted as probabilities.
Constraint (3b) ensures that the table $\tilde{\pi}^{t,t+3}_{c_t c_{t+3}}$ is a bivariate margin of the four-way table $\tilde{\pi}^{t,t+1,t+2,t+3}$.
Constraints (3c)-(3e) enforce the equality of the common univariate margins of the bivariate tables.

3.2 Specific models

For the function D we will consider two alternatives. The first is the usual (weighted) sum of squares, which leads to additive adjustments and an explicit solution to the optimisation problem and the second is the Kullback-Leibler divergence function which leads to multiplicative adjustments, see e.g. Pannekoek and Zhang (2015). These adjustments cannot explicitly be calculated but a simple iterative algorithm can be used. Together with the two choices of constraints, we arrive at the following four reconciliation models.

Model 1. Additive adjustment of register only. This model treats the survey estimates as fixed quantities and only adjusts the register proportions. The objective function for this model is the commonly used least-squares loss-function:

$$D^R_{LS} = \sum_{c_t c_{t+1} c_{t+2} c_{t+3}} \left(p^{R.t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}} - \tilde{\pi}^{t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}}\right)^2. \tag{4}$$

This objective leads to additive adjustments. For the case of the fixed-sum constraints (2) these adjustments are very simple (see Appendix A): if a sum of register proportions is not equal to the corresponding survey proportion, a constant is added to all these register proportions such that the sum does equal the survey proportion:

$$\tilde{\pi}^{t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}} = p^{R.t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}} + \left(p^{S.t,t+3}_{c_t c_{t+3}} - p^{R.t,t+3}_{c_t c_{t+3}}\right) / \, 9 \qquad (5)$$
$$= p^{R.t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}} + a^{t,t+3}_{c_t c_{t+3}}, \text{ say.} \quad t \in (1,4,7,10)$$

Summation of (5) over c_{t+1} and c_{t+2} (9 components) shows that (2) is verified.

For the univariate reconciled proportions we have

$$\tilde{\pi}^{t}_{c_t} = p^{S.t}_{c_t} \quad \text{for } t \in (1,4,7,10,13), \qquad (6)$$
$$\tilde{\pi}^{t}_{c_t} = p^{R.t}_{c_t} \quad \text{for } t \notin (1,4,7,10,13),$$

where the first line is a consequence of (2) and the second line is true as summing (5) over c_t and c_{t+3} yields $p^{R.t+1,t+2}_{c_{t+1} c_{t+2}}$, because $\sum_{c_t c_{t+3}} p^{S.t,t+3}_{c_t c_{t+3}} = \sum_{c_t c_{t+3}} p^{R.t,t+3}_{c_t c_{t+3}} = 1$.

For the adjusted transition proportions for consecutive time-points we can write,

$$\tilde{\pi}^{t,t+1}_{c_t c_{t+1}} = p^{R.t,t+1}_{c_t c_{t+1}} + \left(p^{S.t}_{c_t} - p^{R.t}_{c_t}\right) / \, 3 \qquad (7)$$
$$\tilde{\pi}^{t+1,t+2}_{c_{t+1} c_{t+2}} = p^{R.t+1,t+2}_{c_{t+1} c_{t+2}}$$
$$\tilde{\pi}^{t+2,t+3}_{c_{t+2} c_{t+3}} = p^{R.t+2,t+3}_{c_{t+2} c_{t+3}} + \left(p^{S.t+3}_{c_{t+3}} - p^{R.t+3}_{c_{t+3}}\right) / \, 3,$$

with $p^{S.t}_{c_t}$ and $p^{S.t+3}_{c_{t+3}}$ the row- and column marginals of $p^{S.t,t+3}_{c_t c_{t+3}}$ and $p^{R.t}_{c_t}$ and $p^{R.t+3}_{c_{t+3}}$ the row- and column marginals of $p^{R.t,t+3}_{c_t c_{t+3}}$.

This shows that the univariate reconciled distributions of 'Contract type' for the time-points *t* and *t+3* are equal to those of the survey, which is an obvious consequence of the constraint (2), whereas for the intermediate time points the distributions are equal to the register ones. For the bivariate distribution or transition probabilities, we see that the reconciled transitions from *t* to *t+1* and from *t+2* to *t+3* will differ from the register ones, whereas the intermediate transitions (from *t+1* to *t+2*) will not change. The changes are such that to each row (or column) of the transition table a constant is added. This may be a counterintuitive result, but it follows directly from the cancellation of adjustments when summing (5) over c_t and c_{t+3}.

From the reconciled transition table we can easily obtain marginal proportions. These reconciled marginal time-point distributions are presented in table 3.1. These results confirm that for time-points t and $t+3$ the proportions equal the ones obtained from the survey while for the intermediate time-points the proportions are equal to the register proportions.

Note that this is rather a naive model. The reconciled series have large jumps and are therefore not realistic.

Model 2. Multiplicative adjustment of register only. Just as Model 1, this model treats the survey estimates as fixed quantities and only adjusts the register proportions. The objective function for this model is the Kullback-Leibler divergence or relative entropy. Although this objective function may not be the most common one, the resulting adjustment method is perhaps the most intuitive one. It leads to multiplicative adjustments that guarantee non-negative outcomes and are easy to understand and apply. The Kullback-Leibler divergence is given by

Table 3.1 Reconciled proportions of contract types according to Model 1

Month	Group1			Group 2			Group 3		
	Perm.	Temp.	Oth.	Perm.	Temp.	Oth.	Perm.	Temp.	Oth.
1	0.6568	0.1117	0.2314	-	-	-	-	-	-
2	0.5899	0.1553	0.2548	0.6541	0.1159	0.2300	-	-	-
3	0.5877	0.1562	0.2561	0.5752	0.1649	0.2598	0.6587	0.1121	0.2291
4	0.6643	0.1050	0.2307	0.5745	0.1639	0.2616	0.6032	0.1514	0.2454
5	0.5888	0.1549	0.2563	0.6703	0.1059	0.2238	0.6010	0.1502	0.2488
6	0.5894	0.1518	0.2588	0.5716	0.1660	0.2624	0.6736	0.1034	0.2230
7	0.6597	0.1070	0.2333	0.5695	0.1636	0.2670	0.5972	0.1469	0.2559
8	0.5838	0.1489	0.2673	0.6639	0.1066	0.2296	0.5961	0.1473	0.2566
9	0.5844	0.1474	0.2683	0.5673	0.1634	0.2692	0.6735	0.1030	0.2235
10	0.6624	0.1039	0.2336	0.5705	0.1606	0.2689	0.5950	0.1458	0.2592
11	0.5845	0.1494	0.2661	0.6613	0.1168	0.2218	0.5935	0.1439	0.2626
12	0.5817	0.1482	0.2702	0.5694	0.1609	0.2697	0.6734	0.1044	0.2223
13	0.6611	0.1008	0.2381	0.5764	0.1535	0.2701	0.5961	0.1415	0.2625
14	-	-	-	0.6562	0.1148	0.2290	0.5972	0.1359	0.2669
15	-	-	-	-	-	-	0.6724	0.1009	0.2268

$$D_{KL}^{R} = \sum_{c_t c_{t+1} c_{t+2} c_{t+3}} \tilde{\pi}_{c_t c_{t+1} c_{t+2} c_{t+3}}^{t,t+1,t+2,t+3} \left(\log(\tilde{\pi}_{c_t c_{t+1} c_{t+2} c_{t+3}}^{t,t+1,t+2,t+3} / p_{c_t c_{t+1} c_{t+2} c_{t+3}}^{R.t,t+1,t+2,t+3}) - 1\right). \tag{8}$$

We show in Appendix A that the solution to minimizing D_{KL}^{R} subject to the constraints (2) can be obtained by multiplying the original register proportions by factors varying only over combinations of categories of t and $t+3$ but remaining constant over the combinations of categories of $t+1$ and $t+2$:

$$\begin{aligned} \tilde{\pi}_{c_t c_{t+1} c_{t+2} c_{t+3}}^{t,t+1,t+2,t+3} &= p_{c_t c_{t+1} c_{t+2} c_{t+3}}^{R.t,t+1,t+2,t+3} \times p_{c_t c_{t+3}}^{S.t,t+3} / p_{c_t c_{t+3}}^{R.t,t+3} \\ &= p_{c_t c_{t+1} c_{t+2} c_{t+3}}^{R.t,t+1,t+2,t+3} \times f_{c_t c_{t+3}}^{t,t+3}, \quad t \in (1,4,7,10). \end{aligned} \tag{9}$$

Summation of (9) over c_{t+1} and c_{t+2} shows that (2) is verified.

For the adjusted univariate proportions we then have, for $t \in (1,4,7,10)$

$$\begin{aligned} \tilde{\pi}_{c_t}^{t} &= p_{c_t}^{S.t}, \\ \tilde{\pi}_{c_{t+1}}^{t+1} &= \sum_{c_t c_{t+3}} \left(f_{c_t c_{t+3}}^{t,t+3} \times \sum_{c_{t+2}} p_{c_t c_{t+1} c_{t+2} c_{t+3}}^{R.t,t+1,t+2,t+3} \right), \\ \tilde{\pi}_{c_{t+2}}^{t+2} &= \sum_{c_t c_{t+3}} \left(f_{c_t c_{t+3}}^{t,t+3} \times \sum_{c_{t+1}} p_{c_t c_{t+1} c_{t+2} c_{t+3}}^{R.t,t+1,t+2,t+3} \right), \\ \tilde{\pi}_{c_{t+3}}^{t+3} &= p_{c_{t+3}}^{S.t+3}. \end{aligned} \tag{10}$$

Here we see that, contrary to the additive adjustments, none of the univariate distributions remains equal to the direct register estimates. The distributions for the intermediate time-points $t+1$ and $t+2$ are influenced by the ratios $p_{c_t c_{t+3}}^{S.t,t+3} / p_{c_t c_{t+3}}^{R.t,t+3}$.

For the reconciled transition proportions for consecutive time-points we can write

$$\tilde{\pi}^{t,t+1}_{c_t c_{t+1}} = \sum_{c_{t+3}} \left(f^{t,t+3}_{c_t c_{t+3}} \times \sum_{c_{t+2}} p^{R.t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}} \right), \tag{11}$$

$$\tilde{\pi}^{t+1,t+2}_{c_{t+1} c_{t+2}} = \sum_{c_t c_{t+3}} \left(f^{t,t+3}_{c_t c_{t+3}} \times p^{R.t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}} \right),$$

$$\tilde{\pi}^{t+2,t+3}_{c_{t+2} c_{t+3}} = \sum_{c_t} \left(f^{t,t+3}_{c_t c_{t+3}} \times \sum_{c_{t+1}} p^{R.t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}} \right),$$

where we see again that, contrary to the additive adjustments, there are now no transitions that remain equal to the observed register transitions: all transitions are affected by the differences between survey and register proportions as expressed by the factors $f^{t,t+3}_{c_t c_{t+3}}$.

The results for the marginal time-point distributions for this adjustment model are in table 3.2. These results show that the artificial 'jumps' in the distributuions for the months where survey estimates are available, that were found for model 1, have now disappeared. Not surprisingly, we also see that the level of the intermediate proportions is now much closer to the ones obtained from the survey.

Table 3.2 Reconciled proportions of contract types according to Model 2

Month	Group1			Group 2			Group 3		
	Perm.	Temp.	Oth.	Perm.	Temp.	Oth.	Perm.	Temp.	Oth.
1	0.6568	0.1117	0.2314	-	-	-	-	-	-
2	0.6581	0.1120	0.2299	0.6541	0.1159	0.2300	-	-	-
3	0.6557	0.1117	0.2326	0.6539	0.1166	0.2296	0.6587	0.1121	0.2291
4	0.6643	0.1050	0.2307	0.6546	0.1146	0.2308	0.6672	0.1074	0.2254
5	0.6617	0.1107	0.2276	0.6703	0.1059	0.2238	0.6659	0.1066	0.2275
6	0.6628	0.1084	0.2287	0.6681	0.1076	0.2243	0.6736	0.1034	0.2230
7	0.6597	0.1070	0.2333	0.6650	0.1067	0.2282	0.6693	0.1030	0.2277
8	0.6575	0.1083	0.2343	0.6639	0.1066	0.2296	0.6661	0.1055	0.2284
9	0.6572	0.1071	0.2356	0.6630	0.1104	0.2266	0.6735	0.1030	0.2235
10	0.6624	0.1039	0.2336	0.6645	0.1105	0.2250	0.6703	0.1044	0.2253
11	0.6625	0.1049	0.2325	0.6613	0.1168	0.2218	0.6677	0.1046	0.2278
12	0.6594	0.1039	0.2367	0.6607	0.1153	0.2240	0.6734	0.1044	0.2223
13	0.6611	0.1008	0.2381	0.6592	0.1161	0.2247	0.6709	0.1040	0.2250
14	-	-	-	0.6562	0.1148	0.2290	0.6715	0.0995	0.2290
15	-	-	-	-	-	-	0.6724	0.1009	0.2268

Model 3. Additive adjustment of both survey and register For this model we consider the least-squares objective with components D^R_{LS} and D^S_{LS} :

$$D_{LS} = \sum_{c_t c_{t+1} c_{t+2} c_{t+3}} \left(p^{R.t,t+1,t+2,t+3} - \tilde{\pi}^{t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}} \right)^2 + \sum_{c_t c_{t+3}} \left(p^{S.t,t+3}_{c_t c_{t+3}} - \tilde{\pi}^{t,t+3}_{c_t c_{t+3}} \right)^2 . \tag{12}$$

This function is minimised over $\tilde{\pi}^{t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}}$ and $\tilde{\pi}^{t,t+3}_{c_t c_{t+3}}$ subject to the constraints (3). The solution to this problem can still be obtained in closed form, but is slightly more involved and is given in Appendix A.

The results for the marginal time-point distributions for this adjustment model are in table 3.3. These results show that, similarly to the additive model 1, the proportions for the intermediate time-points are equal the the register proportions but, contrary to model 1, the proportions for the time-points where survey estimates are available are not equal to these survey estimates but are in-between the survey and register estimates.

Table 3.3 Reconciled proportions of contract types according to Model 3

Month	Group1			Group 2			Group 3		
	Perm.	Temp.	Oth.	Perm.	Temp.	Oth.	Perm.	Temp.	Oth.
1	0.6497	0.1159	0.2343	-	-	-	-	-	-
2	0.5899	0.1553	0.2548	0.6463	0.1206	0.2330	-	-	-
3	0.5877	0.1562	0.2561	0.5752	0.1649	0.2598	0.6526	0.1164	0.2311
4	0.6542	0.1102	0.2356	0.5745	0.1639	0.2616	0.6032	0.1514	0.2454
5	0.5888	0.1549	0.2563	0.6537	0.1150	0.2313	0.6010	0.1502	0.2488
6	0.5894	0.1518	0.2588	0.5716	0.1660	0.2624	0.6629	0.1091	0.2279
7	0.6526	0.1106	0.2368	0.5695	0.1636	0.2670	0.5972	0.1469	0.2559
8	0.5838	0.1489	0.2673	0.6549	0.1115	0.2336	0.5961	0.1473	0.2566
9	0.5844	0.1474	0.2683	0.5673	0.1634	0.2692	0.6618	0.1089	0.2293
10	0.6522	0.1089	0.2389	0.5705	0.1606	0.2689	0.5950	0.1458	0.2592
11	0.5845	0.1494	0.2661	0.6529	0.1204	0.2267	0.5935	0.1439	0.2626
12	0.5817	0.1482	0.2702	0.5694	0.1609	0.2697	0.6633	0.1078	0.2289
13	0.6537	0.1048	0.2414	0.5764	0.1535	0.2701	0.5961	0.1415	0.2625
14	-	-	-	0.6481	0.1182	0.2337	0.5972	0.1359	0.2669
15	-	-	-	-	-	-	0.6649	0.1047	0.2303

Model 4. Multiplicative adjustment of both survey and register For this model we consider a Kullback-Leibler objective that contains both the D_{KL}^{R} and D_{KL}^{S} components:

$$D_{KL} = \sum_{c_t c_{t+1} c_{t+2} c_{t+3}} \tilde{\pi}^{t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}} \left(\log(\tilde{\pi}^{t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}} / p^{R.t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}}) - 1\right)$$
$$+ \sum_{c_t c_{t+3}} \tilde{\pi}^{t,t+3}_{c_t c_{t+3}} \left(\log(\tilde{\pi}^{t,t+3}_{c_t c_{t+3}} / p^{S.t,t+3}_{c_t c_{t+3}}) - 1\right). \tag{13}$$

Again this function is minimised over $\tilde{\pi}^{t,t+1,t+2,t+3}_{c_t c_{t+1} c_{t+2} c_{t+3}}$ and $\tilde{\pi}^{t,t+3}_{c_t c_{t+3}}$ and is subject to the constraints (3). The solution to this problem cannot be given in closed form, but a convenient iterative algorithm is given in Appendix A.

The results for the marginal time-point distributions for this adjustment model are in Table 3.4. These results show that the artificial 'jumps' for the months where survey estimates are available, that were found for the additive models (for model 3 to a lesser extent than for model 1), have disappeared (just as for the multiplicative model 2). But, contrary to model 2, the level of all proportions has moved from what was found in the survey towards the level found in the register. For the category permanent contracts this means a decrease while for temporary contracts this is an increase.

Table 3.4 Adjusted monthly proportions of contract types in Model 4

Month	Group1			Group 2			Group 3		
	Perm.	Temp.	Oth.	Perm.	Temp.	Oth.	Perm.	Temp.	Oth.
1	0.6247	0.1311	0.2441	-	-	-	-	-	-
2	0.6275	0.1319	0.2406	0.6213	0.1354	0.2433	-	-	-
3	0.6254	0.1321	0.2425	0.6207	0.1366	0.2427	0.6347	0.1296	0.2357
4	0.6286	0.1267	0.2447	0.6208	0.1350	0.2442	0.6421	0.1257	0.2322
5	0.6259	0.1316	0.2425	0.6205	0.1344	0.2451	0.6404	0.1247	0.2349
6	0.6269	0.1289	0.2442	0.6188	0.1360	0.2452	0.6400	0.1242	0.2358
7	0.6242	0.1254	0.2504	0.6163	0.1344	0.2493	0.6367	0.1227	0.2406
8	0.6220	0.1267	0.2512	0.6167	0.1322	0.2511	0.6345	0.1243	0.2412
9	0.6223	0.1254	0.2524	0.6153	0.1357	0.2491	0.6334	0.1248	0.2418
10	0.6223	0.1240	0.2537	0.6174	0.1347	0.2479	0.6312	0.1251	0.2437
11	0.6223	0.1258	0.2519	0.6156	0.1388	0.2457	0.6292	0.1245	0.2463
12	0.6192	0.1246	0.2561	0.6156	0.1359	0.2484	0.6306	0.1227	0.2467
13	0.6224	0.1202	0.2574	0.6171	0.1339	0.2490	0.6290	0.1228	0.2482
14	-	-	-	0.6146	0.1313	0.2541	0.6298	0.1177	0.2525
15	-	-	-	-	-	-	0.6307	0.1200	0.2493

4 Comparison with results from the measurement error model of Pankowska et al.

The macro-integration or reconciliation approach used in this paper is one of the different methodologies that combine information from multiple sources into single estimates. Another approach is the use of measurement error models. As already mentioned above, these models assume that the estimates from different sources differ from the true values (and each other) due to (differences in) measurement errors. The true variable is called a latent variable. Estimates of proportions based on the true values will provide single estimates of the corresponding population values that are corrected for measurement errors. For longitudinal categorical data hidden or latent Markov models have been proposed, see Pavlopoulos and Vermunt (2015) and Pankowska et al. (2018). Pavlopoulos and Vermunt (2015) analysed similar to our data, but for

Table 4.1 Proportions of 'Contract type' in the LFS for pooled data.

Month	Permanent	Temporary	Other
1	0.6566	0.1131	0.2303
2	-	-	-
3	-	-	-
4	0.6692	0.1048	0.2261
5	-	-	-
6	-	-	-
7	0.6655	0.1056	0.2289
8	-	-	-
9	-	-	-
10	0.6656	0.1081	0.2262
11	-	-	-
12	-	-	-
13	0.6632	0.1052	0.2316

the earlier period of between January 2007 and March 2009 and Pankowska et al. (2018) for the period between January 2009 and March 2010. The data set used by Pankowska et al. (2018) is the same as the one used for the analyses in this paper and it is therefore of interest to compare the results. For this comparison we shall first discuss the differences in combining the data from the different groups, then highlight some properties of the latent Markov model that make this methodology different from our approach and finally compare the numerical results.

4.1 Differences in combining the groups

In our analyses we consider three different groups, that are defined by the starting point of the LFS panel. The groups entered the LFS in January, February and March of 2009. We applied reconciliation models separately to each group. In Pankowska et al. (2018) these three groups were pooled together. Authors ignore time, by combining persons that start in January, February and March in one group. In order to compare our results with the results from Pankowska et al. (2018) we combined the data in the same way. We call this data the pooled data. Tables 4.1 and 4.2 show the monthly distributions of pooled data for LFS and ER. In Table 2.1 we observe that the group 3 has slightly higher proportions of persons with permanent contract than the group 1 and group 2 has slightly lower proportions of persons with permanent contract than group 1. All three groups show decline in temporary contract. In the pooled data these differences for permanent contract are averaged. We still observe decline in temporary contract.

Table 4.2 Proportions of 'Contract type' in the ER for pooled data.

Month	Permanent	Temporary	Other
1	0.5861	0.1571	0.2568
2	0.5892	0.1572	0.2536
3	0.5875	0.1569	0.2556
4	0.5880	0.1550	0.2570
5	0.5858	0.1561	0.2582
6	0.5850	0.1542	0.2608
7	0.5838	0.1517	0.2645
8	0.5819	0.1527	0.2654
9	0.5827	0.1506	0.2667
10	0.5830	0.1508	0.2662
11	0.5831	0.1507	0.2662
12	0.5847	0.1461	0.2691
13	0.5869	0.1433	0.2698

4.2 The latent Markov model approach versus macro-integration approach

The latent Markov model can be described globally with the help of the diagram below (see Table 4.3), which is taken (with some modifications) from Pavlopoulos and Vermunt (2015). It is also similar to the picture of the macro-integration approach in Table 2.3 in Section 2.2 but, contrary to that picture, it describes measurements and true values at the individual level rather than the observed proportions and unknown population proportions aggregated over units.

Here the S^t denote the quarterly survey measurements of 'Contract type' for any individual and the R^t denote the register measurements. The X^t denotes a persons true contract status which is an unobserved or latent variable. Associated with the latent variables are the contract status probability vectors π^t, that have three components for $Permanent, Temporary$ or $Other$

Table 4.3 Path diagram for the latent Markov model with two (partially) observed indicators

…	S^t						S^{t+3}	…
	↑						↑	
…	X^t	→	X^{t+1}	→	X^{t+2}	→	X^{t+3}	…
	↓		↓		↓		↓	
…	R^t	→	R^{t+1}	→	R^{t+2}	→	R^{t+3}	…

values. Each of the cells in this table represents a specific sequence of categories of X in time (a path) and the probability of a specific path is the cell probability, denoted by $\pi_{c_1 c_2 \dots c_T}$. The arrows between the X^t indicate that the 'Contract type' at time-point t depends on the 'Contract type' of the previous time-point $t-1$, but given the 'Contract type' at $t-1$ it does not depend on any of the time-points before $t-1$. This is the Markov property, which means that,

$$\Pi^{t|t-1,t-2,\dots,1}_{c_t,c_{t-1},c_{t-2},\dots,c_1} = \Pi^{t|t-1}_{c_t,c_{t-1}}. \quad (14)$$

Here $\Pi^{t|t-1}_{c_t,c_{t-1}}$ and $\Pi^{t|t-1,t-2,\dots,1}_{c_t,c_{t-1},c_{t-2},\dots,c_1}$ are transition probabilities. $\Pi^{t|t-1}_{c_t,c_{t-1}}$ is a transition probability of the units in each category of c_{t-1} at time-point $t-1$ to the categories of c_t at time-point t. These probabilities are conditional probabilities, summing to 1 for each category of c_t and are defined by:

$$\Pi^{t|t-1}_{c_t,c_{t-1}} = \pi^{t,t-1}_{c_t,c_{t-1}} / \pi^{t-1}_{c_{t-1}}. \quad (15)$$

Similarly $\Pi^{t|t-1,t-2,\dots,1}_{c_t,c_{t-1},c_{t-2},\dots,c_1}$ denote the transition probabilities of units for all combinations of any number of variables by repeated use of superscripts and subscripts.

The consequence of the Markov property in (14) is that the probability of any possible sequence of categories of X in time (a path), is completely specified by the initial state probability and the transition probabilities, since:

$$\pi^{1,\dots,T}_{c_1 \dots c_T} = \pi^{1,\dots,T-1}_{c_1,\dots,c_{T-1}} \Pi^{T|1,\dots,T-1}_{c_T,c_1,\dots,c_{T-1}} = \pi^{1,\dots,T-1}_{c_1,\dots,c_{T-1}} \Pi^{T|T-1}_{c_T,c_{T-1}}$$

and, by recursion on the first term on the right hand site,

$$\pi^{1,\dots,T}_{c_1 \dots c_T} = \pi^1_{c_1} \Pi^{2|1}_{c_2,c_1} \Pi^{3|2}_{c_3,c_2} \dots \Pi^{T|T-1}_{c_T,c_{T-1}}. \quad (16)$$

According to (15) we also have that the transition matrices can be obtained from the bivariate probabilities $\pi^{t+1,t}$ and so, by the Markov assumption, the cell probabilities are completely determined by the bivariate probabilities.

The model links the observed values from each of the sources to the true values by a matrix with misclassification probabilities that gives the probability of an observed category of R^t or S^t given X^t, the diagonal entries of these matrices are the probabilities of correct classification. These matrices, denoted by $p(R^t|X^t)$ and $p(S^t|X^t)$, together with the transition probabilities $\Pi^{t|t-1}_{c_t,c_{t-1}}$, are the parameters of the model and can be estimated provided that some restrictions are imposed. Usually it is assumed that the error matrices $p(R^t|X^t)$ and $p(S^t|X^t)$ only depend on X^t and not on other true or observed values, but Pankowska et al. (2018) use an extensions of the model that allows for certain correlations between classification errors at successive time-points.

In the model-based approach the observed values S^t and R^t are thought to be generated as a function of the true values X^t and added measurement errors, hence the direction of the arrows. The reconciliation approach is not based on a data generating model at the individual level, but begins with estimates at the macro-level and combines these different estimates into a single one. Thus, in the reconciliation approach the observed estimates are the inputs from which the reconciliation model can produce the single estimates whereas in the model based approach the single true values are thought to produce the different measurements. To reflect this difference in reasoning, the direction of the arrows is reversed for the reconciliation approach.

Despite these conceptual differences, both approaches will result in a single estimate of each of the parameters of interest: the univariate and bivariate proportions (or probabilities) of the 'Contract type' categories over time.

4.3 Numerical results

In order to compare results from to different approaches we run the Model 4 defined in Section 3.2 on the pooled data. Results of the reconciled time series for the pooled data are given in Appendix B.

In Table 10 in Pankowska et al. (2018) the average size of 'Contract type' according to the LFS, ER and their optimal latent model, called C", are presented. Here we compare these results with the results we obtained from Model 4. In our model the average size of 'Contract type' is obtained in a different way. In Pankowska et al. (2018) the model estimates new values on a micro level, whereas our models produce adjusted values per month for total group, on a macro level. The average size of 'Contract type' for model 4 in Table 4.4 here is the average over 13 months for Model 4. Observe that the adjusted figures for Model 4 are closer to the average values obtained

Table 4.4 The average size of 'Contract type' according to model C" and model 4

	Survey	Register	Model C"	Model 4
Permanent	0.653	0.585	0.611	0.625
Temporary	0.110	0.151	0.128	0.129
Other	0.237	0.264	0.261	0.246

from these two sources. This is not surprising since in our optimization function we include all initial values as equally reliable. The latent class model gives slightly different results, with the largest deviation from Model 4 in the category "Other".

As mentioned in section 2.2, interest is not only in the marginal distribution π^t over the three contract types for each month, the changes between these categories over time are also of interest. These can be described by the transition probabilities between adjacent time points. Estimates of these probabilities are shown in table 4.5. This table compares conditional probabilities $P(c_t|c_{t-3})$ for observed survey data, observed register data, Model 4 results for pooled data and results from Table 11 in Pankowska et al. (2018). We again observe that the macro-integration model estimates are in between the observed values, but now this is not true for the estimates from Model C". The diagonal value estimates from model C" appear to be overall higher than the observed values in either LFS or ER while the transition probabilities are lower according to this model.

Table 4.5 Conditional probabilities of 'Contract type'

Observed transitions from LFS			
	Contract t		
Contract t-3	Permanent	Temporary	Other
Perm.	0.983	0.006	0.011
Temp	0.058	0.879	0.063
Oth.	0.016	0.037	0.947

Observed transitions from ER register data			
	Contract t		
Contract t-3	Permanent	Temporary	Other
Perm.	0.976	0.012	0.012
Temp	0.073	0.869	0.058
Oth.	0.019	0.043	0.938

Model 4 for pooled data			
	Contract t		
Contract t-3	Permanent	Temporary	Other
Perm.	0.979	0.008	0.012
Temp	0.062	0.874	0.064
Oth.	0.016	0.036	0.948

Model C" in Pankowska et al. (2018), Table 11			
	Contract t		
Contract t-3	Permanent	Temporary	Other
Perm.	0.987	0.004	0.009
Temp	0.017	0.929	0.054
Oth.	0.006	0.030	0.963

5 Discussion

The goal of this research was to investigate if macro-integration methods can contribute to the reconciliation process of the labour force statistics. Four different macro-integration models were defined in Section 3. We presented adjusted proportions of the variable 'Contract type' for these models, see Tables 3.1 - 3.4. From these results, it is easy to argue that model 4 shows the best results. First we observed that the multiplicative adjustments were superior to additive ones. The additive model adjusted only quarterly proportions, while the proportions of the intermediate months were not adjusted, see tables 3.1 and 3.3, resulting in unrealistic monthly time series. Secondly, we assumed that both sources are equally reliable or equally wrong, since we do not have any evidence for one source to be more reliable than the other. Thus we have chosen for a multiplicative adjustment model with adjustments to both sources, denoted as Model 4.

Our macro-integration models resulted in estimated proportions for each group. These estimates can easily be combined to form an overall estimate. From the proportions for the groups we first derive estimates for the totals for each group see Table 5.1. We then combine the totals of all three groups for each month and calculate the adjusted total proportions, see Table 5.2.

For this research a tailored data set with linked individuals for several years was made available to us. In practice it will not be easy to link these data within a reasonable time frame, in order to obtain up to date results. The ER has timeliness issues, it is a complex register including information from different institutions; not all information is available each month. On the other hand, the LFS is a quarterly rotating panel survey, with figures available for each month, see

Table 5.1 Estimated monthly totals of contract types in Model 4

group	1			2			3		
Contract type	Perm.	Temp.	Oth.	Perm.	Temp.	Oth.	Perm.	Temp.	Oth.
Month 1	1983.5	416.4	775.2	-	-	-	-	-	-
2	1992.9	418.9	764.2	1770.1	385.9	693.1	-	-	-
3	1986.2	419.4	770.3	1766.6	388.7	690.8	1708.5	348.9	634.6
4	1996.4	402.3	777.3	1766.9	384.2	694.9	1728.5	338.3	625.2
5	1988.0	417.9	770.2	1765.9	382.6	697.4	1723.8	335.8	632.3
6	1991.7	409.4	776.0	1760.6	386.9	697.5	1721.6	334.1	634.3
7	1983.0	398.4	795.5	1755.8	382.8	710.4	1713.4	330.1	647.5
8	1976.2	402.6	798.1	1755.7	376.5	714.8	1707.4	334.6	649.1
9	1976.4	398.1	801.5	1751.6	386.3	709.1	1703.9	335.7	650.5
10	1976.5	393.7	805.8	1757.1	383.3	705.6	1698.6	336.7	655.7
11	1976.3	399.6	800.0	1752.5	395.0	699.5	1692.5	334.9	662.7
12	1967.2	396.0	813.8	1750.9	386.6	706.6	1695.7	329.9	663.5
13	1975.5	381.6	816.8	1755.0	380.8	708.2	1690.7	330.1	667.2
14	-	-	-	1748.5	373.5	723.0	1693.0	316.3	678.7
15	-	-	-	-	-	-	1694.7	322.6	669.7

Table 5.2 Estimates of combined monthly proportions of contract types.

Contract type	Perm.	Temp.	Oth.
Month 1	0.6247	0.1311	0.2441
2	0.6246	0.1336	0.2419
3	0.6267	0.1328	0.2405
4	0.6302	0.1291	0.2407
5	0.6286	0.1304	0.2410
6	0.6283	0.1297	0.2419
7	0.6255	0.1275	0.2470
8	0.6241	0.1278	0.2481
9	0.6234	0.1286	0.2480
10	0.6235	0.1278	0.2487
11	0.6222	0.1296	0.2482
12	0.6216	0.1277	0.2507
13	0.6227	0.1255	0.2518
14	0.6220	0.1247	0.2533
15	0.6307	0.1200	0.2493

section 2.1. In estimating population proportions, differences between the sources can be caused by differences in definitions, differences in population coverage and systematic errors and other large discrepancies. Coverage problems are completely avoided when using only the linked data with the same set of units from both sources. However, the discrepancies that we observed may still be caused by different definitions or systematic errors. These possible causes of discrepancies should be investigated thoroughly because macro-integration methods should be applied only after the data are cleaned from gross errors.

In Section 4 we described results obtained by Pankowska et al. (2018) by using a different approach on the same data set, the latent class model with the Markov property. In comparison to the macro-integration approach we proposed here, the individual level latent class model-based approach is more flexible because it allows the transition probabilities to vary across individuals according to covariates, which gives insight into the differences in transitions between subgroups and may also result, depending on the validity of the model assumptions, in more accurate estimates at the aggregate level. Moreover, contrary to the macro-integration approach, it also provides estimates of the misclassification probabilities. On the other hand, the macro-integration approach is less restrictive in the sense that it does not impose a Markov property of the integrated time series of proportions that it produces. Moreover, as the inputs for this method are population estimates, the method does not need sources that can be linked at the unit level. The method is also easy to extend and implement. In our model we did not include reliability weights for the data sources. If reliability information of some sort is available, reliability weights should be included in the model. One of the possibilities would be to use the estimates of the misclassification probabilities from the latent class Markov model to define the reliability weights. Another nice feature of our method is the simple algorithm that is easy to implement and runs fast. For the iterative algorithm described in our paper for Model 4 the elapsed system time in R was 2.5 seconds when we required the precision of 1e-09 for the constraints.

Acknowledgement: The authors thank Arnout van Delden and Sander Scholtus for useful comments and suggestions for improvement.

References

Bakker, B. (2011). Micro integration: State of the art. Technical report, ESSnet on Data Integration Report of WP1. State of the art on statistical methodologies for data integration, Eurostat.

Censor, Y. and S. Zenios (1997). *Parallel Optimization. Theory, Algorithms, and Applications.* Oxford University Press, New York.

Cobben, F. (2009). Nonresponse in sample surveys : methods for analysis and adjustment. Technical report, Statistics Netherlands.

Dagum, E. and P. Cholette (2006). *Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series*. Springer-Verlag, 186, Lecture Notes in Statistics, New York.

de Waal, T., A. van Delden, and S. Scholtus (2019). Multi-source statistics: Basic situations and methods. *International Statistical Review*. accepted for publication.

Fosen, J. (2011). Register-based employment statistics. a case of microintegration. the micro-integration process and life-cycle from a quality perspective. Technical report, Statistics Norway. `https://ec.europa.eu/eurostat/cros/system/files/ESSnet\%20DI\%20Wp4.1\%20Micro\%20integration\%20and\%20life\%20cycle.pdf`.

Fosen, J. and L.-C. Zhang (2011). Quality evaluation of employment status in register-based census. Technical report, Int. Statistical Inst.: Proc. 58th World Statistical Congress, 2011, Dublin (Session STS024). `http://2011.isiproceedings.org/papers/650264.pdf`.

Goni, E., A. Serrano, and J. Aramendi (2019). Use of administrative data and data reconciliation in the basque labour force survey. Technical report, EUSTAT – Basque Statistics Office. `https://ec.europa.eu/eurostat/cros/system/files/use-of-admin-data-basque-lfs.pdf`.

Guarnera, U. and R. Varriale (2015). Estimation from contaminated multi-source data based on latent class models. Technical report, ISTAT. https://ec.europa.eu/eurostat/cros/system/files/Guarnera-etal_abstract_NTTS_2015_Guarnera_Varriale.pdf.

Mushkudiani, N., J. Daalmans, and R. Bikker (2018). Solving large-data consistency problems at statistics netherlands using macro-integration techniques. *Statistica Neerlandica. 72*, 553–573.

Pankowska, P., B. Bakker, D. Oberski, and D. Pavlopoulos (2018). Reconciliation of inconsistent data sources by correction for measurement error: The feasibility of parameter re-use. *Statistical Journal of the IAOS. 34(3)*, 317–329. https://doi.org/10.3233/SJI-170368.

Pannekoek, J. and L.-C. Zhang (2015). Optimal adjustments for inconsistency in imputed data. *Survey Methodology.*.

Pavlopoulos, D. and J. Vermunt (2015). Measuring temporary employment. do survey or register data tell the truth? *Survey Methodology. 41(1)*, 197–214.

Zhang, L.-C. (2012). Topics of statistical theory for register-based statistics and data integration. *Statistica Neerlandica 66(1)*, 41–63.

Appendix

A Solutions to the optimisation problems

A.1 General solution

The objective and the constraints can be written more compactly in matrix notation, a formulation that allows for a concise description of the general optimisation problem, the structure of its solution and the algorithms involved. This formulation is also necessary to apply existing software for solving this optimisation problem.

First we collect all observed proportions, from both the register and the survey in one long vector **p** by the following steps. The observed register proportions in a four-way table, for a given t, can be written in vector form as an 81-vector given by (last index running fastest)

$$\vec{\mathbf{p}}^{R.t} = p_{1,1,1,1}^{R.t,t+1,t+2,t+3}, p_{1,1,1,2}^{R.t,t+1,t+2,t+3}, p_{1,1,1,3}^{R.t,t+1,t+2,t+3}, p_{1,1,2,1}^{R.t,t+1,t+2,t+3}, \dots, p_{3,3,3,3}^{R.t,t+1,t+2,t+3}.$$

The observed register proportions for the four-way tables for all time points together are denoted by $\vec{\mathbf{p}}^{R}$, and obtained as the concatenation of the $\vec{\mathbf{p}}^{R.t}$ for $t \in (1, 4, 7, 10)$: $\vec{\mathbf{p}}^{R} = (\vec{\mathbf{p}}^{R.1}, \vec{\mathbf{p}}^{R.4}, \vec{\mathbf{p}}^{R.7}, \vec{\mathbf{p}}^{R.10})$. Similarly, we can re-arrange the $\mathbf{p}^{S.t,t+3}$ in vector form as a concatenation of the vectors collecting the proportions in the four bivariate (3×3) tables $\vec{\mathbf{p}}^{S} = (\vec{\mathbf{p}}^{S.1}, \vec{\mathbf{p}}^{S.4}, \vec{\mathbf{p}}^{S.7}, \vec{\mathbf{p}}^{S.10})$. Note that the vectors $\vec{\mathbf{p}}^{R}$ and $\vec{\mathbf{p}}^{S}$ and the corresponding four-way and two tables only pertain to month 3 to 13, for the last two month there is only a single source (the register) and therefore no reconcilliation problem.

The complete vector of observed proportions, from both the register and the survey, can now be written as $\mathbf{p} = (\vec{\mathbf{p}}^{R}, \vec{\mathbf{p}}^{S})$ with length $360 = 4 \times 81 + 4 \times 9$.The corresponding vectors with reconciled proportions will be denoted by $\tilde{\pi}$,$\vec{\tilde{\pi}}^{R}$ and $\vec{\tilde{\pi}}^{S}$, respectively. The elements of **p** and $\tilde{\pi}$ will be referred to as p_v and $\tilde{\pi}_v$, for v=1...360 .

All constraints are linear equalities in sums of the components of $\tilde{\pi}$ and can therefore be written in the form $\mathbf{a}^T \tilde{\pi} = \mathbf{b}$. Here **a** is a vector defining the constraint. For the reconciliation models considered here, we distinguish two kinds of constraints,

one that sets the reconciled bivariate proportions $\tilde{\pi}^{t,t+3}$ equal to the corresponding survey proportions but allows the register proportions to be adjusted (**Fixed sums**) and one that allows both the survey and the register proportions to be adjusted in order to satisfy equality constraints (**Equalities**).

(i) **Fixed sums**. These are constraints that set sums of parameters equal to fixed constants (constraints (2) and (3a)). Then the vector **a** consists of 1's corresponding to the elements of $\tilde{\pi}$ in the constraint and 0's for the other elements. The corresponding element of **b** contains the fixed constant.

(ii) **Equality**. These are constraints that define equalities between sums of parameters (constraints (3b)-(3e)). We move all elements from right side of equation to left side. Then the elements of **a** are 1's corresponding to the elements of $\tilde{\pi}$ on the left side of the equality sign, -1's for the elements that were on the right side of the equality sign and 0's for all other elements of $\tilde{\pi}$. The corresponding element of **b** is zero in this case.

To refer to all constraints simultaneously, we define the matrix **A** with rows $\mathbf{a}_k$ defining the K constraints.

The objective function can now be expressed as

$$D(\mathbf{p}, \tilde{\pi}) = D^R(\vec{\mathbf{p}}^R, \vec{\tilde{\pi}}^R) + D^S(\vec{\mathbf{p}}^S, \vec{\tilde{\pi}}^S). \tag{A.1}$$

and the reconciled proportions are the solution to the optimisation problem

$$\tilde{\pi} = \underset{\tilde{\pi}}{\operatorname{argmin}}\, D(\tilde{\pi}, \mathbf{p}) \tag{A.2}$$

$$\text{subject to } \mathbf{A}\tilde{\pi} = \mathbf{b},$$

with the different reconciliation models defined by different specifications of constraints (rows of **A**) and objective function D.

The optimisation problem associated with the reconciliation models described in section 3 is a convex optimisation problem with linear constraints for which several algorithms can be used. For the models (1 and 3) that are based on the (weighted) least squares objective, analytic solutions exists. For models based on the Kullback-Leibler divergence D_{KL} no such closed form solutions exists in general, but for the separable model 2 with fixed constraints an analytic solution does exists. For model 4 an iterative algorithm that can be seen as a generalisation of iterative proportional fitting (IPF) is particularly convenient. These solutions will be described below.

In general, the solution to (A.2) can be obtained by using the Lagrangean function for this problem, which can be expressed as:

$$L(\tilde{\pi}, \lambda) = D(\tilde{\pi}, \mathbf{p}) + \lambda^T(\mathbf{A}\tilde{\pi} - \mathbf{b}), \tag{A.3}$$

with λ the K-vector containing the Lagrange multipliers for the K-constraints. The constraint optimum of (A.2) is a stationary point of the Lagrangean (A.3), which can be found by equating the partial derivatives of (A.3) with respect to $\tilde{\pi}$ and λ to zero (see, e.g. Boyd and Vandenberghe, 2004, Ch. 5, Luenberger, 1984, Ch. 10). Thus, we obtain the following equations:

$$\partial L(\tilde{\pi}, \lambda)/\partial \tilde{\pi} = \partial D(\tilde{\pi}, \mathbf{p})/\partial \tilde{\pi} + \mathbf{A}^T\lambda = \mathbf{0} \tag{A.4a}$$

$$\partial L(\tilde{\pi}, \lambda)/\partial \lambda = \mathbf{A}\tilde{\pi} - \mathbf{b} = \mathbf{0} \tag{A.4b}$$

A.2 Quadratic-loss, additive adjustments

If D is the quadratic-loss $\frac{1}{2}(\tilde{\pi} - \mathbf{p})^T(\tilde{\pi} - \mathbf{p})$, we have $\partial D(\tilde{\pi}, \mathbf{p})/\partial \tilde{\pi} = \tilde{\pi} - \mathbf{p}$ so that (A.4a) becomes $\tilde{\pi} = \mathbf{p} - \mathbf{A}^T\lambda$ which by substitution in (A.4b) results in $\lambda = (\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{p} - \mathbf{b})$ and so the reconciled proportions are given by

$$\tilde{\pi} = \mathbf{p} + \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{b} - \mathbf{A}\mathbf{p}). \tag{A.5}$$

For the fixed-sum constraints we have $9 \times 4 = 36$ constraints defined by (2). These constraints are separable in the sense that they each pertain to a different part of $\tilde{\pi}$, they have no $\tilde{\pi}$-parameters in common. The simultaneous equation (A.5) can then be written as a the set of separate equations

$$\tilde{\pi} = \mathbf{p} + \mathbf{a}_k \frac{1}{n_k}(b_k - \mathbf{a}_k^T \mathbf{p}), \tag{A.6}$$

where the $\mathbf{a}_k$ define the fixed-sum type constraints and $n_k = \mathbf{a}_k^T \mathbf{a}_k$ is the number of register values in constraint k. This equation means that reconciled proportions are obtained by adding the difference of the sum of the observed register values in constraint k and the corresponding survey proportion in b_k, divided by n_k, to each of these register proportions. This results in reconciled proportions that add-up to the corresponding survey proportion. For the equality constraints separability does not hold as can be seen from (3) and the solution for $\tilde{\pi}$ must be obtained by the simultaneous equation (A.5).

A.3 KL-loss, multiplicative adjustments

If D is the KL-loss function, no explicit solution is available for the equality constraints but a convenient iterative algorithm will be presented below. For the case of fixed-sum constraints an explicit solution is available and presented before in (9). The iterative algorithm proceeds in a step-by-step manner as follows. It starts by minimising the objective subject to one of the constraints only. In the next step the resulting approximate solution is updated such that a next constraint is satisfied and the difference with the previous approximation is minimised. When all constraints are visited, the first iteration is completed and the next iteration starts that will again sequentially adjust the current approximation to satisfy each of the constraints. The minimisation carried out in each step solves the problem

$$\tilde{\pi}^{r,k} = \underset{\tilde{\pi}^{r,k}}{\operatorname{argmin}} D(\pi^{r,k}, \tilde{\pi}^{r,k-1}) \tag{A.7}$$
$$\text{subject to } \mathbf{a}_k^T \tilde{\pi}^{r,k} = b_k,$$

with r indicating the iterations. This ‘succesive projection algorithm" is known to converge for convex objectives and linear (in)equality constraints, see e.g. Censor and Zenios (1997).

To solve the minimisation problem (A.7) we set up the Lagrangean function for this problem, which can be expressed as

$$L(\tilde{\pi}^{r,k}, \lambda_k^r) = D(\pi^{r,k}, \tilde{\pi}^{r,k-1}) + \lambda_k^r(\mathbf{a}_k^T \tilde{\pi}^{r,k} - b_k), \tag{A.8}$$

with λ_k^r the Lagrange multiplier for constraint k in iteration r. Equating the partial derivatives of (A.8) with respect to $\tilde{\pi}^{r,k}$ and λ_k^r to zero we obtain the following equations:

$$\partial L(\tilde{\pi}^{r,k}, \lambda_k^r)/\partial \tilde{\pi}^{r,k} = \partial D(\pi^{r,k}, \tilde{\pi}^{r,k-1})/\partial(\tilde{\pi}^{r,k}) + \lambda_k^r \mathbf{a}_k = \mathbf{0} \tag{A.9a}$$
$$\partial L(\tilde{\pi}^{r,k} \lambda_k^r)/\partial \lambda_k^r = \mathbf{a}_k^T \tilde{\pi}^{r,k} - b_k = 0, \tag{A.9b}$$

If D is the KL-divergence, $\sum_v \tilde{\pi}_v^{r,k}(\log(\tilde{\pi}_v^{r,k}/\tilde{\pi}_v^{r,k-1}) - 1)$, we have from (A.9a)

$$\log \tilde{\pi}_v^{r,k} - \log \tilde{\pi}_v^{r,k-1} + \lambda_k^r a_{k,v} = 0$$

and so

$$\tilde{\pi}_v^{r,k} = \tilde{\pi}_v^{r,k-1} / \exp(a_{k,v}\lambda_k^r), \tag{A.10}$$

which by substitution in (A.9b) results in

$$\sum_v a_{k,v}\tilde{\pi}_v^{r,k-1} / \exp(a_{k,v}\lambda_k^r) - b_k = 0. \tag{A.11}$$

Since, as noted in section A.1, the elements of $\mathbf{a}_k$ are 0, 1 or -1, we can re-express (A.11) as

$$\frac{1}{\tau_k^r}\sum_{v\in k+}\tilde{\pi}_v^{r,k-1} - \tau_k^r\sum_{v\in k-}\tilde{\pi}_v^{r,k-1} - b_k = 0, \tag{A.12}$$

with $\tau_k^r = \exp(\lambda_k^r)$ and $k+$ and $k-$ the sets of indices v for which $a_{k,v} = 1$ and $a_{k,v} = -1$, respectively.

For the two types of constraints we are considering here, we now have:

(i) **Fixed sums**. In this case, with the $a_{k,v}$ equal to 1 or 0, (A.12) reduces to $\frac{1}{\tau_k^r}\sum_{v\in k+}\tilde{\pi}_v^{r,k-1} = b_k$, and so

$$\tau_k^r = \frac{\sum_{v\in k+}\tilde{\pi}_v^{r,k-1}}{b_k}$$

and from (A.10) we obtain for the update $\tilde{\pi}_v^{r,k}$

$$\tilde{\pi}_v^{r,k} = \tilde{\pi}_v^{r,k-1} \times \frac{b_k}{\sum_{v\in k+}\tilde{\pi}_v^{r,k-1}} \text{ for } v \in k+. \tag{A.13}$$

For a starting value $\tilde{\pi}_v^{r,k-1} = p_v$ we obtain

$$\tilde{\pi}_v^{k} = p_v \times \frac{b_k}{\sum_{v\in k+} p_v} \text{ for } v \in k+, \tag{A.14}$$

which gives the solution without further iterations because the fixed sum constraints are separable and hence the sets k+ for the different constraints are non-overlapping so that (A.14) only applies to different sets of reconciled proportions for each k. This result coincides with the re-scaling expressed by (9) in section 3.

(ii) **Equalities**. In this case, with $b_k = 0$, we obtain from (A.12)

$$\tau_k^r = \left(\frac{\sum_{v\in k+}\tilde{\pi}_v^{r,k-1}}{\sum_{v\in k-}\tilde{\pi}_v^{r,k-1}}\right)^{\frac{1}{2}}$$

and from (A.10) we obtain for the update $\tilde{\pi}_v^{r,k}$

$$\tilde{\pi}_v^{r,k} = \tilde{\pi}_v^{r,k-1} / \tau_k^r \text{ for } v \in k+, \tag{A.15a}$$

$$\tilde{\pi}_v^{r,k} = \tilde{\pi}_v^{r,k-1} \times \tau_k^r \text{ for } v \in k-. \tag{A.15b}$$

A.4 Algorithm

The algorithm to solve the reconciliation problem with D_{KL} can now proceed according to the following steps:

1. Create the constraint vectors $\mathbf{a}_k$, for k ... K. This also defines the sets $k-$ and $k+$ for each k.
2. For fixed-sum constraints set b_k equal to the fixed survey proportions.
3. Initialise the starting value $\tilde{\pi}_v^{1,0} = p_v$

4. For $k = 1$ to K use the update equations (A.14) or (A.15) to update $\tilde{\pi}^{r,k}$ sequentially for each of the K constraints. This completes one iteration.
5. Repeat step 4 until some convergence criterion is met.

B Model outcomes for pooled data

Table B.1 Adjusted proportions of 'Contract type' in the Polis for combined groups for model 4.

Month	Perm.	Temp.	Oth.
1	0.6267	0.1321	0.2412
2	0.6298	0.1316	0.2386
3	0.6285	0.1308	0.2407
4	0.6294	0.1284	0.2421
5	0.6269	0.1300	0.2431
6	0.6257	0.1293	0.2450
7	0.6245	0.1276	0.2479
8	0.6226	0.1292	0.2482
9	0.6227	0.1281	0.2492
10	0.6226	0.1284	0.2490
11	0.6218	0.1287	0.2495
12	0.6219	0.1255	0.2526
13	0.6225	0.1238	0.2537

Table B.2 Adjusted proportions of 'Contract type' in the LFS for combined groups for model 4.

Month	Perm.	Temp.	Oth.
1	0.6267	0.1321	0.2412
2	-	-	-
3	-	-	-
4	0.6294	0.1284	0.2421
5	-	-	-
6	-	-	-
7	0.6245	0.1276	0.2479
8	-	-	-
9	-	-	-
10	0.6226	0.1284	0.2490
11	-	-	-
12	-	-	-
13	0.6225	0.1238	0.2537

Colophon

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Statistics Netherlands, Grafimedia

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information