Methodology Report

Swiss Earnings Structure Survey 2002–2004

0

Statistische Grundlagen und Übersichten
Bases statistiques et produits généraux
Basi statistiche e presentazioni generali
Basic statistical data and overviews

Compositional data in a stratified two-stage sample: Analysis and precision assessment of wage components

Office fédéral de la statistique
Bundesamt für Statistik
Ufficio federale di statistica
Uffizi federal da statistica
Swiss Federal Statistical Office

OFS BFS UST

Neuchâtel, 2006

Statistik der Schweiz

Die vom Bundesamt für Statistik (BFS) herausgegebene Reihe «Statistik der Schweiz» gliedert sich in folgende Fachbereiche:

0 Statistische Grundlagen und Übersichten
1 Bevölkerung
2 Raum und Umwelt
3 Arbeit und Erwerb
4 Volkswirtschaft
5 Preise
6 Industrie und Dienstleistungen
7 Land- und Forstwirtschaft
8 Energie
9 Bau- und Wohnungswesen
10 Tourismus
11 Verkehr und Nachrichtenwesen
12 Geld, Banken, Versicherungen
13 Soziale Sicherheit
14 Gesundheit
15 Bildung und Wissenschaft
16 Kultur, Informationsgesellschaft, Sport
17 Politik
18 Öffentliche Verwaltung und Finanzen
19 Kriminalität und Strafrecht
20 Wirtschaftliche und soziale Situation der Bevölkerung
21 Nachhaltige Entwicklung und Disparitäten auf regionaler und internationaler Ebene

Swiss Statistics

The «Swiss Statistics» series published by the Swiss Federal Statistical Office (SFSO) covers the following fields:

0 Basic statistical data and overviews
1 Population
2 Territory and environment
3 Employment and income from employment
4 National economy
5 Prices
6 Industry and services
7 Agriculture and forestry
8 Energy
9 Construction and housing
10 Tourism
11 Transport and communications
12 Money, banks, insurance companies
13 Social security
14 Health
15 Education and science
16 Culture, information society, sports
17 Politics
18 Public administration and finance
19 Crime and criminal justice
20 Economic and social situation of the population
21 Sustainable development, regional and international disparities

Statistik der Schweiz

Methodenbericht

Swiss Statistics

Methodology Report

Swiss Earnings Structure Survey 2002–2004

Compositional data in a stratified two-stage sample: Analysis and precision assessment of wage components

Author(s) **Monique Graf**
Swiss Federal Statistical Office

Editor **Swiss Federal Statistical Office**

OFS BFS UST

Office fédéral de la statistique
Bundesamt für Statistik
Ufficio federale di statistica
Uffizi federal da statistica
Swiss Federal Statistical Office

Neuchâtel, 2006

Foreword

This study takes the principles and methods of compositional data analysis, initiated by John Aitchison more than 20 years ago and applies them to wage components. Despite the fact that these principles and methods now enjoy considerable support from theorists, they have not yet been extensively applied to survey data. The aim of the present study is to provide insight into the precision and variability of wage components, as measured in the Swiss Earnings Structure Survey (SESS), and at the same time to propose instruments to assess and convey the precision of multivariate data in the framework of official statistics.

Many people should be acknowledged here: the researchers network around the University of Girona, Spain, for their outstanding ability to transmit their innovative new techniques, in particular Vera Pawlovski-Glahn (Univ. of Girona) and Juan-José Egozcue (Univ. Politécnica de Cataluña) for the helpful discussions. Special thanks are also due to Philippe Eichenberger (METH) for his support throughout the period spent working on this project and to Anne Renaud (METH) who read the manuscript. The team from the section «wages and living conditions» (LOHN) should be warmly thanked for the many discussions, and especially Sara Keel for her input in the comments of several charts. Thanks also to Randall Jones from Languages Services (LING) for reviewing the English in this report.

Abstract

Precision of released figures is not only an important quality feature of official statistics, it is also essential for a good understanding of the data. In this paper we show a case study of how precision could be conveyed if the multivariate nature of data needs to be taken into account.

In the official release of the Swiss Earnings Structure Survey, the total salary is broken down into several wage components. For surveys 2002 and 2004, we first investigate the incidence of components and then follow Aitchison's approach for the analysis of compositional data, which is based on logratios of components.

Different multivariate analyses of the compositional data are performed and compared between the years, whereby the wage components are broken down by economic activity classes. Then we propose a number of ways to assess precision.

Key-words

Methodology Report; Swiss Earnings Structure Survey; LSE; ESSS; Compositional data; complex survey; linearization; confidence domain; precision; coefficient of variation.

Published by:	Swiss Federal Statistical Office (SFSO)
Information:	Monique Graf, Phone +41/ (0)32 713 66 15 Monique.Graf@bfs.admin.ch
Realisation:	Statistical Methods Unit, SFSO
Obtainable from:	Swiss Federal Statistical Office CH-2010 Neuchâtel Phone +41/ (0)32 713 60 60 / Fax +41/ (0)32 713 60 61 Order@bfs.admin.ch
Internet:	http://www.statistik.admin.ch
Order number:	338-0038
Price:	free
Series:	Swiss Statistics
Field:	0 Basic statistical data and overviews
Original text:	English
Graphics/Layout:	SFSO
Copyright:	SFSO, Neuchâtel 2006 Reproduction with mention of source authorized (except for commercial purposes)
ISBN:	3-303-00338-6

Contents

1 Introduction **5**

2 Basic notions on compositional vectors **5**

2.1 Aitchison's geometry in the simplex 7

2.1.1 Classical operations 7

2.1.2 Isometric logratio transformation 7

2.2 Statistics on compositions 9

2.2.1 Centre 9

2.2.2 Dispersion 9

2.2.3 Asymptotic distribution 10

2.2.4 Confidence domains 10

2.2.5 Total variance 10

3 The Swiss Earnings Structure Survey **11**

3.1 Design 11

3.2 Calibration - robustification of weights 11

4 Compositional analysis of wage components **12**

4.1 Variance estimation 13

4.2 Proportion of the social contributions within the gross salary 14

4.2.1 Coefficients of variation 15

4.2.2 Tests between activity classes 15

4.2.3 Tests for change 18

4.3 Other components of the gross salary 18

4.3.1 Incidence of components 18

4.3.2 Multivariate analyzes of the estimated components 20

4.3.3 alr and clr metrics 22

4.3.4 Discriminant analysis 24

4.4 Precision 27

4.4.1 Approximations 27

4.4.2 Simplex view of confidence domains 29

4.4.3 Tests between activity classes 30

4.4.4 Tests for change 30

4.4.5 Global CV 36
4.5 Discussion 38

5 Conclusion **41**

Appendix **42**

References **52**

1 Introduction

There are several aspects of quality of surveys in public statistics (see e.g. Eurostat [9] for the definition advocated by Eurostat). One aspect of quality is accuracy, the first feature of accuracy being precision in relation with sampling variability. Among the many ways used to assess precision, the most recommended indicator is the coefficient of variation (CV) (Eurostat [9]). Being dimensionless, it enables easy comparisons of precision among variables with different orders of magnitude. However in the case of multivariate data which are correlated by nature, like parts of a whole, CV's are not enough to assess precision. We seek a generalization of the CV along the lines of multivariate statistics. This global CV will thus be related to the matrix norm of the covariance matrix of estimates. In the present case study, one aspect of the Swiss Earnings Structure Survey is considered, namely the estimation of population wage components. We use the framework of compositional data analysis as developed by Aitchison [1]. The principle is to compute the logarithm of ratios of the components. The total variance is the sum of the variances of all possible logratios of components. If we divide the total variance by the number of ratios, we obtain an average variance. It will be shown that the linearized form of this average variance can be interpreted as an average squared CV. Graf [14] gives a first analysis of the 2002 data.

Other applications of compositional analysis to public statistics have been published by Silva and Smith [18], Brunsdon and Smith [6], Larrosa [10] and Anyadike-Danes [5].

2 Basic notions on compositional vectors

Basic notions and notations on compositional data are recalled here. Compositional data are observations expressed as parts, thus having a unit sum constraint. A good mathematical summary of the principal notions is provided by Aitchison [3], a less formal introduction in Aitchison [2] and a thorough presentation of the theory in Aitchison [1]. A unit-sum compositional vector of length D, $(p_1, p_2, ...p_D)$ has strictly positive components with sum equal to 1:

$$p_1 + p_2 + ... + p_D = 1 \tag{1}$$

The set of these vectors is the simplex S^D.

Figure 1 shows the 3-part simplex S^3. It is the locus of all points such that the sum of the distances to the 3 edges is 1. These distances are equal to the parts of the corresponding composition. If one part is equal to 1, the composition is located at the vertex opposite to the corresponding edge. Each part of a composition is thus associated to a vertex and its coordinate is given by the distance to the opposite edge.

Equation (1) implies that

$$\operatorname{Var}(p_1 + p_2 + ... + p_D) = 0 \Rightarrow \sum_{i \neq j} \operatorname{Cov}(p_i, p_j) = -\operatorname{Var}(p_j) \quad \forall j$$

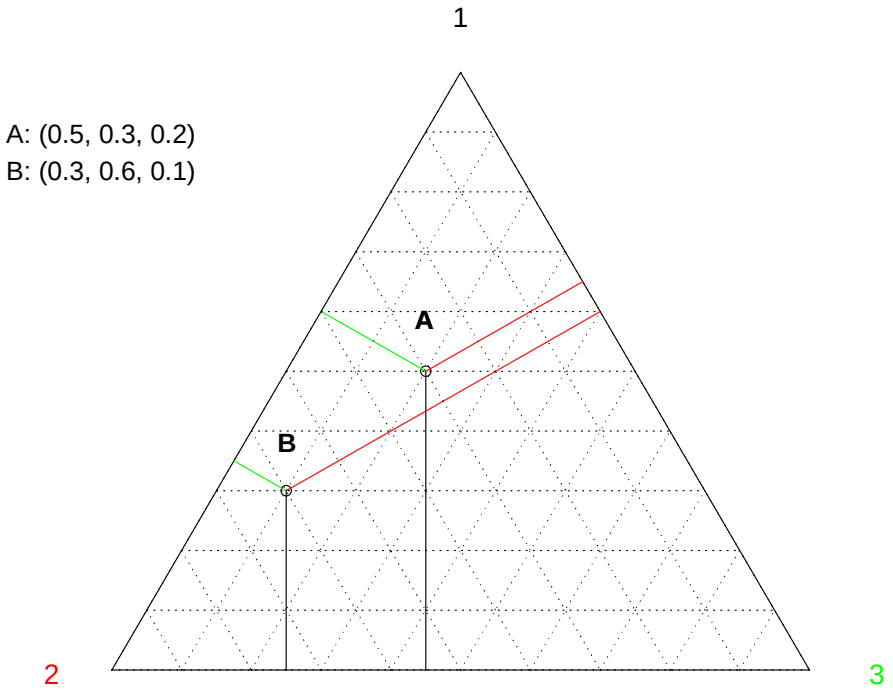


Figure 1: Representation in $D = 3$ dimensions. Dotted lines are 0.1 apart.

so there is necessarily a negative correlation between the components. This shows that the correlations are not directly interpretable. To release this constraint, Aitchison proposes 2 alternative transformations:

- *the additive logratio transform* (alr)
  Consider the vector of ratios of the $d = D - 1$ first components to the last, that is
  $$x = (x_1, ..., x_d) = (p_1, ..., p_d) / p_D = p_{-D}/p_D \tag{2}$$
  The alr is defined by the logarithm
  $$\text{alr}(p) = y = \ln x = (\ln(p_1/p_D), \ln(p_2/p_D), ..., \ln(p_d/p_D))$$
  Applying this transformation, the resulting vector is no longer constrained and correlations between y_i and y_j can be interpreted.

- *the centred logratio transform* (clr)
  It is often desirable that all components are treated symmetrically. In the clr transform, each component is divided by the geometric mean
  $$g(p) = \left(\prod_1^D p_i \right)^{1/D}$$
  $$\text{clr}(p) = (\ln(p_1/g(p)), \ln(p_2/g(p)), ..., \ln(p_D/g(p))) \tag{3}$$

2.1 Aitchison's geometry in the simplex

(This paragraph can be skipped by the non-mathematically inclined reader. It will be referred to only in the discussion of § 4.3.3.)

The *closure* of a vector w with positive coordinates is obtained by dividing each coordinate by their sum:

$$p = \mathrm{clo}(w) = w/\sum w_i$$

The vector space structure in the alr or clr transforms induce a vector space on the simplex. There are two basic operations in a vector space: the addition of two vectors and the multiplication of a vector by a scalar. In the case of compositions, these classical operations have different names.

2.1.1 Classical operations

- *perturbation*:

  $$a \oplus p = \mathrm{clo}(a_1 p_1, a_2 p_2, ..., a_D p_D)$$

  Perturbation corresponds to the vector addition in the alr and the clr representations.

- *powering*:

  $$b \odot p = \mathrm{clo}(p_1^b, p_2^b, ...p_D^b).$$

  Powering corresponds to the scalar multiplication in both representations.

2.1.2 Isometric logratio transformation

- *ilr coordinates* (Egozcue et al [7])

  The space generated by $\mathrm{clr}(p)$ is a vector subspace of $\mathbb{R}^D$ of dimension d. An orthonormal basis is given by the Helmert matrix:

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & ... & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & ... & 0 \\ ... & ... & ... & ... & ... \\ \frac{1}{\sqrt{D(D-1)}} & \frac{1}{\sqrt{D(D-1)}} & \frac{1}{\sqrt{D(D-1)}} & ... & -\frac{D-1}{\sqrt{D(D-1)}} \end{pmatrix} = \begin{pmatrix} e'_1 \\ e'_2 \\ ... \\ e'_d \end{pmatrix}$$

  The image of this basis in the simplex S^D is given by

  $$\mathrm{clr}^{-1}(e_i) = \mathrm{clo}(\exp(e_i)), \quad i = 1..., d$$

  . The $\mathrm{clr}^{-1}(e_i)$ thus form a natural basis for the simplex.

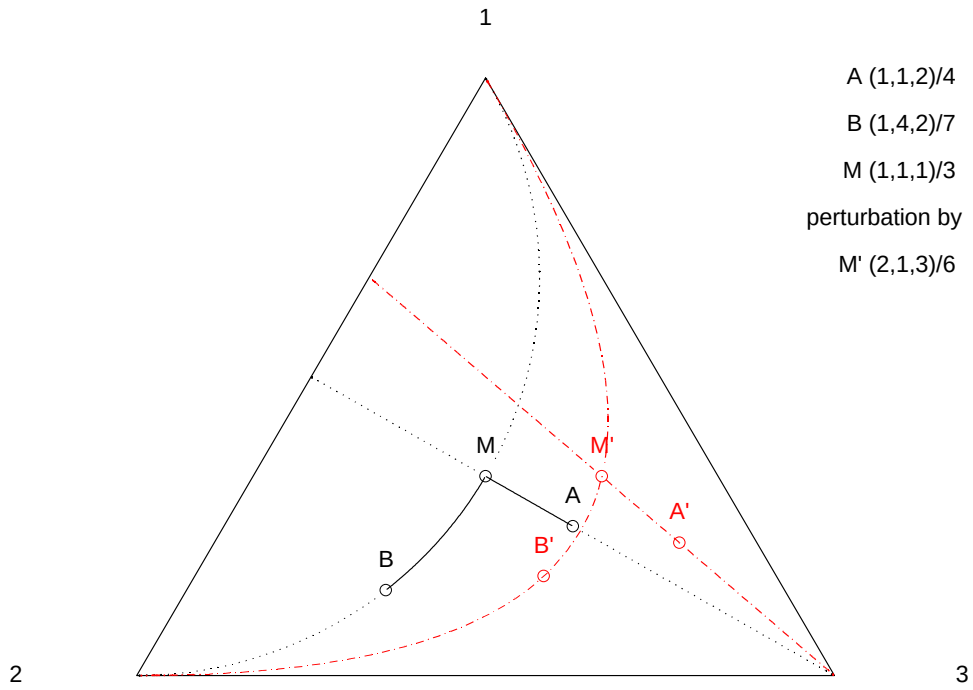


Figure 2: ilr coordinates and perturbation.

Figure 2 shows the 2 ilr axes in 3 dimensions (dotted lines). They intersect in M, which corresponds to the constant composition $(1/3, 1/3, 1/3)$ and is the origin of the vector space. The first axis goes through vertex 3 and is the locus of all compositions such that parts 1 and 2 are equal. The second axis is a curve from vertex 1 to 2 through M. It is the locus of compositions such that part 3 is equal to the geometric mean of parts 1 and 2. If we perturb M, A and B by M', we obtain M', A' and B'. The images of the ilr coordinate axes (dashed lines) are also plotted. It is clear from this picture that perturbation is the concept of translation adapted to the simplex. Other graphical effects of perturbation can be found in von Eymatten et al [19].

2.2 Statistics on compositions

2.2.1 Centre

The centre of the distribution for x is given by the geometric - and not the arithmetic - mean of the compositions. Its theoretical counterpart (expressed with the help of the expectation of the alr coordinates) is:

$$\xi = \exp\left(\mathrm{E}\left(\ln x\right)\right) = \exp(\mathrm{E}(\mathrm{alr}(p))) \tag{4}$$

It can be transformed back to give the centre for p:

$$\begin{aligned} \mathrm{cen}(p_i) &= \frac{\xi_i}{1 + \sum_{j=1}^{d} \xi_j} \quad i = 1, ..., d \\ \mathrm{cen}(p_D) &= \frac{1}{1 + \sum_{j=1}^{d} \xi_j} \end{aligned}$$

That is:

$$\mathrm{cen}(p) = \mathrm{alr}^{-1}\left(\mathrm{E}\left[\mathrm{alr}(p)\right]\right)$$

An equivalent expression can be obtained with the help of clr coordinates:

$$\mathrm{cen}(p) = \mathrm{clr}^{-1}\left(\mathrm{E}\left[\mathrm{clr}(p)\right]\right)$$

2.2.2 Dispersion

There are different equivalent and linearly related dispersion matrices, Aitchison [1, § 4.8].

1. The $d \times d$ - covariance matrix of the logratios:

$$\mathbf{\Sigma} = [\sigma_{ij}] = \mathrm{Cov}\left(\ln x_i, \ln x_j\right) = \mathrm{Cov}\left(\ln \frac{p_i}{p_D}, \ln \frac{p_j}{p_D}\right) \tag{5}$$

   The only drawback in this setting is that the last component p_D is treated differently.

2. All components of p are handled symmetrically, if they are divided by the geometric average $g(p)$. The $D \times D$ - centred covariance matrix is given by:

$$\mathbf{\Gamma} = [\gamma_{ij}] = \mathrm{Cov}\left(\ln \frac{p_i}{g(p)}, \ln \frac{p_j}{g(p)}\right) \tag{6}$$

   which is singular, because $\sum \ln \frac{p_i}{g(p)} = 0$.

3. The last possibility is to use the $D \times D$ -variation matrix, with all elements being variances:

$$\mathbf{T} = [\tau_{ij}] = \mathrm{Var}\left(\ln \frac{p_i}{p_j}\right) \tag{7}$$

   This matrix has a zero principal diagonal and only one positive eigenvalue, corresponding to eigenvector 1_D.

2.2.3 Asymptotic distribution

Under regularity conditions, $y = (\ln x_1, \ln x_2, ..., \ln x_d)'$ is asymptotically normally distributed $N^d(\mu, \mathbf{\Sigma})$ with $\mu = \ln \xi$ and $\mathbf{\Sigma}$ given by Equation (5). The derived distribution for $p = (p_1, p_2, ...p_D)$ on the simplex S^D is called the additive logistic normal distribution and denoted by $L^d(\mu, \mathbf{\Sigma})$.

2.2.4 Confidence domains

Under the asymptotic distribution hypothesis,

1. for y the confidence domain $D_{1-\alpha}(y)$ is limited by a d dimensional ellipsoid. Let $\chi^2_{d;1-\alpha}$ be the $(1-\alpha)$ quantile of the chi-square distribution with d degrees of freedom. Then

$$D_{1-\alpha}(y) = \left\{ y \in R^d \mid (y-\mu)' \mathbf{\Sigma}^{-1} (y-\mu) \leq \chi^2_{d;1-\alpha} \right\}$$

2. for x in Equation (2), the equivalent domain is

$$D_{1-\alpha}(x) = \left\{ x \in R^d_+ \mid (\ln x - \mu)' \mathbf{\Sigma}^{-1} (\ln x - \mu) \leq \chi^2_{d;1-\alpha} \right\} \tag{8}$$

3. for $p = (p_1, ..., p_D)$, the domain is a subset of the simplex S^D:

$$D_{1-\alpha}(p) = \left\{ p \in S^D \mid \left(\ln \frac{p_{-D}}{p_D} - \mu \right)' \mathbf{\Sigma}^{-1} \left(\ln \frac{p_{-D}}{p_D} - \mu \right) \leq \chi^2_{d;1-\alpha} \right\} \tag{9}$$

2.2.5 Total variance

Whereas a thorough study of the precision of a composition implies computing a d-dimensional confidence domain, but we also need a simple global characterization of precision. This is generally given by a matrix norm of the covariance matrix. Aitchison defines (among other measures) the total variance for which different equivalent formulations exist (Aitchison [1, Chapter 4]):

$$\begin{aligned} \text{totvar}(p) &= \text{tr}(\mathbf{\Gamma}) = \sum_{i=1}^{D} \text{Var}\left(\ln \frac{p_i}{g(p)} \right) && (10) \\ &= \frac{1}{D} \sum_{i<j} \text{Var}\left(\ln \frac{p_i}{p_j} \right) = \frac{1}{2D} \sum_{i,j=1}^{D} \tau_{ij} && (11) \end{aligned}$$

or

$$\begin{aligned} \text{totvar}(p) &= \frac{1}{2D} \sum_{i,j=1}^{D} \tau_{ij} = \frac{1}{2D} 2 \left(D \,\text{tr}(\mathbf{\Sigma}) - 1_d' \mathbf{\Sigma} 1_d \right) \\ &= \text{tr}(\mathbf{\Sigma}) - \frac{1}{D} 1_d' \mathbf{\Sigma} 1_d && (12) \end{aligned}$$

3 The Swiss Earnings Structure Survey

The Swiss earnings structure survey (SESS) is a biennial written survey sent out to businesses. The survey is constructed on a stratified two-stage sampling scheme (Graf [13]). The 2002 and 2004 samples are rather large: 1/3 of all businesses in Switzerland are involved. The extrapolation weights and the finite population correction take non response into account (which we suppose is ignorable within the stratum). The variance estimation method applied here is the classical linearization of the estimators, see e.g. Särndal et al [15]. Other aspects of precision computed for the 2000 survey were studied in (Graf [11], [12]), see also Eurostat [8]. A general report on the 2002 survey can be found in (SESS [17], [16]); a description of the sampling design and extrapolation in Graf [13]).

3.1 Design

The sampling frame is the business register (BR) in its latest state at the time of sampling. The stratification was originally designed as a combination of 41 activity classes, 3 business size classes and 13 regional subdivisions. In the SESS, the activity classes are the NOGA at 2 digits level with some grouping in order to avoid the appearance of very small strata (see Table A1)[1]. Class 0 represents the total of all activities considered.

The survey design is a stratified two-stage sampling, with a simple random (SI) sample of businesses in each stratum and a SI sample of salaries within each sampled business. The stratification is a combination of geography, activity and size classes. While the last 2 categorizations are exactly the same in 2002 and 2004, some changes appear in the geographical subdivision, due to differing special demands of the cantons, see Table 1.

The sampling fraction at both stages depends on the size class. The largest businesses form exhaustive strata. For medium and small businesses, a Neyman allocation based on the variance of the mean standardized gross earnings of the preceding survey is computed. The strata sizes are then modified so that a minimum of 10 units (businesses) are sampled (if stratum size permits). Large businesses have to furnish 33% of the earnings paid out in October. Medium size businesses furnish 50% while the smallest businesses give them all. The desired sampling fraction and the expected non-response rate are used to determine the number of businesses to contact.

3.2 Calibration - robustification of weights

The non response is assumed to be ignorable at the stratum level and the Horvitz-Thompson weights at both stages are in principle used (the actual number of salaries paid by the business in October is asked in the questionnaire). Few expansion weights are large due to non-response. To robustify the procedure, these weights were trimmed, first at the cluster level, and then at the stratum level. The resulting

[1]The NOGA is the Swiss version of the Statistical Classification of Economic Activities in the European Community, Revision 1 (NACE Rev. 1). Both classifications are till the 4th level identical.

Table 1: Geographical subdivisions GRS for surveys 2002 and 2004

NUTS II	GR	GRS 2002	Cantons (NUTS III)	GRS 2004
Lake Geneva region	1	11	GE	11
		12	VD	12
		13	VS	19
Central plain	2	22	BE	21
		29	FR	22
		28	NE	23
		22	SO+JU	29
Northwestern Switzerland	3	30	AG+BL+BS	30
Zurich	4	40	ZH	40
Eastern Switzerland	5	51	SG	51
		52	SH+TG+AI+AR+GL+GR	59
Central Switzerland	6	61	ZG	60
		62	UR+SZ+NW+OW+LU	60
Ticino	7	71	TI	70

weights are recalibrated using the CALMAR raking procedure[2], in such a way that the marginal total weights on the 3 stratum classifications remain constant. Thus the "unreliable" estimates are weighted down without changing the total. If the whole population is considered, this procedure changes the results very little. The sampling plan was designed for the main variable, namely the monthly standardized gross earnings. In this study, we are interested in the compositional analysis of the weighted total of monthly non standardized total salary.

4 Compositional analysis of wage components

In the SESS, 5 wage components are published (social security contributions, overtime earnings, hardship allowances, 13th or n-th salary, bonuses), see SESS [16], [17]. They are reproduced here for 2002 and 2004 (Table A1, Appendix). In Table A1, wage mass is defined for an economic branch as the extrapolated sum of all sampled salaries, using the above calibrated weight. The "non standardized total monthly salary", MBLIU, is the sum of the 5 components and the rest (the "naked" salary), which forms a 6th component and is never published as such. The non-standardized gross monthly salary BLIMOK, includes the "naked" salary and social security contributions, but *excludes* components 2 to 5. The defined components are summarized in Table 2. The wage percentage attributed to each component in Table A1 are computed relative to BLIMOK, and not to MBLIU. Thus the published proportions are:

$$(s_1, s_2, s_3, s_4, s_5) / (s_1 + s_6) \tag{13}$$

We see that Table 2 contains two different subcompositions, the first is a 2-

[2]SAS macro written at the French national statistical office INSEE.

Table 2: Wage mass attributed to the different components

Definition	Code	Total amount
Social security contributions	SOZABG	s_1
Overtime earnings	VERDUZ	s_2
Hardship allowances	ZULAGEN	s_3
13th salary (/12)	XIII12E$^{(a)}$	s_4
Special payements, bonuses (/12)	SOND12E$^{(b)}$	s_5
"Naked" salary	-	s_6
Non-standardized gross earnings with social contributions	BLIMOK	$s_1 + s_6$
Monthly non standardized total salary	MBLIU	$\sum_{i=1}^{6} s_i$

$^{(a)}$XIII12E=ROUND(XIIILOHN/12) $^{(b)}$SOND12E=ROUND(SONDERZA/12)

dimensional composition expressed as a part $s_1/(s_1 + s_6)$, and the second is 5-dimensional, expressed as ratios of components $(s_2, s_3, s_4, s_5)/(s_1 + s_6)$.

These two subcompositions will be analyzed separately. We stress that in this framework, the interest is not in the wage composition at the individual level, but in the global composition for segments of the population. The advantage from a mathematical point of view is that no zero components are observed, while they exist at the individual level.

4.1 Variance estimation

The variance-covariance matrix of the estimates is based on the sampling distribution of the wage components. Computing the variance of these global compositions in a stratified double stage cluster sample is a complex task, because no closed formula for the variance is available. The large sample size implies that finite population corrections (fpc) are indispensable for realistic estimates of the precision of the population values. The extrapolation weights and the finite population correction take the non response into account (which we suppose ignorable within the stratum). The variance estimation method applied here relies on the linearization of the estimators and is equivalent to the recovery of the compositional variation array from the crude mean vector and covariance matrix, see Aitchison [1, §4.4]. (An alternative would be to use resampling methods, but it would be extremely cumbersome in this large survey). In fact we simply use the first order approximation, which is a slight overestimation:

$$\mathrm{Var}\left(\ln \widehat{X}\right) \cong \mathrm{E}\left(\ln \widehat{X} - \ln X\right)^2 \cong \mathrm{E}\left(\frac{\widehat{X} - X}{X}\right)^2 = \mathrm{CV}^2\left(\widehat{X}\right) \tag{14}$$

where $\widehat{X}$ is an unbiased estimator for X and CV is the coefficient of variation.

For a ratio:

$$\operatorname{Var}\left(\ln\frac{\widehat{X}}{\widehat{Y}}\right) \cong \mathrm{E}\left(\frac{\widehat{X}-X}{X}-\frac{\widehat{Y}-Y}{Y}\right)^2 \tag{15}$$

$$= \left(\frac{X^2}{Y^2}\right)^{-1}\left\{\frac{1}{Y^2}\mathrm{E}\left(\widehat{X}-\frac{X}{Y}\widehat{Y}\right)^2\right\} \cong \mathrm{CV}^2\left(\frac{\widehat{X}}{\widehat{Y}}\right) \tag{16}$$

We recognize in the left expression in brackets in Equation (16) the formula for the linearization of the variance of a ratio. Practically a program for computing the linearized variance of a ratio will do the job.

In matrix form:

$$\operatorname{Var}\left(\ln\frac{\widehat{X}}{\widehat{Y}}\right) \cong \begin{pmatrix} \frac{1}{X} & -\frac{1}{Y} \end{pmatrix} \Sigma_{\widehat{X},\widehat{Y}} \begin{pmatrix} \frac{1}{X} \\ -\frac{1}{Y} \end{pmatrix} \tag{17}$$

where $\Sigma_{\widehat{X},\widehat{Y}}$ is the covariance matrix of $\widehat{X}$ et $\widehat{Y}$. The covariance is found by a variance computation using

$$\operatorname{Cov}\left(\widehat{X},\widehat{Y}\right) = \frac{1}{2}\left[\operatorname{Var}\left(\widehat{X}+\widehat{Y}\right) - \operatorname{Var}\left(\widehat{X}\right) - \operatorname{Var}\left(\widehat{Y}\right)\right] \tag{18}$$

Once the covariance matrices of the logratio of the wage components to the gross salary are obtained, we are in position to 1. assess the accuracy of the population composition estimates, and 2. test hypotheses on differences in composition between subpopulations. Graphical representations and interpretations of the results will be presented.

4.2 Proportion of the social contributions within the gross salary

With s_1 and s_6 as defined in Table 2, let

$$q = s_1/(s_1+s_6) \tag{19}$$

Our variable of interest is q, which represents the proportion of the social security contributions within the non-standardized gross monthly earnings BLIMOK. $1-q$ is the "naked salary" part.

In this case the compositional vector is of length $D = 2$ and is denoted by $\widetilde{q} = (q, 1-q)$. $\widetilde{q}$ can be replaced by the equivalent form of length $d = 1$

$$x' = q/(1-q)$$

From $\mathrm{CV}(q)$ the following 95% confidence intervals for q are obtained:

1. *Normal approximation CI for q*:

$$[bn_{l95}, bn_{u95}] = q\,(1 \pm 1.96\,\mathrm{CV}(q)) \tag{20}$$

The drawback of approximation (20) is that the inferior bound can be negative.

2. *Log-normal approximation CI for* $\ln\left(\frac{q}{1-q}\right)$, using Equation (17):

$$[bl_{l95}, bl_{u95}] = \ln\left(\frac{q}{1-q}\right) \pm 1.96\frac{\mathrm{CV}(q)}{(1-q)}$$

3. *CI for* q *deduced from 2 (logistic normal approximation)*

$$[b_{l95}, b_{u95}] = \left[\frac{\exp\left(bl_{l95}\right)}{\exp\left(bl_{l95}\right)+1}, \frac{\exp\left(bl_{u95}\right)}{\exp\left(bl_{u95}\right)+1}\right] \tag{21}$$

The approximation (21) is reliable if $\mathrm{CV}(q)$ is less than about 40%. Otherwise the interval will be too wide.

4.2.1 Coefficients of variation

The social security contributions by economic activity classes and their variation coefficients for the years 2002 and 2004 are given in Table A1. The intervals given by Equations (20) and (21) (not shown here) are very similar: the maximum difference is 0.01% (NOGA2=61, water transport, which is marginal in Switzerland). We prefer Equation (21) which has the advantage to guarantee that the bounds are in $[0, 1]$.

4.2.2 Tests between activity classes

Observations in different activity classes being independent, we can easily test 2 by 2 differences in social security contributions. If q_1 and q_2 are the proportions of the social security contributions in classes 1 and 2, then equality of proportions is rejected with risk α, if

$$\frac{\left|\ln\left(\frac{q_1}{1-q_1}\right) - \ln\left(\frac{q_2}{1-q_2}\right)\right|}{\sqrt{\left(\frac{\mathrm{CV}(q_1)}{(1-q_1)}\right)^2 + \left(\frac{\mathrm{CV}(q_2)}{(1-q_2)}\right)^2}} > z_{1-\alpha/2}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution.

Table A2 (Appendix) shows the p-values of all two by two tests, separately for production and services and per year. Classes are ordered within the sector by increasing social security contribution (SOZABG). Results are plotted in Figure 3. Non significant differences in proportions (with $\alpha = 5\%$) are connected by a segment. The horizontal axis has no meaning: a small random quantity is generated for the abscissa so that the connecting segments become distinguishable. It can be seen that in the secondary sector, a larger proportion of the non-standardized gross monthly salary BLIMOK is generally devoted to social security contributions than in the third. Activities 45 (construction), 10-14 (mining and quarrying of stone) and 60 (land transport/pipelines) in 2002 and 62 (air transport) in 2004 have the highest contributions. The contribution for activities 20 and 27 (manufacture of wood and

basic metals, respectively) is among the highest in 2004. At the other extreme, 40 (electricity, gas and water supply) devotes the smallest part of the overall wage bill to social security in 2002. 61 (water transport) is the least precise in 2002, being connected to remote classes on the vertical scale; in 2004 the part devoted to social security contribution in class 61 has increased by 1%.

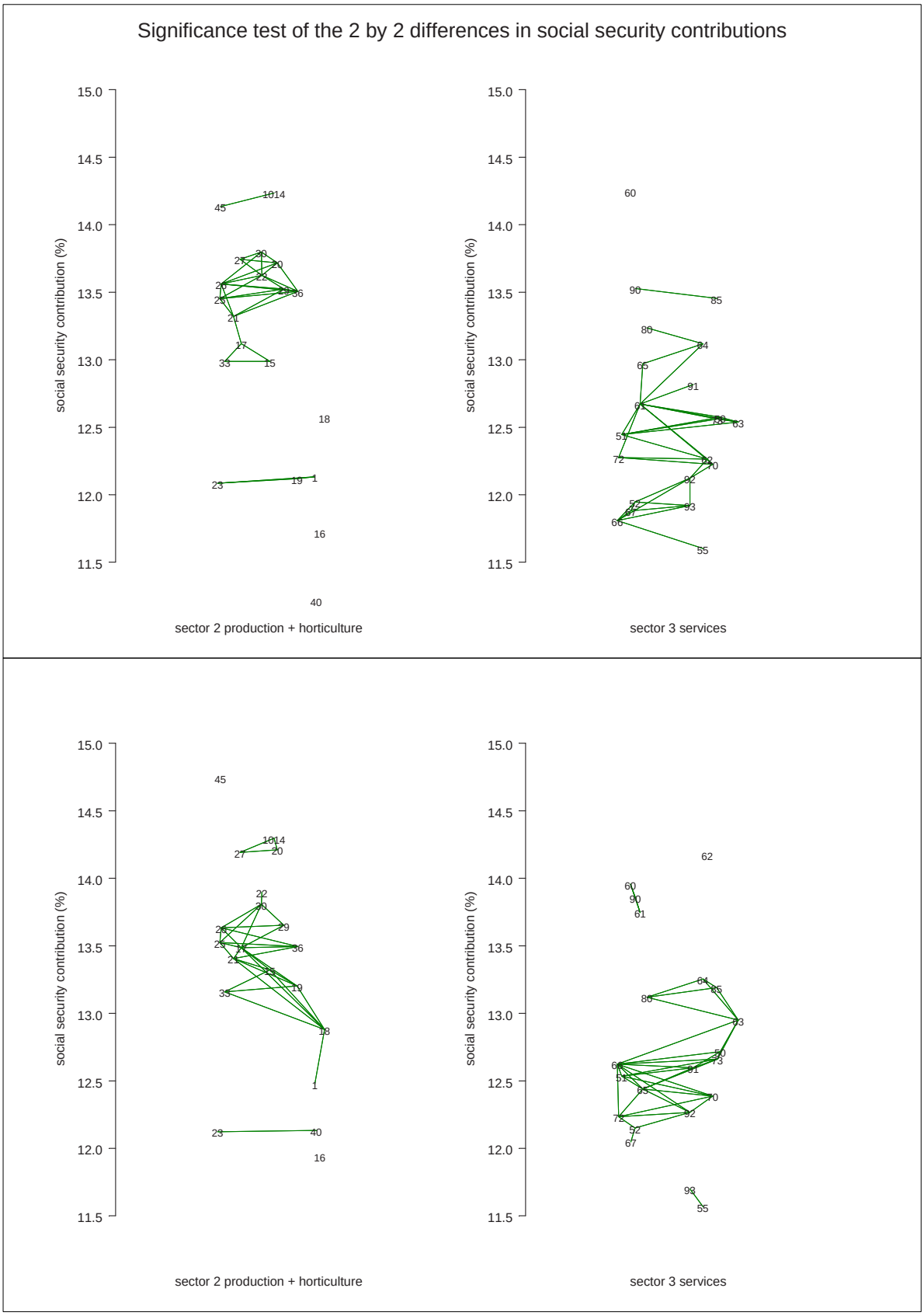


Figure 3: Non significant differences in proportions at the 5 % risk are represented by a connecting segment (2002 (top), 2004 (bottom).

4.2.3 Tests for change

Tests for change were performed (Table A3). Because the 2002 and 2004 samples are drawn independently, the variance of the difference in SOZABG is the sum of the 2002 and 2004 variances. It is interesting to note that all the significant differences in the production sector indicate a larger contribution in 2004 compared to 2002. For the services, a majority of the significant differences are positive too.

4.3 Other components of the gross salary

From the six-parts compositional vector $(p_1, p_2, p_3, p_4, p_5, p_6)$, let us form a new composition by amalgamation of components 1 and 6:

$$p = (p_2, p_3, p_4, p_5, p_1 + p_6) = \frac{(s_2, s_3, s_4, s_5, s_1 + s_6)}{\sum_{i=1}^{6} s_i} \tag{22}$$

This vector can be written in the equivalent form

$$x = (x_1, x_2, x_3, x_4) = \left(\frac{s_2}{s_1 + s_6}, \frac{s_3}{s_1 + s_6}, \frac{s_4}{s_1 + s_6}, \frac{s_5}{s_1 + s_6} \right) \tag{23}$$

Interpreting $s_i, i = 1, ..., 6$ as in Table 2, we see that p represents a decomposition of the non standardized total monthly salary MBLIU into 5 components, and that x equals the ratios of the 4 first components VERDUZ ... SOND12E to the fifth (the non-standardized gross monthly salary BLIMOK). For each economic activity grouping, the main columns 2 to 5 of Table A1 are the components x_i, expressed in %.

4.3.1 Incidence of components

The only component that is always present is the gross earnings BLIMOK. The wage percentages in Table A1 have a different meaning, whether the corresponding mass is distributed over a large or a small number of ultimate units. In order to investigate this point, the average incidence of the components 2-5 was computed. Table 3 shows the weighted incidence of each component. It can be interpreted as a probability that a randomly chosen unit is getting a non-zero amount for that component. So an estimated proportion of 75-76% of wages encompasses a 13th salary. The probability of obtaining the components VERDUZ and ZULAGEN defined in Table 2 is doubled, if the 13th salary is present. On the contrary, the probability of having the SOND12E share given the presence of XIII12E is slightly smaller.
There are 16 possible presence/absence combinations of the 4 components in Table 3. Everyone is observed, but with very different frequencies. The more frequent are "0010" (only XIII12E) and "0000" (nothing at all). Let us combine all the other configurations into one, named "Others". For each economic activity class, the wages can be partitioned into these 3 configurations. In Figure 4, the proportion of wages in each configuration are represented with the help of the barycentric coordinates.It is interesting to note that the triangle can be divided into 2 equal parts, and that except for class 93, the proportion of "Just XIII12E" is always larger than the proportion of "Nothing". Class 93 is the nearest to the edge "Nothing", meaning that the proportion of wages without any supplementary component is the largest. In Table 4

Table 3: Weighted proportion of wages with the considered component: global (first 2 lines) and given presence/absence of XIII12E (4 last lines)

year	XIII12E	VERDUZ	ZULAGEN	SOND12E
2002	0.753	0.031	0.106	0.258
2004	0.760	0.041	0.113	0.255
	given XIII12E			
2002	0	0.018	0.063	0.295
2002	1	0.035	0.120	0.246
2004	0	0.016	0.051	0.273
2004	1	0.049	0.132	0.249

the 13 existing configurations for class 93 are shown, together with N, the number of sampled units and TOTAL, the estimated number in the population. Small numbers are replaced by a * for disclosure control reasons; the sum of the masked values of TOTAL is 783. The 11 first configurations in Table 4 form the category "Others" for that class.
For class 65 (banking), the category "Others" form 67% and is dominated by the configuration "just SOND12E" (44%), whereas for class 16 (manufacture of tobacco products) the category "Others" (77%) consists mainly of 0011 and 0110 (44%).

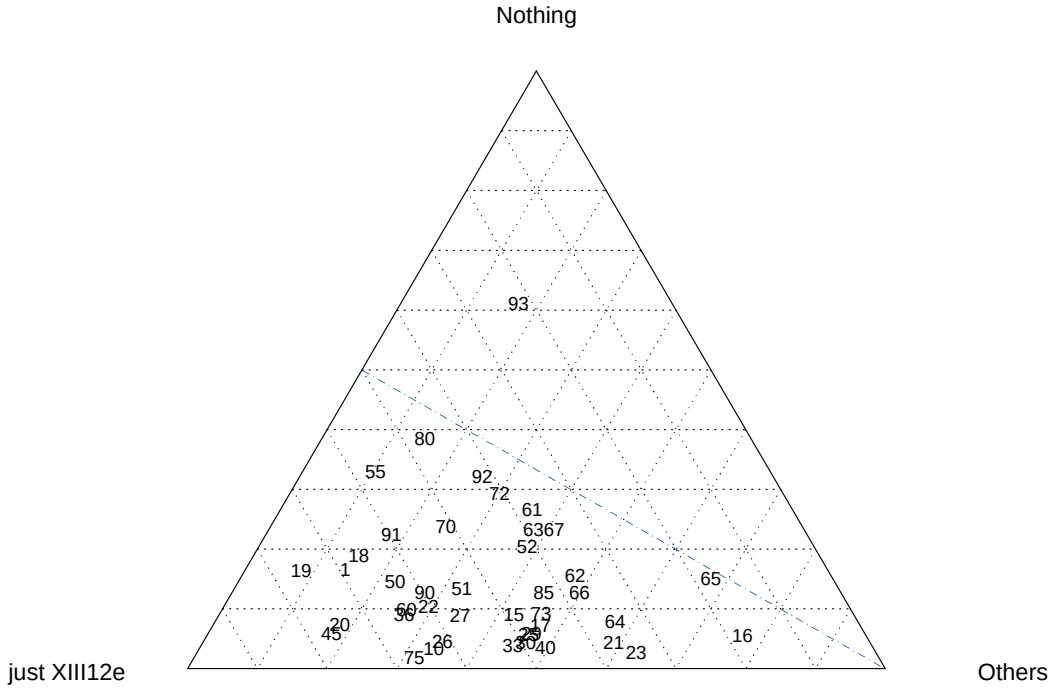


Figure 4: Incidence of component combinations for the 2002 data, ordered by N.

Table 4: Configurations for the activity class 93 (other service activities).

NOGA2	VERDUZ	ZULAGEN	XIII12E	SOND12E	N	TOTAL
93	1	1	1	1	*	*
93	0	1	0	1	*	*
93	0	1	0	0	*	*
93	0	1	1	1	*	*
93	1	0	1	1	*	*
93	1	0	0	0	*	*
93	1	0	0	1	*	*
93	0	1	1	0	*	*
93	1	0	1	0	*	*
93	0	0	1	1	209	770.15
93	0	0	0	1	577	2618.96
93	0	0	1	0	1000	5458.21
93	0	0	0	0	2957	15370.37

4.3.2 Multivariate analyzes of the estimated components

Before we compute the precision, it is important to get a rough idea of the data. A multidimensional scaling on logratio estimates (alr scale, see §1) was performed with the S-plus procedure CMDSCALE on the 2002 data, using as distance between 2 economic activities the Euclidian distance between the corresponding logratio vectors. (The same result would be obtained by principal component analysis). This metrics choice was based on the form of the published Table A1, which gives the ratios of the components to the gross salary BLIMOK.
The 2 panes in Figure 5 represent the same projection onto the first two principal axes[3]. This plane explains 90% of the total variability. Thus the distance between 2 points in the plane can be interpreted as a measure of discrepancy between the corresponding compositional vectors. In the left pane, activity classes are coded by their NOGA2 code, whereas in the right pane, they are represented by a star plot (the half diagonals of the quadrilateral are proportional to the $x_i, i = 1...4$). A partition was also performed (using the 4 new coordinates) by the S-plus procedure PAM (partition around medoids) and an optimal number of 3 groups was obtained. The groups are visible on the left pane (circles, triangles and diamonds).

It is typical for the first group (diamonds) to report a large share of "special payments/bonuses" and a relatively small portion of "13th month salary". It is also typical for the share of "overtime pay" and "hardship allowances" to be practically nonexistent. All eight branches in this group fall in the tertiary sector.

The second group (triangles) counterbalances the first group to a certain extent. Apart from "13th month salary", the branches in this group report practically no wage components. This indicates that in these branches pay is limited to base monthly salary. The nine branches in this group are arranged according to economic sector and number of employees. That said, the vast majority of the branches in

[3]The usual terminology is "principal components principal"; the expression "principal axes" is being used instead, in order to avoid confusion with the salary components.

this group fall in the tertiary sector.

It is typical for the third group (circles) to report relatively small shares of wage components, with the exception of "13th month salary", especially when it comes to "overtime pay". In addition, there are three times more branches of trade in this group than in the first two groups. Most (i.e. sixteen of the twenty-four branches) fall in the secondary sector.

All things considered, it can be said that the shares of the four wage components vary considerably. The share of "13th month salary" is about the same in all branches; the share of "hardship allowances" and "overtime pay" are small to very small, which makes it difficult to assess them in the chart; in contrast, the share of "special payments/bonuses" varies considerably from branch to branch.

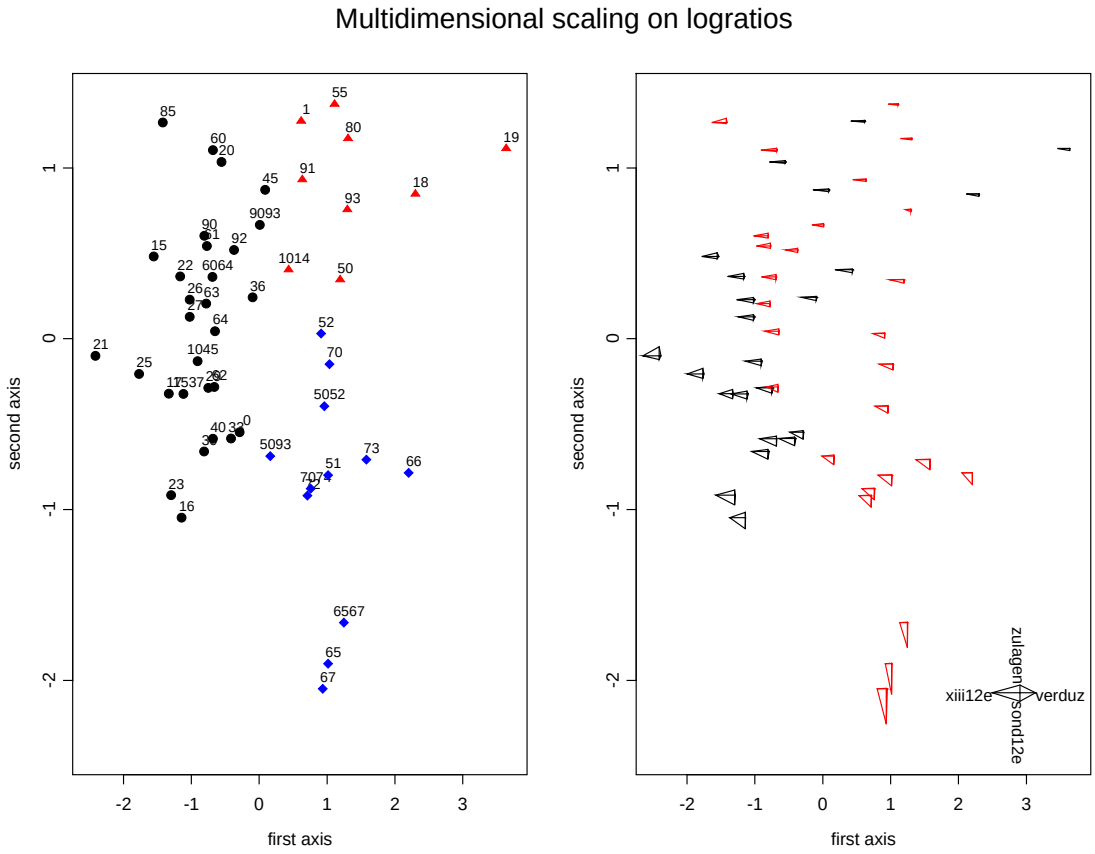


Figure 5: Multidimensional scaling on estimated logratios corresponding to the 5-part composition for all activity classes (2002 data). Circles, triangles and diamonds on the left pane give the group membership computed by PAM; numbers are the NOGA classes. The right pane shows star plots.

4.3.3 alr and clr metrics

(This paragraph can be skipped by the non-mathematically inclined reader.)

Aichison [1] showed that there are three equivalent ways of considering ratios within a compositional vector with linearly related dispersion matrices. They correspond to

1. the alr transformation with the covariance matrix in Equation (5),
2. the clr transformation with covariance matrix in Equation (6),
3. the logratios of all pairs of components, with variation matrix in Equation (7).

Aitchison and Greenacre [4] studied the relationship between the singular value decomposition of the compositional data matrix, when transforms 2. and 3. above are applied, but discarded the alr transform 1. as being least interesting. The reason is that the representation of a composition as a point within the simplex gives all components a symmetrical role and is thus not compatible with the alr transforms for which the last component plays a special role.

The use of the alr representation for the 2002 multidimensional scaling is justified by the form of the published table (Table A1: relative values of the other components to the gross salary). The gross salary BLIMOK indeed plays a special role, because it is the only component that is always present (see preceding paragraph). On the other hand, if we want to make full use of the compositional analysis theory, we should treat all components symmetrically, that is base our analysis on the clr transform. It is thus of importance to make a comparison of the 2 metrics.

Let p and P be 2 compositions.

- The *simplicial distance* Δ between p and P is the Euclidian distance between the clr transforms, given by

$$\Delta^2(p, P) = \sum_{i=1}^{D} (\ln \frac{p_i}{g(p)} - \ln \frac{P_i}{g(P)})^2 \tag{24}$$

- The squared distance between $\mathrm{alr}(p)$ and $\mathrm{alr}(P)$ is given by

$$\| \mathrm{alr}(p) - \mathrm{alr}(P) \|^2 = \sum_{i=1}^{d} (\ln \frac{p_i}{p_D} - \ln \frac{P_i}{P_D})^2 \tag{25}$$

- The *norm* of a composition p is defined as the distance of p to the constant composition $(1/D, ..., 1/D)$. Both alr and clr images of the constant composition is the origin of the corresponding space.

Straightforward computations show that

$$\| \mathrm{alr}(p) - \mathrm{alr}(P) \|^2 = \Delta^2(p, P) + D(\ln \frac{p_D}{g(p)} - \ln \frac{P_D}{g(P)})^2 \tag{26}$$

Table 5: Correlations between the last ilr component and the first two principal components pc1 and pc2 in the alr and clr metrics

Metric	$\text{Cor}(\text{ilr}_d, pc_1)$	$\text{Cor}(\text{ilr}_d, pc_2)$	multiple correlation
alr	0.788	-0.578	0.977
clr	0.418	-0.461	0.623

The last term is proportional to the squared distance between the last isometric logratio $\text{ilr}_d(p)$ and $\text{ilr}_d(P)$, where

$$\text{ilr}_d(p) = e'_d \cdot \ln(p)$$

and e'_d is the last line of the Helmert matrix (see § 2.1). More concretely $\text{ilr}_d(p)$ is the component that contrasts the logarithm of the geometric mean of the d first parts to the logarithm of the last part and has more weight in the alr metric. As expected, because the alr metric gives more weight to the distance in the last ilr, we see in Table 5 (2002 data) that the last ilr component is more correlated with the first 2 alr than clr principal components.

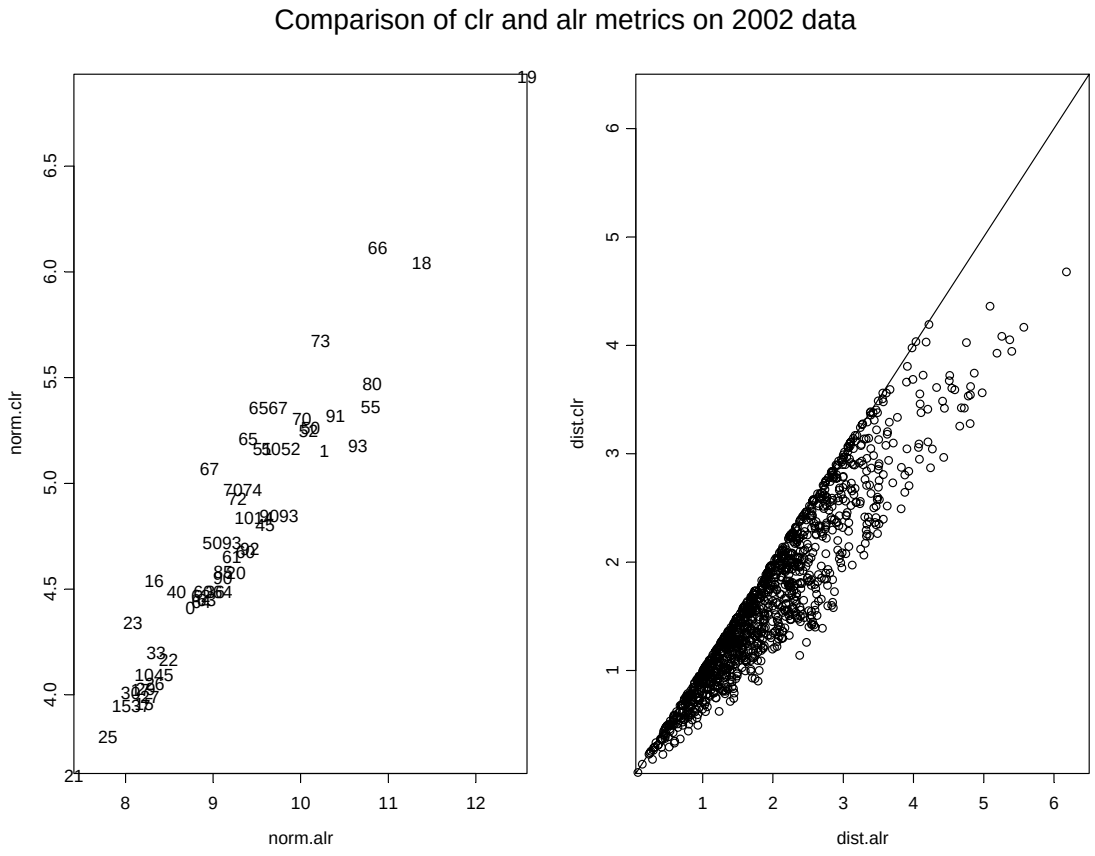


Figure 6: Comparison of clr and alr metrics.

Figure 6 shows the norms (left pane) and the pairwise distances between activity classes (right pane) for both metrics. There is a good agreement between the metrics in this application. We expect similar results on using alternate metrics.

The groups shown in Figure 5 change somehow if the same analysis is made on the clr scale. This is to be expected because they merely form zones in the plane instead of well separated groups, so that very small differences can result in a group change.

4.3.4 Discriminant analysis

In order to see the relationship to clr transforms and also to evaluate the stability of the groups between 2002 and 2004, we performed 2 linear discriminant analyses using the 3 groups shown in Figure 5 that were obtained in 2002 with the help of the alr transformed data[4].

1. The first analysis used as input the 4 principal components computed from the clr transformed 2002 data (not shown). The result of the discriminant analysis is displayed at Figure 7 (top). We see that the point cloud is very similar to the one in Figure 5 that was based on the alr transformed data and that the PAM groupings are equally well separated.

2. In order to see the changes occurred in 2004, the 2004 clr principal components are projected onto the 2002 discriminant functions (Figure 7, bottom). Globally, the groups are stable.

The results of the discriminant analysis are the same, whether we use the alr or the clr transforms. The canonical correlations between the principal axes and the discriminant functions do not depend on the transform used[5]. Only the first 2 discriminant functions are non zero, because 2 functions are enough to distinguish between 3 groups. They are given in Table 6. It can be seen that a very small loss in discriminating power has occurred in 2004: the 2nd canonical correlation is slightly smaller. This is confirmed in Figure 7 where few activity classes changed group in 2004: the largest moving is observed for class 19 (manufacture of leather) that passed from the triangle group to the dot group. Class 52 (retail trade) was at the border between diamonds and triangles and changed to the border between triangles and dots. This implies a moving of aggregate 5052. 30 and 62 moved slightly and changed the group. Class 16 (manufacture of tobacco products) moved down the second discriminant variable. We see in Table A1 that the share in special payments (SOND12E) for class 16 has doubled in 2004.

[4]The first discriminant function is the linear function of the variables that maximizes the ratio of the between-group sum of squares to the within-group sum of squares. The second discriminant function is the linear combination that is uncorrelated with the first, and has the same optimality criterion.

[5]The 1st canonical correlation between two sets of variables is the maximum attainable correlation between a linear combination of the first set of variables and a linear combination of the second. These linear combinations are the first canonical variables. The 2nd canonical correlation is the maximum attainable correlation between 2 linear combinations of the first and second set that are in each set uncorrelated with the preceding canonical variable

Table 6: Canonical correlations of the 2002 discriminant functions and the clr transforms for the 2002 and 2004 surveys

Year	1st canonical correlation	2nd canonical correlation
2002	0.867	0.733
2004	0.867	0.669

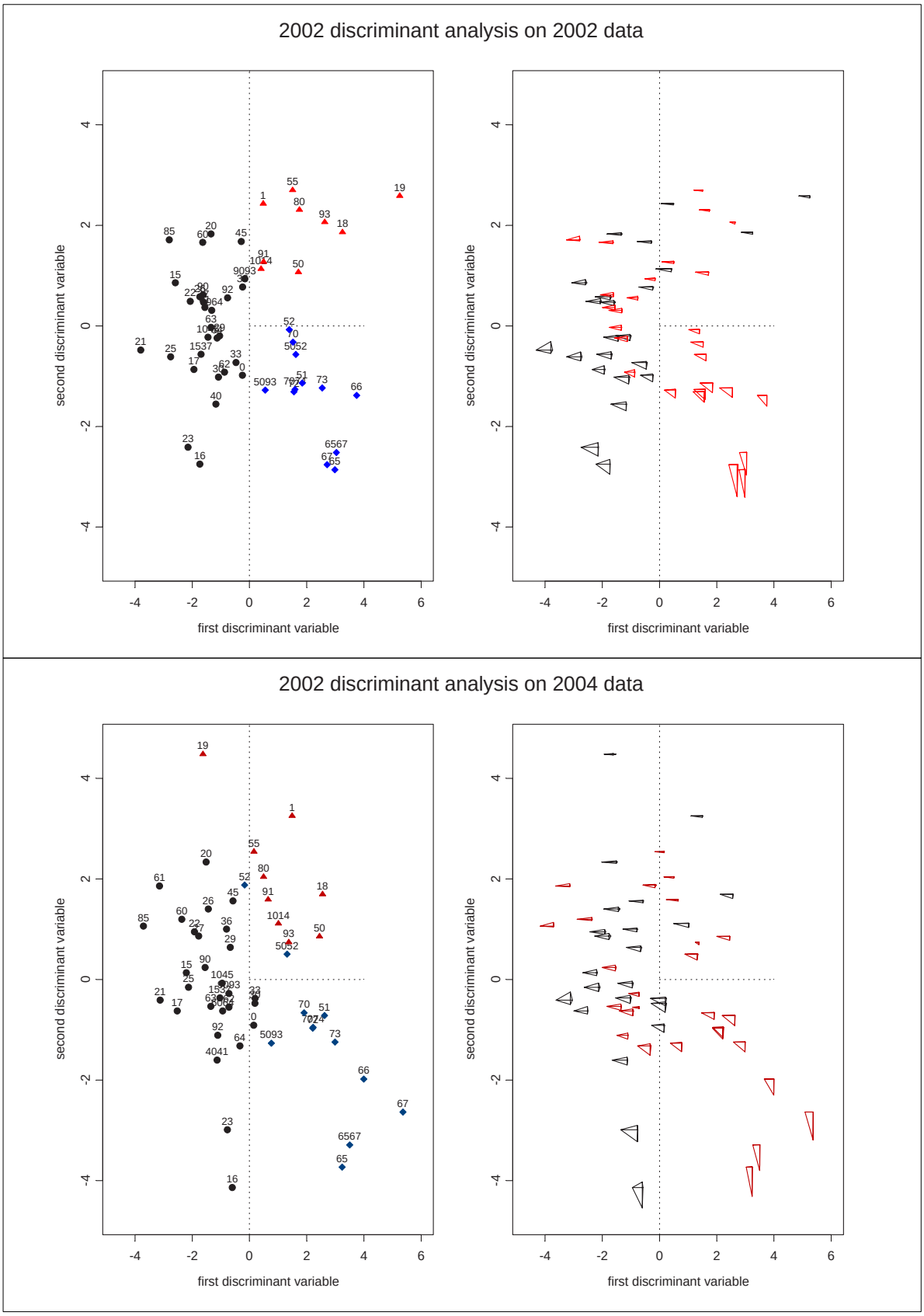


Figure 7: Discriminant analysis of the PAM groupings of the 2002 data (top) and projection onto the 2002 discriminant variables of the 2004 data (bottom).

4.4 Precision

In the preceding paragraph, we did not use the information on the survey design and analyzed the results as if they were a simple data set. Let us return to the concepts described in Chapter 2.

4.4.1 Approximations

The following approximations are obtained by the linearization techniques described in § 4.1.

Univariate statistics Let $\xi_i = \mathrm{E}(x_i)$ and $\mu_i = \mathrm{E}(\ln x_i)$. Denote the coefficient of variation by CV. By linearization and by Equation (14)

$$\mu_i \cong \ln \xi_i \tag{27}$$

$$\ln x_i - \mu_i \cong \frac{x_i - \xi_i}{\xi_i} \tag{28}$$

$$\mathrm{Var}(\ln x_i) = \mathrm{E}(\ln x_i - \mu_i)^2 \cong \frac{\mathrm{Var}(x_i)}{\xi_i^2} = \mathrm{CV}^2(x_i) \tag{29}$$

We get the following 95% CI:

1. *Normal approximation for* x_i:
$$\left[bn_{l95}^{(i)}, bn_{u95}^{(i)}\right] = x_i\left(1 \pm 1.96\,\mathrm{CV}(x_i)\right)$$
2. *Normal approximation for* $\ln(x_i)$:
$$\left[bl_{l95}^{(i)}, bl_{u95}^{(i)}\right] = \ln(x_i) \pm 1.96\,\mathrm{CV}(x_i)$$
3. *Deduced CI for* x_i (*lognormal approximation*):
$$\left[b_{l95}^{(i)}, b_{u95}^{(i)}\right] = \left[\exp\left(bl_{l95}^{(i)}\right), \exp\left(bl_{u95}^{(i)}\right)\right] = x_i \exp\left(\pm 1.96\,\mathrm{CV}(x_i)\right) \tag{30}$$

The CV's can be found in Table A1. If we postulate a lognormal distribution for x, the squared $\mathrm{CV}(x_i)$ is given by $\exp(\sigma_{ii}) - 1$ which slightly overestimates $\sigma_{ii} = \mathrm{Var}(\ln x_i)$.

Covariances and correlations The univariate confidence intervals are misleading because they ignore the dependencies between components. By the linear approximation in Equation (28), matrix $\mathbf{\Sigma} = [\sigma_{ij}]$ in Equation (5) is given by

$$\sigma_{ij} = \mathrm{Cov}(\ln x_i, \ln x_j) \cong \frac{\mathrm{Cov}(x_i, x_j)}{\xi_i \xi_j} \tag{31}$$

The approximation of $\mathbf{\Sigma}$ given by Equation (31) can be seen as a multivariate form of the coefficient of variation. Moreover the correlations are given by correlations of x:

$$\rho_{ij} = \mathrm{Cor}(\ln x_i, \ln x_j) \cong \frac{\mathrm{Cov}(x_i, x_j)}{\xi_i \xi_j\,\mathrm{CV}(x_i)\,\mathrm{CV}(x_j)} = \frac{\mathrm{Cov}(x_i, x_j)}{\sqrt{\mathrm{Var}(x_i)\,\mathrm{Var}(x_j)}} = \mathrm{Cor}(x_i, x_j) \tag{32}$$

which is not surprising, the approximation of $\ln x_i$ being linear in x_i.

These correlations should not be used for sociological interpretations. Indeed the finite population correction implies that exhaustive strata (with full response) are excluded from the calculations and that other strata have different weights. Moreover, differences in the stratification between the 2 surveys (see Table 1) precludes the comparison of correlations matrices between 2002 and 2004. Thus the correlations are only useful for evaluating the precision of the global ratios, and have no other interpretation. Correlation matrices by economic activity class can be found in Table A4 for 2002 and 2004.

Confidence domains In the following confidence domains, we make the simplifying assumption that the correlations above as well as the variances are estimated without error. This assumption leads the use of a chi-square quantile as critical value and is similar to the choice of a normal quantile for the confidence intervals and tests for the social security contributions.
Let $\mathbf{R} = [\rho_{ij}]$ be the 4×4 correlation matrix with elements given by Equation (32), and let us approximate μ by $\ln \xi$ in Equation (27).

The approximate confidence domain in Equation (8) at level $1-\alpha$ is given by:

$$D_{1-\alpha}(x) = \left\{ x \in R_+^4 \mid \mathbf{Q}(x) \leq \chi^2_{4;1-\alpha} \right\} \tag{33}$$

where

$$\mathbf{Q}(x) = \left(\frac{\ln x_1 - \ln \xi_1}{\mathrm{CV}(x_1)} \quad \frac{\ln x_2 - \ln \xi_2}{\mathrm{CV}(x_2)} \quad \frac{\ln x_3 - \ln \xi_3}{\mathrm{CV}(x_3)} \quad \frac{\ln x_4 - \ln \xi_4}{\mathrm{CV}(x_4)} \right) \mathbf{R}^{-1} \begin{pmatrix} \frac{\ln x_1 - \ln \xi_1}{\mathrm{CV}(x_1)} \\ \frac{\ln x_2 - \ln \xi_2}{\mathrm{CV}(x_2)} \\ \frac{\ln x_3 - \ln \xi_3}{\mathrm{CV}(x_3)} \\ \frac{\ln x_4 - \ln \xi_4}{\mathrm{CV}(x_4)} \end{pmatrix} \tag{34}$$

In the coordinates $(\ln x_1, \ln x_2, \ln x_3, \ln x_4)$, this domain is a 4-dimensional ellipsoid. In the coordinates (x_1, x_2, x_3, x_4), the shape of the domain is similar to a drop. There is no direct relationship between the length of the one-dimensional confidence intervals and the limits of the corresponding 4-dimensional confidence domain.

Multivariate test for change Let us denote by $\ln(\mathbf{x}_A)$, $\ln(\mathbf{x}_B)$, $\mathbf{\Sigma}_A$ and $\mathbf{\Sigma}_B$ the estimated alr compositional vectors (in logarithmic scale) and their covariance matrices of a given activity grouping for 2 independent samples A and B respectively. By the independence between the samples, the covariance matrix of the difference between the compositional vectors (in logarithmic scale) is given by the sum $\mathbf{\Sigma}_A + \mathbf{\Sigma}_B$. Under the null hypothesis of no change, the statistics M in Equation (35) follows a chi-square distribution with 4 degrees of freedom:

$$M = (\ln(\mathbf{x}_B) - \ln(\mathbf{x}_A))^t (\mathbf{\Sigma}_A + \mathbf{\Sigma}_B)^{-1} (\ln(\mathbf{x}_B) - \ln(\mathbf{x}_A)) \tag{35}$$

M is known as the Mahalanobis squared distance. If $M > \chi^2_{4;1-\alpha/2}$, the null hypothesis is rejected at the α-level.

4.4.2 Simplex view of confidence domains

3-part compositions can be seen as points within an equilateral triangle with height 1, in which each vertex represents 100% in the corresponding part. To visualize the 95% confidence domains above in Equation (34), let us split the 5-part composition into two 3-part compositions: an amalgamation (VERDUZ+ZULAGEN+SOND12E, XIII13E, BLIMOK), and a subcomposition (ZULAGEN, SOND12E, VERDUZ) (see Aitchison [1] for a thorough description of amalgamation and subcomposition). Both are unit-sum compositional vectors. Because our approximation is linear, it is easy to deduce the corresponding covariance matrices from $\mathbf{\Sigma}$.

The following figures show the results for the economic activity groupings requested by Eurostat in 2002 and 2004. Figures 8, 9 show for years 2002 and 2004 respectively the 95% confidence domains per activity classes for the amalgamation (VERDUZ+ZULAGEN+SOND12E, XIII13E, BLIMOK). The data in the original scale are found at the top panes. At the bottom panes, the data are centred at the national estimate. That is, the compositions are represented by the perturbation to be applied to the national composition in order to get the composition of a particular activity grouping. Thus the national estimate is at the middle of the triangle, and those compositions that are nearer to one edge present a larger share of the corresponding part. It is observed that the sizes of the confidence domains are more similar in the centred representation. Another advantage is that there is no need to zoom the picture. The interpretation is otherwise equivalent.

We see that the amalgamations are very precisely estimated, the worst is for banking, insurance 65-67 for which the uncertainty is essentially in the demarcation between the amalgamation of "gross earnings" and "13th salary" (BLIMOK+XIII12E) and OTHERS. Apart for this group, OTHERS is never larger than 6%. A robust regression line[6] of $\ln(x_3)$ onto $\ln(x_1 + x_2 + x_4)$ in Equation (23)is fitted in the logratio 2-dimensional alr space. It can be seen that groupings from Production are always below the line, indicating a larger share of 13th salary (XIII12E) in Production than in Services. Figures 10 and 11 show how the small amount of OTHERS is distributed among the remaining components. In general, the subcompositions have a minimum of 50% share in bonuses (SOND12E) and very little parts of overtime earnings (VERDUZ) and hardship allowances (ZULAGEN), except grouping 85 "health and social work" which shows narrowly 20% of SOND12E but more than 80% of ZULAGEN. The subcompositions are fairly precisely estimated, as can be seen by the small 95% confidence domains. One exception is the grouping 55 "Hotels and restaurants" which indicates a rather large uncertainty (particularly in 2002) in the separation between SOND12E and VERDUZ. Considered in perspective of Figure 8 or 9, where it is observed that the part OTHERS of the grouping 55 represents only 1%, the rather large uncertainty visible in Figure 10 looses its importance. The centred pictures show a roughly similar distribution of classes around the national estimate between the 2 years. A move towards the VERDUZ edge in 2004 for class 50-52 (sale, repair) can be seen.

[6] An ordinary least squares line would have been attracted by 65-67.

4.4.3 Tests between activity classes

Let us define the samples A and B in Equation (35) as the samples extracted from 2 different and disjoint activity groupings on the same year. Using the independence between the estimates of compositions for these two samples, we can compute the covariance matrix of the difference of compositional logratios. Under the asymptotic distribution, 2 by 2 tests of differences between the 5-dimensional compositions in Equation (22) of economic activity classes within sectors have been processed for the year 2002 at the 5% risk, and show that the null hypothesis of no difference is generally rejected. A small group from 60-64 (transport, storage and communication) are mutually not different in wage compositions, namely 61 (water transport), 63 (supporting and auxiliary transport activities) and 64 (post and telecommunications). Only one other non-significant difference is found between 61 (water transport) and 90 (water processing and other disposal). We conclude that the SESS has a good discriminating power for the wage compositional data.

4.4.4 Tests for change

Did the estimated compositional vector of a given activity class change significantly between 2002 and 2004? Let us now define the samples A and B in Equation (35) as the samples corresponding to the same activity grouping in the years 2002 (A) and 2004 (B). We test the null hypothesis of no change against the alternative that a change occurred in at least one component along the lines described above in Equation (35). Results are found in Table A5. To give a rough idea of the comparative location of the 2002 and 2004 compositional vectors, the geometric means of the 5 components are added. (These geometric means have no direct relationship to the test for change.) The Mahalanobis distance M in Equation (35) shows very large values for "manufacture of tobacco products" (class 16) and "Air transport" (class 62) where structural changes occurred. If we choose the 1% risk level, and the 12 activity groupings required by Eurostat, we have significant differences for the "manufacturing" group 15-37, for "Transport, storage and communication" 60-64 and "Other community, social and personal service activities" 90-93. Table 7 is an extract from Table A1 and shows that the significant differences in these groupings correspond to an increasing proportion of the overall wage bill in some or all the extra components (other than the gross salary BLIMOK) between 2002 and 2004.

Table 7: Significant changes at the 1% level for Eurostat groupings

Activity grouping	Year	VERDUZ	ZULAGEN	XIII12E	SOND12E
15-37	2002	0.6	1.3	7.5	2.5
15-37	2004	0.7	1.3	7.5	2.9
60-64	2002	0.3	1.0	6.8	1.2
60-64	2004	0.4	1.1	6.9	2.5
90-93	2002	0.2	0.4	5.1	1.1
90-93	2004	0.2	0.7	5.3	1.5

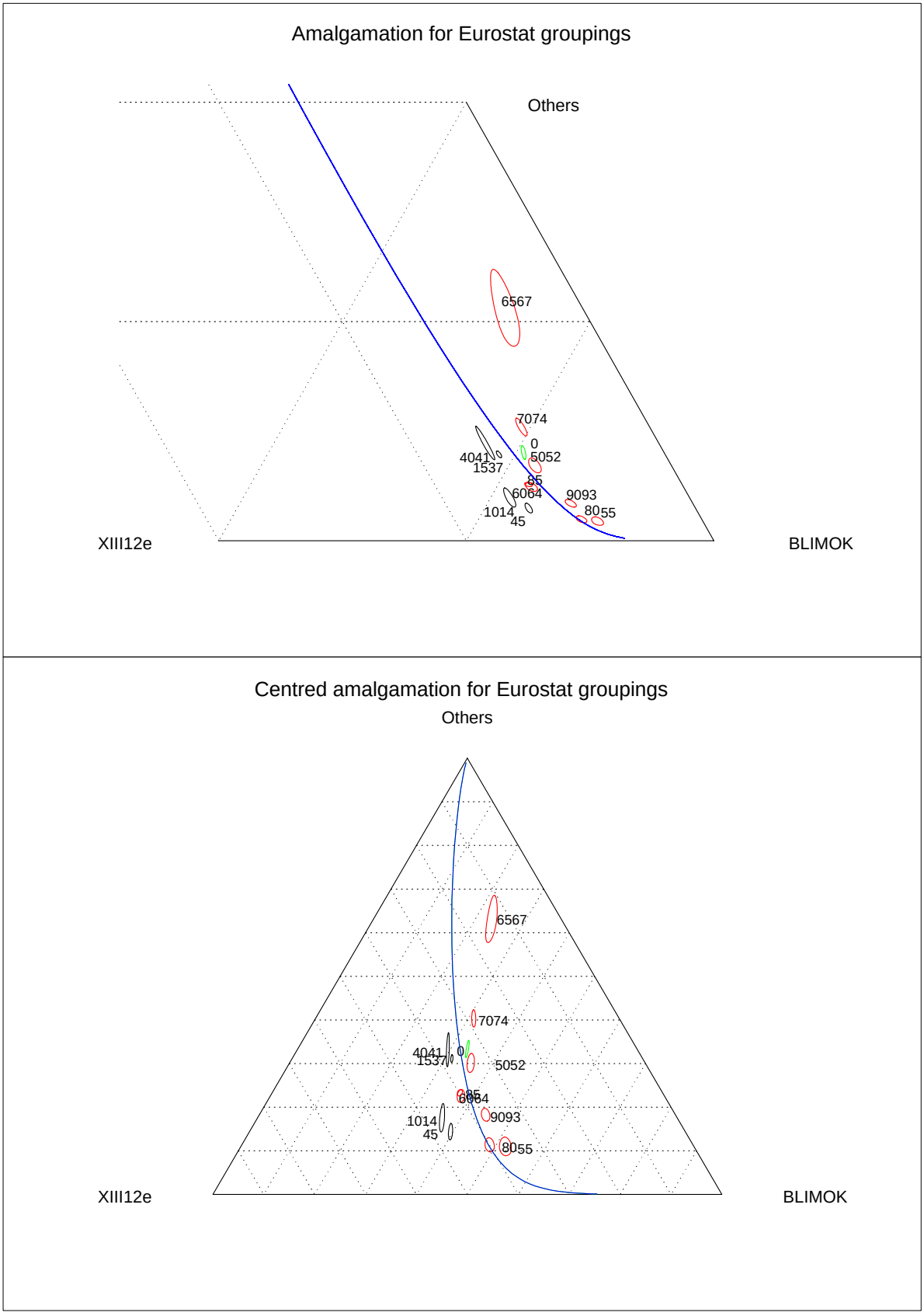


Figure 8: 2002 amalgamations (top) and amalgamations centred on the national estimate given by (O X B)=(4.02 5.68 90.30)% (bottom). Dotted lines are spaced by 10%.

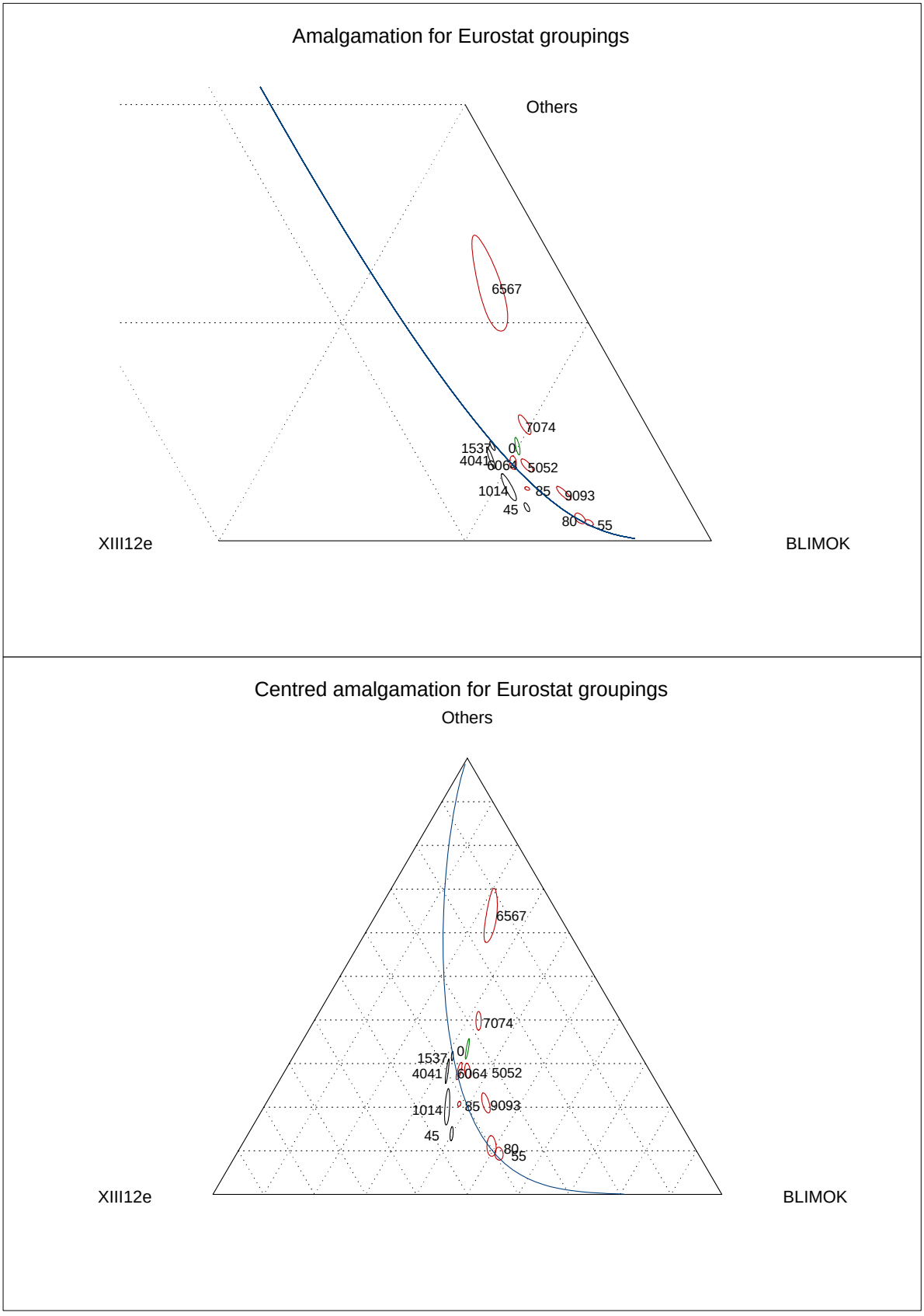


Figure 9: 2004 amalgamations (top) and amalgamations centred on the national estimate given by (O X B)=(4.32 5.71 89.97)% (bottom).

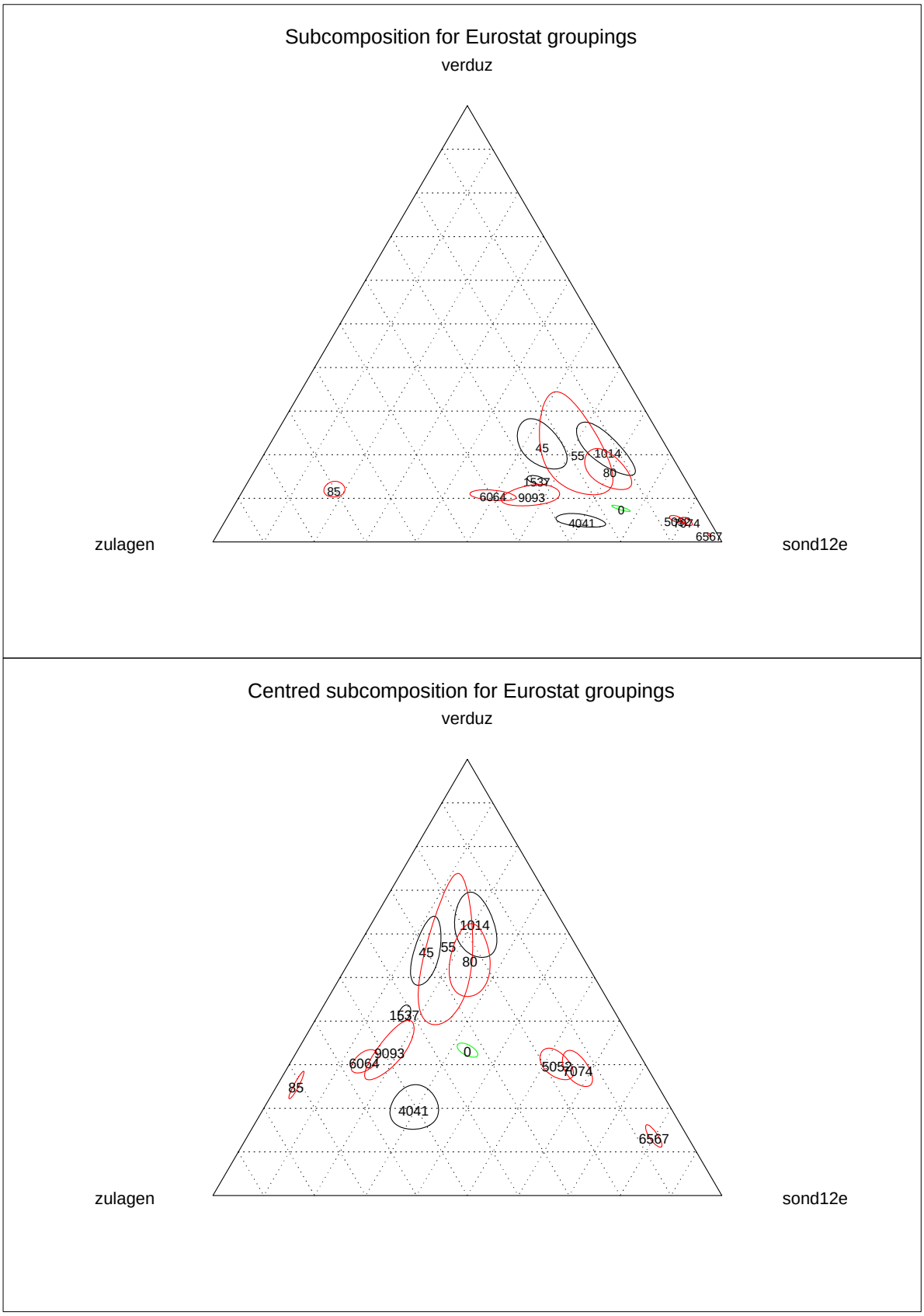


Figure 10: 2002 subcompositions (top) and subcompositions centred on the national estimate given by (v z s)=(7.67 15.95 76.39)% (bottom).

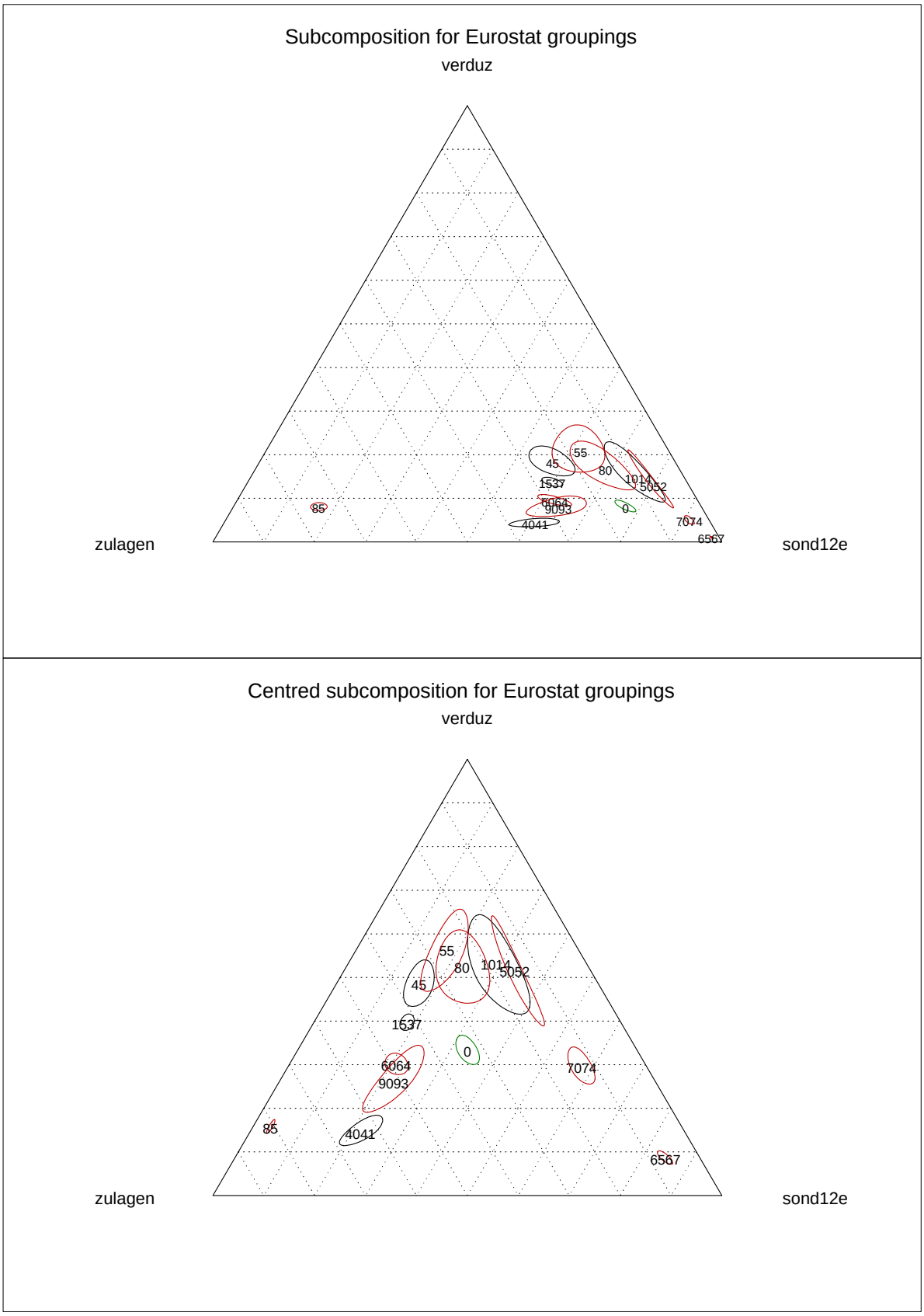


Figure 11: 2004 subcompositions (top) and subcompositions centred on the national estimate given by (v z s)=(8.14 14.81 77.05)% (bottom).

4.4.5 Global CV

Whereas a thorough discussion of precision must take into account the multivariate nature of compositions, it is also of importance to derive a simple summary measure in order to characterize the overall precision. This will be done under the denomination of "global CV".
The total variance per activity class is computed by Equation (12). Taking the first formulation in Equation (11), we define an average standard deviation for logratios by

$$\text{stot}(p) = \sqrt{\frac{\text{totvar}(p)}{(D-1)/2}} = \sqrt{\frac{\sum_{i<j} \text{Var}\left(\ln \frac{p_i}{p_j}\right)}{D\,(D-1)\,/2}} \tag{36}$$

In its linearized form, $\text{Var}\left(\ln \frac{p_i}{p_j}\right) \approx \text{CV}^2\left(\frac{p_i}{p_j}\right)$. We can interpret Equation (36) as a L_2-average of the CV's of all possible ratios of components (i.e. the square root of the mean squared CV's). Practically, it is computed using the linearized form of Equation (5) and the second interpretation of the total variance in Equation (11). It is given in the last column of Table A1 under the heading "global CV". The global CV is a good candidate for an overall assessment of precision and has a theoretical counterpart in the context of compositions. In this application, the global CV is always between the extremes of the 4 corresponding univariate CV's. The representation within the simplex of the 10 economic activity classes with the largest average standard deviation (Figure 12 for 2002) show a differing pattern of variability between classes. For class 67 the observed uncertainty is essentially in the sharing out between gross salary BLIMOK and "others", but category "others" in this case is practically only "Special payments" SOND12E. For class 18 (manufacture of wearing apparel; dressing and dyeing of fur), which has the second largest average standard deviation for logratios, the amalgamation is very precisely estimated. The largest variability is in the subcomposition where we see that the variability lies in the relative parts of SOND12E and VERDUZ. It is interesting to note that the breaking down of the 5-dimensional composition in Equation (22) into the above amalgamation and subcomposition is sufficient for the recovery of the original compositions, but not for the full original covariance matrix. For instance Cov(BLIMOK, VERDUZ) cannot be recovered from the covariance matrices of the provided amalgamation and subcomposition.

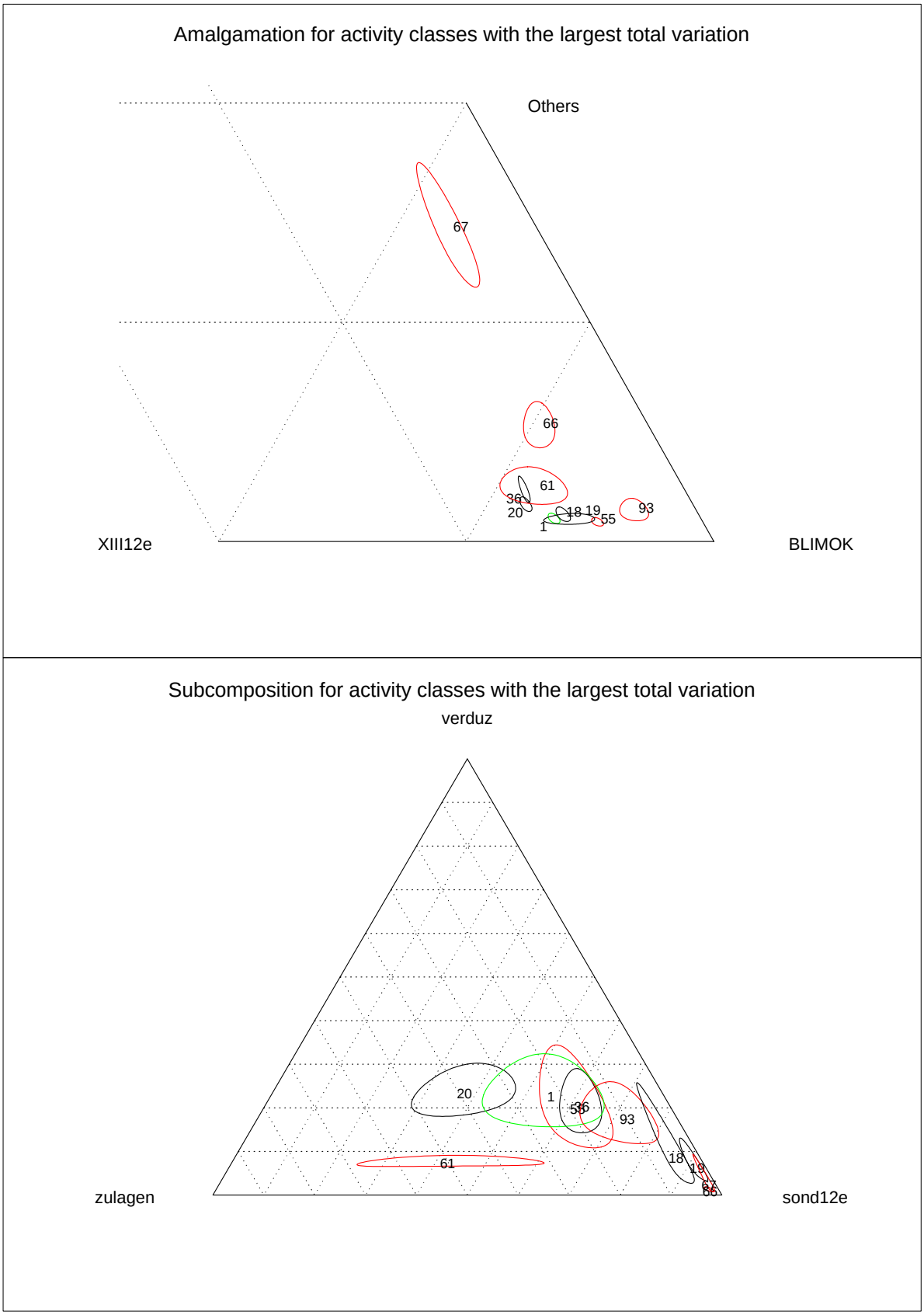


Figure 12: Representation of the activity classes with the 10 largest total variation (2002 data).

4.5 Discussion

The lack of precision in Table A1 is linked with very small proportions, i.e. with a large discrepancy between components. If the discrepancy in components is large, the geometric mean $g(p)$ will be small. For a given dimension D, $\max(g(p)) = 1/D$ is attained for the uniform composition. A plot of $g(p)$ in function of $\text{stot}(p)$ (Figure 13 for the 2002 data) shows a clear relationship between discrepancy and variability. Points are coded by the PAM groupings (Figure 5). This gives a further interpretation of the groups: the dots group has the least discrepant and least variable compositions; the diamonds group has more discrepancy but is still precisely estimated; while for the triangles group, the compositions generally have at least one very small component and also the largest average standard deviation. We see that the top half of the dots group is exclusively formed by activities from the production sector.

Another look at this plot is given at Figure 14 where a comparison between the years is made. Only classes at the most detailed level are shown in Figure 14 (i.e. no groupings). With that remark in mind, top pane is the same plot of $g(p)$ in function of $\text{stot}(p)$ as in Figure 13, but points are coded by the sample size: if n_b and n_w are the number of answering businesses and transmitted wages respectively, the radii of the small and large circles are proportional to $(n_b/1000)^{2/7}$ and $(n_w/1000)^{2/7}$ respectively. The exponent of $2/7$ is chosen by trial and error for a good overall visibility. The factor 1000 has the effect that a sample size of 1000 will be represented by the default size of the plotting character. It is clear from the plot that the very large $\text{stot}(p)$ (i.e. average standard deviation or global CV) are linked to activity classes with a small sample size. Looking at Table A3 where n_b is given, one can see that class 67 has more than doubled its primary unit sample size and the global CV decreased from 25.5 in 2002 to 12.2 in 2004, as can be read in Table A1. This is due to structural changes in this branch. Class 19 has a large move upwards in the geometric mean. Despite its large average standard deviation, we find in Table A5 that the change in the compositional vectors is significant; Table A1 shows that for class 19 there is an increase in "Overtime Earnings" and "Hardship allowance" at the cost of "13th month salary" and "Special payments". The extremely large Mahalanobis distance for class 16 (Table A5) has to be linked with the small number of businesses in the branch (Table A3).

The relationship between the geometric mean and the average standard deviation (or global CV) observed in 2002 is confirmed in 2004.

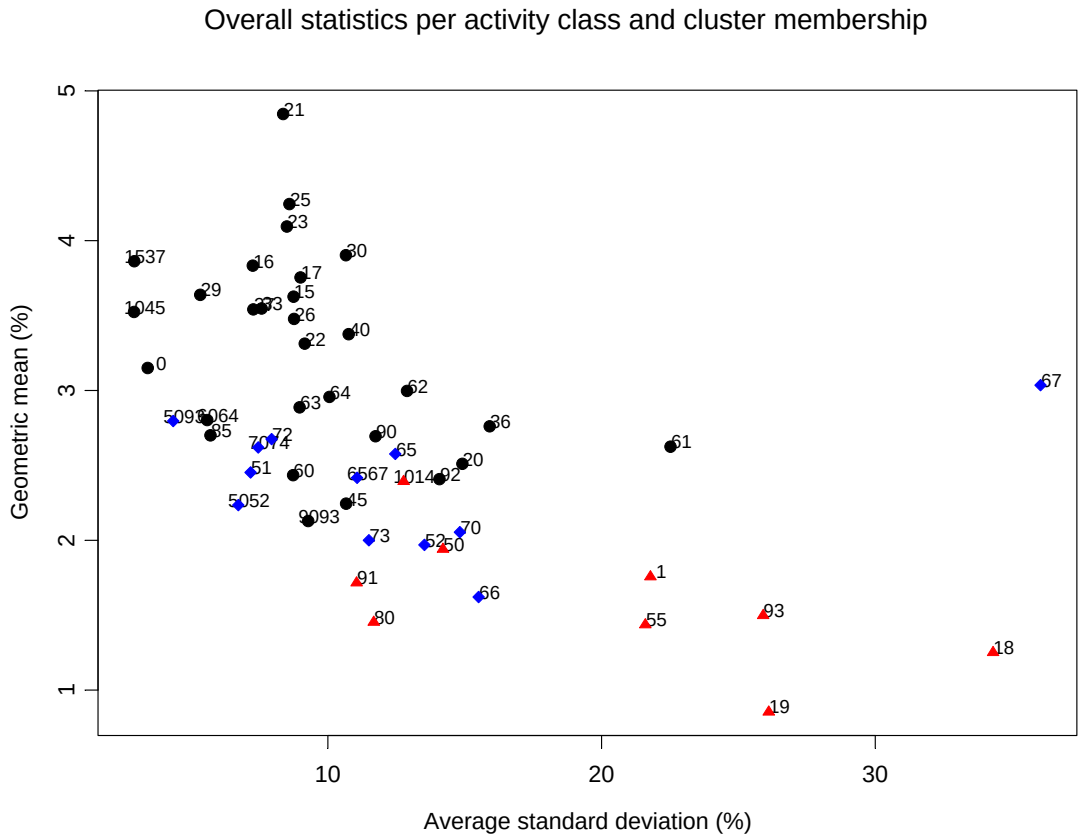


Figure 13: Geometric mean and square root of mean total variation (2002 data). Clustering is based on a partition computed from the multidimensional scaling components (see Figure 5).

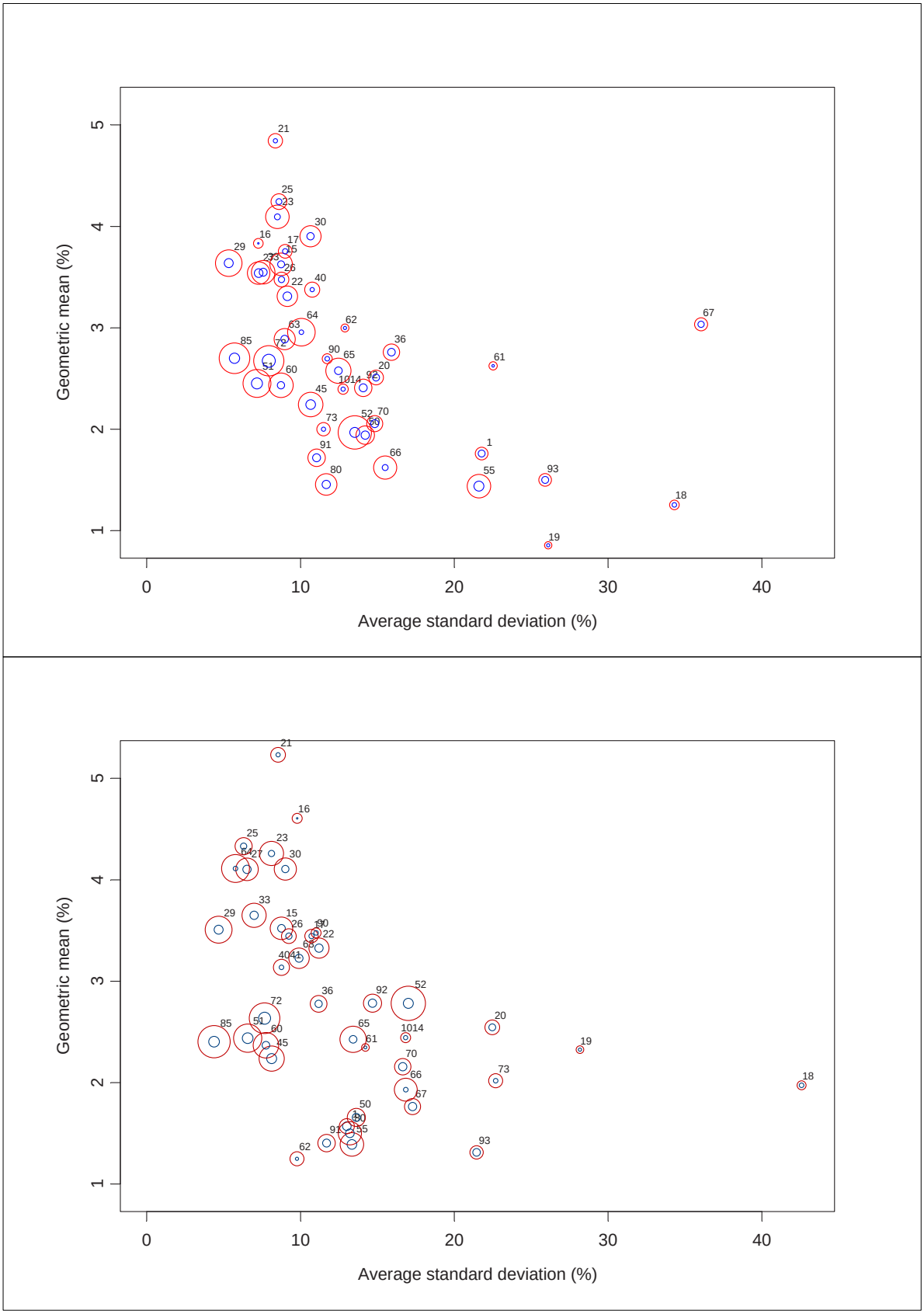


Figure 14: Overall statistics per activity class, 2002 data (top), 2004(bottom). Circles: small radius is a function of the number of answering businesses, large radius of the number of wages.

5 Conclusion

The whole study is based on the interplay between Aitchison's theory of compositional data and the first order approximation of the logratio covariance matrix, interpreted as a multivariate coefficient of variation. If the (univariate) CV is less than 0.3, the approximation is good; otherwise, the computed squared CV overestimates the logratio variance: for a CV = 0.50, the square root of the actual logratio variance would be around 0.47.

The global CV can be viewed as the square root of the average squared CV for all possible ratios of components. It is also the linearized form of Equation (36), which is proportional to the square root of Aitchison's total variance divided by the degrees of freedom. The proportionality factor $\sqrt{2}$ is the norm of the vector c whose components are equal to 0, except $c_i = 1$ and $c_j = -1$, contrasting $\ln(p_i)$ and $\ln(p_j)$.

If the variability of the global estimates of the components is small enough for the linear approximation to be valid, the proposed approach shows a way for generalizing to multivariate compositional data Eurostat's recommendations for communicating precision by CV's. Should the variability be too large, we would suggest that CV's be replaced by the variance of logratios, along the lines given for the analysis of compositional data.

Appendix

Table A1 : Wage components in overall wage bill, in %

Private and public sector (Confederation) combined

Switzerland 2002

TA14	Economic activities	Social security contributions		Overtime earnings		Hardship allowances		13th or nth month wage/salary		Special payments bonuses		Global CV
		in %	CV (%)	in %	CV (%)	in %	CV (%)	in %	CV (%)	in %	CV (%)	in %
0	TOTAL	12.9	***0.1***	0.3	***2.2***	0.7	***1.6***	6.3	***0.9***	3.4	***4.4***	***3.4***
01	Horticulture	12.1	*0.6*	0.3	*17.9*	0.2	*27.8*	6.4	*1.6*	0.6	*10.6*	*21.8*
10-45	SECTOR 2 PRODUCTION	13.5	***0.2***	0.5	***2.9***	1.1	***2.2***	7.5	***0.4***	2.1	***2.8***	***2.9***
10-14	Mining and quarrying of stone	14.2	*0.5*	0.4	*8.9*	0.3	*12.3*	8.0	*0.9*	1.4	*13.5*	*12.8*
15-37	Manufacturing	13.3	***0.2***	0.6	***2.6***	1.3	***2.2***	7.5	***0.5***	2.5	***2.9***	***2.9***
15	Manufacture of food products and beverages	13.0	*0.3*	0.8	*10.3*	1.7	*6.0*	7.1	*0.9*	1.1	*7.3*	*8.7*
16	Manufacture of tobacco products	11.7	*0.1*	0.2	*11.0*	2.4	*1.7*	7.3	*0.2*	5.3	*2.9*	*7.3*
17	Manufacture of textiles	13.1	*0.5*	0.4	*9.5*	1.9	*5.1*	6.2	*3.1*	2.4	*8.5*	*9.0*
18	Manufact. of wearing apparel; dressing and dyeing of fur	12.6	*0.8*	0.1	*50.4*	0.1	*17.3*	5.9	*2.0*	1.1	*11.6*	*34.3*
19	Manufacture of leather and leather products	12.1	*1.0*	0.1	*29.3*	0.0	*26.9*	5.6	*8.6*	1.0	*10.7*	*26.1*
20	Manufacture of wood and wood products	13.7	*0.5*	0.4	*9.9*	0.7	*18.7*	7.4	*1.3*	0.7	*10.1*	*14.9*
21	Manufacture of pulp, paper and paper products	13.3	*0.8*	0.9	*10.5*	4.4	*6.0*	7.7	*1.0*	1.8	*5.2*	*8.4*
22	Publishing, printing, reproduction	13.5	*0.5*	0.5	*8.8*	1.4	*10.4*	7.5	*1.3*	1.3	*5.7*	*9.2*
23,24	Manufacture of coke, chemicals	12.1	*1.1*	0.2	*8.8*	2.4	*5.6*	8.9	*2.2*	4.6	*7.1*	*8.5*
25	Manufacture of rubber and plastic products	13.4	*0.6*	0.7	*8.2*	2.5	*5.3*	7.3	*1.2*	1.9	*9.2*	*8.6*
26	Manufacture of other non-metallic mineral products	13.6	*0.6*	0.8	*7.7*	1.0	*6.7*	7.8	*0.9*	1.5	*10.9*	*8.8*
27,28	Manufacture of basic metals	13.7	*0.3*	0.8	*5.8*	1.0	*5.4*	7.1	*0.9*	1.6	*7.8*	*7.3*
29,34,35	Manufact. of machinery & eq. N.E.C., -of motor vehicles	13.5	*0.3*	0.7	*4.2*	0.8	*5.8*	7.3	*1.5*	2.5	*4.1*	*5.3*
30-32	Manufact. of electrical equipment, precision machinery	13.8	*0.7*	0.6	*12.1*	1.0	*7.2*	7.6	*0.7*	3.6	*7.5*	*10.7*
33	Manufacture of medical and precision instruments	13.0	*0.6*	0.6	*5.2*	0.6	*7.8*	7.6	*0.9*	3.5	*7.6*	*7.6*
36,37	Manufacturing N.E.C.	13.5	*0.7*	0.4	*23.4*	0.4	*7.4*	7.1	*1.1*	1.6	*10.7*	*15.9*
40,41	Electricity, gas and water supply	11.1	*0.8*	0.3	*14.4*	1.3	*4.0*	7.9	*0.7*	4.0	*10.0*	*10.8*
45	Construction	14.1	*0.4*	0.4	*12.8*	0.4	*7.6*	7.3	*0.8*	0.9	*8.9*	*10.7*
50-93	SECTOR 3 SERVICES	12.6	***0.2***	0.2	***3.2***	0.5	***2.2***	5.6	***1.4***	4.0	***5.4***	***4.4***
50-52	Sale, repair	12.2	***0.4***	0.2	***5.3***	0.2	***7.2***	6.0	***1.7***	3.1	***4.8***	***6.7***
50	Sale, repair of motor vehicles	12.6	*0.5*	0.3	*13.7*	0.1	*17.1*	6.7	*1.2*	1.6	*6.2*	*14.2*
51	Brokerage, wholesale trade	12.5	*0.3*	0.2	*5.4*	0.2	*8.1*	6.4	*1.0*	4.5	*5.8*	*7.2*
52	Retail trade, repair of pers. & household goods	11.9	*0.7*	0.1	*10.5*	0.3	*11.9*	5.5	*4.2*	2.2	*11.5*	*13.5*
55	Hotels and restaurants	11.5	*0.7*	0.2	*28.9*	0.2	*15.5*	4.4	*2.2*	0.6	*10.0*	*21.6*
60-64	Transport, storage and communication	13.4	***0.3***	0.3	***3.5***	1.0	***4.7***	6.8	***1.5***	1.2	***6.4***	***5.6***
60	Land transport/pipelines	14.2	*0.2*	0.3	*6.3*	0.9	*7.7*	7.3	*0.6*	0.6	*9.2*	*8.7*
61	Water transport	12.7	*1.9*	0.2	*16.1*	1.3	*5.5*	6.5	*9.7*	1.1	*32.8*	*22.5*
62	Air transport	12.3	*0.7*	0.2	*16.7*	1.3	*8.6*	5.3	*1.8*	2.5	*8.5*	*12.9*
63	Supporting and auxiliary transport activities	12.6	*0.5*	0.3	*9.4*	1.1	*9.2*	6.1	*5.8*	1.6	*8.2*	*9.0*
64	Post and telecommunications	13.2	*0.4*	0.3	*3.2*	1.0	*8.4*	6.9	*1.8*	1.6	*11.7*	*10.1*
65-67	Banking; insurance	12.5	***0.7***	0.2	***10.8***	0.2	***9.0***	3.6	***7.7***	11.7	***7.5***	***11.1***
65	Banking	13.0	*0.5*	0.2	*9.1*	0.3	*8.9*	3.0	*11.5*	14.2	*8.1*	*12.5*
66	Insurance	11.8	*1.5*	0.1	*15.8*	0.1	*16.0*	4.8	*7.3*	5.6	*8.8*	*15.5*
67	Activities relating to banking/insurance	11.9	*0.9*	0.5	*49.9*	0.2	*24.6*	4.3	*6.9*	16.6	*9.8*	*36.0*
70-74	Real estate, computer, research & development	12.3	***0.3***	0.3	***8.0***	0.3	***8.2***	5.8	***1.1***	5.2	***3.7***	***7.5***
70,71	Real estate activities/renting of machinery & equipment	12.2	*0.7*	0.1	*11.7*	0.2	*18.6*	6.5	*3.1*	2.6	*6.4*	*14.8*
72,74	Computer and related activities; other business activities	12.3	*0.3*	0.3	*8.4*	0.3	*8.8*	5.7	*1.2*	5.4	*3.9*	*7.9*
73	Research and development	12.5	*0.5*	0.1	*12.3*	0.2	*8.3*	6.6	*3.3*	4.7	*9.8*	*11.5*
75	Public administration, national defence; social security	14.9	*1)*	0.2	*1)*	0.5	*1)*	7.8	*1)*	0.5	*1)*	
80	Education	13.2	*0.4*	0.2	*12.1*	0.1	*10.4*	5.2	*1.7*	0.7	*9.0*	*11.7*
85	Health and social work	13.4	*0.3*	0.3	*6.5*	2.0	*1.9*	6.8	*1.1*	0.5	*5.4*	*5.7*
90-93	Other community, social and personal service activities	12.4	***0.3***	0.2	***9.0***	0.4	***9.5***	5.1	***1.7***	1.1	***5.7***	***9.3***
90	Waste processing and other disposal	13.5	*0.9*	0.3	*7.5*	1.2	*10.9*	6.6	*2.3*	1.1	*13.1*	*11.7*
91	Activities of membership organizations n.e.c.	12.8	*0.3*	0.1	*11.3*	0.4	*11.3*	6.1	*1.4*	0.9	*7.5*	*11.1*
92	Recreational, cultural and sporting activities	12.1	*0.5*	0.3	*16.2*	0.5	*11.0*	5.3	*2.4*	1.3	*8.5*	*14.1*
93	Other service activities	11.9	*0.8*	0.3	*15.1*	0.1	*35.0*	2.6	*9.9*	1.0	*17.5*	*25.9*

Wage bill : Total of non-standardised gross monthly salary

Non-standardised gross salary : Gross salary in the month of October (incl. employee social insurance contributions, benefits in kind, regularly paid shares in bonuses, turnover or commission), but without any overtime pay, hardship allowances (for shift, night and Sunday work), 13th month salary and annual special payments.

CV: coefficient of variation; 1): not computed, because the results are not based on a random sample.

Global CV: Linearized form of the average standard deviation, that is square root of average total linearized variance of logratios of components (see text).

Source: Swiss Federal Statistical Office, Swiss Earnings Structure Survey (SESS) 2002

Original table: wage components only.

Table A1 (continued) : Wage components in overall wage bill, in %

Private and public sector (Confederation) combined

Switzerland 2004

TA14	Economic activities	Social security contributions		Overtime earnings		Hardship allowances		13th or nth month wage/salary		Special payments bonuses		Global CV
		in %	CV (%)	in %	CV (%)	in %	CV (%)	in %	CV (%)	in %	CV (%)	in %
0	TOTAL	13.0	0.1	0.4	5.4	0.7	1.4	6.3	1.0	3.7	5.1	3.5
01	Horticulture	12.5	0.7	0.4	13.2	0.1	9.4	5.9	1.7	0.8	12.7	9.2
10-45	SECTOR 2 PRODUCTION	13.8	0.2	0.6	2.4	1.1	2.1	7.5	0.3	2.4	3.2	2.0
10-14	Mining and quarrying of stone	14.3	0.6	0.4	20.0	0.2	15.3	7.7	1.0	2.0	13.0	11.9
15-37	Manufacturing	13.5	0.2	0.7	2.5	1.3	2.2	7.5	0.4	2.9	3.4	2.2
15	Manufacture of food products and beverages	13.3	0.4	0.5	7.4	1.7	6.8	7.1	1.0	1.5	9.2	6.2
16	Manufacture of tobacco products	11.9	0.1	0.3	15.4	2.9	1.2	5.2	0.4	10.6	0.7	6.9
17	Manufacture of textiles	13.5	0.8	0.3	9.8	2.3	8.3	7.0	1.7	1.5	10.9	7.6
18	Manufact. of wearing apparel; dressing and dyeing of fur	12.9	2.3	0.4	59.0	0.1	24.5	6.5	2.5	2.2	9.6	30.1
19	Manufacture of leather and leather products	13.2	1.0	1.5	25.9	0.3	33.2	4.6	5.1	0.4	10.5	19.9
20	Manufacture of wood and wood products	14.2	0.5	0.5	11.4	0.6	32.3	7.5	0.9	0.7	10.5	15.9
21	Manufacture of pulp, paper and paper products	13.4	0.6	1.1	12.6	4.3	2.8	7.4	0.9	2.2	3.9	6.0
22	Publishing, printing, reproduction	13.9	0.4	0.6	13.3	1.2	11.4	7.7	0.9	1.2	6.1	7.9
23,24	Manufacture of coke, chemicals	12.1	0.7	0.3	7.0	2.3	5.2	8.8	2.1	6.2	8.0	5.7
25	Manufacture of rubber and plastic products	13.5	0.4	0.8	6.0	2.2	4.5	7.5	0.9	2.2	6.3	4.5
26	Manufacture of other non-metallic mineral products	13.6	0.5	0.9	8.0	0.9	8.8	7.6	1.1	1.4	9.3	6.5
27,28	Manufacture of basic metals	14.2	0.3	1.1	5.7	1.4	6.2	7.4	0.8	1.8	5.9	4.6
29,34,3	Manufact. of machinery & eq. N.E.C., -of motor vehicles	13.7	0.3	0.7	3.7	0.8	5.7	7.3	0.6	2.2	3.1	3.3
30-32	Manufact. of electrical equipment, precision machinery	13.8	1.1	0.8	10.4	0.8	5.1	7.3	1.3	4.6	7.1	6.4
33	Manufacture of medical and precision instruments	13.2	0.6	0.6	4.3	0.7	7.7	7.7	0.5	3.8	6.6	4.9
36,37	Manufacturing N.E.C.	13.5	0.4	0.4	10.1	0.6	13.7	7.1	0.8	1.4	6.2	7.9
40,41	Electricity, gas and water supply	12.1	0.5	0.2	11.3	1.4	4.9	7.9	0.6	2.6	7.8	6.2
45	Construction	14.7	0.3	0.3	7.6	0.4	7.0	7.3	0.6	1.0	8.2	5.7
50-93	SECTOR 3 SERVICES	12.6	0.2	0.3	10.5	0.5	1.8	5.8	1.5	4.4	6.3	5.7
50-52	Sale, repair	12.4	0.3	0.5	22.6	0.3	9.6	6.3	1.3	3.0	3.9	10.3
50	Sale, repair of motor vehicles	12.7	0.5	0.2	11.6	0.1	16.0	6.7	1.2	2.0	9.1	9.6
51	Brokerage, wholesale trade	12.5	0.3	0.3	5.5	0.2	7.8	6.5	0.8	5.3	3.9	4.6
52	Retail trade, repair of pers. & household goods	12.1	0.6	0.8	27.4	0.4	11.4	6.2	2.8	1.4	6.7	12.0
55	Hotels and restaurants	11.6	0.4	0.2	14.5	0.1	14.0	4.8	1.5	0.5	7.7	9.4
60-64	Transport, storage and communication	13.5	0.3	0.4	3.3	1.1	3.2	6.9	1.2	2.5	5.8	3.5
60	Land transport/pipelines	14.0	0.2	0.2	8.5	1.0	6.3	7.7	0.5	0.6	5.8	5.5
61	Water transport	13.7	1.1	0.3	11.7	1.2	6.1	7.4	3.0	0.4	17.6	10.1
62	Air transport	14.2	0.2	0.0	14.2	0.4	5.2	3.4	1.2	0.7	4.0	6.9
63	Supporting and auxiliary transport activities	13.0	1.2	0.3	12.5	1.3	7.8	7.4	0.5	2.0	5.6	7.0
64	Post and telecommunications	13.3	0.5	0.6	4.7	1.2	3.8	6.5	3.2	4.8	5.1	4.1
65-67	Banking; insurance	12.5	0.7	0.1	6.0	0.2	8.9	3.7	9.2	13.3	8.6	7.6
65	Banking	12.4	0.8	0.2	6.0	0.3	7.9	3.1	14.0	15.2	10.3	9.5
66	Insurance	12.6	1.7	0.1	13.6	0.1	17.0	4.9	8.1	8.3	13.0	11.9
67	Activities relating to banking/insurance	12.0	0.4	0.1	23.9	0.1	12.2	4.2	2.8	14.5	4.2	12.2
70-74	Real estate, computer, research & development	12.3	0.4	0.3	7.5	0.2	7.2	5.5	1.4	5.4	4.0	5.1
70,71	Real estate activities/renting of machinery & equipment	12.4	0.7	0.2	15.7	0.2	15.8	6.6	1.9	3.4	14.3	11.8
72,74	Computer and related activities; other business activities	12.2	0.4	0.3	7.9	0.2	7.9	5.4	1.6	5.6	4.3	5.4
73	Research and development	12.7	0.8	0.1	31.5	0.1	16.1	6.0	2.3	5.1	4.4	16.0
80	Education	13.1	0.4	0.2	10.3	0.2	10.8	5.1	1.7	0.7	14.2	9.3
85	Health and social work	13.2	0.2	0.2	4.8	2.0	1.4	6.9	0.6	0.4	4.8	3.1
90-93	Other community, social and personal service activities	12.4	0.3	0.2	8.9	0.7	13.7	5.3	1.6	1.5	6.0	8.0
90	Waste processing and other disposal	13.9	0.7	0.6	10.8	1.2	9.9	6.8	2.1	1.8	9.9	7.8
91	Activities of membership organizations n.e.c.	12.6	0.4	0.1	13.2	0.2	11.0	6.2	1.7	0.8	7.3	8.3
92	Recreational, cultural and sporting activities	12.3	0.5	0.2	15.8	1.1	13.3	5.7	2.3	2.0	7.5	10.4
93	Other service activities	11.7	0.7	0.2	21.0	0.1	23.2	1.8	5.5	1.2	15.2	15.2

Wage bill : Total of non-standardised gross monthly salary

Non-standardised gross salary : Gross salary in the month of October (incl. employee social insurance contributions, benefits in kind, regularly paid shares in bonuses, turnover or commission), but without any overtime pay, hardship allowances (for shift, night and Sunday work), 13th month salary and annual special payments.

CV: coefficient of variation.

Global CV: Linearized form of the average standard deviation, that is square root of average total linearized variance of logratios of components (see text).

Source: Swiss Federal Statistical Office, Swiss Earnings Structure Survey (SESS) 2004

Original table: wage components only.

Table A2: p-values for the 2 by 2 equality tests for social security contributions

1) Sector 2 + horticulture **2002**

Noga2	40	16	23	19	1	18	33	15	17	21	25	36	29	26	22	20	27	30	45	1014
40	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
16	.	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
23	.	.	0	57	62	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
19	.	.	57	0	54	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1	.	.	62	54	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
18	.	.	.	.	.	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.
33	.	.	.	.	.	.	0	50	90	.	.	.	.	.	.	.	.	.	.	.
15	.	.	.	.	.	.	50	0	94	.	.	.	.	.	.	.	.	.	.	.
17	.	.	.	.	.	.	90	94	0	95	.	.	.	.	.	.	.	.	.	.
21	.	.	.	.	.	.	.	.	95	0	84	90	96	96	99	.	.	.	.	.
25	.	.	.	.	.	.	.	.	.	84	0	67	78	83	94	99	.	.	.	.
36	.	.	.	.	.	.	.	.	.	90	67	0	57	67	83	96	99	99	.	.
29	.	.	.	.	.	.	.	.	.	96	78	57	0	65	88	99	.	.	.	.
26	.	.	.	.	.	.	.	.	.	96	83	67	65	0	72	92	98	97	.	.
22	.	.	.	.	.	.	.	.	.	99	94	83	88	72	0	81	92	93	.	.
20	.	.	.	.	.	.	.	.	.	.	99	96	99	92	81	0	63	75	.	.
27	.	.	.	.	.	.	.	.	.	.	.	99	.	98	92	63	0	71	.	.
30	.	.	.	.	.	.	.	.	.	.	.	99	.	97	93	75	71	0	.	.
45	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0	87
1014	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	87	0

2) Sector 3 **2002**

Noga2	55	66	67	93	52	92	70	62	72	51	63	73	50	61	91	65	64	80	85	90	60
55	0	85	98	99	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
66	85	0	63	71	75	95	98	99	99	.	.	.	.	.	.	.	.	.	.	.	.
67	98	63	0	61	69	98	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
93	99	71	61	0	58	97	99	.	.	.	.	.	.	.	.	.	.	.	.	.	.
52	.	75	69	58	0	96	99	.	.	.	.	.	.	.	.	.	.	.	.	.	.
92	.	95	98	97	96	0	85	91	99	.	.	.	.	99	.	.	.	.	.	.	.
70	.	98	.	99	99	85	0	63	71	99	.	.	.	96	.	.	.	.	.	.	.
62	.	99	.	.	.	91	63	0	55	97	99	.	.	95	.	.	.	.	.	.	.
72	.	99	.	.	.	99	71	55	0	.	.	.	.	95	.	.	.	.	.	.	.
51	.	.	.	.	.	.	99	97	.	0	89	95	97	83	.	.	.	.	.	.	.
63	.	.	.	.	.	.	.	99	.	89	0	59	66	70	.	.	.	.	.	.	.
73	.	.	.	.	.	.	.	.	.	95	59	0	57	67	.	.	.	.	.	.	.
50	.	.	.	.	.	.	.	.	.	97	66	57	0	65	.	.	.	.	.	.	.
61	.	.	.	.	.	99	96	95	95	83	70	67	65	0	72	88	96	99	.	.	.
91	.	.	.	.	.	.	.	.	.	.	.	.	.	72	0	99	.	.	.	.	.
65	.	.	.	.	.	.	.	.	.	.	.	.	.	88	99	0	96	.	.	.	.
64	.	.	.	.	.	.	.	.	.	.	.	.	.	96	.	96	0	94	.	.	.
80	.	.	.	.	.	.	.	.	.	.	.	.	.	99	.	.	94	0	.	98	.
85	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0	71	.
90	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	98	71	0	.
60	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0

Less than 97.5% p-values show non-significant differences.
Larger than 99.5% % p-values are replaced by dots

Noga2 class numbers are ordered by increasing social security contributions.
(if several classes are grouped, the first is given. See Table A3 for exact denominations)

Table A2 (continued):

1) Sector 2 + horticulture **2004**

	16	23	40	1	18	33	19	15	21	17	36	25	26	29	30	22	27	20	1014	45
16	0	98	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
23	98	0	54	.	99	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
40	.	54	0	.	99	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1	.	.	.	0	90	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
18	.	99	99	90	0	81	84	92	95	97	98	98	99	99	.	.	.	.	.	.
33	.	.	.	.	81	0	61	97	99	99	.	.	.	.	.	.	.	.	.	.
19	.	.	.	.	84	61	0	80	90	95	98	99	.	.	.	.	.	.	.	.
15	.	.	.	.	92	97	80	0	81	92	99	.	.	.	.	.	.	.	.	.
21	.	.	.	.	95	99	90	81	0	72	81	89	98	.	99	.	.	.	.	.
17	.	.	.	.	97	99	95	92	72	0	53	63	87	92	96	.	.	.	.	.
36	.	.	.	.	98	.	98	99	81	53	0	65	94	99	98	.	.	.	.	.
25	.	.	.	.	98	.	99	.	89	63	65	0	90	98	97	.	.	.	.	.
26	.	.	.	.	99	.	.	.	98	87	94	90	0	59	86	.	.	.	.	.
29	.	.	.	.	99	.	.	.	.	92	99	98	59	0	84	.	.	.	.	.
30	.	.	.	.	.	.	.	.	99	96	98	97	86	84	0	71	99	99	.	.
22	.	.	.	.	.	.	.	.	.	.	.	.	.	.	71	0	.	.	.	.
27	.	.	.	.	.	.	.	.	.	.	.	.	.	.	99	.	0	60	86	.
20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	99	.	60	0	78	.
1014	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	86	78	0	.
45	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0

2) Sector 3 **2004**

	55	93	67	52	72	92	70	65	51	91	66	73	50	63	80	85	64	61	90	60	62
55	0	92	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
93	92	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
67	.	.	0	88	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
52	.	.	88	0	84	90	99	99	.	.	99	.	.	.	.	.	.	.	.	.	.
72	.	.	.	84	0	66	94	97	.	.	97	.	.	.	.	.	.	.	.	.	.
92	.	.	.	90	66	0	89	94	.	.	95	.	.	.	.	.	.	.	.	.	.
70	.	.	.	99	94	89	0	66	95	98	86	98	.	.	.	.	.	.	.	.	.
65	.	.	.	99	97	94	66	0	82	92	79	94	99	.	.	.	.	.	.	.	.
51	.	.	.	.	.	.	95	82	0	84	66	88	99	99	.	.	.	.	.	.	.
91	.	.	.	.	.	.	98	92	84	0	55	71	92	98	.	.	.	.	.	.	.
66	.	.	.	99	97	95	86	79	66	55	0	56	66	89	99	.	.	.	.	.	.
73	.	.	.	.	.	.	98	94	88	71	56	0	67	94	.	.	.	.	.	.	.
50	.	.	.	.	.	.	.	99	99	92	66	67	0	91	.	.	.	.	.	.	.
63	.	.	.	.	.	.	.	.	99	98	89	94	91	0	84	93	96	.	.	.	.
80	.	.	.	.	.	.	.	.	.	.	99	.	.	84	0	90	95	.	.	.	.
85	.	.	.	.	.	.	.	.	.	.	.	.	.	93	90	0	81	.	.	.	.
64	.	.	.	.	.	.	.	.	.	.	.	.	.	96	95	81	0	.	.	.	.
61	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0	73	91	.
90	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	73	0	83	.
60	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	91	83	0	.
62	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0

Table A3: Tests of significance of difference in social security contributions between years 2002 and 2004

Noga2	Economic activities	2002 estimate	2004 estimate	p-value	signi-ficance[1]	2002 no of businesses	2004 no of businesses
0	TOTAL	12.90	13.00	1.00	**	42133	42670
01	Horticulture	12.13	12.48	1.00	**	696	1396
10-45	SECTOR 2 PRODUCTION	13.45	13.76	1.00	**	13842	14189
10-14	Mining and quarrying of stone	14.23	14.30	0.72	-	165	175
15-37	Manufacturing	13.27	13.46	1.00	**	11303	11420
15	Manufacture of food products and beverages	12.99	13.32	1.00	**	732	1000
16	Manufacture of tobacco products	11.72	11.94	1.00	**	11	10
17	Manufacture of textiles	13.12	13.49	1.00	**	311	292
18	Manufact. of wearing apparel; dressing and dyeing of fur	12.57	12.88	0.83	-	202	198
19	Manufacture of leather and leather products	12.12	13.20	1.00	**	73	79
20	Manufacture of wood and wood products	13.72	14.21	1.00	**	754	745
21	Manufacture of pulp, paper and paper products	13.32	13.41	0.74	-	159	160
22	Publishing, printing, reproduction	13.62	13.90	1.00	**	1449	1240
23,24	Manufacture of coke, chemicals	12.08	12.12	0.59	-	499	571
25	Manufacture of rubber and plastic products	13.45	13.52	0.78	-	459	520
26	Manufacture of other non-metallic mineral products	13.56	13.63	0.75	-	507	504
27,28	Manufacture of basic metals	13.74	14.19	1.00	**	1440	1284
29,34,35	Manufact. of machinery & eq. N.E.C., -of motor vehicles	13.52	13.65	0.98	*	1636	1649
30-32	Manufact. of electrical equipment, precision machinery	13.80	13.81	0.52	-	863	916
33	Manufacture of medical and precision instruments	12.99	13.16	0.95	-	1212	1334
36,37	Manufacturing N.E.C.	13.51	13.50	0.54	-	996	918
40,41	Electricity, gas and water supply	11.21	12.13	1.00	**	183	203
45	Construction	14.13	14.75	1.00	**	2191	2391
50-93	SECTOR 3 SERVICES	12.62	12.61	0.56	-	27595	27085
50-52	Sale, repair	12.23	12.37	0.99	*	7025	6405
50	Sale, repair of motor vehicles	12.57	12.72	0.94	-	1323	1087
51	Brokerage, wholesale trade	12.44	12.53	0.97	-	3244	2878
52	Retail trade, repair of pers. & household goods	11.95	12.15	0.97	-	2458	2440
55	Hotels and restaurants	11.60	11.57	0.64	-	2419	2147
60-64	Transport, storage and communication	13.37	13.51	0.99	*	2250	2478
60	Land transport/pipelines	14.25	13.96	1.00	**	870	1053
61	Water transport	12.67	13.75	1.00	**	44	55
62	Air transport	12.26	14.17	1.00	**	72	84
63	Supporting and auxiliary transport activities	12.54	12.95	0.99	*	1063	1076
64	Post and telecommunications	13.12	13.25	0.94	-	201	210
65-67	Banking ; insurance	12.55	12.45	0.79	-	1993	2472
65	Banking	12.97	12.44	1.00	**	997	971
66	Insurance	11.81	12.62	1.00	**	429	207
67	Activities relating to banking/insurance	11.88	12.05	0.92	-	567	1294
70-74	Real estate, computer, research & development	12.29	12.26	0.66	-	6987	5723
70,71	Real estate activities/renting of machinery & equipment	12.23	12.39	0.92	-	1213	1267
72,74	Computer and related activities; other business activities	12.28	12.23	0.75	-	5595	4249
73	Research and development	12.56	12.66	0.80	-	179	208
80	Education	13.24	13.12	0.95	-	1245	1394
85	Health and social work	13.45	13.19	1.00	**	2489	2931
90-93	Other community, social and personal service activities	12.37	12.36	0.55	-	3186	3534
90	Waste processing and other disposal	13.53	13.86	0.98	*	227	217
91	Activities of membership organizations n.e.c.	12.82	12.60	1.00	**	1126	1096
92	Recreational, cultural and sporting activities	12.12	12.27	0.96	-	1072	1220
93	Other service activities	11.92	11.70	0.96	-	761	1001

[1] Legend: ** = significant (risk=1%) * = significant (risk=5%) - = non significant

Table A4: Correlation matrices for components 2 to 5 (total, horticulture and secondary sector), 2002

1
0
1.00 0.14 0.00 0.02
0.14 1.00 0.18 -0.21
0.00 0.18 1.00 -0.74
0.02 -0.21 -0.74 1.00
2
1
1.00 0.16 0.01 0.01
0.16 1.00 0.28 -0.14
0.01 0.28 1.00 -0.18
0.01 -0.14 -0.18 1.00
3
1045
1.00 0.05 -0.04 -0.07
0.05 1.00 0.09 -0.05
-0.04 0.09 1.00 0.09
-0.07 -0.05 0.09 1.00
4
1014
1.00 0.08 -0.04 -0.08
0.08 1.00 -0.03 0.07
-0.04 -0.03 1.00 0.08
-0.08 0.07 0.08 1.00
5
1537
1.00 0.00 -0.10 -0.16
0.00 1.00 0.09 -0.16
-0.10 0.09 1.00 0.08
-0.16 -0.16 0.08 1.00
6
15
1.00 0.21 -0.06 -0.04
0.21 1.00 0.35 0.01
-0.06 0.35 1.00 -0.30
-0.04 0.01 -0.30 1.00
7
16
1.00 0.47 -0.04 -0.11
0.47 1.00 -0.36 -0.37
-0.04 -0.36 1.00 0.34
-0.11 -0.37 0.34 1.00
8
17
1.00 0.12 -0.04 -0.02
0.12 1.00 -0.09 0.11
-0.04 -0.09 1.00 -0.70
-0.02 0.11 -0.70 1.00
9
18
1.00 0.00 -0.11 0.00
0.00 1.00 0.35 0.35
-0.11 0.35 1.00 0.08
0.00 0.35 0.08 1.00
10
19
1.00 0.09 0.16 0.01
0.09 1.00 0.20 -0.04
0.16 0.20 1.00 -0.23
0.01 -0.04 -0.23 1.00
11
20
1.00 0.11 0.09 -0.08
0.11 1.00 0.15 -0.11
0.09 0.15 1.00 -0.36
-0.08 -0.11 -0.36 1.00
12
21
1.00 -0.05 0.11 0.10
-0.05 1.00 0.31 -0.20
0.11 0.31 1.00 -0.12
0.10 -0.20 -0.12 1.00
13
22
1.00 0.40 -0.04 -0.19
0.40 1.00 -0.09 -0.10
-0.04 -0.09 1.00 0.23
-0.19 -0.10 0.23 1.00
14
23
1.00 0.18 -0.17 -0.22
0.18 1.00 -0.09 -0.47
-0.17 -0.09 1.00 -0.37
-0.22 -0.47 -0.37 1.00
15
25
1.00 -0.17 -0.03 0.02
-0.17 1.00 0.03 0.00
-0.03 0.03 1.00 -0.03
0.02 0.00 -0.03 1.00
16
26
1.00 0.25 -0.08 0.24
0.25 1.00 -0.14 0.46
-0.08 -0.14 1.00 -0.15
0.24 0.46 -0.15 1.00
17
27
1.00 0.04 0.02 -0.19
0.04 1.00 0.05 -0.16
0.02 0.05 1.00 0.14
-0.19 -0.16 0.14 1.00
18
29
1.00 -0.09 -0.07 -0.05
-0.09 1.00 0.13 0.01
-0.07 0.13 1.00 0.11
-0.05 0.01 0.11 1.00
19
30
1.00 -0.10 0.30 -0.33
-0.10 1.00 -0.22 -0.40
0.30 -0.22 1.00 0.26
-0.33 -0.40 0.26 1.00
20
33
1.00 -0.07 0.09 -0.09
-0.07 1.00 0.37 0.07
0.09 0.37 1.00 0.36
-0.09 0.07 0.36 1.00
21
36
1.00 0.07 -0.35 0.71
0.07 1.00 0.19 -0.03
-0.35 0.19 1.00 -0.48
0.71 -0.03 -0.48 1.00
22
40
1.00 0.39 0.10 0.26
0.39 1.00 -0.05 0.07
0.10 -0.05 1.00 0.44
0.26 0.07 0.44 1.00
23
45
1.00 0.15 0.06 0.07
0.15 1.00 -0.08 0.11
0.06 -0.08 1.00 -0.02
0.07 0.11 -0.02 1.00

Explanations :
Matrices are represented on 3 columns.

By column :
1st line : sequential number

2nd line : NOGA2 class number (if several classes, the first is given, see Table A3 for exact denominations)

3rd à 6th lines : correlation matrices of (x_1, x_2, x_3, x_4).

Table A4 (continued): Correlation matrices for components 2 to 5 (third sector), 2002

24
5093
1.00 0.12 -0.16 0.20
0.12 1.00 0.06 -0.15
-0.16 0.06 1.00 -0.74
0.20 -0.15 -0.74 1.00
25
5052
1.00 0.08 -0.18 -0.21
0.08 1.00 -0.42 -0.16
-0.18 -0.42 1.00 0.02
-0.21 -0.16 0.02 1.00
26
50
1.00 0.10 -0.03 0.03
0.10 1.00 -0.20 0.12
-0.03 -0.20 1.00 -0.03
0.03 0.12 -0.03 1.00
27
51
1.00 0.11 -0.02 -0.09
0.11 1.00 0.13 -0.06
-0.02 0.13 1.00 0.12
-0.09 -0.06 0.12 1.00
28
52
1.00 0.09 -0.22 -0.45
0.09 1.00 -0.53 -0.32
-0.22 -0.53 1.00 0.24
-0.45 -0.32 0.24 1.00
29
55
1.00 0.04 -0.01 0.00
0.04 1.00 0.12 0.00
-0.01 0.12 1.00 0.04
0.00 0.00 0.04 1.00
30
6064
1.00 0.01 0.27 -0.06
0.01 1.00 0.20 -0.14
0.27 0.20 1.00 -0.03
-0.06 -0.14 -0.03 1.00
31
60
1.00 -0.09 -0.20 -0.02
-0.09 1.00 0.20 -0.08
-0.20 0.20 1.00 -0.02
-0.02 -0.08 -0.02 1.00
32
61
1.00 -0.27 -0.23 0.93
-0.27 1.00 0.48 -0.23
-0.23 0.48 1.00 -0.10
0.93 -0.23 -0.10 1.00

33
62
1.00 0.23 0.02 -0.10
0.23 1.00 -0.06 0.17
0.02 -0.06 1.00 -0.26
-0.10 0.17 -0.26 1.00
34
63
1.00 0.35 0.35 0.24
0.35 1.00 0.67 0.15
0.35 0.67 1.00 0.66
0.24 0.15 0.66 1.00
35
64
1.00 -0.26 0.52 -0.33
-0.26 1.00 -0.30 -0.33
0.52 -0.30 1.00 -0.46
-0.33 -0.33 -0.46 1.00
36
6567
1.00 0.53 -0.43 0.55
0.53 1.00 -0.77 0.70
-0.43 -0.77 1.00 -0.67
0.55 0.70 -0.67 1.00
37
65
1.00 0.65 -0.62 0.69
0.65 1.00 -0.86 0.64
-0.62 -0.86 1.00 -0.58
0.69 0.64 -0.58 1.00
38
66
1.00 0.31 0.47 -0.20
0.31 1.00 0.24 -0.19
0.47 0.24 1.00 -0.54
-0.20 -0.19 -0.54 1.00
39
67
1.00 -0.15 0.49 -0.19
-0.15 1.00 -0.18 0.61
0.49 -0.18 1.00 -0.31
-0.19 0.61 -0.31 1.00
40
7074
1.00 0.17 0.05 0.03
0.17 1.00 0.11 -0.08
0.05 0.11 1.00 0.08
0.03 -0.08 0.08 1.00
41
70
1.00 -0.13 -0.07 -0.09
-0.13 1.00 0.20 -0.10
-0.07 0.20 1.00 0.26
-0.09 -0.10 0.26 1.00

42
72
1.00 0.17 0.05 0.03
0.17 1.00 0.11 -0.08
0.05 0.11 1.00 0.07
0.03 -0.08 0.07 1.00
43
73
1.00 0.41 0.43 -0.22
0.41 1.00 -0.01 -0.59
0.43 -0.01 1.00 0.30
-0.22 -0.59 0.30 1.00
44
80
1.00 -0.05 0.11 0.04
-0.05 1.00 -0.11 -0.06
0.11 -0.11 1.00 0.19
0.04 -0.06 0.19 1.00
45
85
1.00 -0.01 0.12 -0.16
-0.01 1.00 0.37 -0.17
0.12 0.37 1.00 -0.54
-0.16 -0.17 -0.54 1.00
46
9093
1.00 -0.28 -0.19 0.15
-0.28 1.00 0.63 -0.11
-0.19 0.63 1.00 -0.12
0.15 -0.11 -0.12 1.00
47
90
1.00 -0.05 -0.26 0.02
-0.05 1.00 -0.18 0.30
-0.26 -0.18 1.00 -0.53
0.02 0.30 -0.53 1.00
48
91
1.00 0.24 -0.05 -0.04
0.24 1.00 0.25 -0.14
-0.05 0.25 1.00 0.06
-0.04 -0.14 0.06 1.00
49
92
1.00 -0.39 -0.26 0.2
-0.39 1.00 0.67 -0.3
-0.26 0.67 1.00 -0.2
0.20 -0.30 -0.20 1.0
50
93
1.00 0.26 -0.11 0.23
0.26 1.00 -0.07 0.37
-0.11 -0.07 1.00 -0.20
0.23 0.37 -0.20 1.00

Table A4 (continued): Correlation matrices for components 2 to 5 (total, horticulture and secondary sector), 2004

1
0
1.00 0.14 0.26 -0.14
0.14 1.00 0.35 -0.28
0.26 0.35 1.00 -0.82
-0.14 -0.28 -0.82 1.00
2
1
1.00 0.05 -0.47 0.09
0.05 1.00 -0.03 -0.02
-0.47 -0.03 1.00 -0.01
0.09 -0.02 -0.01 1.00
3
1045
1.00 -0.02 -0.01 -0.03
-0.02 1.00 0.08 -0.10
-0.01 0.08 1.00 0.03
-0.03 -0.10 0.03 1.00
4
1014
1.00 0.57 0.13 0.02
0.57 1.00 0.06 0.03
0.13 0.06 1.00 -0.02
0.02 0.03 -0.02 1.00
5
1537
1.00 -0.08 -0.04 -0.11
-0.08 1.00 0.05 -0.20
-0.04 0.05 1.00 0.02
-0.11 -0.20 0.02 1.00
6
15
1.00 -0.14 -0.23 0.08
-0.14 1.00 0.36 -0.14
-0.23 0.36 1.00 -0.26
0.08 -0.14 -0.26 1.00
7
16
1.00 0.13 0.04 0.03
0.13 1.00 0.14 -0.34
0.04 0.14 1.00 -0.29
0.03 -0.34 -0.29 1.00
8
17
1.00 -0.09 0.03 -0.04
-0.09 1.00 -0.31 0.30
0.03 -0.31 1.00 -0.65
-0.04 0.30 -0.65 1.00
9
18
1.00 -0.47 0.73 -0.32
-0.47 1.00 -0.26 0.33
0.73 -0.26 1.00 -0.07
-0.32 0.33 -0.07 1.00
10
19
1.00 -0.02 -0.87 -0.10
-0.02 1.00 0.00 0.00
-0.87 0.00 1.00 0.14
-0.10 0.00 0.14 1.00
11
20
1.00 0.11 0.02 0.03
0.11 1.00 0.30 -0.02
0.02 0.30 1.00 -0.03
0.03 -0.02 -0.03 1.00
12
21
1.00 0.20 0.02 -0.10
0.20 1.00 0.10 -0.16
0.02 0.10 1.00 0.01
-0.10 -0.16 0.01 1.00
13
22
1.00 0.54 -0.10 -0.15
0.54 1.00 0.07 -0.11
-0.10 0.07 1.00 -0.17
-0.15 -0.11 -0.17 1.00
14
23
1.00 0.29 -0.11 -0.35
0.29 1.00 -0.04 -0.51
-0.11 -0.04 1.00 -0.46
-0.35 -0.51 -0.46 1.00
15
25
1.00 0.09 -0.06 -0.11
0.09 1.00 -0.13 -0.02
-0.06 -0.13 1.00 -0.34
-0.11 -0.02 -0.34 1.00
16
26
1.00 0.42 -0.07 -0.05
0.42 1.00 -0.04 0.07
-0.07 -0.04 1.00 -0.11
-0.05 0.07 -0.11 1.00
17
27
1.00 -0.09 0.01 -0.07
-0.09 1.00 0.15 0.08
0.01 0.15 1.00 -0.08
-0.07 0.08 -0.08 1.00
18
29
1.00 0.03 -0.03 0.09
0.03 1.00 -0.08 0.10
-0.03 -0.08 1.00 0.04
0.09 0.10 0.04 1.00
19
30
1.00 -0.32 0.50 -0.19
-0.32 1.00 -0.43 -0.34
0.50 -0.43 1.00 0.41
-0.19 -0.34 0.41 1.00
20
33
1.00 -0.14 -0.06 0.08
-0.14 1.00 0.54 0.00
-0.06 0.54 1.00 0.18
0.08 0.00 0.18 1.00
21
36
1.00 0.02 -0.11 0.06
0.02 1.00 0.19 0.29
-0.11 0.19 1.00 0.02
0.06 0.29 0.02 1.00
22
40
1.00 -0.14 -0.57 0.63
-0.14 1.00 0.15 0.01
-0.57 0.15 1.00 -0.67
0.63 0.01 -0.67 1.00
23
45
1.00 0.14 -0.01 0.06
0.14 1.00 0.01 0.10
-0.01 0.01 1.00 -0.15
0.06 0.10 -0.15 1.00

Table A4 (continued): Correlation matrices for components 2 to 5 (third sector), 2004

24
5093
1.00 0.35 0.30 -0.13
0.35 1.00 0.35 -0.24
0.30 0.35 1.00 -0.83
-0.13 -0.24 -0.83 1.00
25
5052
1.00 0.90 0.80 -0.39
0.90 1.00 0.77 -0.45
0.80 0.77 1.00 -0.51
-0.39 -0.45 -0.51 1.00
26
50
1.00 -0.01 0.04 0.25
-0.01 1.00 -0.10 -0.04
0.04 -0.10 1.00 -0.04
0.25 -0.04 -0.04 1.00
27
51
1.00 0.04 0.00 -0.03
0.04 1.00 0.21 -0.12
0.00 0.21 1.00 -0.08
-0.03 -0.12 -0.08 1.00
28
52
1.00 0.90 0.84 0.22
0.90 1.00 0.81 0.13
0.84 0.81 1.00 -0.17
0.22 0.13 -0.17 1.00
29
55
1.00 0.00 -0.02 0.45
0.00 1.00 -0.03 0.01
-0.02 -0.03 1.00 0.10
0.45 0.01 0.10 1.00
30
6064
1.00 0.03 0.28 -0.43
0.03 1.00 -0.11 -0.03
0.28 -0.11 1.00 -0.62
-0.43 -0.03 -0.62 1.00
31
60
1.00 -0.1 0.03 0.02
-0.10 1.0 0.20 -0.10
0.03 0.2 1.00 -0.29
0.02 -0.1 -0.29 1.00
32
61
1.00 0.02 0.04 -0.18
0.02 1.00 0.38 -0.01
0.04 0.38 1.00 -0.01
-0.18 -0.01 -0.01 1.00

33
62
1.00 0.27 -0.05 -0.04
0.27 1.00 -0.09 -0.06
-0.05 -0.09 1.00 0.08
-0.04 -0.06 0.08 1.00
34
63
1.00 0.26 0.06 -0.18
0.26 1.00 -0.11 -0.19
0.06 -0.11 1.00 0.10
-0.18 -0.19 0.10 1.00
35
64
1.00 0.23 0.53 -0.86
0.23 1.00 -0.33 0.11
0.53 -0.33 1.00 -0.54
-0.86 0.11 -0.54 1.00
36
6567
1.00 0.53 -0.43 0.42
0.53 1.00 -0.69 0.68
-0.43 -0.69 1.00 -0.69
0.42 0.68 -0.69 1.00
37
65
1.00 0.46 -0.54 0.48
0.46 1.00 -0.63 0.66
-0.54 -0.63 1.00 -0.63
0.48 0.66 -0.63 1.00
38
66
1.00 0.31 0.53 -0.35
0.31 1.00 0.40 -0.35
0.53 0.40 1.00 -0.45
-0.35 -0.35 -0.45 1.00
39
67
1.00 -0.01 0.34 -0.14
-0.01 1.00 0.03 -0.05
0.34 0.03 1.00 -0.58
-0.14 -0.05 -0.58 1.00
40
7074
1.00 0.02 -0.06 -0.04
0.02 1.00 0.13 -0.07
-0.06 0.13 1.00 0.02
-0.04 -0.07 0.02 1.00
41
70
1.00 0.12 -0.32 0.03
0.12 1.00 -0.20 0.05
-0.32 -0.20 1.00 -0.49
0.03 0.05 -0.49 1.00

42
72
1.00 0.02 -0.05 -0.04
0.02 1.00 0.13 -0.07
-0.05 0.13 1.00 0.04
-0.04 -0.07 0.04 1.00
43
73
1.00 -0.01 0.03 -0.01
-0.01 1.00 0.12 -0.28
0.03 0.12 1.00 0.24
-0.01 -0.28 0.24 1.00
44
80
1.00 0.00 0.05 -0.07
0.00 1.00 0.08 -0.03
0.05 0.08 1.00 -0.01
-0.07 -0.03 -0.01 1.00
45
85
1.00 0.05 0.01 -0.01
0.05 1.00 0.09 0.06
0.01 0.09 1.00 -0.07
-0.01 0.06 -0.07 1.00
46
9093
1.00 -0.44 -0.31 -0.18
-0.44 1.00 0.68 0.35
-0.31 0.68 1.00 0.31
-0.18 0.35 0.31 1.00
47
90
1.00 -0.20 0.07 0.21
-0.20 1.00 -0.33 0.45
0.07 -0.33 1.00 -0.47
0.21 0.45 -0.47 1.00
48
91
1.00 0.41 -0.48 -0.17
0.41 1.00 -0.45 -0.09
-0.48 -0.45 1.00 0.15
-0.17 -0.09 0.15 1.00
49
92
1.00 -0.54 -0.32 -0.10
-0.54 1.00 0.70 0.07
-0.32 0.70 1.00 0.13
-0.10 0.07 0.13 1.00
50
93
1.00 0.23 0.08 0.14
0.23 1.00 0.21 -0.05
0.08 0.21 1.00 0.12
0.14 -0.05 0.12 1.00

Table A5: Multivariate tests of change in the 5-dimensional compositional vectors between years 2002 and 2004

Noga2	Economic activities	2004 geometric mean (%)	2002 geometric mean (%)	Mahalanobis distance	p-value	signi-ficance
0	TOTAL	3.29	3.15	13.44	0.99	*
01	Horticulture	1.57	1.76	29.44	1.00	**
10-45	SECTOR 2 PRODUCTION	3.63	3.52	15.15	1.00	**
10-14	Mining and quarrying of stone	2.44	2.40	11.24	0.98	*
15-37	Manufacturing	4.03	3.86	20.92	1.00	**
15	Manufacture of food products and beverages	3.52	3.63	27.91	1.00	**
16	Manufacture of tobacco products	4.61	3.83	8133.53	1.00	**
17	Manufacture of textiles	3.45	3.75	29.47	1.00	**
18	Manufact. of wearing apparel; dressing and dyeing of fur	1.97	1.25	37.40	1.00	**
19	Manufacture of leather and leather products	2.33	0.86	185.38	1.00	**
20	Manufacture of wood and wood products	2.54	2.51	4.79	0.69	-
21	Manufacture of pulp, paper and paper products	5.23	4.84	21.73	1.00	**
22	Publishing, printing, reproduction	3.33	3.31	3.54	0.53	-
23,24	Manufacture of coke, chemicals	4.26	4.09	6.09	0.81	-
25	Manufacture of rubber and plastic products	4.33	4.24	6.47	0.83	-
26	Manufacture of other non-metallic mineral products	3.44	3.48	3.23	0.48	-
27,28	Manufacture of basic metals	4.10	3.54	17.56	1.00	**
29,34,35	Manufact. of machinery & eq. N.E.C., -of motor vehicles	3.51	3.64	10.74	0.97	-
30-32	Manufact. of electrical equipment, precision machinery	4.11	3.90	39.11	1.00	**
33	Manufacture of medical and precision instruments	3.65	3.55	7.68	0.90	-
36,37	Manufacturing N.E.C.	2.78	2.76	8.29	0.92	-
40,41	Electricity, gas and water supply	3.14	3.38	10.07	0.96	-
45	Construction	2.24	2.24	2.68	0.39	-
50-93	SECTOR 3 SERVICES	3.01	2.80	14.34	0.99	*
50-52	Sale, repair	2.74	2.23	12.24	0.98	*
50	Sale, repair of motor vehicles	1.66	1.94	29.40	1.00	**
51	Brokerage, wholesale trade	2.44	2.45	8.87	0.94	-
52	Retail trade, repair of pers. & household goods	2.78	1.97	26.09	1.00	**
55	Hotels and restaurants	1.39	1.44	10.35	0.97	-
60-64	Transport, storage and communication	3.36	2.80	133.50	1.00	**
60	Land transport/pipelines	2.37	2.43	51.35	1.00	**
61	Water transport	2.35	2.62	68.71	1.00	**
62	Air transport	1.25	3.00	1209.81	1.00	**
63	Supporting and auxiliary transport activities	3.23	2.89	12.16	0.98	*
64	Post and telecommunications	4.11	2.96	296.38	1.00	**
65-67	Banking; insurance	2.30	2.42	13.41	0.99	*
65	Banking	2.43	2.58	38.04	1.00	**
66	Insurance	1.93	1.62	15.76	1.00	**
67	Activities relating to banking/insurance	1.76	3.03	37.83	1.00	**
70-74	Real estate, computer, research & development	2.59	2.62	14.18	0.99	*
70,71	Real estate activities/renting of machinery & equipment	2.16	2.05	4.39	0.64	-
72,74	Computer and related activities; other business activities	2.63	2.67	13.67	0.99	*
73	Research and development	2.02	2.00	9.43	0.95	-
80	Education	1.50	1.46	0.91	0.08	-
85	Health and social work	2.40	2.70	55.64	1.00	**
90-93	Other community, social and personal service activities	2.31	2.13	33.36	1.00	**
90	Waste processing and other disposal	3.48	2.69	39.16	1.00	**
91	Activities of membership organizations n.e.c.	1.40	1.72	65.34	1.00	**
92	Recreational, cultural and sporting activities	2.78	2.41	37.97	1.00	**
93	Other service activities	1.31	1.50	18.93	1.00	**

References

[1] Aitchison, J. (1986). *The Statistical Analysis of Compositional Data.* Chapman and Hall, Monographs on Statistics and Probability.
Reedited by The Blackburn Press, 2003.

[2] Aitchison, J. (1997). The one-hour course in compositional data analysis or compositional data is simple. *Pawlowsky-Glahn, V., Cimne, eds., Proceedings of Int. Assoc. of Mathematical Geology IAMG'97*, Part **I**, pp.3-35.

[3] Aitchison, J. (2001). Simplicial Inference. *Comtemporary Mathematics* **287**, AMS.

[4] Aitchison, J. and Greenacre, M.(2002). Biplots of Compositional Data. *Appl. Statist.*, **51**, Part 4, 375-392.

[5] Anyadike-Danes, M. (2003). The allometry of non-employment. What can compositional data analysis tell us about labour market performance across the UK's regions?, *Thió-Henestrosa, S. and Martín-Fernández, JA (Eds.), Proceedings of CODAWORK'03.*

[6] Brunsdon, T. M., and Smith, T. M. F. (1998). The time series analysis of compositional data, *Journal of Official Statistics*, **14**, 237-253.

[7] Egozcue, J., Pawlowsky-Glahn, V., Mateu-Figueras, G. and Barleó-Vidal, C. (2003). Isometric Logratio Transformations for Compositional Data Analysis. *Mathematical Geology*, **35**, 3, 279-300.

[8] Eurostat (2002). *Variance estimation methods in the European Union.* Monographs in official statistics. ISSN 17-25 15-67.

[9] Eurostat (2003). *Standard Quality Indicators, Producer-Oriented*, Doc. Eurostat/A4/Quality/03/General/Standard Indicators, Eurostat Working Group "Quality in Statistics" 2003.

[10] Larrosa, J. M. (2003). A compositional statistical analysis of capital stock. *Thió-Henestrosa, S. and Martín-Fernández, JA (Eds.), Proceedings of CODAWORK'03.*

[11] Graf, M. (2002a). Enquête suisse sur la structure des salaires 2000. Plan d'échantillonnage, pondération et méthode d'estimation pour le secteur privé. *Rapport de méthodes 338-0010*, Office fédéral de la statistique.

[12] Graf, M. (2002b). Assessing the Accuracy of the Median in a Stratified Double Stage Cluster Sample by means of a Nonparametric Confidence Interval: Application to the Swiss Earnings Structure Survey. *Proceedings of the Joint Statistical Meeting 2002.*

[13] Graf, M. (2004). Enquête suisse sur la structure des salaires 2002. Plan d'échantillonnage et extrapolation pour le secteur privé. *Rapport de méthodes 338-0025*, Office fédéral de la statistique, Neuchâtel.

[14] Graf, M. (2005). Assessing the Precision of Compositional Data in a Stratified Double Stage Cluster Sample: Application to the Swiss Earnings Structure Survey. *Thió-Henestrosa, S., Pawlowski-Glahn V. and Martín-Fernández, JA (Eds.), Proceedings of CODAWORK'05.*

[15] Särndal, C.-E., Swensson, B. & Wretman, J. (1992). *Model Assisted Survey Sampling.* Springer Series in Statistics.

[16] SESS (2003). Enquête suisse sur la structure des salaires 2002 (ESS 2002). *Actualités OFS,* **3** Vie active et rémunération du travail.

[17] SESS (2004). L'enquête suisse sur la structure des salaires 2002. Résultats commentés et tableaux. *Statistique de la Suisse*, OFS.

[18] Silva, D. B. N., and Smith, T. M. F. (2001). Modelling compositional time series from repeated surveys, *Survey Methodology*, **27** (2), 205-215.

[19] von Eynatten, H., Pawlovski-Glahn, V. & Egozcue, J.J. (2002). Understanding Perturbation on the Simplex: A Simple Method to Better Visualize and Interpret Compositional Data in Ternary Diagrams. *Mathematical Geology*, **34** (3), 249-257.

Methodenberichte des Dienstes Statistische Methoden des BFS
Rapports de méthodes du Service de méthodes statistiques de l'OFS
Methodology reports published by the SFSO's Statistical Methods Unit

Graf, M. (2006). Swiss Earnings Structure Survey 2002-2004. Compositional data in a stratified two-stage sample: Analysis and precision assessment of wage components. Order number: 338-0038

Potterat, J. (2006). Pensionskassenstatistik 2004. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. Bestellnummer: 338-0037

Potterat, J. (2006). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2004. Stichprobenplan, Gewichtung und Schätzverfahren. Bestellnummer: 338-0036

Kilchmann, D. (2006). Vierteljährliche Wohnbaustatistik. Stichprobenplan, statistische Datenaufarbeitung und Schätzverfahren 2005. Bestellnummer: 338-0035

Kilchmann, D. (2006). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2004. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Bestellnummer: 338-0034

Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 2. Bestellnummer: 338-0033

Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 1. Bestellnummer: 338-0032

Graf, M., Matei, A. (2005). Enquête suisse sur la structure des salaires 2002. La précision du salaire brut standardisé médian. Numéro de commande: 338-0031

Graf, E., Renfer, J.-P. (2005). Enquête suisse sur la santé 2002. Plan d'échantillonnage, pondération et estimation de la précision. Numéro de commande: 338-0030

Potterat, J. (2005). Mietpreis-Strukturerhebung 2003. Gewichtung und Schätzverfahren. Bestellnummer: 338-0029

Potterat, J. (2005). Landwirtschaftliche Betriebszählung 2003. Schätzverfahren für die Zusatzerhebung. Bestellnummer: 338-0028

Renaud, A. (2004). Coverage estimation for the Swiss population census 2000. Estimation methodology and results. Order number: 338-0027

Kilchmann, D. (2004). Revision des Schweizerischen Lohnindex. Schätzmethoden der Lohnindices und deren Varianzschätzer. Bestellnummer: 338-0026

Graf, M. (2004). Enquête suisse sur la structure des salaires 2002. Plan d'échantillonnage et extrapolation pour le secteur privé. Numéro de commande: 338-0025

Renaud, A. (2004). Analyse de données d'enquêtes. Quelques méthodes et illustration avec des données de l'OFS. Numéro de commande 338-0024

Renaud, A., Potterat, J. (2004). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sous-couverture (P-sample) et qualité du cadre de sondage des bâtiments. Numéro de commande: 338-0023

Graf, M. (2004). Fusion de données. Etude de faisabilité. Numéro de commande: 338-0022

Potterat, J. (2003). Mietpreis-Strukturerhebung 2003. Entwicklung des Stichprobenplans und Ziehung der Stichprobe. Bestellnummer: 338-0021

Potterat, J. (2003). Landwirtschaftliche Betriebszählung 2003. Stichprobenplan der Zusatzerhebung. Bestellnummer: 338-0020

Renaud, A. (2003). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sur-couverture (E-sample). Numéro de commande: 338-0019

Hulliger, B. (2003). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2000. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Bestellnummer: 338-0018

Renfer, J.-P. (2003). Enquête 2000 sur la recherche et le développement dans l'économie privée en Suisse. Plan d'échantillonnage. Numéro de commande: 338-0017

Potterat, J. (2003). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe. Schätzverfahren. Bestellnummer: 338-0016

Graf, M., Matei, A. (2003). Stratégie de choix des modèles de désaisonnalisation. Application aux séries de l'emploi total. Numéro de commande : 338-0015

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Methoden für die Datenbereinigung. Bestellnummer: 338-0014

Renaud, A. (2002). Programme international pour le suivi des acquis des élèves (PISA). Plans d'échantillonnage pour PISA 2000 en Suisse. Numéro de commande: 338-0013

Renfer, J.-P. (2002). Enquête 2001 sur les coûts et l'utilité de la formation des apprentis du point de vue des établissements. Plan d'échantillonnage. Numéro de commande: 338-0012

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Stichprobenplan und Schätzverfahren für die provisorischen Ergebnisse. Bestellnummer: 338-0011

Graf, M. (2002). Enquête suisse sur la structure des salaires 2000. Plan d'échantillonnage, pondération et méthode d'estimation pour le secteur privé. Numéro de commande: 338-0010

Renaud, A., Eichenberger P. (2002). Estimation de la couverture du recensement de la population de l'an 2000. Procédure d'enquête et plan d'échantillonnage de l'enquête de couverture. Numéro de commande: 338-0009

Kilchmann, D., Hulliger, B. (2002). Stichprobenplan für die Obstbaumzählung 2001. Bestellnummer: 338-0008

Graf, M. (2002). Passage du concept établissement au concept entreprise. Numéro de commande: 338-0007

Salamin, P.A. (2001). La technique de la double enquête pour la statistique du transport routier de marchandise. Numéro de commande: 338-0006

Peters, R., Renfer, J.-P. et Hulliger, B. (2001). Statistique de la valeur ajoutée 1997-1998. Procédure d'extrapolation des données. Numéro de commande: 338-0005

Potterat, J., Hulliger, B. (2001). Schätzung der Sägereiproduktion mit der Sägerei-Erhebung PAUL. Bestellnummer: 338-0004

Graf, M. (2001). Désaisonnalisation. Aspects méthodologiques et application à la statistique de l'emploi. Numéro de commande: 338-0003

Hüsler, J., Müller, S. (2001). Schlussbericht Betriebszählung 1995 (BZ 95), Mehrfach imputierte Umsatzzahlen. Bestellnummer: 338-0002

Renaud, A. (2001). Statistique suisse des bénéficiaires de l'aide sociale. Plan d'échantillonnage des communes. Numéro de commande: 338-0001

Hulliger, B., Eichenberger, P. (2000). Stichprobenregister für Haushalterhebungen: Umstellung auf Telefonnummern ohne Namen und Adressen, Abläufe für Erstellung und Stichprobenziehung. Bestellnummer: 338-0000

de Rossi, F.-X. (1998). Méthodes statistiques pour le compte routier suisse.

Hulliger, B., Kassab, M. (1998). Evaluation of Estimation Methods for the Survey on Environment Protection Expenditures of Swiss Communes.

Salamin, P.A. (1998). Etablissement d'une clef de passage pondérée entre l'ancienne (NGAE 85) et la nouvelle nomenclature (NOGA 95) générale des activités économiques.

Peters, R. (1998). Extrapolation des données de l'enquête de structure sur les loyers.

Bender, A., Hulliger, B. (1997). Enquête suisse sur la population active: rapport de pondération pour 1996.

Salamin, P.A. (1997). Evaluation de la Statistique de l'emploi.

Peters, R. (1997). Etablissement du plan d'échantillonnage pour l'enquête 1996 sur la recherche et le développement dans l'économie privée en Suisse.

Peters, R. (1997). Enquête 1996 sur la structure des salaires en Suisse: établissement du plan d'échantillonnage.

Peters, R. (1996). Pondération des données de l'enquête sur la famille en Suisse.

Comment, T., Hulliger, B., Ries, A. (1996). Gewichtungsverfahren für die Schweizerische Arbeitskräfteerhebung (1991-1995).

Hulliger, B. (1996). Haushalterhebung Familie 1994: Stichprobenplan, Stichprobenziehung und Reservestichproben.

Peters, R., Hulliger, B. (1996). Schätzverfahren für die Lohnstruktur-Erhebung 1994 / Procédure d'estimation pour l'enquête de 1994 sur la structure des salaires.

Peters, R. (1996). Schéma de pondération des indices PAUL.

Hulliger, B., Peters, R. (1996). Enquête sur le comportement de la population suisse en matière de transport en 1994: plan d'échantillonnage et pondération.

Hulliger, B. (1996). Gütertransportstatistik 1993: Schätzverfahren mit Kompensation der Antwortausfälle.

Salamin, P.A. (1995). Estimation des flux pour le module II des comptes globaux du marché de travail.

Peters, R. (1995). Enquête de structure sur les loyers: établissement d'un plan d'échantillonnage stratifié.

Hulliger, B. (1995). Konjunkturelle Mietpreiserhebung: Stichprobenplan und Schätzverfahren.

Schwendener, P. (1995). Verbrauchserhebung 1990 - Vertrauensintervalle.

Peters, R., Hulliger, B. (1994). La technique de pondération des données: application à l'enquête suisse sur la santé.

Hulliger, B., Peters, R. (1994). Enquête sur la structure des salaires en Suisse: stratégie d'échantillonnage pour le secteur privé.

Publikationsprogramm BFS

Das Bundesamt für Statistik (BFS) hat – als zentrale Statistikstelle des Bundes – die Aufgabe, statiastische Informationen breiten Benutzerkreisen zur Verfügung zu stellen.

Die Verbreitung der statistischen Information geschieht gegliedert nach Fachbereichen (vgl. Umschlagseite 2) und mit verschiedenen Mitteln

SFSO Publications

The Swiss Federal Statistical Office (SFSO) is the central and official purveyor of statistical information to the Swiss Government. It is officially mandated to supply this information to a wide range of users.

This statistical data is organized and disseminated on the basis of a subject-matter classification (see inside cover page).

Diffusionsmittel	*Kontakt Phone*	*Distribution medium*
Individuelle Auskünfte	032 713 60 11 info@bfs.admin.ch	Individual information
Das BFS im Internet	www.statistik.admin.ch	The SFSO on the Internet
Medienmitteilungen zur raschen Information der Öffentlichkeit über die neusten Ergebnisse	www.news-stat.admin.ch	Press releases: fast access to the latest results
Publikationen zur vertieften Information (zum Teil auch als Diskette/CD-Rom)	032 713 60 60 order@bfs.admin.ch	Purchase of printed and electronic publications and CD-Rom
Online-Datenbank	032 713 60 86 www.statweb.admin.ch	Online database

Nähere Angaben zu den verschiedenen Diffusionsmitteln liefert das laufend nachgeführte Publikationsverzeichnis im Internet unter der Adresse www.statistik.admin.ch→Aktuell→Publikationen.

The SFSO's detailed online publications catalogue is updated daily (www.statistics.admin.ch→News→Publications).

Methodenberichte des Dienstes Statistische Methoden
Rapports de méthodes du Service de méthodes statistiques
Methodology Reports by the Statistical Methods Unit

Die Methodenberichte beschreiben die mathematischen und statistischen Methoden, die den Resultaten und Analysen der öffentlichen Statistik zu Grunde liegen. Sie enthalten ausserdem die Evaluation und Entwicklung von neuen Methoden im Hinblick auf eine zukünftige Anwendung. Diese Publikationen sollen einerseits die verwendeten Methoden dokumentieren, um Transparenz und Wissenschaftlichkeit sicher zu stellen, und sie sollen andererseits die Zusammenarbeit mit den Hochschulen und der Wissenschaft fördern.

Zur Illustration der beschriebenen mathematischen Konzepte, werden im Bericht numerische Resultate aufgeführt. Diese sind allerdings nicht als offizielle Resultate der betreffenden Erhebungen zu verstehen. Ebenfalls können die tatsächlich angewendeten Methoden leicht von den hier beschriebenen abweichen.

Die Methodenberichte sind auf der Internetseite des BFS in elektronischer Form verfügbar.

Methodology reports describe the mathematical and statistical methods used to produce findings and carry out analysis in relation to official statistics. These reports also contain assessments and descriptions of new methodological approaches that could be implemented in the future. The aim is to ensure clarity and scientific rigor by keeping a record of the methods used and to encourage a closer working relationship with the scientific community and academic circles.

Numerical findings are presented in methodology reports for the purpose of clarifying the mathematical concepts described and should therefore not be taken as official findings for the surveys in question. Likewise, the methods actually used may differ slightly from those described in the reports.

The electronic version of methodology reports can be downloaded directly from the SFSO Web site.

Precision of released figures is not only an important quality feature of official statistics, it is also essential for a good understanding of the data. In this paper we show a case study of how precision could be conveyed if the multivariate nature of data needs to be taken into account.

In the official release of the Swiss Earnings Structure Survey, the total salary is broken down into several wage components. For surveys 2002 and 2004, we first investigate the incidence of components and then follow Aitchison's approach for the analysis of compositional data, which is based on logratios of components.

Different multivariate analyses of the compositional data are performed and compared between the years, whereby the wage components are broken down by economic activity classes. Then we propose a number of ways to assess precision.

Order Number: 338-0038

Orders: +41/(0)32 713 60 60 Fax: +41/(0)32 713 60 61 E-Mail: order@bfs.admin.ch

Price: free

ISBN 3-303-00338-6