RatSWD
Working Paper Series

Working Paper   No. 152

Cumulation of Cross-Section Surveys

Evaluation of Alternative Concepts for the Cumulated Continuous Household Budget Surveys (LWR) 1999 until 2003 compared to the Sample Survey of Income and Expenditures (EVS) 2003

Joachim Merz and Henning Stolze

August 2010

Working Paper Series of the German Data Forum (RatSWD)

The *RatSWD Working Papers* series was launched at the end of 2007. Since 2009, the series has been publishing exclusively conceptual and historical works dealing with the organization of the German statistical infrastructure and research infrastructure in the social, behavioral, and economic sciences. Papers that have appeared in the series deal primarily with the organization of Germany's official statistical system, government agency research, and academic research infrastructure, as well as directly with the work of the RatSWD. Papers addressing the aforementioned topics in other countries as well as supranational aspects are particularly welcome.

RatSWD Working Papers are non-exclusive, which means that there is nothing to prevent you from publishing your work in another venue as well: all papers can and should also appear in professionally, institutionally, and locally specialized journals. The *RatSWD Working Papers* are not available in bookstores but can be ordered online through the RatSWD.

In order to make the series more accessible to readers not fluent in German, the English section of the *RatSWD Working Papers* website presents only those papers published in English, while the the German section lists the complete contents of all issues in the series in chronological order.

Starting in 2009, some of the empirical research papers that originally appeared in the *RatSWD Working Papers* series will be published in the series *RatSWD Research Notes*.

The views expressed in the *RatSWD Working Papers* are exclusively the opinions of their authors and not those of the RatSWD.

The RatSWD Working Paper Series is edited by:

Chair of the RatSWD (2007/2008 Heike Solga; since 2009 Gert G. Wagner)

Managing Director of the RatSWD (Denis Huschka)

Cumulation of Cross-Section Surveys

Evaluation of Alternative Concepts for the
Cumulated Continuous Household Budget Surveys (LWR) 1999 until 2003
compared to the Sample Survey of Income and Expenditures (EVS) 2003

Joachim Merz and Henning Stolze[1]

Abstract

With the development of household budget systems and with regard to the requirements of the European Union with new EU-SILC approaches, the cumulation of cross-section surveys to an integrated information system is recently discussed and required. In particular the reconstruction of household budget surveys should deliver yearly results as well multi-annual sufficient large samples to allow in depth analyses. This study contributes by a general conceptual foundation of the cumulation of cross-sections and an application which in particular evaluates the new cumulation concept with actual large official samples: the cross sectional cumulation of five yearly Continuous Household Budget Surveys (Laufende Wirtschaftsrechnungen, LWR) which will be compared to the large quinquennial Sample Survey of Income and Expenditures (Einkommens- und Verbrauchsstichprobe, EVS) of the German Federal Statistical Office. Therewith the sensitivity of the cumulation concept with its alternatives is evaluated for private household consumption expenditures of selected expenditure groups. A recommendation concludes.

JEL: C42, C81, D10, E20

Keywords: cumulation of cross sections, temporary cumulation, adjustment by information theory, consumption expenditures, Continuous Household Budget Surveys (Laufende Wirtschaftsrechnungen, LWR), Sample Survey of Income and Expenditures (Einkommens- und Verbrauchsstichprobe, EVS) of the German Federal Statistical Office

---

1   Univ.-Prof. Dr. Joachim Merz, Leuphana University Lüneburg, Department of Economics, Research Institute on Professions (Forschungsinstitut Freie Berufe, FFB), Chair ‚Statistics and Professions', 21332 Lüneburg, Tel: 04131/677-2051, Fax: 04131/677-2059, e-mail: merz@leuphana.de, www.leuphana.de/ffb

Dr. Henning Stolze, Research Institute on Professions, Forschungsinstitut Freie Berufe (FFB), Chair ‚Statistics and Professions', Wege&Gehege serverbased computing, e-mail: stolze@wegeundgehege.de. We are grateful to Brigitte Demant, German Federal Statistical Office, for her elaborated work in building the sub-samples of all Continuous Household Budget Surveys 1999 until 2003 (LWR) and the Sample Survey of Income and Expenditures 2003 (EVS) as well as to the microcensus group of the German Federal Statistical Office for their special reporting concerning adjustment population totals.

1   Introduction

With the development of household budget systems and with regard to the requirements of the European Union, the cumulation of cross-section surveys to an integrated information system is discussed.[2] The so far parallel and not connected surveys should be united in an appropriate way to allow analyses of more complex problems in an integrated system of household statistics. Thereby flexible, reasonable, actual and new data requirements should be enabled for the interested public (Ehling 2002a). In particular the reconstruction of household budget surveys should deliver yearly results as well multi-annual sufficient large samples to allow in depth analyses (Ehling 2002b, 22).

Conducted by Merz 2004, the current study provides a general conceptual foundation of the cumulation of cross sections and an application which in particular evaluates the new cumulation concept with actual large official samples.[3] The cumulation concept, at first discussed more general, is applied to the cumulation of several Continuous Household Budget Surveys (Laufende Wirtschaftsrechnungen, LWR) of the German Federal Statistical Office. This temporary cumulation cumulates a series of single cross sections and does not discuss the case of panel data with respondents repeatedly interviewed. Such an approach with overlapping samples and less efficient results requires further processes.[4]

---

2   This study is a contribution to the project „Official Statistics and Socio-economic Questions" of the German Federal Statistical Office, which is embedded into the new EU-SILC approaches (EUROSTAT-Document „Draft Regulation on the Collection of Statistics on Income and Living Conditions in the Community (EU-SILC)", (EUROSTAT 2001, S. 1, European Commission 2001)).

3   The pros and cons of a preferred cumulation of surveys in contrast to alternate samples e.g. are discussed by Ehling (2002b, 24) or Verma (2002, 51-52) in the conference volume of rotating samples (Statistisches Bundesamt 2002).

4   According to the efficiency of cumulated samples: A cumulation of non-overlapping samples (independent samples without repeated questioning the same microunits) in general is ideal from a sample's theoretical perspective, because only these samples deliver efficient results. The variance is the central measure to determine the significance of a value. If an actual sample is combined with a previous sample, the variance of a mean value is the more reduced the larger the overlapping proportion P is. The variance due to Cochran 1977 is reduced by the factor reduction
$$(4) \qquad (1 - (1 - P)\, R^2\, /\, (1 - (1 - P)^2\, R^2)\,,$$
where R ist the Pearson correlation coefficient. In the consequence the smaller variance indicates a higher level of significance when the cumulation has overlapping microunits. Kordos (2002, 60), however, it shows that the maximum variance reduction (with an optimal P and optimal sample weights) is constrained by the factor $(1 + ((1 - R^2)^{0,5})) / 2$. A variance reduction in the case of an overlapping cumulation is not only valid for the original values but also for their rates of changes (Selén 2002, 75; Kish 1999, 136).
Since for our analyses no overlapping information is available, no such aspects have to be considered; the cumulated sample therefore has to be characterised as a sample of independent microunits with respective sample sampling errors. For further remarks according to the accuracy of a cumulated sample in general see Merz 2004.

With the microdata of the Continuous Household Budget Survey (LWR) cross sections 1999, 2000, 2001, 2002 and 2003 we simulate alternative cumulation scenarios over the single years and build an aggregated cumulation sample. These cumulation alternatives are evaluated for private household consumption expenditures of selected expenditure groups by comparing the results of the aggregated cumulation sample with an appropriate even larger sample, the Sample Survey of Income and Expenditures (EVS) 2003. Therewith the sensitivity of the cumulation concept with its alternatives is evaluated on a large empirical base and with regard to a broad spectrum of household expenditure behaviour. We conclude with a recommendation.

2 Cumulation of cross-section surveys – A concept for the cumulation of yearly household budget surveys

Based on general theoretical approaches Merz developed a concrete cumulation concept for household budget surveys in 2004 and put his concept up for discussion to the interested public. This concept is re-capitulated in its essential elements, where further advancements are marked in cursive letters. The following chapters deepen the central elements and cumulation alternatives which then form the simulation and evaluation.

Cumulation concept and tasks:

1. *Price adjustment of economic values (expenditures, income) of all cross sections to the year t=T:* Appropriate price indices (economic multipliers) should adjust all monetary values and convert them into prices of the final evaluation year T. In contrast to demographic weightings, which are dependent of the sociodemographic structure of the respective household in a cross section, such an economic multiplier is independent of the single respondents (households).

2. *Demographic structure and totals*: It has to be decided which demographic structure for the individual as well as for the household structure should be chosen for a demographic representative adjustment (calibration, re-weighting). This is required for the evaluation year T (the year of the large comparison sample, here the EVS 2003) sample as well as for all periods/years before (here the Continuous Household Budget

3

Surveys 1999, 2000, 2001, 2002 and 2003).[5] The demographic totals of the chosen adjustment should be extracted from a large representative population survey (here the German Microcensus).

3. ***Cumulation weighting:*** The aim of a cumulation weighting is to incorporate the information of all previous samples. To account for the different temporary closeness and thereby the different information content of the previous cross sections, we propose different alternatives to determine appropriate depreciation rates (w_t, $t = 1, ..., T$) for all T cross sections. We incorporate assumed as well as data generated weights based on a cluster analysis.

4. ***New adjustments (calibrations) for the cumulated sample CUM at t=T:*** According to the actual totals (margins, aggregated values) **r** at t=T the additive cumulated and so far price adjusted cross sections t (t=1,…,T) – eventually with respective new adjustment weightings – has to be adjusted theoretically based, simultaneous and consistently.[6] According to the Minimum Information Loss (MIL) principle (see Merz 2004, realised by the program package ADJUST by Merz and Stolze 2004) the chosen adjustment procedure takes care of already available, original adjustment weightings within the information theory based objective function. This approach includes already conducted adjustments or given temporary representativeness via the respective adjustment factors and information from the previous cross sections.

Alternative adjustments in principle:

▪ At first the cumulative weightings w_t are multiplied with the individual original adjustment factors of each sub sample (cross section). The original cross sectional adjustment factors might be the original weights q_t or adjustment weights from new adjustments p_t^* for each sub sample based on their respective totals r_t. The entire

---

aggregated cumulated sample CUM at T then is re-weighted to achieve the totals r_T at period T.

- At first there is a new adjustment for each sub sample within the cumulated sample CUM delivering adjustment weights p_T^+ for each sub sample with respect to the totals r_T. Since each adjusted cross section is representing the population N_T, the cumulated sample CUM represents TN_T observations. The adjustment factors then are multiplied by their respective cumulation weights w_t. The cumulation weights should sum up to 1 so that the entire cumulation sample CUM will finally result in N_T.

The second adjustment alternative with a cumulation weighting *after* a demographic adjustment is more flexible since it allows alternative cumulation weightings later on without an additional demographic adjustment.

5. *Model based extrapolation:*

   If a model based extrapolation by microeconometric estimates is chosen then the extrapolation is linked with the adjustment as follows:

   - If the variables with regard to contents are **independent** from the demographic adjustment, then the model based extrapolation can be applied after the adjustment.
   - If the variables with regard to contents however are **dependent** of the demographic adjustment, then the model based extrapolation has to be considered within the adjustment as a further characteristic.

6. ***Evaluation of CUM compared to another large sample (like EVS):*** With the final cumulation file CUM then the evaluation by comparing its substantive results with the results of another large sample, here the EVS at t=T has to be done.

3   The cumulation concept at work

The above cumulation concept is based on four central building blocks

- Price adjustment of economic values (like expenditures, income)
- Alternative cumulation weighting
- Model based extrapolation
- New demographic adjustment of the cumulated sample(s)

which will be discussed in the following.

3.1 Price adjustment of economic values

Price adjustments of economic variables – here the expenditures and incomes of private households – take into account the price development by appropriate price indices. A price index (economic multiplier) – if not different by regions – is equal for all households and is either a general price index – like the consumer price index – or group specific. The price adjustment of economic values therefore is not a computational problem.

3.2 Alternative cumulation weightings

Our temporary cumulation combines all T given cross sections, here the Continuous Household Budget Surveys (Laufende Wirtschaftsrechnungen, LWR). Since the (yearly) cross sections are delayed by T-t (t=1,…,T-1) we face „outdated" information compared to the actual situation at T.

The aim of a cumulation weighting is to incorporate the information of all samples, in particular former samples with appropriate depreciation rates. The depreciation rates of all cross sections, further called **cumulation weights** w_t, are not to be mixed up with the weights of a demographic adjustment, which will achieve demographic representativeness.

Four approaches to calculate cumulation weights will be discussed briefly:

- Approaches from the computer sciences
- Information theory based approach
- Alternative distance measuring: weighting by similarity-(proximity-) measures
- Model based econometric extrapolation by the AIDS complete demand system and calibration
- Alternative fixed temporary cumulative weighting.

These approaches will be linked and determine the simulation alternatives.

3.2.1 Approaches from the computer sciences: the information value of a data base

The value of information in databases is discussed in informatics with regard to its aging and optimal updating intervals. For instance, the value of a customer database for marketing purposes will decline if the database is older and some of the addresses are not valid any further. Another example is the steering of the information flow: For the caching of network

information certain information is buffered. If the cached information is wrong because of being too old the, wrong information generates costs of additional accesses. From a certain point in time the risk to generate costs because of too old data will outbalance the chance for a direct access to the desired information and potential cost minimization. To evaluate the „risk", a method is necessary to find a measure for „actuality". With address data this is relatively simple: New invalid address data for some point in time are taken to approximate rates of invalid addresses. This is not as easy for other constellations.

Altogether, the idea of estimating the risk to use outdated information is portable to our problem of a temporary cumulation. Different consumer behaviour from different cross sections could be the base to estimate changes in consumer behaviour by a similarity index by distance measures or naturally by econometric approaches. The result could be a certain time dependent depreciation rate $d(\Delta t)$ which could be used for the different cross sections of the cumulation.

Respective approaches from an information theory based perspective, data generated proximity measures, a model based econometric extrapolation and calibration and fixed alternative weightings now will be discussed.

3.2.2 Information theory based weighting

Following the information aspect the information theory based approach with the entropy as a measure of information novelty could help.[7]

The entropy of the information content of a set of objects j (j=1,…,n) on a pro-rata basis $\mathbf{p}=(p_1,...,p_n)'$, $(p_j>0)$, $\sum_j p_j=1$, there is characterized by

(1) $\qquad H(\mathbf{p}) = H(p_1,...,p_n) = \sum_j p_j \log(1/p_j).$

If $\mathbf{p}$ would measure all variable values, then the aggregated information of this cross section could be measured one dimensional by $H(\mathbf{p})$. The **information loss** (respectively the information gain) of a former cross section – with respective pro-rata based $\mathbf{q}=(q_1,...,q_n)'$ – compared to the actual situation $\mathbf{p}$ then could be evaluated by

(2) $\qquad I(\mathbf{p}{:}\mathbf{q}) = \qquad \sum_j p_j \log(1/q_j) - \sum_j p_j \log(1/p_j)$

$\qquad\qquad\qquad = \qquad \sum_j p_j \log(p_j/q_j),$

---

7 Background information about information theory and its applications are provided e.g. by Golan, Judge and Miller (1996).

where $\mathbf{p} = (p_1,...,p_n)'$, $\mathbf{q} = (q_1,...,q_n)'$ with $(p_j,q_j > 0)$, $\Sigma_j p_j = \Sigma_j q_j = 1$,   (j=1,...,n).

This approach corresponds to our demographic adjustment/calibration minimum information loss principle.

For each former cross section an entropy value H_t repectively a distance measure I_t compared to the actual situation at T would be given and a information theory based temporary cumulation weighting could be constructed for the cross sections t(t=1,…,T) by

$$(3) \qquad w_t = I(p_T : q_t) / \sum_{i=1}^{T} I(p_T : q_t) \,.$$

The cross section which is most different to the actual situation could get the highest – or inverse eventually the lowest – weight in the cumulated sample.

For using the entropy concept to characterize a sample, note the following: The entropy is measuring the information content. If the entropy is equal one, the information is distributed at random, with small values redundancies or statistical regularities are given. H(I) is an average information about the regularity structure of the data. Therefore it is questionable if a measure of such a structure is the right weighting approach by content when further socioeconomic behaviour is surrendered.

However, the entropy and its information loss could be regarded as a general measure of distance if the original relative frequencies (p and q) would be further developed as metric survey variables.

3.2.3   Data generated alternative distance measuring: proximity measures

In addition to the discussed information theory based approach there are many alternative distance measures, which detect the distance of an entire sample by proximity measures. As proximity measures – dependent on the scale of measurement – well known are

- **Proximity measures based on a nominal scale** Tanimoto-coefficient, M-coefficient, Kulczynski-coefficient, RR-coefficient, Dice-coefficient, chi^2-coefficzient, …
- **Proximity measures based on a metric scale** L_1- and L_2-Norm, Q-correlation-coefficient, Mahalanobis-distance, Minkowski-metrick (with special case of the quadratic Euclidian distance), generalized least squares, minimum information loss, raking ratio, minimum entropy, Hellinger-distance, modified chi-square, …

All these measures are generated by the samples and its information itself and take into account – similar to the information theory based weighting – differences of all variable

values between two or more samples. A temporary cumulation weighting aspect is caught by the degree of variable value changes as revealed changed behaviour. The proximity approach delivers distances between every cross section at t compared to the actual situation at T. A greater distance shows a relative great change of (consumption) behaviour. We argue that therefore the situation at t then is of lower interest for the actual situation (which has changed a lot); the situation at t because of its particular loss of actuality should be considered by a lower degree. Since not a great distance but the similarity is of final interest, our final proximity based cumulative weight is inverse constructed: The more similar (and probably more actual) a sub sample is, the higher will be its weight.

In the end our concern is to evaluate the impacts of alternative cumulation weightings of private household expenditures in a cumulated sample. The base of any proximity measure, thus are expenditures for certain commodity groups like food, drinks or other services etc. Since these are variables with a metric scale, different metric distances (z.B. Minkowski-metric, cosinus-distance or Tschebyscheff-distance) and proximity measures (e.g. Q-correlation) come into consideration.

Concretely, we apply the Euclidean distance which is underlying the analysis of variance in general. For our case we compute four distances as a respective distance between a Continuous Household Budget Survey (LWR) 1999, 2000, 2001 and 2002 (t=1,…,4) compared to the last available LWR 2003 (t=5=T). Since a distance matrix is needed between the respective cross sections and not between the single observations, the question how to deal with groups (cluster) with regard to their centre has to be answered. Analogous to fusion algorithms of an analysis known approaches like the single or complete linkage, the Centroid or the Ward method can be applied. If practical considerations like the group size and handling with available statistic programs could be neglected the Ward method would be the optimal choice; it is robust and credibly assigns cluster centres and distances to other clusters without causing problems like chain building.

However, hierarchical methods with 20,000 and more observations like in the LWR will meet computational limits of desktop computers. In addition, own fusion routines have to be programmed – because of the given group dependency of the cross section years – since implemented fusion algorithms of common statistical packages are not available. Due to reasons of transparency and practicability a distance measuring between cross sections based on mean values of the expenditure variables is chosen.

Finally the calculated distances have to be transformed into appropriate weights, which have to fulfil the restriction of $\sum_t w_t = 1$. Here we take the respective share of the whole distance as the information loss. The cumulation weights – like in the other approaches – then have to be normalized to the sum of 1.

A data generated cluster analytic cumulation weight then is

$$w_t = \frac{1 - \dfrac{d_{t,T}}{\sum_t d_{t,T}}}{\sum_t \left(1 - \dfrac{d_{t,T}}{\sum_t d_{t,T}}\right)}$$

where $d_{t,T}$ is the squared Euclidean distance between cross section at t and the cumulation year 2003 (T).

Steps of the data generated cluster analytic cumulation weights for our simulations

These are the steps within the cluster analysis to achieve the respective cumulation weights for our simulations:

- Aggregation of single expenditures from the LWR 1999 to 2003 according to desired central commodity groups (here 12 commodity groups).
- Compute arithmetic means of the expenditures of the 12 commodity groups for all cross sections as the basis for the distance matrix.
- Specific price adjustment of the mean values for the expenditures of all 12 commodities expenditures in every survey period.
- Clusteranalysis and calculation of the distances of the cross sections 1999 till 2002 respectively to 2003 (squared Euclidean distances).
- Building cumulation weights from the distance matrix.

The concrete extensive computations finally result in the following weights of the LWRs 1999, 2000, 2001, 2002 and T=2003:

Data generated cluster analytic cumulation weights

w_t = {0.156; 0.177; 0.194; 0.224; 0.250}.

As the result shows, more recent samples here produced higher weights because they are more similar to the sample at T. However, an increasing data generated cluster analytic cumulation weight from t=1 to t=T has not always to be expected necessarily, though more similar data in more recent samples compared to T could be expected.

10

3.2.4 Model based econometric extrapolation with the AIDS demand system and calibration

A model based approach will be understood as an approach supported by economic theory and forming the basis for microeconometric estimates. From a multitude of microeconomic based models (see Merz 2004) we briefly regard the flexible AIDS complete demand system (Almost Ideal Demand System, Deaton and Muellbauer 1980), which has been used already within the framework of cumulation approaches and the analysis of expenditures.

Cassel, Granström, Lundquist und Selén 1997 have proposed such a model based estimation connected with a calibration (adjustment) when cumulating the Swedish household survey HBS from 1985, 1988 and 1992.[8] They apply the AIDS model within their calibration for seven commodity groups out of 6 months and 10 household types. The idea: Expenditure shares for certain commodity groups are estimated from an aggregate (e.g. total expenditures) by a regression analysis and calibrated at the same time.

The central equation of a generalized regression estimator is

$$(5) \qquad t_c(z) = t_z + (t_x^* - t_x)' \beta_{zx}$$

where $t_c(z)$ are the estimated consumption expenditures of a subgroup depending of total expenditures z, $t_z = \sum_{i=1}^{n} (1/\pi_i) z_i = \sum_{i=1}^{n} d_i z_i$ is the weighted expenditure sum (weighted by the Horvitz-Thompson estimator as the reciprocal of the selection probability π) and β_{zx} a coefficient for variable x out of z with $\beta_{zx} = (\sum d_i x_i z_i)/(\sum d_i x_i^2)$.

The linkage to the AIDS model is realized via $t_x^* = R_m t_y$, the estimated expenditures from an expenditure share R_m of income t_y, say. With the AIDS model the expenditure shares R_m are estimated by

$$(6) \qquad R_m = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_j \log(x/P)$$

and its parameters $\alpha_i, \gamma_{ij}, \beta_j$ (P is the price level).

The results from different AIDS applications and their calibration with

---

8   With respectively the same sample plan, same sample size; samples are drawn from the "Register of the Total Population"', largely a random sample

- a simple randomized sample techniques
- calibration with register data (CRD)
- calibration with model supported data (CMD)
- calibration with model supported data and register data (CMRD)

yields the following conclusion (Cassel et al. 1997, S. 19): „it can be expected that the model based calibration methods CMD and CMRD with respect to the variance and the systematic error will yield good results."

If a model based extrapolation is chosen, either by such an expenditure model[9] or by a time series approach etc., then such an extrapolation would be connected with a demographic adjustment in general by

- if the variables of interest with regard to contents are independent from the demographic adjustment the model based extrapolation could be applied after the demographic adjustment of the cumulated sample,
- if these variables are dependent, then the model based extrapolation has to be considered within the demographic adjustment.

Though a model based extrapolation of a sample – here by extrapolation of the expenditure behaviour – has its content driven merits, however and to be critical, in many results with the AIDS application by Cassel et al. 1997, no significant improvement will be visible by their model based estimation and calibration approach (see also the discussion in Selén 2002, 83 pp).

Of course, an improvement might be found with another model type/expenditure system. Since the sample results are dependent on the chosen model and the scientific discussion about the "best" indeed is not finally concluded (if ever), it could be justified, if an institution like the Federal Statistical Office is not following such a model based extrapolation.

The following simulation and evaluation therefore do not include such a model based extrapolation.

---

9   Examples for expenditure systems are the complete demand systems with flexible functional form like the Translog-Model, the mentioned Almost Ideal Demand System (AIDS, QAIDS), the Rotterdam Model etc., or Stone's Linear Expenditure System LES (Stone 1954) the extensions ELES Lluch 1973 and FELES Merz 1983. A good survey about demand systems is given e.g. by Deaton 1990.

3.2.5   Alternative fixed cumulation weights

There is a multitude of cumulation weights as information depreciators when they are pretended without any consideration of the data structure externally. To cover a certain spectrum of such externally fixed cumulation weightings we propose the following three alternatives of cumulation weightings for the samples at t=1,…,T, where T characterizes the actual sample:

- **_Uniform cumulation weighting:_** All samples, the youngest as well the oldest sample is considered by the same weight:

$$w_t = 1/T, \ (t=1,...,T)$$

- **_Linear progressive weighting:_** The oldest sample has the smallest weight, the younger samples have proportional growing weights:

$$w_t = t / \sum_{i=1}^{T} i, \qquad (t=1,...,T)$$

- **_Exponential progressive weight:_** Like the linear progressive weighting, but with an even greater, exponential progression. The actual sample again gets the highest weight. An exponential progression to the base of x is:

  Progression to the base of x

$$w_t = x^{t-1} \Big/ \sum_{i=0}^{T-1} x^i \qquad (t=1,...,T).$$

  Of course, a larger base x strengthens the progression. As alternative c we will choose an exponential progressive weighting to the base of 2, since a higher base would insufficiently consider the first (oldest) samples.[10]

3.2.6   Choosen alternative cumulation weightings

To summarize: The following evaluation encompasses three externally fixed weightings as well as a data generated cluster analytic cumulation weighting. With five sequential samples of the LWRs 1999, 2000, 2001, 2002 and 2003 (t=1,…,T=5) they are:

- a)   **_Uniform cumulation weighting_**

$$w_t = 1/T, \ (t=1,...,T),$$ $w_t = \{0.20; 0.20; 0.20; 0.20; 0.20\}$.

---

10   By a weighting to the base of 3 (and higher) the information from the first samples practically would be lost, since the last sample would have a weight which is 80 times higher than the weight from the first sample.

b)   ***Linear progressive weighting***

$$w_t = t / \sum_{i=1}^{T} i, \qquad (t = 1,...,T), \qquad w_t = \{0.067; 0.133; 0.200; 0.267; 0.333\}$$

c)   ***Exponential growing weighting (base 2)***

$$w_t = t / \sum_{i=0}^{T-1} 2^i, \qquad (t = 1,...,T) \qquad w_t = \{0.032; 0.065; 0.129; 0.258; 0.516\}$$

d)   ***Data generated cluster analytic weighting (Euclidean distance)***

$w_t = \{0.156; 0.177; 0.194; 0.224; 0.250\}$.

Alternative cumulation weights without LWR 2000

When mean and variances are compared between the different LWRS from 1999, 2000, 2001, 2002, 2003 extraordinary deviations of the 2000 LWR will be evident. A deeper inspection shows that e.g. even with a threefold standard deviation more than 35% (and more than 15% with a fivefold standard deviation) of all values are beyond that deviation around the mean. Based on that and on further evidence, the LWR 2000 will not be considered further on because of its restricted data quality.

So the discussed weightings have to be changed: The LWR 2000 will be deleted by a weight of zero and the other weights are changed to sum up to 1. Table 1 shows the final used cumulation weightings.

Table 1: Alternative cumulation weightings without LWR 2000

New cumulation weightings (without LWR00, t=2)	Alternative cumulation weightings			
	a: uniform	b: linear progressive	c: exponential progressive	d: data generated cluster analytic
t=1 (1999)	25.0%	7.7%	3.4%	18.9%
2 (2000)	0.0%	0.0%	0.0%	0.0%
3 (2001)	25.0%	23.1%	13.8%	23.5%
4 (2002)	25.0%	30.8%	27.6%	27.2%
5 (2003)	25.0%	38.5%	55.2%	30.4%

As Table 1 shows, our alternative cumulation weightings cover a broad spectrum with lower and higher weights of older and younger samples which allow pre-estimates for other weighting proposals, too.

14

Figure 1: Alternative cumulation weightings without LWR 2000
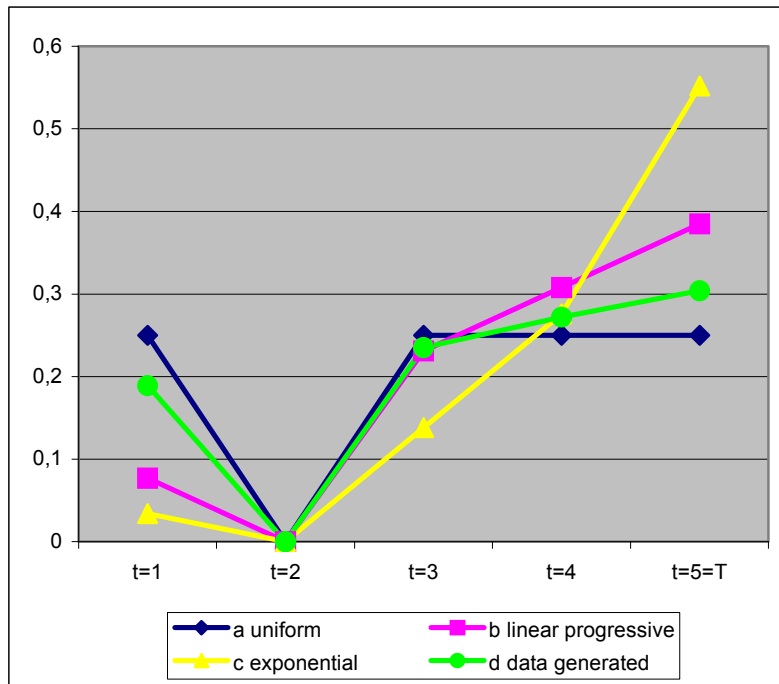
3.3 Alternative demographic adjustments/calibrations

A new adjustment (calibration) as a demographic weighting to achieve available totals in general is necessary if a sample is not at random finally. Representativeness is obtained by an observation (microunit) dependent on weighting, which takes into account the individual characteristics of each household. Such an adjustment is going by far beyond an identical weight for all observations (as the reciprocal of the selection rate).

Our demographic adjustment within alternative cumulation concepts is based on information theory and the Minimum Information Loss (MIL) principle where the information loss in the objective function is minimized when the distribution of available weights is substituted by new weights. An information theory based approach was already discussed in chapter 3.2.2 when a whole sample's information is used to determine a depreciation weight. When applying information theory to the adjustment/calibration task the new adjustment factors then are the solution of a non-linear optimization problem under constraints:

(10) $Z(\mathbf{p},\mathbf{q}) = \min_{\mathbf{p}} \{\sum_j p_j \log(p_j/q_j)\}$ $0 < p_j, q_j < 1, \sum_j p_j = \sum_j q_j = 1,$

subject to

(11) $\mathbf{Sp} = \mathbf{r}$.

15

where **p** is the n (=number of sample units) vector of the quested adjustment factors/weights. The objective function Z(**p,q**) is minimizing the distance to already available, original weights **q**. The restriction assures the externally given totals with m adjustment characteristics via the weighted by **p** aggregation over the sample information matrix **S** (m,n-matrix). The Lagrange multipliers **λ** (m-vector) as the solution the determine the new adjustment factors by

(12) $p_j = q_j \exp(\mathbf{\lambda}'\mathbf{s^j}-1)$ (j=1,…,n)

where $\mathbf{s^j}$ is describing the characteristics of the j-th observation (household) as a column vector.

Each single adjustment factor p_j (only one factor per observation/household) is simultaneously weighting all m adjustment characteristics. Such an adjustment factor describes the final number of households in the population with the same characteristics like household j from the sample.

The information based approach assures the necessary positivity condition of the new adjustment factors (maintaining the observation) and allows a simultaneous weighting also of hierarchical data (here household and personal data). A detailed discussion of the methodological background and the efficient algorithm for even large micro data sets is given in Merz 1983a, 1985, 1994 and Merz and Stolze 2008. Information about the used ADJUST adjustment program package is available in http://ffb.uni-lueneburg.de/adjust and Merz and Stolze 2004.

As mentioned within the discussion of the cumulation weights this adjustment approach is used for the demographic calibration of the single samples at t=1,…,T and/ort he new demographic calibration of the cumulated sample CUM with alternative cumulation weights.

Three fundamental variants of the demographic adjustment with regard to the cumulation weights come into consideration (see Table 2) which refines the above adjustment alternatives:

I.   Start with a new demographic adjustment for each sample to achieve updated totals $\mathbf{r_t}$. The new cross section weights $\mathbf{p_t^*}$ then has to be multiplied by the above cumulation weights $\mathbf{p_t^*}\mathrm{w_t}$ and serve as available weights for a final demographic adjustment at t=T (2003) of the entire cumulation sample CUM with totals $\mathbf{r_T}$.

II. Use already available, original adjustment factors $\mathbf{q_t}$ which might be provided by the surveying institution. Multiply them by the above cumulation weights $\mathbf{q_t}\mathrm{w_t}$. These

16

weights then serve as available weights for a final demographic adjustment at T= 2003 of the entire cumulation sample CUM with totals $\mathbf{r}_T$.

III. Adjust each sub sample at t (t=1,…,T) of the cumulation sample CUM to the totals $\mathbf{r}_T$ (T=2003). Since each adjusted cross section is representing the population N_T, the cumulated sample CUM represents TN_T observations. The adjustment factors then are multiplied by their respective cumulation weights w_t. The cumulation weights should sum up to 1 so that the entire cumulation sample CUM will finally result in N_T.

As mentioned, variant III is more flexible than the others because another cumulation weighting later on is possible without a new overall demographic adjustment.

Table 2: Alternative temporal cumulation procedures

Cumulation weighting ex ante of a (final) demographic adjustment		Cumulation weighting ex post of a demographic adjustment to r_T
With ex ante demographic adjustment to r_t	Without ex ante demographic adjustment to r_t	
(Variant I)	(Variant II)	(Variant III)
Adjustment (single samples)		
1. Adjustment of single samples to the respective totals at t q $\downarrow$ Adjustment to r_t p^* $p^* = f(q; r_t)$		1. Adjustment of single samples to the totals at T q $\downarrow$ Adjustment to r_T p^+ $p^+ = f(q; r_T)$
Cumulation weighting		
2. Consideration of the information loss of older samples by cumulation weighting (4 methods) of the new adjustment factors $p^* \cdot w_t$	1. Consideration of the information loss of older samples by cumulation weighting (4 methods) of the available, original adjustment factors $q \cdot w_t$	2. Consideration of the information loss of older samples by cumulation weighting (4 methods) of the new adjustment factors of the cumulated sample $p_{III} = f(p^+; w_t)$ $= p^+ \cdot w_t$

Adjustment (cumulated sample)	3. Adjustment of the cumulated sample to the totals at T $p^* \cdot w_t$ ↓ Adjustment to r_T p_I $p_I = f\left(p^* \cdot w_t; r_T\right)$	2. Adjustment of the cumulated sample to the totals at T $q \cdot w_t$ ↓ Adjustment to r_T p_{II} $p_{II} = f\left(q \cdot w_t; r_T\right)$	

q: available, original weights of the single samples (here: Continuous Household Budget Surveys LWR, variable name: HRD)

Chosen adjustment totals: Household and person information from the German Microcensus 1999 to 2003

With many thanks to the Federal Statistical Office and its special summary tabulations we could choose German Microcensus results 1999, 2000, 2001, 2002 and 2003 for appropriate demographic totals. Our aim was to include adequate structural data which are connected to the content driven analysis of private household consumption expenditures. So we consider different household types, the labour force participation as purchasing power background, as well personal information like occupational status and the population age structure according to gender:

- Occupational status of persons (civil servant, employee, blue collar worker, pensioner)
- Private households according to household type (single household, households with two, three, four and more persons)
- Private households according to the number of persons in the labour force (no such person, one and two active persons)
- Population according to gender and age (with respective 7 age classes).

The vector of totals **r** therefore consists of m=25 adjustment characteristics, which tells us the number of the respective households or persons for Germany in the respective years (for details see the Appendix Table A1 in Merz and Stolze 2010).

The adjustment procedure for instance delivers more than 24,000 individual demographic adjustment factors for the cumulation sample CUMLWR 2003. Each single adjustment factor simultaneously takes care of all 25 adjustment characteristics and is the number of respective population households/persons.

4   Chosen cumulation alternatives

The cumulation alternatives unify the alternative methods of cumulation weighting (chapter 3.2) with the alternative demographic adjustment variants (chapter 3.3).

With four **methods of cumulation weighting** (depreciation weights) as

a)   Uniform cumulation weighting

b)   Linear progressive weighting

c)   Exponential growing weighting (base 2)

d)   Data generated cluster analytic weighting

and three different **demographic adjustment variants**

I)   Cumulation weighting *ex ante* of the (final) adjustment
      with previous adjustment to current totals r_t

II)   Cumulation weighting ex ante of the (final) adjustment
      without previous adjustment to current totals r_t

III) Cumulation weighting ex post of the adjustment
      with adjustment only to the totals r_T at T

We face 12 simulation alternatives with **12 individual specific cumulation factors**

$$P_{adjustment\ variant}^{cumulation\ alternative}$$

which are signed by their respective indices. Note, these 12 cumulation factors are different for each microunit (here household) in the cumulated sample and are the basis for our following evaluation.

5   Impacts of cumulation alternatives on private household expenditures – Evaluation of CUMLWR 2003 compared to EVS 2003 and LWR 2003

To evaluate the information gain of a cumulation, here the cumulation of Continuous Household Budget Surveys (LWR) (smaller samples), we choose a comparison to a content near large sample, here the Sample Survey of Income and Expenditures (EVS), as well as to the LWR at the survey period of the large sample, too. We analyse the individual quarterly consumption expenditures by the 12 main expenditure categories as of the German Federal Statistical Office.

The additive cumulated sample CUMLWR 2003 of the five respective four LWRs 1999, (2000), 2001, 2002, and 2003 encompasses 30,480 household data records (6,169 thereof (20.2%) of the LWR 2000 which will be erased by appropriate zero adjustment factors). The comparison survey EVS 2003 consists of 50,511 households with respective expenditure information.

Which cumulation alternative from CUMLWR will be „better" than the Sample Survey of Income and Expenditures EVS 2003? Is the cumulation sample superior to the original Continuous Household Budget Survey 2003? These questions will be answered empirically based as follows:[11]

Expenditure categories

We analyze the individual expenditures of the Federal Statistical Office's 12 main expenditure categories as well total expenditures as respective quarterly values:

Total expenditures	Sum W01 to W12
Food and alcohol free beverages	W01
Alcohol beverages and tobacco	W02
Clothing and shoes	W03
Accomodation, water, electricity, gas etc.	W04
Equipment, instruments, devices etc.	W05
Health	W06
Transport	W07
Media	W08
Leisure, entertainment, culture	W09
Education	W10
Lodging and catering industry	W11
Other goods and services	W12

Rather rare expenditure groups (lady's trousers and PC complete systems (that time)) are only available in CUMLWR and can not be evaluated any further unfortunately.

Cumulation variants

We compare three adjustment variants : (I) cumulation weighting *ex ante* of the (final) adjustment *with* previous adjustment to current totals r_t ; (II) cumulation weighting ex ante of the (final) adjustment *without* previous adjustment to current totals r_t ; (III) Cumulation weighting ex post of the adjustment with adjustment only to the totals r_T at T with the respective four alternative cumulation weights: a   uniform, b   linear progressive, c

---

11 An additional evaluation of the cumualation factors themselves with measures of central tendency and variance including Gini-coefficients to the brief of exposition is not shown here but can be found in Merz and Stolze 2010.

exponential, d   data generated cluster analytic. Every household thus has 12 cumulation alternative weights.

Evaluation criteria

The cumulation sample consists of all households of the T cross sections. In principle and known so far, the comparison sample EVS relies on different households. Therefore an individual comparison of the same households out of CUMLWR and EVS is not possible due to its independency. Known measures of forecasting accuracy (see e.g. Merz 1980) which compare individual forecasts (LWR, say) with actual values (EVS, say) cannot be applied. Thus, only aggregates over households can be evaluated.

With reference to the efficiency of an estimator as a desirable estimation property with unbiasedness and minimum variance as the evaluation criteria for the goodness of fit of the CUMLWR 2003 to EVS 2003 and LWR 2003 we choose the arithmetic mean and the variance of the respective expenditure aggregates.

The following evaluation results/Tables provide the respective means and variances of the EVS 2003 and the LWR 2003 compared by the relative deviation to all cumulation alternatives out of CUMLWR 2003 as well as maximum and minimum and range (max-min) of these alternative specific deviations (not the range of individual expenditures).

5.1   Cumulation alternatives in comparison – Mean values of private household expenditures

The mean value comparison results are organized with respect to the three adjustment variants I, II and III in the evaluation Tables 3, 4, 5.

Total expenditures: Comparison of means

The EVS 2003 has mean total quarterly expenditures of about 6,.392.70 EURO. All twelve cumulation alternatives of CUMLWR 2003 quite closely achieve this value. Though they underestimate this value between -5.3% and -6.6% according to the alternative, the CUMLWR results, however, all are closer to the EVS than the current LWR 2003 (-8.4%).

Thus there is a cumulation gain for all twelve cumulation alternatives compared to the single LWR 2003. All alternative specific aggregates are close together, their difference is max 1.3 percentage points.

Table 3: Comparison of means of selected consumption expenditures: CUMLWR 2003 versus EVS 2003 and LWR 2003 – Adjustment variant I

| Means | | EVS 2003* (Var.-name) | LWR 2003 | LWR-EVS | CUMLWR 2003 ** Adjustment variant I | | | | | | |
					Ia	Ib	Ic	Id	Minimum	Maximum	Max-Min
Sum W01 to W12	Mean	6.392,70	5.858,18		5.972,00	6.022,25	6.045,50	5.990,05	5.972,00	6.045,50	73,51
Total expenditures	Rel. dev. to EVS			-8,40%	-6,60%	-5,80%	-5,40%	-6,30%	-6,60%	-5,40%	1,15%
W01	Mean	778,08	709,78		686,41	697,02	700,03	690,07	686,41	700,03	13,61
Food and alcohol free beverages	Rel. dev. to EVS	(NG)		-8,80%	-11,80%	-10,42%	-10,00%	-11,30%	-11,80%	-10,00%	1,75%
W02	Mean	119,47	143,45		134,65	132,82	131,86	133,98	131,86	134,65	2,78
Alcohol beverages and tobacco	Rel. dev. to EVS	(AlkGT)		20,10%	12,70%	11,17%	10,40%	12,10%	10,40%	12,70%	2,33%
W03	Mean	326,25	343,74		337,57	333,21	330,22	335,94	330,22	337,57	7,34
Clothing and shoes	Rel. dev. to EVS	(BeklSch)		5,40%	3,50%	2,10%	1,20%	3,00%	1,20%	3,50%	2,25%
W04	Mean	2.043,75	1.844,24		1.792,28	1.814,30	1.823,80	1.800,29	1.792,28	1.823,80	31,51
Accomodation, water, electricity, gas, etc.	Rel. dev. to EVS	(W04selbst)		-9,80%	-12,30%	-11,20%	-10,80%	-11,90%	-12,30%	-10,80%	1,54%
W05	Mean	375,01	410,79		415,67	414,11	410,52	414,64	410,52	415,67	5,16
Equipment, instruments, devices etc.	Rel. dev. to EVS	(W05selbst)		9,50%	10,80%	10,40%	9,50%	10,60%	9,50%	10,80%	1,38%
W06	Mean	251,55	343,72		265,4	276,54	281,49	269,43	265,4	281,49	16,1
Health	Rel. dev. to EVS	(Gesundpflege)		36,60%	5,50%	9,90%	11,90%	7,10%	5,50%	11,90%	6,40%
W07	Mean	890,61	928,53		943,26	946,35	957,01	945,24	943,26	957,01	13,75
Transport	Rel. dev. to EVS	(Verkehr)		4,30%	5,90%	6,30%	7,50%	6,10%	5,90%	7,50%	1,54%
W08	Mean	200,48	173,19		161,52	170,85	174,08	164,77	161,52	174,08	12,56
Media	Rel. dev. to EVS	(Nachrichten)		-13,60%	-19,40%	-14,80%	-13,20%	-17,80%	-19,40%	-13,20%	6,26%
W09	Mean	770,16	662,9		655,8	664,9	666,23	658,88	655,8	666,23	10,43
Leisure, entertainment, culture	Rel. dev. to EVS	(Freizeit)		-13,90%	-14,80%	-13,70%	-13,50%	-14,40%	-14,80%	-13,50%	1,35%
W10	Mean	55,3	262,83		301,66	305,01	297,04	302,65	297,04	305,01	7,97
Education	Rel. dev. to EVS	(Bildung)		375,30%	445,50%	451,60%	437,20%	447,30%	437,20%	451,60%	14,41%
W11	Mean	292,87	382,43		367,51	368,38	367,08	367,63	367,08	368,38	1,3
Lodging and catering services	Rel. dev. to EVS	(Beherbg)		30,60%	25,50%	25,80%	25,30%	25,50%	25,30%	25,80%	0,44%
W12	Mean	289,17	249,66		250,29	243,81	243,84	248,14	243,81	250,29	6,48
Other goods and services	Rel. dev. to EVS	(AndWarDl)		-13,70%	-13,40%	-15,70%	-15,70%	-14,20%	-15,70%	-13,40%	2,24%
W0312226	Mean	n/a***	132,25		135,01	132	131,31	133,62	131,31	135,01	3,7
Ladies pants (long, cotton)	Rel. dev. to EVS										
W0913011	Mean	n/a***	2.404,34		2.377,93	2.277,60	2.262,16	2.337,07	2.262,16	2.377,93	115,77
PC- complete systems and notebooks	Rel. dev. to EVS										

*   weighted data by EVS internal variable HRB (results in N=35,.899,946 cases)
**   cumulated survey out of LWR surveys 1999-2003 with alternative adjustment variants and cumulation weightings; price adjusted to 2003; re-calculated in EURO and quarter yearly values
Source: Cumulation survey CUMLWR 2003 out of Continuous Household Budget Surveys (LWR) 1999, 2001, 2002, 2003; Own calculations

Table 4: Comparison of means of selected consumption expenditures: CUMLWR 2003 versus EVS 2003 and LWR 2003 – Adjustment variant II

Means		EVS 2003* (Var.-name)	LWR 2003	LWR-EVS	CUMLWR 2003 ** Adjustment variant II						
					IIa	IIb	IIc	IId	Minimum	Maximum	Max-Min
Sum W01 to W12	Mean	6.392,70	5.858,18		5.981,60	6.036,07	6.056,53	6.000,61	5.981,60	6.056,53	74,93
Total expenditures	Rel. dev. to EVS			-8,40%	-6,40%	-5,60%	-5,30%	-6,10%	-6,40%	-5,30%	1,17%
W01	Mean	778,08	709,78		688,42	698,84	701,37	691,98	688,42	701,37	12,95
Food and alcohol free beverages	Rel. dev. to EVS	(NG)		-8,80%	-11,50%	-10,20%	-9,90%	-11,10%	-11,50%	-9,90%	1,66%
W02	Mean	119,47	143,45		135,08	133,09	132,03	134,36	132,03	135,08	3,05
Alcohol beverages and tobacco	Rel. dev. to EVS	(AlkGT)		20,10%	13,10%	11,40%	10,50%	12,50%	10,50%	13,10%	2,55%
W03	Mean	326,25	343,74		338,36	334,3	331,11	336,84	331,11	338,36	7,25
Clothing and shoes	Rel. dev. to EVS	(BeklSch)		5,40%	3,70%	2,50%	1,50%	3,20%	1,50%	3,70%	2,22%
W04	Mean	2.043,75	1.844,24		1.796,45	1.819,88	1.828,62	1.804,85	1.796,45	1.828,62	32,16
Accomodation, water, electricity, gas, etc.	Rel. dev. to EVS	(W04selbst)		-9,80%	-12,10%	-11,00%	-10,50%	-11,70%	-12,10%	-10,50%	1,57%
W05	Mean	375,01	410,79		417,08	415,78	411,74	416,15	411,74	417,08	5,34
Equipment, instruments, devices etc.	Rel. dev. to EVS	(W05selbst)		9,50%	11,20%	10,90%	9,80%	11,00%	9,80%	11,20%	1,42%
W06	Mean	251,55	343,72		264,36	276,42	281,31	268,46	264,36	281,31	16,95
Health	Rel. dev. to EVS	(Gesundpflege)		36,60%	5,10%	9,90%	11,80%	6,70%	5,10%	11,80%	6,74%
W07	Mean	890,61	928,53		942,42	947,44	958,1	944,96	942,42	958,1	15,68
Transport	Rel. dev. to EVS	(Verkehr)		4,30%	5,80%	6,40%	7,60%	6,10%	5,80%	7,60%	1,76%
W08	Mean	200,48	173,19		162,19	171,11	174,16	165,3	162,19	174,16	11,97
Media	Rel. dev. to EVS	(Nachrichten)		-13,60%	-19,10%	-14,60%	-13,10%	-17,50%	-19,10%	-13,10%	5,97%
W09	Mean	770,16	662,9		657,75	666,28	667,1	660,62	657,75	667,1	9,35
Leisure, entertainment, culture	Rel. dev. to EVS	(Freizeit)		-13,90%	-14,60%	-13,50%	-13,40%	-14,20%	-14,60%	-13,40%	1,21%
W10	Mean	55,3	262,83		298,04	301,9	295,25	299,13	295,25	301,9	6,65
Education	Rel. dev. to EVS	(Bildung)		375,30%	439,00%	445,90%	433,90%	440,90%	433,90%	445,90%	12,02%
W11	Mean	292,87	382,43		367,42	368,81	367,38	367,64	367,38	368,81	1,43
Lodging and catering services	Rel. dev. to EVS	(Beherbg)		30,60%	25,50%	25,90%	25,40%	25,50%	25,40%	25,90%	0,49%
W12	Mean	289,17	249,66		250,74	244,43	244,26	248,67	244,26	250,74	6,49
Other goods and services	Rel. dev. to EVS	(AndWarDl)		-13,70%	-13,30%	-15,50%	-15,50%	-14,00%	-15,50%	-13,30%	2,24%
W0312226	Mean	n/a***	132,25		135,44	132,36	131,57	134,04	131,57	135,44	3,87
Ladies pants (long, cotton)	Rel. dev. to EVS										
W0913011	Mean	n/a***	2.404,34		2.378,35	2.273,51	2.253,11	2.335,31	2.253,11	2.378,35	125,24
PC- complete systems and notebooks	Rel. dev. to EVS										

*   weighted data by EVS internal variable HRB (results in N=35,.899,946 cases)

**   cumulated survey out of LWR surveys 1999-2003 with alternative adjustment variants and cumulation weightings; price adjusted to 2003; re-calculated in EURO and quarter yearly values

Source: Cumulation survey CUMLWR 2003 out of Continuous Household Budget Surveys (LWR) 1999, 2001, 2002, 2003; Own calculations

Table 5: Comparison of means of selected consumption expenditures: CUMLWR 2003 versus EVS 2003 and LWR 2003 – Adjustment variant III

Means		EVS 2003* (Var.-name)	LWR 2003	LWR-EVS	CUMLWR 2003 ** Adjustment variant III				Minimum	Maximum	Max-Min
					IIIa	IIIb	IIIc	IIId			
Sum W01 to W12	Mean	6.392,70	5.858,18		5.977,89	6.029,70	6.053,98	5.996,58	5.977,89	6.053,98	76,09
Total expenditures	Rel. dev. to EVS			-8,40%	-6,50%	-5,70%	-5,30%	-6,20%	-6,50%	-5,30%	1,19%
W01	Mean	778,08	709,78		687,68	698,6	701,66	691,46	687,68	701,66	13,99
Food and alcohol free beverages	Rel. dev. to EVS	(NG)		-8,80%	-11,60%	-10,22%	-9,80%	-11,10%	-11,60%	-9,80%	1,80%
W02	Mean	119,47	143,45		134,81	132,95	131,99	134,13	131,99	134,81	2,82
Alcohol beverages and tobacco	Rel. dev. to EVS	(AlkGT)		20,10%	12,80%	11,28%	10,50%	12,30%	10,50%	12,80%	2,36%
W03	Mean	326,25	343,74		338,24	333,87	330,87	336,63	330,87	338,24	7,37
Clothing and shoes	Rel. dev. to EVS	(BeklSch)		5,40%	3,70%	2,30%	1,40%	3,20%	1,40%	3,70%	2,26%
W04	Mean	2.043,75	1.844,24		1.794,49	1.817,75	1.827,47	1.802,93	1.794,49	1.827,47	32,99
Accomodation, water, electricity, gas, etc.	Rel. dev. to EVS	(W04selbst)		-9,80%	-12,20%	-11,10%	-10,60%	-11,80%	-12,20%	-10,60%	1,61%
W05	Mean	375,01	410,79		416,71	414,42	410,81	415,47	410,81	416,71	5,9
Equipment, instruments, devices etc.	Rel. dev. to EVS	(W05selbst)		9,50%	11,10%	10,50%	9,50%	10,80%	9,50%	11,10%	1,57%
W06	Mean	251,55	343,72		264,25	275,84	281,01	268,37	264,25	281,01	16,76
Health	Rel. dev. to EVS	Gesundpflege)		36,60%	5,10%	9,70%	11,70%	6,70%	5,10%	11,70%	6,66%
W07	Mean	890,61	928,53		943,55	947,07	958,19	945,69	943,55	958,19	14,64
Transport	Rel. dev. to EVS	(Verkehr)		4,30%	5,90%	6,30%	7,60%	6,20%	5,90%	7,60%	1,64%
W08	Mean	200,48	173,19		161,68	171,1	174,38	164,96	161,68	174,38	12,7
Media	Rel. dev. to EVS	(Nachrichten)		-13,60%	-19,40%	-14,70%	-13,00%	-17,70%	-19,40%	-13,00%	6,34%
W09	Mean	770,16	662,9		656,13	665,19	666,52	659,21	656,13	666,52	10,39
Leisure, entertainment, culture	Rel. dev. to EVS	(Freizeit)		-13,90%	-14,80%	-13,60%	-13,50%	-14,40%	-14,80%	-13,50%	1,35%
W10	Mean	55,3	262,83		302,75	306,34	298,45	303,86	298,45	306,34	7,88
Education	Rel. dev. to EVS	(Bildung)		375,30%	447,50%	454,00%	439,70%	449,50%	439,70%	454,00%	14,25%
W11	Mean	292,87	382,43		367,5	368,36	367,16	367,62	367,16	368,36	1,2
Lodging and catering services	Rel. dev. to EVS	(Beherbg)		30,60%	25,50%	25,80%	25,40%	25,50%	25,40%	25,80%	0,41%
W12	Mean	289,17	249,66		250,67	244,1	244,09	248,5	244,09	250,67	6,58
Other goods and services	Rel. dev. to EVS	(AndWarDl)		-13,70%	-13,30%	-15,60%	-15,60%	-14,10%	-15,60%	-13,30%	2,28%
W0312226	Mean	n/a***	132,25		135,1	132,17	131,49	133,74	131,49	135,1	3,62
Ladies pants (long, cotton)	Rel. dev. to EVS										
W0913011	Mean	n/a***	2.404,34		2.374,31	2.269,17	2.252,54	2.331,82	2.252,54	2.374,31	121,78
PC- complete systems and notebooks	Rel. dev. to EVS										

*   weighted data by EVS internal variable HRB (results in N=35,.899,946 cases)
**   cumulated survey out of LWR surveys 1999-2003 with alternative adjustment variants and cumulation weightings; price adjusted to 2003; re-calculated in EURO and quarter yearly values
Source: Cumulation survey CUMLWR 2003 out of Continuous Household Budget Surveys (LWR) 1999, 2001, 2002, 2003; Own calculations

Which cummulant weighting (depreciation rate) is in favour?
Single expenditure categories: comparison of means

Although the cumulation alternatives are close together with regard to total expenditures, there are varying differences with respect to single expenditure categories.

An outlier will be visible: The relative low education expenditures within the EVS 2003. With quarterly 55 EURO this value is very different compared to the LWR 2003 with 262 EURO and about 300 EURO from all alternatives from CUMLWR 2003. Here definition problems have to be assumed so that we neglect education expenditures further on.

From all three adjustment variants I, II and III an identical pattern of a *best[12] cumulation weighting* is crystallizing with Table 5: The best results are by *exponential (c) and uniform (a) weighting* for the following expenditure categories:

Best cumulation weighting c (exponential)

Food and alcohol free beverages	W01
Alcohol beverages and tobacco	W02
Clothing and shoes	W03
Accomodation, water, electricity, gas etc.	W04
Equipment, instruments, devices etc.	W05
Media	W08
Leisure, entertainment, culture	W09
Education	W10
Lodging and catering industry	W11

Best cumulation weighting a (uniform)

Health	W06
Transport	W07
Other goods and services	W12

All other cumulation weightings (b linear progressive und d data generated) yield in all adjustment variants some greater relative deviations. However, the mean expenditures of all twelve cumulation alternatives for all expenditures are relatively close; the ranges of the twelve relative deviations CUMLWR 2003 to EVS 2003 are between -0.41 percentage points (lodging) and 6.74 percentage points (health).

The goodness of fit thus is dependent to a certain extent of the expenditure category. An uniform weighting of the temporal depreciation rates – and thus a relative stronger weighting even for the oldest information (LWR 1999) – for health, transport and other goods and services yield better results compared to all other nine categories where a strong prompt exponential weighting yield better results.

---

12 In the sense of minimal absolute deviation to the respective EVS value.

Probably there might be more habit persistence for health, transport and other goods and services expenditures whereas for the other expenditure categories over time a more rapid behavioural change could be deducted from the stronger prompt weighting. Further research is necessary here.

Which adjustment variant is in favour?
Single expenditure categories: comparison of means

Central result so far: The best cumulation weightings are c: Exponential weighting with a strong weight of the most actual sample and a: Uniform weighting of all samples dependent on the expenditure category.

Now, which *adjustment variant* is the best? We combine the respective single results from Tables 3, 4 and 5 into Table 6. The respective columns show the just discussed best weighting alternatives exponential c respectively uniform a.

All three adjustment variants (I, II, III) again are close together: The range between the best relatively worst adjustment variant (measured as relative deviation to EVS 2003) over all expenditure categories is only between 0.1% and 0.4%.[13]

If we take the computational efforts as an additional evaluation criteria, ~~so~~ adjustment variant III is to be favoured, which at first is adjusting the entire cumulation sample CUMLWR at T=2003 and then is applying the cumulation weighting. In particular, variant III allows a subsequent alternative cumulation weighting of the single cross sections without a new adjustment when a new cumulation weighting is of interest (a new adjustment for each new cumulation weighting would be required by the adjustment variants I and II).

Thus, with reference to the necessary computational efforts, the close results between variants I, II and III as well as respective best results in some expenditure categories

> adjustment variant III: cumulation weighting with alternative depreciation
>
> rates *after* the final adjustment of the entire cumulation sample

is recommended.

---

13 Education with 5.8%, but regard the discussed specific discrepancy between the EVS and the LWR values in general.

CUMLWR 2003 better than LWR 2003?
Single expenditure categories: comparison of means

With the EVS 2003 as the reference the information gain of the cumulation with CUMLWR 2003 compared to the actual LWR 2003 is in favour not only for total expenditures but also for most of the single expenditure categories (eight out of twelve). Even if for single categories the difference is not large, there are cumulation gains up to more than 30 percentage points (health),

Taking into account former cross sections improve the results – for total expenditures as well as for single expenditure categories – compared to a singular current Continuous Household Budget Surveys (LWR), a result which clearly favours a cumulation of cross section samples.

Table 6: Comparison of means of selected consumption expenditures: CUMLWR 2003 versus EVS 2003 and LWR 2003 – Best results of the adjustment variants I, II and III

Means		EVS 2003* (Var.-name)	LWR 2003	LWR-EVS	CUMLWR Cumulation weightings c,a	2003 ** I	II	III	Minimum	Maximum	Max-Min
Sum W01 to W12	Mean	6.392,70	5.858,18			6.045,50	6.056,53	6.053,98	6.045,50	6.056,53	11,03
Total expenditures	Rel. dev. to EVS			-8,40%	c	-5,40%	-5,30%	-5,30%	-5,40%	-5,30%	0,17%
W01	Mean	778,08	709,78			700,03	701,37	701,66	700,03	701,66	1,63
Food and alcohol free beverages	Rel. dev. to EVS	(NG)		-8,80%	c	-10,00%	-9,90%	-9,80%	-10,00%	-9,80%	0,21%
W02	Mean	119,47	143,45			131,86	132,03	131,99	131,86	132,03	0,17
Alcohol beverages and tobacco	Rel. dev. to EVS	(AlkGT)		20,10%	c	10,40%	10,50%	10,50%	10,40%	10,50%	0,14%
W03	Mean	326,25	343,74			330,22	331,11	330,87	330,22	331,11	0,88
Clothing and shoes	Rel. dev. to EVS	(BeklSch)		5,40%	c	1,20%	1,50%	1,40%	1,20%	1,50%	0,27%
W04	Mean	2.043,75	1.844,24			1.823,80	1.828,62	1.827,47	1.823,80	1.828,62	4,82
Accomodation, water, electricity, gas, etc.	Rel. dev. to EVS	(W04selbst)		-9,80%	c	-10,80%	-10,50%	-10,60%	-10,80%	-10,50%	0,24%
W05	Mean	375,01	410,79			410,52	411,74	410,81	410,52	411,74	1,22
Equipment, instruments, devices etc.	Rel. dev. to EVS	(W05selbst)		9,50%	c	9,50%	9,80%	9,50%	9,50%	9,80%	0,33%
W06	Mean	251,55	343,72			265,4	264,36	264,25	264,25	265,4	1,14
Health	Rel. dev. to EVS	(Gesundpflege)		36,60%	a	5,50%	5,10%	5,10%	5,10%	5,50%	0,45%
W07	Mean	890,61	928,53			943,26	942,42	943,55	942,42	943,55	1,13
Transport	Rel. dev. to EVS	(Verkehr)		4,30%	a	5,90%	5,80%	5,90%	5,80%	5,90%	0,13%
W08	Mean	200,48	173,19			174,08	174,16	174,38	174,08	174,38	0,3
Media	Rel. dev. to EVS	(Nachrichten)		-13,60%	c	-13,20%	-13,10%	-13,00%	-13,20%	-13,00%	0,15%
W09	Mean	770,16	662,9			666,23	667,1	666,52	666,23	667,1	0,87
Leisure, entertainment, culture	Rel. dev. to EVS	(Freizeit)		-13,90%	c	-13,50%	-13,40%	-13,50%	-13,50%	-13,40%	0,11%
W10	Mean	55,3	262,83			297,04	295,25	298,45	295,25	298,45	3,2
Education	Rel. dev. to EVS	(Bildung)		375,30%	c	437,20%	433,90%	439,70%	433,90%	439,70%	5,79%
W11	Mean	292,87	382,43			367,08	367,38	367,16	367,08	367,38	0,3
Lodging and catering services	Rel. dev. to EVS	(Beherbg)		30,60%	c	25,30%	25,40%	25,40%	25,30%	25,40%	0,10%
W12	Mean	289,17	249,66			250,29	250,74	250,67	250,29	250,74	0,46
Other goods and services	Rel. dev. to EVS	(AndWarDl)		-13,70%	a	-13,40%	-13,30%	-13,30%	-13,40%	-13,30%	0,16%

* weighted data by EVS internal variable HRB (results in N=35,.899,946 cases)

** cumulated survey out of LWR surveys 1999-2003 with alternative adjustment variants and cumulation weightings; price adjusted to 2003; re-calculated in EURO and quarter yearly values

Source: Cumulation survey CUMLWR 2003 out of Continuous Household Budget Surveys (LWR) 1999, 2001, 2002, 2003; Own calculations

5.2 Cumulation alternatives in comparison – Variances of private household expenditures

Whilst a preferably small deviation of the cumulation file and comparison file means is desirable according to an unbiased estimator, this is discussable when the deviation of their variances is regarded. In order to better meet the expenditure heterogeneity in the population, it could be argued that even a greater variance in the cumulation file is better than in the larger comparison file (EVS). This is an argument in particular for durable goods expenditures which are bought more seldom and thus might be captured to a lesser extent in a cross-section sample.

However, we adhere to the argument that the larger sample is ‚nearer' to the population than the smaller cumulation file; a smaller deviation of the variances between the cumulation file and the comparison file (EVS) then is seen as the desired property.

Since we give the unbiasedness property a higher value than the minimum variance property, the following variance results will be discussed less detailed based on the Evaluation Tables 7, 8, 9 and summarized by Table 10.

Total expenditures: comparison of variances

The standard deviations of all cumulation alternatives are close together and depart from the EVS-deviation only by -6% up to 5.4%. In contrast, the LWR 2003 overestimated the deviation of the total expenditures by 33%.

The cumulation file CUMLWR for all cumulation alternatives is remarkable better than the continuous household budget survey LWR at the survey period of the EVS.

Which cumulation weighting (depreciation rate) is in favour?
Single expenditure categories: comparison of variances

As addressed within the evaluation of mean expenditures, the variance of education expenditures in our EVS sub-sample is disproportional low. We therefore disregard this expenditure category in the following discussion.

Table 7: **Comparison of variances of selected consumption expenditures: CUMLWR 2003 versus EVS 2003 and LWR 2003 – Adjustment variant I**

Standard deviation		EVS 2003* (Var. Name)	LWR 2003	LWR-EVS	CUMLWR 2003 ** Adjustment variant I						
					Ia	Ib	Ic	Id	Minimum	Maximum	Max-Min
Sum W01 to W12	Standard deviation	4.382,21	5.865,27		4.119,83	4.130,58	4.145,27	4.124,38	4.119,83	4.145,27	25,44
Total expenditures	Rel. dev. to EVS			33,80%	-6,00%	-5,70%	-5,40%	-5,90%	-6,00%	-5,40%	0,58%
W01	Standard deviation	450,02	436,95		396,19	401,43	399,66	397,85	396,19	401,43	5,24
Food and alcohol free beverages	Rel. dev. to EVS	(NG)		-2,90%	-12,00%	-10,80%	-11,20%	-11,60%	-12,00%	-10,80%	1,16%
W02	Standard deviation	165,23	186,7		179,97	170,53	170,47	176,91	170,47	179,97	9,5
Alcohol beverages and tobacco	Rel. dev. to EVS	(AlkGT)		13,00%	8,90%	3,21%	3,20%	7,10%	3,20%	8,90%	5,75%
W03	Standard deviation	339,62	425,26		327,59	326,06	321,8	327,01	321,8	327,59	5,79
Clothing and shoes	Rel. dev. to EVS	(BeklSch)		25,20%	-3,50%	-4,00%	-5,20%	-3,70%	-5,20%	-3,50%	1,71%
W04	Standard deviation	1.431,80	1.011,63		880,84	879,43	866,02	879,8	866,02	880,84	14,81
Accomodation, water, electricity, gas, etc.	Rel. dev. to EVS	(W04selbst)		-29,30%	-38,50%	-38,60%	-39,50%	-38,60%	-39,50%	-38,50%	1,03%
W05	Standard deviation	892,86	1.603,00		1.007,00	1.005,77	961,66	1.001,75	961,66	1.007,00	45,34
Equipment, instruments, devices etc.	Rel. dev. to EVS	(W05selbst)		79,50%	12,80%	12,60%	7,70%	12,20%	7,70%	12,80%	5,08%
W06	Standard deviation	796,36	1.392,08		683,62	724,59	755,11	699,72	683,62	755,11	71,49
Health	Rel. dev. to EVS	(Gesundpflege)		74,80%	-14,20%	-9,00%	-5,20%	-12,10%	-14,20%	-5,20%	8,98%
W07	Standard deviation	2.411,58	4.558,18		2.422,22	2.417,26	2.452,11	2.423,68	2.417,26	2.452,11	34,85
Transport	Rel. dev. to EVS	(Verkehr)		89,00%	0,40%	0,20%	1,70%	0,50%	0,20%	1,70%	1,45%
W08	Standard deviation	158,21	156,24		127,07	131,57	132,08	128,63	127,07	132,08	5,01
Media	Rel. dev. to EVS	(Nachrichten)		-1,20%	-19,70%	-16,80%	-16,50%	-18,70%	-19,70%	-16,50%	3,17%
W09	Standard deviation	866,61	1.286,38		848,31	878,54	877,97	858,92	848,31	878,54	30,23
Leisure, entertainment, culture	Rel. dev. to EVS	(Freizeit)		48,40%	-2,10%	1,40%	1,30%	-0,90%	-2,10%	1,40%	3,49%
W10	Standard deviation	177,1	314,94		760,59	816,42	765,04	780,94	760,59	816,42	55,83
Education	Rel. dev. to EVS	(Bildung)		77,80%	329,50%	361,00%	332,00%	341,00%	329,50%	361,00%	31,52%
W11	Standard deviation	404,9	692,91		467,57	471,96	472,72	468,93	467,57	472,72	5,15
Lodging and catering services	Rel. dev. to EVS	(Beherbg)		71,10%	15,50%	16,60%	16,70%	15,80%	15,50%	16,70%	1,27%
W12	Standard deviation	417,58	642,48		433,86	405,93	403,31	424,77	403,31	433,86	30,55
Other goods and services	Rel. dev. to EVS	(AndWarDl)		53,90%	3,90%	-2,80%	-3,40%	1,70%	-3,40%	3,90%	7,32%
W0312226	Standard deviation	n/a***	105,41		97,78	96,42	95,66	97,13	95,66	97,78	2,12
Ladies pants (long, cotton)	Rel. dev. to EVS										
W0913011	Standard deviation	n/a***	1.671,28		1.630,60	1.606,43	1.575,59	1.617,93	1.575,59	1.630,60	55,01
PC- complete systems and notebooks	Rel. dev. to EVS										

*   weighted data by EVS internal variable HRB (results in N=35,.899,946 cases)
**   cumulated survey out of LWR surveys 1999-2003 with alternative adjustment variants and cumulation weightings; price adjusted to 2003; re-calculated in EURO and quarter yearly values
Source: Cumulation survey CUMLWR 2003 out of Continuous Household Budget Surveys (LWR) 1999, 2001, 2002, 2003; Own calculations

Table 8:   **Comparison of variances of selected consumption expenditures: CUMLWR 2003 versus EVS 2003 and LWR 2003 – Adjustment variant II**

Standard deviation		EVS 2003* (Var. Name)	LWR 2003	LWR-EVS	CUMLWR 2003 ** Adjustment variant II						
					IIa	IIb	IIc	IId	Minimum	Maximum	Max-Min
Sum W01 to W12	Standard deviation	4.382,21	5.865,27		4.133,02	4.151,34	4.164,36	4.140,25	4.133,02	4.164,36	31,34
Total expenditures	Rel. dev. to EVS			33,80%	-5,70%	-5,30%	-5,00%	-5,50%	-5,70%	-5,00%	0,72%
W01	Standard deviation	450,02	436,95		398,01	402,25	400,15	399,35	398,01	402,25	4,24
Food and alcohol free beverages	Rel. dev. to EVS	(NG)		-2,90%	-11,60%	-10,60%	-11,10%	-11,30%	-11,60%	-10,60%	0,94%
W02	Standard deviation	165,23	186,7		178,46	170,08	170,28	175,69	170,08	178,46	8,38
Alcohol beverages and tobacco	Rel. dev. to EVS	(AlkGT)		13,00%	8,00%	2,90%	3,10%	6,30%	2,90%	8,00%	5,07%
W03	Standard deviation	339,62	425,26		327,83	326,93	322,67	327,43	322,67	327,83	5,16
Clothing and shoes	Rel. dev. to EVS	(BeklSch)		25,20%	-3,50%	-3,70%	-5,00%	-3,60%	-5,00%	-3,50%	1,52%
W04	Standard deviation	1.431,80	1.011,63		881,23	883,24	870,13	881,23	870,13	883,24	13,11
Accomodation, water, electricity, gas, etc.	Rel. dev. to EVS	(W04selbst)		-29,30%	-38,50%	-38,30%	-39,20%	-38,50%	-39,20%	-38,30%	0,92%
W05	Standard deviation	892,86	1.603,00		1.024,72	1.028,66	979,46	1.021,20	979,46	1.028,66	49,2
Equipment, instruments, devices etc.	Rel. dev. to EVS	(W05selbst)		79,50%	14,80%	15,20%	9,70%	14,40%	9,70%	15,20%	5,51%
W06	Standard deviation	796,36	1.392,08		669,35	714,14	746,03	685,58	669,35	746,03	76,67
Health	Rel. dev. to EVS	Gesundpflege)		74,80%	-15,90%	-10,30%	-6,30%	-13,90%	-15,90%	-6,30%	9,63%
W07	Standard deviation	2.411,58	4.558,18		2.425,31	2.429,86	2.466,34	2.430,00	2.425,31	2.466,34	41,03
Transport	Rel. dev. to EVS	(Verkehr)		89,00%	0,60%	0,80%	2,30%	0,80%	0,60%	2,30%	1,70%
W08	Standard deviation	158,21	156,24		128,65	132,67	132,79	130,03	128,65	132,79	4,15
Media	Rel. dev. to EVS	(Nachrichten)		-1,20%	-18,70%	-16,10%	-16,10%	-17,80%	-18,70%	-16,10%	2,62%
W09	Standard deviation	866,61	1.286,38		856,57	883,17	880,4	866,12	856,57	883,17	26,6
Leisure, entertainment, culture	Rel. dev. to EVS	(Freizeit)		48,40%	-1,20%	1,90%	1,60%	-0,10%	-1,20%	1,90%	3,07%
W10	Standard deviation	177,1	314,94		723,8	780,63	742,42	744,12	723,8	780,63	56,83
Education	Rel. dev. to EVS	(Bildung)		77,80%	308,70%	340,80%	319,20%	320,20%	308,70%	340,80%	32,09%
W11	Standard deviation	404,9	692,91		469,63	475,01	475,17	471,26	469,63	475,17	5,54
Lodging and catering services	Rel. dev. to EVS	(Beherbg)		71,10%	16,00%	17,30%	17,40%	16,40%	16,00%	17,40%	1,37%
W12	Standard deviation	417,58	642,48		448,36	412,14	406,83	437,43	406,83	448,36	41,53
Other goods and services	Rel. dev. to EVS	(AndWarDl)		53,90%	7,40%	-1,30%	-2,60%	4,80%	-2,60%	7,40%	9,95%
W0312226	Standard deviation	n/a***	105,41		97,59	96,37	95,6	96,98	95,6	97,59	1,99
Ladies pants (long, cotton)	Rel. dev. to EVS										
W0913011	Standard deviation	n/a***	1.671,28		1.634,80	1.606,20	1.571,39	1.620,26	1.571,39	1.634,80	63,41
PC- complete systems and notebooks	Rel. dev. to EVS										

*   weighted data by EVS internal variable HRB (results in N=35,.899,946 cases)
**   cumulated survey out of LWR surveys 1999-2003 with alternative adjustment variants and cumulation weightings; price adjusted to 2003; re-calculated in EURO and quarter yearly values
Source: Cumulation survey CUMLWR 2003 out of Continuous Household Budget Surveys (LWR) 1999, 2001, 2002, 2003; Own calculations

Table 9: Comparison of variances of selected consumption expenditures: CUMLWR 2003 versus EVS 2003 and LWR 2003 – Adjustment variant III

Standard deviation		EVS 2003* (Var. Name)	LWR 2003	LWR-EVS	CUMLWR 2003 ** Adjustment variant III				Minimum	Maximum	Max-Min
					IIIa	IIIb	IIIc	IIId			
Sum W01 to W12	Standard deviation	4.382,21	5.865,27		4.126,40	4.136,73	4.151,49	4.130,94	4.126,40	4.151,49	25,09
Total expenditures	Rel. dev. to EVS			33,80%	-5,80%	-5,60%	-5,30%	-5,70%	-5,80%	-5,30%	0,57%
W01	Standard deviation	450,02	436,95		397,92	403,58	402,02	399,75	397,92	403,58	5,66
Food and alcohol free beverages	Rel. dev. to EVS	(NG)		-2,90%	-11,60%	-10,30%	-10,70%	-11,20%	-11,60%	-10,30%	1,26%
W02	Standard deviation	165,23	186,7		181,34	171,06	170,78	178,04	170,78	181,34	10,56
Alcohol beverages and tobacco	Rel. dev. to EVS	(AlkGT)		13,00%	9,70%	3,53%	3,40%	7,80%	3,40%	9,70%	6,39%
W03	Standard deviation	339,62	425,26		327,98	326,45	322,15	327,43	322,15	327,98	5,83
Clothing and shoes	Rel. dev. to EVS	(BeklSch)		25,20%	-3,40%	-3,90%	-5,10%	-3,60%	-5,10%	-3,40%	1,72%
W04	Standard deviation	1.431,80	1.011,63		882,59	882,98	870,24	882,22	870,24	882,98	12,75
Accomodation, water, electricity, gas, etc.	Rel. dev. to EVS	(W04selbst)		-29,30%	-38,40%	-38,30%	-39,20%	-38,40%	-39,20%	-38,30%	0,89%
W05	Standard deviation	892,86	1.603,00		1.010,76	1.004,77	959,44	1.003,88	959,44	1.010,76	51,32
Equipment, instruments, devices etc.	Rel. dev. to EVS	(W05selbst)		79,50%	13,20%	12,50%	7,50%	12,40%	7,50%	13,20%	5,75%
W06	Standard deviation	796,36	1.392,08		674,81	717,5	749,03	691,12	674,81	749,03	74,22
Health	Rel. dev. to EVS	(Gesundpflege)		74,80%	-15,30%	-9,90%	-5,90%	-13,20%	-15,30%	-5,90%	9,32%
W07	Standard deviation	2.411,58	4.558,18		2.424,60	2.420,19	2.455,25	2.426,26	2.420,19	2.455,25	35,06
Transport	Rel. dev. to EVS	(Verkehr)		89,00%	0,50%	0,40%	1,80%	0,60%	0,40%	1,80%	1,45%
W08	Standard deviation	158,21	156,24		127,63	132,2	132,76	129,22	127,63	132,76	5,13
Media	Rel. dev. to EVS	(Nachrichten)		-1,20%	-19,30%	-16,40%	-16,10%	-18,30%	-19,30%	-16,10%	3,24%
W09	Standard deviation	866,61	1.286,38		848,95	878,31	876,79	859,31	848,95	878,31	29,36
Leisure, entertainment, culture	Rel. dev. to EVS	(Freizeit)		48,40%	-2,00%	1,40%	1,20%	-0,80%	-2,00%	1,40%	3,39%
W10	Standard deviation	177,1	314,94		767,56	822,95	773,61	788,02	767,56	822,95	55,39
Education	Rel. dev. to EVS	(Bildung)		77,80%	333,40%	364,70%	336,80%	345,00%	333,40%	364,70%	31,27%
W11	Standard deviation	404,9	692,91		467,54	472,22	472,96	468,98	467,54	472,96	5,42
Lodging and catering services	Rel. dev. to EVS	(Beherbg)		71,10%	15,50%	16,60%	16,80%	15,80%	15,50%	16,80%	1,34%
W12	Standard deviation	417,58	642,48		436,48	407,13	404,03	427,05	404,03	436,48	32,44
Other goods and services	Rel. dev. to EVS	(AndWarDI)		53,90%	4,50%	-2,50%	-3,20%	2,30%	-3,20%	4,50%	7,77%
W0312226	Standard deviation	n/a***	105,41		97,9	96,58	95,81	97,26	95,81	97,9	2,09
Ladies pants (long, cotton)	Rel. dev. to EVS										
W0913011	Standard deviation	n/a***	1.671,28		1.627,26	1.601,44	1.571,55	1.614,02	1.571,55	1.627,26	55,7
PC- complete systems and notebooks	Rel. dev. to EVS										

*   weighted data by EVS internal variable HRB (results in N=35,.899,946 cases)

**   cumulated survey out of LWR surveys 1999-2003 with alternative adjustment variants and cumulation weightings; price adjusted to 2003; re-calculated in EURO and quarter yearly values

Source: Cumulation survey CUMLWR 2003 out of Continuous Household Budget Surveys (LWR) 1999, 2001, 2002, 2003; Own calculations

Table 10: Comparison of variances of selected consumption expenditures: CUMLWR 2003 versus EVS 2003 and LWR 2003 – Best results of adjustment variants I, II and III

Standard deviation		EVS 2003* (Var.-name)	LWR 2003	LWR-EVS	CUMLWR Cumulation weightings c,a	2003 ** I	II	III	Minimum	Maximum	Max-Min
Sum W01 to W12	Mean	4.382,21	5.865,27			4.145,27	4.164,36	4.151,49	4.145,27	4.164,36	19,09
Total expenditures	Rel. dev. to EVS			33,80%	c	-5,40%	-5,00%	-5,30%	-5,40%	-5,00%	0,44%
W01	Mean	450,02	436,95			399,66	400,15	402,02	399,66	402,02	2,36
Food and alcohol free beverages	Rel. dev. to EVS	(NG)		-2,90%	c	-11,20%	-11,10%	-10,70%	-11,20%	-10,70%	0,52%
W02	Mean	165,23	186,7			170,47	170,28	170,78	170,28	170,78	0,5
Alcohol beverages and tobacco	Rel. dev. to EVS	(AlkGT)		13,00%	c	3,20%	3,10%	3,40%	3,10%	3,40%	0,30%
W03	Mean	339,62	425,26			321,8	322,67	322,15	321,8	322,67	0,87
Clothing and shoes	Rel. dev. to EVS	(BeklSch)		25,20%	c	-5,20%	-5,00%	-5,10%	-5,20%	-5,00%	0,26%
W04	Mean	1.431,80	1.011,63			866,02	870,13	870,24	866,02	870,24	4,21
Accomodation, water, electricity, gas, etc.	Rel. dev. to EVS	(W04selbst)		-29,30%	c	-39,50%	-39,20%	-39,20%	-39,50%	-39,20%	0,29%
W05	Mean	892,86	1.603,00			961,66	979,46	959,44	959,44	979,46	20,02
Equipment, instruments, devices etc.	Rel. dev. to EVS	(W05selbst)		79,50%	c	7,70%	9,70%	7,50%	7,50%	9,70%	2,24%
W06	Mean	796,36	1.392,08			683,62	669,35	674,81	669,35	683,62	14,27
Health	Rel. dev. to EVS	(Gesundpflege)		74,80%	a	-14,20%	-15,90%	-15,30%	-15,90%	-14,20%	1,74%
W07	Mean	2.411,58	4.558,18			2.422,22	2.425,31	2.424,60	2.422,22	2.425,31	3,09
Transport	Rel. dev. to EVS	(Verkehr)		89,00%	a	0,40%	0,60%	0,50%	0,40%	0,60%	0,16%
W08	Mean	158,21	156,24			132,08	132,79	132,76	132,08	132,79	0,71
Media	Rel. dev. to EVS	(Nachrichten)		-1,20%	c	-16,50%	-16,10%	-16,10%	-16,50%	-16,10%	0,45%
W09	Mean	866,61	1.286,38			877,97	880,4	876,79	876,79	880,4	3,62
Leisure, entertainment, culture	Rel. dev. to EVS	(Freizeit)		48,40%	c	1,30%	1,60%	1,20%	1,20%	1,60%	0,42%
W10	Mean	177,1	314,94			765,04	742,42	773,61	742,42	773,61	31,19
Education	Rel. dev. to EVS	(Bildung)		77,80%	c	332,00%	319,20%	336,80%	319,20%	336,80%	17,61%
W11	Mean	404,9	692,91			472,72	475,17	472,96	472,72	475,17	2,45
Lodging and catering services	Rel. dev. to EVS	(Beherbg)		71,10%	c	16,70%	17,40%	16,80%	16,70%	17,40%	0,61%
W12	Mean	417,58	642,48			433,86	448,36	436,48	433,86	448,36	14,5
Other goods and services	Rel. dev. to EVS	(AndWarDl)		53,90%	a	3,90%	7,40%	4,50%	3,90%	7,40%	3,47%

* weighted data by EVS internal variable HRB (results in N=35,.899,946 cases)

** cumulated survey out of LWR surveys 1999-2003 with alternative adjustment variants and cumulation weightings; price adjusted to 2003; re-calculated in EURO and quarter yearly values

Source: Cumulation survey CUMLWR 2003 out of Continuous Household Budget Surveys (LWR) 1999, 2001, 2002, 2003; Own calculations

The frequencies of the best cumulation weightings according to the three adjustment variants out of the total expenditures and the eleven expenditure categories are summarized in Table 11. As by the mean evaluation – though not as in a similar uniqueness – the exponential weighting is comparably the best weighting procedure in all three adjustment variants (single results are given in Tables 7, 8, 9 and 10)

Table 11: Comparison of variances: Frequencies of best cumulation weightings of total expenditures and all eleven expenditure categories* – CUMLWR 2003 to EVS 2003

Frequencies of smallest variance deviations*			
	Adjustment variant		
Cumulation weighting	I	II	III
a uniform	3	3	3
b linear progressive	2	4	2
c exponential progressive	5	4	5
d data generated cluster analytic	2	1	2

*without education
Source: Cumulation survey CUMLWR 2003 out of Continuous Household Budget Surveys (LWR) 1999, 2001, 2002, 2003; Own calculations

Yet, the discrepancy between all cumulation weightings in CUMLWR 2003 is small. In most instances (9 out of 12) all CUMLWR 2003 deviations meet the EVS deviations by far better than the deviations of the single LWR 2003.

To record: According to the variances as well as to the means for the respective expenditure categories there is a distinct information gain by the cumulation of cross-section surveys and our cumulation approach.

Following the best cumulation alternatives form the mean evaluation – exponential weighting (c) and uniform weighting (a) throughout all adjustment variants I, II and III (Table 6) – at first also according to the variances the differences between all adjustment variants are small (all ranges are between 0.16 and 3.5 percentage points (Table 10).

However, the variance goodness of fit of the cumulation alternatives is dependent from the single expenditure categories. Expenditures for health and other goods and services and here equipment expenditures fit less. Nonetheless, with a maximum difference of 7.4% the CUMLWR 2003 variances are still near to the EVS variances. Further single results can be found in the Tables above.

To summarize: The best cumulation weightings with respect to their variances are

c: exponential weighting with a strong weighting of the actual survey

a: uniform weighting of all surveys according to the expenditure category.

Which adjustment variant is in favour?
Single expenditure categories: comparison of variances

As mentioned, all three adjustment variants (I, II and III) show similar variances (Table 10); a result similar to the mean evaluation. With respect to the computational burden, the near variance results and according to the best results in several expenditure categories, again the best variance goodness of fit is recommended by

adjustment variant III: cumulation weighting with alternative depreciation

rates *after* the final adjustment of the entire cumulation sample

CUMLWR 2003 better than LWR 2003?
Single expenditure categories: comparison of variances

Conspicuous are the relative large percentage differences between the EVS 2003 and the LWR 2003; there are differences up to 89% (Transport, Table 10). The differences of all cumulation weightings by CUMLWR 2003, for instance, are less than 1%. Also for other expenditure categories, though not as impressive, the message is: There is a distinct information gain by the cumulation of cross-sectional surveys with CUMLWR 2003.

To take into account the information of former cross-sections enhances the results of the single Continuous Household Budget Survey and does ask for a cumulation of surveys.

8   Conclusion: Evaluation in summary and perspectives

Within the project „Official Statistics and Socio-economic Questions" of the German Federal Statistical Office and embedded into the new EU-SILC approaches we developed a concept for cumulating cross-section surveys and realized and evaluated the approach for the cumulation of five (respective four) cross sections of the Continuous Household Budget Surveys 1999, (2000), 2001, 2002 and 2003.

The aim was to provide in depth structural data out of an appropriate linkage and to analyse the information gain compared to another large survey (Ehling 2002). To meet this aim we compared individual household expenditures of the cumulated survey CUMLWR 2003 with the expenditures of the Sample Survey of Income and Expenditures 2003 (EVS

2003) as the large survey as well with the expenditures of the single Continuous Household Budget Surves at the period of the large survey 2003.

The theoretical foundation and cumulation brick stones were discussed in Merz 2004 and further developed in the study at hand.

Beyond the development of appropriate cumulation weightings, which incorporate the information of former cross sections, the central task of a structural new demographic adjustment was realized by an adjustment procedure based on information theory (Minimum Information Loss (MIL) Principle). The particular advantage of this procedure is the theoretical based structural and representative adjustment also to hierarchical microdata by a simultaneous approach. In addition to the demographic adjustment 'economic multiplies' ('inflators') for considering price changes and item related statements (economic variables in relation to period dependent means) were embraced and are possible.

We analyzed three adjustment variants (I, II and III)

(I) Cumulation weighting *before* (final) adjustment *with* previous adjustment to period specific totals r_t

(II) Cumulation weighting *before* (final) adjustment *without* previous adjustment to period specific totals r_t

(III) Cumulation weighting *after* adjustment with adjustment only to totals r_T at period T

with respective four alternative cumulation weightings

a uniform, b linear progressive, c exponential, d data generated cluster analytic.

The twelve cumulation alternatives allocate an individual weight to each survey household in each survey period within the aggregated cumulation file. This cumulation file CUMLWR 2003 embodies all five respective four Continuous Household Budget Surveys from 1999 till 2003 with 30,480 respective 24,311 data records.

Result: Comparison CUMLWR 2003 with EVS 2003 and LWR 2003

With respect to the efficiency of an estimator (unbiased results and minimum variance as desired estimation properties) we have chosen as evaluation criteria for the goodness of fit of CUMLWR 2003 to the EVS 2003 and LWR 2003 the arithmetic mean and variance for twelve selected expenditure categories of private consumption.

The results for the total expenditures as well as for the single expenditure categories all twelve cumulation alternatives are close together and meet the EVS 2003 values better compared to the single LWR 2003. This holds for the mean value as well as for the variance comparison. The developed cumulation approach thus produces a distinct information gain so measured.

Cumulation weights: With all three adjustment variants (I, II and III) and with respect to the mean and variance indicator an identical pattern of the best cumulation weights with *c: exponential* and *a: uniform weighting* is given. The goodness of fit, however, is dependent of the single expenditure category: a uniform cumulation weighting, not accounting for some period dependent depreciation (some 'habit persistence'), results in better values for health, transport and other goods and services expenditures. For all other eight expenditure categories the exponential weighting is in favour and recommends a strong period depending depreciation.

Adjustment variants: All three adjustment variants with their four alternative cumulation weightings result in a similar way. If the computational burden is accounted for an additional evaluation criteria, then the adjustment variant III is the best variant, which at first adjusts the entire cumulation file at T=2003 and then allocates alternative cumulation weightings to each cross section. This variant also allows another ex post cross section weighting without the necessity of a new demographic adjustment (as in variants I and II). These results are hold by the mean and variance evaluation.

Result: Method comparison of alternative cumulation factors

Each cumulation factor comprises the cumulation weight of the respective cross section and the adjustment factor of a certain adjustment variant. Such a cumulation factor finally is the number of microunits (here households) in the population (here Germany) which is represented by one household in the sample (here CUMLWR 2003). Result: all cumulation factors of all twelve cumulation alternatives are close together between the first and third quartile of the distribution. Different maximum values, however, are pointing to necessary adjustments of strong underrepresented groups in the sample.

Our cumulation weightings comprise fixed approaches (a uniform, b linear progressive, c exponential) as well as a data generated approach. The data generated approach (without the model based methods) in fact has to be favoured because of its theoretical foundation. However, the computational burden of the cluster analytical approach is expensive. As a data

generated approach the data itselves define the depreciation rates, the valuation of the former cross sections. On the other hand, the fixed alternatives are transparent and reveal the evaluation criteria of the user.

General conclusion

The adjustment variant III, the adjustment of the aggregated cumulation file to its totals before a cumulation weighting has to be proven the best cumulation alternative for the description of private household expenditures. Dependent on the expenditure category, the exponential weighting (c) with a high weight of the most actual cross section and information as well as the uniform weighting (a) of all former cross sections and information show to be the best cumulation weightings.

With the recommended cumulation alternative by our analysis, the adjustment variant III, and by using of a powerful and efficient adjustment procedure (like the MIL principle and the available associated ADJUST-Software, http://ffb.uni-lueneburg.de/adjust) it is relatively easy to adjust an aggregated cumulation file of different cross sections to actual demographic totals. After such a demographic adjustment, the single adjustment factors can easily be modified by the chosen alternative cumulation weights delivering final cumulation factors. Then analyses by content might examine variables of interest – here expenditures of private households – which are representative by its cumulation factors and informative by all the cross sections behind.

Our cumulation concept results in a distinct information gain by cumulation of cross section surveys. The results also indicate that a socio-economic model based extrapolation of former cross sections will lead to further information gains. This has to be reserved to further consumption analyses of single expenditure categories by content and theory. Panel data, with repeated information of the same interviewed household will further enhance the possibilities of targeted microanalyses.

References:

Cassel, C.M., Granström, F., Lundquist, P. und J. Selen, (1997), Kumulierung von Daten aus Erhebungen über die Wirtschaftsrechnung privater Haushalte, EUROSTAT-Arbeitsgruppe 'Wirtschaftsrechnungen privater Haushalte', DOK E2/WPH/109/97

Cochran, W.G. (1977), Sampling Techniques, New York.

Deaton, A. (1990), Demand Analysis, in: Griliches, Z. und M.D. Intriligator, eds., Handbook of Econometrics, Chapt. 30, 1768-1839

Deaton, A. and J. Muellbauer (1980), An Almost Ideal Demand System, in: The American Economic Review, Vol. 70(3), 312-326.

Ehling, M. (2002a), Considerations on Changing the Sytem of Household Statistics in Germany, in: Statistisches Bundesamt (Hg.), Rotierende Stichproben – Datenkumulation und Datenqualität, Spektrum der Bundesstatistik, Band 21, Stuttgart, 5-10.

Ehling, M. (2002b), Weiterentwicklung der Wirtschaftsrechnungen und Integration neuer Datenanforderungen der Europäischen Union (EU), in: Statistisches Bundesamt (Hg.), Rotierende Stichproben – Datenkumulation und Datenqualität, Spektrum der Bundesstatistik, Band 21, Stuttgart, 20-34.

Golan, A., Judge, G. und D. Miller (1996), Maximum Entropy Econometrics: Robust Estimation with Limited Data, New York: John Wiley & Sons Inc.

Kish, L. (1999), Cumulating/Combining population surveys, in: Survey Methodology, Vol. 25, No. 2, 129-138

Kordos, J. (2002), Estimation Problems and Data Quality, in: Statistisches Bundesamt (Hg.), Rotierende Stichproben – Datenkumulation und Datenqualität, Spektrum der Bundesstatistik, Band 21, Stuttgart, 57-71.

Krug, W., Ehling, M. und N. Ernst (2002), Ergebnis der Abschlussdiskussion, in: Statistisches Bundesamt (Hg.), Rotierende Stichproben – Datenkumulation und Datenqualität, Spektrum der Bundesstatistik, Band 21, Stuttgart, 18-19.

Lluch, C. (1973), The Extended Linear Expenditure System, in: European Economic Review, 4, 21-32.

Merz, J. (1980), Prognosegüte und Spektraleigenschaften ökonomischer Modelle, in: Stöppler, S. (Hg.), Dynamische ökonomische Analyse und Steuerung, Wiesbaden 1980b, 2. Auflage, 31-66

Merz, J. (1983a), Die konsistente Hochrechnung von Mikrodaten nach dem Prinzip des minimalen Informationsverlustes, in: Allgemeines Statistisches Archiv, 76, 4, 342-366.

Merz, J. (1983b), FELES: The Functionalized Extended Linear Expenditure System – Theory, Estimation Procedures and Application to Individual Household Consumption Expenditures involving Socioeconomic and Sociodemographic Characteristics, in: European Economic Review, Vol. 23, 359-394

Merz, J. (1994), Microdata Adjustment by the Minimum Information Loss Principle, FFB-Diskussionspapier Nr. 10, Forschungsinstitut Freie Berufe, Fachbereich Wirtschafts- und Sozialwissenschaften, Universität Lüneburg

Merz, J. (2004), Kumulation von Mikrodaten – Konzeptionelle Grundlagen und ein Vorschlag zur Realisierung, in: Allgemeines Statistisches Archiv, 88, 451-472

Merz, J. und H. Stolze (2004), Adjust for Windows Version 1.1 – A Software Package to Achieve Representative Microdata by the Minimum Information Loss Principle, Program Manual, Lüneburg (http://ffb.uni-lueneburg.de/adjust).

Merz, J. und H. Stolze (2008), Representative Time Use Data and New Harmonised Calibration of the American Heritage Time Use Data (AHTUD) 1965-1999, in: electronic International Journal of Time Use Research, 5/1, 90-126.

Selén, J. (2002), Some comments on the use of cumulated surveys and a presentation of a model-based calibration, in: Statistisches Bundesamt (Hg.), Rotierende Stichproben – Datenkumulation und Datenqualität, Spektrum der Bundesstatistik, Band 21, Stuttgart, 57-71.

Statistisches Bundesamt (Hg.) (2002), Rotierende Stichproben – Datenkumulation und Datenqualität, Spektrum der Bundesstatistik, Band 21, Stuttgart.

Statistisches Bundesamt, Auswahlfiles der laufenden Wirtschaftsrechnungen 1999, 2000, 2001, 2002, 2003 und der Sample Survey of Income and Expenditures 2003, Wiesbaden.

Stone, J.R.N. (1954), Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of Britisch Demand, in: Economic Journal, 64, 511-527

Verma, V. (2002), The Case for a Continuous Household Budget Survey, in: Statistisches Bundesamt (Hg.), Rotierende Stichproben – Datenkumulation und Datenqualität, Spektrum der Bundesstatistik, Band 21, Stuttgart, 51-56.