Computational methods for optimum stratification: A review

Er, Sebnem
Istanbul University, Quantitative Methods Department
Avcilar
Istanbul (34320),Turkey
E-mail: er.sebnem@gmail.com

Abstract

Over the past number of years, many computational procedures have been developed to obtain optimum stratification boundaries, those that give the minimum variance (for example, Keskinturk & Er (2007), Kozak (2004), Lavallée & Hidiroglou (1988)). In this paper, I examine these methods in terms of efficiency, ease-of-use and success in arriving at the optimum boundaries, and compare them with each other using the statistical computing system R. Keskinturk & Er's method is available in the GA4Stratification package whereas the other two methods are available in the stratification package in R. The methods are applied to the data set available in these packages as well as the data set used in Kozak's (2004) paper. For a given set of circumstances (e.g. skewed populations, take-all stratum, number of strata, precision level), there appears to be a unique optimum set of bounds, and apart from some difficulties with reaching a local optimum rather than a global optimum, the numerical algorithms considered appear to reach the optimum set of bounds if allowed to run long enough. Given the availability of high powered computing, the number of iterations required to reach the optimum does not pose a problem.
Keywords: Stratified Sampling, Optimum boundaries.

1. Introduction

In stratified sampling, a heterogeneous population is divided into subpopulations, each of which is internally homogeneous in order to gain more precision than other methods of sampling. As a result the main problem arising in stratified sampling is to obtain optimum boundaries. Several numerical and computational methods have been developed for this purpose. Some apply to highly skewed populations and some apply to any kind of populations. An early and very simple method is the cumulative square root of the frequency method (cum$\sqrt{f}$) of Dalenius & Hodges. More recently Lavallée & Hidiroglou algorithm and Gunning & Horgan's (2004) geometric method have been proposed for highly skewed populations whereas Kozak's (2004) random search method and Keskinturk & Er's (2007) genetic algorithm (GA) method have been proposed for even non-skewed populations. Very recently, Brito et.all (2010-a) proposed an exact algorithm for the stratification problem with only proportional allocation based on the concept of minimum path in graphs (Brito et.all, 2010-a, pp.190) and they called their method StratPath (Brito et.all, 2010-a, pp.190). Moreover, (Brito et.all, 2010-b, pp.190) developed an iterated local search method to solve the stratification problem of variables with any distribution with Neyman allocation Neyman (1934). All these methods aim to achieve the optimum boundaries that maximise the level of precision or equivalently minimise the variance of the estimate or the sample size required to reach a level of precision and some of them are available in the stratification package *stratification* for use with the statistical programming environment R; freely available on the Comprehensive R Archive Network (CRAN) at http://CRAN.R-project.org/package=stratification.
In this paper, a GA approach for boundary determination and sample allocation will be reviewed and then the use of GA4Stratification package *GA4Stratification*, that is freely available on CRAN at http://CRAN.R-project.org/package=GA4stratification will be explained in detail.

2. Determination of the stratum boundaries and sample sizes

2.1 An exact solution by Dalenius (1950)

Dalenius (1950) considers a density $f(x)$ with mean

$$\mu = \int_{-\infty}^{\infty} t f(t) dt \tag{1}$$

The range $(x_{min} - x_{max})$ of the stratification variable x is divided into L parts at points $b_1 < b_2 < \ldots < b_{L-1}$, each part corresponding to a stratum. When a sample of $n = \sum n_h$ observations is selected from $f(x)$, the true mean

$$\mu = \sum_{h=1}^{L} W_h \mu_h \tag{2}$$

is estimated by Cochran (1977, pp.91-92)

$$\overline{x}_{st} = \sum_{h=1}^{L} W_h \overline{x}_h, \tag{3}$$

where for the hth stratum Cochran (1977, pp.90)

$$W_h = \int_{b_{h-1}}^{b_h} f(t) dt = \frac{N_h}{N}, \tag{4}$$

and the true mean Cochran (1977, pp.90)

$$\mu_h = \frac{\sum_{i=1}^{N_h} x_{hi}}{N_h}, \tag{5}$$

and the sample mean Cochran (1977, pp.90)

$$\overline{x}_h = \frac{\sum_{i=1}^{n_h} x_{hi}}{n_h}. \tag{6}$$

The estimate $\overline{x}_{st}$ has variance Cochran (1977, pp.92)

$$\sigma^2(\overline{x}_{st}) = \sum_{h=1}^{L} W_h^2 \frac{\sigma_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right), \tag{7}$$

where the true variance is

$$\sigma_h^2 = \frac{\sum_{i=1}^{N_h} (x_{hi} - \mu_h)^2}{N_h - 1}. \tag{8}$$

If the sampling fractions n_h/N_h are negligible then the variance could be written in short

$$\sigma^2(\overline{x}_{st}) = \sum_{h=1}^{L} W_h^2 \frac{\sigma_h^2}{n_h}. \tag{9}$$

It is well-known that the variance of the estimate given in (9) is minimum

$$\sigma^2_{min}(\overline{x}_{st}) = \frac{1}{n} \left(\sum_{h=1}^{L} W_h \sigma_h \right)^2, \tag{10}$$

when total sample size n is allocated using Neyman's optimum allocation method Neyman (1934):

$$n_h = n \frac{N_h \sigma_h}{\sum_{h=1}^{L} N_h \sigma_h}. \tag{11}$$

The variance of the estimate in (10) is a function of the boundaries b_h of stratification which makes it difficult to be minimised. Dalenius (1950) showed that when the boundaries satisfy the equations

$$\frac{\sigma_h^2 + (b_h - \mu_h)^2}{\sigma_h} = \frac{\sigma_{h+1}^2 + (b_h - \mu_{h+1})^2}{\sigma_{h+1}}, \tag{12}$$

then the variance with Neyman allocation given in (10) is the minimum among the variances of all possible boundary combinations. Dalenius equations given in (12) are computationally difficult to solve, since both μ_h and σ_h depend on b_h Cochran (1977, pp.128), and since then many approximate methods have been developed to solve (12).

2. Overview of the Approximate Stratification Methods

2.1 Dalenius & Hodges' (1959) cum$\sqrt{f}$ rule

Dalenius and Hodges' cum$\sqrt{f}$ rule is based on constructing equal intervals on the cumulative of the square roots of the frequencies of the stratification variable so that nearly optimum points are obtained (Dalenius & Hodges (1959), pp.88). They introduce the transformation (Dalenius & Hodges (1959), pp.90-91)

$$Y(u) = \int_{-\infty}^{u} \sqrt{f(t)}dt. \tag{13}$$

When $u \to \infty$, $Y(u)$ approaches an upper bound H such that,

$$Y(u) = \frac{h}{L}H,\ h = 1, \ldots, L-1 \tag{14}$$

When L is large, the strata will be narrow and therefore the distribution in each strata is approximately uniform, so that the variance within each stratum is (Dalenius & Hodges (1959), pp.91),

$$\sigma_h \approx \frac{b_h - b_{h-1}}{\sqrt{12}} \tag{15}$$

and the W_h in equation (4) is defined as

$$W_h \approx f_h\,(b_h - b_{h-1}) \tag{16}$$

Then, by the mean value theorem there exists a constant value f_h of $f(x)$ in stratum h, such that (Dalenius & Hodges (1959), pp.91)

$$\sqrt{12}\sum W_h\sigma_h \approx \sum \left[\sqrt{f_h}\,(b_h - b_{h-1})\right]^2 \approx \sum \left[Y_h - Y_{h-1}\right]^2 \tag{17}$$

Since $Y_L - Y_0$ is fixed, it is easy to verify that the sum on the right of (17) is minimized by making $Y_h - Y_{h-1}$ constant which is roughly equivalent to making $W_h\sigma_h$ constant (Cochran (1977) pp.130). Given $f(x)$, the rule is to form the cumulative of $\sqrt{f(x)}$ and choose the b_h so that it creates equal intervals on the cum$\sqrt{f(x)}$ scale (Cochran (1977) pp.129) to obtain an approximate minimum variance. An illustration of the method could be found in Cochran (1977) pp.129-130 and in Gunning, Horgan and Yancey (2004). Even though it is a very simple method, the main problem with it is the arbitrariness in deciding the value of number of classes. Furthermore, since the approximation highly depends on the uniform distribution within strata, Cochran (1961) cautions that it is advisable to have a substantial number of classes in the original frequency, otherwise the true optimum stratification may be missed and the calculation of the within-stratum boundaries becomes affected by grouping errors. Hedlin (2000) notes that the final stratum boundaries depend on the initial choice of the number of classes M, and that there is no theory which gives the best number of classes.

2.2. Lavallée & Hidiroglou (1988)

Lavallée & Hidiroglou (1988) derived an iterative algorithm for skewed populations, such that the sample size is minimised for a given level of precision expressed in terms of coefficient of variation and usually valued between 1% and 10% considering a take-all stratum (Lavallée & Hidiroglou (1988), pp.33). The algorithm is a modification to Sethi's (1963) method for stratifying a population. The method is numerically compared, in terms of boundary values and sample size, to the Dalenius & Hodges' (1959) cumulative square root f rule. The algorithm, which is recursive in nature, is simple to program and converges rapidly to the optimum boundary points. It also offers substantial savings in terms of sample size for given reliability criteria.
Lavallée & Hidiroglou tries to find such values (Lavallée & Hidiroglou (1988), pp.36)

$$b_1 < b_2 < \cdots < b_{L-1} \tag{18}$$

that minimizes n considering a take-all top stratum,

$$n = N_L + \left(\sum_{h=1}^{L-1} N_h^2 \sigma_h^2 / a_h\right)\left(N^2\mu^2 CV^2 + \sum_{h=1}^{L-1} N_h {\sigma_h}^2\right)^{-1}. \tag{19}$$

given the level of precision (CV) and the specified allocation scheme represented by a_h (Lavallée & Hidiroglou (1988), pp.36).
In their paper Lavallée & Hidiroglou (1988) mainly consider X-proportional power allocation:

$$a_h = \frac{(W_h\mu_h)^p}{\sum_{h=1}^{L-1}(W_h\mu_h)^p} \tag{20}$$

where $0 < p < \infty$. Putting (20) in (19)

$$n = NW_L + \frac{N\left[\sum_{h=1}^{L-1}(W_h\sigma_h)^2(W_h\mu_h)^p\right]\left[\sum_{h=1}^{L-1}(W_h\mu_h)^p\right]}{NCV^2\mu^2 + \sum_{h=1}^{L-1} W_h\sigma_h^2}. \tag{21}$$

is obtained. Since this function will be minimised in order to find the boundaries, the first derivatives are taken with respect to $b_1 < b_2 < \cdots < b_{L-1}$ and equated to zero. The resulting equations could be seen in the paper of Lavallée & Hidiroglou (1988) pp.37. Using Sethi's (1963) method, these equations are solved using an iterative method starting with some arbitrary boundaries till two consecutive sets are either identical or differ by negligible amounts (Lavallée & Hidiroglou (1988), pp.38-39.).
Lavallée & Hidiroglou (1988) observed that, under relatively simple assumptions, the coefficients of variation in each stratum tend to be equalised without a significant increase in the variance of the stratified sample mean, and went on to say that equality of coefficients of variation is often asked by users of the survey data.
It has been shown that numerical difficulties may occur when using the Lavallée & Hidiroglou algorithm. Detlefsen & Veum (1991) found that the resulting boundaries depend on where the initial boundaries are set, so that the minimum sample size attained is a local but not necessarily a global minimum. They also found that convergence occurs faster for the lower number of strata. Slanta & Krenzke (1994, 1996) encountered numerical difficulties when using the Lavallée & Hidiroglou algorithm with Neyman allocation. Rivest (2002) reported numerical difficulties, failure to reach the global minimum sample size, and non-convergence of the algorithm when the number of strata was large.

2.3. Kozak's (2004) Random Search Method

Kozak's (2004) Random Search Method tries to find such values

$$b_1 < b_2 < \cdots < b_{L-1} \tag{22}$$

that minimizes the objective function considering a take-all top stratum with Neyman allocation

$$n = N_L + \left(\sum_{h=1}^{L-1} W_h \sigma_h\right)^2 \left(\mu^2 CV^2 + \frac{1}{N}\sum_{h=1}^{L-1} W_h {\sigma_h}^2\right)^{-1}; \tag{23}$$

under the constraints $N_h \geq 2$ for $h = 1, 2, \ldots, L$ and $2 \leq n_h \leq N_h$ for $h = 1, 2, \ldots, L-1$, where n is the minimizing sample size required to achieve the given precision CV of the mean estimate of the strafication when the strata boundaries are $\mathbf{b} = (b_1, b_2, \ldots, b_{L-1})^T$.
Kozak's (2004) random search method chooses an initial set of strata boundaries and calculates the function values of n. Then for a certain number of iterations the following steps are repeated:

1. Generate the set of boundaries $\mathbf{b}'$ by choosing one stratum boundary b_h and changing it as follows:

   $$b_h{}' = b_h + j \tag{24}$$

   where j is the random integer, $j \in \langle -p; -1\rangle \cup \langle 1; p\rangle$ and p is a given integer according to the size of the population.

2. Calculate the function value of n'.

3. If the constraints are satisfied and $n' < n$, the new set of boundaries $b_h{}'$ are accepted.

The algorithm is finished if the stopping criteria is fulfilled. The details of this algorithm could be found in Kozak (2004). Since this method does not guarantee that the global minimum is achieved, the performance of the algorithm has been compared with other available methods. Kozak & Verma (2006) compared the efficiency of the optimization approach of Kozak (2004) (random search method) with the geometric stratification method using five positively skewed artificial populations of various sizes between $2,000$ and $10,000$ stratified into $4, 5, 6$ and 7 strata. They found that for all the populations, the optimization method was more efficient than the geometric method and the more strata constructed the greater the gain efficiency (Kozak & Verma (2006), pp.159). They pointed out that with Geometric method, some strata may be empty or/and sample sizes from some strata may be smaller than two or greater than their population sizes (Kozak & Verma (2006), pp.161).

2.4. Genetic Algorithm

Keskinturk & Er (2007) developed a genetic algorithm approach for the determination of stratum boundaries and sample sizes to be allocated among strata that minimises the objective function given in (7) with equal, proportional, Neyman and genetic algorithm allocation methods. Neyman and genetic algorithm results for the sample sizes happen to be the same when there is no need for a take-all top stratum.
In order to solve the stratification problem with GA, stratification values are encoded into chromosomes. The range of ascending values subject to stratification are divided into L parts by points $b_1, b_2, \ldots, b_{L-1}$, each such part corresponding to a stratum boundary. In this method binary encoding is used for boundary determination with equal, proportional and Neyman sample allocation methods, whereas both binary and realvalued encoding are used for GA boundary determination and sample allocation. The algorithm begins with a randomly produced generation where each chromosome is evaluated by the objective function, referred to as a fitness function, given in (7). The next step after determination of the fitness values is the selection process, where whether chromosomes will survive in

the next generation or not, according to their fitness values. Chromosomes with a better fitness value have more chance to survive. After selection, the parents are exchanged with a crossover probability to form new offspring and are mutated with a mutation probability to mutate new offspring. This process is repeated unless the stopping criterion is not reached. Finally the best solution in the current generation is returned as a solution for the problem of boundary determination and sample allocation. Further details on the method could be obtained in the paper of Keskinturk & Er (2007).
Keskinturk & Er (2007) compared their method's efficiency with the Geometric and cum$\sqrt{f}$ methods using the sales data of the first 500 largest corporations in Turkey, the 1975 population of the 284 Swedish municipalities that is now available in the stratification package R, 6 randomly generated data, 4 of which were generated from Chi-square distribution with different degrees of freedom, 1 of which was generated from a normal distribution and 1 of which was generated from the Beta distribution. They found that the best results for all of the numerical examples are obtained when both stratum boundaries and strata sample sizes are determined with GA. As a result, their proposal confirms that GA can be efficiently utilized in the stratification of heterogeneous populations.

2.5. Other Methods

Brito, Maculan, Lila & Montenegro (2010) proposed an exact algorithm for the stratification problem with only proportional allocation based on the concept of minimum path in graphs (pp.190) and they called their method StratPath (pp.192). They claimed that so far no previous algorithm had been developed considering proportional allocation (Brito, et all. (2010), pp. 189) whereas Keskinturk & Er in 2007 had developed a genetic algorithm approach with proportional allocation as well as equal and Neyman allocation (Keskinturk & Er, (2007), pp.55). In order to test the performance of their method they applied geometric method with proportional allocation to 10 populations with 3, 4 and 5 strata case with a sample size of 100. They found that their method outperformed geometric results in all of the data used with an efficiency ratio of more than 1.1.
Brito, Ochi, Montenegro, Maculan (?) developed an Iterated Local Search (ILS) method to solve the stratification problem of variables with any distribution with Neyman allocation. They compared the performance of their method with geometric method, cum$\sqrt{f}$ and Kozak's (2004) random search using 16 populations with a skewness changing in between 1.4 and 34.8 (pp.7-8). In their comparisons they used geometric starting points for initial boundaries in Kozak's (2004) random search method and they tested their method by dividing the 16 populations into 3, 4, and 5 strata with varying sample sizes. They claimed that their ILS method provided the minimum coefficients of variations among the methods applied and when they compared the variance ratios of geometric method with their method they found that the ratios were 1.0 or 1.1 for 2 sets of data for all the strata cases. This meant that geometric method performed well enough in some data cases and poorly in other data sets (pp.9).

3. Comparisons of Different Stratification Methods

3.1. The Populations

The data used for the comparison of the available stratification methods are those given in the stratification package R and also the ones used in the papers of Kozak & Verma (2006) and Keskinturk & Er (2007). They can be summarized as follows:

Debtors: An accounting population of debtors in an Irish firm, detailed in Horgan (2003).

UScities: The population in thousands of US cities.

UScolleges: The number of students in four-year US colleges.

USbanks: The resources in millions of dollars of a large commercial bank in the US.

ME84: Number of municipal employees in 1984 in the 284 municipalities in Sweden.

Table1. Summary Statistics for the Populations

Pop	Source	N	Range	Skew	Kurt	Mean	SD
1	Debtors	3369	40-28,000	6.44	59.00	838.64	1873.99
2	UScities	1038	10-198	2.87	9.12	32.57	30.40
3	UScolleges	677	200-9,623	2.45	5.80	1563.00	1799.06
4	USbanks	357	70-977	2.07	4.06	225.62	190.46
5	ME84	284	173-47,074	8.64	84.04	1779.07	4253.13
6	P75	284	4-671	8.43	88.56	28.81	52.87
7	REV84	284	347-59,877	7.83	81.33	3088.09	4746.16
8	MRTS	2000	141.23-486,366.5	8.61	136.20	16882.80	21574.88
9	HHINCTOT*	16025	100-690,000	2.71	18.79	52123.73	41120.41
10	iso2004	487	63582908-10446591755	10.03	137.91	278237616.44	637769009.37
11	iso2005	485	69121110-14239223472	12.63	206.49	305852522.35	785107451.87
12	Kozak1	4000	3-72	1.40	3.73	16.11	6.77
13	Kozak2	4000	243-28578	2.66	13.18	2823.95	2201.14
14	Kozak3	2000	6-2793	3.54	20.08	224.12	245.08
15	Kozak4	10000	62-74398	4.20	29.27	3616.41	4530.39
16	Kozak5	2000	259-186685	5.01	50.62	9265.36	10688.04

*Data has 32 zero values which were excluded.

P75: 1975 population (in thousands) in the 284 municipalities in Sweden.

REV84: Real estate values according to 1984 assessment (in millions of kronor) in the 284 municipalities in Sweden.

MRTS: Simulated Data from the Monthly Retail Trade Survey of Statistics Canada.

HHINCTOT: Household income before taxes from the 2001 Survey of Household Spending (SHS) carried out by Statistics Canada.

iso2004: Net sales data of 487 largest Turkish manufacturing companies in 2004 according to Istanbul Chamber of Industry (ICI).

iso2005: Net sales data of 485 largest Turkish manufacturing companies in 2005 according to ICI.

Kozak1: Data set generated by Kozak & Verma (2006).

Kozak2: Data set generated by Kozak & Verma (2006).

Kozak3: Data set generated by Kozak & Verma (2006).

Kozak4: Data set generated by Kozak & Verma (2006).

Kozak5: Data set generated by Kozak & Verma (2006).

4. Overview of the Approximate Stratification Methods

In order to compare the efficiencies of the three algorithms available, all the populations mentioned are stratified into 3,4,5 and 6 strata with a sample size of 100 for the first 11 populations and with a sample size of 600, 600, 300, 1500 and 300 for Kozak & Verma's (2006) data respectively. This is done in order to obtain consistency with the paper of Kozak & Verma (2006). After stratifying the populations, the ratio of the variances of the estimates or the square of the ratio of the coefficient of variations are calculated as follows:

$$eff_{GA/Kozak} = \frac{\sigma^2_{GA}(\overline{x}_{st})}{\sigma^2_{Kozak}(\overline{x}_{st})} = \frac{(CV_{GA} \times \mu_{\overline{x}_{st}})^2}{(CV_{Kozak} \times \mu_{\overline{x}_{st}})^2} = \left(\frac{CV_{GA}}{CV_{Kozak}}\right)^2 \tag{25}$$

$$eff_{GA/LH} = \left(\frac{CV_{GA}}{CV_{LH}}\right)^2 \tag{26}$$

$$(27) \qquad eff_{Kozak/LH} = \left(\frac{CV_{Kozak}}{CV_{LH}}\right)^2$$

Since Lavallée & Hidiroglou's algorithm only considers a take-all top stratum, all three algorithms are compared with each other only when they all give a take-all top stratum result.
As it could be seen from Table2 that the efficiency ratios between GA and Kozak's random search method mostly equal to 1 with some exceptions whereas LH's algorithm gives greater coefficient of variations in many populations and strata cases.

5. Conclusion

Having stratified many populations with different characteristics reveals that Kozak's random search method and Keskinturk & Er's genetic algorithm gives more or less the same results both from the point of coefficient of variations and boundaries. This result shows that either genetic algorithm or Kozak's random search method could be efficiently applied in order to obtain near optimum boundaries in stratified sampling. Since both of the methods are available in **R** it is not a constraint for choosing the method.

REFERENCES (RÉFERENCES)

1 Brito, J., Ochi, L., Montenegro, F., Maculan, N. (?). An ILS Approach Applied to the Optimal Stratification Problem. (URL) http://www.ic.uff.br/ satoru/conteudo/artigos /ITOR2010-Andre

2 Brito, J., Maculan, N. Lila, M., Montenegro, F., (2010). An Exact Algorithm for the Stratification Problem with Proportional Allocation. Optimization Letters, 4, 2, pp.185-195.

3 Dalenius, T., Hodges, J.L.Jr. (1959). Minimum Variance Stratification, Journal of the American Statistical Association, 54, 285, pp.88-101.

4 Gunning, P., Horgan, J.M. (2004). A Simple Algorithm for Stratifying Skewed Populations, Survey Methodology. 30, 2, 159-185.

5 Horgan, J.M. http://www.janehorgan.com

6 Horgan, J.M. (2009) *Probability with R: An Introduction with Computer Science Applications*, Wiley, Hoboken, New Jersey, website: http://www.janehorgan.com

7 Keskinturk, T., Er, S. (2007). A Genetic Algorithm Approach to Determine Stratum Boundaries and Sample Sizes of Each Stratum in Stratified Sampling, Computational Statistics & Data Analysis, 52, 1, 53-67.

8 Kozak, M. (2004). Optimal Stratification Using Random Search Method in Agricultural Surveys, Statistics in Transition, 6, 5, 797-806.

9 Lavallée, P., Hidiroglou, M. (1988). On the Stratification of Skewed Populations, Survey Methodology, 14, 1, 33-43.

10 R Development Core Team (2005). R: A language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, (URL) http://www.r-project.org.

11 Singh, R., Mangat, N.S. Elements of Survey Sampling.

12 Venables, W.N., Smith, D.M. and the R Development Core Team (2005), An Introduction to R: A Programming Environment for Data Analysis and Graphics. At http://www.r-project.org/

Table2. Coefficients of Variations Obtained with Lavallée & Hidiroglou, Kozak and Keskinturk & Er

H	CVLH	CVGA	CVKozak	effGA/Kozak	effGA/LH	effKozak/LH
Debtors						
3	0.0693	0.05554	0.05554	0.9999	-	-
4	0.04721	0.04049	0.04049	1	-	-
5	0.03331	0.03131	0.03131	0.9998	-	-
6	0.02678	0.02562	0.02587	0.9801	-	-
UScities						
3	0.03217	0.02649	0.02649	1	-	-
4	0.02249	0.01927	0.01934	0.9928	-	-
5	0.01943	0.01437	0.0168	0.7312	-	-
6	0.01552	0.01214	0.01209	1.0076	-	-
	n=110					
UScolleges						
3	0.0346	0.02749	0.02749	0.9998	-	-
4	0.02399	0.02018	0.02018	1	-	-
5	0.01995	0.01607	0.01726	0.8672	-	-
6	0.01715	0.01323	0.01324	0.9995	-	-
USbanks						
3	0.01839	0.01802	0.01802	1	-	-
4	0.0127	0.0127	0.01270*	1	0.9991	0.9991
5	0.01094	0.00861	0.00861*	1	0.6198	0.6198
6	0.0071	0.0071	0.00711*	0.9981	0.9997	1.0016
ME84						
3	0.01296	0.01296*	0.01296*	1	0.9998	0.9998
4	0.0087	0.00870*	0.00870*	1	0.9991	0.9991
5	0.00657	0.00661*	0.00661*	1	0.9944	0.9944
6	0.00525	0.00577*	0.00522*	1.2217	1.2064	0.9875
P75						
3	0.01514	0.01459	0.01459*	1	0.9278	0.9278
4	0.01068	0.00966	0.00966*	1	0.8179	0.8179
5	0.00765	0.00835	0.00713*	1.3705	1.1904	0.8686
6	0.00608	0.00623	0.00552*	1.2735	1.0521	0.8261
REV84						
3	0.01618	0.01607	0.01607*	1	0.9954	0.9954
4	0.01112	0.01112	0.01112*	1	0.9996	0.9996
5	0.00832	0.00836	0.00837*	0.9971	0.9896	0.9924
6	0.00692	0.00666	0.00675*	0.9759	0.9074	0.9298
MRTS						
3	0.04559	0.04167	0.04168	0.9994	-	-
4	0.03025	0.0296	0.0296	0.9999	-	-
5	0.02307	0.02485	0.02297	1.1704	-	-
6	0.01837	0.01836	0.01836*	0.9995	0.9984	0.9988
HHINCTOT						
3	0.04503	0.03184	0.03184	1	-	-
4	0.03114	0.0243	0.02429	1.0012	-	-
5	0.02379	0.01979	0.01977	1.0012	-	-
6	0.01974	0.01629	0.0163	0.9995	-	-
iso2004						
3	0.01895	0.01894	0.01894*	1	0.9982	0.9982
4	0.01208	0.01206	0.01206*	1	0.9973	0.9973
5	0.00927	0.00908	0.00925*	0.9626	0.9584	0.9956
6	0.0082	0.00703	0.00811*	0.7516	0.7346	0.9773
iso2005						
3	0.01833	0.01833	0.01833*	0.9999	0.9997	0.9998
4	0.01245	0.01244	0.01244*	1	0.9973	0.9973
5	0.00912	0.00903	0.00910*	0.9852	0.981	0.9958
6	0.00808	0.00706	0.00805*	0.7689	0.763	0.9924
Kozak1						
3	0.00833	0.00636	0.00634	1.0066	-	-
4	0.00607	0.00481	0.0048	1.0062	-	-
5	0.0051	0.00398	0.00411	0.9397	-	-
6	0.00379	0.00331	0.00326	1.0267	-	-
Kozak2						
3	0.01237	0.0108	0.0108	1.0001	-	-
4	0.00841	0.00798	0.00798	1	-	-
5	0.00648	0.00636	0.00636	1	-	-
6	0.00511	0.00511	0.00511	1	-	-
Kozak3						
3	0.01966	0.01852	0.01852	1	-	-
4	0.01334	0.01328	0.01328	1.0001	-	-
5	0.01003	0.00999	0.00999*	1.0002	0.9923	0.9921
6	0.00836	0.00811	0.00811*	1	0.9403	0.9403
Kozak4						
3	0.00931	0.00888	0.00888	1	-	-
4	0.00634	0.00634	0.00634*	1	0.9998	0.9998
5	0.00473	0.00478	0.00473*	1.0195	1.0195	1
6	0.0038	0.00385	0.00380*	1.026	1.026	0.9999
Kozak5						
3	0.02003	0.01944	0.01944	1	-	-
4	0.01338	0.01338	0.01338*	1	0.9997	0.9997
5	0.00998	0.00998	0.00998*	1	1	1
6	0.0081	0.00806	0.00810*	0.9888	0.9888	1

* A take-all top stratum is observed.