STATISTICS IN TRANSITION new series, March 2012
Vol. 13, No. 1, pp. 7-20

DYNAMIC K-COMPOSITE ESTIMATOR FOR AN ARBITRARY ROTATION SCHEME

Przemysław Ciepiela[1], Małgorzata Gniado[2], Jacek Wesołowski[3] and Małgorzata Wojtyś[4]

ABSTRACT

Classical K-composite estimator was proposed in Hansen et al. (1955). Its optimality properties were developed in Rao and Graham (1964). This estimator gives an alternative solution to quasi-optimal estimation under rotation sampling when it is allowed that units leave the sample for several occasions and then come back. Such situations happen frequently in real surveys and are not covered by the recursive optimal estimator introduced by Patterson (1955). However the K-composite estimator suffers from certain disadvantages. It is designed for a stable situation in the sense that its basic parameter is kept constant on all occasions. Additionally it is restricted only to a certain family of rotation designs. Here we propose a dynamic version of the K-composite estimator (DK-composite estimator) without any restrictions on the rotation pattern. Mathematically, the algorithm, we develop, is much simpler than the one for the classical K-composite estimator with optimal weights. Moreover, it is precise, in the sense that it does not use any approximate or asymptotic approach (opposed to the method used in Rao and Graham (1964) for computing optimal weights).

1. Introduction

It is well known that, while looking for optimal estimators in surveys which repeat in time with the same time spacing, taking under account observations not only from the present edition of the survey (occasion) but also from previous occasions may significantly improve the quality of estimation.

[1] Bank PEKAO, Warszawa, POLAND
[2] Towarzystwo Ubezpieczeń na Życie "Warta", Warszawa, POLAND
[3] Główny Urząd Statystyczny and Wydział Matematyki i Nauk Informacyjnych, Politechnika Warszawska, Warszawa, POLAND, e-mail: wesolo@mini.pw.edu.pl
[4] Wydział Matematyki i Nauk Informacyjnych, Politechnika Warszawska, Warszawa, POLAND, e-mail: wojtys@mini.pw.edu.pl

For the best linear unbiased estimators (BLUEs) of the mean on a given occasion, to reduce time and memory requirements, it is desirable to have a recursive form for such an estimator which refers only to certain (possibly, small) number of optimal estimators from recent occasions and, additionally, observations from those occasions. Such a problem was completely solved in a seminal paper by Patterson (1950) for a family of rotation patterns which do not allow for a come-back of a unit to the sample, after leaving it for some occasions. The solution gives a formula for the BLUE of μ_h, the mean on the hth occasion, as a linear combination of the BLUE of μ_{h-1} and observations from $(h-1)$th and hth occasions.

However, in many practical surveys, the rotation pattern allows holes, i.e. some units stay in a sample for a number of occasions, leave it for a number of occasions, then return to the survey for a number of occasions. Important examples include the Current Population Survey (CPS) in the US, where the units follow the pattern 1111000000001111 (a unit is in the sample for subsequent 4 occasions, leaves it for subsequent 8 occasions, is again in the sample for subsequent 4 occasions and then never returns to the sample), or polish Labour Force Survey with the pattern 110011 (see Szarkowski and Witkowski (1994) or Popiński (2006)). Unfortunately the recurrent form of the BLUE in such situations of rotation patterns with holes is not known in general, see for instance Yansaneh and Fuller (1998). (Actually, the recurrent form of the optimal estimators for any rotation pattern with holes of size 1 has been derived only recently in Kowalski (2009) and for the Szarkowski scheme 110011 even more recently in Wesołowski (2010).) A widely accepted solution in the general situation is the K-composite estimator introduced in Hansen et al. (1955). Its optimality properties were studied for several models in Rao and Graham (1964) (shortened to RG in the rest of this paper).

By definition K-composite estimator makes use only of the most recent past composite estimator and observations from the present and the most recent past occasions. More precisely K-composite estimator on hth occasion, $\hat{\mu}_h$, has the following form

$$\hat{\mu}_h = Q\left(\hat{\mu}_{h-1} + \bar{X}^{(h)}_{h-1,h} - \bar{X}^{(h-1)}_{h-1,h}\right) + (1-Q)\bar{X}_h \ , \tag{1}$$

where $\hat{\mu}_{h-1}$ is the K-composite estimator on $(h-1)$th occasion, $\bar{X}^{(h)}_{h-1,h}$ is the sample mean for the units common to both $(h-1)$th and hth occasions calculated for the hth occasion, $\bar{X}^{(h-1)}_{h-1,h}$ is the sample mean of for the units common to both $(h-1)$th and hth occasions calculated for the $(h-1)$th occasion, $\bar{X}_h$ is the sample mean for all the units on hth occasion and $Q \in [0,1)$ is a numerical parameter which does not depend on h(!). Additionally in RG only a restricted though natural family of rotation patterns is investigated:

a group of units remains in the sample for r occasions, then leaves it for m occasions, comes back to the sample for r occasions, leaves it for m occasions, and so on. In such a setting strengthened by assuming exponential (Model 1) or arithmetic (Model 2) correlation pattern the optimal choice of Q is considered in that paper (in passing, let us note that Model 3 for correlation pattern is impossible since the resulting covariance matrix may not be positive definite). To attain this goal in RG it is taken $h \to \infty$, since otherwise, apparently, the optimal Q has to depend on h. Numerical solutions are then obtained since the resulting formula ((14) in RG) for the variance of the estimator is analytically non-treatable.

As it already has bee mentioned K-composite estimator has been used for years with some adjustments in the CPS - see for instance Bailar (1975), Breau and Ernst (1983) or Lent et al. (1994). A complete description can be found for instance in Current Population Survey (2002). The adjustments known as AK-composite estimator introduced in Gurney and Daly (1965) has been further developed, e.g. in Cantwell (1988) and Cantwell and Caldwell (1998). A more recent approach through regression composite estimator has been considered in Bell (2001), Fuller and Rao (2001), Singh et al. (2001) (with implications for Canadian Labour Force Survey). It is based on modified regression method proposed in Singh (1996). The difficulty in recursive estimation in repeated surveys for patterns with holes was raised in Yansaneh and Fuller (1998), who analyzed variances of composite estimators in several rotation schemes. For a relatively current description of the state of art in the area one can consult Steel and McLaren (2008), in particular Sec. IV on different rotation patterns and Sec. V on composite estimators. A very recent paper on optimal estimation under rotation is by Towhidi and Namazi-Rad (2010).

In the present paper we develop the idea of K-composite estimator in two new directions. First, $Q = Q_h$ is allowed to depend on the number of occasion. Then it appears that the optimal solution for Q_h is very simple: it is attained through minimizing certain quadratic function F_h (which has to be determined on each occasion). Second, any rotation pattern is allowed. The price for such a development is surprisingly cheap: we only have to keep track of subsequent Q_h's (to be able to determine F_h's).

2. Dynamic K-composite estimator

We consider a double array of random variables $(X_{i,j})$ which may be column-wise or row-wise infinite or finite, where the rows are for values of the variable of interest for different units on the same occasion, while columns are for values of the variable for the same unit on different occasions. Thus $X_{i,j}$ represents the value of the variable on ith occasion for the jth unit of the population. We assume that on a given occasion all the variables have the same mean, which is the parameter we want to estimate, i.e.

$$\mathbb{E}\, X_{i,j} = \mu_i \ , \quad i,j = 1,2,\ldots$$

Also it is assumed that there is no correlation between different units, i.e.

$$\mathbb{C}\mathrm{ov}(X_{i,j}, X_{l,k}) = 0 \quad \text{for} \ \ j \neq k, \ \ i,j,k,l = 1,2,\ldots$$

These two assumptions are crucial for further development of our result. The remaining two are not important for the derivation we propose but, first, make formulas somewhat simpler, second, since they include some parameters which are assumed to be known, it is desirable to have as few such parameters as possible. Thus, additionally we assume the exponential correlation pattern between the values of the variable for the same unit on different occasions (which, while not so important here, is a crucial condition for the Patterson scheme), i.e.

$$\mathbb{C}\mathrm{orr}(X_{i,j}, X_{i+k,j}) = \rho^k, \ \text{for any} \ k = 0,1,\ldots, \ i,j = 1,2,\ldots,$$

for some $\rho \in [-1,1]$. Finally it is assumed that the variances of all variables are constant, i.e.

$$\mathbb{V}\mathrm{ar}\, X_{i,j} = \sigma^2 > 0 \ , \quad i,j = 1,2,\ldots$$

The dynamic version of K-composite estimator, which called here DK-composite estimator, has the form

$$\hat{\mu}_h = Q_h \left(\hat{\mu}_{h-1} + \bar{X}^{(h)}_{h-1,h} - \bar{X}^{(h-1)}_{h-1,h} \right) + (1 - Q_h)\bar{X}_h \ , \quad h = 2,3,\ldots \tag{2}$$

while $\hat{\mu}_1 = \bar{X}_1$, where all the symbols where introduced in (1) except of Q_h which plays the role of the former Q. Let us point out the we do not impose any a priori restrictions on the range of (Q_h) (restriction imposed in RG on the range of Q, $Q \in (0,1)$, made it possible to pass to the limit with $h \to \infty$ in the expression for the variance of $\hat{\mu}_h$). Our goal is to choose Q_h in a dynamic way, i.e. on each occasion $h \geq 2$, the value Q_h has to minimize the variance of $\hat{\mu}_h$.

The rotation scheme is described by the rotation matrix $R = (r_{i,j})$, where $r_{i,j} = 1$ if the jth unit is in the sample on the ith occasion, otherwise $r_{i,j} = 0$. There is absolutely no restriction on the rotation pattern. By n_i we denote the sample size on the ith occasion, and m_i denotes the size of overlap between samples on occasions $(i-1)$th and ith.

Denote also for $k = 2,3,\ldots$

$$D_{i,k} = \begin{cases} Q_i Q_{i+1} \cdot \ldots \cdot Q_k, & \text{for } i = 1, \ldots, k, \\ \\ 1, & i = k+1\ . \end{cases} \tag{3}$$

Then we define weights which will be responsible for the form of quadratic functions (F_k) to be minimized:

$$w_{i,j}^{(1)} = \frac{1}{n_1}$$

and for any $i = 1, \ldots, k > 1$ and any $j = 1, 2, \ldots$

$$w_{i,j}^{(k)} = r_{i,j} \left[D_{i,k} \left(\frac{r_{i-1,j}}{m_i} - \frac{1}{n_i} \right) + D_{i+1,k} \left(\frac{1}{n_i} - \frac{r_{i+1,j}}{m_{i+1}} \right) \right] , \tag{4}$$

where in the last expression we adopt the rule that $r_{k+1,j} = r_{0,j} = 0$. Note that with such a little abuse of notation the formula for $w_{i,j}^{(1)}$ agrees with (4). Let us emphasize that to find the weights $\left(w_{i,j}^{(k)}\right)$ for a given occasion k nothing more is needed but $k-1$ numbers $Q_2, \ldots, Q_k$ (note that $Q_1 = 0$, by the definition of $\hat{\mu}_1$).

Now we are ready to present our main result which explains how to choose, occasion by occasion, the values (Q_h) which make the estimator $\hat{\mu}_h$ optimal in the model we consider here. Though the formulas, in particular (6), do not look very friendly, it has to be emphasized that actually to find Q_h one needs just $Q_2, \ldots, Q_{h-1}$ to calculate $w_{i,j}^{(h-1)}$ and consequently, A_h, B_h and C_h.

Theorem 1. *In the model described above the optimal value of Q_h which minimizes the variance of DK-composite estimator $\hat{\mu}_h$, $h \geq 1$, is*

$$Q_h = \frac{C_h - B_h}{A_h - 2B_h + C_h} \tag{5}$$

with $C_1 = B_1$ and $A_h - 2B_h + C_h > 0$, where for $h \geq 2$

$$A_h = \sigma^2 \sum_j \left[\sum_{i=1}^{h-1} \left(w_{i,j}^{(h-1)} \right)^2 r_{i,j} + 2 \sum_{1 \leq i_1 < i_2 \leq h-1} w_{i_1,j}^{(h-1)} w_{i_2,j}^{(h-1)} r_{i_1,j} r_{i_2,j} \rho^{i_2 - i_1} \right]$$

$$+ 2 \frac{(1-\rho)\sigma^2}{m_h} \left[1 - \sum_j \sum_{i=1}^{h-1} w_{i,j}^{(h-1)} r_{h-1,j} r_{h,j} r_{i,j} \rho^{h-1-i} \right] , \tag{6}$$

$$B_h = \frac{\sigma^2}{n_h} \left[1 - \rho + \sum_j \sum_{i=1}^{h-1} w_{i,j}^{(h-1)} r_{i,j} r_{h,j} \rho^{h-i} \right] , \tag{7}$$

$$C_h = \frac{\sigma^2}{n_h}. \tag{8}$$

Moreover,

$$\hat{\mu}_h = \sum_{i=1}^{h} \sum_{j} w_{i,j}^{(h)} r_{i,j} X_{i,j}$$

with the weights $(w_{i,j}^{(h)})$ *defined in (4).*

Actually, as it will be observed during the proof, which is given in Section 3,

$$A_h - 2B_h + C_h = \mathbb{V}\mathrm{ar}\left(\hat{\mu}_{h-1} + \bar{X}_{h-1,h}^{(h)} - \bar{X}_{h-1,h}^{(h-1)} + \bar{X}_h\right)$$

which is always positive since $\sigma^2 > 0$.

3. Proof

The proof is by induction with respect to h. For $h = 1$ the result holds true since then $C_1 = B_1$ yields $Q_1 = 0$. Moreover, (4) for $k = 1$ agrees with the formula for $w_{i,j}^{(1)}$. We assume that it holds for $h - 1$ and we will prove it for h.

Compute the variance of $\hat{\mu}_h$:

$$\mathbb{V}\mathrm{ar}\,\hat{\mu}_h = Q_h^2 \Big[\mathbb{V}\mathrm{ar}\,\hat{\mu}_{h-1} + \mathbb{V}\mathrm{ar}\,\bar{X}_{h-1,h}^{(h)} + \mathbb{V}\mathrm{ar}\,\bar{X}_{h-1,h}^{(h-1)} + 2\mathbb{C}\mathrm{ov}\left(\hat{\mu}_{h-1}, \bar{X}_{h-1,h}^{(h)}\right)$$

$$-2\mathbb{C}\mathrm{ov}\left(\hat{\mu}_{h-1}, \bar{X}_{h-1,h}^{(h-1)}\right) - 2\mathbb{C}\mathrm{ov}\left(\bar{X}_{h-1,h}^{(h)}, \bar{X}_{h-1,h}^{(h-1)}\right)\Big]$$

$$+2Q_h(1-Q_h)\Big[\mathbb{C}\mathrm{ov}\left(\hat{\mu}_{h-1}, \bar{X}_h\right) + \mathbb{C}\mathrm{ov}\left(\bar{X}_{h-1,h}^{(h)}, \bar{X}_h\right) - \mathbb{C}\mathrm{ov}\left(\bar{X}_{h-1,h}^{(h-1)}, \bar{X}_h\right)\Big] +$$

$$(1-Q_h)^2\,\mathbb{V}\mathrm{ar}\,\bar{X}_h = Q_h^2 A_h + 2Q_h(1-Q_h)B_h + (1-Q_h)^2 C_h,$$

where the last equality defines the quantities A_h, B_h and C_h.

By the induction assumption

$$\hat{\mu}_{h-1} = \sum_{i=1}^{h-1} \sum_{j} w_{i,j}^{(h-1)} r_{i,j} X_{i,j}\ .$$

Then a direct computation gives

$$\mathbb{V}\mathrm{ar}\,\hat{\mu}_{h-1} = \sigma^2 \sum_j \left[\sum_{i=1}^{h-1} \left(w_{i,j}^{(h-1)} \right)^2 r_{i,j} + 2 \sum_{1 \le i_1 < i_2 \le h-1} w_{i_1,j}^{(h-1)} w_{i_2,j}^{(h-1)} r_{i_1,j} r_{i_2,j} \rho^{i_2 - i_1} \right] ,$$

$$\mathbb{V}\mathrm{ar}\,\bar{X}_{h-1,h}^{(h)} = \mathbb{V}\mathrm{ar}\,\bar{X}_{h-1,h}^{(h-1)} = \frac{\sigma^2}{m_h} ,$$

$$\mathbb{C}\mathrm{ov}\left(\hat{\mu}_{h-1}, \bar{X}_{h-1,h}^{(h)}\right) = \frac{\sigma^2}{m_h} \sum_j \sum_{i=1}^{h-1} w_{i,j}^{(h-1)} r_{i,j} r_{h-1,j} r_{h,j} \rho^{h-i} ,$$

$$\mathbb{C}\mathrm{ov}\left(\hat{\mu}_{h-1}, \bar{X}_{h-1,h}^{(h-1)}\right) = \frac{\sigma^2}{m_h} \sum_j \sum_{i=1}^{h-1} w_{i,j}^{(h-1)} r_{i,j} r_{h-1,j} r_{h,j} \rho^{h-1-i} ,$$

$$\mathbb{C}\mathrm{ov}\left(\bar{X}_{h-1,h}^{(h)}, \bar{X}_{h-1,h}^{(h-1)}\right) = \frac{\sigma^2 \rho}{m_h} .$$

Combining the last five formulas we observe that the definition of A_h agrees with the expression (6).

Similarly to check if (7) holds we have to compute

$$\mathbb{C}\mathrm{ov}\left(\hat{\mu}_{h-1}, \bar{X}_h\right) = \frac{\sigma^2}{n_h} \sum_j \sum_{i=1}^{h-1} w_{i,j}^{(h-1)} r_{i,j} r_{h,j} \rho^{h-i} ,$$

$$\mathbb{C}\mathrm{ov}\left(\bar{X}_{h-1,h}^{(h)}, \bar{X}_h\right) = \frac{\sigma^2}{n_h} ,$$

$$\mathbb{C}\mathrm{ov}\left(\bar{X}_{h-1,h}^{(h-1)}, \bar{X}_h\right) = \frac{\sigma^2 \rho}{n_h} .$$

Finally, (8) follows since

$$\mathbb{V}\mathrm{ar}\,\bar{X}_h = \frac{\sigma^2}{n_h} .$$

Minimizing

$$F_h(x) = (A_h - 2B_h + C_h)x^2 + 2(B_h - C_h)x + C_h$$

we get the solution (5).

The DK-composite estimator is a linear estimator so in general it has a form

$$\hat{\mu}_h = \sum_{i=1}^{h} \sum_{j} v_{i,j} r_{i,j} X_{i,j}$$

with some weights $(v_{i,j})$. To finish the proof we have to show that $v_{i,j} r_{i,j} = w_{i,j}^{(h)} r_{i,j}$ as defined in (4) for any $i = 1, \ldots, h$ and any $j = 1, 2, \ldots$.

Note that by the definition of $\hat{\mu}_h$ given in (2) we have

$$\hat{\mu}_h = Q_h \left(\sum_j \sum_{i=1}^{h-1} w_{i,j}^{(h-1)} r_{i,j} X_{i,j} + \frac{1}{m_h} \sum_j r_{h-1,j} r_{h,j} X_{h,j} \right.$$

$$\left. - \frac{1}{m_h} \sum_j r_{h-1,j} r_{h,j} X_{h-1,j} \right) + (1 - Q_h) \frac{1}{n_h} \sum_j r_{h,j} X_{h,j}.$$

Comparing the coefficients of $X_{i,j}$ in the last two expressions we get for $i = h$

$$v_{h,j} r_{h,j} = \frac{Q_h}{m_h} r_{h-1,j} r_{h,j} + \frac{1 - Q_h}{n_h} r_{h,j}$$
$$= r_{h,j} \left[D_{h,h} \left(\frac{r_{h-1,j}}{m_h} - \frac{1}{n_h} \right) + \frac{1}{n_h} \right] = w_{h,j}^{(h)} r_{h,j} \ ,$$

for $i = h - 1$

$$v_{h-1,j} r_{h-1,j} = Q_h \left(w_{h-1,j}^{(h-1)} r_{h-1,j} - \frac{1}{m_h} r_{h-1,j} r_{h_j} \right)$$

$$= Q_h r_{h-1,j} \left[D_{h-1,h-1} \left(\frac{r_{h-2,j}}{m_{h-1}} - \frac{1}{n_{h-1}} \right) + \frac{1}{n_{h-1}} \right] - Q_h \frac{1}{m_h} r_{h-1,j} r_{h,j}$$

$$= r_{h-1,j} \left[D_{h-1,h} \left(\frac{r_{h-2,j}}{m_{h-1}} - \frac{1}{n_{h-1}} \right) + D_{h,h} \left(\frac{1}{n_{h-1}} - \frac{r_{h,j}}{m_h} \right) \right] = w_{h-1,j}^{(h)} r_{h-1,j} \ ,$$

and for any $i < h - 1$

$$v_{i,j} r_{i,j}$$

$$= Q_h w_{i,j}^{(h-1)} r_{i,j} \left[D_{i,h-1} \left(\frac{r_{i-1,j}}{m_i} - \frac{1}{n_i} \right) + D_{i+1,h-1} \left(\frac{1}{n_i} - \frac{r_{i+1,j}}{m_{i+1}} \right) \right] = w_{i,j}^{(h)} r_{i,j}$$

since $Q_h D_{k,h-1} = D_{k,h}$ for $k = i, \ i + 1$.

Thus the proof is completed. $\square$

4. Numerical examples

Below, similarly to numerical comparisons in RG, we consider percentage gain in efficiency for the DK-composite estimator compared to the mean of the observations from the last hth occasion. It is defined as

$$g_h = \frac{\mathbb{V}\mathrm{ar}\,\bar{X}_h - \mathbb{V}\mathrm{ar}\,\hat{\mu}_h}{\mathbb{V}\mathrm{ar}\,\hat{\mu}_h} \times 100. \tag{9}$$

We took $h = 20$, since the parameter Q_h behaves quite stable with respect to occasion number h. In Tables 1 and 2 we give the optimal weight Q_h, the variance of μ_h and g_h for different values of correlation ρ in two schemes: Szarkowski's 110011 (Table 1) and CPS 1111000000001111 (Table 2). We can easily see that the largest gain is achieved for strong correlations and the smallest when there is no correlation between occasions for the same unit.

Table 1. Szarkowski scheme

ρ	Q_{20}	$\mathbb{V}\mathrm{ar}\,\hat{\mu}_{20}$	g_{20}
-0.9	-0.16	0.235	6.275
-0.8	-0.14	0.238	5.173
-0.7	-0.13	0.240	4.138
-0.6	-0.12	0.242	3.181
-0.5	-0.10	0.244	2.317
-0.4	-0.08	0.246	1.560
-0.3	-0.07	0.248	0.926
-0.2	-0.05	0.249	0.437
-0.1	-0.02	0.250	0.116
0	0.00	0.250	0.000
0.1	0.03	0.250	0.135
0.2	0.06	0.249	0.592
0.3	0.09	0.246	1.473
0.4	0.13	0.243	2.940
0.5	0.17	0.238	5.256
0.6	0.22	0.230	8.877
0.7	0.29	0.218	14.695
0.8	0.38	0.200	24.797
0.9	0.51	0.171	46.499

Table 2. CPS scheme

ρ	Q_{20}	$\mathbb{V}\mathrm{ar}\,\hat{\mu}_{20}$	g_{20}
-0.9	-0.3	0.116	7.469
-0.8	-0.27	0.118	5.772
-0.7	-0.24	0.120	4.341
-0.6	-0.20	0.121	3.149
-0.5	-0.17	0.122	2.172
-0.4	-0.14	0.123	1.390
-0.3	-0.11	0.124	0.787
-0.2	-0.07	0.125	0.355
-0.1	-0.04	0.125	0.091
0	0.00	0.00	0.000
0.1	0.04	0.250	0.098
0.2	0.08	0.249	0.414
0.3	0.12	0.124	1.000
0.4	0.17	0.123	1.947
0.5	0.23	0.121	3.417
0.6	0.29	0.118	5.729
0.7	0.36	0.114	9.586
0.8	0.45	0.107	16.908
0.9	0.58	0.092	35.564

Source: own calculations

Consider now a cascade rotation scheme which is defined through a rotation pattern $(1, \epsilon_2, \ldots, \epsilon_{k-1}, 1)$, $\epsilon_l \in \{0, 1\}$, $l = 2, \ldots, k-1$, which moves one unit down the rotation matrix with subsequent occasions, that is

$$(r_{i,i}, \ldots, r_{i,i+k}) = (1, \epsilon_2, \ldots, \epsilon_{k-1}, 1)$$

for any $i = 1, 2, \ldots$, otherwise $r_{i,j} = 0$. The number k is called the rotation pattern length.

Taking the advantage of the fact that the DK-composite estimator allows for any rotation scheme, we calculated percentage gain (as defined in (9)) in efficiency for all possible cascade schemes with rotation patterns of length up to 10. Table 3 contains 10 schemes with the smallest and 10 with the largest gain among such $2^8 = 256$ schemes. Here, again, the results for $h = 20$ are presented. The largest gain is achieved for "sparse" schemes with small number of elements in the rotation pattern (and strong correlations) while the lowest gain is observed for schemes with complete or almost complete rotation patterns (and for weak correlations). Similar comparisons of variances for particular rotation cascade patterns in the time series framework can be found in McLaren and Steel (2000) (see also Steel and McLaren (2002)).

Table 3. The worst and the best rotation patterns

worst patterns	ρ	g_{20}	best patterns	ρ	g_{20}
1111111111	-0.1	0.045	1000010001	0.9	84.087
1111111111	0.1	0.046	1000100001	0.9	84.087
111111111	-0.1	0.049	101	0.9	93.570
111111111	0.1	0.051	1001	0.9	108.519
11111111	-0.1	0.054	10001	0.9	117.448
11111111	0.1	0.056	100001	0.9	122.970
1111111	-0.1	0.060	1000001	0.9	126.075
1111111	0.1	0.063	10000001	0.9	127.970
1101010101	-0.1	0.064	1000000001	0.9	129.264
1011010101	-0.1	0.064	100000001	0.9	129.305

Source: own calculations

Table 4. Comparison between K-composite and DK-composite estimators

ρ	0.5			0.6			0.7		
	Q_{20}	g_{20}	diff	Q_{20}	g_{20}	diff	Q_{20}	g_{20}	diff
m	$r = 2$								
2	0.23	3.42	1.74	0.29	5.73	2.77	0.36	9.59	4.64
4	0.17	5.26	0.01	0.22	8.88	-0.19	0.29	14.70	0.58
8	0.17	5.29	-0.03	0.23	9.02	-0.33	0.29	15.23	0.09
∞	0.17	5.29	-0.03	0.23	9.02	-0.33	0.29	15.23	0.09
m	$r = 3$								
3	0.24	2.34	1.83	0.30	3.97	3.28	0.38	6.56	5.41
6	0.21	4.27	-0.04	0.27	7.20	0.13	0.34	12.03	0.40
9	0.21	4.27	-0.04	0.27	7.23	0.10	0.34	12.16	0.28
∞	0.21	4.27	-0.04	0.27	7.23	0.10	0.34	12.16	0.28
m	$r = 4$								
4	0.25	1.84	1.47	0.32	3.02	2.76	0.39	4.96	4.82
8	0.23	3.42	-0.11	0.29	5.73	0.06	0.36	9.56	0.32
12	0.23	3.42	-0.11	0.29	5.73	0.06	0.36	9.59	0.29
∞	0.23	3.42	-0.11	0.29	5.73	0.06	0.36	9.59	0.29
m	$r = 6$								
6	0.25	1.25	1.08	0.32	2.04	1.92	0.39	3.32	3.38
12	0.24	2.40	-0.07	0.30	3.97	-0.01	0.38	6.56	0.14
18	0.24	2.40	-0.07	0.30	3.97	-0.01	0.38	6.56	0.14
∞	0.24	2.40	-0.07	0.30	3.97	-0.01	0.38	6.56	0.14
m	$r = 8$								
8	0.26	0.94	0.86	0.32	1.54	1.45	0.40	2.50	2.53
16	0.25	1.84	-0.03	0.31	3.02	-0.03	0.39	4.96	0.07
∞	0.25	1.84	-0.03	0.31	3.02	-0.03	0.39	4.96	0.07

Source: own calculations

Table 4. Comparison between K-composite and DK-composite estimators, continuation

ρ	0.8			0.9		
	Q_{20}	g_{20}	diff	Q_{20}	g_{20}	diff
m	$r = 2$					
2	0.45	16.91	5.94	0.58	35.72	3.06
4	0.38	24.80	1.88	0.51	46.50	4.15
8	0.38	26.75	0.37	0.52	54.65	1.01
∞	0.38	26.77	0.35	0.52	55.07	1.18
m	$r = 3$					
3	0.47	11.46	8.74	0.60	24.23	12.42
6	0.43	20.77	1.53	0.56	40.48	4.80
9	0.43	21.44	1.00	0.57	44.28	2.43
∞	0.43	21.47	0.98	0.57	44.89	2.09
m	$r = 4$					
4	0.48	8.57	8.55	0.61	17.93	14.47
8	0.45	16.69	1.10	0.58	33.67	4.92
12	0.45	16.91	0.89	0.58	35.56	4.23
∞	0.45	16.91	0.89	0.58	35.72	4.30
m	$r = 6$					
6	0.49	5.67	6.31	0.61	11.61	14.41
12	0.47	11.44	0.78	0.60	23.79	4.56
18	0.47	11.46	0.76	0.60	24.22	4.28
∞	0.47	11.46	0.76	0.60	24.23	4.28
m	$r = 8$					
8	0.49	4.23	5.00	0.62	8.55	12.08
16	0.48	8.57	0.70	0.61	17.83	3.50
∞	0.48	8.57	0.70	0.61	17.93	3.42

Source: own calculations

Table 4 shows Q_h, g_h and the difference ($diff = g - g_h$) between the gains obtained in two ways: g for the K-composite estimator as computed in Table 1 of RG and g_h for the DK-composite estimator as proposed in the present paper. We took $h = 100$ though in many particular cases the values for Q_h and g_h stabilized much earlier. The numbers r and m are responsible for the rotation pattern, i. e. a unit stays in the sample for r occasions, leaves the sample for m occasions, comes back into the sample for r occasions, and so on. Table 4 is named "comparison" nevertheless we cannot actually in strictly mathematical sense compare these two values of g and g_h because the two methods involve different models: RG considered a finite population case in which a given unit returns to the survey infinitely often whereas in the present paper an infinite population model is investigated and a unit returns to the survey after a gap of m occasions for another sequence of r occasions and then leaves the survey. In the course of simulations we noted that Q_h for the DK-composite estimator are quite stable, even for relatively small values of h. Moreover, their values observed in simulations were quite similar to those of Q, obtained in Table 1 of RG. In our Table 4 it is visible that differences in gains of efficiency are remarkable for strong correlations and small gaps m, while for small correlations and large gaps m they are insignificant. Small negative values which appear in some cases are due to the fact that the two methods are not precisely equivalent, otherwise negative values would not be possible since the method we present here is optimal within considered class of estimators.

REFERENCES

BAILAR, B. (1975). The effects of rotation group bias on estimates from panel surveys. J. Amer. Statist. Assoc. 70, 23-30.

BELL, P. (2001). Comparison of alternative Labour Force Survey estimators. Survey Meth. 27(1), 53-63.

BREAU, P., ERNST, L. (1983). Alternative estimators to the current composite estimator. Proc. Sec. Survey Res. Meth., Amer. Statist. Assoc., 397-402.

CANTWELL, P.J. (1988). Variance formulae for the generalized composite estimator under balanced one-level rotation plan. SRD Research Report Census/SRD/88/26, Bureau of the Census, Statistical Research Division, 1-16.

CANTWELL, P.J., CALDWELL, C.V. (1998). Examining the revisions in monthly retail and wholesale trade surveys under a rotation panel design. J. Offic. Statist. 14, 47-54.

Current Population Survey (2002). Design and Methodology, Technical Paper 63RV, Bureau of Labour Statistics, U.S. Census Bureau.

FULLER, W., RAO, J.N.K. (2001). A regression composite estimator with application to the Canadian Labour Force Survey. Survey Meth. 27(1), 45-51.

GURNEY, M., DALY, J.F. (1965). A multivariate approach to estimation in periodic sample surveys. Proc. Amer. Statist. Assoc., Sect. Soc. Statist., 242-257.

HANSEN, M.H., HURWITZ, W.N., NISSELSON, H., STEINBERG, J. (1955). The redesign of the census current population survey. J. Amer. Math. Assoc. 50, 701-719.

KOWALSKI, J. (2009). Optimal estimation in rotation patterns. J. Statist. Plan. Infer. 139(4), 2429-2436.

LENT, J., MILLER, S., CANTWELL, P. (1994). Composite weights for the Current Population Survey. Proc. Sec. Survey Res. Meth., Amer. Statist. Assoc., 867-872.

MCLAREN, C.H., STEEL, D.G. (2000). The impact of different rotation patterns on the sampling variance of seasonally adjusted and trend estimates. Survey Meth. 26(2), 163-172.

PATTERSON, H.D. (1950). Sampling on successive occasions with partial replacement of units. J. Royal Statist. Soc., Ser. B 12, 241-255.

POPIŃSKI, W. (2006). Development of the Polish Labour Force Survey. Statist. Transit. 7(5), 1009-1030.

RAO, J.N.K., GRAHAM, J.E. (1964). Rotation designs for sampling on repeated occasions. Ann. Math. Statist. 35, 492-509.

SINGH, A.C. (1996). Combining information in survey sampling by modified regression. Proc. Sect. Survey Res. Meth., Amer. Statist. Assoc., 120-129.

SINGH, A.C., KENNEDY, B., WU, S. (2001). Regression composite estimation for the Canadian Labour Force Survey with a rotating panel design. Survey Meth. 27, 33-44.

STEEL, D., MCLAREN, C. (2002), In search of a good rotation pattern. In: Advances in Statistics, Combinatorics and Related Areas. Singapore, World Scientific, 309-319.

STEEL, D., MCLAREN, C. (2008). Design and analysis of repeated surveys. Centre for Statist. Survey Meth., Univ. Wollonong, Working Paper 11-08, 1-13, http://ro.uow.edu.au/cssmwp/10

SZARKOWSKI, A., WITKOWSKI, J. (1994), The Polish labour force survey. Statist. Transit. 1(4), 467-483.

TOWHIDI, M., NAMAZI-RAD, M.-R. (2010). An optimal method of estimation in rotation sampling. Adv. Appl. Statist. 15(2) , 115-136.

WESOŁOWSKI, J. (2010). Recursive optimal estimation in Szarkowski rotation scheme. Statist. Transit. 11(2), 267-285.

YANSANEH, I.S., FULLER, W. (1998). Optimal recursive estimation for repeated surveys. Survey Meth. 24, 31-40.