STATISTICS IN TRANSITION new series, March 2021
Vol. 22, No. 1 pp. 207–216, DOI 10.21307/stattrans-2021-012
Received – 05.05.2020; accepted – 27.11.2020

A new family of robust regression estimators utilizing robust regression tools and supplementary attributes

Irsa Sajjad[1], Muhammad Hanif[2], Nursel Koyuncu[3], Usman Shahzad[2,4], Nadia H. Al-Noor[5]

ABSTRACT

Zaman and Bulut (2018a) developed a class of estimators for a population mean utilising LMS robust regression and supplementary attributes. In this paper, a family of estimators is proposed, based on the adaptation of the estimators presented by Zaman (2019), followed by the introduction of a new family of regression-type estimators utilising robust regression tools (LAD, H-M, LMS, H-MM, Hampel-M, Tukey-M, LTS) and supplementary attributes. The mean square error expressions of the adapted and proposed families are determined through a general formula. The study demonstrates that the adapted class of the Zaman (2019) estimators is in every case more proficient than that of Zaman and Bulut (2018a). In addition, the proposed robust regression estimators based on robust regression tools and supplementary attributes are more efficient than those of Zaman and Bulut (2018a) and Zaman (2019).The theoretical findings are supported by real-life examples.

Key words: supplementary attributes, ratio-type estimators, SRS, robust regression tools, percentage relative efficiency..

1. Introduction

The estimation theory is important in different interdisciplinary territories of research including financial matters, clinical preliminaries, population studies, agriculture, engineering, and so on. Additionally, the issue of estimation of mean is critical in research, for example, the estimation of average crop yield, normal life

[1] Department of Lahore Business School - University of Lahore, Islamabad, Pakistan.
E-mail: irsasajjad@yahoo.com. ORCID: https://orcid.org/0000-0002-7246-7043.
[2] Department of Mathematics and Statistics - PMAS-Arid Agriculture University, Rawalpindi, Pakistan.
E-mail: mhpuno@hotmail.com. ORCID: https://orcid.org/0000-0002-1976-4452.
[3] Hacettepe University, Department of Statistics, Beytepe, Ankara, Turkey. E-mail: nkoyuncu@hacettepe.edu.tr. ORCID: https://orcid.org/0000-0003-1065-3411.
[4] Department of Mathematics and Statistics, International Islamic University, Islamabd, Pakistan.
Corresponding E-mail: usman.stat@yahoo.com. ORCID: https://orcid.org/0000-0002-0178-5298.
[5] Department of Mathematics, College of Science, Mustansiriyah University, Baghdad, Iraq.
E-mail: nadialnoor@uomustansiriyah.edu.iq. ORCID: https://orcid.org/0000-0002-4433-9044.

expectancy of people in an area and many others. To improve the efficiency of estimation of parameters, an auxiliary variable is widely used in the literature. An alternate way to enhance the efficiency of an estimator is to utilize supplementary attributes (for more details, interested readers may refer to Naik and Gupta (1996), Shahzad (2016), and Shahzad et al. (2018)). Ratio and product estimation techniques are widely used under SRS (simple random sampling) scheme. Both of these schemes have their own advantages and disadvantages. For instance, a ratio estimator is suitable for a positive linear relationship between the study and supplementary attribute, a product estimator is suitable for a negative linear relationship between the study and supplementary attribute. The usual regression estimator solves this issue and provides much better results for both positive and negative correlations. Note that the usual regression estimator based on ordinary least square (OLS) regression coefficient. However, when data are contaminated with outliers, OLS will not perform well, and as a result we get poor results. For solving this issue, Zaman and Bulut (2018a) utilized one of the robust regression technique namely LMS (Least Median of Squares), and provide a set of estimators utilizing auxiliary attribute under SRS scheme. In the current article, a class of robust ratio estimators is constructed by adapting the estimators of Zaman (2019), and a new class of robust regression estimators is introduced utilizing the supplementary attributes and robust-regression tools (least absolute deviations (LAD), Huber's M-estimator (H-M), least median of squares (LMS), least trimmed squares (LTS), Huber's MM-estimator (H-MM), Hampel-M estimator, Tukey-M estimator) under simple random sampling.

Zaman and Bulut (2018a) developed a class of estimators utilizing the known parameters of a supplementary attribute and LMS regression coefficient, as given below:

$$\bar{y}_{zb_1} = \frac{\bar{y} + b_{\varphi(lms)}(P_1 - p_1)}{p_1} P_1 \,, \tag{1}$$

$$\bar{y}_{zb_2} = \frac{\bar{y} + b_{\varphi(lms)}(P_1 - p_1)}{\left(P_1 + C_{\varphi}(\varphi)\right)} \left(P_1 + C_{\varphi}\right), \tag{2}$$

$$\bar{y}_{zb_3} = \frac{\bar{y} + b_{\varphi(lms)}(P_1 - p_1)}{\left(P_1 + \beta_2(\varphi)\right)} \left(P_1 + \beta_2(\varphi)\right), \tag{3}$$

$$\bar{y}_{zb_4} = \frac{\bar{y} + b_{\varphi(lms)}(P_1 - p_1)}{\left(P_1 C_{\varphi} + \beta_2(\varphi)\right)} \left(P_1 C_{\varphi} + \beta_2(\varphi)\right), \tag{4}$$

$$\bar{y}_{zb_5} = \frac{\bar{y} + b_{\varphi(lms)}(P_1 - p_1)}{\left(P_1 \beta_2(\varphi) + C_{\varphi}\right)} \left(P_1 \beta_2(\varphi) + C_{\varphi}\right), \tag{5}$$

where C_φ, the coefficient of variation, $\bar{y}$, the sample mean, $\beta_2(\varphi)$, the coefficient of kurtosis, and $b_{\varphi(lms)}$ is the LMS robust regression coefficient. Further, P_1 and p_1 represent population and sample proportions, respectively. For more details about proportions, interested readers may refer to Zaman and Bulut (2018a).

The MSE of Zaman and Bulut (2018a) family of estimators is given below:

$$MSE(\bar{y}_{zb_i}) = \gamma[S_y^2 + g_i^2 S_\varphi^2 + 2B_i g_i S_\varphi^2 + B_i^2 S_\varphi^2 - 2g_i S_{y\varphi} - 2B_i S_{y\varphi}]; i = 1, \ldots, 5 \tag{6}$$

where

$g_1 = 1,\ g_2 = \frac{\bar{Y}}{P_1 + C_\varphi}, g_3 = \frac{\bar{Y}}{P_1 + \beta_2(\varphi)}, g_4 = \frac{C_\varphi \bar{Y}}{C_\varphi P_1 + \beta_2(\varphi)}, g_5 = \frac{\beta_2(\varphi)\bar{Y}}{C_\varphi P_1 + C_\varphi}, S_{y\varphi} = \rho S_y S_\varphi$ and $= \left(\frac{1}{n} - \frac{1}{N}\right)$. Further, S_y^2 and S_φ^2 are the unbiased variances of Y and P_1 respectively, ρ is the coefficient of correlation.

2. Adapted Family of Estimators

Zaman (2019) developed a class of estimators utilizing known characteristics of supplementary information. By analogy to the approach of Zaman (2019), a supplementary attribute is utilized here, as given below:

$$\bar{y}_{z_1} = k\frac{\bar{y} + b_{\varphi(i)}(P_1 - p_1)}{p_1} P_1 + (1-k)\frac{\bar{y} + b_{\varphi(i)}(P_1 - p_1)}{\left(P_1 + C_\varphi(\varphi)\right)}\left(P_1 + C_\varphi\right), \tag{7}$$

$$\bar{y}_{z_2} = k\frac{\bar{y} + b_{\varphi(i)}(P_1 - p_1)}{p_1} P_1 + (1-k)\frac{\bar{y} + b_{\varphi(i)}(P_1 - p_1)}{(P_1 + \beta_2(\varphi))}\left(P_1 + \beta_2(\varphi)\right), \tag{8}$$

$$\bar{y}_{z_3} = k\frac{\bar{y} + b_{\varphi(i)}(P_1 - p_1)}{p_1} P_1 + (1-k)\frac{\bar{y} + b_{\varphi(i)}(P_1 - p_1)}{\left(P_1 C_\varphi + \beta_2(\varphi)\right)}\left(P_1 C_\varphi + \beta_2(\varphi)\right), \tag{9}$$

$$\bar{y}_{z_4} = k\frac{\bar{y} + b_{\varphi(i)}(P_1 - p_1)}{p_1} P_1 + (1-k)\frac{\bar{y} + b_{\varphi(i)}(P_1 - p_1)}{(P_1 \beta_2(\varphi) + C_\varphi)}\left(P_1 \beta_2(\varphi) + C_\varphi\right), \tag{10}$$

In general form, we can write the adapted class of estimators as given below:

$$\bar{y}_{z_i} = k\frac{\bar{y} + b_{\varphi(i)}(P_1 - p_1)}{p_1} P_1 + (1-k)\frac{\bar{y} + b_{\varphi(i)}(P_1 - p_1)}{(P_1 U_\varphi + V_\varphi)}\left(P_1 U_\varphi + V_\varphi\right), \tag{11}$$

where U_φ and V_φ are the known characteristics of an auxiliary attribute. The MSE of $\mathrm{MSE}(\bar{y}_{z_i})$ is as follows:

$$MSE(\bar{y}_{z_i}) = \gamma[S_y^2 - 2\delta S_{y\varphi} + \delta^2 S_\varphi^2], \tag{12}$$

where $\delta = [k(B_{\varphi(i)} + g_1) + (1 - k)(B_{\varphi(i)} + g_i)]$.

By replacing $(\delta = B)$ in the above MSE expression, we get the minimum MSE of $\bar{y}_{z_i}$ as follows:

$$MSE(\bar{y}_{z_i}) = \gamma S_y^2 (1 - \rho^2), \tag{13}$$

which is the MSE of traditional regression estimator, i.e. $\bar{y}_{reg} = \bar{y} + b_{\varphi(i)}(P_1 - p_1)$. Note that Abd-Elfattah et al. (2010) consider same class in the same context utilizing OLS regression coefficient. However, Zaman (2019) introduced robust regression techniques instead of OLS regression in the presence of outliers.

3. Proposed Estimators

Taking motivation from ratio type estimators of Zaman (2019), we propose the following family of robust regression estimators as given by

$$\hat{\bar{y}}_{p_1} = w_1[\bar{y} + b_{\varphi(mm)}(P_1 - p_1)] + w_2[\bar{y} + b_{\varphi(lms)}(P_1 - p_1)], \tag{14}$$

$$\hat{\bar{y}}_{p_2} = w_1[\bar{y} + b_{\varphi(lad)}(P_1 - p_1)] + w_2[\bar{y} + b_{\varphi(tuket)}(P_1 - p_1)], \tag{15}$$

$$\hat{\bar{y}}_{p_3} = w_1[\bar{y} + b_{\varphi(lts)}(P_1 - p_1)] + w_2[\bar{y} + b_{\varphi(huber)}(P_1 - p_1)], \tag{16}$$

$$\hat{\bar{y}}_{p_4} = w_1[\bar{y} + b_{\varphi(mm)}(P_1 - p_1)] + w_2[\bar{y} + b_{\varphi(hample)}(P_1 - p_1)], \tag{17}$$

$$\hat{\bar{y}}_{p_5} = w_1[\bar{y} + b_{\varphi(huber)}(P_1 - p_1)] + w_2[\bar{y} + b_{\varphi(mm)}(P_1 - p_1)], \tag{18}$$

$$\hat{\bar{y}}_{p_6} = w_1[\bar{y} + b_{\varphi(huber)}(P_1 - p_1)] + w_2[\bar{y} + b_{\varphi(lms)}(P_1 - p_1)], \tag{19}$$

In general form, we can write the proposed family of estimators as

$$\hat{\bar{y}}_{p_i} = w_1[\bar{y} + b_{\varphi(i)}(P_1 - p_1)] + w_2[\bar{y} + b_{\varphi(j \neq i)}(P_1 - p_1)]; (i, j \neq i) = 1, \ldots, 6 \tag{20}$$

However, it is interesting to note that if we put $(w_1, w_2) = (0, 1)$, $\hat{\bar{y}}_{p_i}$ will be converted into a traditional robust regression type estimator introduced by Nasir et al. (2018) under SRS for quantitative sensitive study variable. Hence, these estimator is a special case of the proposed class. The proposed family relies on the robust-regression tools, i.e. $b_{\varphi(i)}$ (LAD, H-M, LMS, LTS, H-MM, Hampel-M, Tukey-M) for $(i = 1, \ldots, 6)$ respectively. For deep knowledge of $b_{\varphi(i)}$, interested readers may refer to Zaman and Bulut (2018b).

To obtain MSE, let us define $\bar{y} = (1+\eta_y)\bar{Y}$, $p_1 = (1+\eta_\varphi)P_1$. Utilizing these notations $\eta_i (i = y, \varphi)$, we can write
$E(\eta_y) = E(\eta_\varphi) = 0, E(\eta_y^2) = \gamma C_y^2, E(\eta_\varphi^2) = \gamma C_\varphi^2$ and $(\eta_y \eta_\varphi) = \gamma C_{y\varphi}$.

Now, expanding $\hat{\hat{y}}_{p_i}$ in terms of η_y and η_φ as

$$\hat{\hat{y}}_{p_i} = [w_1\bar{Y}(1+\eta_y) - b_{\varphi(i)}P_1\eta_\varphi] + [w_2\bar{Y}(1+\eta_y) - b_{\varphi(j\neq i)}P_1\eta_\varphi]. \quad (21)$$

Squaring (21), applying expectation, we get theoretical MSE of the estimator $\hat{\hat{y}}_{p_i}$ up to the order n^{-1} , as

$$MSE(\hat{\hat{y}}_{p_i}) = \bar{Y}^2 + w_1^2\delta_A + w_2^2\delta_B + 2w_1w_2\delta_C - 2w_1\delta_D - 2w_2\delta_E\,, \quad (22)$$

where

$$\delta_A = [\bar{Y}^2 + \gamma\{S_y^2 + B_{\varphi(i)}\,(B_{\varphi(i)}S_\varphi - 2\rho S_y)S_\varphi\}],$$

$$\delta_B = [\bar{Y}^2 + \gamma\{S_y^2 + B_{\varphi(j\neq i)}\,(B_{\varphi(j\neq i)}S_\varphi - 2\rho S_y)S_\varphi\}],$$

$$\delta_C = [\bar{Y}^2 + \gamma\{S_y^2 - (B_{\varphi(i)} + B_{\varphi(j\neq i)})S_{y\varphi} + B_{\varphi(i)}B_{\varphi(j\neq i)}S_\varphi^2\}],$$

$$\delta_D = \delta_E = \bar{Y}^2.$$

By partially differentiating (22) w.r.t. w_1 and w_2, we obtained the optimum values as given by

$$w_1^{opt} = \left[\frac{\delta_B\delta_D - \delta_C\delta_E}{\delta_A\delta_B - \delta_C^2}\right],$$

and

$$w_2^{opt} = \left[\frac{\delta_A\delta_E - \delta_C\delta_D}{\delta_A\delta_B - \delta_C^2}\right],$$

Substitution of w_1^{opt} and w_2^{opt} in (22) provides the minimum MSE of $\hat{\hat{y}}_{p_i}$ as

$$MES_{min}(\hat{\hat{y}}_{p_i}) = \left[\bar{Y}^2 - \frac{\delta_B\delta_D^2 - 2\delta_C\delta_D\delta_E + \delta_A\delta_E^2}{\delta_A\delta_B - \delta_C^2}\right]. \quad (23)$$

The general theoretical condition of proposed vs. existing estimators as given below:

$$MSE(\bar{y}_{zi}) - MSE(\hat{\hat{y}}_{pi}) > 0$$

4. Numerical Illustration

A numerical illustration is performed utilizing the previous studies of Koyuncu (2012).

Data 1 [Source: Sukhatme and Sukhatme (1970)]

$\emptyset$ = A circle consisting of more than five villages.

Y = Number of villages in the circles.

$B_{lad} = 4, B_{lms} = 5, B_{huber} = 4.660824, B_{hample} = 4.672494,$
$B_{lts} = 5, B_{tukey} = 4.655754, B_{mm} = 4.647839, \text{var}(\bar{y}) = 4.0738$.

Data 2 [Source: Sukhatme and Sukhatme (1970)]

$\emptyset$ = A circle consisting of more than five villages.

Y = Area under the wheat crop within the circles.

$B_{lad} = 1678.281, B_{lms} = 1896, B_{huber} = 1462.839, B_{hample} = 1438.403,$
$B_{lts} = 1896, B_{tukey} = 1574.684, B_{mm} = 1573.993, \text{var}(\bar{y}) = 513592$.

For remaining characteristics of the data sets, interested readers may refer to Koyuncu (2012). The data sets are also available in the appendix.

Figures 1- 4 clearly show that our considered data sets suffer from non-normality and the presence of outliers. Hence, suitable for robust regression tools. In Table 1, results of PRE, which are figured utilizing PRE equations displayed in Sections 1, 2, and 3, are provided. Note that by ignoring fractional values in the proposed class, all members of the proposed class are providing the same results. So, we ignore the fractional part and provide a single value of PRE in Table 1. When we look at Table 1, we see that the proposed class has the maximum PRE among all estimators given in Sections 1 and 2. From the consequence of this numerical delineation, it is unmistakably concluded that all new estimators are more effective than existing and adapting estimators.

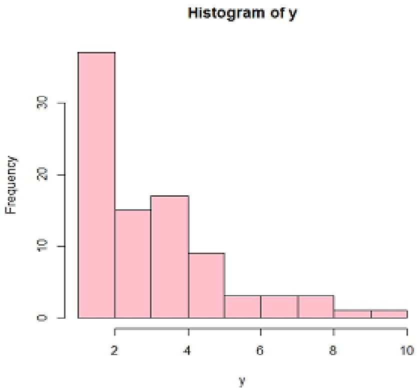

Figure 1. Population-1

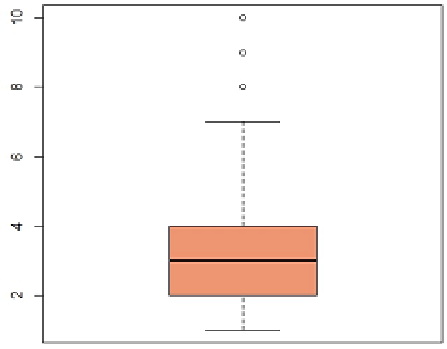

Figure 2. Boxplot Population-1

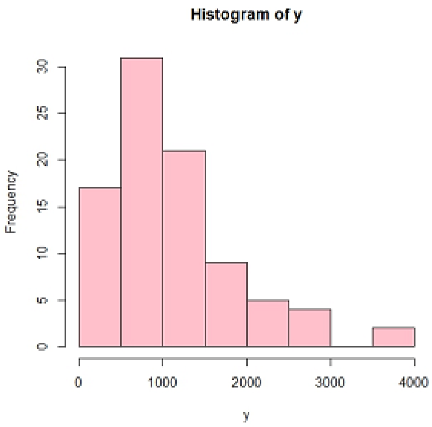

Figure 3. Population-2

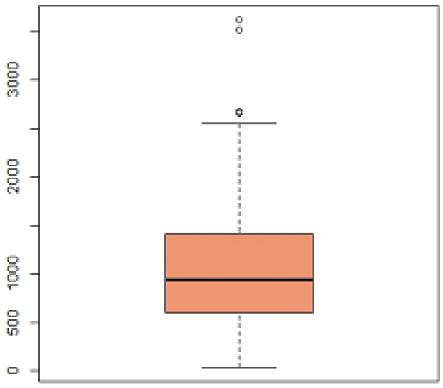

Figure 4. Boxplot Population-2

Table 1. The MSE and PRE of Proposed and Existing Estimators w.r.t. var($\bar{y}$)

Estimator	Pop-1		Pop-2	
	MSE	PRE	MSE	PRE
$\bar{y}_{zb_1}$	2.72724	149.37	326355.7	157.37
$\bar{y}_{zb_2}$	0.06246	6521.38	13236.41	3880.14
$\bar{y}_{zb_3}$	0.05682	7169.33	11947.89	4298.59
$\bar{y}_{zb_4}$	0.06021	6764.89	12766.2	4023.06
$\bar{y}_{zb_5}$	0.06442	6323.61	13621.23	3770.52
$\bar{y}_{reg}$	0.05423	7511.55	10120.65	5074.69
$\bar{y}_p$	0.05397	7547	10037	5116

5. Conclusion

This paper proposes two classes of estimators. It was discovered that the proposed robust regression estimators were more efficient than the estimators of Zaman and Bulut (2018a) and Zaman (2019). The outcomes displayed here support this conclusion by hypothetical improvement and numerical examination.

References

ABD-ELFATTA,H A. M., EL-SHERPIENY, E. A., MOHAMED, S. M., ABDOU, O. F., (2010). Improvement in estimating the population mean in simple random sampling using information on auxiliary attribute, Appl. Math. Comput., Vol. 215, pp. 4198–4202.

KOYUNCU, N., (2012). Efficient estimators of population mean using auxiliary attributes. Applied Mathematics and Computation, Vol. 218, pp. 10900–10905.

NAIK, V. D., GUPTA, P. C., (1996). A note on estimation of mean with known population of an auxiliary character, J. Indian Soc. Agr. Stat., Vol. 48, pp. 151–158.

NASIR, A., AHMAD, I., HANIF, M., SHAHZAD, U., (2018). Robust-regression-type estimators for improving mean estimation of sensitive variables by using auxiliary information, Commun. Stat. Theory Methods, doi:10.1080/03610926.2019.1645857.

SHAHZAD, U., (2016). On the Estimation Of Population Mean Under Systematic Sampling Using Auxiliary Attributes, Oriental Journal Physical Sciences, Vol. 1 (1&2), pp. 17–22.

SHAHZAD, U., HANIF, M., KOYUNCU, N., SANAULLAH, A., (2018). On the estimation of population variance using auxiliary attribute in absence and presence of non-response, Electronic Journal of Applied Statistical Analysis, Vol. 11, pp. 608–621.

SUKHATME, P. V., SUKHATME, B. V., (1970). Sampling Theory of Surveys with Applications, Iowa State University Press, Ames, USA,.

ZAMAN, T., BULUT, H., (2018a). Modified ratio estimators for population mean using robust regression based on auxiliary attribute, IJMR, Vol. 4, pp. 1–6.

ZAMAN, T., BULUT, H., (2018b). Modified ratio estimators using robust regression methods, Commun. Stat. Theory Methods, doi:10.1080/03610926.2018.1441419.

ZAMAN, T., (2019). Improvement of modified ratio estimators using robust regression methods. Applied Mathematics and Computation, Vol. 348, pp. 627–631.

APPENDIX

Data 1

$\emptyset$ = 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0.

$\boldsymbol{Y}$ = 6, 5, 4, 5, 4, 2, 4, 2, 5, 1, 3, 4, 3, 1, 1, 3, 4, 8, 2, 4, 3, 4, 4, 3, 5, 2, 3, 1, 2, 4, 3, 2, 4, 4, 1, 7, 3, 2, 2, 5, 2, 4, 6, 1, 2, 2, 3, 1, 1, 10, 5, 9, 2, 2, 3, 8, 3, 2, 5, 2, 2, 2, 2, 3, 4, 1, 5, 4, 3, 5, 7, 3, 2, 7, 4, 1, 2, 1, 1, 8, 2, 1, 3, 2, 6, 1, 1, 4, 4.

Data 2

$\emptyset$ = 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0.

$\boldsymbol{Y}$ = 1562, 1003, 1691, 271, 458, 736, 1224, 996, 475, 34, 1027, 1393, 692, 524, 602, 1522, 2087, 2474, 461, 846, 1036, 948, 1412, 438, 2111, 977, 814, 319, 583, 1150, 670, 499, 714, 1081, 389, 2675, 868, 1412, 445, 706, 642, 2050, 2530, 247, 421, 687, 941, 710, 387, 3516, 2002, 3622, 1400, 1584, 830, 167, 622, 591, 273, 781, 1101, 799, 601, 928, 1141, 1208, 1633, 902, 1286, 1299, 1947, 741, 574, 2554, 669, 1187, 852, 51, 1265, 1423, 794, 1604, 1621, 1764, 2668, 1076, 348, 1224, 1490.