

# Journal of Economic and Social Thought

[www.kspjournals.org](http://www.kspjournals.org)

Volume 5

June 2018

Issue 2

## A note on computing the standard errors of estimate of composite index

By Sudhanshu K. MISHRA <sup>†</sup>

**Abstract.** This short note proposes working out of the standard errors of estimate of composite indices when they are constructed by using intrinsically derived weights. It illustrates the proposed method, using the jackknife re-sampling technique, by an example that relates to crime of different types in Uttar Pradesh (India). Improvements are suggested through bootstrapping.

**Keywords.** Composite index, Standard error of estimate, Jackknife resampling, Crime data, Uttar Pradesh, India.

**JEL.** C43, C61, C71.

### 1. Introduction


A composite (or synthetic) index is  $Z=Xw$ , where  $Z$  is an array of  $n$  elements,  $X$  is an  $n \times m$  matrix (of  $m$  variables, called the indicator variables or the constituent variables, each one being an array of  $n$  values called replicates, cases or observations) and  $w$  is a row vector of  $m$  elements, often called weights. The weight vector may be extraneous, based on certain concept or criteria. Alternatively it may be determined on the basis of certain criteria pertaining to the properties of  $X$ . In any case, an index value is a weighted mean.


There could be many methods to obtain weights ( $w$ ) intrinsically from  $X$  (standardized to have zero mean and unit Std. deviation). For example, one may set up the criterion as  $C = \sum_{j=1}^m r^2(Z, x_j)$  or the sum of the squared values (squared Euclidean norm) of the coefficients of correlation between the composite index,  $Z$ , and the constituent variables,  $x_j$ , for all  $j$ .  $C$  is minimized. Composite indices using such weights are called Principal Component scores. This is the most popular method that has a history of over 50 years of its use. Instead of using the Euclidean norm, one may also use absolute or Chebyshev norm (Mishra, 2011). Weights based on minimization of the Shapley value norms that best equalize the mean expected marginal contribution of the constituent variables to the composite index ( $s(Z, x_j)$ ) or  $C = \sum_{j=1}^m s^2(Z, x_j)$  has also been proposed (Mishra, 2016; 2017).

### 2. A need to obtain the standard errors of estimate of composite index values

Were the values of a composite index used purely for descriptive purposes, it was not necessary to raise the issue of their standard errors of estimate. But in practice, composite indices (their values) are used for inferential purposes. They are compared very often and such comparisons are used for inference. This use necessitates obtaining their standard errors of estimate.

<sup>†</sup> North-Eastern Hill University, Shillong, 91 - 793022, New Delhi, India.

 918130397754

 . mishrasknehu@hotmail.com

A mathematical method to obtain such standard errors of estimate may be based on many assumptions that might not be realistic with respect to the data being analysed. They may also be cumbersome to work out and use. This short note suggests a simple method based on jackknife resampling to resolve this problem. Jackknife resampling is a well-established and amply applied method to obtain standard error of estimate of statistically estimated parameters where deductive (mathematically derived) methods are either inapplicable or cumbersome (Efron, 1979, 1981; Wolter, 1985; Efron & Tibshirani, 1993; Shao & Tu, 1995).

### 3. Jackknife resampling method to obtain the standard errors of estimate

Jackknife is a re-sampling method that leaves one observation (case) at a time and thus constructs n sample indices by using the weights obtained from the samples. More elaborately, let  $Y_k(n-1, m)$  be a subset of  $X(n, m)$  such that it excludes the  $k^{\text{th}}$  case. This  $Y_k$  is used to obtain weight vector  $w_k(m)$  and using this weight the composite index  $Z_k = Xw_k$  is constructed. Since  $w_k$  is based on sample,  $Z_k$  inherits its sample nature. This is done for  $k=1,2,\dots,n$  and thus we have  $Z_k(n); k = 1,2,\dots,n$ . From these  $Z_{k=1,\dots,n}$  we may obtain mean and standard deviation such that  $\bar{Z}(n) = \sum_{k=1}^n Z_k(n)$  or stated more elaborately,  $\bar{Z}_i = (1/n) \sum_{k=1}^n Z_{ik}$  where  $i=1,2,\dots,n$  refer to cases and  $k$  refers to the sample of  $n-1$  size drawn from  $X$ . Similarly,  $s^2(Z_i) = (n-1)[(1/n) \sum_{k=1}^n Z_{ik}^2 - \bar{Z}_i^2]$ .

### 4. An illustrative example

By way of giving an example, we use the crime data for 68 districts of Uttar Pradesh (India), presented in Table 1. For purpose of analysis, all crime statistics for a particular district have been divided by population so that the crime rate per lakh population is obtained. From this set, 68 samples of 67 cases (leaving 1 out of  $n= 68$ ) have been drawn and for them the sample correlation matrices are computed. For each correlation matrix, eigen values are computed and sample weight vectors are obtained. Those weights are used for computing 68 sample indices. Their mean and standard deviation are computed. The detailed results are presented in Table 2.

**Table 1. District-Wise Statistics for Major Crimes in Uttar Pradesh (India) for the Year 2014**

District	Murder	Rape	Kidnap	Robbery	Theft	Auto Theft	Riots	Crim Brch	Cheating	Griev Hurt	CrueltyHusb	Populn(Lakh)
Agra	178	77	463	218	3512	2824	413	99	382	317	389	36.11
Aligarh	179	112	471	242	2297	1660	436	129	354	285	455	29.90
Allahabad	132	109	330	125	2245	1673	129	124	511	353	355	49.42
Ambed. Ngr	24	20	59	13	86	39	37	19	43	126	41	20.25
Auraiya	34	21	99	14	165	119	20	28	78	7	151	11.79
Azamgarh	68	42	183	73	387	247	203	53	134	313	158	39.51
Badaun	120	71	164	53	331	165	9	27	68	323	62	30.69
Baghpat	83	35	121	45	332	184	56	23	61	9	90	11.64
Bahraich	53	68	202	14	258	126	73	39	93	169	0	23.84
Ballia	36	21	116	28	287	188	113	30	83	188	95	27.52
Balrampur	28	19	67	4	68	33	22	13	27	45	2	16.85
Banda	42	71	151	25	240	119	86	31	65	40	96	15.00
Barabanki	57	35	65	17	113	60	3	64	148	306	36	26.73
Bareilly	132	76	337	80	886	509	153	147	462	181	139	35.99
Basti	31	15	53	11	85	42	20	21	19	86	40	20.69
Bijnor	84	80	184	59	434	181	77	42	203	16	0	31.31
Bulandshr	171	82	291	122	841	483	158	65	17	17	0	29.23
Chandoli	20	26	78	27	156	88	39	39	152	207	113	16.40
Chitrakoot	33	23	46	6	81	45	12	29	40	16	12	8.01
Deoria	42	35	137	13	182	117	81	19	35	209	6	27.30
Etah	65	54	172	63	399	236	110	38	26	322	120	27.88
Etawah	55	30	136	113	506	385	26	54	119	100	104	13.40
Faizabad	32	37	120	25	295	156	28	48	138	208	8	20.88
Fatehpur	55	58	131	21	217	135	42	70	93	273	93	23.06
Firozabad	140	61	284	92	687	453	201	62	175	347	146	20.46
GautamB.Ngr	104	54	210	227	3680	2344	191	106	382	17	198	11.91
Ghaziabad	166	118	482	114	3392	2635	70	176	493	19	575	32.90





## 5. Conclusion

This short note proposes working out of the standard errors of estimate of composite indices when they are constructed by using intrinsically derived weights. It illustrates the proposed method, using the jackknife resampling technique, by an example that relates to crime of different types in Uttar Pradesh (India). Jackknife re-sampling results may further be refined by bootstrapping. It may give further insight as to the nature of variation in the composite index values.

## References

- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1), 1-26.
- Efron, B. (1981). Nonparametric estimates of standard error: The Jackknife, the bootstrap and other methods. *Biometrika*, 68(3), 589-599. doi. [10.1093/biomet/68.3.589](https://doi.org/10.1093/biomet/68.3.589)
- Efron, B., & Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall, New York.
- Mishra, S.K. (2011). A comparative study of various inclusive indices and the index constructed by the principal component analysis. *IUP Journal of Computational Mathematics*, 4(2), 7-26.
- Mishra, S.K. (2016). A note on construction of a composite index by optimization of Shapley value shares of the constituent variables. *Turkish Econ. Review*, 3(3), 466-472.
- Mishra, S.K. (2017). Almost equi-marginal principle based composite index of globalization: China, India and Pakistan. *Journal of Economic and Social Thought*, 4(3), 335-351.
- Shao, J., & Tu, D. (1995). *The Jackknife and Bootstrap*, Springer Verlag, New York.
- Wolter, K.M. (1985). *Introduction to Variance Estimation*, Springer Verlag, New York.



## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by-nc/4.0>).

