Application of Principal Component Analysis to Investigate the Proliferation of Armed Movements and Insurgencies in Recent Time in Sudan.

Mohamed Hassan, Sharaf Obaid Ali, Abdallah M Mohamed Zain.

[1.] Department of Statistics, College of Economics and Political Sciences, Omdurman Islamic University, Sudan, Department of Mathematics, College of Arts and Sciences, Shagra University, kingdom of Saudi Arabia
[2.] College of Computer Science, Alzaeim alazhari University, Sudan, Department of Mathematics, College of Sciences, Shaqra University, kingdom of Saudi Arabia
[3.] Department of Statistics, College of Economics and Political Sciences, Omdurman Islamic University, Sudan , Department of Mathematics, College of Sharia and Islamic Studies, Department of Management , Al-Imam Muhammad Ibn Saud Islamic University, kingdom of Saudi Arabia.
sharafobaid@gmail.com

Abstract: This article describes thepplication of principal component analysis to investigate the proliferation of armed movements and insurgencies in recent time in Sudan.
[Mohamed Hassan, Sharaf Obaid Ali, Abdallah M Mohamed Zain. **Application of Principal Component Analysis to Investigate the Proliferation of Armed Movements and Insurgencies in Recent Time in Sudan.** *Nat Sci* 2014;12(3):66-72]. (ISSN: 1545-0740). http://www.sciencepub.net/nature. 9

Keyword: Principal component, Proliferation, armed

1- Introduction:

The difficulty of the study of analyzing multi variables can be attributed in the first place to the need to understand and explain the tangled relationships between the variables that affect the phenomena understudy. A second obstacle is the plethora of data that has to be analyzed. As well as the need to include many advanced mathematical model in order to derive statistical methods that can aid statistical inference regarding multi variables reduction. Factor analysis, based mainly on the possibility to aggregating variables regarding their correlation coefficient is viewed as one of method analyze of multi variables.(5)

The motivation of this paper comes from the proliferation of armed movement in recent time in Sudan. The researcher has employed a multi variable method –factor analysis using principal component method to reduce large number of variables into considerably viewer one, however the difficulties of this approach lies in problem of naming listed variables from among larger.

It is believed that the rise of armed insurgency against ruling regimes is a factor that hampering development in all fields, equally to these rebellions bloodshed among citizen of and, indeed wars of every kind weaken internal solidarity.(7)

This study is intended to identify the major factors related to the rise of armed movements, be they economic political, foreign,…ect. Moreover , we aim to furnish our library with reference on this little studied subjects.

The importance springs from the fact that it attempts to reach solutions for a problem that has preoccupied intellectuals, leaders, government and parties in Sudan on addition solution proposed are grounded.

The researcher adopted the statistical data including factor analysis, considered one most important approaches which fact that testifies to accuracy of the results. The study follows the deductive approach in order to pinpoint the theoretical frame is used to collect, classify and present the study data through the appropriate statistical methods.

2- Hypothesis of study:

In the first place "in this research following hypothesis postulated: mainly factors can played for the rise of armed movement in Sudan,". A second hypothesis to be testified is that:"there are many potential factors that can draw recruit to these armed forces".

3- Data and Methodology:

Source of Data: A questionnaire was employed to obtain the primary data of the study, in the form of triple Likert scale with options ranging from(agree (three points), neutral (2points) and disagree (one point).

A sample method was used to determine a size on the following equation:

$$n = \frac{pqz^2}{d^2} \quad \Rightarrow \quad n = \frac{(0.5)*(0.5)*(1.96)^2}{(0.05)^2} = 384$$

Such that: n=sample size, z= confidence interval, p=proportion of those previously joined to armed movements and, d^2 = estimation error.(6)

3-1 PCA:

Principal component analysis is a variable reduction procedure. It is useful when you have obtained data on a number of variables (possibly a large number of variables), and believe that there is some redundancy in those variables. In this case, redundancy means that some of the variables are correlated with one another, possibly because they are measuring the same construct. Because of this redundancy, you believe that it should be possible to reduce the observed variables into a smaller number of principal components (artificial variables) that will account for most of the variance in the observed variables.

The main components method leads to disclosure of relationships and new interpretations have never thought of existence.(5)

Principal component analysis (PCA) involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components.(1) The best results can be obtained from the analysis is when the original variables high correlation positively or negatively The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible.

The other main advantage of PCA is that once you have found these patterns in the

data, and you compress the data, ie. by reducing the number of dimensions, without

much loss of information. This technique used in image compression, as we will see

in a later section.(3)

Apply PCA to Reduce the Dimension of the Armed Forces:

Step 1: Organize the dataset in a matrix X.

Step 2: Normalize the data set using Z-score.

Step 3: Calculate the singular value decomposition of the data matrix. X =UDV T

Step 4: Calculate the variance using the diagonal elements of D.

Step 5: Sort variances in decreasing order.

Step 6: Choose the p principal components from V with largest variances.

Step 7: Form the transformation matrix W consisting of those p PCs.

Step 8: Find the reduced projected dataset Y in a new coordinate axis by applying W to X. (2)

3-2 Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy Kaiser-Meyer-Olkin's sampling adequacy criteria (usually abbreviated as KMO) with MSA (individual measures of sampling adequacy for each item): Tests whether there are a significant number of factors in the dataset:

Technically, tests the ratio of item-correlations to partial item correlations. If the partials are similar to the raw correlations, it means the item doesn't share much variance with other items. The range of KMO is from 0.0 to 1.0 and desired values are > 0:5*5. Variables with MSA being below 0.5 indicate that item does not belong to a group and may be removed form the factor analysis.(4)

If two variables share a common factor with other variables, their partial correlation (a_{ij}) will be small, indicating the unique variance they share.

$$a_{ij} = (r_{ij} \bullet 1,2,3,......k)$$

$$KMO = (\sum\sum r_{ij}^2) / (\sum\sum r_{ij}^2 + \sum\sum a_{ij}^2)$$

If $a_{ij} = 0$, The variables are measuring a common factor, and KMO ≅ 1.0, while

KMO ≅ 0.0 if $a_{ij} = 1$, therefore The variables are not measuring a common factor.

4-Results:

The study showing that correlation matrix-which is two large-to present in this study, it apparent there is a proportional correlation and inverse one between the various variables, equally only a weak or moderate correlation exist probably due to the employment of triple Likert Scale used in the questionnaire.

Table(1): Kaiser-Meyer-Olkin)KMO) Measure of Sampling Adequacy and Bartlett's Test of Sphericity:

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.872
Bartlett's Test of Sphericity	Approx. Chi-Square	14638.191
	Degree of freedom	2145
	Sig.	.000

From table(1) display KMO measure adequacy and Bartlett's test Sphericity. It can shown that value equals to 0.872 which is more than 0.05 and this is an evidence of increase reliability of factors obtained from Factor Analysis and so, we can be sure of a adequacy of the sample. Moreover the probability value reach through Bartlett's test is 0.00 which is less than 0.05 and this can be interpreted as meaning that correlation is not equal to the identity matrix, and that some measure of correlation exist and consequently the factor analysis for the data is visible.

In table (2) : illustrating initial and extract values for the communalities the common factors can explain high proportion of variance of variable since all of these exceed 0.5 with exception of the value x48 which is equals 0.493 which approximates 0.5, indicating that 49.3% of variance within the variable value of x48 is explicable through the common factor.

Tables - (3-1), (3-2) and (3-3) display the total explained variance which comprises three parts:

Table(2): Table():Extraction Method: Principal Component Analysis (Communalities):

Communalities				Communalities		
Var	**Initial**	**Extraction**		**Var**	**Initial**	**Extraction**
X1	1	.785		**X34**	1	.756
X2	1	.738		**X35**	1	.688
X3	1	.724		**X36**	1	.682
X4	1	.715		**X37**	1	.624
X5	1	.712		**X38**	1	.666
X6	1	.734		**X39**	1	.729
X7	1	.614		**X40**	1	.632
X8	1	.722		**X41**	1	.662
X9	1	.603		**X42**	1	.729
X10	1	.568		**X43**	1	.702
X11	1	.698		**X44**	1	.710
X12	1	.496		**X45**	1	.751
X13	1	.641		**X46**	1	.594
X14	1	.536		**X47**	1	.667
X15	1	.743		**X48**	1	.678
X16	1	.740		**X49**	1	.652
X17	1	.635		**X50**	1	.656
X18	1	.550		**X51**	1	.614
X19	1	.636		**X52**	1	.683
X20	1	.702		**X53**	1	.678
X21	1	.621		**X54**	1	.637
X22	1	.582		**X55**	1	.672
X23	1	.727		**X56**	1	.676
X24	1	.687		**X57**	1	.706
X25	1	.689		**X58**	1	.660
X26	1	.672		**X59**	1	.659
X27	1	.563		**X60**	1	.529
X28	1	.702		**X61**	1	.573
X29	1	.565		**X62**	1	.696
X30	1	.731		**X63**	1	.760
X31	1	.704		**X64**	1	.683
X32	1	.814		**X65**	1	.613
X33	1	.781		**X66**	1	.548
Extraction Method: Principal Component Analysis.				Extraction Method: Principal Component Analysis.		

The first part deal with initial values which related to imaginary roots of the correlation matrix and which, specifies the factors remaining for the analysis as the factors having values less than 1 shall be excluded. Moreover the initial solution shall be undertaken in the assumption that a number of factors equals to the number of variables entered. Thus we have the following:

1- "The total column" include the Eigen roots of each factor, taken into consideration the fact that the total value of this column is equal to the number of the variables, i.e 13.325+7.666+4.027+2.728+2.728+…+0.088=66

2- "The displaying ratio of the variance explained through each factor will be calculated as follows": the variance ratio of each factor equal to:

$$\frac{Eigen\,root}{No.of.\,\mathrm{var}\,ibles} \times 100\%$$

Table(3-1)

Total Variance Explained			
Component	**Initial Eigenvalues**		
	Total	**% of Variance**	**Cumulative %**
1	13.325	20.190	20.190
2	7.666	11.616	31.806
3	4.027	6.102	37.907
4	2.728	4.133	42.041
5	2.428	3.679	45.719
6	2.125	3.219	48.938
7	1.672	2.533	51.471
8	1.644	2.490	53.962
9	1.458	2.209	56.171
10	1.305	1.977	58.148
11	1.270	1.924	60.072
12	1.191	1.804	61.876
13	1.154	1.749	63.625
14	1.078	1.633	65.258
15	1.022	1.549	66.807
16	.959	1.453	68.260
17	.932	1.413	69.673
18	.909	1.377	71.049
19	.885	1.341	72.390
20	.823	1.247	73.637
21	.777	1.177	74.815
22	.762	1.154	75.969
23	.727	1.101	77.070
24	.690	1.046	78.115
25	.681	1.032	79.147
26	.655	.993	80.140
27	.640	.969	81.109
28	.623	.943	82.052
29	.589	.893	82.945
30	.569	.862	83.807
31	.523	.792	84.599
32	.514	.780	85.378
33	.494	.749	86.127
34	.486	.737	86.864
35	.458	.694	87.558
36	.448	.679	88.237
37	.424	.643	88.880
38	.410	.621	89.501
39	.409	.619	90.120
40	.391	.592	90.713
41	.370	.561	91.273
42	.354	.537	91.810
43	.346	.525	92.335
44	.336	.509	92.844
45	.322	.488	93.331
46	.309	.469	93.800
47	.300	.455	94.255
48	.287	.435	94.691
49	.280	.424	95.115
50	.273	.413	95.528
51	.261	.396	95.924
52	.246	.373	96.297
53	.230	.348	96.645
54	.229	.347	96.992
55	.217	.330	97.322
56	.207	.314	97.636
57	.205	.311	97.947
58	.195	.296	98.243
59	.185	.281	98.524
60	.169	.257	98.781
61	.162	.245	99.026
62	.153	.231	99.257
63	.147	.223	99.480
64	.139	.210	99.690
65	.116	.176	99.866

Table(3-2)

Total Variance Explained			
Component	Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	13.325	20.190	20.190
2	7.666	11.616	31.806
3	4.027	6.102	37.907
4	2.728	4.133	42.041
5	2.428	3.679	45.719
6	2.125	3.219	48.938
7	1.672	2.533	51.471
8	1.644	2.490	53.962
9	1.458	2.209	56.171
10	1.305	1.977	58.148
11	1.270	1.924	60.072
12	1.191	1.804	61.876
13	1.154	1.749	63.625
14	1.078	1.633	65.258
15	1.022	1.549	66.807

Table(3-3)

Total Variance Explained			
Component	Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	11.154	16.900	16.900
2	7.265	11.007	27.907
3	3.372	5.109	33.016
4	2.568	3.891	36.907
5	2.238	3.390	40.298
6	2.225	3.372	43.670
7	2.219	3.362	47.032
8	1.924	2.916	49.947
9	1.770	2.681	52.629
10	1.657	2.511	55.140
11	1.634	2.476	57.615
12	1.588	2.406	60.022
13	1.584	2.399	62.421
14	1.447	2.193	64.614
15	1.447	2.193	66.807

Table(4): Unrotated Component Matrix [a]

	Component Matrix(a)							
	Component							
	1	2	3	4	5	6	7	8
X47	.713							
X64	.712							
X34	.710							
X45	.708							
X56	.698							
X58	.679							
X31	.650							
X55	.640							
X62	.639							
X57	.627							-.405
X36	.624							
X38	.617							
X63	.617							
X6	.609							
X61	.574							
X19	.572							
X44	.557	-.398						
X23	.534			.461		-.397		
X65	.533							
X37	.526							
X4	.523							
X43	.517							
X30	.510	-.446						
X32	.507	-.492		-.457				
X40	.500							
X27	.500							
X51	.497	-.366						
X33	.497	-.402		-.461				
X25	.482			.482		-.374		
X5	.482							
X49	.403	.362						
X13		-.582						
X42	.362	-.554						
X20		-.541						
X21		-.541						
X28		-.539						
X41	.353	-.535						
X35	.399	-.515						
X22		-.504						
X46	.451	-.461						
X17		.455	.423					
X39	.439	-.453						
X12		.441						
X24	.380	-.416						
X9		.412						
X66		-.409						
X11		.406	.395					
X7	.352	-.391						
X3		.384	.350					
X60		.383						
X29		.379						
X10		.364						
X48			.532					
X1			.476			.406		
X50		.387	.460					
X18			.433			.351		
X2		.399	.410					
X15				.459				
X54	.359			-.454				
X59	.351				-.371			
X26	.400					-.438		
X8			.424			.426		
X14								
X52								
X16		-.371						

a 15 components extracted.

For instance: the variance ratio explained through the first factor is $\frac{13.325}{66} \times 100\% = 20.19$

3- "the cumulative Ratio column": represent the ascending correlate variance ratio for the cumulative column.

Part two: is the Extraction Sums of Squared Loadings before rotating of factors, this section includes the same data as in part one, except for the factor that have been extracted, the second section contains the same data as of the first one. Hence, this relates to the factor making up the Eigen roots with values larger than 1.(Tables 4,5).

There are only 15 factors of that kind while the residual factors have been excluded, these factors constitutes 61% of the total variance.

	Component Matrix(a)						
	Component						
	9	10	11	12	13	14	15
X47							
X64							
X34							
X45							
X56							
X58							
X31							
X55							
X62							
X57							
X36							
X38							
X63							
X6							
X61							
X19							
X44							
X23							
X65							
X37							
X4							
X43							
X30							
X32							
X40							
X27							
X51							
X33							
X25							
X5							
X49							
X13							
X42							
X20							
X21							
X28							
X41							
X35							
X22							
X46							
X17							
X39							
X12							
X24							
X9		.394					
X66							
X11							
X7							
X3	.369						
X60							
X29							
X10							
X48							
X1	.377						
X50							
X18							
X2							
X15					-.358		
X54							
X59							
X26							
X8							
X14		.404					
X52			.432				
X16				.430			
X53				.427			

Part three: contains the Rotation Sums of Squared Loadings table include the same set of data extracted factors as in part 2 only after rotation. Here the variance ratios explained by the extracted factors after rotation have been redistributed in an equal manner, using the varimax technique for rotation.

Table (4) shows Un rotated Component Matrix.

Table(5): rotated Component Matrix [a]

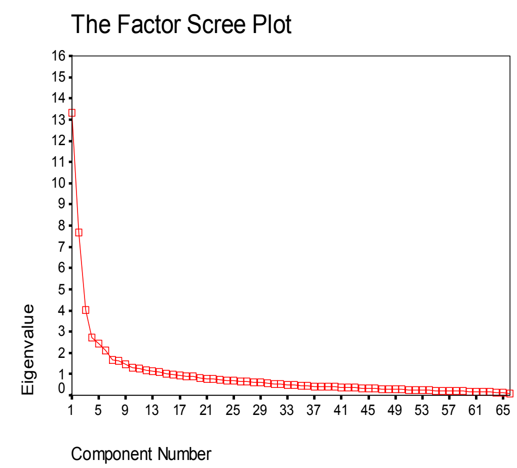
	Component Matrix(a)							
	Component							
	1	2	3	4	5	6	7	8
X64	.797							
X56	.770							
X47	.769							
X34	.759							
X38	.729							
X45	.723							
X19	.706							
X62	.688							
X58	.671							
X63	.664					.360		
X57	.660							
X37	.659							
X55	.648							
X36	.644			.359				
X31	.635			.440				
X6	.631							
X27	.567							
X61	.556					.428		
X40	.479		.377					
X20		.779						
X28		.729						
X21		.702						
X13		.668						
X35		.660						
X22		.626						
X24		.586						.440
X66		.545						
X7		.544						
X44	.364	.517						-.374
X41		.511						
X46		.510						
X43		.486						-.396
X18		.482						
X51	.381	.442						
X50			.717					
X17			.705					
X29			.636					
X49	.394		.588					
X8			.410		.356			-.352
X33		.400		.692				
X32		.466		.680				
X30		.457		.608				
X14					.663			
X11					.646			
X9					.613			
X10					.490			
X12					.395			
X59						.663		
X60						.532		
X65	.444					.483		

X2							.789	
X3							.743	
X1							.729	
X23	.444							.603
X25	.408							.459
X4	.439							
X5	.356							
X48		.376						
X42		.506						
X39		.544						
X52								
X53								
X26			.390					
X16		.393						
X15								
a 15 components extracted.								

	Component Matrix(a)						
	Component						
	9	10	11	12	13	14	15
X64							
X56							
X47							
X34							
X38							
X45							
X19							
X62							
X58							
X63							
X57							
X37							
X55							
X36							
X31							
X6							
X27							
X61							
X40							
X20							
X28							
X21							
X13							
X35							-.396
X22							
X24							
X66							
X7	.438						
X44							
X41			.482				
X46							
X43							
X18					-.372		
X51							
X50							
X17							
X29							
X49							
X8		-.382					
X33							
X32							
X30							
X14							
X11							
X9							
X10							

X12							
X59							
X60							
X65							
X2							
X3							
X1							
X23							
X25					.379		
X4	.696						
X5	.680						
X48		.605					
X54		.389					
X42			.629				
X39			.588				
X52				.776			
X53				.683			
X26					.596		
X16						.732	
X15						.715	
a Rotation converged in 12 iterations.							

Shape (1): The Factor Scree Plot for Eigen Roots



Table(5)-Rotated Component Matrix- and Shape (1) show the scree plot of Eigen values of the imaginary roots to the corresponding various factor. The figure displays the Eigen values of each factor extracted. It equally illustrated that the amount of variance to which the change in Eigen value is attributable within each of the factor rapidly diminishes in relation to the successive variable. The figure also, shows the accumulation begin to emerge between the 16 and 15 factors correspond to the Eigen value of less than 1.consequantely variables up to 17 are retained.

The first factor includes 11 variables, namely (x_{44},x_{45},x_{47},x_{46},x_{43},x_{42},x_{39},x_{41},x_{40},x_{64},x_{22}) this could designated as the religious factor.

The second factor includes 7 variables, namely (x_{32},x_{31},x_{33},x_{30},x_{35},x_{34},x_{36},x_{51}) this could designated as the foreign factor.

The third factor includes 7 variables, namely (x_{10},x_{11},x_8,x_9,x_{14},x_{17},x_{12}) this could designated as the political factor.

The fourth factor includes 7 variables, namely (x_{54},x_{53},x_{52},x_{50},x_{55},x_{49},x_{29}) this could designated as the military factor.

The fifth factor includes 3 variables, namely (x_3,x_2,x_1) this could designated as the distribution of recourses and development policies factor.

The sixth factor includes 4 variables, namely (x_{21},x_{23},x_{24},x_{25}) this could designated as the conflict between tribes among themselves on one side and central government on the other side factor.

The seventh factor includes 3 variables, namely (x_7,x_5,x_4) this could designated as economic pressure factor.

The eighth factor includes 2 variables, namely (x_{38},x_{37}) this could designated as the overlapping of border and tribes with neighboring countries factor.

The ninth factor includes 3 variables, namely (x_{62},x_{63},x_{61}) this could designated as the poor level of education factor.

The tenth factor includes 3 variables, namely (x_{26},x_{65},x_{59}) this could designated as the rise cost of limited education factor.

The eleventh factor includes 2 variables, namely (x_{57},x_{58}) this could designated as the proliferation of weapon and easy access by people factor.

The twelfth factor includes 2 variables, namely (x_{16},x_{15}) this could designated as the ineffectiveness of Sudanese diplomacy factor.

The thirteenth factor includes one variables, namely (x_{18}) this it hardly designated as it contain only one factor.

The fourteenth factor includes one variables, namely (x_{60}) this it hardly designated as it contain only one factor.

The fifteenth factor includes 2 variables, namely (x_{13},x_{56}) this could designated as the colonial heritage of militaristic tendencies factor.

The data analysis resulted the following:

The data of those who have joined the armed movements:

The most factors that lead to the emergence of the armed movements were: the take off educational, military, religious, political culture, the civil conflicts that caused by the pretext of new and old colonization, internal policy, religious conflict, the bias for some states, foreign greediness, political factor, educational policies as well as the distribution for resource and development policies, the tribes tensions and conflicts on power, economical living pressures, the take of nationality of armed forces and it rudeness, Islamic groups and parties, the weakness and underestimate of the armed forces to settle the situations, and the failure of Sudanese diplomacy.

5-Conclusion:

Regarding the factors of supporting and joining armed movements, statistically significant difference was found for sex, age, and educational level. The study reveals that, the ratio of males was 81.1% compared to that 18.9% for females. Also, we show that the age group "less than 20", :21-30", "31-40", "41-50", "51 and over were represented by ratios of 4.6%, 41.5%, 37.8%, 11.1%, and 5% respectively. In addition, we notice that the ratios of educational levels among those who joined the armed movements were 4.2%, 12.6%, 34.4%, 36.1% and 12.8% for illiterate, basic, higher secondary, university and graduate levels respectively. Also, the ratio of those supporting the rise of armed movement, compared to 49.6% for those opposing that opinion. Among the most significant factors for this phenomenon there are the religious factors, inter-border tribal conflicts, political factors such as the relation between the army and political power. Equally, there are influences of popular administration, foreign intervention, and professionalism of the army as well as prior resources of the army and its incapacity to resolve conflicts. There are economic factors such as economic hardship, unfair distribution of wealth and poor development and educational policies regarding curricula and fees.

References:

1. Brian P. C. Manley: Translation of Dr. Abdul Rahman Mohammed Abu uncle of English, The Basis for the Statistical Methods Multivariate, scientific publishing and printing presses, King Saud University, Riyadh, Saudi Arabia, 2001.
2. D.Napoleon, S.Pavalakodi, A New Method for Dimensionality Reduction using K-Means Clustering Algorithm for High Dimensional Data Set, Bharathiar University, International Journal of Computer Applications (0975 – 8887), Volume 13– No.7, January 2011.
3. Lindsay I Smith, A tutorial on Principal Components Analysis, 2002.
4. Minato Nakazawa, R practice: Factor Analysis, 2011.
5. Richard Johnson, Dean and Sharan, Arabization: Abdul Hamid patients Azzam, Dr. Review: Boualem Ben Djilali: Statistical Analysis of Multiple Variables of Applied Destination, Mars Publishing House, Riyadh, Saudi Arabia.1998.
6. William Cochran, translation Anis Kengo: Statistical Sampling Technique, third edition, King Saud University Press, 1416, Saudi Arabia.
7. A.M.Zain. Using Factor Analysis to Study the Movements of Armed in Sudan, Ph.D. thesis, 2012.

2/13/2014