NORTHERN ARIZONA UNIVERSITY
College of Business Administration

Summarizing the Effect of a Wide Array of Amenity Measures into Simple Components

Working Paper Series—05-05 | April 2005

Ronald Gunderson
Professor of Economics
Ronald.Gunderson@nau.edu

Pin Ng
Associate Professor of Economics
Pin.Ng@nau.edu

Northern Arizona University
College of Business Administration
Box 15066
Flagstaff, AZ 86011-5066

Summarizing the Effect of a Wide Array of Amenity Measures into Simple Components

Introduction

Originally proposed by Hotelling (1933), principal component analysis (PCA) is a useful variable reduction technique for situations when large numbers of observed variables can be compressed into a smaller number of artificial variables. Use of the principal components (PC) allows the researcher to reduce the number of variables into a smaller number of linear combinations of the original variables in instances where some redundancy exists among the original variables resulting from correlation among them. The resulting components contain the maximum amount of the variance occurring in the original variables with no statistical redundancy (uncorrelated components) and are geometrically non-redundant (component coefficients, also called loadings or eigenvectors, being orthogonal or perpendicular). Hence, they are considered as the benchmark for linear dimension reduction.

Since the original variables are usually standardized to a mean of zero and a variance of one in situations where variables are measured on scales with widely different ranges or where units of measurement are not commensurate, the total variance will equal the number of observed variables. The first principal component contains the largest amount of the overall variance, measured by the eigenvalue, in the original variables, whereas the second component contains the highest amount of the remaining variance after the extraction of the initial component. Subsequent components are derived in a similar manner, and each of the extracted components is uncorrelated with the others.

Principal components analysis was used by Miller (1976) to detect the underlying regional pattern of Nebraska county economies. More recently, Wagner and Deller (1998) used this technique to measure the effects of economic diversity on growth and stability. English, Marcouiller and Cordell (2000), and Gunderson and Ng (2005) employed principal component techniques to derive a smaller set of resource factors to estimate the recreation component of tourism-sector exports. Dorf and Emerson (1978), and Barkley, Henry and Bao (1998) used a similar technique (factor analysis) to arrive at factors to examine manufacturing plant location in selected nonmetropolitan communities and to determine the role that local school quality might exhibit in rural employment and population growth. Deller (2004) used the first principal component of nine different groups of amenity variables as an index measure for amenities classified into campgrounds, clubs, coastal, climate, tourism, rivers, land, tour operators and skiing in a study to link rent, wages and unemployment to amenities.

The PCA, however, suffers from significant impediments in applications. The major difficulty is that the interpretation of the PCs is generally neither easy nor straightforward (Chatfield and Collins, 1980; Kendall 1980; Jolliffe, 1986; Jackson, 1991; and Cadima and Jolliffe, 1995). To see this, we attempt to reduce the 13 variables that English, Marcouiller and Cordell (2000) used to measure urban amenities into a smaller number of PCs. The results are presented in Table 1 and the definitions of these variables are shown in Appendix A. The first PC (PC1) captures about 17.13% of the total variability of the original 13 variables and is a "block component", in which the loadings (coefficients of the linear combination of the original variables or the coefficients of the eigenvector) are all of the same sign. This first PC is interpreted generally as a weighted average of the 13 original variables. However, the weights are not all the same. The variables ABIPARKD, ABITEN2, AMUSE and ABIGOLF2 carry more weight than the others while the effects of ABIPLAY2 and FAIR are negligible. The second PC (PC2) captures about 10.27% of the total variability and is called a difference component, in which the variables associated with the positive loadings contrast those with negative loadings. The variables ABITOUR, CAMPS, ABITATT2, AMUSE and ESTURBAN oppose all the remaining ones. All the remaining PCs (PC3 through PC13) are also difference components because of the mixture of positive and negative loadings in each component.

All of the 13 original variables have very different coefficients in the PCs; some loadings are much larger than others. This makes it difficult to assess the relative importance of the original variables in affecting the PCs in terms of weighted averages or contrasts. The difficulty of interpreting the PCs stems from the weighting scheme.

Loadings of Principal Components
And Simple Components

	PCA					SCA		
Variables	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**	**B1**	**B2**	**D3**
ABIPARKD	**<u>-0.40</u>**	0.27	-0.27	0.10	0.02	0.32	0.00	0.33
ABITOUR	<u>-0.29</u>	<u>-0.33</u>	-0.05	**<u>-0.37</u>**	0.25	0.32	0.00	-0.44
ABIPLAY2	-0.02	0.01	-0.16	0.08	**<u>0.49</u>**	0.00	0.58	0.00
ABISWIM2	-0.10	<u>0.28</u>	**<u>0.46</u>**	**<u>-0.49</u>**	-0.14	0.32	0.00	0.00
ABITEN2	**<u>-0.46</u>**	0.20	<u>-0.32</u>	0.08	0.06	0.32	0.00	0.33
CAMPS	<u>-0.29</u>	-0.20	<u>0.28</u>	**<u>0.46</u>**	-0.17	0.32	0.00	0.00
ABITATT2	-0.17	**<u>-0.55</u>**	0.00	**<u>-0.36</u>**	0.24	0.32	0.00	-0.44
AMUSE	**<u>-0.35</u>**	<u>-0.33</u>	-0.08	0.20	-0.06	0.32	0.00	0.00
FAIR	-0.01	0.19	**<u>0.36</u>**	-0.05	**<u>0.54</u>**	0.00	0.58	0.00
PKLOC	-0.03	0.01	**<u>0.41</u>**	**<u>0.44</u>**	**<u>0.42</u>**	0.00	0.58	0.00
ABIGOLF2	**<u>-0.39</u>**	0.23	0.24	-0.05	-0.13	0.32	0.00	0.33
ISTEA	<u>-0.32</u>	<u>0.31</u>	-0.06	-0.18	0.08	0.32	0.00	0.33
ESTURBAN	-0.22	-0.24	**<u>0.37</u>**	-0.01	<u>-0.30</u>	0.32	0.00	-0.44
% Variance	17.13	10.27	9.23	8.60	7.90	16.13	8.13	9.97
Cumulative % Variance	17.13	27.40	36.63	45.23	53.14	16.13	24.26	34.23
Eigenvalues	2.23	1.33	1.20	1.12	1.03			

			B1	**B2**	**D3**
Optimality SCA (%):	93.33	**Correlations among simple components**	**B1**	1	0.03	0.05
			B2	0.03	1	0.02
Max (abs) correlation:	0.05 between B1 and D3		**D3**	0.05	0.02	1

Table 1. Loadings of principal components (PC1 through PC5) and simple components (B1, B2 and D3) for the group of 13 variables that measure urban amenities. Loadings in bold faces are the ones retained using Jeffers' (1967) criterion while the underlined ones are those that are retained using the less stringent rule in Rousson and Gasser (2003) of keeping those that are larger or equal to $p^{-1/2} = 0.28$, where p = 13 is the number of the original variables.

A common practice in PCA is to see whether a PC can be adequately approximated by a subset of the original variables. As a result, the original variables in the linear combination which have smaller loadings are dropped to form the "truncated PC" or "approximated PC." This truncated PC is considered "meaningful" if the remaining variables possess common features of relevance to the original problem of interest. If we follow Jeffers' (1967) criterion of keeping those variables whose loadings exceed 70% of the largest absolute loading for that PC, the first truncated PC for urban amenities will be a weighted average of ABIPARKD, ABITEN2, AMUSE, and ABIGOLF2. The second truncated PC now changes from a difference component to a simple block component with a weighted ABITATT2, the third truncated PC has also become a block component which is a weighted average of ABISWIM2, FAIR, PKLOC and ESTURBAN while the forth truncated PC remains a difference component which contrasts ABITOUR, ABISWIM2 and ABITATT2 to CAMPS and PKLOC. The fifth truncated PC has also turned into a block component which is a weighted average of ABIPLAY2, FAIR and PKLOC. On the other hand, if we adopt a less stringent truncating rule used in Rousson and Gasser (2003) by dropping those variables with a loadings smaller than $p^{-1/2} = 13^{-1/2} = 0.28$, where p is the number of the original variables, the first truncated PC becomes a weighted average of ABIPARKD, ABITOUR, ABITEN2, CAMPS, AMUSE, ABIGOLF2, and ISTEA. The second truncated PC is a difference component that contrasts ABITOUR, ABITATT2, and AMUSE to ABISWIM2 and ISTEA. The third truncated PC is also a difference component which opposes ABITEN2 to ABISWIM2, CAMPS, FAIR, PKLOC and ESTURBAN. The fourth truncated PC is the same as that obtained using the more stringent criterion while the fifth PC turns into a difference component which contrasts ABIPLAY2, FAIR and PKLOC to ESTURBAN.

It is apparent that the truncated PCs are very sensitive to the threshold that is used to truncate the smaller loadings. Also, the weighting scheme is still rather complex for both the block and difference components. Some variables carry more weight than the others. Ideally, one would like a simpler weighing scheme, in which the block component will just be an average of the variables while the positive weights will balance the negative weights in difference components. Cadima and Jolliffe (1995) argued vehemently against using truncated PCs as a precursor to interpretation and provided evidence that loadings are not reliable for determining whether or not some subset of the original variables can provide an acceptable truncated PC. The loading along with the measurement of each variable determines the importance of that variable in the linear combination, and using truncated PCs to approximate PCs can be highly unreliable.

The fact that all the PCs except PC1 are difference components, which is not an unusual phenomenon in PCA, compounds the problem because they are usually more difficult to interpret. In fact, when the values of the correlation matrix of the original variables are all positive (all variables are positively correlated), only the first PC will be a block component and all the remaining PCs will be difference components (Proposition 1, Rousson and Gasser, 2004). Correlation matrices with only positive entries arise quite often in practice when the variables measure different aspects of the same theme, e.g. the different categories of amenities in this paper. If the correlation matrix is block diagonal (has a block structure) with b blocks so that correlations between variables in the same block are larger than correlations between variables of different blocks, there will be b block components and the remaining PCs will be within-block difference components. This will be desirable so that the PCs can reveal the block structure. However, in most practical cases that involve real data sets, the sample correlation matrices will be approximately rather than exactly block diagonal and, as a result, only the first PC will be a block component. Ideally, we would like to have simple components that will have b block components to capture the block structure of the correlation matrix. In practice, one would like to replace the unequal weights and not so simple contrasts by equal weights and simple contrasts like the block components (B1 and B2) and difference components (D3 and D4) of the simple component analysis (SCA) in Table 1.

The choice of the number of PCs to keep is another difficult issue. A few criteria have been proposed over the years to help determine the number of components to be retained. The one that is probably most widely used is the Kaiser (1960) criterion which recommends keeping only the components whose eigenvalues are greater than one. Since the eigenvalue of a PC measures the variance it captures, the Kaiser criterion essentially suggests keeping the PC that extracts at least as much of the variation of the original variable. Using the Kaiser criterion, the first five PCs could be kept. The first five PCs capture about 53.14% of the variability of the original 13 variables. Another commonly use criterion is the *scree plot* first proposed by Cattell (1966). Scree is the geological term referring to the debris at the bottom of a cliff. A scree plot is a plot of the magnitude of eigenvalues against their positions. Cattell suggested keeping the PCs that were in front of the elbow or knee of the scree plot. Using the scree plot that refers to the Urban Amenities grouping in Figure 1, only the first PC would be kept. Both criteria have been studied in detail (Browne, 1968; Cattell & Jaspers, 1967; Hakstian, Rogers, & Cattell, 1982; Linn, 1968; Tucker, Koopman & Linn, 1969). The general finding is that the Kaiser criterion sometimes retains too many components while the scree plot sometimes retains too few, which is the situation we have encountered in Table 1 and Figure 1. In practice, several solutions with fewer or more PCs are investigated and the one that makes the most "sense" in interpretation is then chosen. This ambiguity in choosing the number of components is another impediment to applying PCA in practice.

The difficulty of interpreting the block PCs due to the non-uniform weighting scheme, as well as the difficulty of interpreting the difference PCs and the ambiguity of deciding the number of PCs to use prompted us to search for alternative dimension reduction techniques such as the simple component analysis (SCA) by Rousson and Gasser (2004).

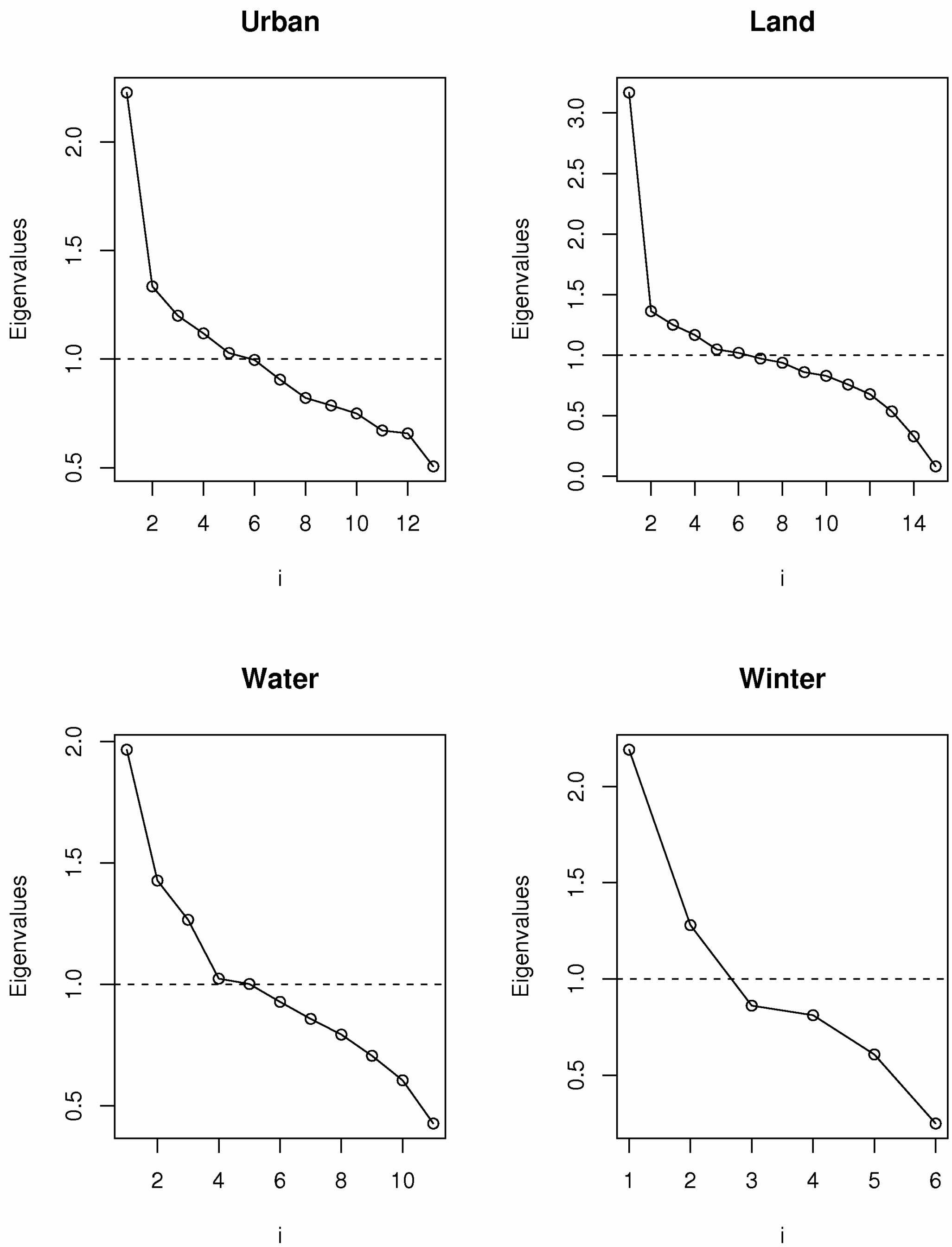

Figure 1. Screen plots for the principal components for urban, land, water and winter amenities.

Simple Component Analysis

Even though PCA is efficient in dimension reduction, extracting the maximum amount of variability and imposing exactly uncorrelated components can entail a rigid structure and is somewhat idealistic from a statistical point of view. In practical applications, less variability being extracted and slight correlation among the components might be a small price to pay for an increase in interpretability. The goal of simple component analysis is to replace the principal components by a suboptimal system of simple components (SC) which are easier to interpret. If the loss of extracted variability is low and the correlations between the SCs are low, the trade-off between optimality and simplicity will be favorable.

The SCA seeks a system of simple components that maximizes a criterion of optimality. The "simplicity" in Rousson and Gasser (2004) is characterized by (1) if there is an approximate block structure in the correlation matrix (i.e., there is more than one block of variables that are closely related to each other), there will be more than one block component so that each block component captures the block of variables that are highly correlated to each other within the block and the non-zero weights of a block component are all the same, and (2) a difference component will have "proper" contrasts in that all positive weights will be equal, all negative weights will be equal, and the sum of all weights will be 0. The simplicity can be further enhanced by requiring that all non-zero weights refer to variables of the same block.

The SCA algorithm in Rousson and Maechler (2003) allows users to choose interactively an interpretable system of SCs. This consists of three stages. In stage one, an agglomerative hierarchical procedure is used to classify the original variables into *b* blocks that correspond to the approximate block structure of the correlation matrix. In stage two, simple difference components are computed one by one by first obtaining the principal components of the residual variables given the existing simple components and then shrinking them toward simple structure by retaining the large loadings (in absolute values) and shrinking the small ones to zero. Stage three of the algorithm allows users the option to remove some of the difference-components from the system.

Data

For purposes of this study, we used the National Outdoor Recreation Supply Information System (NORSIS) data set developed and maintained by the USDA Forest Services' Wilderness Assessment Unit, Southern Research Station, Athens, Georgia. The variables in this data set are designed to capture the amenities that contribute to the overall quality of life in the 2,260 non-metropolitan counties in the contiguous United States. Following English, Marcouiller and Cordell (2000), we selected 45 of the NORSIS variables and grouped them into four amenity categories - urban facilities, land resources, water resources and winter resources. These are shown in Appendix A.

Results

All computations were performed in the *R* language and environment of statistical computing and graphics (R Development Core Team, 2004) on a Sun Microsystem's dual Sparc III 1280MHz Sun V240. The principal component analysis was performed using the function "prcomp" in the "stats" package of *R* while simple component analysis was performed using the extension packages "sca" version 0.8-5 contributed by Rousson and Maechler (2003).

We began stage 1 in the interactive SCA procedure with as many block components as the original number of variables (p) so that each block component explained $(100/p)\%$ of the total variance. We then allowed the algorithm to agglomerate blocks using the "complete linkage" in cluster analysis until the % of variance explained by every one of the agglomerated block components was higher than $(100/p)\%$ of the total variance. This is similar in spirit to Kaiser's criterion in choosing the number of PCs. In stage 2, the algorithm constructed simple difference components one at a time and reported the % of total variance explained by each of the difference components. Stage 2 was terminated when the % of total variance captured by the additional new difference component was smaller than $(100/p)\%$. In stage 3, we deleted the difference components with lower than $(100/p)\%$ of total variance explained. We also deleted any difference component that did not have meaningful contrast interpretation. The results for the four categories of amenity measures on urban, land, water and winter resources are reported in Table 1 through Table 4.

Urban Facilities

The first three simple components from the SCA for the 13 urban amenity variables were presented earlier in this paper. (See Table 1) There are two block components and one difference component. The first block component is an average of ABIPARKD, ABITOUR, ABISWIM2, ABITEN2, CAMPS, ABITATT2, AMUSE, ABIGOLF2, ISTEA and ESTURBAN which corresponds more or less to the less stringently truncated PC1. The second block component is an average of the remaining variables ABIPLAY2, FAIR and PKLOC which corresponds roughly to the more stringently truncated PC3. The difference SC does not correspond to any of the difference PCs. The correlations among all the SCs are small with the largest absolute correlation occurring between B1 and D3 at 0.05. There is practically no redundancy in the system. The three SCs account for 34.23% of the total variability compared to 36.63% for the first three PCs. The simple system of components is, hence, 34.23/36.63 = 93.33% optimal compared to the system of principal components. So we lose about 2.5% of the total variability, or less than 7% of optimality, for a simple and interpretable system.

As mentioned earlier in the paper, using the scree plot analysis only one principal component would be retained while Kaiser's criterion would include five PCs. However, under the SC analysis, we retained two block components and two difference components that can be interpreted as follows:

- B1: Urban and recreation activities and attractions
- B2: Playgrounds and fairgrounds in local parks
- D3: Contrasts swimming pools, tennis and golf activity with tour-operated and sightseeing activities, and number of organized camps and amusement places.

Therefore, while the first block focuses primarily on generally structured activities such as tennis courts, golf courses and tour related operations, the second block focuses on less structured activities such as playgrounds and fairgrounds. The difference component looks at activities within the first block component and contrasts activities that are targeted more at local users with activities that attract tourists.

Land Resources

	PCA						SCA				
Variables	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**B1**	**B2**	**B3**	**B4**	**D5**
GUIDE	0.30	0.12	-0.07	0.33	-0.28	0.06	0.41	0.00	0.00	0.00	0.00
ABIHUNT2	0.01	0.13	0.23	0.28	**-0.64**	0.14	0.00	0.00	0.58	0.00	0.00
MOUTAINS	**0.48**	-0.12	0.15	-0.11	0.07	0.02	0.41	0.00	0.00	0.00	-0.29
PVTAGAC	0.08	0.19	**0.59**	0.21	-0.04	-0.19	0.00	0.00	0.58	0.00	0.00
FSACRE	**0.51**	-0.21	0.08	-0.06	0.05	0.01	0.41	0.00	0.00	0.00	-0.29
FWSREFUG	0.03	0.22	**0.51**	0.05	0.23	-0.34	0.00	0.00	0.58	0.00	0.00
PRICAMPG	**0.36**	-0.03	-0.18	0.15	-0.05	-0.06	0.41	0.00	0.00	0.00	0.00
PUBCAMPG	0.13	0.21	-0.31	**0.54**	0.20	0.06	0.00	0.00	0.00	0.71	0.00
NPSFEDAC	0.11	0.36	-0.36	0.01	-0.25	-0.33	0.00	0.50	0.00	0.00	0.00
NRIFORAC	0.16	0.37	-0.09	**-0.39**	0.10	0.10	0.00	0.50	0.00	0.00	0.00
OFEDAC	0.00	0.06	-0.05	**0.42**	**0.55**	-0.09	0.00	0.00	0.00	0.71	0.00
RTCTRTM	0.12	**0.57**	-0.04	-0.21	0.10	0.12	0.00	0.50	0.00	0.00	0.00
STPARKS	0.10	0.32	0.16	-0.10	0.10	**0.53**	0.41	0.00	0.00	0.00	0.87
TNCPUBAC	0.03	0.14	-0.11	-0.20	-0.10	**-0.63**	0.00	0.50	0.00	0.00	0.00
NWPSAC	**0.44**	-0.25	0.00	-0.13	0.02	-0.05	0.41	0.00	0.00	0.00	-0.29
% Variance	21.13	9.09	8.34	7.78	6.98	6.80	18.15	8.87	7.97	7.42	8.43
Cumulative % Variance	21.13	30.21	38.55	46.33	53.32	60.12	18.15	27.01	34.98	42.40	50.83
Eigenvalues	3.17	1.36	1.25	1.17	1.05	1.02					
							B1	**B2**	**B3**	**B4**	**D5**
Optimality SCA (%):	92.20			**Correlations among simple components**		**B1**	1	0.2	0.08	0.11	-0.36
						B2	0.2	1	0.03	0.05	-0.03
Max (abs) correlation:	0.36 between B1 and D5					**B3**	0.08	0.03	1	0	0.01
						B4	0.11	0.05	0	1	-0.03
						D5	-0.36	-0.03	0.01	-0.03	1

Table 2. Loadings of principal components (PC1 through PC6) and simple components (B1, B2, B3, B4 and D5) for the group of 15 variables that measure land amenities. Loadings in bold faces are the ones retained using Jeffers' (1967) criterion while the underlined ones are those that are retained using the less stringent rule in Rousson and Gasser (2003) of keeping those that are larger or equal to $p^{-1/2} = 0.26$, where p = 15 is the number of the original variables.

The first block SC (B1) corresponds to the less stringently truncated PC (PC1) while the second block SC (B2) corresponds roughly to the less stringently truncated block PC (PC2). The more stringently truncated PC (PC3) corresponds roughly to the third block SC (B3). The fourth block SC (B4) does not have a close counterpart in PC. The difference component D5, which contrasts MOUTAINS, FSACRE and NWPSAC with STPARKS, also does not have a corresponding PC counterpart. While the largest absolute correlation of 0.36 occurs between B1 and D5 and is higher than the largest correlation of 0.18 between B1 and D4 in the urban amenity category, it is still quite small since only $0.36^2 = 12.96\%$ of the information is shared by the two SCs. The simple system of the five SCs is about 92% optimal compared to the system of the five PCs.

If we use the scree plot analysis for the land resources, only one principal component would be retained compared to six using Kaiser's criterion. However, under the SC analysis, we retained four block components and one difference component that can be interpreted as:

B1: Acres set aside as national forests, state parks and wilderness preservation areas

B2: Acres in national parks and other forest acreage

B3: Acres defined for specific uses including hunting, fishing, crops, pasture, grazing and recreation

B4: Public campgrounds and other federally managed acreage

D5: Contrasts state park acreage with mountain areas and national forests and grasslands

The first three block components look at various land-oriented amenities from the perspectives of the forests, parks and specific sporting and agricultural uses while the final block component is concerned with acreage available for public camping. On the other hand, the difference component identifies activities within the first block by contrasting the state parks which are more established and regularly maintained with the natural resources such as mountains and forests which require considerably less human intervention.

Water Resources

	PCA					SCA			
Variables	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**	**B1**	**B2**	**D3**	**D4**
ABIMARIN	0.26	**-0.49**	0.26	-0.06	0.10	0.35	0.00	0.33	0.29
ABICANO2	0.29	-0.11	0.34	0.05	0.44	0.35	0.00	0.00	0.29
ABIDIVE2	0.08	-0.10	0.29	-0.12	**-0.86**	0.35	0.00	0.33	0.00
ABIFISH2	0.35	-0.26	**0.44**	-0.07	-0.01	0.35	0.00	0.33	0.29
AWAWHITE	**0.48**	**0.41**	-0.04	0.11	-0.06	0.35	0.00	-0.44	0.00
WSRIVER	0.28	0.33	0.04	0.21	0.05	0.35	0.00	-0.44	0.00
LAKE	0.27	**-0.43**	**-0.37**	0.18	0.00	0.00	0.58	0.00	0.00
RUNWATER	0.18	0.00	-0.27	**-0.56**	0.00	0.00	0.58	0.00	0.00
BAYEST	0.04	-0.19	-0.18	**0.73**	-0.19	0.35	0.00	0.33	-0.87
NRIWETLD	0.21	-0.30	**-0.53**	-0.19	-0.05	0.00	0.58	0.00	0.00
RIVERML	**0.52**	0.30	-0.14	-0.03	-0.10	0.35	0.00	-0.44	0.00
% Variance	17.87	12.97	11.51	9.30	9.10	14.98	11.83	12.54	12.54
Cumulative % Variance	17.87	30.84	42.35	51.65	60.76	14.98	26.81	39.35	49.34
Eigenvalues	1.97	1.43	1.27	1.02	1.00				
						B1	**B2**	**D3**	**D4**
Optimality SCA (%):	93.83		**Correlations among simple components**		**B1**	1	0.17	-0.14	0.17
					B2	0.17	1	-0.02	0.01
Max (abs) correlation:	0.17 between B1 and D4				**D3**	-0.14	-0.02	1	-0.07
					D4	0.17	0.01	-0.07	1

Table 3. Loadings of principal components (PC1 through PC5) and simple components (B1, B2, D3 and D4) for the group of 11 variables that measure water amenities. Loadings in bold faces are the ones retained using Jeffers' (1967) criterion while the underlined ones are those that are retained using the less stringent rule in Rousson and Gasser (2003) of keeping those that are larger or equal to $p^{-1/2} = 0.30$, where $p = 11$ is the number of the original variables.

The SCs are quite different from the PCs. The SCA, however, provides much cleaner block and difference components. The loss of optimality from the four SCs is only 6% compared to the PCA with very little redundancy as measured by the largest correlation of 0.17 between B1 and D4, and also between B1 and B2.

Using the scree plot analysis for water resources would suggest that three principal components be retained while Kaiser's criterion again would result in keeping five PCs. Under the methodology used in the SC analysis, we retained two block components and two difference component that can be interpreted as follows:

- B1: Existence of manmade water-related activities
- B2: Presence of lakes, streams and wetlands
- D3: Contrasts marinas, fishing camps and organized water activities with the number of river miles in each county
- D4: Contrasts marinas and organized water activities with the number of bays and estuaries greater than 40 acres in size

In this case, the majority of the variables in the first block component reference organized water sports such as canoeing, rafting, diving and snorkeling. The second block appears to center more on the number of lakes, streams and wetlands where activities of a more unstructured nature could occur. Thus the markets in these two blocks likely consist of different users. When we consider the difference components for the water resources, it is apparent that each of these focuses on contrasts within the first block. D3 contrasts the more organized maintained activities in the marinas with the resources suitable for more remote water activities, while D4 contrasts the manmade marinas with natural bays and estuaries. In both of these instances, the types of activities and the expected users or beneficiaries are likely to differ as well.

Winter Resources

	PCA					SCA	
Variables	**PC1**	**PC2**				**B1**	**B2**
CCSFIRM2	0.30	**0.60**				0.00	0.71
ISSSACRE	0.16	**0.66**				0.00	0.71
SNOWLAND	**0.55**	-0.22				0.50	0.00
SNOWAG	0.33	-0.32				0.50	0.00
SNOWMTN	**0.58**	-0.18				0.50	0.00
SNOWFOR	0.36	0.11				0.50	0.00
% Variance	36.52	21.33				33.36	22.24
Cumulative % Variance	36.52	57.85				33.36	55.60
Eigenvalues	2.19	1.28					
						B1	**B2**
Optimality SCA (%):	94.93		**Correlations among simple components**		**B1**	1	0.18
					B2	0.18	1
Max (abs) correlation:	0.18 between B1 and B2						

Table 4. Loadings of principal components (PC1 and PC2) and simple components (B1 and B2) for the group of 6 variables that measure winter amenities. Loadings in bold faces are the ones retained using Jeffers' (1967) criterion while the underlined ones are those that are retained using the less stringent rule in Rousson and Gasser (2003) of keeping those that are larger or equal to $p^{-1/2} = 0.41$, where $p = 6$ is the number of the original variables.

The first block SC (B1) is an average of SNOWLAND, SNOWAG, SNOWMTN and SNOWFOR while the first truncated block PC (PC1) is an average of only SNOWLAND and SNOWMTN. The second block SC (B2) corresponds to the second truncated block PC (PC2). It is apparent that the block structures in B1 and B2 are much cleaner than those on PC1 and PC2. The largest correlation between B1 and B2 is only 0.18 indicating about 3% of the information is shared by these two SCs. The simple system is nearly 95% optimal. The gain in interpretability of the SCA definitely outweighs the 5% loss in optimality.

Finally, when we use both the scree plot analysis and Kaiser's criterion for winter resources we would retain two principal components. Using the SC methodology we also retain two block components but no difference components. These can be interpreted as follows:

B1: Acreage in all areas where snowfall averages over 24 inches annually

B2: Acreage devoted to cross country and downhill skiing

Although both block components focus on winter activities, the first block is primarily a measure of acreage in snow covered areas that contain considerable open space including forests, farms and mountains. The second block is primarily a measure of ski-related activities which are utilized by a separate subset of winter sports enthusiasts.

Conclusion

While PCA is optimal in alleviating the curse of dimensionality in multivariate data sets, it lacks the interpretability that is often crucial and desirable in empirical applications. The SCA, while forgoing optimality by introducing correlation among components, offers much cleaner interpretability through the construction of simple systems with sufficient block components and a simple weighting scheme. In this paper, we demonstrate that the gain in interpretability from the SCA is worth the price of sacrificing a small amount of extracted variability while introducing slightly redundant information among the components.

By employing the simple components approach for each of the four amenity groups we were able to specifically identify various amenities within specific groups. For example, the presence of manmade urban amenities such as marinas and playgrounds oppose the natural amenities such as forest acreage, river miles and state and national parks. This technique illustrates the gain in interpretation that was not evident using principal components.

Appendix A

Abbreviations and Definitions of Variables

(County level data)

Urban Facilities:

ABIPARKD	Number of parks and recreation departments
ABITOUR	Number of tour operators & sightseeing tour operators
ABIPLAY2	Number of playgrounds & number of recreation centers
ABISWIM2	Number of private & public swimming pools
ABITEN2	Number of private & public tennis courts
CAMPS	Number of organized camps
ABITATT2	Number of tourist attractions & number of historical places
AMUSE	Number of amusement places
FAIR	Number of fairgrounds
PKLOC	Number of local or county parks
ABIGOLF2	Number of private & public golf courses
ISTEA	Number of ISTEA funded greenway trails
ESTURBAN	Estimate of acres of urban/built up land from 1995 National Resources Inventory (NRI)

Land Resources:

GUIDE	Number of guide services
ABIHUNT2	Number of hunting/fishing preserves, clubs, lodges
MOUTAINS	Acres of Mountains
PVTAGAC	Acres of cropland, pasture and rangeland
FSACRE	USDA-Forest Service National Forest and Grassland acres
FWSREFUG	FWS refuge acres open for recreation
PRICAMPG	Woodall’s number of private campground sites
PUBCAMPG	Woodall’s number of public campground sites
NPSFEDAC	NPS federal acres
NRIFORAC	NRI estimate of forest acres
OFEDAC	Acres managed by Bureau of Reclamation, TVA, Corps of Engineers
RTCTRTM	Total rail-trail miles
STPARKS	State park acres
TNCPUBAC	The Nature Conservancy acres with public access
NWPSAC	National Wilderness Preservation System acreage (1993)

(Continued on next page)

Appendix A - Continued

Water Resources:

ABIMARIN	Number of marinas
ABICAN02	Number of canoe/raft outfitters & raft trip firms
ABIDIVE2	Number of diving & snorkel tours & outfitters
ABIFISH2	Number of fish camps & private/fish lakes, piers, ponds
AWAWHITE	American Whitewater Association total whitewater river miles
WSRIVER	Designated Wild and Scenic River miles (1993)
LAKE	National Resources Inventory (NRI) acres in water bodies 2-40 acres, <2 acres, >/= 40 acres (lake or reservoir)
RUNWATER	River and Stream Acres (%)
BAYEST	NRI water body >/= 40 acres (bay, gulf, estuary)
NRIWETLD	NRI wetland acres (%)
RIVERML	Nationwide Rivers Inventory total river miles, any size

Winter Resources:

CCSFIRM2	Cross-country Ski Areas Association number of XC ski firms & public XC centers
ISSSACRE	International Ski Service ski able acreage
SNOWLAND	Federal land acres in counties with > 24 inches of snowfall
SNOWAG	Agricultural acres in counties with > 24 inches annual snowfall
SNOWMTN	Acres of mountains in counties with > 24 inches annual snowfall
SNOWFOR	Acres of forestland in counties with > 24 inches annual snowfall

Original Source: USDA – Forest Service (1997). Abbreviations: FWS = Fish and Wildlife Service, NPS = National Parks Service, NRI = National Resources Inventory, XC = Cross-Country.

References

Barkley, D. L., Henry, M. S. and Bao, S. (1998). "The Role of Local School Quality in Rural Employment and Population Growth", *The Review of Regional Studies*, 28, 81-102.

Browne, M. W. (1968). "A Comparison of Factor Analytic Techniques", *Psychometrika*, 33, 267-334.

Cadima, J. and Jolliffe, I. T. (1995). "Loadings and Correlations in the Interpretation of Principal Components", *Journal of Applied Statistics*, 22, 203-214.

Cattell, R. B. (1966). "The Scree Test for the Number of Factors", *Multivariate Behavioral Research*, 1, 245-276.

Cattell, R. B. and Jaspers, J. A. (1967). "A General Plasmode For Factor Analytic Exercises And Research", *Multivariate Behavioral Research Monographs*.

Chatfield, C. and Collins, A. J. (1980). Introduction to Multivariate Analysis (London, Chapman and Hall).

Deller, S. (2004). "Wages, Rent, Unemployment and Amenities", manuscript.

Dorf, R. J. and Emerson, M. J. (1978). "Determinants of Manufacturing Plant Location for Nonmetropolitan Communities in the West North Central Region of the U.S.", *Journal of Regional Science*, 18, 109-120.

English, D. B. K., Marcouiller, D. W. and Cordell, H. K. (2000). "Linking Local Amenities with Rural Tourism Incidence: Estimates and Effects", *Journal of Society and Natural Resources*, 13, 185-202.

Hakstian, A. R., Rogers, W. D. and Cattell, R. B. (1982). "The Behavior of Numbers of Factors Rules with Simulated Data", *Multivariate Behavioral Research*, 17, 193-219.

Hotelling, H. (1993). "Analysis of a Complex of Statistical Variables into Principal Components", *Journal of Educational Psychology*, 24, 417-441 and 498-520.

Jackson, J. E. (1991). A User's Guide to Principal Components. New York: Wiley.

Jeffers, J. N. R. (1967). "Two Case Studies in the Application of Principal Component Analysis", *Applied Statistics*, 16, 225-236.

Jolliffe, I. T. (1986). Principal Component Analysis (New York, Springer).

Kaiser, H. F. (1960). "The Application of Electronic Computers to Factor Analysis", *Educational and Psychological Measurement*, 20, 141-151.

Kendall, M. (1980). Multivariate Analysis, 2nd Ed (London, Charles Griffin).

Linn, R. L. (1968). "A Monte Carlo Approach To The Number Of Factors Problem", *Psychometrika*, 33, 37-71.

Miller, R. E. (1976). "A Taxonomy of Nebraska County Economies: an Application of the Optimization Approach to Identifying a System of Regions", *Journal of Regional Science*, 16, 225-235.

R Development Core Team (2004). R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org.

Rousson, V. and Gasser, T. (2003). "Some Case Studies of Simple Component Analysis. Department of Biostatistics, University of Zurich, Zurich. (Available from http://www.unizh.ch/biostat/Manuscripts/.)

Rousson, V. and Gasser, T. (2004). "Simple Component Analysis", *Applied Statistics*, 53, Part 4, 539-555.

Rousson, V. and Maechler, M. (2003). Simple Component Analysis, Version 0.8-5, http://cran.cnr.berkeley.edu/src/contrib/Descriptions/sca.html

Tucker, L. R., Koopman, R. F. and Linn, R. L. (1969). "Evaluation of Factor Analytic Research Procedures By Means Of Simulated Correlation Matrices", *Psychometrika*, 34, 421-459.

Wagner, J. E. and Deller, S.C. (1998). "Measuring the Effects of Economic Diversity on Growth and Stability", *Land Economics*, 74, 541-556.