Clusters of Attributes and Well-Being in the US[1]

Joseph G. Hirschberg[2], Esfandiar Maasoumi and Daniel J. Slottje[3]

February, 2000
(wp)

ABSTRACT

Using ARIMA models and entropy, the dynamic evolution of several functions of aggregate income and other attributes of well-being is analyzed for statistical "similarity" in order to determine potentially distinct dimensions in multidimensional analysis of welfare and quality of life in the US. The entropy metric compares entire distributions and is more general than principal components and other correlation-based techniques for clustering. To help macroeconomic policy makers, we compare the distribution of several composite measures of well-being, including income, with the distribution of some common measures of aggregate income over the period 1915-1995.

Key Words: Time Series, Information Measures, Aggregate Well-being, Entropy.

JEL classification: C82, I31, C14

---

[1] This paper is a extended and revised version of WP#596.

[2] Department of Economics, University of Melbourne, Melbourne, Victoria, 3010, Australia ,j.hirschberg@unimelb.edu.au.

[3] Department of Economics, Southern Methodist University, Dallas, TX 75275. maasoumi@mail.smu.edu. This paper is dedicated to the memory of Denis Sargan as it deals with fundamental issues that arise in a natural multidimensional extension of his work on the evolution of wealth distribution in Sargan (1957). We thank without implicating, Tom Fomby, David Hendry, Hashem Pesaran, and many participants at our seminars

1 Introduction

Much of modern empirical macroeconomics is concerned with time series models of the conditional mean or variance of income. Less attention is paid to the unconditional distribution of income (or other variables) from which our observations may have been drawn. The welfare basis of economics requires that we learn about the determination of income (and employment), and its distribution Almost all of the classic work on distribution, both theoretical and empirical, is in one dimension, either income or wealth. Some early work developed diffusion processes which produced steady state distributions for income or wealth. Examples of this include Champernowne (1953), Gibrat (1931), Rutherford (1955) and Sargan (1957). Sargan considered a relationship representing the effects of four separate causes of change in wealth which will explain its evolution over time. This was a causal extension of the model suggested by Champernowne for income. It produced differential equations whose steady state solutions have provided durable support for log normal and Pareto distributions in macroeconomics. Similarly, Langley (1950) had analyzed the distribution of private capital. Each of these attributes can be considered as an "indicator" of well-being. There are many more that impinge upon well-being at the aggregate level, such as health, education, environment, and freedoms.

In recent times attempts have been made to formally acknowledge that well-being is a function of several arguments, including income and the GNP. Sen (1985, 1987) has examined this question. He points out that capabilities are as important as commodities in analyzing the quality of life. Thus, examining only income or only expenditures as arguments in a utility or social welfare function may be inadequate.[4]Sargan and his contemporaries did not have access to data on other indicators of well-being. We do, and well-being is the central question of macroeconomic policy that requires a better understanding of the dynamic evolution of income as well as other attributes of well-being.

Indeed there is now a rather extensive literature on the relationship between "growth" (of incomes) and "life". One interpretation of this literature is that it attempts to reduce all attribute dimensions to that of income (GNP). Easterly (1997) provides a critical survey of the largely cross section evidence in favor of such propositions. His own panel data model for 95 "life" indicators, controlling for country specific factors, finds scant supportive evidence for a positive and significant relation between growth and most measures of well-being. Correlation analysis may be missing the whole point. Firstly, for most inequality measures, such as those in the Atkinson family or the Generalized Entropy (see Maasoumi (1986)), the indices aim to represent the whole distribution of welfare attributes, not just their correlations or serial correlations. Secondly, all

[4]Kolm (1977), Sen (1985, 1987), Atkinson and Bourguignon (1982), Maasoumi (1986) and others have discussed these issues. In recent years there have been several attempts at practical implementation of this philosophy, aimed generally at providing summary indices of well-being. Also, major attempts at compilation of data, such as the United Nation supported Basic Needs (BN) and Physical Quality of Life Indicators (PQLI), may be cited. See Ram (1980, 1982a).

of this evidence suggests the existence of distinct welfare dimensions. On the other hand, 95 indicators will surely lead to some double counting of similar dimensions.

It follows that a major issue is the number of "distinct" attributes that should and could be included. When data are available on desired attributes, the determination of distinct attributes is a statistical question. The question is whether a candidate attribute adds to the statistical information set. This question is too serious to be determined by only an analysis of correlations between variables, or by the integration/cointegration properties of the attributes. One must also look at the whole distribution of each attribute and uncover their similarities. Using information theory methods, kernel density estimation, as well as ARIMA modeling, we show it is possible to do so.

Most of the work that has been done on analyzing the quality of life within a country has focused on comparisons of urban locations using hedonic price models, cf. Rosen (1974,1979), Roback (1980,1982), Berger et al. (1988) and Gyourko and Tracy (1991) for examples of work in this vein. Nordhaus and Tobin (1972) created a measure of economic welfare which adjusted GNP for leisure, women engaged in household production and urban disamenities.. Their measure still focused on GNP, however. Jorgensen and Slesnick (1990) and Slesnick (1991) have discussed the idea of a standard-of-living index by defining their index as the ratio of two levels of aggregate expenditure per capita. They posit a social welfare function and base their analysis on that aggregate function.

In this paper we analyze fifteen apparently distinct variables indicating well-being and quality of life in the US. We use an entropy metric to measure the "distances" between attributes in order to identify distinct clusters of attributes. Almost all other entropy measures record "divergence" and are not metric. This would complicate comparisons across populations and time. We first fit univariate time series models to each series. Similarity in the conditional means is revealed in this first step. Then a non-parametric kernel density is estimated for the residuals of each attribute obtained from the first step. Next we center each series at their observed value, and measure entropy distances in order to determine likely clusters. We are also able to shed light on a decomposition of the distance between these attributes into two components. One is the entropy distance between the non parametric residual distributions. The second is due to the mean differences over time.

At the empirical level, there is a tradition for this type of analysis, but it has usually been based on cross section data.[5] To our knowledge, this paper is the first to attempt to explore the quality of life in a particular geographic region over a relatively long time period [6]. For international comparisons, Ram

[5] For instance, Ram (1984) suggested Principal Components (PC) of the PQLI and the BN data, Maasoumi and Jeong (1985) and Maasoumi (1989) proposed information theoretic indices of the same data as Ram, Maasoumi and Nickelsburg (1988) proposed similar indices for the U.S. based on the Michigan Panel data on incomes as well as housing equity and education, and Slottje (1991) studied several indexing techniques based on hedonic regressions, PCs, and "ranked attributes", using many more economic and social attributes than usual.

[6] Slesnick (1991) analyzes the living standard over time. He relies of expenditure surveys

(1984) studied composite measures from cross section (international) data on quality of life and well being, as well as per capita GNP. Hirschberg et al. (1991) studied several economic and social indicators, including civil liberties and labor force participation, in a cross section analysis of international data. As they noted, while a single attribute such as income has been criticized as an inadequate welfare index, a multivariate measure may suffer from a problem of "double counting". Put differently, two apparently distinct attributes may offer almost identical "information" to the information set inevitably utilized by any indexing technique.

Our results shed light on this "double counting" question. By uncovering similar attributes and thereby allowing a reduction in dimension, a related problem is also lessened. This is the difficult and subjective question of how to weight distinct attributes (i.e., choosing a cardinal welfare function). It is easier to attach weights to a few clusters of attributes than many of their constituent variables. It is also easier to conduct sensitivity studies that determine the robustness of any qualitative inferences within sensible bounds for these weights.

The hierarchical clustering method is adopted here, revealing the successively increasing distances that must be tolerated to reduce the dimension; i.e., the number of distinct clusters. Our findings reveal that two important thresholds must be crossed in reducing welfare dimensionality. One is going from 15 dimensions to 10, the other is in reducing to fewer than four clusters. Relatively little distance needs to be tolerated to reduce ten clusters to four. We think there are at least four distinct dimensions in these 15 attributes.

The plan of this paper is as follows. In section two the 15 attributes and the data sources used by us are described. Section three describes the entropy based method of cluster analysis employed here, and section four reports our empirical findings. We provide graphs in which the evolution of national income is compared with that of several aggregate measures of well-being. Section five concludes the study.

2 Attributes of Economic Well-Being

Since our objective is to measure well being as comprehensively as possible, a total of fifteen indicators of the quality of life were selected. The attributes are from a number of sources which are listed in the appendix. The time series are for the years 1915-1995. One important selection criterion was that the observed series went back sufficiently long enough in time.

Attribute L1 is annual per capita Gross Domestic Product of the United States in real 1958 dollars. Per capita real GDP is still one of the best representations of the command over resources and how the overall economy is doing. All of the attributes that are measured in dollar terms have been deflated by the GNP deflator to put them in real terms and in per capita terms where appropriate. This is necessary to eliminate the momentum effects which would always

and does not have contiguous data nor does he go back in time as we do here.

give later years higher values and consequently higher rankings. We adopt the point of view that per capita GDP contributes positively to well being.

L2 is the inverse of the Infant Mortality Rate (IMR). This variable is included because it may be interpreted as a proxy for deprivation suffered by those at the lower end of the wealth distribution. It also serves as a painful reminder that while the U.S. may have a powerful economy in other ways, it also has one of the highest IMR's among developed western countries.

Attributes L3 and L4 are the life expectancies for males and females, respectively. Life expectancy is associated with better medical, sanitary, and other quality of life conditions. We are aware that this view is not universally held.

The employment rate is L5 (1 minus the unemployment rate) and mean income per household is L6. Persistent unemployment is highly correlated with poverty and many other social ills. Mean real income per household is an indicator of the household unit's control over resources. It is assumed that as both L5 and L6 increase, the quality of life increases. L7 is the number of physicians for every 1,000,000 people in the population. The eighth attribute, L8, is the total number of rural and urban federal highway miles per capita. L9 and L10 are the number of telephones per 1000 people, and the total number of households with radio receivers which are more of interest in a historical context than for predictive purposes since both technologies have long since reached almost 100% saturation. These attributes are included to capture the quality of health care, the ability to be mobile, and the ability to communicate at the simplest level. It is assumed that all four indicators are positively related to the living standard. The inverse murder rate and % of children ages 5 to 17 enrolled in school are attributes L11 and L12. An environment that is stressful from fear of attack is less desirable. Dropout rates and average number of years of schooling completed may be more appropriate than L12, but the data are not available back to 1915. Higher levels of schooling suggest an increased ability to enjoy life. Increased schooling also has positive externalities such as lowering the crime rate and the unemployment rate.

The circulation of daily newspapers per capita in the United States is L13. It is included as an indicator of access to public information and as an indicator of the ease of acquiring this information. This variable is presumed to be positively correlated with the quality of life although it appears to be declining as it becomes an obsolete form of communication. The annual rate of real GNP growth is L14. This is a proxy for productivity which is assumed to be welfare enhancing. This position is open to debate and may be reassessed by future philosophers and historians. L15 is % of GNP not for defense expenditures in real terms. It is assumed the less a country is in war, or spends on defense, the more it has to support other welfare needs.

The plot of each time series is given in Figure 1. All of these attributes have been rescaled to have a mean of zero and a variance of one and the values on the horizontal axis are the last two digits of the year. In the next section the cluster method is defined and its construction explained.

(figure 1)

3 Cluster Analysis

Cluster analysis has been described as "... the art of finding groups in data" (Kaufman and Rousseeuw 1990). In its most common form it is used to define groups of observations which have multidimensional measures. Early forms of such clustering have been used in the classification of plants and animals as surveyed by Holman (1985). In these cases each of T observations is defined by a k by 1 vector of attributes (x_i) such as number of teeth, the length of bones, the weight, etc. Thus the entire data set can be defined as the T by k matrix X with the x_is as the rows.

In economics one of the first applications of cluster analysis was for the aggregation of industrial sectors within an input-output table (see for example Blin and Cohen 1977). More recently with the advent of large panel data sets there has arisen an difficulty in the interpretation of a large number of estimated models. Such data sets as the Penn World Tables (Summers and Heston 1991), supermarket scanner data and large scale panel surveys such as the Michigan Panel Study of Income Dynamics allow for the estimation of separate models for each cross section observational unit. For example, Hirschberg and Dayton (1996) estimate a set of 49 models of intra-industry trade patterns in processed food commodities using a combination of 24 years of United Nations bilateral trade data between 30 countries and the Penn World Tables. Given the difficulties of interpretation for 49 sets of 92 parameters a cluster analysis was designed to determine those commodities that were most similar using a similarity measure based on the Wald test statistic of the estimated parameters. In another recent econometric application, Hobijn and Franses (2000) also use data from the Penn World Tables estimate the time series properties of the per capita productivity for 112 countries over 29 years to establish which economies are converging to "clubs" of similar forms of productivity growth.

Clustering of time series has a long history. Primarily these methods have been applied to the problem of identifying regime changes in a particular series. One of the earliest papers in this area is by McGee and Carlton (1970). Again this is similar to the animal classification problem in which observations are clustered although here the clusters determine similarities in time periods. The clusters are formed with the additional condition that clusters can only be formed with adjacent observations. The difference between these types of methods and the method proposed here is in the form of the data. The cluster analysis used in this paper also starts with a multidimensional set of measures that can define a T by k matrix X with time as the observations and the attributes of well-being as the columns. However, in this case we use the transpose of X and cluster the attributes instead of the observations. Thus the pattern of time change defines the interrelationship between the attributes. The objective of this analysis is similar to the use of principal components analysis to reduce the dimensionality of a set of data except that our method is transparent in the process of combination and the new summaries are defined by individual series and not on combinations of all series. Our analysis could be done using a distance metric based on the correlation matrix. However methods that use the estimated second moments matrix are prone to outlier influences.

The objective of our method is to cluster different time series. This is also the objective of the work of Piccolo (1990) and Maharaj (1995). Both propose grouping time series based on the parameters of an ARIMA model fit to the time series. Piccolo looks only at the parameter estimates while Maharaj uses the Wald-test statistic as employed in Hirschberg and Dayton (1996) to cluster the estimated ARIMA parameters with the estimated standard errors as well. A difficulty with both the Piccolo and Maharaj studies is that they assume that a parameterized model explains the series under review entirely. In the present case some of our series are integrated and some are not. But comparisons across ARIMA models of different orders of integration involve a comparison of dissimilar parameters. In addition, the Wald-test distance used by Maharaj although it includes the accuracy of the estimate as well as the parameter values violates the triangle inequality. The violation of the triangle inequality means that the distance between series A and C can be greater than the distance from A to B plus the distance from B to C. Thus, one cannot use geometric analogies to interpret the results. In our case the central characteristic (information) is the distribution of the attribute over time which we model in a nonparametric manner. Clusters are formed by groups of attributes that are most "similar", according to our metric entropy, in their distribution across years.

The method we use falls under the general class of hierarchical agglomerative clustering techniques (see Kaufman and Rousseeuw 1990 for details). To begin with, each attribute is the only member of each cluster. Then, using the entropy affinity measure of fusion or distance, each cluster is considered for association (clustering) with every other cluster, in successive stages. At first, 15 clusters/dimensions are available, then 14, 13, . . . , etc., until there remains one cluster with all the attributes in it. We use this class of methods due to the ability to describe the entire set of clusters with graphical methods. The measure of distance between clusters we use is the entropy affinity measure (ρ^*_{ij} defined below) which provides a means for comparing the distribution of two random variables. This measure requires that a distribution be defined for each attribute. We do this by estimating a nonparametric distribution for the unpredictable innovations to each series and then moving the location of this distribution of innovations to reflect the location of the series. The shapes of the distributions are estimated using a kernel density estimate based on the residuals from an ARIMA specification fit to the series and the locations of each distribution are based on the observed values of the levels of the series for that year. In this way we have captured not only the long-run shifts in the location but also the nature of the innovations to the series that may not be anticipated. If the series were all assumed to have the same distribution of innovations then the clustering would be based solely on the location of the series. In contrast if the series all had very different distributions of their innovations then the location differences would not matter. This method provides a method that allows both the location and the characteristics of the distribution of the unexplained variation in the variable to be accounted for in the combinations.

The Bhattacharyya (1943) and Matusita (1967) entropy affinity measure between two distributions (ρ^*_{ij}) is defined as (see Maasoumi (1993) for details):

$$\rho^*_{ij} = \int_{-\infty}^{\infty} f_i(x)^{1/2} f_j(x)^{1/2} dx$$

where $f_i(x)$ and $f_j(x)$ are the densities of the two attributes being compared. The distance measure we use is given by $D(i, j) = 1 - \rho^*_{ij}$. It is zero if the two densities are identical and one if they are entirely different.

The process of clustering is detailed below following the 5 steps given below.

(1) We estimate a density for the innovations in each attribute's time series. This is done by fitting a set of time series models to each attribute's time series after each indicator has been scaled with a mean of 0 and a variance of 1. The residual densities can be very similar even for independent innovations. But the attributes can have trending behavior and conditional means that are similar or completely unrelated[7]. We center the distributions on the actual observed values at each point in time and compute the corresponding $D(i, j)$ values for the entire sample.

The models were either ARMA(1,1) or ARIMA(1,1,1) depending on the potential presence of a unit root in the series.

$$x_{i,t} = a_i x_{i,t-1} + d_i war_t + \epsilon_{i,t} + b_i \epsilon_{i,t-1},$$
$$\text{or, } \Delta x_{i,t} = a_i \Delta x_{i,t-1} + d_i war_t + \epsilon_{i,t} + b_i \epsilon_{i,t-1}$$

Where a, b, and d are the parameters which we estimate, $x_{i,t}$ is the series, $\epsilon_{i,t}$ is the random error, and war_t is a dummy variable for the years during the second world war. These models are found to be the most parsimonious models that capture the non-random component of these series.

Except for L5 (the employment rate), L13 (the newspaper circulation per capita), L14 (the rate of GNP growth) and L15 (the % of GNP not for defence), we could not reject the hypothesis of a unit root for all the other series. In each case we used the Portmanteau test to test for randomness of the estimated residuals from the model fit and we found that we could not reject the null hypothesis of randomness at better than a 90% level in almost all cases.

(2) Using a normal kernel and 1/4 the window width specification recommended by Scott (1979),[8] we estimated a density function for the estimated innovations of each series[9]. Figure 2 displays the plots of these densities evalu-

[7] In contrast to correlation analysis where only variation around the means matters, in this analysis the entire distribution is considered in the comparison.

[8] The window or band width (h) is $h = 3.49 \text{ sd } n^{-1/3}$, where sd is the standard deviation and n is the number of observations. This bandwidth often resulted in an overly smooth density for these series so we used $h/4$ as the bandwidth here. Other bandwidths were used such as the one found by Bowman (1985) $h = 2 \text{ sd } n^{-1/5}$ however they resulted in no difference in the clustering results reported here.

[9] Other kernels were tried including the Epanechnikov (1969), the triangular, and the rectangluar with vary little variation in the resulting distances computed.

ated at 1000 equally spaced points [10]. The estimated densities shown in figure 2 were then located according to the observation for a particular year in order to construct $\hat{f}_{it}(x)$. For example figure 3 plots the density estimate for GNP Growth (series 14).

(figure 2)

(figure 3)

(3) We compute the average of the $\overline{D}(i,j)$ over the entire time frame. This metric not only compares the values of the attribute's time series as it shifts over time but it also incorporates the distributional information in the innovations. Even though two attribute's time series may move together, if they have different innovation densities (shocks or volatilities, for instance), they may not be considered to be "close" to each other. Specifically, the overall distance measure $\overline{D}(i,j)$ is defined as the scaled average of our entropy distance for the entire series.

$$\overline{D}(i,j) = 100\left\{1 - \frac{1}{T}\sum_{t=1}^{T}\left[\int_{-\infty}^{\infty} \hat{f}_{it}(x)^{1/2}\hat{f}_{jt}(x)^{1/2}dx\right]\right\}$$

where $\hat{f}_{it}(x)$ is the estimated density for the innovations for series i located at the observation for time t, $\hat{f}_{jt}(x)$ is the corresponding estimated density for series j at time t, and $T = 81$. Table 1 reports the overall distances between the attribute. The permissible range is 0 to 100.

(table 1)

(4) We apply the hierarchical process to sequentially form clusters based on the series that are closest in proximity. From Table 1 the two closest series are female and male life expectancy (L4 and L5) with a distance of 15. These two are combined to form the 14th cluster with the other thirteen separate attributes/clusters now being considered as candidates for entry for the case of 13 clusters. To form 13 clusters we either add another variable to the two life expectancies to form a new cluster with three members, or we form a new cluster consisting of two or more of the remaining series. In this case we find that the employment rate (L5) and the GNP growth rate (L14) are the closest to each other with a distance of 32 and they are combined to form a the next cluster. We use the average linkage method which computes the value of $\overline{D}(i,j)$ between each candidate attribute and the two members of the existing clusters. The attribute(s) with the smallest such distance joins in the existing cluster. If none is closer to an existing cluster than to some of the remaining variables, closest combinations amongst the latter will form new clusters.

(5) Determine the number of clusters based on a stopping rule. As new clusters are formed the distance between the new members and the existing members gets larger and larger. From Table 1 the first distance was 15 then

[10] The number of points used did not have an effect on the clustering when at least 200 were used. The horizontal scale in these figures is the same and the density plots are centered at 500. The vertical axis in every case is scaled differently so that the integral of each density is equal one.

32 then 49 thus each cluster requires a greater distance be spanned. Based on this progression of distances we determine when to stop clustering. This is done by determining a maximum distance that must be tolerated for a new entry to be placed into an exiting cluster. This question has been the subject of a number of studies that deal with this "stopping rule problem". Here we have adopted the same kind of methods used in factor analysis based on the "scree" plots. These normally plot the eigen values of variables for elbow points. Mojena (1977) suggests a similar method for the examination of the distances needed to form successive clusters. Accordingly, we plot the distances between the variables that formed a given cluster at each clustering stage, along with the change in these distances to find an inflection point[11].

4 Clustering Results

The results of the cluster analysis are summarized in the dendrograms in Figures 4 and 5. The dendrograms (or tree diagrams) show the genealogy of the clusters as they are formed. In Figure 4 the number of clusters is on the horizontal axis and the cluster membership is on the vertical axis. This figure shows how the clustering method combines the clusters from the case where each is in its own cluster (when there are 15 clusters at the far left hand edge of the figure) to the case in which all the variables are in the same cluster (at the far right hand edge). For example, from this diagram we can see that the first two series to be combined (the move from 15 to 14 clusters) were the two life expectancy series (L3 and L3). The transition from 14 clusters to 13 clusters was accomplished by the formation of a new cluster as the combination of the employment rate (L5) and the GNP growth rate (L14). This diagram can also be used to investigate the genealogy of any of the clusters to show how they were combined and which series are contained in each.

(figures 4 and 5)

Figure 5 is another version of the dendrogram. In this case distance is depicted on the horizontal axis. We can see that the distances between most series that are included in the same cluster lie in a range of 15 and 85. Note from this figure that the within-cluster distances are relatively small until one reaches four clusters. But as we move to combine the series into 3 or fewer clusters we are combining series that are much further apart. From Figure 4 it can be noted that the distance needed to form the cluster that includes all the series was 85 (note this is less than the maximum in Table 1 because we are averaging distances between clusters). Figure 6 shows a plot of the distance

[11] The decision to stop may be made on the basis of the statistical significance of the distances. In related work with a normalized version of our entropy measure, Granger, Maasoumi, and Racine (1999) find that it has predictably large bootstrapped standard errors. Density estimators tend to exhibit this behaviour. Such a bootstrap study is beyond the scope of this paper. But our experience suggests that the major kinks in the scree diagrams correspond to the likely significant changes in the distances.

between the two closest clusters which was spanned to form the next cluster. Thus this plot starts goes from 85 for the distance to form 1 cluster to 15 for the formation of 14 clusters from 15 series. Figure 7 shows the first differences from the plot in Figure 6 and these show when the distances needed to span to combine the next clusters changes. The plot in Figure 6 show a considerable drop between the distances needed to combine to 3 clusters and 4 clusters. There is a marked spike in Figure 7 at 3 clusters indicating that a much greater distance was tolerated in order to combine the 4th and the 3rd clusters than was needed in the formation of any new cluster since the reduction from 11 to 10 clusters. Note that all of this information is contained in the distance scaled dendrogram in Figure 5 and we only show these for the other clustering discussed below.

(figures 6 and 7)

It would seem that the following three clusters, at least, should be regarded as distinct welfare dimensions:The first cluster contains Highway miles per capita (L8), and the number of radios per household (L10). The second cluster includes GNP per capita (L1), the inverse Infant Mortality Rate (L2), Disposable income per capita (L6), and the number of physicians per capita (L7). In comparing the distances between the series in this cluster we find that Disposable income per capita (L6) has the smallest average distance to any of the other members of the second cluster. Finally, the third cluster contains male and female life expectancy (L3 and L4), the employment rate(L5), the inverse homicide rate (L11), the % of children aged 5 to 17 in school (L12), the newspaper circulation per capita (L13), the growth rate of GNP (L14) and the % of GNP that is non-defense spending. In the third cluster we find that the growth rate of GNP (L14) is the series that has the smallest average distance to any of the other series in this cluster.

The number of phones per household (L9) is the one series that is still not clustered with any other series at this stage. From table 1 we note that this series is closest to female life expectancy (L4) but it is further from this series than the other members of the third cluster.

The four groups of series in these clusters can then be identified as: 1) the Highway miles per capita (L8)[12], 2) the Disposable income per capita (L6) cluster, 3) the growth rate of GNP (L14) cluster, and 4) the number of phones per household (L9) cluster.

We can decompose the influence of the innovations from the changes in the locations of the densities. Consider the hypothetical case where all the series are assumed to have the same innovation distributions with locations that change. For example if the series have densities that are Normal and their variances are equal to one then the distance measure $\overline{D}(i,j)$ can be shown to be given by[13]:

[12] The Highway miles per capita is less subject to potential error introduced by the approximation of the radios per household series as detailed in the Appendix.

[13] This is due to the result that: $\int_{-\infty}^{\infty}\left(\frac{1}{\sqrt{2\pi}}\exp(\frac{-(x-a)^2}{2})\right)^{1/2}\left(\frac{1}{\sqrt{2\pi}}\exp(\frac{-(x-b)^2}{2})\right)^{1/2} = e^{-\frac{1}{8}(a-b)^2}$

$$\overline{D}_n(i,j) = 100\left\{1 - \frac{1}{T}\sum_{t=1}^{T}\left[e^{-\frac{1}{8}(\mu_{it}-\mu_{jt})^2}\right]\right\},$$

where μ_{it} is the location of the ith series at time t. Figure 8 is the dendrogram based on $D_n(i,j)$ where the values for the series by year are the locations. We see that the time profiles that are the most similar in Figure 1 cluster together in this analysis. Note that these clusters are closest to the type of analysis that would be performed in traditional cluster analysis on the basis of distance measures such as the Euclidean. From Figure 8 we note that the a number of series are very close together while the relative distances of others is much greater. The cluster that combines the life expectancies, highway miles per capita, radios per household, phones per household, physicians per capita, GNP per capita, inverse infant mortality, and disposable income per capita is much more homogenous than any cluster made up of the other series. Note this is very different from Figure 5.

(figure 8)

In order to isolate the effect of the innovation densities on these result, we can cluster only the estimated densities of the innovations. In this case we define a new distance :

$$\overline{D}_r(i,j) = 100\left\{1 - \int_{-\infty}^{\infty} \hat{f}_i(x)^{1/2}\hat{f}_j(x)^{1/2}dx\right\}$$

where the estimated densities $\hat{f}_i(x)$ and $\hat{f}_j(x)$ are located at zero and compared directly. These clusters are given by the dendrogram in Figure 9. One interpretation of this case is that locations do not matter - thus the long-run aspect of the series are removed in this case and the equivalent to figure 3 would all line up with the same location. It can be seen that the most similar densities from Figure 2 (inverse infant mortality and the number of radios per household) are clustered together first in this diagram. In addition, we can see that GNP growth (L14), with the most diffuse density estimate from Figure 2, is one of the last series to be included in a cluster! In this way one can evaluate to what extent the clustering using $\overline{D}(i,j)$ is influenced by the nature of the innovations to the series as compared to the distances in the locations.

(figure 9)

We note that two I(1) variables may or may not be considered as "similar" in the sense of our information measure. Indeed integrated variables are similar in terms of trending properties, and once differenced appropriately, they are I(0) variables which can have very different distributions. It seems perhaps obvious to note that not all stationary variables are similar! The same variables, should they be also cointegrated, produce I(0) variables in linear combinations. This too does not necessarily guarantee "similarity" in the larger sense. Only if the innovation densities are also very similar, or their differences dominated by the means of the series, is cointegration predictive in our context. There are many series that are known to be cointegrated. But as members of welfare enhancing clusters they would make for very strange bedfellows.

5 Conclusions

Our study suggests that there are at least four distinct dimensions in the 15 attribute set we have analysed. This casts further doubt on unidimensional analyses of well-being. Correlation based studies that entirely focus on linear co-movements between variables may neglect important informational content in the whole distribution of variables. This is especially important in inequality studies.

From the first row of Table 1 one can find the distances between the traditional measure of well-being in an economy, GNP per capita (L1) and all the other series considered here. Not surprisingly Disposable income per capita (L6) is the closest. Figure 10 provides a graphic view of the similarity between the series that best typifies each cluster and GNP per capita. Note that except for cluster 2 which includes GNP per capita most of the clusters appear to climb more slowly in the last 15 years. The most interesting difference between these series is that clusters 1, and 4 seem to be approaching a plateau. And series in cluster 2 appear to be stabilizing. From Table 1 we can determine which series is closest to all the other series. The closest on average is the growth rate for GNP (L14) with an average distance of 61.99 with Male life expectancy (L3) as a close second with an average of 62.06 and a lot lower variability. Alternatively by using the minimax criterion we find that Male life expectancy has the smallest maximum distance to any other series in the set. Figure 11 shows how these series compare to the GNP per capita series. Again we note that the other series are quite different from the GNP per capita. The different behavior of these other series indicates that to some degree increases in aggregate well-being may be bounded by physical limits. Thus increases in production as measured by GNP does not increase life expectancy at a constant same rate.

(figures 10 and 11 here)

In addition to the identification of clusters we can also aggregate these clusters to create new summary series that incorporate the information contained in the four individual clusters. Using the averaging methods that can be applied here as proposed in Maasoumi (1986) we may create a new combined series. In Figure 12 we show how a simple average of the 4 representative series, a weighted average based on the number of series in each cluster, and a weighted average based on the inverse of the sum of the square error (SSE) from the ARIMA model fit to the series. This average down-weights those series with large relative variances such as the growth of GNP which are hard to predict and increases the contribution of those series with smaller relative variances, those that are more predictable. From figure 12 it can be seen that the new summary formed by weighting with the inverse of the SSE is closest to the GNP per capita series however it also displays the plateau effect.

(figure 12 here)

Clustering techniques can provide useful non-parametric means of identifying attribute groups that may then be used in multidimensional welfare analyses. Aggregation by clustering may reduce the chance of double counting highly similar attributes. Secondly, aggregation may be desirable if measurement errors and noise in some attribute data is suspected. Thirdly, although the entire time period was used to establish the distance matrix in table 1 (the limit in the summation that defines $D(i,j)$ is T), distance matrices for subperiods of the total data series could have been computed to establish a how the information content of these series as changed over time.

6 Acknowledgements

We are grateful to Thomas Fomby, the anonymous referees and Hashem Pesaran for many helpful suggestions and comments. Also we thank the seminar participants at Southern Methodist University and the conference participants at the 1996 Australasian meetings of the Econometric Society at University of Western Australia for comments on earlier versions of this work.

References

[1] Atkinson, A. B. and F. Bourguignon (1982), "The comparison of multidimensional distributions of economic status," Review of Economic Studies, 12, 183-201.

[2] Blin, J. M. and C. Cohen (1977), "Technological similarity and aggregation in input-output Systems: A Cluster-Analytic Approach", Review of Economics and Statistics, 59, 82-91.

[3] Blomquist, G., M. Berger and J. Hoehn (1988), "New estimates of the quality of life in urban areas," American Economic Review, 78, 89-107.

[4] Bhattacharyya, A. (1943), "On a measure of divergence between two statistical populations defined by their population distributions," Bulletin Calcutta Mathematical Society, 35, 99-109.

[5] Bowman, A. W. (1985), "A comparative study of some kernel-based nonparametric density Estimators," Journal of Statistical Computation and Simulation, 21, 313-327.

[6] Champernowne, D. G. (1953),"Model of income distribution," Economic Journal , 63, 318-51

[7] Easterly, W. (1997),"Life During Growth," the World Bank (mimeo).

[8] Epanechnikov, V. A. (1969), "Nonparametric estimation of a multidimensional probability density", Theory of Probability and its Applications, 14, 153-158.

[9] Gibrat, R. (1931), Les inegalites economiques, Paris.

[10] Granger, C. W., E. Maasoumi, and J. Racine (1999), "A new metric entropy measure of dependence," Working paper (in progress), Department of Economics, Southern Methodist University, Dallas, TX 75275-0496.

[11] Gyourko, J. and J. Tracy (1991), "The structure of local public finance and the quality of life," Journal of Political Economy, 99, 774-806.

[12] Hirschberg, J. G., E. Maasoumi and D. J. Slottje, (1991), "Cluster analysis for measuring welfare and quality of life across countries," Journal of Econometrics, 50, 131-150.

[13] Hirschberg, J. G., and J. R. Dayton, (1996) "Detailed patterns of intraindustry trade in processed food," in Industrial Organization and Trade in the Food Industries, I M. Sheldon and P. C. Abbott eds., Westview Press, Boulder, Colorado, 141-159.

[14] Hobijn, B. and P. H. Franses, (2000), "Asymptotically perfect and relative convergence of productivity", Journal of Applied Econometrics, 15, 59-81.

[15] Holman, E. W. (1985), "Evolutionary and psychological effects in pre-evolutionary classifications", Journal of Classification, 2, 29-39.

[16] Jorgensen, D. W. and D. Slesnick (1990), "Inequality and the standard of living," Journal of Econometrics, 43, 103-120.

[17] Kolm, S-Ch. (1977), "Multidimensional egalitarianism," Quarterly Journal of Economics, 91, 1-13.

[18] Kaufman, L. and P. J. Rousseeuw, (1990), Finding Groups in Data: An Introduction to Cluster Analysis, John Wiley & Sons, New York.

[19] Langley, K. M. (1950),"The distribution of capital in private hands in 1936-38 and 1946-47," Bulletin of the Oxford University Institute of Statistics, 12, 339-59.

[20] Maasoumi, E. (1986), "The measurement and decomposition of multidimensional inequality," Econometrica, 54, 991-997.

[21] ____(1989), "Composite indices of income and other developmental indicators: a general approach," Research on Economic Inequality, 1, 269-286.

[22] ____(1993), "A compendium of information theory in economics and econometrics", Econometric Reviews, 12, 137-181.

[23] Maasoumi, E. and J-H. Jeong (1985), "The trend and the measurement of world inequality over extended periods of accounting," Economics Letters, 19, 295-301.

[24] Maasoumi, E. and G. Nickelsburg (1988), "Multivariate measures of well-being and an analysis of inequality," Journal of Business and Economic Statistics, 6, 327-334.

[25] Maharaj, E. A., (1995), "A significance test for classifying ARMA models," Proceedings of the 1995 Econometrics Conference at Monash, Monash University, Department of Economics, 219-251.

[26] Matusita, K. (1967), "On the notion of decision functions," Annals of the Institute of Statistical Mathematics, 19, 181-192.

[27] Mojena, R. (1977) "Hierarchical grouping methods and stopping rules: An evaluation," Computer Journal, 20, 359-363.

[28] McGee, V. E. and W. T. Carlton (1970), "Piecewise regression," Journal of the American Statistical Association, 65, 1109-1124.

[29] Nordhaus, W. and J. Tobin (1972), "Is growth obsolete?," Fiftieth Anniversary Colloquium V., National Bureau of Economic Research, New York: Columbia University Press.

[30] Piccolo, D., (1990), "A distance measure for classifying ARIMA models," Journal of Time Series, 11, 153-164.

[31] Ram, R. (1984), "Composite indices of physical quality of life, basic needs fulfillment and income," Journal of Development Economics, 11, 227-247.

[32] Roback, J. (1980), The Value of Local Amenities: Theory and Measurement, Ph.D. dissertation, University of Rochester.

[33] ____(1982), "Wages, rents, and the quality of life," Journal of Political Economy, 90, 1257-1278.

[34] Rosen, S. (1974), "Hedonic prices and implicit markets: product differentiation in pure competition," Journal of Political Economy, 82, 34-55.

[35] ____(1979), "Wages-based indexes of urban quality of life," in Current Issues in Urban Economics, edited by P. Mieszkowshi and M. Straszheim, Baltimore: Johns Hopkins University Press.

[36] Rutherford, R. S. G. (1955), "Income distributions: a new model," Econometrica, 23, 277-94.

[37] Sargan, J. D., (1957), "The distribution of wealth," Econometrica, 25, 568-90.

[38] Scott, D. W. (1979), "On optimal and data-based histograms," Biometrika, 66, 605-610.

[39] Sen, A. (1985), Commodities and Capabilities, Amsterdam: North-Holland.

[40] ____(1987), The Standard of Living, Cambridge: Cambridge University Press.

[41] Slesnick, D. (1991), "The standard of living in the U.S.," Review of Income and Wealth, 37, 363-386.

[42] Slottje, D. J. (1991), "Measuring the quality of life across countries," Review of Economics and Statistics, 73, 513-519.

[43] Summers, R. and A. Heston (1991), "The Penn World Table (mark 5): an expanded set of international comparisons, 1950-1988", Quarterly Journal of Economics, 106, 327-368.

Appendix: Data Sources and Notes

The data used in this study are from the following sources:

Economic Report of the President, various editions. Superintendent of Documents, US GPO.

Historical Statistics of the United States: Colonial Times to 1970, Parts 1 and 2, Bicentennial Edition, Bureau of the Census, Dept. of Commerce.

Statistical Abstract of the United States, various editions. Bureau of the Census, Dept. of Commerce.

Notes on particular values in the series used when annual values were missing for series L6, L7, L9, L10, L12 and L13.

L6 For the disposable income per capita (L6) the years 1912-1916 and 1917-1921 were given as one value so the annual values used here were interpolated.

L7 For the number of physicians used to construct the physicians per capita (L7) series the value for the year before was used for the odd numbered years from 1915 to 1941. For the years 1943-1948, 1971-1974, and 1976-1979 the mean was used. The last value was used for 1956,1957, 1961, 1981, 1984 and 1988.

L9 The percent of households with phones (L9) were estimated for 1915 to 1919 using the means from later values. From 1982 to 1995 a growth curve estimate is used based on the earlier data and a limiting value of 100%.

L10 The percent of households with radios (L10) were estimated for 1915 to 1921 with a fixed value of .01 then from 1988 on a growth curve estimate based on the earlier data and a limiting value of 100% was used.

L12 The percentage of persons aged 5 to 17 enrolled in school (L12) was linearly interpolated for the odd numbered years from 1917 to 1943. From 1988 to 1993 these values are linearly interpolated from the proximate values.

L13 Newspaper circulations (L13) were estimated for the period from 1915 to 1919 as the mean for the period.

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	
L1		66	58	57	78	39	77	88	69	90	91	80	88	71	83	L1
L2			75	77	89	63	58	95	91	97	91	90	91	82	84	L2
L3				15	55	66	68	65	66	71	74	65	74	50	67	L3
L4					59	66	73	61	61	69	78	65	79	55	74	L4
L5						82	81	76	83	76	64	61	65	32	43	L5
L6							62	92	76	93	87	82	88	73	79	L6
L7								93	91	94	81	84	89	71	71	L7
L8									91	60	88	81	91	76	86	L8
L9										93	94	82	93	80	90	L9
L10											89	84	89	78	88	L10
L11												63	63	56	73	L11
L12													68	49	63	L12
L13														60	72	L13
L14															35	L14
L15																L15

Table 1: Distances $D(i,j)$ computed between each series.

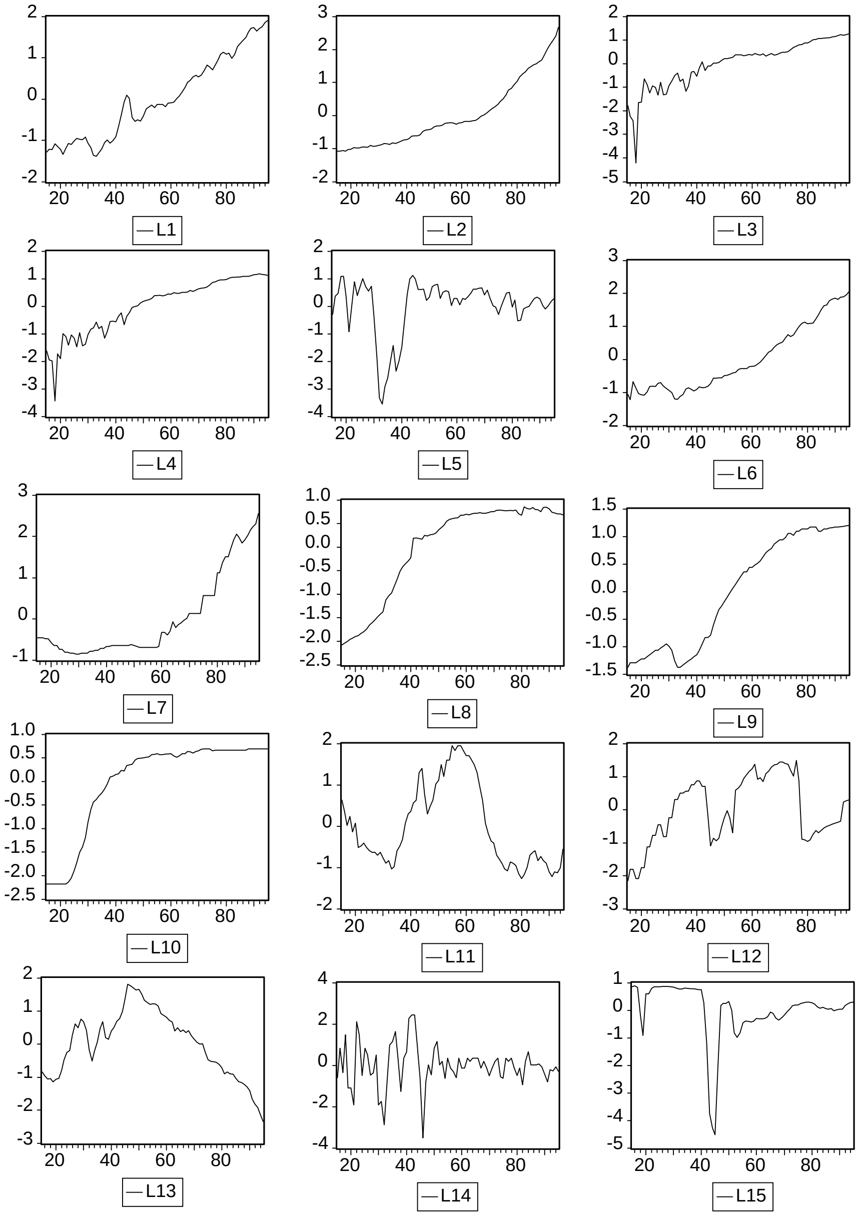

Figure 1: The series used (mean=0, and variance=1)

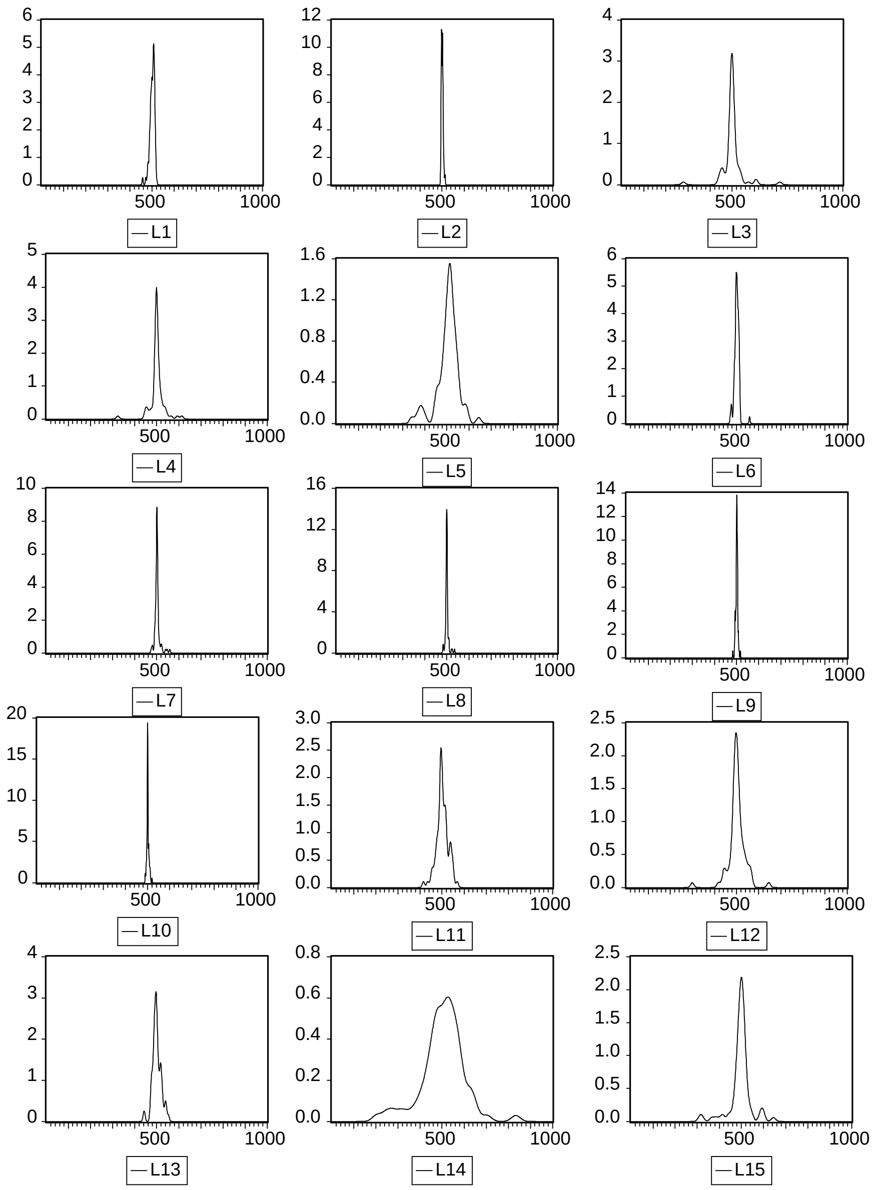

Figure 2: Estimated densities of the innovations using a Normal kernel. The innovations are centered at 500 for these plots.

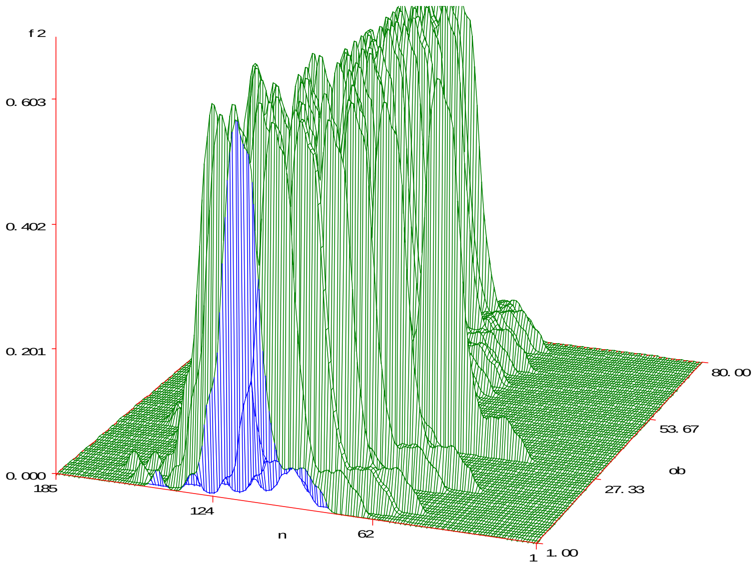

Figure 3: Estimated density for GNP Growth (series 14). The axis labelled Ob is the year number.

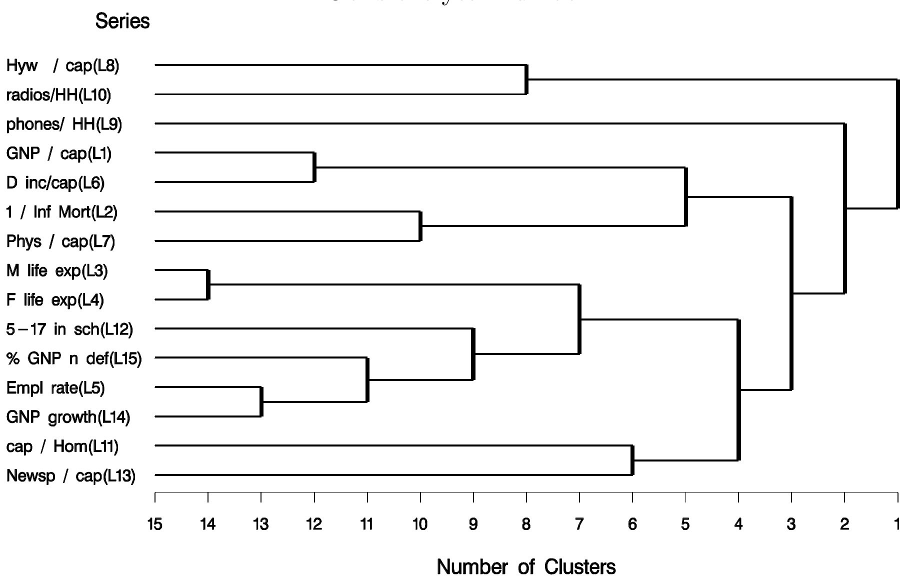

Figure 4: Dendrogram by number of clusters based on $\overline{D}(i, j)$.

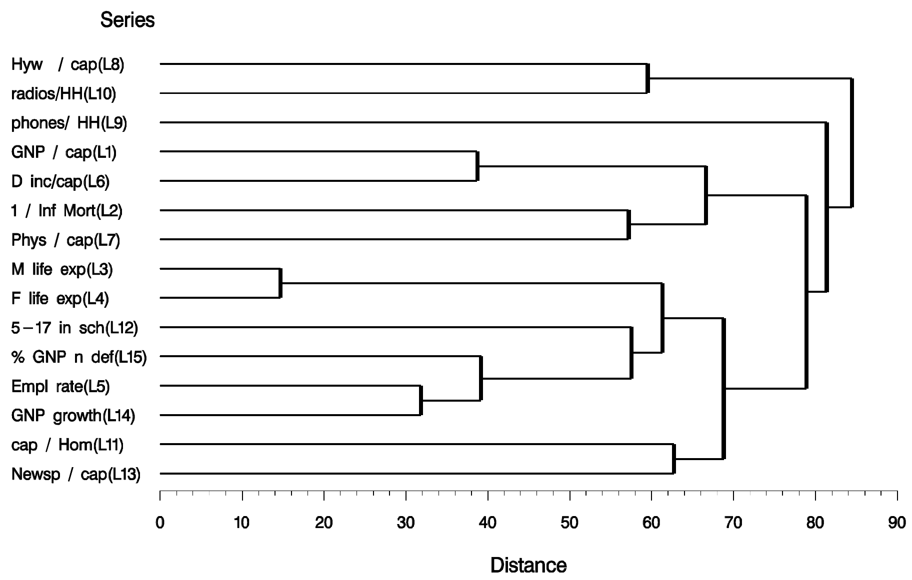

Figure 5: Dendrogram by distances based on $\overline{D}(i,j)$.

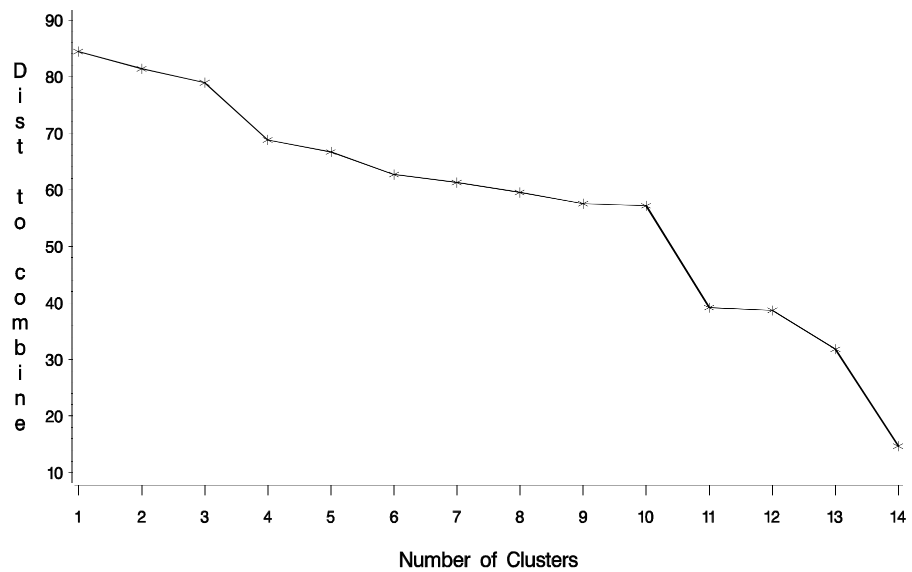

Figure 6: Distance to last member added to the cluster.

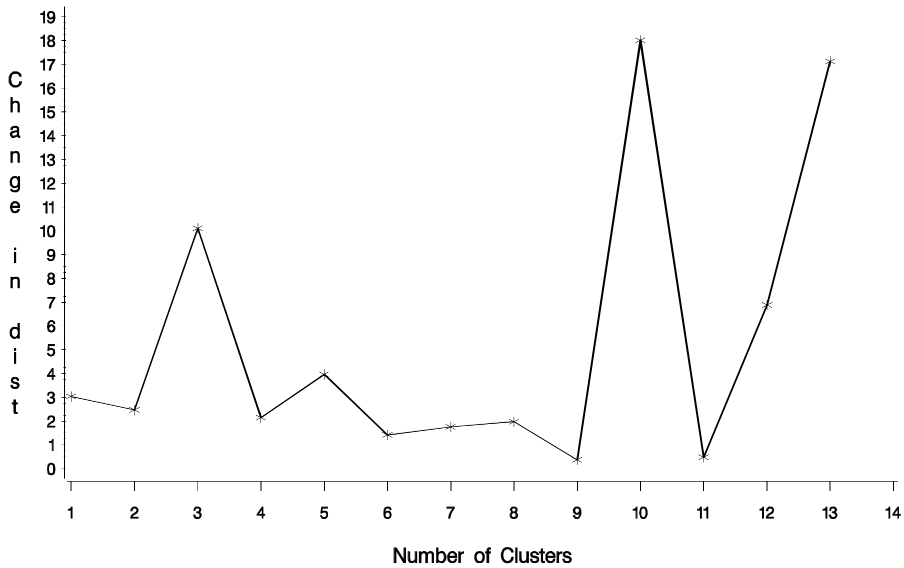

Figure 7: Change in distance to last member added to the cluster.

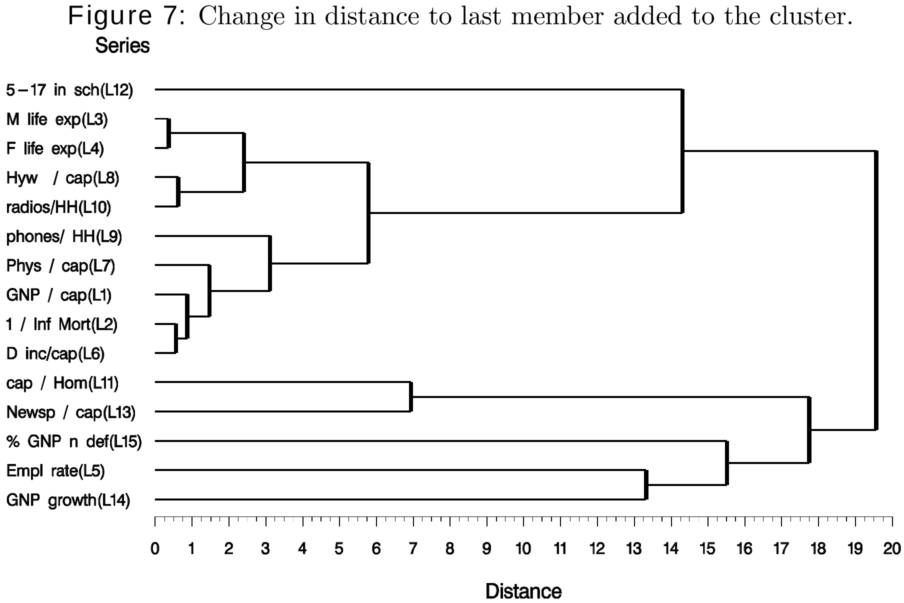

Figure 8: Dendrogram of series using $\overline{D}_{\mathsf{n}}(i,j)$ (under the assumption that each series is normally distributed with a mean of zero and a variance of one).

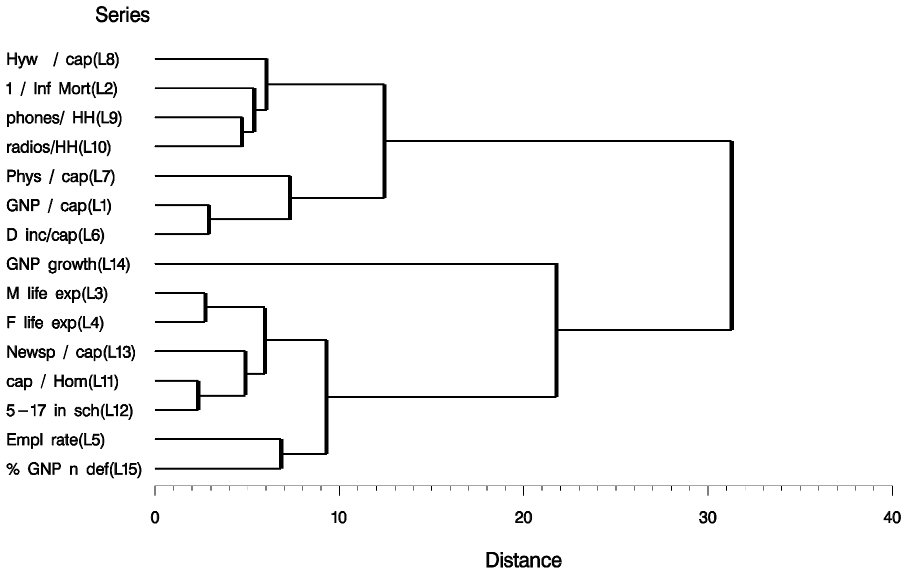

Figure 9: Dendrogram of series using $\overline{D}_{\mathsf{r}}(i,j)$ the part of the distance due to the distances between the distributions of the innovations.

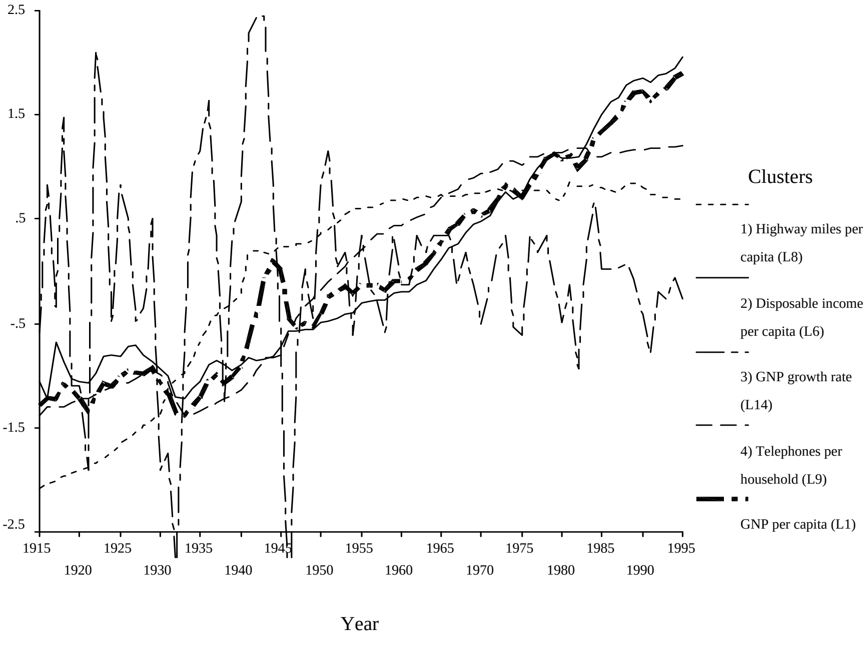

Figure 10: Plots of the series that typify each cluster.

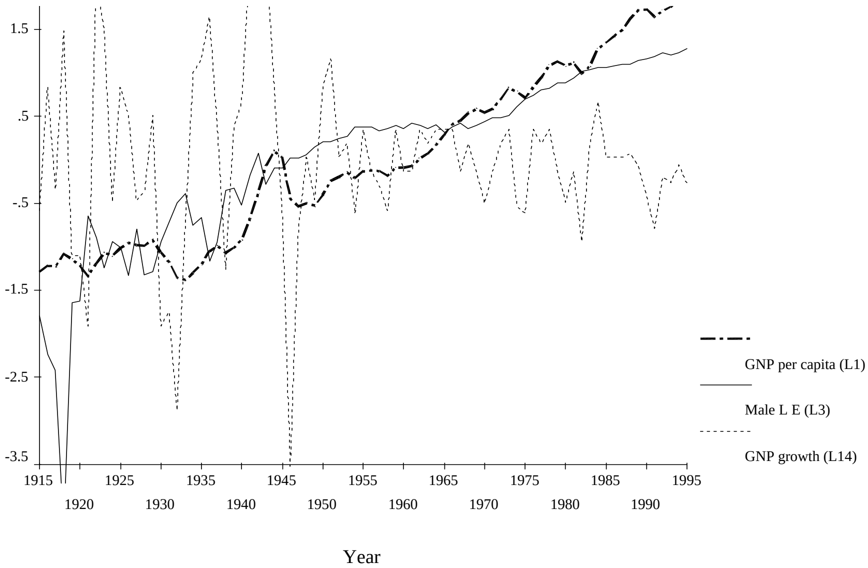

Figure 11: Alternate aggregate series.

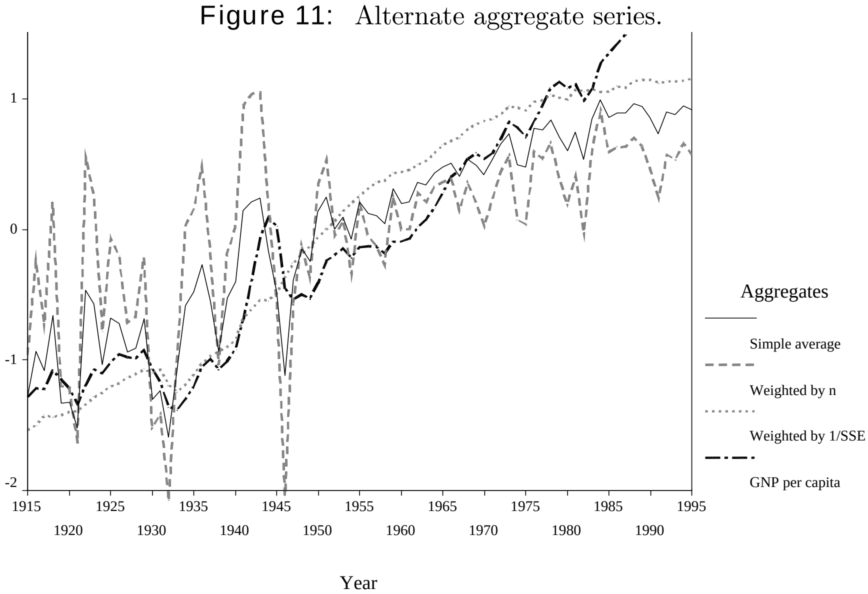

Figure 12: Comparisons of GNP per capita to three alternative weighted averages of the clusters.