# Efficiency Analysis and the Measurement of Multidimensional Well-Being

Jacques G. Silber and Joseph Deutsch (Bar-Ilan University, Israel)

Paper prepared for the 34th IARIW General Conference

Dresden, Germany, August 21-27, 2016

Session 3 (Plenary): Well-Being I

Time: Tuesday, August 23, 2016 [Morning]

# Efficiency Analysis and the Measurement of Multidimensional Well-Being

**Joseph Deutsch***

**and**

**Jacques Silber****

**Preliminary version.**

* Department of Economics, Bar-Ilan University, Israel

** Department of Economics, Bar-Ilan University, Israel, and Senior Research Fellow, LISER, Esch-sur-Alzette, Luxembourg.

**Abstract**

Using data from the Caucasus Barometer survey this paper derives measures of overall well-being for Armenia, Azerbaijan and Georgia. Correspondence analysis is first implemented to aggregate variables in each domain of well-being. Overall well-being is then derived via efficiency analysis (stochastic production frontier). This well-being is then regressed on several explanatory variables and the contribution of each of these explanatory variables to the R-square of the regression for overall well-being is obtained via the so-called Shapley decomposition. The list of determinants of overall well-being is then compared with that of material well-being.

**Key Words**: Caucasus barometer; correspondence analysis; efficiency analysis; Shapley decomposition; well-being.

## 1. Introduction: On the measurement of well-being

The *Report by the Commission on the Measurement of Economic Performance and Social Progress* (Stiglitz et al., 2009) recommended that in order to measure economic performance and social progress one should look at income and consumption rather than at production; consider income and consumption jointly with wealth; emphasize a household perspective; give more prominence to the distribution of income, consumption and wealth; and broaden income measures to non-market activities.

This report was influenced by previous work by Amartya Sen who described the complexity of the concept of well-being as: 'One could be *well-off*, without being *well* (due to health problems). One could be *well*, without being able to lead the life he or she *wanted* (due to cultural restrictions and bounds). One could have got the life he or she *wanted*, without being *happy* (due to psychological problems). One could be *happy*, without having much *freedom* (due to society's norms). One could have a good deal of *freedom*, without *achieving* much (due to lack of self-confidence or self-esteem). We can go on' (Sen, 1985:3)

There are thus many dimensions of well-being and these can certainly not be captured by some measure of income or wealth. In fact, Sen (1985) advocated taking a 'capability approach' for analyzing well-being. Such a view considers individual well-being as a combination of various 'functionings', and refers to the achievements of a person, that is, to what she manages to do or to be, and reflecting a part of the 'state' of that person. In other words, according to Sen, the mere command over commodities cannot determine the valuation of the goodness of the life that one can lead for 'the need of commodities for any specified achievement of living conditions may vary greatly with various physiological, social, cultural and other contingent features' (Sen, 1985).

As a consequence commodity command is only a means to the end of well-being and the latter should be measured by the set of capabilities with which an individual is endowed. Sen, however, did not propose a list of the relevant capabilities, not even one of 'functionings', whereas Nussbaum (2006: 76-78) has prepared such a list of capabilities.

The purpose of this paper is to derive measures of overall well-being for Armenia, Azerbaijan and Georgia for which enough data were available to take a broad enough view of well-being, to compare the findings concerning overall well-being with those based on its narrow view, one whose focus is only on material well-being, and to isolate the determinants of overall and material well-being, in particular those which seem to play a crucial role.

The paper is organized as follows. Section 2 presents the main features of the database, the Caucasus barometer survey, and defines the various domains of well-being.[1] Section 3 shows how it is possible, via a correspondence analysis, to derive a synthetic index for each dimension of well-being. Section 4 indicates how an efficiency analysis, more precisely the stochastic production frontier approach, allows one to determine well-being at the individual level. In Section 5, using regression analysis, we explain overall well-being and material well-being. On the basis of the results of such an empirical investigation we then implement the so-called Shapley decomposition in order to determine the specific contributions of the various explanatory variables of the regression to its R-square. Concluding comments are given in Section 6.

## 2. The database and the different domains of well-being

### 2.1. The database: The Caucasus Barometer Survey

Since in the South Caucasus are reliable data on social, political and economic issues are relatively rare, the Caucasus Research Resource Centers began in 2004 a coordinated data gathering effort to obtain reliable, comparable data on household composition, knowledge, social and political attitudes, and practices across the South Caucasus. This nationwide household survey named Caucasus Barometer (Data Initiative in 2004-2010) with over 6,800 respondents across the South Caucasus is conducted annually in Armenia, Azerbaijan, and Georgia and the same methodological approach and the same survey instrument are used in the three countries..

Although the emphasis of these surveys is on trends in public attitudes, these surveys have the advantage of including many questions which can be used to derive estimates of the material and overall well-being of individuals.

[1] The choice of domains and variables was dictated by the availability of data.

## 2.2. The different domains of individual well-being

The following four domains of well-being have been distinguished on the basis of the information available in the Caucasus Barometer surveys: Participation and trust in the public sphere, social relations, health and material well-being.

The list of questions on the basis of which an indicator of well-being was derived for each domain is given in Appendix A.

# 3. Correspondence Analysis (CA) and the derivation of a synthetic index for each dimension of well-being

Correspondence analysis was introduced by Benzécri (see, for example, Benzécri & Benzécri, 1980) and his French school. It is an exploratory data analytic technique which aims at analyzing simple two-way (or multi-way) tables where some measure of correspondence is assumed to exist between the rows and columns. Correspondence analysis is extremely useful for transforming a set of complex data into a simple description of almost all the implicit information provided by the data.

A useful characteristic of CA is that it allows one to obtain a graphical display of row and column points in biplots, which helps in discovering some structural relationships that may exist between the variables and the observations.[2]

Although CA may be defined as a special case of a principal components analysis (PCA) of the rows and columns of a table, one should stress that CA and PCA each have specific uses. PCA is a useful tool when the variables are continuous, whereas CA is typically applied to a case of contingency tables.

While the Chi-square test is the usual procedure adopted for analyzing the degree of association between rows and columns in a cross-tabulation, this test does not allow us to find out the important individual associations between a specific pair of rows and columns. CA on the contrary indicates how the variables are related and not simply whether there is such a link.

Assume a contingency table that has $I$ rows and $J$ columns. The plot given by CA then gives a set of ($I$+$J$) points, $I$ points corresponding to the rows and $J$ points to the columns. If two row points

[2]See Appendix B for more details on this technique.

are close, one can then conclude that their conditional distributions across the columns are similar. Given the symmetry of the role played by lines and columns in CA, we can also conclude that if two column points are close on the biplot provided by CA, this implies that their conditional distributions across the rows are similar. Like PCA, CA provides a researcher with principal components which are orthogonal. More specifically each component is a linear combination of the variables on the one hand and observations on the other. The coefficients of these variables (observations) for the first two components give us the coordinates that allow us to plot these variables (observations) in the graph previously mentioned. In this paper devoted to an analysis of multidimensional well-being, we first defined different domains of well-being and in each domain several variables were assumed to characterize well-being in the domain. CA was therefore applied separately to each domain and the first factor in each domain was then assumed to summarize the features of well-being in this domain and was then used in the second stage of the analysis.

## 4. The stochastic production frontier approach and the determination of individual wellbeing

On the basis of the 'inputs' (first factor) derived by CA in each domain of well-being an efficiency analysis was then implemented and an 'output' score (degree of well-being) attributed to each individual. More precisely the (first) factors derived separately from CA for each domain were considered as inputs in the production of a latent variable reflecting the overall degree of well-being of the individual. Such a latent variable is evidently not observed and to implement a stochastic production frontier analysis we used a technique originally proposed by Lovell et al. (1994) and later adopted by Deutsch & Silber (1999), Deutsch et al. (2003) and Ramos & Silber (2005).

Let $x = (x_1, ..., x_k)$ denote the vector of the $k$ aggregated 'inputs' (first factors) derived from CA for each of the $k$ domains. Lovell et al.'s (1994) approach (see, Appendix C for more details) amounts to estimating a translog input distance function expressed as:

$$\ln(1/x_M) = a_0 + \sum_{j \neq M}^{k} a_j \ln x_j + (1/2) \sum_{j \neq M}^{k} \sum_{h \neq M}^{k} a_{jh} \ln x_j \ln x_h + \varepsilon$$

where the sub-index *M* refers to one of the domains of well-being previously defined (see, Lovell et al., 1994, for more details on the procedure).

Note that the value of the (first) factors derived from CA for the various domains were negative for some of the individuals. In order to be able to use a translog production function we transformed these inputs as:

$$x'_{ji} = \frac{[x_j - Min\{x_{j1},...,x_{jI}\}]}{[Max\{x_{j1},...,x_{jI}\} - Min\{x_{j1},...,x_{jI}\}]}$$

where $x_{ji}$ is the value of input $j\,(j=1$ to $k)$ for individual $i\,(i=1\ to\ I)$ and $x'_{ji}$ is the value of the 'transformed input'.

The technique of corrected least squares (COLS) was then used to obtain estimates of the various coefficients (see Appendix C for more details on the COLS technique). The modified residuals which were then derived to provide input distance functions for each individual by means of the transformation:

$$d_i = e^{[(\max imum negative residual)-(residual\ \ for\ \ individual\ \ i)]}$$

This distance will by definition be greater than one so that all individual input vectors lie on or beyond the isoquant (frontier).

This input distance function will in fact measure the extent of well-being for individual $i$. More precisely the further outside the isoquant the point corresponding to the degree of well-being of individual $i$ in the various domains is, the more it must be radially contracted in order to reach the isoquant.

## 5. Explaining individual well-being

In previous sections, we identified nine broad domains of well-being. As discussed, we considered well-being as a latent factor which was quantified by means of a multi-step process that involved correspondence and efficiency analyses. The process that led to the quantification of our well-being variable used a set of variables that we judged to be proxies of the latent concept of well-being and finally delivered a well-being variable that we can now use as dependent variable in a regression.

Given the individual degree $d_i$ of well-being, we can then estimate the following OLS regression:

$$d_i = \alpha + z_i\beta + u_i$$

where $z_i$ is a vector of determinants of individual well-being and $u_i$ is the normally distributed error term.

**6. Results of the empirical investigation:**

Table 1 gives the mean values of the variables introduced in the regressions. We can see that the proportions of individuals living in the capital city (variable which does not appear in the regression and in Table 2), urban and rural areas are quite similar in Armenia and Azerbaijan: about one third of the population lives in each of these areas. But in Georgia 42% of the population still lives in rural areas, 24% in urban areas and 34% in the capital city. The average age of the respondents is 46 in Armenia, 42 in Azerbaijan and 48 in Georgia. As far as educational levels are concerned, we observe that the average numbers of years of education are quite similar in the three countries: 12 in Armenia, 11.5 in Azerbaijan and 13 in Georgia. There are however very important differences in the percentage of respondents who are married: 26% in Armenia, 66% in Azerbaijan and 29% in Georgia. The percentage of male respondents varies also from one country to the other: 33.5% in Armenia, 46% in Azerbaijan and 41% in Georgia. Differences between the three countries in the size of the household are smaller: the average size is 3.85 in Armenia, 4.31 in Azerbaijan and 3.59 in Georgia. Note also that the percentage of respondents who know Russian is 84% in Armenia, 38% in Azerbaijan and 73% in Georgia. The percentage of respondents knowing English is 17.5% in Armenia, 9% in Azerbaijan and 21.5% in Georgia. Finally the percentage of respondents having a good knowledge of computers is 29% in Armenia, 17% in Azerbaijan and 36% in Georgia.

Table 2 presents the results of the regressions for the various countries. When the dependent variable is overall well-being we observe that the impact of the area of residence varies from one country to the other. Other things constant well-being is highest in the capital city (Yerevan) and lowest in other urban areas in Armenia. In Azerbaijan ceteris paribus overall well-being is highest in the capital city of Baku and lowest in rural areas or in Baku, the capital city. Finally in Georgia well-being, ceteris paribus, seems to be highest in rural areas. In general, well-being decreases and then increases with age, reaching its minimal value at the age of 36 in Armenia, 23 in Azerbaijan and 36.5 in Georgia. Education, on the contrary, has always a significant positive effect on well-being. It is interesting to note that whereas in Armenia and Georgia male respondents have, ceteris paribus, a lower level of well-being, the opposite is true in Azerbaijan. There is no significant difference in the level of well-being between married and non-married respondents. Religion (measured via a dummy variable equal to 1 for Muslims) has also no significant impact on well-being. Note that well-being first increases, then decreases with the size of the family in Armenia and Georgia but this effect is however not really significant in Azerbaijan. In Armenia and Georgia knowing Russian has a significant positive impact on well-being (ceteris paribus) while knowing English does not have any significant effect on well-being in any of the three countries. Finally being knowledgeable in computers has, other things constant, a positive impact on well-being in Azerbaijan and Georgia.

Looking at the regressions as a whole, it appears that the R-square is generally quite low (0.07 for Armenia, 0.03 for Azerbaijan and 0.07 for Georgia) though significant (when looking at the F-value). This should not be too surprising since first we deal with individual data, second well-being is somehow defined here as a weighted average of well-being in different domains and may hence be the consequence of countervailing influences.

Let us now look at the determinants of material well-being, which is defined here as the value for the individual of the first axis of the correspondence analysis applied to the set of variables that were assumed to determine material well-being. It appears first that material well-being is higher in the capital city of Yerevan in Armenia but lowest in other urban areas while in Azerbaijan material well-being is highest in other urban areas. In Georgia finally material well-being is lowest in rural areas. Material well-being seems to first decrease, then increase with age but such an effect is significant only in Armenia where the minimal level of material well-being is reached at the age of 58-59. The only case where gender has a significant impact on material well-being is in Azerbaijan where it is higher among male respondents. Education on the contrary has always a significant positive impact on material well-being. Being married has a significant (and positive) effect on material well-being only in Azerbaijan. In all three countries material well-beings increases and then decreases with the size of the household. Religion, measured via a dummy variable equal to 1 for Muslims, has a negative and significant impact on material well-being in Azerbaijan and a positive one in Georgia. Note that in Azerbaijan 99.7% of the respondents are Muslims while 7% of the respondents are Muslims in Georgia. Knowing Russian has always a positive impact on material well-being (in all three countries) while knowing English has a positive effect in Azerbaijan and Georgia. As expected, being knowledgeable in computers has always a positive impact on material well-being.

Table 3 presents the results of the so-called Shapley decomposition of the R-square of the regressions whose dependent variables are respectively overall and material well-being. As far as the overall well-being is concerned, in Armenia the greatest impacts are those of gender (28%), education (21%), age (18%), the area of residence (11%) and knowledge (of Russian, English and computers). In Azerbaijan the greatest impact is that of the area of residence (34%), education (24%) and knowledge, as it was defined previously (21%). In Georgia the categories of variables having the greatest impact are knowledge (28%), the size of the household (25%), education (22%) and age (12.5%). Finally in the regression regrouping all three countries, the greatest impact is that of the dummy variable for the country (26%), knowledge (24%), the area of residence (14.5%) and education (10.5%).

If we now look at the Shapley contributions for the material well-being regressions we observe that in Armenia the greatest contributions are those of knowledge (28%) and education (23%) as well as those of the size of the household (20%), age (16%) and the area of residence (11%). In Azerbaijan the greatest Shapley contributions to material well-being are those of knowledge (42%) and education (36%). In Georgia the most important contributions to material well-being are those of knowledge (42.5%), education (22%) and the size of the household (18%). Finally, if we take a look at the Shapley contributions to the R-square of the regression regrouping all three countries, we see that the greatest contributions are those of knowledge (38%), education (25.5%) and the size of the household (15.5%). Note that for the material well-being regressions the R-square is quite higher than that in the overall well-being regression (0.12 for Armenia, 018 for Azerbaijan, 0.19 for Georgia and 0.16 for the combined regression).

## Table 1: Summary Statistics (Mean of the variables)

| | Armenia | Azerbaijan | Georgia | Three countries together |
|---|---|---|---|---|
| Urban | 0.3676 | 0.3682 | 0.2411 | 0.3235 |
| Rural | 0.3594 | 0.3182 | 0.4222 | 0.3687 |
| Male | 0.3347 | 0.4621 | 0.4061 | 0.3991 |
| Age | 46.1646 | 42.1826 | 47.9633 | 45.5649 |
| Square of age | 2427.9973 | 2024.1265 | 2617.1242 | 2369.5209 |
| Education | 12.0302 | 11.4432 | 13.1617 | 12.2451 |
| Islam | 0.0069 | 0.9970 | 0.0688 | 0.3343 |
| Married | 0.2586 | 0.6576 | 0.2912 | 0.3932 |
| Household size | 3.8464 | 4.3136 | 3.5872 | 3.8999 |
| Square of household size | 18.3690 | 21.7576 | 16.2057 | 18.6578 |
| Knowledge of Russian | 0.8402 | 0.3795 | 0.7281 | 0.6587 |
| Knowledge of English | 0.1742 | 0.0886 | 0.2151 | 0.1621 |
| Knowledge of computer | 0.2888 | 0.1735 | 0.3594 | 0.2779 |
| Azerbaijan | | | | 0.3088 |
| Georgia | | | | 0.3502 |

**Table 2: Regression Results by country**

**A- Armenia**

| Variable | Dependent variable: Overall well-being Coefficients | t-values | Dependent variable: material well-being coefficients | t-values |
|---|---|---|---|---|
| **Constant** | 1.3878345 | 47.20 | 0.7975123 | 38.75 |
| **Urban** | -0.0193421 | -2.52 | -0.0203804 | -3.80 |
| **Rural** | -0.0133349 | -1.65 | -0.0102882 | -1.82 |
| **Male** | -0.0362753 | -5.59 | 0.0037819 | 0.83 |
| **Age** | -0.0025229 | -2.44 | -0.0020426 | -2.82 |
| **Square of age** | 0.0000349 | 3.36 | 0.0000174 | 2.39 |
| **Education** | 0.0046908 | 3.90 | 0.0042913 | 5.10 |
| **Islam** | 0.0225894 | 0.62 | -0.0029065 | -0.11 |
| **Married** | 0.0043465 | 0.62 | -0.0031629 | -0.64 |
| **Household size** | 0.0167576 | 3.12 | 0.0152953 | 4.07 |
| **Square of household size** | -0.0012400 | -2.19 | -0.0010819 | -2.73 |
| **Knowledge of Russian** | 0.0187395 | 2.07 | 0.0187275 | 2.95 |
| **Knowledge of English** | 0.0120687 | 1.28 | 0.0061636 | 0.93 |
| **Knowledge of computers** | 0.0054163 | 0.65 | 0.0130104 | 2.24 |

**Note: Number of observations: 1458**

**Overall well-being: R-square = 0.0707; Adjusted R-square = 0.0623. F-value for the regression: 8.45**

**Material well-being: R-square=0.1205; Adjusted R-square=0.1126; F-value for the regression: 15.22**

**Table 2: Regression Results by country**

**B-Azerbaijan**

| Variable | Dependent variable: Overall well-being Coefficients | t-values | Dependent variable: material well-being coefficients | t-values |
|---|---|---|---|---|
| **Constant** | 11.8316226 | 26.51 | 0.4824276 | 4.87 |
| **Urban** | -0.1362290 | -2.57 | 0.0249705 | 2.13 |
| **Rural** | -0.2341676 | -4.24 | -0.0194569 | -1.59 |
| **Male** | 0.0851656 | 1.93 | 0.0216113 | 2.21 |
| **Age** | -0.0038454 | -0.49 | -0.0026678 | -1.52 |
| **Square of age** | 0.0000828 | 1.00 | 0.0000278 | 1.52 |
| **Education** | 0.0282320 | 2.91 | 0.0151228 | 7.03 |
| **Islam** | -0.1374738 | -0.35 | -0.1614601 | -1.88 |
| **Married** | -0.0304890 | -0.61 | 0.0220557 | 2.00 |
| **Household size** | 0.0685128 | 1.60 | 0.0307573 | 3.25 |
| **Square of household size** | -0.0053741 | -1.33 | -0.0020082 | -2.25 |
| **Knowledge of Russian** | 0.0085835 | 0.17 | 0.0299646 | 2.69 |
| **Knowledge of English** | -0.0136726 | -0.15 | 0.0731208 | 3.63 |
| **Knowledge of computers** | 0.1444519 | 2.08 | 0.0620521 | 4.02 |

**Note: Number of observations: 1320**

**Overall well-being: R-square = 0.0412; Adjusted R-square = 0.0316. F-value for the regression: 4.31**

**Material well-being: R-square=0.1847; Adjusted R-square=0.1766; F-value for the regression: 22.76**

**C-Georgia**

| Variable | Dependent variable: Overall well-being Coefficients | t-values | Dependent variable: material well-being coefficients | t-values |
|---|---|---|---|---|
| **Constant** | 1.7960513 | 28.81 | 0.6083504 | 21.63 |
| **Urban** | 0.0213191 | 1.28 | -0.0094580 | -1.26 |
| **Rural** | 0.0281276 | 1.74 | -0.0147864 | -2.03 |
| **Male** | -0.0298425 | -2.36 | 0.0095116 | 1.67 |
| **Age** | -0.0046623 | -2.18 | -0.0010999 | -1.14 |
| **Square of age** | 0.0000640 | 3.05 | 0.0000137 | 1.45 |
| **Education** | 0.0100649 | 4.14 | 0.0059003 | 5.39 |
| **Islam** | 0.0654692 | 2.62 | 0.0212132 | 1.88 |
| **Married** | -0.0203037 | -1.45 | -0.0053207 | -0.84 |
| **Household size** | 0.0633456 | 5.66 | 0.0363038 | 7.19 |
| **Square of household size** | -0.0057162 | -4.70 | -0.0031963 | -5.83 |
| **Knowledge of Russian** | 0.0397967 | 2.52 | 0.0335963 | 4.71 |
| **Knowledge of English** | 0.0240622 | 1.28 | 0.0160914 | 1.90 |
| **Knowledge of computers** | 0.0530750 | 3.15 | 0.0424203 | 5.58 |

**Note: Number of observations: 1497**

**Overall well-being: R-square = 0.0737; Adjusted R-square = 0.0656; F-value for the regression: 9.07**

**Material well-being: R-square=0.1944; Adjusted R-square=0.1874 F-value for the regression: 27.53**

**D- Three countries (Armenia, Azerbaijan and Georgia) together**

| Variable | Dependent variable: Overall well-being<br>Coefficients | t-values | Dependent variable: material well-being<br>coefficients | t-values |
|---|---|---|---|---|
| **Constant** | 1.8186486 | 64.51 | 0.6671486 | 44.39 |
| **Urban** | -0.0247937 | -3.42 | -0.0069910 | -1.81 |
| **Rural** | -0.0365099 | -4.93 | -0.0104569 | -2.65 |
| **Male** | -0.0206972 | -3.49 | 0.0097911 | 3.10 |
| **Age** | -0.0024706 | -2.50 | -0.0017401 | -3.30 |
| **Square of age** | 0.0000372 | 3.73 | 0.0000171 | 3.22 |
| **Education** | 0.0044152 | 3.81 | 0.0063834 | 10.33 |
| **Islam** | 0.0151289 | 0.84 | 0.0132073 | 1.38 |
| **Married** | 0.0039488 | 0.61 | 0.0019607 | 0.57 |
| **Household size** | 0.0265374 | 5.12 | 0.0237303 | 8.58 |
| **Square of household size** | -0.0023871 | -4.48 | -0.0017815 | -6.27 |
| **Knowledge of Russian** | 0.0256199 | 3.53 | 0.0250988 | 6.48 |
| **Knowledge of English** | 0.0006214 | 0.07 | 0.0178109 | 3.54 |
| **Knowledge of computers** | 0.0376108 | 4.62 | 0.0302372 | 6.96 |
| **Azerbaijan** | -0.0561101 | -2.87 | -0.0048072 | -0.46 |
| **Georgia** | -0.0603073 | -8.39 | -0.0229567 | -5.99 |

**Note: Number of observations: 4275**

**Overall well-being: R-square = 0.0652; Adjusted R-square = 0.0619; F-value for the regression: 19.79**

**Material well-being: R-square=0.1582; Adjusted R-square=0.1552; F-value for the regression: 53.35**

**Table 3: Shapley contributions (in percentage) to the R-square of the regression**

**A- Dependent variable: overall well-being**

| Determinants | Armenia | Azerbaijan | Georgia | Three countries together |
|---|---|---|---|---|
| **AREA OF RESIDENCE (URBAN OR NOT)** | 11.04 | 33.55 | 2.38 | 14.40 |
| **GENDER** | 28.12 | 7.85 | 4.39 | 4.68 |
| **AGE** | 18.15 | 9.47 | 12.63 | 9.52 |
| **EDUCATION** | 20.76 | 24.09 | 22.18 | 10.50 |
| **RELIGION** | 0.16 | 0.32 | 3.67 | 3.76 |
| **MARITAL STATUS** | 0.33 | 0.81 | 1.58 | 0.30 |
| **HOUSEHOLD SIZE** | 9.03 | 2.89 | 24.97 | 6.41 |
| **KNOWLEDGE (of Russian, English, Computers)** | 12.42 | 21.01 | 28.19 | 24.39 |
| **COUNTRY** | | | | 26.04 |
| **TOTAL** | 100.00 | 100.00 | 100.00 | 100.00 |

### B- Dependent variable: material well-being

| Determinants | Armenia | Azerbaijan | Georgia | Three countries together |
|---|---|---|---|---|
| **AREA OF RESIDENCE (URBAN OR NOT)** | 11.15 | 7.34 | 8.87 | 4.64 |
| **GENDER** | 0.97 | 3.53 | 1.01 | 1.89 |
| **AGE** | 16.39 | 2.74 | 6.16 | 7.92 |
| **EDUCATION** | 23.01 | 36.07 | 21.89 | 25.60 |
| **RELIGION** | 0.23 | 1.16 | 0.61 | 0.96 |
| **MARITAL STATUS** | 0.21 | 1.02 | 0.65 | 0.67 |
| **HOUSEHOLD SIZE** | 19.70 | 6.04 | 18.41 | 15.66 |
| **KNOWLEDGE (of Russian, English, Computers)** | 28.33 | 42.10 | 42.41 | 37.76 |
| **COUNTRY** | | | | 4.92 |
| **TOTAL** | 100.00 | 100.00 | 100.00 | 100.00 |

## 7. Concluding comments:

The purpose of this paper was to derive measures of material and overall well-being for three countries in West Asia: Armenia, Azerbaijan and Georgia. Our measure of overall well-being was derived in two stages: in a first stage we made a distinction between four domains of well-being: participation and trust in the public sphere, social relations, health and material wealth. For each of these four domains we had a certain number of variables that were aggregated, separately for each domain, via correspondence analysis. Then we used these four aggregated variables (first axis of the correspondence analysis implemented on each domain) to estimate via the stochastic frontier approach (corrected least squares) the level of overall individual well-being. We then regressed this individual overall well-being on a set of explanatory variables (area of residence, gender, age, education, religion, marital status, size of household and familiarity with the Russian and English languages and the computer). Finally we used the Shapley decomposition technique to determine the impacts of the explanatory variables on the R-square of these regressions. In all cases education and knowledge (familiarity with Russian, English and the computer) had important contributions. Depending on the country the area of residence, age and the size household had also significant contributions.

We then analyzed the determinants of material well-being, the latter being estimated via the first factor derived from a correspondence analysis based on various variables measuring the material well-being. This individual material well-being was then regressed on the same set of explanatory variables that were used in the regressions relative to the overall individual well-being. It then appears that in all the three countries the most important Shapley contributions to the R-square of these regressions were education and knowledge (familiarity with Russian, English and the computer). The size of the household played generally also a role as well as age and the area of residence.

It is then clear that the determinants of the overall level of individual well-being are not the same as those of individual material well-being, education and knowledge (as defined previously) being, as expected, more important determinants of material than overall well-being.

## References


Aigner, D., C. A. K. Lovell and P. Schmidt (1977) "Formulation and estimation of stochastic frontier production function models," *Journal of Econometrics*, **6**: 21- 37.

Benzécri, J.P. and F. Benzécri (1980). *Pratique de L'Analyse des Données, I, Analyse des Correspondances, Exposé Elémentaire*, Paris: Dunod Bordas.

Coelli, T. (1992). 'A computer program for Frontier Production Function Estimation: Version Frontier 2.0', *Economic Letters*, **39**: 29-32.

Coelli, T., D.S. Prasada Rao and G.E. Battese (1998). *An Introduction to Efficiency and Productivity Analysis.* Boston: Kluwe Academic Publishers.

Deutsch, J., X. Ramos and J. Silber (2003). 'Poverty and Inequality of Standard of Living and Quality of Life in Great Britain', in J. Sirgy, D. Rahtz and A.C. Samli (eds), Kluwer Dordrecht, The Netherlands: Academic Publishers, pp. 99-128.

Deutsch, J. and J. Silber (1999). 'Religion, Standard of Living and the Quality of Life', *Contemporary Jewry*, 20:119-37.

Farrell, M.J. (1957). 'The measurement of productive efficiency', *Journal of the Royal Statistical Society*, Series A, CXX: 253-90.

Green, W.H. (1992). *LIMDEP Version 6.0: User's Manual and Reference Guide*. New York: Econometric Software Inc.

Kakwani, N. and J. Silber (eds), (2008).*Quantitative Approaches to Multidimensional Poverty Measurement.* Palgrave-Macmillan, New York.

Lovell, C.A.K, S. Richardson, P. Travers and L. Wood (1994). 'Resources and functionings: A new view of inequality in Australia', in W. Eichhorn (ed.), *Models and Measurement of Welfare and Inequality.* Heidelberg: Springer-Verlag.

Meeusen, W. and J. van den Broeck (1977). 'Efficiency estimates from Cobb-Douglas production functions with composed error', *International Economic Review*, **18**: 435-44.

Nussbaum, M. C. (2006). *Frontiers of Justice. Disability, Nationality, Species Membership*. Cambridge, MA: The Belknap Press of Harvard University Press.

Ramos, X. and J. Silber (2005). 'On the Application of Efficiency Analysis to the Study of the Dimensions of Human Development', *Review of Income and Wealth*, 51(2): 285-309.

Richmond, J. (1974). 'Estimating the efficiency of production', *International Economic Review*, **15(**2): 515-21.

Sen, A. (1985). *Commodities and Capabilities.* Amsterdam and New York: North-Holland.

Stiglitz, J.,A. Sen and J.P. Fitoussi (2009). *Report of the Commission on the Measurement of Economic Performance and Social Progress* (CMEPSP).

**Appendix A: List of Domains of Well-Being and of Variables in each Domain**

**(Caucasus Barometer 2013)**

## 1) PARTICIPATION AND TRUST IN PUBLIC SPHERE

ACTPBLM: Activities during the past 6 months: Attended a public meeting

ACTMEDIA: Activities during last 6 months: Written a letter / called a newspaper, TV or radio

ACTSPET: Activities during last 6 months: Signed a petition, including online petitions

GALLTRU: Most people can be trusted?

TRUHLTH: Trust - Healthcare system

## 2) SOCIAL RELATIONS

ACTREST: Activities during last 6 months: Went to a restaurant

ACTTHEA: Activities during last 6 months: Went to a theatre / cinema

ACTCHAR: Activities during last 6 months: Made a contribution to a charity

ACTVLNT: Activities during last 6 months: Volunteered without compensation

ACTRESDT: Activities during last 6 months: Helped someone to resolve a dispute

ACTCHORE: Activities during last 6 months: Helped a neighbor/friend with household chores

ACTCLEAN: Activities during last 6 months: Cleaned/helped to clean public space

DISCPOL: How often do you discuss politics/current events with friends/close relatives?

## 3) HEALTH

HLTHRAT: How would you rate your health?

NUMCIGW: How many cigarettes do you smoke per week?

REGEXCER: Are you currently exercising on a regular basis, for at least 2 hours per week?

## 4) MATERIAL WEALTH

SAVPERS: Do you have any personal savings?

NODEBT:

OWEMON: Does anybody owe you money?

ECONSTN: HH economic situation

OWNCOTV: HH owns - Color television

OWNDIGC: HH owns - Digital photo camera

OWNWASH: HH owns - Automatic washing machine

OWNFRDG: HH owns - Refrigerator

OWNAIRC: HH owns - Air conditioner

OWNCARS: HH owns - Car

OWNLNDP: HH owns - Land line phone

OWNCELL: HH owns - Cell phone

OWNCOMP: HH owns - Personal computer

RELCOND: Perceived relative economic condition

## Appendix B: On Correspondence Analysis

Correspondence analysis (CA) was originally introduced by Benzécri & Benzécri (1980). It is strongly related to the principal components analysis (PCA) but while PCA assumes that the variables are quantitative, CA has been designed to deal with categorical variables. More precisely CA offers a multidimensional representation of the association between the row and column categories of a two-way contingency table. In short CA's goal is to find scores for both the row and column categories on a small number of dimensions (axes) that will account for the greatest proportion of the *chi²* measuring the association between the row and column categories. There is thus a clear parallelism between CA and PCA, the main difference being that PCA[3] accounts for the maximum *variance*. A clear presentation of CA is given in Asselin & Vu Tuan Anh (2008: chapter 5) and in Kakwani & Silber (2008).

Let us first recall the main features of PCA. It is a data reduction technique that consists of building a sequence of orthogonal and normalized linear combinations of the K primary indicators that will exhaust the variability of the set primary indicators. These orthogonal linear combinations are evidently latent variables and are usually called 'components'. In PCA the first component has the greatest variance and all subsequent components have decreasing variances.

Let $N$ be the size of the population, $K$ the number of indicators $I_k$ . The first component $F^1$ may be expressed for observation *i* as:

$$F_i^1 = \sum_{k=1}^{K} \omega_k^1 I_i^{*k} \ .$$

where $I^{*k}$ refers to the standardized primary indicator $I^k$. Note that $\omega_k^1$ is the (first) factor score coefficient for indicator *k*. It turns out that the scores $\omega_k^1$ are in fact the multiple regression coefficients between the component $F^1$ and the standardized primary indicators $I^{*k}$. It is important to understand that PCA has some limitations, of which the most important is probably the fact it has been developed for *quantitative* variables.

[3] For an illustration of the use of PCA, see, for example, Berrebi & Silber (1981).

It is therefore better not to use PCA when some of the variables are of a qualitative nature. (Multiple) Correspondence Analysis (MCA) is in fact a data reduction technique that should be used in the presence of categorical variables.

Let us therefore assume now that the *K* primary indicators are categorical ordinal and that the indicator $I^k$ has $J^k$ categories. Note that if some of the variables of interest are quantitative, it is always possible to transform them into a finite number of categories. To each primary indicator $I^k$ we therefore associate the set of $J^k$ binary variables that can only take the value 0 or 1.

Let us now call $X(N,J)$ the matrix corresponding to *N* observations on the *K* indicators which are now decomposed into $J^k$ variables. Note that $J=\sum_{k=1}^{K} J^k$ represents now the total number of categories. Call $N_j$ the absolute frequency of category *j*. Clearly $N_j$ is equal to the sum of column *j* of the matrix *X*. Let $N_{..}$ refer to the sum of all the $(N \;\; by\, K)$ elements of the matrix *X*. Let also $f_j$ be the relative frequency $(N_j / N_{..})$, $f^i$ be the sum of the $i^{th}$ line of matrix *X*, $f_{ij}$ be the value of cell (*i*,*j*) and $f_j^i$ be equal to the ratio $(f_{ij} / f^i)$. Finally call $\{f_j^i\}$ the set of all $f_j^i$'s for a given observation *i* (*j* = 1 to *J*). This set will be called the profile of observation *i*.

As stressed previously CA is a PCA process applied to the matrix *X*, but with the $\chi^2$ - metric on row/column profiles, instead of the usual Euclidean metric. This $\chi^2$ - metric is in fact a special case of the Mahalanobis distance developed in the 1930s. This metric defines the distance $d^2(f_j^i, f_j^{i'})$ between two profiles *i* and *i'* as:

$$d^2(f_j^i, f_j^{i'}) = \sum_{j=1}^{J} (1/f_j)(f_j^i - f_j^{i'})^2$$

Note that the only difference with the Euclidean metric lies in the term $(1/f_j)$. This term indicates that categories which have a low frequency will receive a higher weight in the computation of distance. As a consequence CA will be overweighting the smaller categories within each primary indicator. It can be shown that:

$$\omega_j^{1,k} = \frac{1}{(N_j^k / N)} Cov(F^{1*}, I_j^k)$$

where $\omega_j^{1,k}$ is the score of category $j_k$ on the first (non-normalized) factorial axis, $I_j^k$ is a binary variable taking the value 1 when the population unit belongs to the category $j_k$, $F^{1*}$ is the normalized score on the first axis and $N_j^k$ is the frequency of the category $j_k$ of indicator *k*.

It is also interesting to note that CA offers a unique duality property since it can be shown that:

$$F_1^i = \frac{\sum_{k=1}^{K}\sum_{j=1}^{J_k}\frac{w_j^{1,k}}{\lambda_1}I_{i,j}^k}{K}$$

where *K* is the number of categorical indicators, $J_k$ is the number of categories for indicator *k*, $w_j^{1,k}$ is the score of category $j_k$ on the first (non-normalized) factorial axis, $I_{i,j}^k$ is a binary variable taking the value 1 when unit *i* belongs to category $j_k$ and $F_1^i$ is the (non-normalized) score of observation *i* on the first factorial axis.[4]

Reciprocally it can be shown that:

$$\omega_j^{1,k} = \frac{\sum_{i=1}^{N}\frac{F_1^i}{\lambda_1}}{N_j^k}$$

This duality relationship implies thus that the score of a population unit on the first factor is equal to the average of the standardized factorial weights of the K categories to which it belongs. Conversely the weight of a given category is equal to the average of the standardized scores of the population units belonging to the corresponding category.

[4] Very similar results can be derived for the other factorial axes.

## Appendix C: On Frontier Efficiency Measurement:

### 1) Duality and the Concept of Input Distance Function in Production Theory:

Let $x_i = (x_{1i},...,x_{ji},...,x_{ki})$ denote the vector of levels of social exclusion in the various $k$ domains of social exclusion for individual $i$ and let $y_i$ denote the overall level of social exclusion for individual $i$. An individual's performance, as far as social exclusion is concerned, may hence be represented by the pair $(x_i, y_i)$, $i=1, \ldots I$.

A theoretical social exclusion index $SE$ can then be estimated using a Malmquist input quantity index:

$$SE(y, x^s, x^t) = D_{input}(y, x^s) / D_{input}(y, x^t)$$

where $x^s$ and $x^t$ are two different 'social exclusion inputs' vectors and $D_{input}$ is an input distance function. The idea behind the Malmquist index is to provide a reference set against which to judge the relative magnitudes of the two vectors of 'social exclusion inputs'. That reference set is the isoquant $L(y)$ and the radially farther $x_i$ is from $L(y)$, the higher the overall level of social exclusion of individual $i$ is, for $x_i$ must be shrunk more to move back onto the reference set $L(y)$.

There is, however, a difficulty because the Malmquist index depends generally on $y$. One could use an approximation of this index such as the Tornquist index, but such an index requires price vectors as well as behavioral assumptions.[5] Since we do not have prices for 'social exclusion inputs', we have to adopt an alternative strategy. The idea is to get rid of $y$ by treating all individuals equally and assume that each individual has the same overall level of social exclusion: one unit for each 'social exclusion input'. Let $e$ represent such a vector of 'social exclusion inputs' —a $k$-dimensional vector of ones. Thus, the reference set becomes $L(e)$ and bounds the vectors of 'social exclusion inputs' from below. Individuals with 'social exclusion vectors' on to $L(e)$ share in fact the lowest level of 'overall social exclusion', with an index value

[5] This is also the case of other indices that are usually used to approximate the Malmquist index such as the Paasche index, the Laspeyres index or the Fisher index.

of unity, whereas individuals with large vectors of 'social exclusion inputs' will then have higher overall levels of social exclusions, with index values above unity.

To estimate the distance function, let $\lambda = (1/x_k)$ and define a $(k-1)$ dimensional vector $z$ as $z = \{z_j\} = (x_j / x_k)$ with $j = 1,...,k-1$. Then $D_{input}(z,e) = (1/x_k) D_{input}(x,e)$ and, since $D_{input}(x,e) \geq 1,$ , we have:

$$(1/x_k) \leq D_{input}(z,e)$$

This implies that we may also write it as:

$$(1/x_k) = D_{input}(z,e)\exp(\varepsilon),\ \varepsilon \leq 0.$$

By assuming that $D_{input}(z,e)$ has a translog functional form, we have:

$$\ln(1/x_k) = \alpha_0 + \sum_{j=1}^{k-1} \alpha_j \ln z_j + (1/2) \sum_{j=1}^{k-1} \sum_{h=1}^{k-1} \alpha_{jh} \ln z_j \ln z_h + \varepsilon$$

Estimates of the coefficients $\alpha_j$ and $\alpha_{jh}$ may be obtained using COLS (corrected Ordinary Least Squares) or maximum likelihood methods (see later) while the input distance function $D_{input}(z_i, e)$ for each individual *i* is provided by the transformation

$$D_{input}(z_i, e) = \exp\{\max(\varepsilon_i) - \varepsilon_i\}.$$

This distance will, by definition, be greater than or equal to one (since its logarithm will be positive) and will hence indicate by how much an individual's 'social exclusion input vector' must be scaled back in order to reach the 'social exclusion inputs' frontier. This procedure guarantees therefore that all 'social exclusion input vectors' lie on or above the resource frontier $L(e)$. The overall level of social exclusion for individual $i$ will then be obtained by dividing $D_{input}(z_i, e)$ by the minimum observed distance value —which by definition equals 1.

**3) Estimation Procedures: The Stochastic Production Frontier Approach**

Let us take as a simple illustration, the case of a Cobb-Douglas production function. Let $\ln y_i$ be the logarithm of the output of a firm $i$ ($i = 1$ to $N$) and $x_i$ a $(k+1)$ row vector, whose first element is equal to one and the others are the logarithms of the $k$ inputs used by the firm. We may then write that:

$\ln(y_i) = x_i\beta - u_i \; i = 1 \text{ to } N.$

where $\beta$ is a $(k+1)$ column vector of parameters to be estimated and $u_i$ a non-negative random variable, representing technical inefficiency in production of firm $i$.

The ratio of the observed output of firm $i$ to its potential output will then give a measure of its technical efficiency $TE_i$ so that:

$$TE_i = y_i / \exp(x_i\beta) = \exp(x_i\beta - u_i) / \exp(x_i\beta) = \exp(-u_i)$$

One of the methods that allows the estimation of this output-oriented Farrell measure of technical efficiency $TE_i$ (see, Farrell 1957) is to use an algorithm proposed by Richmond (1974) which has become known as Corrected Ordinary Least Squares (COLS). This method starts by using Ordinary Least Squares to derive the (unbiased) estimators of the slope parameters. Then in a second stage the (negatively biased) OLS estimator of the intercept parameter $\beta_0$ is adjusted up by the value of the greatest negative residual so that the new residuals all become non-negative. Naturally the mean of the observations does not lie any more on the estimated function: the latter has become in fact an upward bound to the observations.

One of the main criticisms of the COLS method is that it ignores the possible influence of measurement errors and other sources of noise. All the deviations from the frontier have been assumed to be a consequence of technical inefficiencies. Aigner et al. (1977) and Meeusen & van den Broeck (1977) have independently suggested an alternative approach called the stochastic production frontier method in which an additional random error $v_i$ is added to the non-negative random variable $u_i$. We can therefore write:

$$\ln(y_i) = x_i\beta + v_i - u_i$$

The random error $v_i$ is supposed to take into account factors such as the weather and luck and it is assumed that the $v_i$'s are i.i.d. normal random variables with mean zero and constant variance $\sigma_v^2$. These $v_i$'s are also assumed to be independent of the $u_i$'s, the latter being taken generally to be i.i.d. exponential or half-normal random variables. For more details on this Maximum Likelihood estimation procedure, see Battese & Corra (1977) and Coelli et al. (1998) as well as programs such as FRONTIER (Coelli, 1992) or LIMDEP (Green, 1992). The same methods (COLS and Maximum Likelihood) may naturally also be applied when estimating distance functions.