Geir Øvensen

An asset index for the Syrian 2003 Unemployment Survey

A background paper on the construction of an asset index for measuring households' long-term wealth

Fafo

Geir Øvensen

An asset index for the Syrian 2003 Unemployment Survey

A background paper on the construction of an asset index for measuring households' long-term wealth

Fafo-paper 2006:01

ISSN 0804-5135

Contents

1 Summary

The purpose of this paper is to document and investigate the asset index estimated for the CBS Syria 2003 Unemployment Survey. Information about the economic situation of households improves our understanding of their labour market strategies; however, the 2003 Unemployment Survey contained no questions about households' incomes and expenditures. Hence, it was decided to develop an asset index as an indicator for the households' *long-term* wealth.

When constructing an asset index, one must first decide which items to include, secondly, one must estimate weights for the various items used in the index, and finally one must check the reliability and performance of the index.

The 2003 Unemployment Survey questionnaire naturally suggests three main groups of items for use in the index: There are 7 variables about ownership of various consumer durables, three questions about households' land ownership, and 8 ordinal or scale level variables about the quality and characteristics of dwelling and its infrastructure. The simplest type of an asset index would be one that just sums up the household assets, giving all assets equal weights, regardless of their value and type. However, this is a completely arbitrary method, and instead we *estimate* the weights of the asset index, using the statistical procedure of *principal components*. The crucial condition for using this approach is that for a list of asset variables, long-run household wealth is what causes this most common variation in the variable set. Hence, the assumption is that the *first* principal component estimated may be labelled "long-run household wealth". The mean value of the index is zero by definition, but those who prefer an index that is non-negative by definition, such as the simple additive one, may simply add the difference between 0 and the lowest household index score to the estimated index value for all households. The resulting asset index in both cases will only rank households according to long-term economic resources, and not be defined at the *scale* measurement level.

The last part of the paper deals with tests of the performance and reliability of the index. This is particularly important when there is a complete absence of income and expenditure data, as in our case. We first test that the index is internally coherent, i.e. that it produces clear separations across the poor, the middle and the rich households for each asset included in the index.

The most problematic finding is that land ownership is associated with a *low* rather than a high score on the index. This is even so for ownership of *irrigated* land, which one would assume were of higher value and quality than non-irrigated or non-cultivated land. The reason is probably that the index has a fairly strong urban bias because it gives a strong "reward" for owning items associated with modern society (e.g. computers and mobile phones) rather than traditional items, such as land. Hence, even though land ownership is probably a positive feature at the village level, living in a *rural* area is *not* associated with a high asset index score in the *national* Syrian context. The index is generally "robust" to the assets in-

cluded. The lowest 20% of households on the full index are also mostly classified among the lowest 20% on two sub-indices. Finally, the index seems to produce reasonable comparisons with indicators related to household welfare, e.g. with male and female levels of education, and with subjective judgement of the welfare distribution across regions and urban and rural locality types.

In conclusion, the index seems to perform relatively well, in spite of a fairly limited number of asset variables in the questionnaire, and in spite of the problem that only a small minority of households own these assets. However, the intention of the index is to serve as a reference for employment data, and we strongly warn against using the asset index as an independent indicator of welfare. For good measurement of household welfare and poverty one should rather use the analysis of the household income and expenditure survey, which recently has been published by the UNDP (El Laithy and Abu-Ismail 2005).

2 Introduction

When analyzing household surveys we are almost always interested in obtaining reliable information about the economic situation of households. Because income from employment is the most important income source, households' ability to generate incomes to a large extent depends on their employment situation. We would thus expect information about to be an important indicator for households' *current* incomes.

However, due to short reference periods for the employment questions, and because of frequently changing labor market conditions, we must use other indicators than employment for assessing the *long*-run economic situation of households. Moreover, the conceptual framework that we apply in the analysis of the employment and unemployment data, the so called ILO "Labor Force Framework", utilizes *time* and *activity types* rather than *wage* and *income levels* as its key classification parameters.

In many cases, questions about household incomes and consumption expenditures are included in the questionnaires, and analysis of these variables allows at least a *ranking* of households according to their disposable economic resources. In the CBS Syria 2003 Unemployment Survey there were no questions about incomes and expenditures. However, some questions about housing characteristics, public infrastructure, possession of consumer durable, and ownership of agricultural land were included in the questionnaire. We thus decided to use an asset and infrastructure index as an indicator for household "long-run wealth".

The simplest type of an asset index would be one that just sums up the household assets, giving the score of 1 if a household has the asset, and the score 0 otherwise. This simple "additive" asset index implies giving all assets equal weights, regardless of their value and type. The classical problem of how to add "apples and oranges" is thus solved by labeling everything "fruits". However, giving equal weights to all assets is a simple, but completely arbitrary method, because assets have varying prices, quality and importance. We should thus look for an alternative weighting approach.

Instead of giving equal weights to each item in the index, Filmer and Pritchett (1998) suggest to *estimate* the weights of the asset index, using the statistical procedure of *principal components*. The construction of this index is explained in more detail in Section 3. Section 4 deals with the testing of the performance and reliability of the index. This is particularly important when there is a complete absence of income and expenditure data, as in our case. We first test that the index is internally coherent, i.e. that it produces clear separations across the poor, the middle and the rich households for each asset included in the index. Second, we investigate whether the index is "robust" to the assets included, by testing out sub-indices. Third, we evaluate whether the index produces reasonable comparisons with indicators related to household welfare, e.g. with education levels and poverty or GDP per capita data.

Section 5 concludes the discussion, while an appendix contains some additional tables on the performance of the index[1].

[1] It should be noted that due to the current Syrian Statistical law, the full dataset should not be distributed to foreign institution. Hence, the current index is estimated on the basis of a 15% simple randomly selected sample from the full dataset. However, *relative to the full 100% sample*, we expect the estimators to be unbiased and consistent, although with larger variance.

3 The Construction of the Asset Index

Although we would have preferred to have questions about household income and expenditures in the Unemployment Survey questionnaire we would not necessarily use income or expenditure as our primary indicators of the households' economic resources. The reason is that income and expenditure data contains common shortcomings when used as indicators for households' long-term wealth:

Income data are usually prone to severe measurement errors, in particular underreporting of income levels. Such underreporting may be deliberate, from households that fear taxation, or hope for public support. In some households members may sometimes also conceal their true incomes for each other. However, underreporting of income may also be caused by memory lapse, in particular in households where there are many sources of income.

Another problem is that income data tend to fluctuate, both in a random and in a systematic manner. Systematic fluctuation in income is typical for those economic activities where work efforts are conducted during a longer period, but where sales, and hence incomes, take place during a relatively short period. The problem is largest in climate-dependent productive sectors such as agriculture, and in seasonal service activities, such as tourism.

Households' *expenditures* are also commonly underreported. As for incomes the underreporting may be deliberate. For example, some household members may not want to reveal spending on controversial items for other members. However, for expenditures the biggest problem is probably memory lapse. It is usually easier to keep track of a few substantial income sources than expenditures on numerous large and small items. Hence, it is usually required that detailed information is collected in order for expenditure data to be accurate. Relative to income, the advantage with (true) household expenditure is that it usually fluctuates less than household incomes, due to consumption smoothing.

From the discussion above we realize that it is not evident that an asset index must be a secondary choice as an indicator of households' economic resources when household income and expenditure data are available. This is particularly the case when we need a measure for *long-term* household wealth in societies with stable economic conditions. Because many items may be observed, there are probably less measurement errors associated with the items included in an asset index although some items, such as bank savings and gold, are usually not observable.

Some authors have maintained that the main aim of an asset index is to create a proxy for current consumption expenditures (Montgomery, Burke Paredes and Zaidi, 1997). To the contrary, Filmer and Pritchett (1998) maintain that *both* an asset index and current consumption expenditures are proxies for a household's true, but unobserved, long run wealth or "economic status". Possible discrepancies between the two indicators with respect to ranking of households cannot only to be ascribed to the "mistakes" of the asset index. In principle, the "current consumption expenditure" only serves as a perfect measure for long-run wealth under the unrealistic assumption of perfect foresight and perfect capital markets.

However, even though we would sometimes make an asset index our first choice as indicator for household economic resources, we would always prefer to test its performance against household income and expenditure data whenever possible.

When the decision has been made to construct an asset index several issues emerge: First, which items should be included? Second, how should one weight the various items up against each other? Third, how should one control for price and quality differences among items in the index? Fourth, what is the best way to check the reliability and performance of the index?

With respect to the question of which items to include, the limited number of candidates in the 2003 Unemployment Survey questionnaire naturally suggests that three main groups are used: First, there are 7 variables about ownership of various consumer durables. Second, there are some nominal-level questions about the quality and characteristics of dwelling and its infrastructure than can be transformed into 8 ordinal or scale level variables. Third, there are 3 questions about households' land ownership, respectively the size of their irrigated, non-irrigated and non-cultivated land.

The second main challenge in the asset index construction is how one should weight the various items up against each other? Filmer and Pritchett (1998) outline three common solutions to the problem in the literature: The *first* solution would be the simple additive one, where all assets are given equal weights. As mentioned above, this method is quickly written off. Although giving equal weights to all assets is a simple approach, it is a completely arbitrary method in the common situation when assets have varying prices, quality and importance. (However, the method is at least "open" about this shortcoming). By definition, an additive index will also give positive weights to all items in the index.

The *second* approach is to enter all asset variables individually in a multivariate regression equation. The problem with this method is that many assets may have both a *direct* (i.e. wealth) and an *indirect* effect on the phenomenon of interest. For example they present a case where one wants to study the effect of household wealth on children's education. In this case, having electricity is an indicator of household wealth, but also facilitates reading at night. Similarly, having piped water is an indicator of household wealth, but also reduces the need for water collection, and hence reduces the opportunity cost of schooling for children (in particular girls, whose task is often to collect water). In an unemployment survey one may similarly assume that some items affects employment both through their wealth dimension, and because they facilitates certain types of employment taking place in the home. The main point here is that we cannot uniquely identify the wealth effect from *other effects* on the phenomenon we are investigating. The implication is that the regression coefficients in addition to showing the effects of a change in wealth on our phenomenon of interest, e.g. labour force participation, also represent "something more" that is not related to wealth.

The *third*, and perhaps most desirable solution is to estimate the index weights from the prices of the various assets. However, such price data are rarely available (and *not* in our case), and it is also difficult to estimate prices of old, and sometimes partially non-functioning assets. For some items, such as agricultural land, there may further be weakly developed commercial markets from which to collect prices.

Alternatively, Filmer and Pritchett suggest estimating the weights of the asset index, using the statistical procedure of *principal components*. Principal component analysis (PCA) involves a mathematical procedure that transforms a number of (possibly) correlated varia-

bles into a (smaller) number of uncorrelated variables called *principal components*. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible.

The crucial assumption is that for a list of asset variables, long-run household wealth is what causes this most common variation in the variable set. Hence, the assumption is that the *first* principal component estimated may be labelled "long-run household wealth". One may easily think of situations where this assumption is not valid. Assume for example that the asset index only comprises items requiring grid electricity or running water. In this case we could as well risk that the first principal component turns out to be "connection to the water network", or "connection to the electricity grid", rather than household wealth.

The "scoring factors" of the first principal component among the asset variables are listed in the left column in Table 1. These factors are "normalized" by their standard deviation, and then used as asset weights in the index. Since most assets contribute positively to household wealth, it is reasonable that they have positive weights. However, as we can see from the left column in Table 1, some of the "assets", such as e.g. "living in dar[2]", "living in clay house", and land ownership, etc. have a negative effect on a household's total score on the index. The reason for the latter result is that the items included in the index give it an urban bias. Although land ownership makes a household wealthy in their local context, it usually implies that the household is a *rural* household, and hence has a low score on all other items included in the *national* index. (We will come back to this discussion below).

Table 1 Scoring factors and summary statistics for variables entering the computation of the first principal component (long-term wealth)

Asset variable	Scoring factors (F1)	Mean (a1)	Std.dev (s1)	(F1)/(s1)
Live in flat	0.717	0.350	0.477	1.503
Live in dar	-0.561	0.569	0.495	-1.133
Live in clay house	-0.251	0.072	0.258	-0.972
Live in other dwelling type	0.047	0.005	0.477	0.098
No. of rooms used for sleeping	0.419	1.297	0.793	0.528
Connected to water network	0.392	0.874	0.332	1.183
Connected to sewerage network	0.504	0.744	0.436	1.154
No. of (ordinary) telephone lines	0.593	0.611	0.506	1.172
No. of mobile phones	0.558	0.089	0.340	1.643
No. of private cars	0.448	0.116	0.343	1.306
No. of rooms with air-conditioner	0.510	0.084	0.348	1.465
No. of washing machines	0.705	0.272	0.469	1.503
No. of freezers	0.520	0.144	0.371	1.404
No. of computers	0.568	0.068	0.252	2.255
Size of irrigated land (dunums)	-0.129	3.483	16.438	-0.008
Size of non-irrigated land (dunums)	-0.154	7.914	33.611	-0.005
Size of non-cultivated land (dunums)	-0.029	0.293	2.970	-0.010
Own commercial, industrial or service establishment	0.400	0.163	0.370	1.082
Asset index		0.003	3.890	

[2] Traditional Syrian house.

With respect to "living in dar", and "living in clay house", the negative sign implies that these dwelling types are estimated to be strongly negative, relative to living in flats[3].

The formula for the asset index suggested by Filmer and Pritchett is that household j's value on the index, *Aj*, is calculated as follows:

$$A_j = \frac{f_1 * (a_{j1} - a_1)}{s_1} + \ldots\ldots\ldots + \frac{f_n * (a_{jn} - a_n)}{s_n},$$

where f_1 is the "scoring factor" for the first (of a total of *n*) assets in the index, determined by the principal component analysis, a_{j1} is the j'th household's variable value for the first asset, and a_1 and s_1 are, respectively, the mean and the standard deviation of the first asset variable over all households.

For binary variables, with values 0-1, the interpretation is acquiring an asset (or a feature such as connected to sewerage network) changes the index by f_i/s_i (i=1,…,n). For example, Table 1 shows that living in a "flat" increases the index by 1.50 units, while living in a "dar" lowers the index by 1.13 units.

The mean value of the index over all households should be zero by definition, (here it was 0.003). This implies that many households will take negative values on the index. This is unproblematic, but readers used to simple additive indices, which by definition are non-negative, may find it strange to assign households negative index values. One may solve this "problem" by normalizing the minimum score of the index to 0. This can do this by adding the difference between 0 and the lowest household index score to the index value for all households. The main issue here is that the asset index is defined at the *ordinal*, and not at the *scale* measurement level.

[3] The original (nominal) question in the questionnaire about type of dwelling was recoded into four (ordinal) binary variables, for living in "flat", "dar", "clay house" and "other type".

4 The reliability of the asset index

In this section we will test the performance and the reliability of the index[4]. Such checks are particularly important when there is a complete absence of income and expenditure data in the survey, as is the situation in our case. We will firstly test that the index is internally coherent, i.e. that it produces clear separations across the poor, the middle and the rich households for each asset included in the index. Secondly, we will investigate whether the index is "robust" to the assets included, by testing out various sub-indices. Thirdly, we will evaluate whether the index produces reasonable comparisons with indicators we know are related to household welfare, i.e. with male and female education levels, and across provinces (mohafazat) and type of locality.

4.1 Internal coherence of the asset index

Let us first turn to the question about internal coherence of the index. Table 2 contains four columns, one for each of four groups that are constituted by their score on the overall asset index. The first column from the left shows the mean value for each asset among those 20% (of households) who have the *lowest* score on the asset index. The second column contains the mean asset value for the next 30% of households, and so on. We would then expect the mean asset value to increase when we move from the very left, towards the very right column. The bottom line of Table 2 shows the mean values for the full asset index.

From Table 1 we can see the gap in the mean value of the index between the two highest groups is relatively large, at more than 5 units. To move from the second highest to the highest, a household would have to acquire a computer, an air-conditioner, and a mobile telephone, which would raise its score on the asset index by roughly 5 points.

First in Table 1 come seven variables related to housing and infrastructure, of which water and sewerage network connection to some extent are provided on a community basis. This helps to explain the somewhat odd phenomenon, that the mean value of "living in a dar" decreases as the total index score increases. The "dars" are to a large extent old houses situated in rural areas, where community infrastructure such as piped water, sewage and electricity grids are sparse, while "apartments" are typically newer dwellings situated in more developed areas. In contrast, the mean number of rooms used for sleeping in the households' dwelling increases nicely with the index scores[5].

[4] Most of these consistency checks are due to Filmer and Pritchett (1998).

[5] Five percent of the households had illegal missing value for the number of rooms used for sleeping (0 was a legal value). This problem has been carried onto the full asset index, but we still chose to include the variable because of the general scarcity of candidate variables for the index.

Next comes ownership of seven consumer durables, mostly requiring electricity connection. These items vary across the groups formed by the households' index scores, as one would expect. Many of the assets, like air-conditioner or personal computer are not common in Syrian households, even among households whose score fall in the upper 20% group on the asset index. However, the mean value for all assets increases systematically as we move from the left to the right column[6]. This is particularly the case for items that we know are commonly owned by the rich, but not by the poor, such as ordinary telephone, washing machine and car.

The most problematic part of the index is the odd phenomenon that land ownership is associated with a *low* rather than a high score on the index. This is even so for ownership of *irrigated* land, which one would assume were of higher value and quality than non-irrigated or non-cultivated land. As mentioned above, the reason is probably that the index has a fairly strong urban bias. Many of the consumer durables included are not common in the countryside. Even though land ownership is probably a positive feature at the village level, living in a *rural* area is *not* associated with a high asset index score in the *national* Syrian context. However, the index seems to contain a strong "reward" for owning items associated

Table 2 Group means for the asset index with weights derived by principal components analysis

Asset variable	Lowest 20%	Lower middle 30%	Upper middle 30%	Upper 20%
Live in flat	0.000	0.009	0.537	0.912
Live in dar	0.757	0.929	0.424	0.079
Live in clay house	0.233	0.052	0.029	0.002
Live in other dwelling type	0.000	0.003	0.007	0.007
No. of rooms used for sleeping	0.867	1.166	1.343	1.852
Connected to water network	0.492	0.952	0.978	0.989
Connected to sewerage network	0.184	0.775	0.933	0.977
No. of (ordinary) telephone lines	0.062	0.530	0.793	1.003
No. of mobile phones	0.000	0.007	0.038	0.378
No. of private cars	0.023	0.032	0.112	0.337
No. of rooms with air-conditioner	0.001	0.004	0.045	0.343
No. of washing machines	0.001	0.014	0.270	0.920
No. of freezers	0.015	0.020	0.118	0.495
No. of computers	0.000	0.000	0.016	0.315
Size of irrigated land (dunums)	8.825	3.179	1.595	1.403
Size of non-irrigated land (dunums)	20.143	6.646	4.227	3.011
Size of non-cultivated land (dunums)	0.487	0.293	0.211	0.222
Own commercial, industrial or service establishment	0.020	0.058	0.206	0.396
Asset index	-4.250	-2.120	0.845	6.083

[6] Since the asset variables take the value 1 if the household owns the asset, and 0 otherwise, one may simply multiply the column figures by 100 in order to obtain the *percentage* of households in each group that owns the asset. For ordinary telephones the interpretation is that the average number of telephones in the highest group is 1.03.

with modern society (e.g. computers and mobile phones) rather than traditional items, such as land. Below we will investigate this phenomenon closer by estimating one principal component asset index for an *urban* sample only, and one similar index for the *rural* sample.

As an alternative to the asset index with weights derived from principal component analysis we also constructed a simple additive index[7]. Although this index, by definition, assures that land ownership has positive weights, we already remarked above that giving equal weights to all assets is a simple, but completely arbitrary method for assets with varying prices, quality and importance. The mean value of the index across the items included is shown in Table 3. We note that even for this index the share of households owning irrigated land hardly increases from the lower 20% to the upper 20%.

Table 3 Group means for the simple additive asset index

Asset variable	Lowest 20%	Lower middle 30%	Upper middle 30%	Upper 20%	Total
Live in flat	0.042	0.234	0.485	0.703	0.350
No. of rooms used for sleeping	0.750	1.170	1.438	1.918	1.297
Connected to water network	0.698	0.866	0.956	0.984	0.874
Connected to sewerage network	0.468	0.722	0.871	0.931	0.744
No. of (ordinary) telephone lines	0.144	0.550	0.811	0.979	0.611
No. of mobile phones	0.003	0.009	0.054	0.343	0.089
No. of private cars	0.011	0.036	0.076	0.394	0.116
No. of rooms with air-conditioner	0.001	0.013	0.043	0.327	0.084
No. of washing machines	0.002	0.071	0.401	0.719	0.272
No. of freezers	0.011	0.051	0.145	0.428	0.144
No. of computers	0.001	0.001	0.033	0.281	0.068
Household owns irrigated land?	0.130	0.161	0.112	0.138	0.137
Household owns non-irrigated land?	0.080	0.497	0.812	0.933	0.570
Has non-cultivated land?	0.000	0.034	0.056	0.315	0.091
Own commercial, industrial or service establishment	0.021	0.082	0.178	0.423	0.163
Additive asset index	2.362	4.497	6.470	9.817	5.613

The bi-variate correlation coefficient between the asset index with weights derived from principal component analysis, and the simple additive asset index was as high as 0.837. One may argue that this confirms the performance of the first index type. However, one may also argue that one could as well replace the former with the latter, which is simpler to construct, because the results do not differ much between the indices. Still, we maintain that one should prefer the asset index with weights derived from principal component analysis to the simple additive asset index because the weighing of items in the latter index lack any theoretical or empirical qualified justification.

[7] For the simple additive index we selected only "live in flat" from the nominal variable about type of dwelling. Moreover, we replaced the *size* of land of various types owned, by binary yes-no variables for the three land types.

4.2 Robustness of the asset index

In order to check the *robustness* of the index we constructed four sub-indices. First, we made two sub-indices for sub-groups of the assets included in the full index, i.e. one sub-index for housing, infrastructure and land ownership, and one sub-index for the seven consumer durables. Then we made one index for the full range of items in the national index, but estimated from *urban* households only, and a similar index estimated from *rural* households only.

First, Table 4 shows the bi-variate correlation matrix between respectively the full index with weights estimated by principal component analysis, the four sub-indices, and the simple additive index[8]. The correlations between the full principal component index and its four sub-indices are quite high, and higher than for the simple additive index. To some extent this follows by definition, since the sample of households consisted of a roughly equal number of urban and rural households, and since the consumer durable, and the housing/ infrastructure sub-indices contributed roughly equally to the total index.

Table 4 The bi-variate correlation matrix*

	Full PC index	Full PC index urban sample	Full PC index rural sample	PC Sub-index for 7 consumer durables	PC Sub-index for 11 housing, land and infrastructure features	Simple additive index, full sample
Full PC index	1.000					
Full PC index urban sample	0.997	1.000	*			
Full PC index rural sample	0.977	*	1.000			
PC Sub-index for 5 consumer durables	0.882	0.905	0.860	1.000		
PC Sub-index for 11 housing, land and infrastructure features	0.815	0.721	0.751	0.459	1.000	
Simple additive index, full sample	0.837	0.911	0.694	0.782	0.600	1.000

* Households that are in the urban sample are not in the rural sample and vice versa. Hence this coefficient cannot be computed.

[8] By definition, the correlation coefficient varies between –1 and 1. A value of 1 implies "perfect" positive correlation, while a coefficient of –1 implies "perfect" *negative* correlation (high values on one index goes together with low values on the other).

The lowest correlation coefficient was found between the consumer durable, and the housing/ infrastructure sub-indices. One reason for this is that the consumer durable index contained big "jumps" in the households' scores. This phenomenon occurred because only a small minority of the households owned most of the consumer durables in the sub-index, except ordinary telephone[9]. Another reason is that ownership of these items was typically low among households who owned land. For example, the correlation coefficients between the consumer durables sub-index and all types of land ownership were negative.

For each of these indices, as well as for the simple additive index we have also investigated how those households *who belonged to the lowest 20%* on the full-item, full-sample asset index were (re-) classified by each of the four *sub*-indices. Moreover, we have estimated the simple bi-variate correlation coefficients between the altogether 6 sub-indices[10].

Table 5shows those households *who belonged to the lowest 20%* on the full-item, full-sample asset index were re-classified by each of the four *sub*-indices, and the simple additive index. As a first point we note the encouraging fact that hardly any household was re-classified into the upper two groups on any of the four the sub-indices. For the simple additive index approximately 5% of the households who belonged to the lowest 20% on the full-item, full-sample principal component asset index were re-classified into the highest 50% on the simple additive index. We expect most of these households to be landowning, rural households.

As expected the sub-index for the urban sample behaves "perfectly" as no households at all were re-classified. However, this also shows the urban bias of the full index. For the rural sub-index the results are rather different as almost half the households were reclassified from the lowest to the second lowest group. However, even for this index hardly any households were reclassified into the two highest groups.

The two sub-indices for respectively the consumer durables and the housing, land and infrastructure performed better than the rural sample sub-index, but not as "good" as the urban sample sub-index. Hence, it seems that the problem with the full index is mainly related to its classification of *rural* households, rather than the internal consistency with

Table 5: Classification of bottom 20% group for full index

	Full PC index urban sample	Full PC index rural sample	PC Sub-index for 7 consumer durables	PC Sub-index for 11 housing, land and infrastructure features	Simple additive index, full sample
Lowest 20%	100.0	52.8	89.9	82.7	61.5
Lower middle 30%	0.0	46.0	9.9	17.3	33.6
Upper middle 30%	0.0	1.2	0.0	0.0	4.8
Upper 20%	0.0	0.0	0.2	0.0	0.1
Total	100.0	100.0	100.0	100.0	100.0

[9] The index should ideally have contained both items owned by nearly all households, items owned by half the households and items owned by only few households. However, almost all questions about consumer durable in the questionnaire referred to the latter type of rarely found items, such as air-conditioner and computer.

[10] By definition, the correlation coefficient between the urban and rural sub-indices could not be calculated because their two samples were mutually exclusive.

respect to the items included in the full index. The problem occurs because possession of traditional asset items, such as agricultural land, is negatively correlated with possession of modern asset items, such as computers, air –conditioners and mobile telephones.

4.3 The asset index and other welfare related indicators

The third of the reliability checks recommended by Filmer et al. is to see whether the asset index produces *reasonable comparisons with related indicators*, e.g. with poverty or GDP per capita. Because the data from the CBS 2004 Income and Expenditure Survey has not yet been analyzed, we do not have indicators for household poverty at the mohafaza level. Instead, we suggest to use the individual education levels of members of households grouped by their asset index score as an indicator. The basic idea is that there is usually a close relation between a household's level of economic resources and the education levels of its members. However, also here we expect that this argument is more applicable for urban than for rural areas. Because of the large historical difference in male and female education levels, we have also chosen to split the sample by gender.

First, Table 6 shows how male education levels increase systematically as we move from the lowest household asset index group (left bar) towards the highest group (right bar). While less than 20% of the males in the lowest group has preparatory education or higher, this figure increases to more than 60% in the highest asset index group. The corresponding results

Table 6 Male education levels by four asset index groups

	Lowest 20%	Lower middle 30%	Upper middle 30%	Upper 20%	Total
Illiterate	21	13	8	3	11
Can read and write	14	13	12	6	12
Elementary	46	43	41	28	40
Preparatory	11	17	19	23	18
Secondary	4	8	10	17	10
Inter-mediate institute	3	3	5	6	4
University and above	1	3	5	16	6
Total	100	100	100	100	100

Table 7 Female education levels by ten asset index groups

	Asset index deciles (1= lowest 10 percent, 10 = highest 10 percent)										
	1	2	3	4	5	6	7	8	9	10	Total
Illiterate	53	44	37	35	27	25	22	16	12	6	28
Can read and write	8	10	10	9	12	10	9	10	8	7	9
Elementary	33	33	37	34	36	36	38	31	27	19	32
Preparatory	4	8	10	14	14	17	16	21	20	24	15
Secondary	1	3	4	5	8	8	9	12	17	21	9
Intermediate institute	0	1	2	2	3	3	4	7	10	10	4
University and above	0		0	1	1	1	2	3	6	12	3
Total	100	100	100	100	100	100	100	100	100	100	100

for women are shown in Table 7, except that here, we have chosen to divide the households into *ten* rather than four groups according to their asset index score. For women the positive relation between asset possession and education is similar, but even clearer than for men. Less than 5% of the women in the lowest group have preparatory education or higher, this figure increases to almost 60% in the highest asset index group.

In the absence of reliable and updated data about the economic situation at the *local* level in Syria, our final type of check of the performance of the asset index is to check whether the scores on the asset index by mohafaza and type of locality correspond with insiders' *subjective* perceptions of differences in welfare levels between various areas. We would for example expect the largest urban areas as Damascus and Aleppo to have higher scores than predominantly rural areas in the north and east of the country.

First, Table 8 shows how almost all households who belong to the lowest 20% on the full-item, full-sample asset index were situated in rural areas, while to the contrary, almost all households belonging to the *highest* 20% group lived in urban areas. As we will se below, there is also a large difference within the group of urban households, i.e. between the large cities, as for example mohafaza capitals, and the smaller cities and towns.

Keeping in mind the large difference between the score of urban and rural areas on the asset index, we also note from Table 8 that there are fairly large variations in the distribution of the six Syrian regions' shares of the four asset index groups. While 50% of the households in the lowest group are situated in the Southern and Eastern regions, only 23% of the households in the highest group live in these regions. For the Damascus region these figures are respectively 2% and 26%. Table 9 (overleaf) shows the same distribution across Syria's 14 mohafazat.

The distributions of the asset index both by region and mohafaza are, however, deceiving because there are large difference within almost all regions and mohafazat. To illustrate this point we have broken down the mohafazat in two stages. First, we split up each mohafaza into urban and rural households. Second, we sub-divide the urban households into households living in the mohafaza capital, and households living in other cities and towns, creating three groups within each mohafaza (Table 10, overleaf).

Table 8 Regions and urban-rural localities by four asset index groups

	Lowest 20%	Lower middle 30%	Upper middle 30%	Upper 20%	Total
Damascus	2	7	15	26	12
Southern	18	27	20	16	21
Middle	16	16	18	13	16
Costal	6	8	15	11	10
Northern	26	26	23	27	25
Eastern	32	16	10	7	16
Total	100	100	100	100	100
Urban	15	41	74	90	56
Rural	85	59	26	10	44
Total	100	100	100	100	100

We note from Table 10 that some mohafazat (in particular Aleppo and Hasakeh) shows very different results for urban and rural areas. For both mohafazat their population shares are higher among the upper 20% than the lower 20% in the *urban* areas, while the opposite is the case in the rural areas. This difference is particularly marked in the mohafaza capitals, while other cities and towns come out in the middle range on the asset index. It is our impression that the distribution of the asset index scores across mohafazat split up by mohafaza capital, other cities and towns, and rural areas do not deviate much from the common perceptions of the distribution of household welfare in Syria. That being said we strongly warn against using the asset index as an independent indicator of welfare. The intention of the index is to serve as a reference for employment data. For measurement of household welfare and poverty, and comparison of these indicators across socio-economic groups and regions, one should wait for the analysis of the household income and expenditure survey.

Table 9 Mean value of the asset index by mohafaza

	Lowest 20%	Lower middle 30%	Upper middle 30%	Upper 20%	Total
Damascus city	2	7	15	26	12
Damascus other	8	16	16	14	14
Homs	8	8	12	10	10
Hama	9	9	6	3	7
Tartos	3	4	5	4	4
Latakia	3	4	9	7	6
Edleb	5	8	5	2	5
Aleppo	21	18	17	25	20
Rakka	6	6	2	0	4
Der Elzor	10	5	4	3	5
Hasakeh	17	5	4	4	7
Sweda	3	3	1	1	2
Dara	6	7	3	1	4
Qunitra	0	0	0		0
Total	100	100	100	100	100

Table 10 Mean value of the asset index by mohafaza, mohafaza capitals, other cities and towns, and rural localities

	Lowest 20%	Lower middle 30%	Upper middle 30%	Upper 20%	Total
Damascus city	2	7	15	26	12
Damascus gvt urban	3	6	9	9	7
Damascus gvt rural	4	10	6	5	7
Homs center	1	1	7	9	5
Homs other urban	0	2	1	0	1
Homs rural	7	5	4	1	4
Hama center	0	2	2	2	2
Hama other urban	0	0	1	0	1
Hama rural	8	6	2	0	4
Tartos center	0	0	1	1	1
Tartos other urban	0	0	1	1	1
Tartos rural	3	3	3	1	3
Latakia center	0	1	6	5	3
Latakia other urban			1	1	1
Latakia rural	2	4	3	1	3
Edleb center		0	1	1	1
Edleb other urban	0	1	2	1	1
Edleb rural	5	7	2	0	4
Aleppo center	2	9	15	24	12
Aleppo other urban	2	3	1	1	2
Aleppo rural	18	6	1	0	6
Rakka center	1	3	1	0	1
Rakka other urban	0	0	0		0
Rakka rural	5	3	0		2
Der Elzor center	0	0	2	3	2
Der Elzor other urban		0	1	0	0
Der Elzor rural	9	4	1	0	3
Hasakeh center	0	0	1	1	1
Hasakeh other urban		2	3	3	2
Hasakeh rural	17	3	0		4
Sweda center		0	0	1	0
Sweda other urban		0	0	0	0
Sweda rural	3	3	1		2
Dara center	1	0	0	0	0
Dara other urban	1	2	1	0	1
Dara rural	5	5	2	0	3
Qunitra rural	0	0	0		0
Total	100	100	100	100	100

5 Conclusion

The aim of this paper was to document and investigate the asset index estimated for the CBS Syria 2003 Unemployment Survey. In the complete absence of questions about households' incomes and expenditures we decided to develop an asset index as an indicator for the households' *long-term* wealth. The index included 7 variables about ownership of various consumer durables, 3 questions about households' land ownership, and 8 ordinal or scale level variables about the quality and characteristics of dwelling and its infrastructure. In order to avoid the completely arbitrary method of giving all assets equal weights regardless of their value and type as in the simple additive asset index, we *estimated* the weights of our asset index, using the statistical procedure of *principal components*. The assumption behind using this method that the *first* principal component, that causes the most common variation for a list of asset variables, may be labelled "long-run household wealth", and should then be used for weighting each item in the index.

Testing the internal coherence of the index we made the problematic finding that land ownership is associated with a *low* rather than a high score on the index. The reason that the index has a fairly strong urban bias in the sense that it "rewards" items associated with *modern* society (e.g. computers and mobile phones) rather than *traditional* assets, such as land. Even though land ownership gives status at the village level, living in a *rural* area is *not* associated with a high asset index score in the *national* Syrian context.

The index is generally "robust" to the assets included, and the lowest 20% of households on the full index are also mostly classified among the lowest 20% on two sub-indices for respectively consumer durables, and housing and infrastructure variables. Moreover, the index produces reasonable comparisons with indicators related to household welfare, e.g. with male and female levels of education, which, however, are also related to "modern" society. Finally, the distribution of the scores of the asset index across regions and urban and rural locality types seem to be on accord with most common perceptions of the distribution of household welfare in Syria.

Our conclusion is that the index performs relatively well, in spite of a fairly limited number of asset variables in the questionnaire, and in spite of the problem that only a small minority of households owns these assets. Hence, we recommend that it may be used as an *ordinal* reference variable for households' *long-term* economic wealth in the CBS 2003 Unemployment Survey. However, we want to stress that the intention of the index is solely to serve as a reference for employment data. *We thus strongly warn against using the asset index as an independent indicator of welfare.* Only the forthcoming analysis of the household income and expenditure survey may be expected to produce reliable measurement of household welfare and poverty across regions and socio-economic groups.

References

El Laithy, H. and K. Abu-Ismail (2005) *Poverty in Syria: 1996-2004. Diagnosis and pro-poor policy considerations*, Damascus: UNDP.

Filmer, D. and L. Pritchett (1998) "Estimating wealth effects without expenditure data - or tears: with an application to educational enrollments in states of India", World Bank Policy Research Working Paper 1994, Washington DC: World Bank.

Montgomery, M, K. Burke, E. Paredes, and S. Zaidi (1997) "Measuring Living Standards with DHS Data: Any Reason to Worry?" mimeo, Research Division, The Population Council. New York: The Population Council

Tables for Sub-Indices

Table 11 Scoring factors and summary statistics for the urban sample

Urban Syria, all items					Group means			
Asset variable	Scoring factors (F1)	Mean (a1)	Std.dev (s1)	(F1)/(s1)	Lowest 20%	Lower middle 30%	Upper middle 30%	Upper 20%
Live in flat	0.639	0.547	0.498	1.284	0.004	0.272	0.884	0.906
Live in dar	-0.624	0.425	0.494	-1.262	0.971	0.679	0.099	0.078
Live in clay house	-0.068	0.021	0.144	-0.468	0.019	0.042	0.016	0.002
Live in other dwelling type	0.034	0.004	0.066	0.512	0.000	0.005	0.001	0.013
No. of rooms used for sleeping	0.477	1.400	0.786	0.607	0.918	1.199	1.511	1.955
Connected to water network	0.184	0.968	0.175	1.052	0.866	0.989	0.990	0.993
Connected to sewerage network	0.152	0.956	0.205	0.742	0.856	0.976	0.978	0.981
No. of (ordinary) telephone lines	0.528	0.739	0.475	1.111	0.148	0.671	0.945	1.038
No. of mobile phones	0.610	0.138	0.419	1.455	0.002	0.011	0.059	0.577
No. of private cars	0.585	0.132	0.370	1.580	0.000	0.025	0.063	0.521
No. of rooms with air-conditioner	0.540	0.141	0.445	1.215	0.003	0.038	0.058	0.556
No. of washing machines	0.664	0.411	0.523	1.269	0.000	0.158	0.529	0.967
No. of freezers	0.473	0.235	0.452	1.046	0.039	0.140	0.194	0.619
No. of computers	0.594	0.107	0.309	1.925	0.000	0.003	0.029	0.483
Size of irrigated land (dunums)	0.042	1.273	9.884	0.004	1.300	1.135	0.810	2.201
Size of non-irrigated land (dunums)	-0.010	3.958	29.967	0.000	6.251	3.635	2.733	4.375
Size of non-cultivated land (dunums)	0.042	0.150	1.462	0.029	0.094	0.099	0.114	0.334
Own commercial, industrial or service establishment	0.382	0.210	0.407	0.937	0.043	0.140	0.214	0.459
Urban asset index		0.000	3.540		-4.184	-2.009	0.812	5.458

Table 12 Scoring factors and summary statistics for the rural sample

Rural Syria, all items					Group means			
Asset variable	Scoring factors (F1)	Mean (a1)	Std.dev (s1)	(F1)/(s1)	Lowest 20%	Lower middle 30%	Upper middle 30%	Upper 20%
Live in flat	0.581	0.103	0.303	1.916	0.000	0.002	0.024	0.475
Live in dar	-0.148	0.750	0.433	-0.341	0.484	0.908	0.943	0.489
Live in clay house	-0.351	0.135	0.342	-1.027	0.504	0.083	0.023	0.011
Live in other dwelling type	0.129	0.005	0.070	1.853	0.000	0.000	0.003	0.020
No. of rooms used for sleeping	0.283	1.169	0.782	0.362	0.990	0.947	1.287	1.518
Connected to water network	0.400	0.756	0.430	0.931	0.154	0.873	0.925	0.928
Connected to sewerage network	0.479	0.478	0.500	0.959	0.004	0.423	0.623	0.823
No. of (ordinary) telephone lines	0.590	0.449	0.497	1.185	0.005	0.138	0.783	0.888
No. of mobile phones	0.430	0.029	0.180	2.385	0.000	0.000	0.005	0.136
No. of private cars	0.332	0.096	0.305	1.089	0.026	0.022	0.088	0.293
No. of rooms with air-conditioner	0.357	0.012	0.124	2.873	0.000	0.000	0.002	0.059
No. of washing machines	0.651	0.097	0.312	2.085	0.000	0.002	0.009	0.470
No. of freezers	0.449	0.030	0.172	2.616	0.000	0.006	0.012	0.126
No. of computers	0.533	0.020	0.140	3.812	0.000	0.000	0.000	0.100
Size of irrigated land (dunums)	-0.104	6.258	21.750	-0.005	12.018	5.236	5.302	3.476
Size of non-irrigated land (dunums)	-0.206	12.882	37.099	-0.006	35.607	7.283	7.224	7.079
Size of non-cultivated land (dunums)	-0.012	0.473	4.142	-0.003	0.790	0.199	0.602	0.392
Own commercial, industrial or service establishment	0.414	0.105	0.306	1.352	0.005	0.021	0.092	0.351
Rural asset index		0.000	2.873		-2.926	-1.197	0.246	4.413

Table 13 Scoring factors and summary statistics for the consumer durables sub-index

Total Syria, consumer durable items					Group means		
Asset variable	Scoring factors (F1)	Mean (a1)	Std.dev (s1)	(F1)/(s1)	Lowest 37%	Middle 44%	Upper 20%
No. of (ordinary) telephone lines	0.522	0.596	0.508	1.028	0.000	0.935	0.954
No. of mobile phones	0.681	0.086	0.333	2.045	0.000	0.003	0.434
No. of private cars	0.605	0.112	0.338	1.788	0.000	0.070	0.417
No. of rooms with air-conditioner	0.635	0.081	0.341	1.861	0.000	0.001	0.411
No. of washing machines	0.681	0.262	0.463	1.472	0.000	0.223	0.837
No. of freezers	0.602	0.139	0.365	1.649	0.000	0.062	0.573
No. of computers	0.674	0.065	0.247	2.724	0.000	0.000	0.334
Consumer durable index		0.002	2.787		-1.932	-0.408	4.535

Table 14 Scoring factors and summary statistics for the housing, infrastructure and land sub-index

Total Syria, housing, infrastructure and land					Group means			
Asset variable	Scoring factors (F1)	Mean (a1)	Std.dev (s1)	(F1)/(s1)	Lowest 20%	Lower middle 30%	Upper middle 30%	Upper 20%
Live in flat	0.817	0.350	0.477	1.713	0.000	0.014	0.516	0.998
Live in dar	-0.582	0.569	0.495	-1.175	0.749	0.919	0.460	0.000
Live in clay house	-0.383	0.072	0.258	-1.485	0.238	0.056	0.014	0.000
Live in other dwelling type	0.016	0.005	0.068	0.231	0.001	0.004	0.010	0.002
No. of rooms used for sleeping	0.345	1.297	0.793	0.436	0.958	0.996	1.373	2.012
Connected to water network	0.608	0.874	0.332	1.834	0.470	0.969	0.987	1.000
Connected to sewerage network	0.704	0.744	0.436	1.612	0.065	0.860	0.951	1.000
Size of irrigated land (dunums)	-0.281	3.483	16.438	-0.017	10.289	2.666	1.382	0.438
Size of non-irrigated land (dunums)	-0.299	7.914	33.611	-0.009	23.252	6.488	2.665	1.163
Size of non-cultivated land (dunums)	-0.056	0.293	2.970	-0.019	0.482	0.293	0.249	0.152
Own commercial, industrial or service establishment	0.324	0.163	0.370	0.878	0.026	0.044	0.226	0.403
Housing, infrastructure and land asset index		0.001	2.414		-3.418	-0.810	1.213	3.166

An asset index for the Syrian 2003 Unemployment Survey

Fafo

P.O.Box 2947 Tøyen
N-0608 Oslo
http://www.fafo.no/english/

Fafo-paper 2006:01
Order no.: 792
ISSN 0804-5135