ECONOMIC RESEARCH FORUM

منتدى البحوث الاقتصادية

2010

working paper series

GENERATING RELIABLE DATA TO PERFORM DISTRIBUTIONAL ANALYSIS IN THE ARAB REGION

Sami Bibi and AbdelRahmen El Lahga

Working Paper No. 561

GENERATING RELIABLE DATA TO PERFORM DISTRIBUTIONAL ANALYSIS IN THE ARAB REGION

Sami Bibi and AbdelRahmen El Lahga

Working Paper 561

October 2010

Send correspondence to:
Sami Bibi
Department of Economics, University of Laval, Quebec, Canada
Email: sami.bibi@ecn.ulaval.ca

First published in 2010 by
The Economic Research Forum (ERF)
7 Boulos Hanna Street
Dokki, Cairo
Egypt
www.erf.org.eg

Abstract

The objective of this paper is twofold. Firstly, we show how reliable datasets can be generated to perform distributional analysis in the Arab region, based on a useful procedure of desegregating grouped data published by official statistical agencies. Secondly, using accessible raw micro data and synthetic datasets, we rely on the existing conventional set of inequality indices to assess the degree of disparity and its evolution over time in the Arab region. The paper fills in an important gap of knowledge on data access and the level and patterns of inequality in the Arab region.

ملخص

الهدف من هذه الورقة البحثية ينقسم الى شقين. أولا: تعرض الورقة كيف يمكن إنشاء مجموعات من البيانات الموثوق بها يتم الاعتماد عليها في إجراء تحليل توزيعي في المنطقة العربية على أساس طريقة مفيدة لإلغاء التمييز العنصري في تجميع البيانات التي تنشرها الوكالات الإحصائية الرسمية. ثانيا: باستخدام مجموعة بيانات دقيقة وكذلك مجموعة من البيانات المؤلفة مع بعضها البعض، فنحن نعتمد على البيانات التقليدية الموجودة لمؤشرات التفاوت الاجتماعي لتقييم درجة التفاوت وتطورها على مر الزمن في المنطقة العربية. واخيرا: فهذه الورقة تهدف الى ملئ فراغ كبير في معرفة كيفية الوصول إلى البيانات ومستوى وأنماط التفاوت الاجتماعي في المنطقة العربية.

1. Introduction

Reliable distributional analyses of well-being should ideally be based on raw micro data of household surveys which typically include information on consumption expenditure. They also involve various demographic and socioeconomic characteristics of the household (household size and structure, region of residence, occupation status, etc.), and living conditions such as dwelling characteristics, possession of durable goods, etc. The availability of and accessibility to detailed household surveys are therefore a prerequisite for undertaking reliable distributive analyses.

Bibi and Nabli (2009, 2010) noted that distributive analyses in Arab countries (henceforth ACs) are scarce mainly because raw micro data is inaccessible.[1] Although household surveys are available in ACs, access to primary raw data is restricted for scholars and independent researchers for confidentiality concerns or political sensibility. Even for international organizations, such as the World Bank, access to the primary data is more restricted in ACs than the rest of the world (Iqbal, 2006). Unlike common practice in developed countries and many developing countries, we are aware of no country in the Arab region which provides unfettered or even easy access to the primary data for researchers. In some cases, *ad-hoc* access to such data for individual researchers is available but is not based on transparent and predictable rules.

However, statistical agencies in charge of distributive analyses publish summary statistics and grouped frequency tables by applying simple computational procedures to individual records. Scholars interested in distribution related issues often have to be content with these secondary materials, which seriously limits the in depth and spatial coverage of poverty and inequality analysis.[2] Some techniques that consist of fitting various parametric Lorenz curves to grouped income data to obtain estimates of the Gini index have certainly been proposed.[3] The most popular one is the POVCAL software developed by the World Bank.[4] In addition to providing summary statistics, the software enables inferring different poverty and inequality indices. Unfortunately, it seems that most of the fitted functions are unreliable, i.e., they often produce summary statistics which are very different from those obtained from individual micro data.[5] Furthermore, Minoiu and Reddy (2008) provided a systematic analysis of performances of parametric fitting based on grouped data. They concluded that accuracy of estimation depended on the shape of income distribution, functional form used and poverty and inequality indices estimated. Therefore conclusions based on such approach must be taken with some cautions.

More recently, Shorrocks and Wan (2008)—hereafter SW—have suggested a useful procedure to reconstruct the living standards distribution from any set of Lorenz curve coordinates. According to SW the characteristics of the generated sample exactly match the reported grouped values. The simulated data can then be used to estimate various inequality statistics. Preliminary experiments following SW and based on the US Current Population Survey show that the estimated results for the Gini coefficient are very encouraging.

The objective of this paper is twofold. First, we assess the reliability of the SW's procedure by contrasting the Lorenz curves and the Gini indices estimated using primary data with their

[1] Among the few studies on the extent of inequality in some Arab countries, see Abdelkhalek (2005) for Morocco, El-Laithy et al. (2003) for Egypt, El-Laithy and Abu-Ismail (2005) for Syria, World Bank (2004) for Jordan, and Zouari-Bouattour and Jallouli (2001) for Tunisia.

[2] For an illustrative example on Arab countries, see Adams and Pages (2003).

[3] See Kakwani and Podder (1973), Basmann et al. (1990), Ortega et al. (1991), among many others.

[4] http://iresearch.worldbank.org/PovcalNet/jsp

[5] For a discussion on this, see Schader and Schmid (1994), Minoiu and Reddy (2008), and Shorrocks and Wan (2008).

synthetic counterparts, i.e., those computed by implementing the SW's procedure to grouped data. Grouped data will be either computed using primary data from several ACs' household surveys or collected from the POVCAL website. As we will see below, results are conclusive, and we extend the implementation of the SW's procedure to ACs where only grouped data can be collected from POVCAL or the national bureaus of statistics. Second, we use the set of accessible primary data and the synthetic data to portray the level and the trend of overall inequality in the Arab region.

The rest of the paper is structured as follows. Section 2 briefly summarizes the algorithm proposed by SW to reconstruct the whole distribution from grouped data. Section 3 evaluates the effectiveness of SW's algorithm in providing reliable synthetic data. Section 4 offers a global portray of the level and trend of inequality in a selected Arab countries. Section 5 concludes the paper.

2. Desegregating Grouped Data

Shorrocks and Wan (2008) have suggested an algorithm to construct a synthetic self-weighted sample of *n* observations. The properties of the generated distribution are admittedly as close as possible to those of the grouped data.

The SW's procedure can be summarized as follows. Consider a partition of income distribution into *m* disjoint income classes. Let (p_k^*, L_k^*) be *m*+ 1 Lorenz coordinates calculated from raw data, with ($k = 1, ..., m$). We denote by p_k^* the aggregate proportion of the population in the poorest income classes 1 to *k*, L_κ^* is the corresponding cumulative income share, and $(p_0^*, L_0^*) = (0, 0)$. The asterisks are used to distinguish the target (true) values from the synthetic sample values (non-asterisked), which may not exactly match the target figures. Given that Lorenz coordinates are typically calculated from published quantile shares, details of the absolute levels of income are lost in the construction of Lorenz curves. Thus, during the reconstruction of the synthetic distribution, the overall mean income is assumed to be unity. The mean income of a class *k* is therefore given by

$$\mu_k^* = \frac{L_k^* - L_{k-1}^*}{p_k^* - p_{k-1}^*}, \; k = 1, ..., m. \tag{1}$$

The synthetic sample of size *n* can be partitioned into *m* non-overlapping groups, where each group *k* contains $m_k = n(p_k^* - p_{k-1}^*)$ observations. The value of the i^{th} observation in class *k* is denoted by $x_{ki} \; (k = 1, ..., m; i = 1, ..., m_k)$, and the (normalized) sample mean of the class k is given by μ_κ.

The ungrouping procedure involves two stages. Using grouped data, we can fit, in the first stage, a suitable parametric income distribution, the lognormal distribution, to construct a preliminary sample of size *n*. The obtained sample values are, in the second stage, adjusted until that the synthetic sample statistics match those of the reported grouped data. Shorrocks and Wan (2008) illustrate their procedure by using the lognormal distribution to generate 1000 observations. The standard deviation of log incomes, σ, is obtained by averaging the $m - 1$ estimates:

$$\sigma_k = \Phi^{-1}(p_k^*) - \Phi^{-1}(L_k^*), \quad k = 1, ..., m - 1, \tag{2}$$

where Φ is the standard normal distribution function. The raw sample may then be generated by the percentile points 0.001, 0.002, ..., 0.999 corresponding to the fitted lognormal distribution.

In the second stage, observations of the initial sample are adjusted until the sample statistics match the true values. To do this, the sample observations are adjusted so that each of the class k means incomes, μ_k, is transformed into the corresponding 'true' values, μ_κ^*, and the appropriate changes are made to the intermediate values. More specifically, consider any interval $[\mu_k , \mu_{k+1})$, for $k = 1,...,m - 1$, and convert the initial sample value $x_j \in [\mu_k , \mu_{k+1})$ into the intermediate value $\hat{x}_j$ according to the rule

$$\frac{\hat{x}_j - \mu_k^*}{\mu_{k+1}^* - \mu_k^*} = \frac{x_j - \mu_k}{\mu_{k+1} - \mu_k}, \quad for\ k = 1, ..., m-1\ and\ x_j \in [\mu_k, \mu_{k+1}), \tag{3}$$

or equivalently

$$\hat{x}_j = \mu_k^* + \frac{\mu_{k+1}^* - \mu_k^*}{\mu_{k+1} - \mu_k}(x_j - \mu_k), \quad for\ k = 1, ..., m-1 \quad and x_j \in [\mu_k, \mu_{k+1}) \tag{4}$$

Similar adjustments are made at the bottom and top of the distribution using the rule

$$\hat{x}_j = \frac{\mu_1^*}{\mu_1} x_j for x_j < \mu_1; \hat{x}_j = \frac{\mu_m^*}{\mu_m} x_j \quad for\ x_j < \mu_k \tag{5}$$

Note that the transformation given by (3) or (4) is well defined because the raw sample from the first stage is both distinct and ordered, so that $\mu_{k+1} > \mu_k$. Note also that the transformations defined by (3), (4) and (5) are (weakly) monotonic, so that the non-decreasing order of the sample is preserved.

Such adjustments ensure that, within each income class, the true mean lies within the range of sample values. Putting it analytically,

$$\min_i \hat{x}_{ki} \le \mu_k^* \le \max_i \hat{x}_{ki}, \quad for\ k = 1, ..., m. \tag{6}$$

The second stage ensures that the group bounds are kept fixed and that the gaps between the sample values and the upper (lower) bound of the group are compressed when the sample mean is below (above) the true value. To see this, define the lower bound of each group as

$$c_1 = 0; c_k = \frac{1}{2}(\max \hat{x}_{k-1,i} + \min \hat{x}_{k,i}), \quad k > 1, \tag{7}$$

and convert the intermediate value $\hat{x}_{\kappa i}$ into the final value $\hat{x}_{\kappa i}^*$ according to the rule

$$\hat{x}_{ki}^* = \begin{cases} c_{k+1} - \frac{c_{k+1} - \mu_k^*}{c_{k+1} - \hat{\mu}_k}(c_{k+1} - \hat{x}_{ki}), & \text{if } \mu_k^* > \hat{\mu}_k \text{ and } k < m, \\ c_{k+1} - \frac{\mu_k^* - c_k}{\hat{\mu}_k - c_k}(\hat{x}_{ki} - c_k), & \text{if } \mu_k^* < \hat{\mu}_k \text{ or } k = m. \end{cases} \tag{8}$$

Applying this procedure ensures preserving the sample ordering both within and between income groups and that the synthetic distribution exactly replicates the properties of the published grouped frequency table.

3. Assessing the Reliability of Shorrocks and Wan's Procedure

We turn now to assessing the effectiveness of SW's procedure. For this end, we check the reliability of SW's procedure by contrasting the Gini indices estimated using household micro data with their synthetic counterparts, i.e., those computed by the implementation of SW's procedure to grouped frequency tables attached to the same surveys. Grouped data is either computed using primary data from several ACs household surveys or collected from the POVCAL website. If the results are conclusive, we extend the implementation of the

SW's procedure to ACs, where only grouped data is affordable or accessible from national or international providers of secondary material.

The assessment procedure works as follows.

(i) For each raw dataset, we generate 10 points of Lorenz curve coordinates.

(ii) We apply SW's procedure by fitting a lognormal distribution to the 10 Lorenz curve coordinates in order to generate a self-weighted sample of *n* observations.[6]

(iii) We compare the estimated inequality statistics based on synthetic samples obtained in step *(ii)* to the estimated ones based on the household micro data.

Table 1 reports the results of estimated Gini and its standard error from different sources. The results are encouraging since the estimated values from different sources are very close to each other. Using t-statistics, we formally test for the differences between estimated Gini indices from each data source. Table 2 presents test results and show that none of the estimated indices from the ungrouped data is statistically different from the indices calculated using the primary raw data.

In order to test the robustness of our findings, we compare the estimated density of raw data (deflated by the distribution mean) to that generated from grouped data based on decile shares and/or POVCAL data (when available), respectively. Looking at the cases of Yemen 1998, Tunisia 1990, Syria 2003 and United Arab Emirates (UAE) 2008, Figures 1 to 4 show that raw data density of household expenditure per capita is practically undistinguishable from that of synthetic samples generated from grouped data. Other experiments, not reported here for the sake of expositional simplicity, show that the same findings apply to all other countries. This means that the effectiveness of SW's procedure goes beyond the calculation of the Gini indices to other poverty or inequality statistics or curves. Summing up, our results show that:

1. The difference between any distributive statistics estimated from household micro data and their synthetic counterparts are not statistically significant;
2. The grouped data published by POVCAL is a reliable source that can be used to conduct any distributional analysis.

4. Spatial and Temporal Comparisons of Inequality in the Arab Region

We now use the dataset generated following the procedure described in Section 2 to characterize the dynamics of inequality across and within 10 Arab countries, namely Algeria, Djibouti, Egypt, Jordan, Mauritania, Morocco, Syria, Tunisia, the United Arab Emirates (UAE) and Yemen. Except for the UAE, grouped data on household expenditure per capita—for two time periods or more—is accessible. Although only one dataset is accessible in the case of the UAE, it is included because we have grouped data on both income and expenditure. Recall that the synthetic distributions are self-weighted. However, when household micro data is used, the results presented are weighted by raw household sample weights multiplied by the household size. Household expenditures include all durable and non-durable expenditure plus the implicit rent from own housing. Despite efforts made to harmonize datasets, we note that comparisons should be taken with caution. Indeed, survey designs and temporal coverage are in many case very different (period covering expenditures, recall bias.. etc.). These differences cannot be treated ex-post. However, within each country, all surveys used are fairly comparable over time.

[6] Shorrocks and Wan (2008) Show that the lognormal function used to generate synthetic sample performs much better than other specification like General Quadratic (GQ), Beta, Generalized Beta (GB), and Singh-Maddala (SM) functions.

Table 3 presents expenditure shares of 10 deciles of the distribution. That is to say, household expenditure per capita is sorted in ascending order and divided into 10 groups of equal size. Remarkably enough, the expenditure share of the poorest decile is always below 3.7 percent in all 10 countries. For most countries, this share is even lower than three percent. The richest individuals of the tenth decile always hold at least 25 percent of total expenditure, but it ranges from 25.8 percent in Yemen (1998) to 42.5 percent in Mauritania (1993). For almost all countries, the expenditure share of the poorest decile has remained roughly stable over time. The gaps between deciles are more pronounced in the upper tail of the welfare distribution, especially between the ninth and tenth deciles, as the last column of Table 3, which reports the ratio of the ninth to the tenth decile, clearly illustrates.

Now if we compare three North-African countries, namely Egypt, Algeria and Tunisia in 1995, it seems that Egypt is the most equal society where the shares of the first six deciles are greater than those observed in Algeria and Tunisia and where the same shares are lower. Making the same comparison between Syria and Jordan in 1997 and Yemen in 1998, we can show that Syria displays lower inequality than the other two countries.

In parallel to the descriptive analysis presented above, Table 4 presents a set of usual inequality measures along with their standard errors. The Gini, the Atkinson's (1970) and the generalized entropy index for different parameterizations.[7] The general pattern of findings is that all indices rank countries in the same manner, for a given period.

If we focus on the Gini inequality and the 1990s, it ranges from 0.497 for Mauritania in 1993 to 0.299 for Egypt in 1995. In 1995, Tunisia was the most unequal country with a Gini coefficient of about 0.416, followed by Mauritania, Algeria, Djibouti and Egypt. With a Gini index of about 0.299, Egypt's is the lowest inequality index observed over time in the Arab region. The remaining countries are not included in this comparison due to the absence of data for the same year.

By the mid 2000s, Tunisia remains the most unequal society with a Gini index of 0.41, followed by Mauritania in 2004, Yemen in 2006 and Egypt in 2005 (table 4). Meanwhile, figures 5 and 6 map the inequality level in ACs in the early 1990s and mid 2000s. Note that these figures present a rough comparison of inequality level across countries. One of the natural extensions of this study is to offer a more rigorous comparison of inequality across countries by constructing, for each country, a synthetic counterfactual distribution in 1990 and 2005.

We now turn our attention to the trend of inequality in each individual country with more than a single published household survey. We find Egypt, Morocco and Tunisia showing stable expenditure distributions between 1990–2004, 1984–2007 and 1980–2005, respectively. Syria experiences a significant increase of inequality, of about 3.6 percentage points, between 1997 and 2003 However, it appears from the most recent survey (2007) that inequality has decreased again to the 1997 level, i.e., a Gini index of 0.338. Similar fluctuations are observed in Yemen, Mauritania and Jordan. For Algeria, the somewhat outdated data of 1995 shows a moderate level for the Gini index (approximately 0.353) which is lower than 0.398 observed in 1988. Djibouti exhibits a relatively high inequality level (Gini=0.399) in 2002.

As for the United Arab Emirates (UAE), the only accessible household survey is the one conducted in 2008. Table (4) shows that this country displays a high level of inequality, with values of inequality indices that are close to those observed in Tunisia for example. As expected for any country, income inequality is clearly greater in UAE than expenditure inequality.

[7] For a formal presentation of each index see appendix A.

5. Conclusion

Access to reliable sources of microeconomic data in the Arab world is one of the serious obstacles faced by researchers in development economics. In this paper we evaluated the reliability of a simple method that generates individual data used in distributional analysis. We have shown that the procedure proposed by Shorrocks and Wan (2008) can, at least in part, overcome the difficulties of access to data. However we emphasize the need for improvements in (i) access to raw data surveys and (ii) standardization of survey questions and sampling design in order to help scholars conduct other relevant comparative studies across Arab countries. For instance, the generated data is used to portray an initial picture on levels and trends of monetary inequality in the Arab region. Statistical tables produced in this paper show that Arab countries witnessed different experiences in the evolution of overall inequality during the 1990s and the early 2000s. Clearly, a deeper analysis of each country is needed to understand other aspects of inequality such as contributions of socioeconomic groups, the main determinants of the observed inequality and its spatial dimensions. Some of these issues will be addressed in companion papers.

References

Abdelkhalek, T. 2005, "La Pauvreté au Maroc." Chapter and contribution for theme 7 of the thematic group "Les Phénomènes d'Exclusion, de Pauvreté, et d'Analphabétisme". du Rapport du Cinquantenaire sur *50 ans de développement humain et perspectives pour 2025. (RDH50)*.

Adams, R. and J. Pages. 2003. "Poverty, Inequality and Growth in Selected Middle East and North Africa Countries: 1980–2000." *World Development* Vol. 31, N. 12, pp. 2027–2048.

Atkinson, A. B. 1970. "On the Measurement of Inequality." *Journal of Economic Theory* Vol. 2, pp. 244–263.

Basmann, R. L., Hayes, K. J., Slottje, D. J., and J. D. Johnson. 1990. "A General Functional Form for Approximating the Lorenz Curve." *Journal of Econometrics* Vol. 43, pp. 77–90.

Bibi, S. 2009. "Complete and Partial Analysis of Pro-Poorness in Some Arab Countries." *Mimeo*. Université Laval and the World Bank.

Bibi, S., and M. K. Nabli. 2009. "Income Inequality in the Arab Region: Data and Measurement, Pattern and Trends." *Middle East Development Journal* Vol. 1, N.2, pp. 275-314.

Bibi, S, and M. K. Nabli. 2010. "Equity and Inequality in the Arab Region." *ERF Policy Research Report PPR No 33*. Egypt.

Cowell, F. 2000. *Measuring Inequality.* Third edition. Oxford University Press.

El-Laithy, H., and K. Abu-Ismail. 2005. "Poverty in Syria: 1996 – 2004." *Diagnosis and Pro-Poor Policy Considerations*. United Nation Development Programme (UNDP), Syria.

El-Laithy, H., M. Lokshin, and A. Banerji. 2003. "Poverty and Economic Growth in Egypt: 1995–2000." *World Bank Policy Research Working Paper 3068*.

Iqbal, F. 2006. *Sustaining Gains in Poverty Reduction and Human Development in the Middle East and North Africa*. Orientations in Development Series. World Bank, Washington.

Kakwani, N. C., and N. Podder. 1973. "On the Estimation of Lorenz Curves from Grouped Observations." *International Economic Review* Vol. 14, pp. 278–291.

Minoiu, C., and S. Reddy. 2008. "Kernel Density Estimation Based on Grouped Data: The Case of Poverty Assessment." *IMF Working Papers 08-183*. International Monetary Fund.

Ortega, P., Fernandez, M. G. A., Ladoux, M., and A. Garcia. 1991. "A New Functional Form For Estimating Lorenz Curves." *Review of Income and Wealth* Vol. 37, pp. 447–452.

Schader, M., and F. Schmid. 1994. "Fitting Parametric Lorenz Curves to Grouped Income Distributions - A Critical Note." *Empirical Economics* Vol. 19, pp. 361–370.

Shorrocks, A. F. 1984. “Inequality Decomposition by Population Subgroups.” *Econometrica* Vol. 52, N. 6, pp. 1369–1385.

Shorrocks, A. F., and G Wan. 2008. “Ungrouping Income Distributions: Synthesising Samples for Inequality and Poverty Analysis.” *UNU-WIDER Research Paper RP2008/16*. World Institute for Development Economic Research.

Theil, H. 1967. *Economics and Information Theory*. North Holland, Amsterdam.

World Bank. 2004. “Jordan: Poverty Assessment. Social and Economic Development Group.” *Middle East and North Africa Region Report N. 27658*- JO.

Zouari-Bouattour, S., and K. Jallouli. 2001. “Inequality of Expenses: The Tunisian Case.” *ERF Working Paper N. 0118*.

Figure 1: Density Estimates Yemen 1998 (Various Sources)

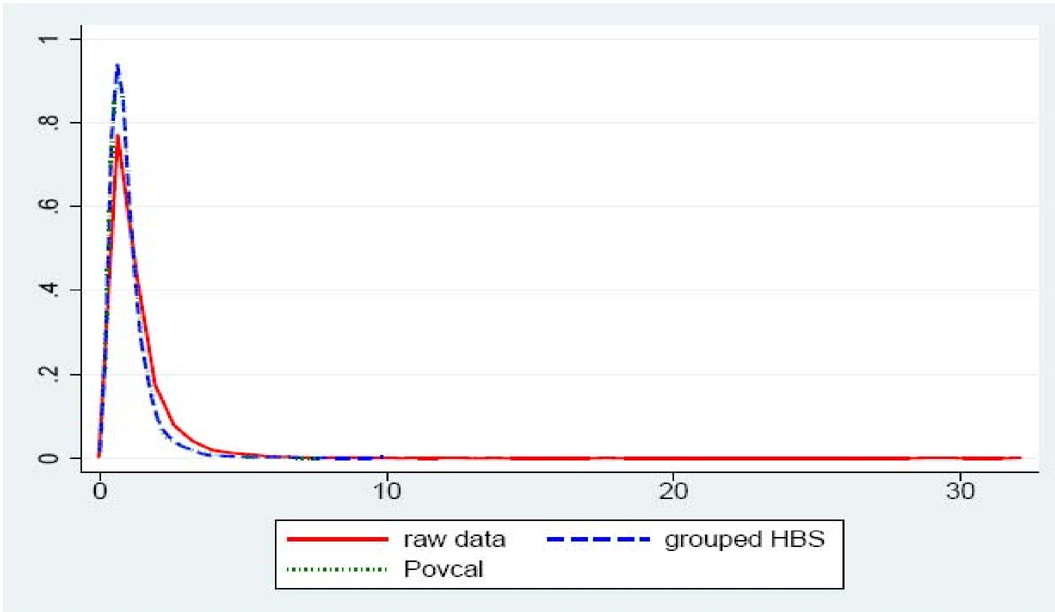

Figure 2: Density Estimates Tunisia 1990 (Various Sources)

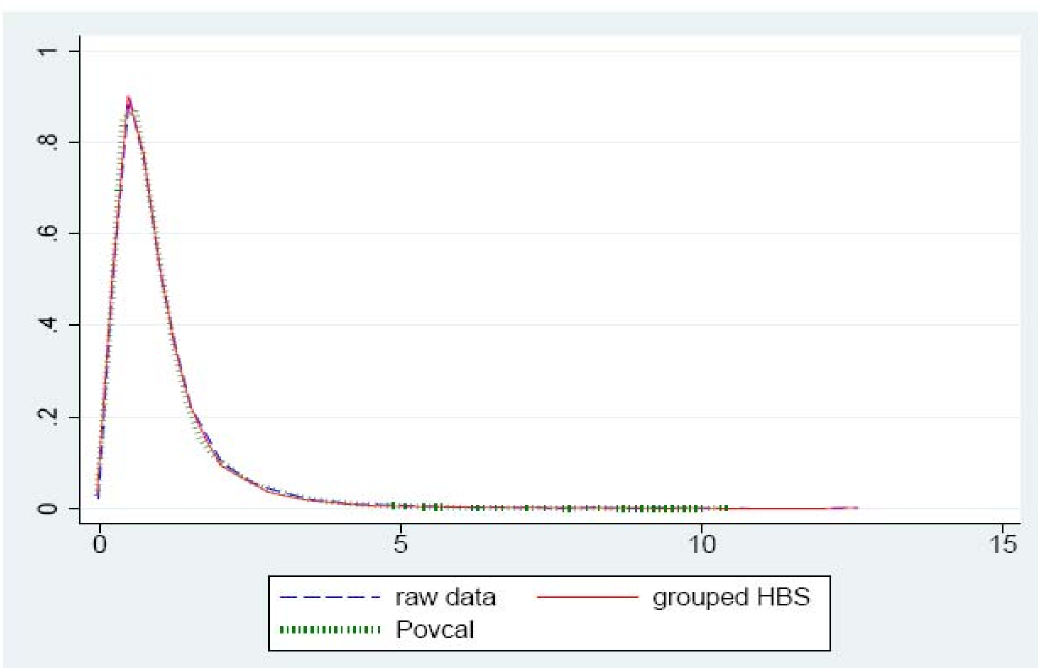

Figure 3: Density Estimates U.A.E 2008 (Various Sources)

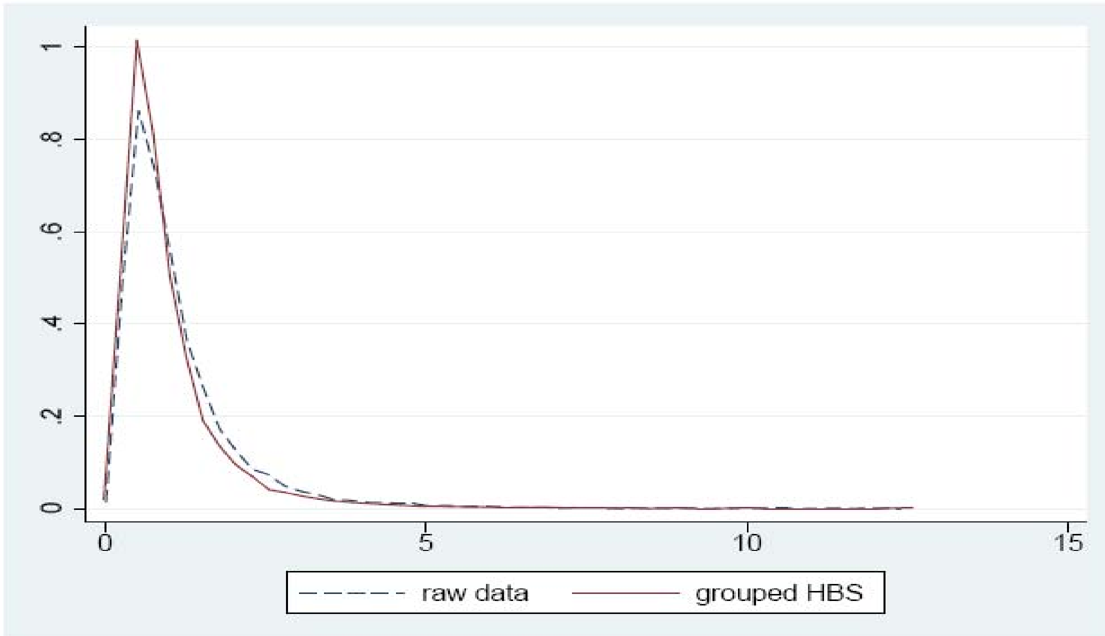

Figure 4: Density Estimates Syria 2003 (Various Sources)

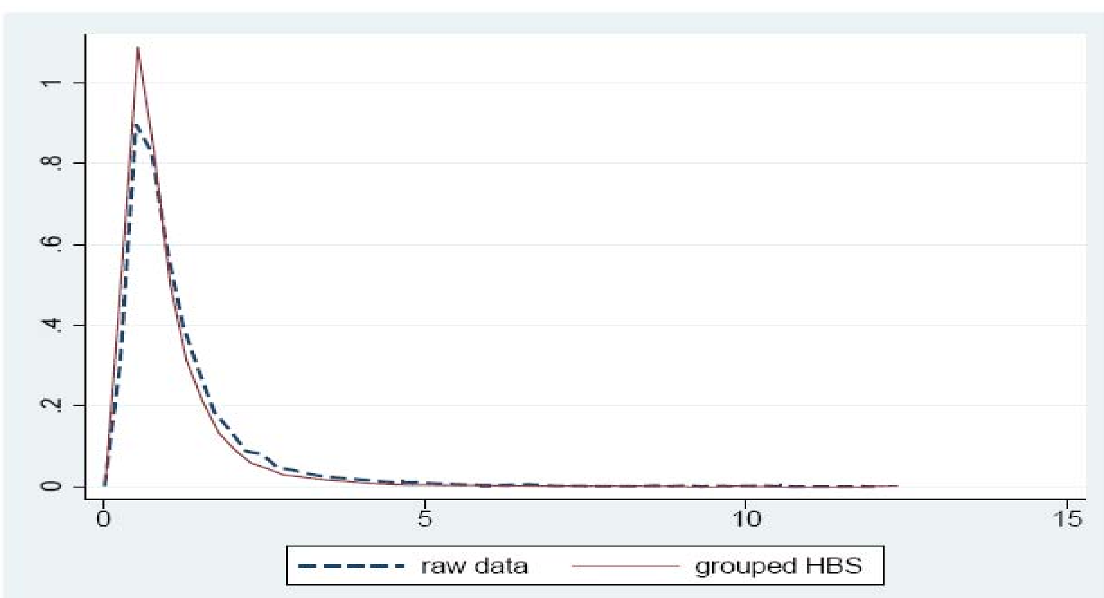

Figure 5: Gini Index Early 1990

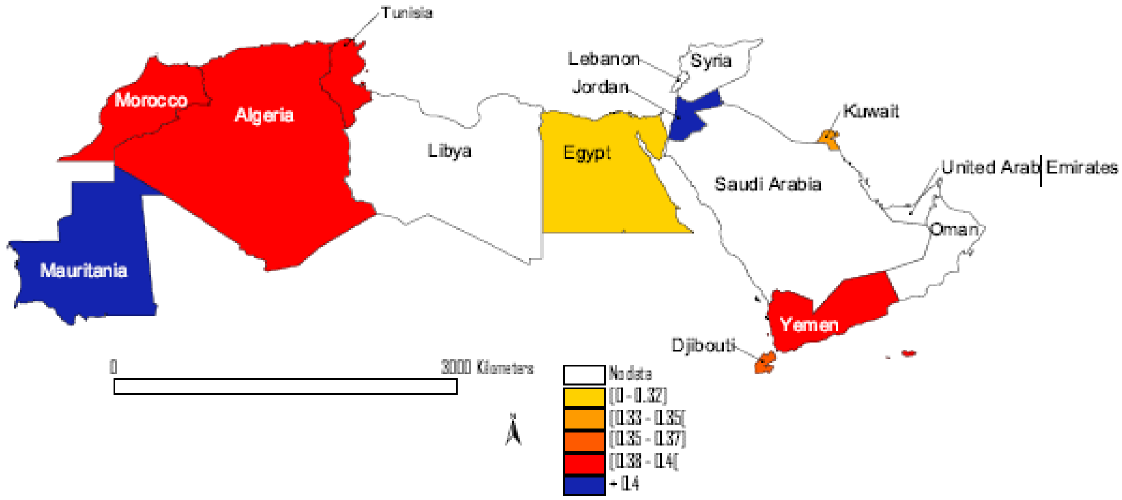

Figure 6: Gini Index Early 2000

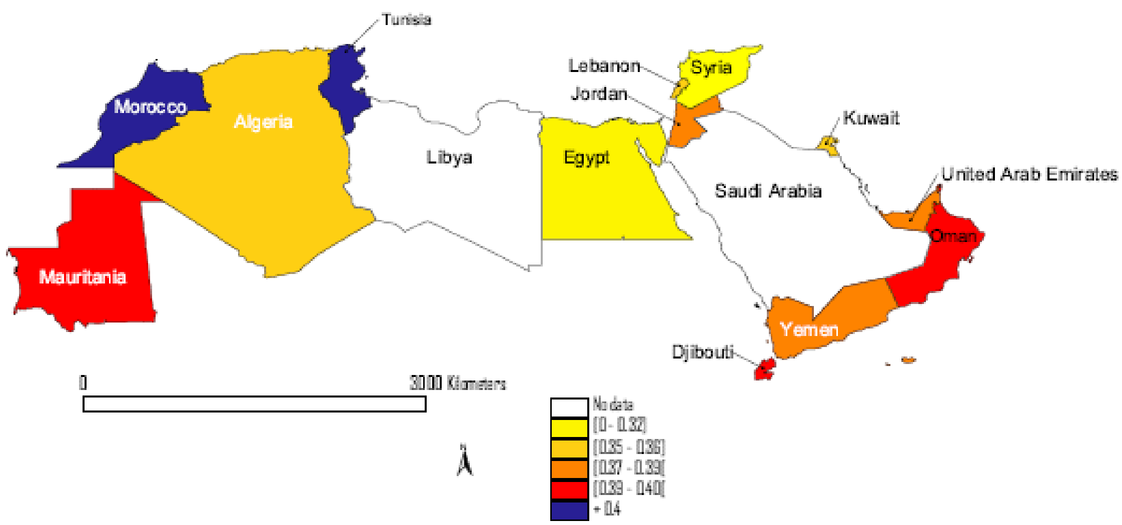

Table 1: Estimated Gini Indices Using SW Procedure and Raw Data.

Country	Type	Source	Year	Gini	Sd-Err
Morocco					
	Consumption	Povcal deciles	1991	.389	.006
	Consumption	Survey deciles	1991	.393	.008
		Raw data	1991	.393	.007
	Consumption	Povcal deciles	1999	.394	.008
	Consumption	Survey deciles	1999	.394	.008
		Raw data	1999	.395	.005
Tunisia					
	Consumption	Survey deciles	1985	.440	.011
	Consumption	Raw data	1985	.441	.005
	Consumption	Povcal deciles	1985	.431	.008
	Consumption	Survey deciles	1990	.401	.008
	Consumption	Povcal deciles	1990	.400	.007
	Consumption	Raw data	1990	.401	.004
	consumption	Survey deciles	1995	.416	.009
	Consumption	Povcal deciles	1995	.416	.007
	Consumption	Raw data	1995	.416	.004
	Consumption	Survey deciles	2000	.404	.009
	Consumption	Povcal deciles	2000	.407	.008
	Consumption	Raw data	2000	.404	.007
Yemen					
	Consumption	Survey deciles	1998	.344	.007
	Consumption	Povcal deciles	1998	.335	.006
	Consumption	Raw data	1998	.344	.003

Table 2: T-Statistics for the Difference between Gini Indices

Morocco 1991				Morocco 1999			
	Povcal	raw_data	Survey		Povcal	raw_data	Survey
Povcal	0	.5	.45	Povcal	0	.042	.034
raw_data		0	-.0018	raw_data		0	.0001
Survey			0	Survey			0

Tunisia 1985				Tunisia 1990			
	Povcal	raw_data	Survey		Povcal	raw_data	Survey
Povcal	0	.95	.69	Povcal	0	.14	.1
raw_data		0	-.004	raw_data		0	-.0024
Survey			0	Survey			0

Tunisia 1995				Tunisia 2000			
	Povcal	raw_data	Survey		Povcal	raw_data	Survey
Povcal	0	-.027	-.021	Povcal	0	-.37	-.32
raw_data		0	-.0018	raw_data		0	-.0052
Survey			0	Survey			0

Yemen 1998			
	Povcal	raw_data	Survey
Povcal	0	1.3	.97
raw_data		0	-.0027
Survey			0

Table 3: Distribution of Per Capita Household Expenditure: Share of Deciles and Expenditure Ratios

Country				Share of deciles										Income ratios	
				1	2	3	4	5	6	7	8	9	10	10/1	9/10
Name	Type	Source	Year												
Algeria															
	Consumption	Povcal	1988	2.57	3.94	4.94	5.90	6.87	7.99	9.32	11.26	14.57	32.66	12.71	0.45
	Consumption	Povcal	1995	2.76	4.20	5.28	6.33	7.45	8.68	10.30	12.40	15.80	26.81	9.71	0.59
Djibouti															
	Consumption	Povcal	1996	2.31	4.18	5.31	6.32	7.45	8.54	9.87	12.13	15.84	28.03	12.13	0.57
	Consumption	Povcal	2002	2.32	3.71	4.78	5.80	6.91	8.20	9.80	11.98	15.73	30.78	13.27	0.51
Egypt															
	Consumption	Povcal	1990	3.64	5.02	5.90	6.73	7.60	8.55	9.78	11.58	14.54	26.67	7.33	0.55
	Consumption	Povcal	1995	4.16	5.35	6.15	6.92	7.67	8.61	9.75	11.33	14.08	25.97	6.24	0.54
	Consumption	Povcal	1999	3.88	5.05	5.88	6.64	7.44	8.34	9.48	11.09	13.90	28.31	7.30	0.49
	Consumption	Povcal	2004	3.85	5.09	5.94	6.73	7.55	8.47	9.62	11.16	13.98	27.60	7.17	0.51
Jordan															
	Consumption	Povcal	1986	3.08	4.19	5.15	6.11	7.19	8.44	10.04	11.93	15.89	27.99	9.09	0.57
	Consumption	Povcal	1992	2.37	3.57	4.47	5.36	6.37	7.49	8.99	11.32	15.32	34.74	14.66	0.44
	Consumption	Povcal	1997	3.08	4.45	5.33	6.22	7.16	8.18	9.55	11.46	14.87	29.68	9.64	0.50
	Consumption	Povcal	2002	2.70	3.99	4.94	5.87	6.88	8.07	9.55	11.74	15.66	30.61	11.34	0.51
	Consumption	Povcal	2006	2.98	4.26	5.19	6.05	6.98	8.04	9.45	11.47	15.10	30.50	10.23	0.50
Mauritania															
	Consumption	Povcal	1987	1.32	2.95	4.14	5.31	6.69	8.26	10.13	12.65	16.51	32.04	24.27	0.52
	Consumption	Povcal	1993	2.04	3.09	3.88	4.69	5.62	6.65	8.01	10.06	13.44	42.50	20.83	0.32
	Consumption	Povcal	1995	2.39	4.03	5.05	6.16	7.25	8.58	10.22	12.23	15.77	28.32	11.85	0.56
	Consumption	Povcal	2000	2.45	3.72	4.76	5.83	6.97	8.27	10.06	12.26	16.15	29.54	12.06	0.55
	Consumption	RINS	2004	2.69	4.04	5.00	5.93	6.90	7.94	9.37	11.34	14.76	32.03	11.91	0.46
Morocco															
	Consumption	Povcal	1984	2.64	4.04	5.02	5.97	6.99	8.16	9.60	11.94	13.17	32.47	12.30	0.41
	Consumption	Povcal	1990	2.74	3.87	4.75	5.65	6.77	8.15	9.79	12.16	16.65	29.46	10.75	0.57
	Consumption	Povcal	1998	2.63	3.88	4.83	5.79	6.82	8.03	9.61	11.83	15.65	30.93	11.76	0.51
	Consumption	Povcal	2000	2.63	3.83	4.70	5.58	6.63	7.83	9.41	11.62	15.65	32.12	12.21	0.49
	Consumption	Povcal	2007	2.64	3.90	4.81	5.69	6.66	7.79	9.20	11.27	14.94	33.10	12.54	0.45
Syria															
	Consumption	RINS	1997	3.28	4.63	5.60	6.52	7.47	8.59	9.98	11.87	14.99	27.07	8.25	0.55
	Consumption	RINS	2003	3.03	4.23	5.13	6.01	6.94	8.11	9.59	11.76	15.35	29.84	9.85	0.51
	Consumption	RINS	2006	3.46	4.71	5.58	6.43	7.38	8.45	9.82	11.81	15.04	27.32	7.90	0.55

Table 3: *Continued*

Tunisia															
	Consumption	HBS	1980	1.93	3.29	4.32	5.36	6.51	7.93	9.60	11.93	15.97	33.15	17.18	0.48
	Consumption	Povcal	1985	2.20	3.37	4.32	5.34	6.45	7.71	9.31	11.61	16.06	33.63	15.29	0.48
	Consumption	Povcal	1990	2.28	3.60	4.66	5.77	6.91	8.25	9.90	12.10	16.26	30.27	13.28	0.54
	Consumption	Povcal	1995	2.24	3.45	4.43	5.48	6.62	7.99	9.76	12.15	16.14	31.72	14.16	0.51
	Consumption	Povcal	2000	2.34	3.61	4.62	5.63	6.76	8.04	9.71	11.97	15.82	31.49	13.46	0.50
	Consumption	RINS	2005	2.34	3.57	4.58	5.61	6.72	7.97	9.53	11.79	15.51	32.39	13.84	0.48
Yemen															
	Consumption	Povcal	1992	2.29	3.84	4.93	5.93	7.03	8.22	9.76	11.86	15.32	30.82	13.46	0.50
	Consumption	Povcal	1998	2.98	4.44	5.56	6.65	7.74	8.89	10.34	12.27	15.33	25.80	8.66	0.59
	Consumption	Povcal	2005	2.90	4.25	5.22	6.13	7.09	8.19	9.46	11.35	14.45	30.96	10.68	0.47
UA Emirates															
	Consumption	HBS	2008	2.63	3.87	4.83	5.79	6.82	8.03	9.61	11.82	15.65	30.94	11.76	0.51
	Income	HBS	2008	1.91	3.14	4.11	5.12	6.21	7.56	9.35	11.94	16.42	34.24	17.93	0.48

Table 4: Inequality Indices: Distribution of Household Per Capita Expenditure

Country				Gini	Theil	A (0.5)	A (1)	E (0)	E (2)
Name	Type	Source	Year						
Algeria									
	Consumption	Povcal	1988	.398	.294	.130	.232	.263	.445
				(.008)	(.015)	(.006)	(.008)	(.011)	(.035)
	Consumption	Povcal	1995	.353	.212	.100	.188	.209	.268
				(.006)	(.009)	(.004)	(.006)	(.007)	(.017)
Djibouti									
	Consumption	Povcal	1996	.367	.234	.110	.209	.235	.305
				(.007)	(.010)	(.004)	(.007)	(.009)	(.021)
	Consumption	Povcal	2002	.399	.288	.130	.237	.271	.420
				(.008)	(.015)	(.005)	(.008)	(.011)	(.036)
Egypt									
	Consumption	Povcal	1990	.320	.189	.085	.154	.167	.268
				(.007)	(.011)	(.004)	(.007)	(.008)	(.022)
	Consumption	Povcal	1995	.299	.162	.073	.134	.144	.215
				(.006)	(.008)	(.003)	(.005)	(.006)	(.013)
	Consumption	Povcal	1999	.325	.197	.088	.157	.170	.276
				(.007)	(.010)	(.004)	(.006)	(.007)	(.018)
	Consumption	Povcal	2004	.319	.187	.084	.152	.164	.258
				(.007)	(.009)	(.004)	(.006)	(.007)	(.017)
Jordan									
	Consumption	Povcal	1986	.360	.224	.103	.190	.211	.295
				(.006)	(.009)	(.004)	(.006)	(.008)	(.019)
	Consumption	Povcal	1992	.430	.345	.151	.265	.307	.540
				(.009)	(.017)	(.006)	(.009)	(.013)	(.046)
	Consumption	Povcal	1997	.363	.245	.109	.195	.217	.360
				(.008)	(.013)	(.005)	(.008)	(.010)	(.029)
	Consumption	Povcal	2002	.388	.271	.122	.220	.248	.390
				(.008)	(.013)	(.005)	(.008)	(.010)	(.030)
	Consumption	Povcal	2006	.375	.255	.114	.205	.229	.364
				(.007)	(.012)	(.005)	(.007)	(.009)	(.026)

Table 4: *Continued*

Mauritania									
	Consumption	Povcal	1987	.440	.341	.161	.307	.367	.474
				(.008)	(.016)	(.006)	(.009)	(.013)	(.041)
	Consumption	Povcal	1993	.497	.546	.213	.343	.420	1.233
				(.014)	(.038)	(.012)	(.016)	(.024)	(.148)
	Consumption	Povcal	1995	.373	.241	.113	.212	.238	.318
				(.007)	(.011)	(.004)	(.007)	(.009)	(.022)
	Consumption	Povcal	2000	.390	.263	.122	.225	.255	.353
				(.007)	(.011)	(.004)	(.007)	(.009)	(.025)
	Consumption	RINS	2004	.393	.302	.130	.227	.258	.515
				(.010)	(.021)	(.007)	(.010)	(.013)	(.062)
Morocco									
	Consumption	Povcal	1984	.389	.291	.127	.225	.255	.457
				(.009)	(.016)	(.006)	(.009)	(.012)	(.040)
	Consumption	Povcal	1990	.389	.259	.119	.219	.247	.345
				(.006)	(.010)	(.004)	(.006)	(.008)	(.022)
	Consumption	Povcal	1998	.394	.281	.126	.227	.257	.411
				(.008)	(.014)	(.005)	(.008)	(.010)	(.033)
	Consumption	Povcal	2000	.406	.303	.134	.238	.271	.463
				(.008)	(.016)	(.006)	(.009)	(.011)	(.040)
	Consumption	Povcal	2007	.405	.305	.134	.237	.271	.463
				(.008)	(.015)	(.006)	(.009)	(.011)	(.037)
Syria									
	Consumption	RINS	1997	.338	.207	.093	.171	.187	.295
				(.007)	(.012)	(.005)	(.007)	(.008)	(.028)
	Consumption	RINS	2003	.374	.259	.114	.204	.228	.398
				(.008)	(.016)	(.006)	(.008)	(.011)	.041)
	Consumption	RINS	2006	.338	.210	.094	.169	.185	.312
				(.008)	(.014)	(.005)	(.007)	(.009)	(.033)

Table 4: *Continued*

Tunisia									
	Consumption	HBS	1980	.433	.350	.155	.278	.326	.575
				(.010)	(.022)	(.007)	(.010)	(.014)	(.070)
	Consumption	Povcal	1985	.431	.335	.150	.268	.312	.498
				(.008)	(.016)	(.006)	(.009)	(.012)	(.040)
	Consumption	Povcal	1990	.400	.279	.128	.238	.271	.379
				(.007)	(.012)	(.005)	(.007)	(.010)	(.027)
	Consumption	Povcal	1995	.416	.306	.139	.253	.292	.433
				(.007)	(.014)	(.005)	(.008)	(.011)	(.033)
	Consumption	Povcal	2000	.407	.295	.134	.243	.278	.420
				(.008)	(.014)	(.005)	(.008)	(.010)	(.032)
	Consumption	RINS	2005	.413	.309	.138	.248	.285	.458
				(.008)	(.015)	(.006)	(.008)	(.011)	(.037)
Yemen									
	Consumption	Povcal	1992	.394	.284	.128	.233	.266	.419
				(.008)	(.015)	(.006)	(.008)	(.011)	(.036)
	Income	HBS	1998	.429	.368	.156	.273	.318	.709
				(.012)	(.030)	(.010)	(.013)	(.018)	(.109)
	Consumption	Povcal	1998	.335	.193	.090	.171	.187	.245
				(.006)	(.009)	(.003)	(.006)	(.007)	(.016)
	Consumption	Povcal	2005	.376	.272	.119	.210	.235	.430
				(.009)	(.016)	(.006)	(.009)	(.011)	(.038)
UA Emirates									
	Consumption	HBS	2008	.395	.285	.127	.228	.228	.431
				(.008)	(.016)	(.006)	(.009)	(.009)	(.041)
	Income	HBS	2008	.448	.371	.164	.293	.347	.600
				(.009)	(.021)	(.007)	(.010)	(.015)	(.068)

Table 5: A List of Household Budget Surveys

Country	Survey	Year	Already available?
Tunisia	HBCS	1975, 1980, 1985,	Yes
	HBCS	1990, 1995, 2000	Yes
Egypt	EIHS	1997	Yes
Morocco	HBCS	1991, 1998	Yes
Yemen	HBCS	1998, 2006	Yes
Syria	HIES	1997, 2004, 2007	Yes
Mauritania	EPCV	2004	Yes
Jordan	HEIS	1997, 2002	Yes
Emirates	HBS	2008	Yes

Table 6: A List of Lorenz Curves Coordinates (Already Available)

Country	Year
Algeria	1988, 1995
Djibouti	1996, 2002
Egypt	1990, 1995, 1999, 2004
Jordan	1986, 1992, 1997, 2002
Morocco	1984, 1991, 1998, 2000, 2007
Tunisia	1985, 1990, 1995, 2000, 2005
Yemen	1992, 1998, 2005

Appendix A: Measuring Inequality

In this appendix, we present analytical tools that are used to portray inequality levels and changes of overall inequality within each country included in our study.

A.1 Inequality indices

A natural starting point is to define a synthetic measure of income inequality. Consider a vector $Y = (y_1, y_2, \ldots, y_n)$ of living standards y_i (income, for short) for a population of n individuals, where y_i are ordered in increasing values, such that $y_1 \leq y_2 \leq \ldots \ldots \leq y_n$.[8] Further, let $y(p)$ be the quantile function giving the individual's income whose percentile in the distribution Y is p. The most common inequality measure used in the literature is the Gini index defined as:

$$I^{Gini}(Y) = \frac{1}{2n(n-1)\bar{y}} \sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j|$$
$$= \frac{n+1}{n-1} - \frac{2}{n(n-1)} \sum_{i=1}^{n} (n+1-i) y_i$$

where $\bar{y}$ is the average income. The Gini index can alternatively be given by the ratio of the area between the well known Lorenz curve and the perfect equality 45^o line to the whole area below the 45^o line, i.e.,

$$I^{Gini}(Y) = 2 \int_0^1 (p - L(p))\, dp$$

where $L(p)$ is the Lorenz curve giving the cumulative percentage of total income held by the poorest p proportion of the population:

$$L(p) = \frac{1}{\bar{y}} \int_0^p y(p) dp.$$

The main drawback of the Gini index is the very limited possibilities of its decomposition between population subgroups (areas of residence, social, ethnic groups, gender, etc.).[9] Such issues limit any study aiming to analyze the contribution of socioeconomic groups to overall inequality of a given income distribution. A natural solution to overcome such a problem is to resort to more appealing measures allowing an exact decomposition of inequality between groups. Shorrocks (1984) and Cowell (2000) show that any inequality statistic that fulfills some desirable principles such as the Pigou-Dalton transfer principle and the decomposability principle is a member of the Generalized Entropy (GE) class of inequality indices:[10]

$$I^{GE}(\theta) = \frac{1}{\theta^2 - \theta} \left[\frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i}{\bar{y}} \right)^{\theta} - 1 \right], \forall y_i > 0, \theta \neq 0, 1$$

[8] Although a discussion will be made in terms of income among individuals, any alternative measure of socioeconomic position (consumption, expenditure, earnings, wages, assets, land, education, health, occupational status index) among any well-defined recipient unit (households, workers, generations, per capita, per equivalent adult) would do.

[9] The Gini index is fully decomposable if (and only if) the incomes in one subgroup are all less than those in the other subgroups, i.e., under non-overlapping partitions of the income distribution.

[10] The Pigou-Dalton transfer principle suggests that an appropriate inequality measure should decrease following a progressive transfer from a rich to a poorer person.

where θ is a known parameter which captures the aversion to inequality and n, y_i, and $\bar{y}$ are as defined above. In contrast to most inequality indices that lie between 0 and 1 (like the Gini index), the values of GE range from zero (perfect equality) to infinity (high level of inequality).[11] The parameter θ can take any integer value. Commonly used values of θ are 0, 1 and 2. For $\theta = 0$, $I^{GE}(0)$ is simply the mean log deviation given by

$$I^{GE}(0) = \frac{1}{n}\sum_{i=1}^{n} \ln \frac{\bar{y}}{y_i}, \forall y_i > 0$$

Notice that $I^{GE}(0)$ is, in accordance with the transfer sensitivity principle, more sensitive to changes that occur in the bottom distribution. $I^{GE}(1)$ is the well known Theil (1967) index. It is formally defined as

$$I^{GE}(1) = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i}{\bar{y}} \ln \frac{y_i}{\bar{y}}, \forall y_i > 0$$

However, for $\theta > 1$, GE measures are more sensitive to changes that affect the upper tail of the distribution which make them less appealing for distributional judgments.

A.2 Ethical principles

Ideally, the functional form of an inequality measure should directly depend upon what we want to know about the specified aspect of income distribution. One has first to set the purpose of the analysis and then find a suitable measure within the framework. The axiomatic approach, which has been largely developed in the literature, seeks to fit this framework. To evaluate different aspects of welfare, it is necessary to examine various axioms, with regard to the purpose of the analysis and then, to select the basic axioms for a pertinent yardstick. For expositional simplicity, and in view of our interest in this paper, we present the most common used axioms.

Axiom 1 Principle of Population: the pooling of several identical income distributions does not affect the level of income inequality. Hence, for any integer k > 1, I(x[k]) = I(x) whenever x[k] is obtained from x by any k replications.

Axiom 2 Anonymity or Symmetry: any characteristic other than the individuals' income does not matter for measuring inequality. Hence, I(y) = I(x) whenever y is obtained from x by a permutation.

These two axioms are the common core axioms for poverty, inequality, and social welfare comparisons. In the context of disparity judgments, the core axioms also include the Pigou-Dalton transfer principle and the scale invariance.

Axiom 3 The Pigou-Dalton Transfer: a mean preserving transfer from a given person to a less well-off one (i.e., a progressive (equalizing) transfer, should not increase the inequality level); while a disequalizing (regressive) transfer should not decrease the inequality measure. Hence, $I(y)^2I(x)$ ($I(y)^3I(x)$) whenever y is obtained from x by a progressive (regressive) transfer.

Axiom 4 Scale Invariance: inequality level is invariant to any uniform proportional change of individual's income (or the welfare indicator for which the distribution is being

[11] However, one may normalize $I^{GE}(\theta)$ by its hypothetic maximum value, obtained when only one person owns all available resources, to make these indices range between 0 and 1. This is important for the purpose of an integrated analysis of inequality and social welfare (or poverty).

characterized). Hence, for any scalar $\alpha > 0$, $I(\alpha x) = I(x)$ for any inequality index I(.) is in line with this principle.

B. Data

In this section we present various lists of data used in this research paper. There are mainly two types of datasets. Conventional household budget surveys and grouped data extracted from the World Bank's POVCAL website. All surveys contain detailed information on household socioeconomic characteristics. For some countries, household income is also reported.