WPS4009

Azerbaijan's Household Survey Data: Explaining Why Inequality is So Low

Lire Ersado[1]
Human Development Network
Europe and Central Asia Region
World Bank

Abstract

While the Azerbaijan household income and expenditure survey (HIES) data satisfy most empirical regularities expected in a typical household survey data, the inequality measures based on the data are unusually low. For example, for the latest three years for which we have data (2002-2004), the consumption Gini coefficient—the commonly used summary measure of inequality—is in the range of 16—18 percent. This is among the lowest Gini coefficients ever observed in any country, and is extremely low even with the standard of countries generally considered as most equal in the world. Azerbaijan, a transitional economy with a significant natural resource base, is very unlikely to be the most equal country in the world. The objective of this paper is to investigate why inequality measures are unusually low in the Azerbaijan household survey data. The paper presents a methodology for diagnosing and identifying the potential sources of low inequality in the data, including cluster analysis at the primary sampling unit level. The main inference from the findings of the cluster analysis is that the observed low inequality indices are not due to poor supervision of the interviewers and the data collection process. We find that the main culprits for the observed low inequality in the HIES data are (1) the low participation rates of wealthy households in the household surveys, and (2) the widespread availability of well targeted public and private transfers.

1. Introduction

Azerbaijan has experienced appreciable economic growth over the last several years. Since 1998, by most macroeconomic indicators, it has enjoyed one of the highest rates of economic growth among the CIS countries. Between 1998 and 2005, average GDP growth rate is in double-digits (Figure 1). Macroeconomic stability, coupled with the recent boom in oil income and prudent management of the windfall, have contributed significantly to the improved economic performance. However, the lion's share of business and economic activities of the country is concentrated in the capital city.

Figure 1: Azerbaijan's economic growth has been impressive since 1998
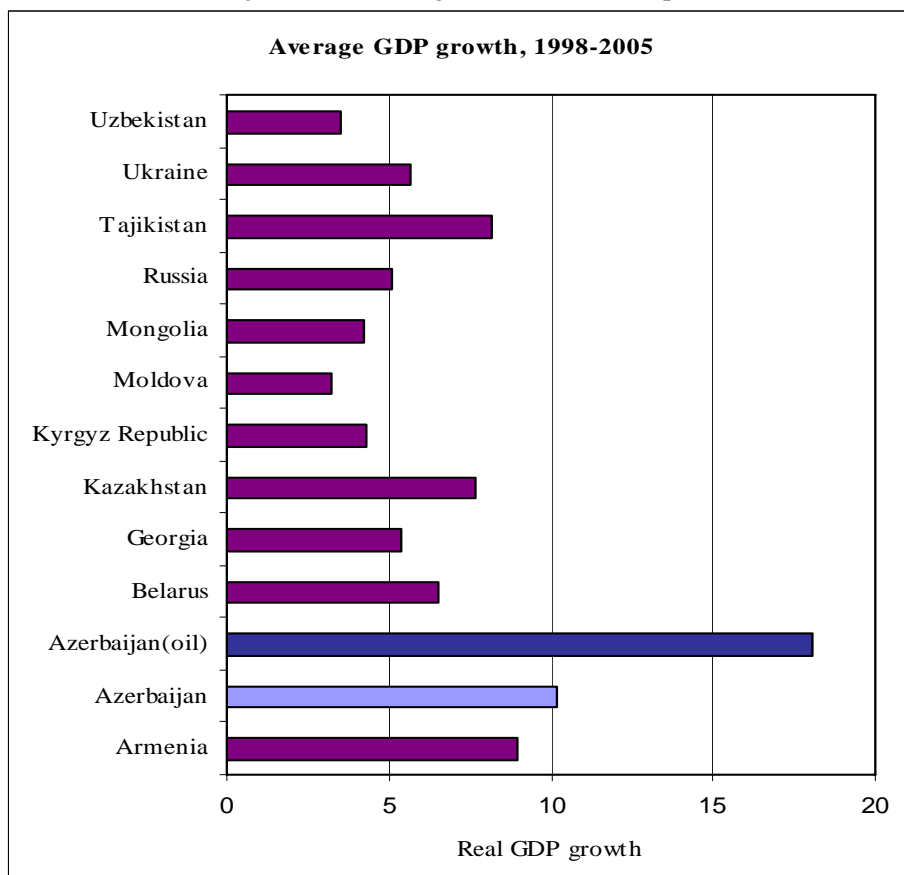
Source: Asian Development Bank; IMF; DDP, World Bank.

The impact of these developments on distribution of wealth and poverty remains largely unclear, however. Although the government of Azerbaijan undertakes a regular annual household survey, the resulting data have been considered by many as unreliable for measuring poverty due to extremely low inequality estimates in the Azeri survey data. The Azerbaijan Household Income and Expenditure Survey (HIES) data, which are the primary source of economic and social information for poverty analysis in the country, exhibit extremely low consumption inequalities. For instance, the measures of inequality such as the Gini coefficient of consumption expenditures—the commonly used summary measure of inequality—are unusually low. Azerbaijan's per capita consumption Gini coefficients estimable from the HIES data are low—by

a large margin—from those in other transition countries and countries with significant natural resource base such as the Russian Federation (see Table 1). For the last three years for which we have data, the Gini coefficients of consumption expenditures were in the range of 16—18 percent. These are among the lowest Gini coefficients ever observed in any country. These are extremely low even with the standard of countries generally considered as most equal in the world such as Denmark, where the Gini coefficient was about 25 percent in 2004. It is very unlikely that Azerbaijan is the most equal country in the world. This oddly low inequality figures has cast serious doubts on the quality and the reliability of the Azerbaijan HIES data.

Table 1. Most recent Gini indices of Azerbaijan and neighboring countries

Country	GINI Coefficient	Year
Azerbaijan	**18.4**	**2002**
Azerbaijan	**18.2**	**2003**
Azerbaijan	**16.3**	**2004**
Armenia	33.8	2003
Estonia	35.8	2003
Georgia	40.4	2003
Latvia	37.7	2003
Lithuania	36.0	2003
Macedonia, FYR	39.0	2003
Moldova	36.9	2002
Poland	34.5	2002
Russian Federation	39.9	2002
Tajikistan	32.6	2003
Turkey	43.6	2003

Source: Azerbaijan HIES data; WB Poverty Assessment Reports; WB Development Data Platform (DDP).

The objective of this paper is to investigate why inequality figures in the Azerbaijan household survey data are so low to extent of putting the quality and reliability of the data in question. Is it plausible that Azerbaijan has such a low inequality and that the data are merely reflecting the underlying reality in the country? Is the data generating mechanism systematically biased rendering inequality artificially low? Is the variability in the data suppressed due to disproportionately less representation of households in the various income and consumption echelons? Have lack of or poor supervision of the data collection and entry processes caused unnaturally high homogeneity in the data? Is a bad sampling design, which led to unrepresentative sample, to blame for the extremely low inequality? Are respondents, especially in the upper income echelon, concealing their incomes and expenditures, and by so doing depressed income and consumption expenditure disparity in the data? In this paper, we explore the data of the 2002-2004 HIES and seek answers to some or all of the above questions.

The paper puts forward various hypotheses why inequality is low in the Azerbaijan data and empirically tests them. We first employ ANOVA and cluster analysis to investigate if poor supervision of the data collection process had led to increased homogeneity in the resulting data. Secondly, we test the impact of public and private transfers on inequality by calculating

consumption inequality with and without various forms of transfers. Thirdly, we examine if inadequate representation of various wealth groups, particularly the poor and the rich households, explains the usually low inequality in the data. This requires the knowledge of the distribution of the non-responding or missing households. Once the missing observations are identified, the next step is how to account for them. We employ the Pareto distribution assumption for accounting for missing consumption per capita of households in the top and bottom wealth brackets (see Section 3). Finally, combination of the effects of transfers and missing households are tested. The results are reported in Section 4.

2. Background

The Azerbaijan Household Survey

The history of household survey data collection based on international practice is relatively recent in Azerbaijan. The Azerbaijan HIES was first introduced in 2001 and it has been undertaken on a regular annual basis ever since. Several factors are cited as affecting the quality of the Azerbaijan HIES data. To begin with, the current sampling frame, which is based on the 1999 population census, has relatively large primary sampling units (PSU) with, on average, over 16,000 people. Such big primary units are unsuitable for periodic updating and have the potential of putting at risk the selection of a random sample. The survey instrument is deficient for measuring certain aspects of household consumption such as expenditures on durables and does not allow differentiating between emergency medical care expenditures from regular health medical expenditures.

Another peculiar aspect of Azerbaijan household practice is the interviewers' hiring and work practice. The arrangement of interviewers may affect data collection practice and data quality. In Azerbaijan, one interviewer is permanently assigned to each selected PSUs. There is also a widespread understanding that some or most of the interviewers need further training to improve their ability to conduct the interviews. Moreover, there is unequal hiring practice for interviewers: out of a total 85 interviewers, 65 are regular staff of the SSC, while the rest are on contract basis and with less compensation. Such uneven pay and hiring practices could endanger the consistency and uniformity in the data collection process.

Finally, there is belief among those who are involved in the collection and use of the HIES data that the surveys do not adequately represent all income groups of the country. It has been widely acknowledged that rich households tend to decline participation in the surveys.

Empirical regularities in the HIES data

The Azerbaijan HIES data satisfy most empirical regularities expected in a typical household survey data. The data have a lot of the characteristics typically found in a good household survey data. In terms of the composition of household consumption, expenditure on food as a share of total household income consistently declines from poorest deciles to the richest deciles. While the poorest tend to spend nearly over 90 percent of their total income on food consumption, the richest spend, on average, a little over 40 percent of their total income on food. The Engel's Law,

which is accepted as a basic principle of income and consumption, is strongly evident in the Azeri household data for all three years considered in this analysis. As expected, rural areas spend a bigger share of their total expenditures on food. Households with significantly higher human capital such as educational attainment have higher income generating opportunity, other things being equal. Per capita income, expenditures and wage earning for households with household heads having higher level educational attainments are consistently higher than those corresponding to less educated household heads. These observations and empirical regularities hold true across time and geographic locations (see Ersado, 2006 for details).

Table 2. Consumption share of income decreases with income

	INCOME DECILES									
	1	2	3	4	5	6	7	8	9	10
2002										
Income (2003 prices)	95,205	125,080	141,808	155,152	167,062	178,533	191,894	208,677	235,342	332,082
Total consumption (% income)	144	112	102	96	93	89	87	86	83	80
Food (% income)	93	72	65	60	58	55	53	51	50	44
2003										
Income (2003 prices)	90,468	120,808	138,132	152,568	166,420	181,555	199,594	221,262	252,612	355,712
Total consumption (% income)	161	121	110	105	99	94	90	85	82	73
Food (% income)	107	81	73	67	62	58	54	51	47	40
2004										
Income (2003 prices)	109,357	134,651	149,794	163,564	175,837	188,862	202,804	220,803	254,302	365,693
Total consumption (% income)	132	110	105	100	99	97	93	91	88	79
Food (% income)	91	76	71	67	65	62	59	55	52	42

Source: Azerbaijan HIES data.

Another good aspect of the Azerbaijan data is that the survey made a reasonable capture of the Gross National Income (GNI). For instance, the average consumption per capita in the 2002 survey is $386 annually vs. the 2002 GDI per capita of $742. The capture ratio is thus 52 percent, which is slightly below the average 57 percent capture ratio for Eastern Europe and former Soviet Union countries. But the capture ratio of the Azerbaijani survey is much better than Armenia's (34 percent) and is similar to Georgia's (54 percent).

However, as discussed above, the extremely low inequality in the data is implausible for Azerbaijan. The ratio of per capita consumption of the top 5 percent of people compared to the bottom 5 percent (also known as ventile ratio) is extremely low. The ventile ratios are respectively, 4.2, 4.1 and 3.5 in 2002, 2003 and 2004 (Table 3). The corresponding ratio for the top and bottom 10 percent of the people was only 3.2, 3.1 and 2.8 respectively for 2002, 2003 and 2004. These values are significantly smaller than in comparable countries. For example, in 2002 consumption ventile ratio was 7 in Armenia, 22 in Georgia, 10 in Russia, and 7 in Ukraine (see Milanovic, 2005). It is not obvious whether this low value in Azerbaijan is due to an unrealistically high income of the bottom, or due to an unrealistically low income of the top.

4

Table 3. Azerbaijan has extremely low top-bottom consumption ratio

	Consumption (Azeri manats)		
	2002	**2003**	**2004**
bottom 5% (2003 prices)	81,826	87,207	102,157
Top 5% (2003 prices)	345,022	353,932	361,873
95/5 ratio	**4.2**	**4.1**	**3.5**
bottom 10%(2003 prices)	91,293	96,641	110,952
Top 10%(2003 prices)	299,803	308,527	315,816
90/10 ratio	**3.3**	**3.2**	**2.8**

Source: Azerbaijan HIES data.

The unusually low level of inequality is not just limited to the national aggregate data; it equally applies at the regional level. For example, the average per capita consumption in Baku is only 7 percent higher than the country-wide average (Table 4) in 2002. In 2004, it was about 12 percent higher than the national average. The average per capita consumption in the poorest region in 2002 and 2003 (i.e., Nakhchivan) is only about 12 percent lower than the national average. The regional Ginis are, similar to the national-wide Gini, uniformly low. Baku's inequality of close to only 20 percent would make it one of the most (perhaps the most) equal city in the world.

Table 4. Similar to the national inequality, regional inequalities are extremely low

	2002			2003			2004		
	Gini	Consumption (% of national average)	Income (% of national average)	Gini	Consumption (% of national average)	Income (% of national average)	Gini	Consumption (% of national average)	Income (% of national average)
Azerbaijan	18.4	100	100	18.2	100	100	16.3	100	100
Baku	21.4	107	114	20.0	107	115	17.9	112	117
Non-Baku urban	17.4	95	100	18.5	94	101	15.7	94	98
Rural	16.8	99	92	16.7	100	91	14.9	98	92
Nakhchivan	14.9	88	89	16.7	87	89	15.5	95	92
Absheron-Guba	19.4	97	105	19.3	97	103	15.3	98	99
Mughan_Salyan	15.9	96	94	15.5	95	94	15.5	89	93
Ganja-Gazakh	16.7	97	92	18.7	99	92	15.8	97	94
Sheki-Zagatala	14	102	93	14.7	99	93	12	99	95
Lankaran_Astara	15	99	94	13.8	101	93	14.5	101	94
Shirvan	17.7	95	96	17.4	96	96	16.6	91	93
Qarabagh-Mil	18.2	103	95	18.1	102	96	14.9	96	93

Source: Azerbaijan HIES data.

Did poor supervision lead to low inequality?

To investigate whether poor supervision of the data collection process led to low inequality, an ANOVA and cluster analysis techniques on the inequality estimates at the primary sampling unit level were employed. As each interviewer is permanently assigned to a certain primary sampling

unit, these analyses are tantamount to analyzing the effect of the interviewers on inequality measures. The main findings of this analysis are summarized below.[2]

The main inference from the findings of the ANOVA and cluster analysis is that the observed low inequality indices are not likely due to interviewers' work practice, nor do they reflect poor supervision of the data collection process. The results of the ANOVA and cluster analysis of the three years data were compared across years and interviewers to see if any pattern emerges by cluster of interviewers. The interviewers were very consistent for all three years 2002, 2003 and 2004 examined. Gini coefficients were uniformly low for all interviewers and for all years. The interviewers appear to be a reasonably homogeneous group in terms of enumeration level inequality. Furthermore, interviewers in the extreme cases (i.e., reporting relatively small or large Ginis) did not remain extreme in all years. For example, only two interviewers, each in 2003 and 2004, had a Gini of above 30 percent and these were not the same. And three interviewers had a Gini that was below 10 percent, two interviewers in 2003 and a different interviewer in 2004. The overall implication that interviewers not no meaningful contribution towards the observed low inequality holds true unless one is to argue that all interviewers through out the country are colluding, a very improbable proposition. Colluding is not practical and physically impossible for interviewers who are geographically isolated and dispersed through out the country.

3. Potential Sources of Low Inequality —Hypotheses

The observed low inequality in the HIES data could be indicative of various underlying conditions. Some of these possible conditions and associated hypotheses are presented below. In the following sections, we test the hypotheses that are testable with the available information in the HIES data. It is anticipated that these will likely help us judge the quality of the HIES data for using to analyze poverty in Azerbaijan as well as identify the potential sources of low inequality in the data.

First, the reason for low inequality in Azerbaijan could be that inequality is in fact low and the low Gini is just reflecting the underlying reality in the country. For anyone with the understanding of the economic structure of Azerbaijan and regional distribution of welfare, such explanation is implausible. The observed Gini coefficients are too low to be credible for Azerbaijan. It is implausible to think that Azerbaijan is the most equal country in the world, and Baku the most equal city. However, certain government and donor programs for the poor and vulnerable households could provide, albeit not sufficient, explanation for the apparently low inequality in the Azeri data. This led us to our first hypothesis.

- **Hypothesis 1:** The government's social protection programs, the involvement of donors in support of these programs, and family support networks have played an important role in reducing welfare disparity among the population. Well targeted social transfers to the poor and the vulnerable groups, such as the displaced population, lifted the consumption levels of these groups without affecting the welfare of the people in the upper

---

[2] For more details, see Ersado (2006).

consumption brackets. Evidence abound in the HIES data that transfers from government in the form of child allowance, pensions and other social assistance programs are important components of household consumption (see next section). Transfers from family members abroad and within the country are also important sources of consumption in Azerbaijan. International donor supports are significant, particularly for the displaced population. The outcome of such measures would undoubtedly lead to reduction in inequality, while at the same time alleviating poverty. It is unlikely this explains away all the observed extremely low inequality in Azerbaijan, but it could be a contributing factor.

Secondly, the data generating mechanism may be systematically biased and does not represent well the various segments of the Azeri population. This scenario gives rise to the following three hypotheses:

- **Hypothesis 2:** The inequality in the data is suppressed due to the fact that the HIES data have fewer representatives of households in the upper income echelon. This situation could arise, for instance, if some households in the richest category, decline participation in the survey. From communications with Azeri officials with the knowledge of the HIES data, there is high likelihood and clear evidence that the rich households tend to decline responding to survey questionnaires.

- **Hypothesis 3:** Households in the lower income echelon are inadequately represented in the data. Several factors could lead to survey data unrepresentative of the poorest households such as their location. The poorest households tend to live in inaccessible locations and in areas with poor infrastructures such as roads. In such conditions, lack of or poor supervision of the interviewers and data collection process, may lead to replacement of the poor households with better off ones in more accessible location or to many missing households in the data, most of which are poor. Also it is not implausible to think the poor may decline participation in the survey.

- **Hypothesis 4:** The data do not fully represent both the upper and lower income echelons. In the scenario where households in the upper and lower welfare echelon are "truncated", the resulting data would have unnaturally high homogeneity and therefore low inequality.

Thirdly, various forms of transfers and deficiency in the data generating mechanism could lead to the observed low inequality:

- **Hypothesis 5:** Combination of factors stated as conditions for hypotheses 1 through 4 are responsible for extremely low inequality in Azerbaijan.

Finally respondents, although they participated in the survey, were less forthcoming on the questions about their wealth. This is a common problem in most household surveys through out the developing world; Azerbaijan is not an exception. This led us to our final hypothesis:

- **Hypothesis 6:** Respondents, especially those participating households in the upper income echelon, were concealing their incomes and expenditures, thus depressing income and consumption disparity in the data.

4. Hypotheses Tests and Results

Hypotheses 1 through 5 are testable using the available information in the surveys. The last hypothesis does not yield itself to statistical testing. The null hypothesis for each of the testable hypotheses is that *measures to correct for factors hypothesized to lower inequality will not lead to significant increase in inequality*. The alternative is *these measures will significantly increase inequality in the Azerbaijan data*. As there is no established lower or upper limit to the level of inequality, a *significant* level of increase would be one that makes the Gini coefficient "reasonable" for Azerbaijan and comparable to that of other countries in the region and with similar economic structure.

Hypothesis 1

We test **hypothesis 1** by calculating consumption inequality with and without various forms of transfers such as social transfers (including pensions, state benefits and certain privileges), and transfers from family and friends abroad and within country. As can be seen in Table 5, over 90 percent of all households report receiving some kind of transfer income. Over half of urban and over 55 percent of rural households were reported to receive pension income. Transfers from friends and family members reached about 73 percent of Azerbaijan households, with nearly 80 percent of rural households reporting receipt of transfer income from friends and family members. Transfers also constituted a large share of household income. For example, in 2004 total transfer income accounted for a non-trivial 29 percent of total household income per capita nationwide.

Table 5. Most Azeri households report receipt of transfer income in 2004

	Azerbaijan	*Baku*	*Non-Baku Urban*	*Rural*
	Share of households receiving transfers			
All transfer	91.9	84.3	91.3	96.3
Social transfer	78.4	66.8	77.7	84.9
of which pensions	52.7	45.6	54.8	55.2
Transfers from family/friends	73.0	64.1	69.2	79.9
	Share of transfers in the household income			
Total transfer	29.1	23.7	29.0	32.5
Social transfer	13.3	11.2	13.0	14.8
of which pensions	10.9	9.5	11.0	11.7
Transfers from family/friends	15.8	12.5	16.1	17.7

Source: Azerbaijan 2004 HIES data.

Inequality increases substantially after subtracting transfers from household per capita consumption. Table 6 presents Gini coefficients without transfers over time and by area of residence. This holds true nationally as well as at regional levels. Nationally the inequality increases by over 73 percent in 2002, by over 81 percent in 2003 and by over 94 percent in 2004 if no transfers were available for consumption (Table 6). If only social transfers were subtracted from consumption, the Gini coefficient increases by about 50, 53 and 67 percent in 2002, 2003 and 2004, respectively. The findings suggest that social transfers had welfare equalizing effect, which is reflected in lower inequality with transfers than without transfers. There are evidences

elsewhere that a sharp increase in the level of government cash transfers to individuals can lead to decline in inequality (Keane and Prasad, 2001). Keane and Prasad (2001) show that despite growing labor earnings inequality following transition to market economy, the rise in transfer expenditures in Poland in the early years of transition mitigated the increase in overall income inequality. Transfers from families and friends have equally important effect on inequality. For instance, without transfers from family members and friends networks, the Gini coefficient would increase by 59, 67 and 67 percent respectively in 2002, 2003 and 2004.

There is important policy implication from the findings under hypothesis 1. The fact that inequality increases significantly without transfers implies that transfer incomes (particularly those from government sources) are reasonably targeted well. As equalizing effect of increased from 2002 to 2004 further suggests that the efficiency of targeting improved over time. As a result, the incomes of households in the lower wealth echelons were lifted more so than the incomes of the rich.

Table 6. Transfers have welfare equalizing effect

Gini coefficient	2002			2003			2004		
	without all transfers	without social transfers	without transfers from family and friends	without all transfers	without social transfers	without transfers from family and friends	without all transfers	without social transfers	without transfers from family friends
Azerbaijan	31.8	27.5	28.1	32.9	27.9	29	31.8	27.2	27.2
Baku	33.4	30.3	30.8	33.8	29.1	30.8	29.5	27.5	25.7
Non-Baku urban	31.6	26.6	27.5	34.2	28.7	29.8	31.2	27.2	26.8
Rural	30.5	25.9	26.7	31	26.1	27.1	31.8	25.8	26.8

Source: Azerbaijan HIES data.

Table 7. Percent increase in the Gini coefficients after transfers are removed

Gini coefficient	2002			2003			2004		
	without all transfers	without social transfers	without transfers from family and friends	without all transfers	without social transfers	without transfers from family and friends	without all transfers	without social transfers	without transfers from family friends
Azerbaijan	73	50	53	81	53	59	95	67	67
Baku	56	52	44	69	45	72	65	38	43
Non-Baku urban	82	44	75	85	55	90	99	47	71
Rural	81	55	79	86	56	82	114	55	80

Source: Azerbaijan HIES data.

Possible Reasons for Missing Observations

Testing hypotheses 2, 3 and 4 are relatively more complicated than hypothesis 1. It requires knowledge of the number of missing observations, preferably at the lower level of geographic aggregation. Fortunately the sampling framework for Azerbaijan HIES established identical

sample size for all primary sampling units (PSUs). In all surveys, an interviewer is assigned to each PSU and required to collect data from 104 sampled households per year. As pointed out in the introduction, there are 85 interviewers throughout Azerbaijan, leading to a total 8840 observations if all selected households were interviewed and their data retained. The fact that we have interviewer identification numbers in the data allows us to count the number of missing households (non responding households or households deemed to have unusable data) at the interviewer (i.e., PSU) level:

$$Nmis_p = 104 - Nobs_p$$

Where $Nmis_p$ denotes the number of missing households at PSU p, and $Nobs_p$ denotes the number of usable observations (households) available at PSU p. If the part of the population that is left out is different than the part that is observed, there will be differences between the survey results and what is actually true in the population. Table 8 below shows the distribution of missing households at the regional level. The percentage of missing households is higher in urban areas than rural. There were more than twice as many missing households in 2002 than either 2003 or 2004, suggesting some improvement in response rate over time.

Table 8. Distribution of missing households in Azerbaijan HIES data

|	2002			2003			2004		
	Total Sample	Missing	% Missing	Total Sample	Missing	% Missing	Total Sample	Missing	% Missing
Azerbaijan	8840	683	7.7	8840	316	3.6	8736	311	3.6
Baku	2392	200	8.4	2392	159	6.6	2392	133	5.6
Non-Baku urban	2808	250	8.9	2704	84	3.1	2288	91	4.0
Rural	3640	233	6.4	3744	73	1.9	4160	87	2.1
Nakhchivan	520	49	9.4	520	23	4.4	520	10	1.9
Absheron-Guba	1040	152	14.6	1040	52	5.0	1040	53	5.1
Mughan_Salyan	728	52	7.1	728	10	1.4	728	12	1.6
Ganja-Gazakh	1144	63	5.5	1144	24	2.1	1144	32	2.8
Sheki-Zagatala	624	37	5.9	624	6	1.0	624	20	3.2
Lankaran_Astara	832	36	4.3	832	17	2.0	832	12	1.4
Shirvan	624	20	3.2	624	8	1.3	624	16	2.6
Qarabagh-Mil	1144	74	6.5	1144	17	1.5	1144	23	2.0

Source: Azerbaijan HIES data.

Non-response or missing observations at the household level occur when a household or person selected for the survey does not participate in the survey or does participate but does not provide complete information. There can be many reasons for missing observations or non-response in survey data. A listed housing unit chosen for the sample may be found unoccupied when an interviewer visits the housing unit. Even if it is occupied, several adverse events may prevent data collection. Although a housing unit is occupied, its residents may be away from home during the entire survey period. Health conditions, whether permanent, such as hearing impairment or blindness, or temporary, such as an episode of a severe acute illness, may preclude an individual from responding as well. More importantly, a non-response also may occur when a person at home during the survey refuses to participate as an individual or as a representative of the entire unit.

From communications with Azeri officials with the knowledge of the HIES data, the predominant reason for non-response was refusal of the rich households to participate in the

surveys. There is clear evidence that the rich households tend to decline responding to survey questions. Azerbaijan State Statistical Committee (SSC) informed us that most other causes of missing observations have been addressed by replacing originally selected households that could not be interviewed at the time of interview, through a random draw from the census list. A cursory examination of the proportion of the missing observations and income shows positive association. A two-way scatter plots of the share of missing households and mean consumption expenditures at the economic regions level show positive correlation between wealth and refusal to participate in the survey in 2004 (see Figure 2). There are other empirical evidences in the literature that high-income households are less likely to participate because of a high opportunity cost of their time or concerns about intrusion in their affairs (Atkinson, 1987; Korinek, et al. 2005). Groves and Couper (1998) found that higher socio-economic status tended to be associated with lower response rate in the US surveys.

Figure 2. Rate of refusal is higher in the richer regions
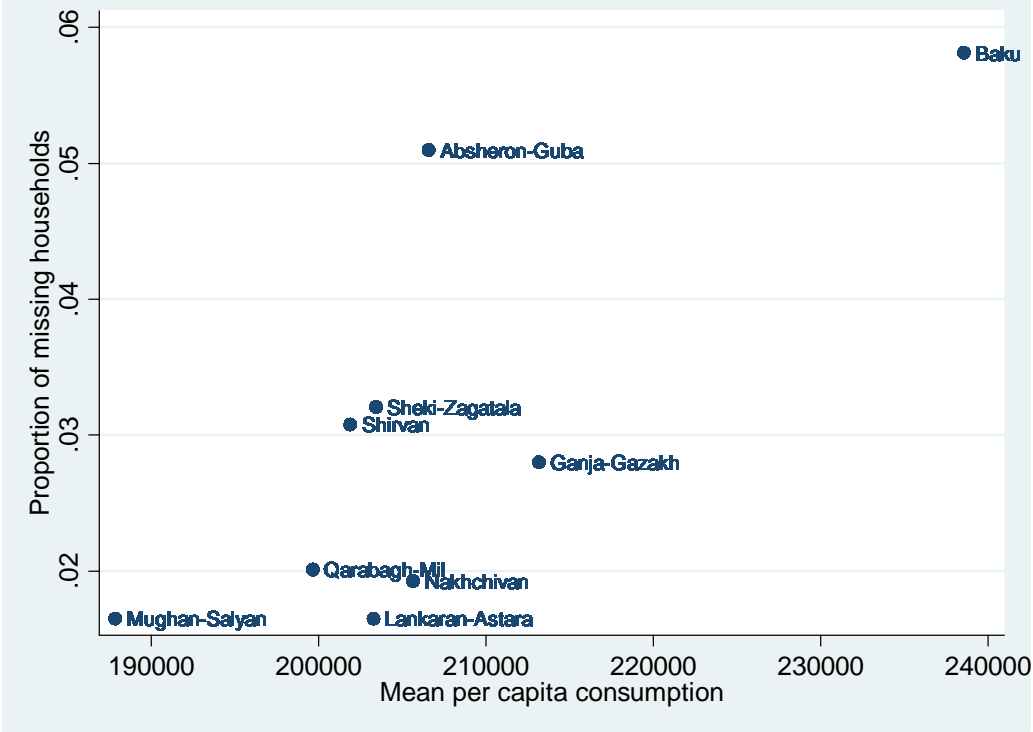
Source: Azerbaijan HIES, 2004.

With this backdrop, for the purpose of hypothesis 2, we assume that all missing households came from the upper wealth echelon due to their unwillingness to participate in the survey. The assumption that all missing households are due to either lack of participation of the rich is quite plausible for reasons stated in the last paragraph. But we also make an opposite assumption that all missing households came from the bottom wealth echelon for reasons stated under hypothesis 3. Moreover, we embrace the possibility that the missing households came from both upper and lower end of the wealth ladder. This scenario is explored for testing hypothesis 4.

11

Hypotheses 2 and 3

Once the missing observations are identified, the next step is how to account for them. In the statistical literature, there are a number of alternative ways of dealing with missing data. If data are missing at random, missingness is ignorable. However, if data are not missing at random, but are missing as a function of some other variable, a complete treatment of missing data would have to include a model that accounts for missing data. There are several ad hoc as well as purposeful models for treating missing observations. By far the most common approach is to simply omit those cases with missing data and to run our analyses on what remains. Some ad hoc approaches to replace missing data include carrying last observation forward, replacing observations by the mean of the variable and mean imputation using regression.

In our specific case, (i.e., for replacing households in the top wealth bracket), we avoid ad hoc approaches such as replacing the missing observations on income or consumption with the highest value in the data. Instead we use a more insightful approach by making some distributional assumptions. Let's assume that wealth, particularly in the top part of the distribution, has a *Pareto distribution* (see Box 1).

Box 1: Pareto Distribution

The Pareto distribution is particularly useful to describe the allocation of wealth among individuals since it shows well the way that a larger portion of the wealth of any society is owned by a smaller percentage of the people in that society. It can be seen from the probability density function (PDF) graph below, that the "probability" or fraction of the population f(x) that owns a small amount of wealth per person (x) is rather high, and then decreases steadily as wealth increases.

Figure 3. Example of Pareto probability density function
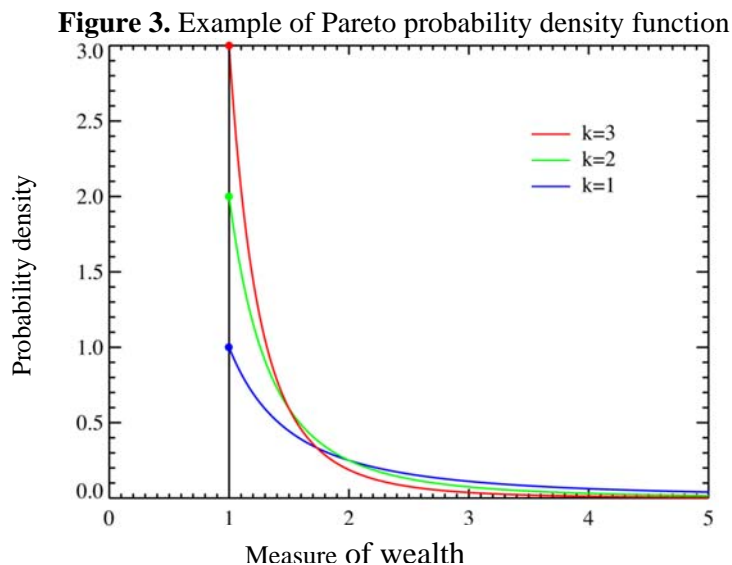
Pareto probability density functions ($f(x) = \dfrac{kx_m^{\,k}}{x^{k+1}}$ $x \geq x_m$ and $k > 0$) for various values of k (shape parameter) with x_m (location parameter) = 1

Source: *Wikipedia 2006 (http://en.wikipedia.org/wiki/Pareto_distribution).*

The values of consumption aggregates for households who did not participate in the survey were simulated using the *Pareto distribution* assumption. Let x_m be equal to the highest consumption per capita in the data (i.e., the consumption per capita for missing households is equal to or greater than the highest consumption per capita in the existing HIES data). Also assume that k (shape parameter) is equal to 2. We test hypothesis 2 by calculating consumption inequality after accounting for upper truncation in the distribution as described above. Similarly Hypothesis 3 can be tested by calculating consumption inequality after accounting for lower truncation in the distribution. For hypothesis 3, consumption per capita for missing households, which are now assumed to be the poorest, was assumed to be equal to or less than the lowest consumption per capita in the existing HIES data.

The likelihood that exclusion of households in the lower wealth echelon had led to lower inequality is very low. Table 9 presents the Gini coefficients and Figure 3 presents percent changes in them under hypotheses 2 and 3. The Gini coefficient does not exhibit appreciable increase with the assumption that the missing households were in the bottom wealth distribution. On other hand, the assumption of upper truncation (i.e., the missing households belong to the wealthiest upper echelon) appears to contribute to apparent low inequality. For example, accounting for exclusion of rich households in the data will lead to over 120 percent increase in Gini coefficient in 2002 (from 0.184 to 0.412). The impact of upper truncation is even higher in non-Baku urban areas, where the Gini coefficient would increase by close 150 percent (0. 172 to 0.432) in 2002 (Figure 4). The Gini coefficients would also increase appreciably in 2003 and 2004 after accounting for the missing households. Note that the effect of lower truncation is extremely low compared to the assumption of upper truncation.

Therefore, the most likely factor for the observed low inequality in the Azerbaijan HIES data appears to be the under-representation of households in the higher income brackets. This is likely so due to inadequate participation of the richer households in the surveys. However, the inequality figures after accounting for top truncation are still too low to be credible for Azerbaijan, particularly for 2003 and 2004. This suggests the most credible source of lower inequality may be the combination of both lack of participation of rich households and the transfers. This possibility is tested under hypothesis 5 below.

Table 9. Lack of participation of rich households partly explains low inequality in Azerbaijan

Gini Coefficients	2002		2003		2004	
	Top Truncation	Bottom Truncation	Top Truncation	Bottom Truncation	Top Truncation	Bottom Truncation
Azerbaijan	41.2	18.4	28	20.4	32.4	18.4
Baku	43.2	23.2	34.8	23.9	37.7	20.9
Non-Baku urban	43.3	26.5	26.9	20.3	33.3	17.9
Rural	38.2	22.9	22.8	18.1	26.6	16.5

Source: Azerbaijan HIES data.

Figure 4. Top truncation is the more likely source of low inequality
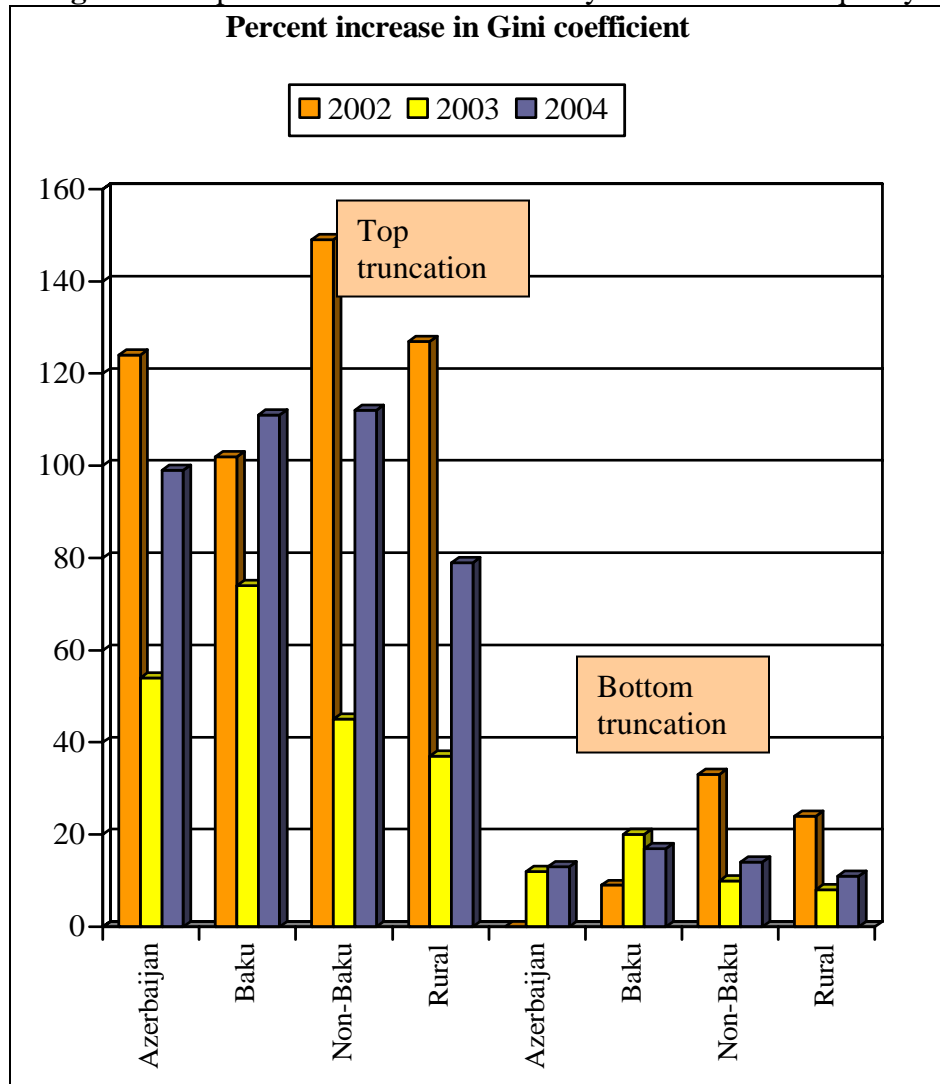
Source: Azerbaijan HIES data.

Hypothesis 4

The Gini coefficients under hypothesis 4 are not large enough for Azerbaijan, compared to its comparators in Table 1, particularly for 2003 and 2004, to warrant that the missing households were from both the top and the bottom of the wealth distribution. Under hypothesis 4 we assumed that there was truncation both in the top and bottom wealth groups. Here we substituted the consumption per capita of half of the missing households following hypothesis 2 and the remaining half were replaced following hypothesis 3. The resulting Gini coefficients are reported in Table 10. The findings thus reinforce the conclusions under hypotheses 2 and 3 that the likely source of low inequality is upper truncation rather than lower truncation or mixed truncation.

Table 10. There is little evidence of exclusion of poor households

	Gini Coefficients		
	2002	**2003**	**2004**
Azerbaijan	36.6	24.8	26.8
Baku	36.3	27.9	27.5
Non-Baku urban	38.0	25.5	28.4
Rural	35.8	22.2	24.9

Source: Azerbaijan HIES data.

Hypothesis 5

Under the assumptions for hypothesis 5, the Gini coefficient for Azerbaijan increase more appreciably than under any other hypotheses. Here we made the assumption that the extremely low inequality in Azerbaijan HIES data arose from lack of and/or inadequate participation of rich households and significant amount of transfers among Azeri households. Table 11 presents the Gini coefficients after removing transfers and replacing the consumption per capita for missing households following the Pareto distribution assumption under hypotheses 2 and 3. After accounting for upper truncation and transfers, the Gini coefficient more than double in most cases (see Table 12). After removing all transfer incomes and accounting for refusal of rich households, the Gini coefficient in 2002 increases to over 200 percent for Azerbaijan as the whole. Inequalities in 2003 and 2004 similarly climb by about 145 and 194 percent respectively.

The overall findings here suggest that the main culprit for extremely low inequality in the Azerbaijan HIES data is not only refusal of rich households to participate in the survey, but also significant infusion of transfers to the Azeri household consumption. Transfers from both government and private sources are important sources of income. Furthermore the transfers disproportionately benefited those households in the lower wealth category than those in the upper income groups.

Table 11. The most likely reasons for low inequality are lack of participation of the rich and significant transfers from government and family networks

	2002			2003			2004		
	Gini Coefficients								
	without all transfers	without social transfers	without transfers from family and friends	without all transfers	without social transfers	without transfers from family and friends	without all transfers	without social transfers	without transfers from family and friends
Azerbaijan	55.8	46.0	49.4	44.5	33.3	37.4	47.9	38.0	40.3
Baku	57.5	47.7	51.4	51.3	39.8	44.9	50.7	42.9	44.0
Non-Baku urban	58.0	48.0	51.8	45.7	32.3	37.8	49.7	39.0	41.7
Rural	52.8	43.2	46.2	37.4	27.8	30.7	42.6	32.2	35.1

Source: Azerbaijan HIES data.

Table 12. Percent increase in the Gini coefficients after transfers are removed

	2002			2003			2004		
	without all transfers	without social transfers	without transfers from family and friends	without all transfers	without social transfers	without transfers from family and friends	without all transfers	without social transfers	without transfers from family and friends
Azerbaijan	203	150	168	145	83	106	194	133	147
Baku	169	139	140	157	99	151	183	114	146
Non-Baku urban	233	159	230	147	75	141	216	111	165
Rural	214	159	210	124	67	106	186	93	136

Source: Azerbaijan HIES data.

5. Conclusions

The Azerbaijan Household Income and Expenditure Survey (HIES) data show extremely low inequality measures, which wrongly suggest that Azerbaijan is one of the most or perhaps the most equal country in the world. Such oddly low inequality figures have cast serious doubts on the quality and the reliability of the HIES data for poverty analysis. This paper analyzed the HIES data to find out the main culprits for low inequality. The results of this analysis are summarized in Table 13.

The study finds two main sources of low inequality in the data. First, Azerbaijan HIES data are somewhat unrepresentative of living conditions of the population because of the fact that the richer households are disproportionately less willing to participate in the surveys. While this would have no impact on the aggregate number of poor households, it would affect the FGT poverty measures (i.e., headcount ratio, poverty gap ratio and poverty severity). The missing households add to the pool of the non-poor, so including them would lower the headcount poverty rates. Secondly, transfers have inequality reducing effect and there is significant amount of transfer incomes going into the hands of Azeri households. The finding suggest that households in the lower wealth echelon are benefiting more from transfers than those in the richer category, which in turn implies government and other transfer programs are relatively well targeted. Inequality in the HIES data would increase significantly—more than two-fold—if the effects of top truncation and transfers were accounted for (see Table 13).

Table 13. Inequality in Azerbaijan HIES data under various hypotheses

	2002	2003	2004
Base case	*18.4*	*18.2*	*16.3*
Hypothesis 1: transfer effect	31.8	32.9	31.8
Hypothesis 2: top truncation	41.2	28.0	32.4
Hypothesis 3: bottom truncation	18.4	20.4	18.4
Hypothesis 4: bottom and top truncations	36.6	24.8	26.8
Hypothesis 5: transfer and bottom truncation	36.1	35.3	31.9
Hypothesis 6: transfer and top truncation	55.8	44.5	47.9

Source: Azerbaijan 2002-2004 HIES data.

References

Atkinson, A.B.1987. "On the measurement of poverty," *Econometrica,* 55:749–764

Ersado, L. 2006. Azerbaijan Poverty Profile: 2002-2004. Human Development Unit, ECA Region, Washington, D.C

Korinek, A., J. A. Mistiaen and M. Ravallion. 2005. "Survey non-response and the distribution of income", *Journal of Economic Inequality* (forthcoming)

Groves, R.E. and Couper, M.P. 1998. *Non-response in Household Interview Surveys*, Wiley, New York

Milanovic, B. 2005. *Worlds Apart: Measuring International and Global Inequality*, Princeton, NJ: Princeton University Press

Wikipedia. 2006. http://en.wikipedia.org/wiki/Pareto_distribution

Michael P. Keane and Eswar S. Prasad. 2001. "Poland: Inequality, Transfers, and Growth in Transition", Finance and Development , Vol. March 2001, 38 (march)