Top incomes and the measurement of inequality in Egypt

Vladimir Hlasny

Paolo Verme

Top incomes and the measurement of inequality in Egypt[*]

Vladimir Hlasny

Ewha Womans University, Seoul, Korea

Paolo Verme

World Bank

Abstract

By all accounts, income inequality in Egypt is low and had been declining during the decade that preceded the 2011 revolution. As the Egyptian revolution was partly motivated by claims of social injustice and inequalities, this seems at odds with a low level of income inequality. Moreover, while income inequality shows a decline between 2000 and 2009, the World Values Surveys indicate that the aversion to inequality has significantly increased during the same period and for all social groups. This paper utilizes a range of recently developed statistical techniques to assess the true value of income inequality in the presence of a range of possible measurement issues related to top incomes, including item and unit non-response, outliers and extreme observations, and atypical top income distributions. The analysis finds that correcting for unit non-response significantly increases the estimate of inequality by just over 1 percentage point, that the Egyptian distribution of top incomes follows rather closely the Pareto distribution, and that the inverted Pareto coefficient is located around median values when compared with 418 household surveys worldwide. Hence, income inequality in Egypt is confirmed to be low while the distribution of top incomes is not atypical compared with what Pareto had predicted and compared with other countries in the world. This would suggest that the increased frustration with income inequality voiced by Egyptians and measured by the World Values Surveys is driven by factors other than income inequality.

Keywords: Top incomes, inequality measures, survey nonresponse, Pareto distribution, parametric estimation, Egypt.

JEL Classification: D31, D63, N35.

Contents

1. Introduction

A recent study of inequality in Egypt (World Bank 2013) has shown that there is an important discrepancy between income inequality as measured by household expenditure surveys and the perception of income inequality as reported by people in values surveys. This is no small issue given that part of the frustration voiced by the people of Egypt and culminated with the Egyptian revolution in 2011 has been explained in terms of inequality and social injustice. The Egyptian Center for Economic Studies (ECES) for example, in a note shortly after the revolution argued that *"Social inequality and inadequate human development coupled with the lack of political reforms have been among the main factors that led to the outbreak of the revolution."* (p. 7, ECES, Policy Viewpoint, May, 2011). The World Bank (2013) study also shows that the aversion to income inequality as measured by the World Values Surveys in 2000 and 2008 has significantly increased for all social groups, which is at odds with the apparently declining values of income inequality during the same period.

This discrepancy may be explained by a variety of factors, including the various determinants of *feelings* of inequality such as expectations about the future, or factors related to the measurement of the *facts* of inequality such as whether household surveys are able to capture incomes well. The World Bank (2013) study provided some initial leads on what could constitute an explanation of such a paradox and one of these leads concerned the measurement of top incomes. It is rather well known that household surveys are not particularly accurate at measuring top incomes because richer households tend to either underreport income or expenditure, or are less likely to participate in household surveys altogether. When this happens, measures based on incomes such as the Gini index for income inequality are biased and do not reflect the actual extent of inequality in a country. However, beyond anecdotal evidence, there is little research that has shown convincingly that household surveys worldwide underreport income inequality while there is no evidence as yet of this being the case in Egypt.

Studying the relation between top incomes and inequality using statistical techniques is important not just for statistical reasons. It has been observed, for example, that GDP growth in national accounts statistics is at odds with the growth of household incomes inferred from the Egyptian Household Income, Expenditure and Consumption Survey (HIECS). While GDP growth has shown a consistent cumulative growth over the period 2000-2009, household incomes have shown a slight decline. This may be explained by the fact that growth occurred only among top income households and that these households are not well captured by household surveys. As shown by Atkinson et al. (2011), for example, despite strong GDP growth, US household income measured by tax records grew by only 1.2% on average between 1976 and 2007, and by only 0.6% if the top 1% of earners are excluded. The relative performance of top earners also has an impact on perceptions. In the same example for the US, Atkinson et al. (2011) noted that while the top 1% of incomes grew at similar rates during the Clinton (1993-2000) and Bush (2002-2007) administrations – around 10% per year – the bottom 99% had very different growth rates during the two administrations – of 2.7% and 1.3% respectively. According to these authors, this could explain why the public outcry over top incomes was much louder during the Bush years than during the Clinton years. Therefore, studying the relation between top incomes and inequality is important to better understand who benefits from GDP growth and how top incomes affect the measurement and perception of inequality, two issues that may help to understand the Egyptian revolution.

This paper focuses on top incomes in an effort to determine how top incomes affect the measurement of inequality in Egypt. We will attempt to investigate a number of well-known issues related to the measurement of inequality and top incomes including: 1) item non-response; 2) unit non-response; 3) the role of extreme observations and 4) the shape of the top income distribution. In doing so, we will draw on three separate bodies of literature that we join into a consistent framework. In our knowledge, this is the first time that these three bodies of literature are considered jointly.

The first body of literature we consider, on survey non-response, is vast and part of a long tradition in statistics and economics addressing questions related to item or unit non-response biases in household surveys. Deaton (2005), for example, has shown how unit non-response may well be one the factors that can explain the discrepancy between national accounts and household surveys when it comes to the measurement of household consumption. We will focus here only on a recent strand of this literature that provides guidance on how to assess and correct measures of inequality in the presence of income biases determined by household non-response (Korinek et al. 2006 and 2007). This literature essentially provides two instruments that could help our investigative effort on inequality in Egypt: an instrument to detect whether an income bias exists due to non-responses, and an instrument to correct for such a bias.

The previous paragraph suggests that top incomes may be systematically under-represented in surveys. On the other hand, a large body of literature (Cowell and Victoria-Feser, 1996, Cowell and Flachaire, 2007, and Davidson and Flachaire, 2007) has found that extreme values of income can greatly influence the measurement of inequality. This literature tests how extreme observations (such as top incomes) affect measures of poverty and inequality, and proposes a method for correcting such measures in the presence of biases induced by extreme observations. The correction is facilitated by a semi-parametric approach that combines a parametric method for the uppermost part of the income distribution and a classic non-parametric method using actual household data for the rest of the distribution. This approach has been shown to be effective in correcting sample distributions that do not capture top incomes precisely. Comparing the results from this exercise with those for the correction of non-response bias allows us to understand better the role of top incomes in household surveys.

The third body of literature is that on top incomes, summarized in a recent paper by Atkinson et al. (2011). This literature uses the Pareto distribution and the Pareto coefficients to study the distribution of top incomes across the world using tax records. We borrow from this literature and apply the same tools to household data instead of tax records. In our context, given our findings regarding the role of top incomes in the Egyptian data, studying the shape of the top income distribution in Egypt can provide some clues on whether this distribution departs from Pareto's law regarding top incomes, and whether the distribution is very different from those in other countries.

The Pareto distribution used in Atkinson et al. (2011) is one of the distributions suggested by the Cowell et al. literature mentioned above, while the Pareto and inverted Pareto parameters used by the Atkinson et al. (2011) literature can be evaluated along with the Gini index using the Korinek et al. methods, which may correct these parameters for income biases caused by unit non-responses. In essence, these three bodies of literature can be nicely combined in a consistent framework to provide a robust assessment and correction of the measures of inequality estimated with household data. We will also be able to benchmark some of our results by comparing the Pareto parameters in Egypt with those estimated using a data set of 418 household surveys administered in 107 countries worldwide.

The paper is organized as follows. The next section discusses the key issues and literature that we use in the study. The following section outlines the main models and methods used in the empirical section. Section four briefly describes the data under analysis. Section five presents and discusses results. Section six compares Egyptian data with the rest of the world and section seven summarizes results and discusses policy implications.

2. Measurement issues

Our objective is to understand how top incomes affect the measurement of income inequality in Egypt. In this section we discuss the main issues to consider as pinpointed by studies focusing on top incomes and inequality. In the next section, we will outline some of the models that can help us in addressing these issues.

Sub-sample random extraction. In the particular case of Egypt, the national statistical agency provides to researchers 25% or 50% of the sample extracted randomly from the four quarterly independent subsamples of the full sample. As we know from sampling theory, random extraction is the best option for extracting a sub-sample in the absence of any information on the underlying population. However, only one sub-sample is extracted from the full sample and given to researchers and this implies that a particularly "unlucky" random extraction can potentially provide skewed estimates of the statistics of interest. This is a question that we will test with a simple Monte Carlo experiment later in the paper.

Data errors. Extreme observations in an income distribution can sometime be explained in terms of errors. Before any analysis with the available sample, it is worth checking whether extreme observations among top incomes are simply errors such as data input errors or they are plausible data particularly distant from the central moments of the distribution (outliers or extreme values). Statistical agencies are usually quite thorough on this issue and clear data of errors before providing the data to researchers. In our experience with the Central Agency for Public Mobilization and Statistics (CAPMAS), this is also the case in Egypt. We will, however, report top observations and briefly discuss this issue before carrying out any further analysis.

Item non-response. Item non-response occurs when households participating in the survey do not reply to an item of interest (income or expenditure in our case). Item non-responses may be related to households' particular factors such as wealth or education, and this may bias statistics based on the surveyed incomes or expenditures. The standard practice to address this problem is to impute the value of the item by predicting this value based on a number of socio-economic characteristics observed for households with the missing item. Alternative practices include assigning the mean or median values to the missing items using information for households responding to the item, or information from external sources. In our case, we do not have households that do not report any income or expenditure. It is possible that some of the components of income or expenditure are not reported but the data do not distinguish between sub-item non-response and sub-item nonexistence. For example, a household may report no income for rent but the interviewer may not be able to distinguish whether the household does not own rented properties or whether the household does not wish to reply to the question. This is a problem similar to underreporting (we have a part of income or expenditure that is supposedly not observed) and that will be treated as any other non-observed factor (in the error term).

Unit non-response. Unit non-response refers to households that were selected into the sample but did not participate in the survey. The reasons for non-participation can be many such as a change of address or non interest on the part of the household. Interviewers generally have lists of addresses that can be used to replace the missing household but this practice is not always sufficient to complete the survey with the full expected sample. Most of the available household survey data, particularly in developing countries, suffer from substantial unit non-response. For some surveys, the reason for non-response is recorded and sometimes this reason is used to correct the weights when the survey is completed if households have not been replaced. In the case of Egypt, we did not have any information at our disposal concerning the reason for non-response while we have about 4% of the sampled households that did not participate in the 2009 survey.

Unit non-response may or may not affect the statistics of interest. We therefore need to understand first whether unit non-response affects income inequality. If this is the case, we can attempt to correct the bias so as to obtain more accurate statistics. Korinek et al. (2006, 2007) have developed a method to estimate whether unit non-response affects the measurement of inequality and also a method to correct for such a bias if it exists. In this paper, we will follow these methods to address these issues as discussed in the next section.

Top incomes distribution. Vilfredo Pareto introduced long ago the notion that the top observations in an income series follow a particular distribution and pattern represented respectively by the Pareto distribution and by the Pareto coefficient. More recently, Piketty and a number of other authors have used these tools in conjunction with tax records to study top incomes across countries and across time, a literature neatly summarized in Atkinson et al. (2011). In this paper, we will follow this literature to study the shape of the top income distribution in Egypt and use the Pareto measures in three different contexts. First, we will apply the Korinek et al. (2007) approach to the Pareto coefficients and identify how these are affected by unit non-response and its correction. Second, we will use the parametric properties of the Pareto distribution to evaluate how representative are the top income observations in our sample to the underlying income distribution. And third, we will use the Pareto and inverted Pareto coefficients to compare the top income distribution in Egypt with those in the rest of the world using a unique database of 418 household budget surveys administered in 107 countries.

Extreme values and inequality. How sensitive is the Gini to extreme values? Cowell and Victoria-Feser (1996), and Cowell and Flachaire (2007) have shown that, unlike poverty measures, inequality measures are very sensitive to extreme observations, to the extent that even a single observation can significantly affect the measurement of inequality. What constitutes extreme observations is a matter of judgment of course. Neri et al. (2009), for example, define outliers as observations exceeding the median 4-5 times or more. Working with the EU Surveys on Income and Living Conditions (EU-SILC), they find that this typically comprises 0.1-0.2% of households. Cowell and Flachaire (2007) and Davidson and Flachaire (2007) define extreme values as those values that can significantly change the value of inequality, and propose a methodology to test and address the problem. In this paper, we will use this methodology to evaluate the role of extreme values for the measurement of inequality with our data.

The same literature also shows that the choice of the measure of inequality and the choice of method to estimate income inequality are very important. Measures of income inequality are many and some of these measures are more sensitive to extreme values than others. The Gini index, for example, is known to

give more weight to central observations in a distribution and consequently discounts observations in the tails. Cowell and Victoria-Feser (1996) have found that the Gini index is more robust to contamination of extreme values than two members of the generalized entropy family, a finding later confirmed by Cowell and Flachaire (2007). For these reasons and throughout the paper, we will use only the Gini index as a measure of inequality while we would expect measures of the generalized entropy family to exhibit sharper sensitivity to extreme income observations.

Cowell and Victoria-Feser (1996) and Cowell and Flachaire (2007) have also shown that even the Gini index can be consistently underestimated with household surveys that cannot capture top observations precisely. These authors concord in finding that inequality estimates imputed from a parametric distribution function are less sensitive to extreme observations than non-parametric observations from actual household data, and suggest combining parametric Pareto estimates for the top of the distribution with non-parametric statistics for the rest of the distribution. This approach complements the Korinek et al. (2009) method for correcting for unit non-response of high-income households and overlaps with the Atkinson et al. (2011) method to model top incomes. We will use this method to correct the Gini coefficient for the potential influence of top observations, so as to compare the results with non-corrected Ginis or Ginis corrected for other statistical issues. This will allow us to comment on the relative influence of extreme observations and other statistical issues in our data.

3. Models

Unit non-response

To test for the presence of a systematic non-response bias properly, we can use a formal model to estimate the relationship between household income and its probability of response. Unfortunately, unlike in the case of item non-response, we cannot simply infer households' unreported income from their other reported characteristics, because we don't observe any information for the non-responding households. Assigning the mean or median values to the missing items would be inappropriate, as the missing values may be systematically very different from the rest of the distribution. However, following a technique developed by Korinek et al. (2006 and 2007), we can still use information about household-response rates at a higher level of geographic aggregation to infer the propensity of households with different characteristics, such as different incomes, to participate in the survey. This approach essentially takes advantage of the variation in household response rates and the variation in the distribution of observable variables (income or expenditure per capita) across geographical areas. We estimate the response probability for households as a function of their characteristics by observing the propensity of households with similar characteristics across all regions to participate in the survey, and by fitting regional population imputed from the participating households' response probabilities to the regions' actual population.

We assume that the probability of a household i to respond to the survey, P_i, is a logistic function of its arguments (Korinek et al. 2006, 2007):

$$P_i(x_i, \theta) = \frac{e^{g(x_i, \theta)}}{1 + e^{f(x_i, \theta)}}, \qquad (1)$$

where $g(x_i,\theta)$ is a stable function of x_i, the observable characteristics of responding households i that are used in estimations, and of θ, the corresponding vector of parameters from a compact parameter space. Variable-specific subscripts are omitted for conciseness. $g(x_i,\theta)$ is assumed to be twice continuously differentiable in θ. The parameters θ can be estimated by fitting the estimated and actual number of households in each region using the generalized method of moments (GMM) estimator

$$\hat{\theta} = \arg\min_{\theta} \sum_{j} \left[(\hat{m}_j - m_j) w_j^{-1} (\hat{m}_j - m_j) \right]. \tag{2}$$

Here m_j is the reported number of households in region j, $\hat{m}_j$ is the estimated true number of households in the region, and w_j is a region-specific analytical weight proportional to m_j. The estimated number of households, $\hat{m}_j$, can be imputed as the inverse of the estimated response probability of responding households in the region, $\hat{P}_{ij}$, summed over all N_j households. If the sample is extracted from a larger population, the imputed true number of households should be divided by the sampling rate for the underlying population in each region, s_j, to obtain population estimates. Finally, if the available sample includes only a fraction of the households responding to the full survey – such as the 25% random extraction from the full HIECS sample – we should divide by the sub-sampling rate for each region, ss_j:

$$\hat{m}_j = s_j^{-1} ss_j^{-1} \sum_{i=1}^{N_j} \hat{P}_{ij}^{-1}. \tag{3}$$

Under the assumptions of random sampling within and across regions, representativeness of the sample for the underlying population in each region, and stable functional form of $g(x_i,\theta)$ for all households, the estimator $\hat{\theta}$ is consistent for the true θ. Estimated values of $\hat{\theta}$ that are significantly different from zero would serve as an indication of a systematic non-response bias. In that case, we can use the imputed household response probabilities to correct for the bias. In the absence of any information about non-responding households, we have two options for correcting the bias: imputing the income of non-responding households, or re-weighting households that responded to the survey according to their inferred probability of response. Under the first option, estimation of the expected value of income for non-responding households would entail integrating incomes weighted by the corresponding probabilities of non-response across all possible incomes. With the imputed incomes of non-responding households, we would obtain the full income distribution on which to estimate measures of inequality. The problem with this method is that the results are sensitive to our assumption regarding the domain of incomes, and representativeness of the estimated income–probability relationship to counterfactual income levels. The second option entails imputing the true distribution of incomes by correcting the mass of each observation for its probability of being sampled. In this study we take the latter approach. Inverses of the estimated response probabilities serve as the appropriate household weights. In the income distribution imputed in this way, the derived measures of inequality converge to their true values as long as our sample is representative of the underlying population.

The model presented in equations 1-3 above uses *within-j* information as well as *between-j* information. It uses within-*j* information because the estimated number of households $\hat{m}_j$ is estimated within-*j* and it uses between-*j* information because the number of households observed within-*j* and the distribution of explanatory variables vary across *j*s. The choice of geographic disaggregation involves a trade-off between the number of *j* data points, and the number and distribution of within-*j* observations vis-à-vis the underlying population. On the one hand, observations should be behaviorally similar to non-responding households within-*j*, calling for smaller geographic units. On the other hand, Equation 3 requires that the sample encompass the entire range of values of relevant characteristics of the underlying population, potentially calling for larger geographic units. In this paper we opt to use 2,526 Primary Sampling Units (PSU) as *j* regions with an average of 18.6 responding households per region, as compared to the 51 US states with an average of 1,649 households per state used by Korinek et al. (2006 and 2007). These are clearly two different approaches with different implications.

In our case, the primary sampling units have relatively homogeneous households, with similar behavioral responses and presumably also similar survey-response probabilities. Because of a high response rate to the HIECS survey (96.3%), the observed range of household characteristics in each PSU is expected to comprise the values of the few non-responding households. A higher level of geographic aggregation would make behavioral responses less likely to be stable within *j* areas, while offering little additional assurance that values of characteristics of responding households encompass values of non-responding units. In our case, households' response probabilities are essentially inferred by comparing regions with similar, narrow ranges of explanatory variables. The response probability curve is constructed using 2,526 sets of probability estimates that are little overlapping on the curve. In Korinek et al. case, response probabilities are inferred by comparing fewer regions with greater ranges of explanatory variables. The response probability curve is constructed using 51 sets of probability estimates largely overlapping. In our case, the non-response bias correction is limited by the low observed non-response rate and by homogeneity of households in each PSU, which prevent the response probabilities to be estimated too low. In Korinek et al. case, response probabilities can be very low for some households, because other households in the same region can be assigned very high probabilities in compensation. This difference in methodologies is important because model errors are at the level of regions *j*. We think that our approach represents a more appropriate bias correction of the Gini coefficients in the HIECS data, that it is less likely to overshoot the correction, and that it is more consistent with the Pareto corrections illustrated in the next section. We will test these claims in the results section.

Extreme values

To evaluate the distribution of topic incomes and study the presence of extreme values in our data, we follow the approach pioneered by Pareto (1896) and recently rediscovered by the work of Piketty and others summarized in Atkinson et al. (2011). The Pareto distribution is a particular type of distribution which is skewed and heavy-tailed. It has been used to model various types of phenomena and it is thought to be suitable to model incomes, particularly upper incomes. The Pareto distribution can be described as follows:

$$F(x) = 1 - \frac{1}{x^\alpha} \, , \ \ 1 \leq x \leq \infty \, , \tag{4}$$

where α is a fixed parameter called the Pareto coefficient and x is the variable of interest, which in our case will be income or expenditure per capita. It follows that the probability density function can be described as:

$$f(x) = \frac{\alpha}{x^{\alpha+1}} , 1 \leq x \leq \infty .\qquad(5)$$

The probability density function has the properties of being decreasing, tending to zero as x tends to infinity and with a mode equal to 1. Intuitively, as income becomes larger, the number of observations declines following a law dictated by the constant parameter α. Clearly, this is not a distribution function that suits well all incomes under all income distributions, but should be thought as one possible alternative to model the right hand tail of a general income distribution, which is the focus of this paper.

In the application that follows, and for empirical purposes, we will use a slightly different definition of the Pareto coefficient (α) as well as the Inverted Pareto coefficient (β) as proposed in Atkinson et al. (2011):

$$\alpha = \frac{1}{1 - \left[\log(\frac{s10}{s1})/\log(10)\right]}\qquad(6)$$

$$\beta = \frac{\alpha}{\alpha - 1},\qquad(7)$$

where $s10$ and $s1$ represent the income shares of the top 10% and 1% of the population respectively. With tax records, it is generally more common to use the top 1% and 0.1% respectively but with household data, where samples are typically in the thousands of observations, the top 0.1% of households is a sample too small to be representative of the very top of the distribution as it may comprise extreme observations, hence the choice of the top 1% of the population.

The interpretation of the beta coefficient is that larger betas correspond to larger top income shares while the opposite is true for the alpha coefficient. In what follows, we will report both coefficients but, as a rule of thumb, the beta coefficient is what provides a snapshot indication of top incomes. Research on top incomes has shown that the alpha and beta coefficients are effectively stable for any income distribution, and in any given year and country, as originally predicted by Pareto. The work by Piketty and others, which used much longer time-spans than previous research, has shown that the beta coefficient can vary over time and that this variation can be explained by a combination of economic and political factors.

Measures of inequality can be influenced by the presence of even few observations with unusually high values. To evaluate the possible presence of extreme observations in our sample, and to evaluate the sensitivity of our Gini coefficients to these observations, we follow a procedure proposed by Cowell and Flachaire (2007) and Davidson and Flachaire (2007) to replace highest-income observations with values estimated under an expected distribution, and to combine the corresponding parametric inequality measure for these incomes with a non-parametric measure for lower incomes. As the afore-mentioned

literature has confirmed, top incomes appear to be distributed as under the Pareto distribution with an estimable coefficient α. Cowell and Flachaire (2007), propose the following formulation of α

$$\alpha = \frac{1}{k^{-1} \sum_{i=0}^{k-1} \log X_{(n-i)} - \log X_{(n-k+1)}}, \qquad (8)$$

where $X_{(j)}$ is the *j*th order statistic in the sample of incomes *n*, and *k* is the delineation of top incomes such as the top 10% of observations. We could also estimate α using maximum-likelihood methods to obtain the estimate with its robust standard error. All these methods allow weighting of observations by their sampling probability. These estimation methods yield results that are similar to the formulation proposed by Atkinson et al. (2011) in Equation 6 above. We will therefore use this formulation in the rest of the paper for consistency. The Gini coefficient under the estimated Pareto distribution for the *k* top-income households can be derived from the expression for the corresponding Lorenz curve (expression inside of the integral below) as

$$Gini = 1 - 2 \int_0^1 1 - [1 - F(x)]^{1-1/\alpha} \, dF(x) = \frac{1}{2\alpha - 1}. \qquad (9)$$

Finally, this parametric Gini coefficient can be combined with the non-parametric Gini coefficient for the *n-k* lower-income observations using geometric properties of the Lorenz curves as

$$Gini = (1 + Gini_k)\frac{k}{n} s_k - (1 - Gini_{n-k})\left(1 - \frac{k}{n}\right)(1 - s_k) + \left(1 - \frac{2k}{n}\right). \qquad (10)$$

Here s_k refers to the share of aggregate income held by the richest *k* percent of households. As long as it was correct to assume that top incomes in the population are Pareto-distributed, this semi-parametric Gini coefficient can be compared to an uncorrected non-parametric estimate for the observed income distribution. A difference between the semi-parametric and non-parametric estimates would indicate that some observed high incomes may have been generated by a statistical process other than Pareto, and that our inequality measure is sensitive to this. A semi-parametric Gini that is lower than the non-parametric Gini can be interpreted as evidence that some top incomes in the sample are 'extreme' compared to those predicted under the Pareto distribution. A higher semi-parametric Gini would indicate that the observed top incomes are lower than what the Pareto distribution would predict, potentially implying under-representation of high-income units in the sample.

The unit non-response and extreme-observations conjectures thus yield opposite predictions about the influence of top incomes on inequality measures, to the extent that they may even cancel each other out. The former conjecture is that the observed top incomes are valid for the measurement of inequality, and should be even used to stand for unobserved incomes of non-responding households. The latter conjecture is that the observed top incomes may have been generated by processes different from those in the underlying population, by error or by different accounting practices, and should be replaced by values imputed from the data generating process in the population.

To comment on the validity of these opposite predictions and evaluate their relative size, we can compare a set of four Gini coefficients: semi-parametric Gini accounting for the possibility of extreme observations but not for the non-response bias (i.e., Equation 10 where $Gini_k$ is derived from Equations 6 and 9 in an unweighted income distribution); semi-parametric Gini accounting for them both (Equation 10 where $Gini_k$ is derived from Equations 6 and 9 in an income distribution weighted as per Model 4); non-parametric Gini correcting only for the non-response bias (Gini observed in a Model 4-weighted income distribution); and the baseline uncorrected non-parametric Gini (observed in the unweighted income distribution). This comparison can inform us about the relative importance of extreme income observations versus non-response bias among high-income households and about their combined effect for the measurement of inequality in Egypt.

4. Data

This study relies on the Household Income, Expenditure and Consumption Survey (HIECS) administered by the Central Agency for Public Mobilization and Statistics (CAPMAS). The survey was conducted every five years until 2009 and is now implemented every two years. In this study, we use four rounds of the HIECS: 1999-2000, 2004-2005, 2008-2009 and 2010-2011 (2000, 2005, 2009 and 2011 for short). Survey samples comprise four quarterly independent subsamples that are nationally representative and stratified by governorate, and urban and rural substrata. The original full samples of the 2000, 2005 and 2009 surveys included 48,000 households, but starting from 2011 the survey includes a smaller sample of 16,000 households. All samples are stratified by governorate, and urban and rural substrata, and they are multi-stage random samples based on the most recent population censuses: the 1996 census for the 2000 and 2005 HIECS, and the 2006 census for the 2009 and 2011 HIECS. For a full description of the data and for a discussion of comparability issues over time see World Bank (2013).

The CAPMAS traditionally provided researchers with access to only 25% of observations in the HIECS. Since May 2013, however, the agency decided to grant researchers access to 50% of the data for selected years and posted these data on the internet. Extraction of the 25% or 50% subsamples is carried out randomly within each of the quarterly independent subsamples. For the purpose of this study, the CAPMAS has also granted exceptional access to 100% of the 2005 sample and allowed us to investigate 100% of the 2009 sample on site in Cairo. We therefore use 25% of the sample for 2000 (12,000 observations), 100% for the 2005 and 2009 samples (48,000 observations each) and 25% for 2011 (4,000 observations). From a methodological perspective, these samples and subsamples provide a complex combination of challenges for the measurement of inequality that, in our view, make Egypt a very good case study.

In this paper we put special emphasis on the 2009 HIECS. This is the only sample for which we have information on household response rates for all Primary Sampling Units (PSU), which is essential to implement some of the tests conducted in this paper. The 2009 sample is based on the last national census (November 21, 2006), and follows it most closely, which implies better accuracy of the sampling frame. To the extent that new residential developments may have arisen or people changed their residence since the latest census, this proximity in time minimizes any distortions or biases in sample coverage. Improvements to the sampling and methodology made by the CAPMAS between 2000 and 2009 as well as the fact that the 2010-2011 survey was carried out during the revolution make the 2009 survey the most

accurate of all surveys implemented by the agency to date. We will use the other rounds of the HIECS to carry out several additional tests and compare statistics over time.

The main welfare measure used in this paper is income per capita. It is common practice in developing countries to use consumption as a proxy of income rather than income itself given that income tends to be underreported and given that consumption is smoother than income, especially in rural areas. The World Bank (2013) report on income inequality in Egypt and our own work have shown that the income variable in the HIECS is actually good. The distribution of income is very similar in shape to that of consumption while the central moments of the distribution of income are higher than those of consumption. The difference between income and consumption (savings) is also an increasing linear function of income as one should expect. Therefore, while we will use also expenditure to compare our results for income, our preference in the Egyptian case is for the income variable. Income includes six main groups of items: wages and salaries, income from non-agricultural activities, cash transfers, income from agricultural activities, income from non-financial assets and income from financial assets.

5. Results

Data errors

Our inspection of the HIECS data, done by evaluating the distributions of income, expenditure and other socio-economic characteristics of households, did not reveal any likely data errors. Nevertheless, it is worth inspecting top values of income and expenditure for anomalies. We inspect the top observations of income and expenditure in each of the four years under analysis using either the top 100 observations or the top 1% of observations given that our samples are of different sizes (figure A1 in Annex). None of the samples show implausibly high observations or implausibly steep distribution functions. However, the 2009 sample is consistently the most extreme from the standpoint of top observations as compared to other years and for both income and expenditure. This is an important finding as our focus in this paper is on the 2009 income distribution.

Subsampling

Can sub-samples randomly extracted from the full surveyed sample bias the measurement of inequality? The Egyptian national statistical agency provides to researchers 25% or 50% of the full surveyed sample extracted randomly from the full sample. Extracting 25% of observations randomly from the full sample reduces the number of top and bottom income observations in the sample. This is similar to the problem of sampling of top observations already discussed. The probability of capturing top-income observations in a sub-sample follows the same probability laws as in the original sampling and we cannot predict ex-ante whether inequality will be under or over-estimated in the sub-sample randomly selected.

However, we can conduct a simple Monte Carlo experiment and extract 25% or 50% of observations randomly from the 2005 sample (which is available in full) 100 times, and then recalculate the Gini for each subsample. The figure below shows the results with the 100 Ginis sorted in the ascending order. As expected, the Gini of the full sample falls right in the middle of the distribution of Ginis calculated from the sub-samples for both income and expenditure per capita. However, the CAPMAS provides only one

sub-sample to researchers, and this sub-sample could yield any Gini in the range depicted in the Figures below. As can be seen, for both income and expenditure and for the 25 percent subsample, there are about 20 extractions that provide a Gini below the 95% lower bound of the value estimated from the full sample and about ten extractions that provide a Gini above the upper bound. This means that the 25% random sample can potentially provide biased Ginis about a third of the times, although we cannot predict *ex-ante* the direction or size of the bias. With the 50% subsample the problem persists but is reduced by about half. There is about a 15% chance that a Gini from an extracted subsample will be outside of the 95% confidence interval of the full-sample Gini. The final estimations in this paper will rely on 100% of the 2009 sample as we were able to run our programming codes on the full sample in the CAPMAS offices in Cairo. However, researchers who are currently using the 25% and 50% subsamples should be aware of this potential issue.

Figure 1. Monte Carlo experiment Ginis (100 repetitions, 25% or 50% random sample extractions)
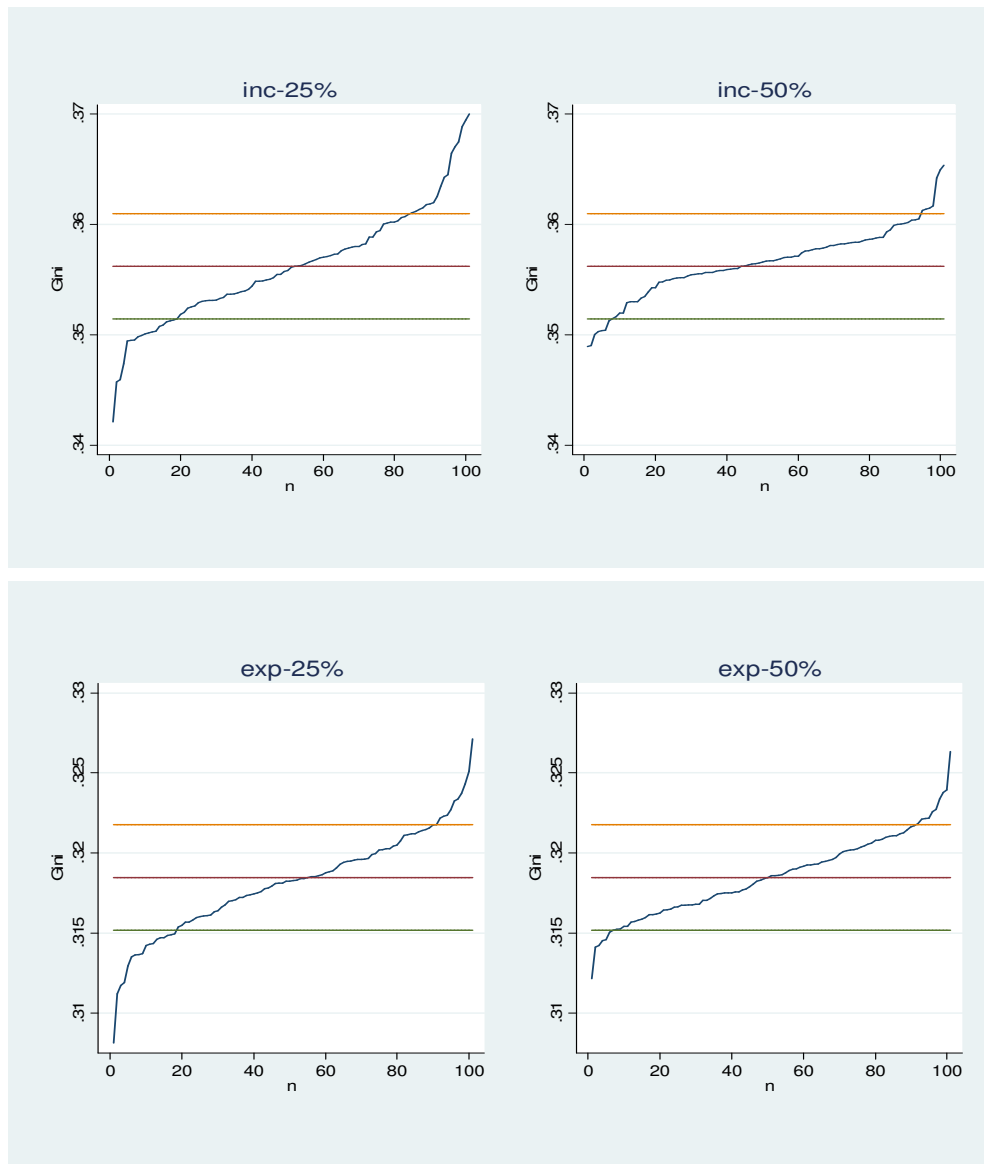
Unit non-response

Unit non-response is a problem in the HIECS data, particularly in some regions. Across governorates, the survey non-response rate in 2009 ranged from 0.0% to 10.5% with a mean of 3.7%. While the nationwide average non-response rate in the HIECS data is lower than in household surveys in other countries (for instance, refer to the literature surveyed in Korinek et al. 2006), it still leads to biases in statistics based on the observed sample. Out of 48,635 households contacted for the 2009 survey, only 46,857 responded to the survey, while 1,778 reportedly did not respond, a large number. Secondly, the problem may be more serious in some governorates than in others, and so interregional demographic comparisons based on the sample may be flawed. Table 1 illustrates the interregional differences in non-response rates and mean incomes of reporting households.

**Table 1. Non-response rates and mean incomes and expenditures by governorate**

Governorate	PSUs in the 100% Sample	Households	Non-Response Rate (%)	Mean Household Income	Mean Income per Capita	Mean Household Expenditure	Mean Expenditure per Capita
Alexandria	149	2,801	6.0	22,094.95	5,393.10	20,815.49	5,082.83
Assiut	101	1,872	2.4	14,188.56	2,665.06	11,800.88	2,216.75
Aswan	52	978	1.0	17,442.17	3,635.79	13,018.19	2,713.95
Behera	152	2,871	0.6	17,268.48	3,680.44	14,240.29	3,035.94
Beni Suef	69	1,294	1.3	15,258.93	2,887.36	13,514.71	2,557.90
Cairo	285	5,194	8.9	26,693.58	6,499.94	23,781.25	5,794.74
Dakahlia	176	3,289	1.6	18,852.61	4,467.94	15,898.13	3,768.32
Damietta	52	959	2.9	21,379.38	5,460.37	18,202.50	4,654.69
Fayoum	78	1,466	1.1	17,120.80	3,071.68	15,523.90	2,784.29
Gharbia	139	2,584	2.2	20,925.32	4,606.58	18,255.12	4,025.31
Giza	215	3,939	6.5	19,684.33	4,347.80	17,270.96	3,821.73
Ismailia	52	967	2.1	25,295.13	5,401.84	17,843.52	3,810.80
Kafr ElSheikh	85	1,547	4.2	25,035.71	4,279.37	20,465.43	3,497.10
Kalyoubia	145	2,668	3.2	20,178.65	4,137.20	17,753.81	3,642.90
Luxor	14	263	1.1	20,629.04	4,704.10	15,746.41	3,591.63
Matrouh	11	209	0.0	28,858.18	5,861.38	22,282.55	4,525.81
Menia	128	2,371	2.5	19,469.71	3,451.37	16,205.61	2,876.04
Menoufia	107	1,977	2.8	19,622.80	4,147.15	15,742.03	3,324.27
New Valley	8	146	3.9	26,562.99	5,322.18	22,243.22	4,458.13
North Sinai	14	243	10.5	17,891.85	3,768.41	13,423.69	2,829.52
Port Said	50	925	7.4	28,091.89	6,501.37	25,207.07	5,844.91
Qena	88	1,628	2.6	17,655.77	3,302.03	14,099.05	2,637.08
Red Sea	13	239	3.2	30,745.62	7,050.69	22,396.95	5,151.85
Shrkia	175	3,262	1.9	16,454.62	3,662.45	13,896.70	3,093.52
South Sinai	4	69	9.2	52,438.13	10,969.95	29,246.05	6,357.09
Suez	50	951	4.9	31,069.54	7,269.37	27,198.66	6,370.75
Suhag	114	2,145	1.0	13,961.63	2,809.37	11,880.76	2,391.82
Mean	94	1,735	3.7	20,549.65	4,653.03	17,375.99	3,974.44

Note: Non-response rate, reported in the survey at the PSU level, is weighted by the number of responding households in each PSU. Household income and expenditure, reported in the survey at the household level, are also weighted by the number of responding households in each PSU. Per-capita income and expenditure are further

weighted by household size. These mean incomes and expenditures may not be representative of those for the entire governorates, as they omit non-responding households.

Figure 2. Mean household non-response rates versus mean incomes per capita at the PSU level
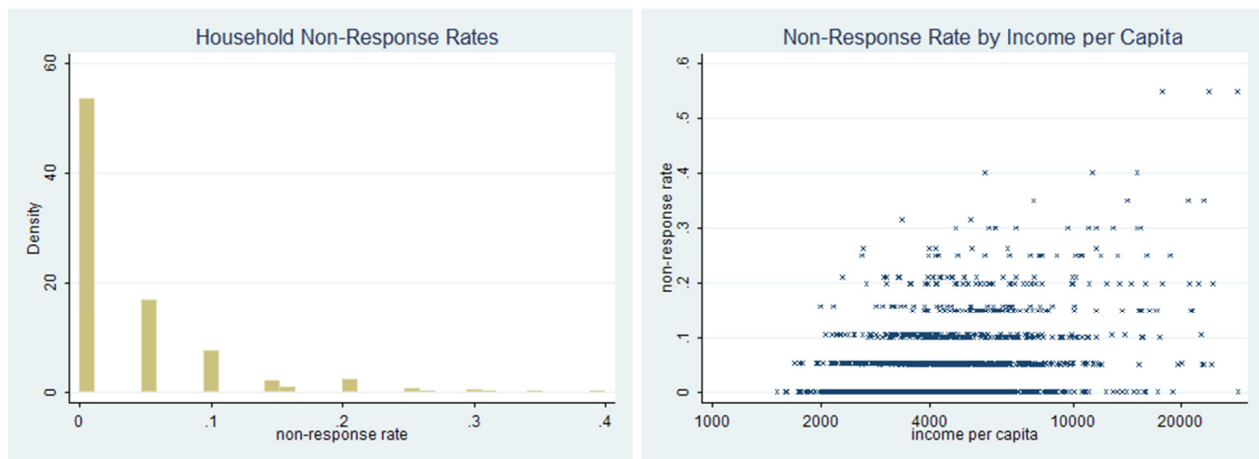
(a) Histogram of mean non-response rates   (b) Mean non-response rate by income per capita

Note: The unit of observation in this figure is a PSU. Average household non-response rate and average income per capita in a PSU are shown.

Unit non-response in a region is associated positively with income of responding households in that region. At the level of governorates, the Pearson correlation of the non-response rate with per-capita income of reporting households is 0.53, and 0.54 with per-capita expenditure of reporting households. At the level of individual primary sampling units (PSU), the correlations are 0.39 and 0.46, respectively. Figure 2a reports that survey non-response rate ranges from 0.0% to 55% with a heavy right tail. Figure 2b shows the systematic relationship between household non-response rate and mean per-capita income of responding households at PSUs. Non-response rates greater than 33% occur only among the richest 25% of PSUs in terms of income per capita, and only among the richest 15% of PSUs in terms of expenditure per capita. Because of these findings, it is likely that mean incomes and expenditures are even higher in the underlying populations of regions with high non-response rates, and that the associations are even stronger with the incomes and expenditures of the underlying populations.

Table 2 shows the results of estimation of households' survey response as a function of household income or expenditure. Our sample covers the 100% sample of the 2009 HIECS. Response-probability is thus estimated for 46,857 households, by fitting population in 2,526 PSUs. Following Korinek et al.'s (2006, 2007) lead, all models estimate survey-response probability as a nonlinear function of income or expenditure. Models 1 and 2 make $g(x)$ in Equation 1 a function of household income or expenditure. Models 3-10 use imputed income or expenditure per capita as explanatory variables, by dividing household-level variables by household size. Model specifications in Table 2 were selected in concurrence with Korinek et al.'s models, and with the aim to evaluate a variety of functional forms, from linear to highly non-linear.

The basic finding is that households' survey response is related negatively to income and expenditures. The coefficients on income and expenditures are consistently negative, and statistically very significant.

The simplest uni-variate logarithmic functions exhibit better fit than more complex or polynomial functions. They yield greater significance of all coefficients, lower value of the minimization objective function, and lower values of the Akaike and Schwarz Information Criteria, implying more efficient overall model fit.

Household expenditure appears to have a better explanatory power than household income, yielding lower values of the Akaike and Schwarz Information Criteria. Income and expenditure per capita provide better fit than household-level income and expenditure, implying that dividing household-level values by household size yields variables that are more predictive of householders' probability to respond than the household-level equivalents, without introducing additional noise into the model.

The negative relationship between income (expenditure) and response probability is particularly strong at high incomes (expenditures). The estimated relationship is highly nonlinear, with the response rate dropping rapidly in the highest range of expenditures. Models using linear, quadratic or polynomial functions (such as square-root or cubic-root of expenditures) rather than logarithmic functions achieve inferior measures of fit. Linear, quadratic and square-root models (Models 7-9) exhibit the poorest fit.

The various models correcting for non-response bias yield similar estimates for the measure of income inequality. The last two columns in Table 2 report the estimated Gini coefficients for income and expenditure per capita across models. They range from 0.329 to 0.351, for income, and from 0.305 to 0.320, for expenditure. Considering the differences in specifications used and fit achieved, these ranges are quite narrow, particularly for expenditure. Across models, 95% confidence intervals of the income Gini coefficients have lower bounds of 0.324-0.336 and upper bounds of 0.333-0.365. Expenditure Gini coefficients have lower bounds of 0.302-0.313 and upper bounds of 0.309-0.327. With the exception of the Gini coefficients from the poorly performing Models 7-9, all Ginis fit within the 95% confidence intervals of each other. This provides some evidence of consistency of the estimates.

Table 2. Estimation results for various logistic models of response probability

Specification of $g(x)$	$E(\theta_1)$ / s.e.	$E(\theta_2)$ / s.e.	Objective Value: Sum of Squared Weighted Errors	Factor of Proportio-nality (σ^2)	Akaike Informat. Criterion	Schwarz Informat. Criterion	Per-Capita Income Gini / s.e.	Per-Capita Expendit. Gini / s.e.
Household level								
1: $\theta_1 + \theta_2 \log(income)$	14.9909 (.0169)	-1.1853 (.0016)	85,079.65	.0776	8,887.82	8,885.20	.3506 (.0072)	.3151 (.0024)
2: $\theta_1 + \theta_2 \log(expenditure)$	17.2057 (.0184)	-1.4232 (.0017)	81,219.50	.0753	8,770.53	8,767.92	.3426 (.0035)	.3200 (.0033)
Per capita								
3: $\theta_1 + \theta_2 \log(income)$	11.6554 (.0122)	-.9939 (.0013)	83,400.47	.0757	8,837.46	8,834.85	.3488 (.0062)	.3151 (.0023)
4: $\theta_1 + \theta_2 \log(expenditure)$	13.0790 (.0142)	-1.1742 (.0015)	80,554.84	.0737	8,749.77	8,747.16	.3423 (.0035)	.3181 (.0025)
5: $\theta_1 + \theta_2 \log(exp.)^2$	7.4535 (.0066)	-.0603 (.0001)	81,623.97	.0744	8,783.08	8,780.46	.3421 (.0039)	.3176 (.0026)
6: $\theta_1 \log(exp.) + \theta_2 \log(exp.)^2$	1.5485 (.0013)	-.1391 (.0001)	83,644.60	.0757	8,844.85	8,842.23	.3418 (.0045)	.3168 (.0028)
7: $\theta_1 + \theta_2 10^{-3} expenditure$	3.3528 (.0019)	-.0254 (.0000)	95,919.03	.0845	9,190.73	9,188.11	.3338 (.0044)	.3084 (.0023)
8: $\theta_1 + \theta_2 10^{-9} expenditure^2$	3.2832 (.0020)	-.0026 (.0189)	99,480.83	.0873	9,282.83	9,280.21	.3289 (.0023)	.3054 (.0017)

|								
9: $\theta_1+\theta_2 \text{expenditure}^{1/2}$	4.0854 (.0023)	-.0137 (.0000)	88,808.82	.0792	8,996.18	8,993.56	.3388 (.0052)	.3130 (.0029)
10: $\theta_1+\theta_2 \text{expenditure}^{1/3}$	5.1798 (.0035)	-.1224 (.0001)	85,366.91	.0768	8,896.33	8,893.72	.3408 (.0049)	.3153 (.0029)

Note: Sample size is 2,526 PSUs, containing 46,857 household observations. PSU populations are fitted using response probabilities estimated for all households. Standard errors on Gini coefficients are bootstrapped estimates.

Beside the ten models in Table 2, we have considered other polynomial specifications as well as a model controlling for the four quarterly rounds in the 2009 HIECS. While some coefficients in these models were statistically significant, the models' overall fit was worse than in Models 1-4, and the corresponding Gini coefficients did not depart significantly from those in Table 2. The imputed household response probabilities and Gini coefficients are thus not too sensitive to the addition of more variables into *g(x)*.

In the rest of the analysis, we will use Model 4 as a benchmark specification, due to its superior fit, and similarity to the model used by Korinek et al. (2006, 2007). The following figures provide additional results for this model, as well as other comparison models. Figure 3 shows households' probability of survey response by income or expenditure per capita estimated in Models 3 and 4. In agreement with negative estimates of θ_2 in the logarithmic specifications, the estimated response-probability falls with income, most rapidly in the highest range of incomes (expenditures). Figure 3 thus confirms the central premise of this analysis, that richer households are systematically less likely to participate in surveys, and that this issue is particularly grave for top-income households. The response probabilities shown here will be used as the appropriate household weights for the imputation of income distribution and measures of inequality.

Figure 3. Estimated household response probability by income or expenditure per capita (Models 3, 4)

(a) Model 3   (b) Model 4
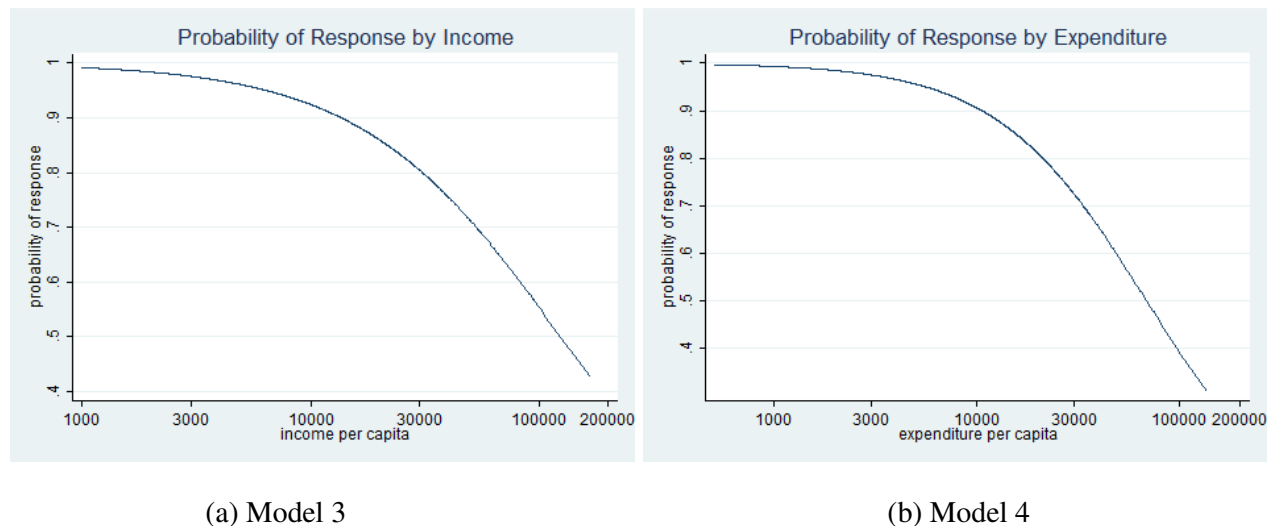
The corrected weights differ significantly from the CAPMAS-provided sampling weights. The CAPMAS provides sampling weights that correct for unit non-response by simply expanding the weight for the non-response rate at PSUs. CAPMAS-provided sampling weights are normalized to 1, have standard deviation of 0.173, and are identical for all households within a PSU. Weights from Model 4, obtained as the inverse response probabilities estimated in that model, have a mean of 1.041, standard deviation of 0.057,

and vary across all households even within PSUs. Figure 4 reports the distribution of households' sampling weights provided by the CAPMAS and those derived from Model 4 (demeaned for ease of comparison).

Figure 4. Distribution of CAPMAS-provided sampling weights and weights correcting for non-response bias from Model 4
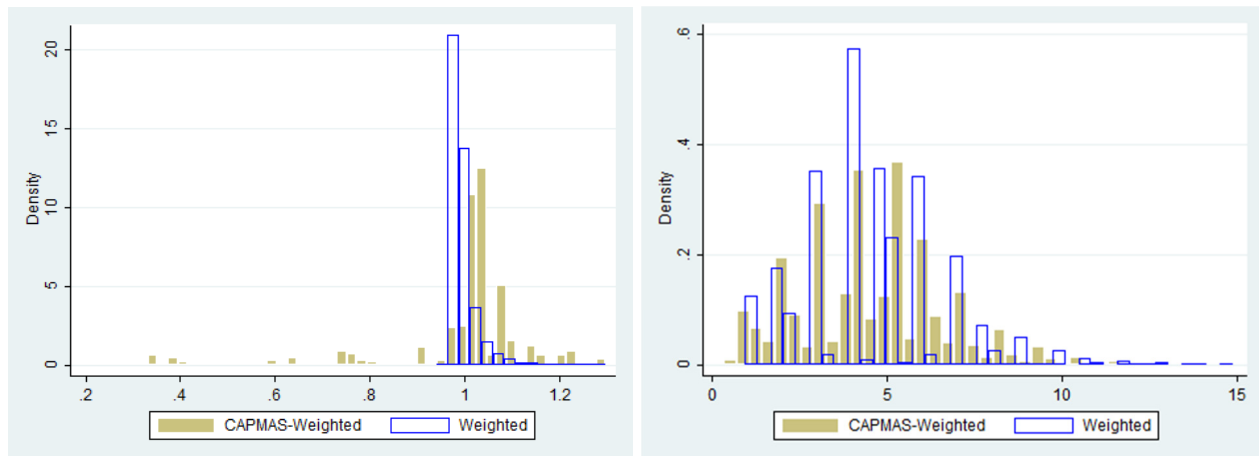

(a) Weights for household-level variables   (b) Weights for per-capita variables

Note: Weights are normalized to have a mean of 1, and of mean household size (4.665), respectively.

Use of the corrected weights affects the imputed income distribution. Figures 5-6 show the implications of our estimation for the imputed distribution of per-capita incomes and the corresponding Lorenz curves, for the entire population as well as for the poorest and richest households. (Similar results for expenditure per capita are available on request.) These figures show that our correction of the survey-nonresponse bias increases our measurement of income-inequality. The Lorenz curve imputed using our weights first-order dominates both the uncorrected Lorenz curve as well as the CAPMAS sampling-weights corrected Lorenz curve on the entire domain. The uncorrected and CAPMAS-corrected Lorenz curves do not exhibit clear dominance over one another. Under our corrected income distribution, the estimated fraction of households in the highest income range increases, and the fraction of households in all lower income ranges – including the lowest-income range (less than LE2,500 in Figure 5 panel b) – falls.
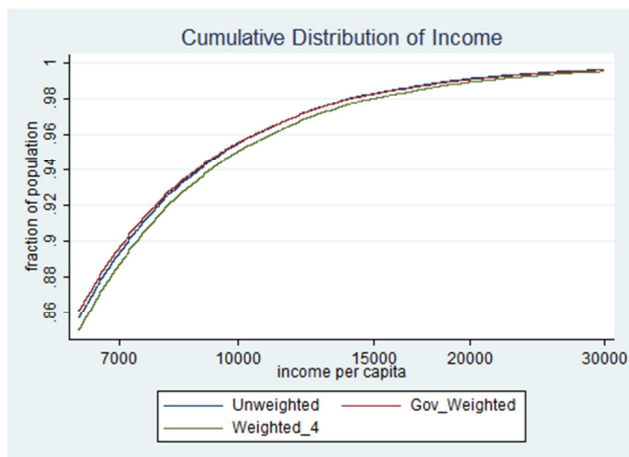
Figure 5. Cumulative distribution of income per capita across population, and among the poorest 25% and richest 10% of households (Model 4)
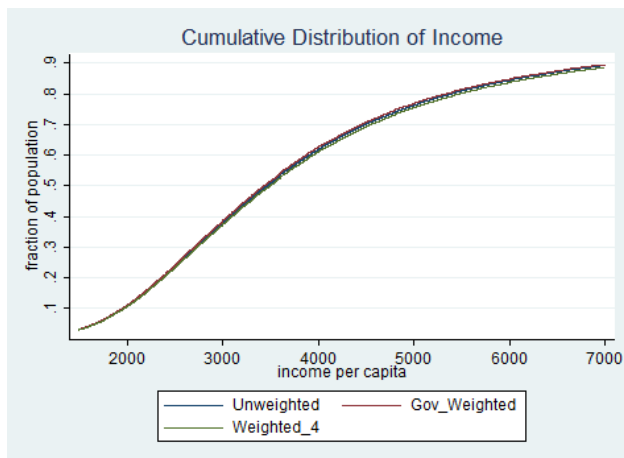
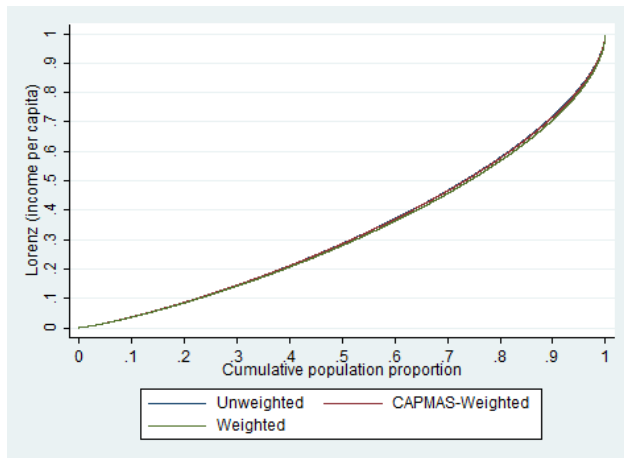(a) Per-capita income distribution (Model 4)
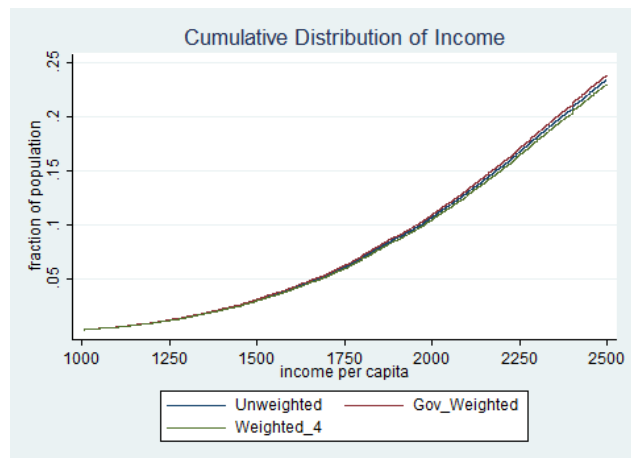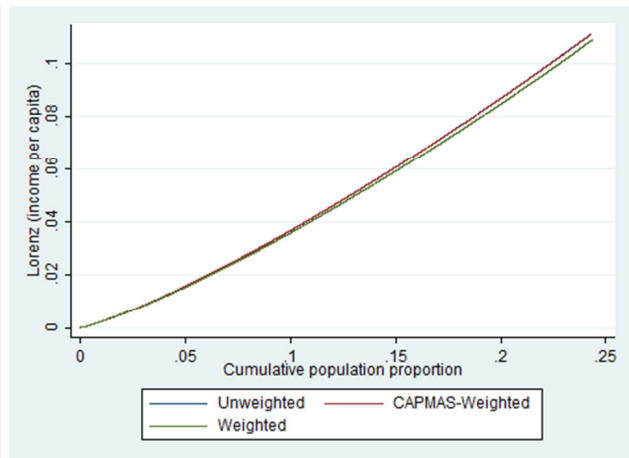

(b) Poorest 25% per-capita incomes (Model 4)

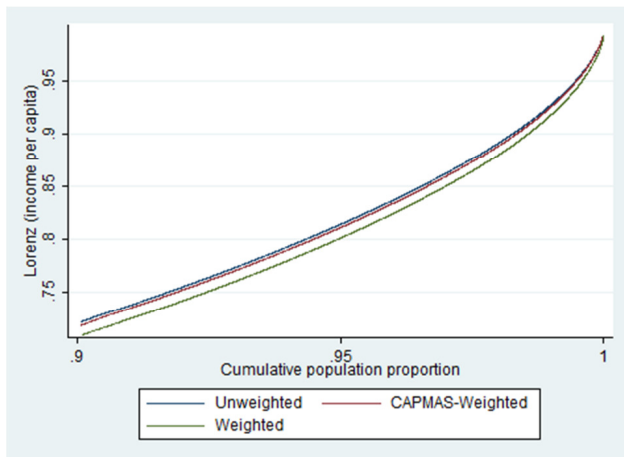(c) Richest 10% per-capita incomes (Model 4)

Figure 6. Lorenz curves in the population, and for the poorest 25% and richest 10% of households (Model 4)

(a) Lorenz curve in the population (Model 4)

(b) Lorenz curve for the poorest 25% (Model 4)


(c) Lorenz curve for the richest 10% (Model 4)

Correspondingly, use of the corrected weights affects the imputed Gini index of inequality positively. By reweighting income distribution to account for households' endogenous survey response, we obtain significantly higher measures of income inequality. The Gini coefficient for per-capita incomes using simple household-size weights is 0.3289 (s.e. 0.0023). The Gini coefficient using the CAPMAS-provided sampling weights is 0.3305 (s.e. 0.0024). The Gini coefficient using response-probability weights estimated in our Model 4 is 0.3423 (s.e. 0.0035). This corrected Gini coefficient is statistically higher than both of the uncorrected ones at the 1% level of significance (p-values of 0.002).

For per-capita expenditure, the Gini coefficient for the unweighted distribution is 0.3054 (s.e. 0.0017), while that using the CAPMAS-provided sampling weights is 0.3070 (s.e. 0.0019). The Gini coefficient using response-probability weights estimated in Model 4 is 0.3181 (s.e. 0.0025). Again, this corrected Gini coefficient is statistically higher than either of the uncorrected ones at the 1% level of significance (p-values of 0.001).

21

Use of the corrected weights also significantly affects the estimated distribution of top incomes. The Pareto coefficient for unweighted per-capita incomes is 2.428, and the inverted Pareto coefficient is 1.700. For incomes weighted by the CAPMAS-provided weights, these coefficients are 2.392 and 1.718, respectively. For incomes weighted by the response-probability weights estimated in Model 4, these coefficients are 2.250 and 1.800. For per-capita expenditure, the Pareto and inverted Pareto coefficients are 2.685 and 1.593 in the unweighted income distribution, 2.606 and 1.623 in the income distribution weighted using the CAPMAS weights, and 2.478 and 1.677 in the income distribution weighted as per Model 4.

The corrected weights estimated across the alternative models in Table 2 give rise to very different estimates of top-income distribution. The Pareto coefficients for per-capita incomes estimated in Models 1-3 and 5-10 are, respectively: 2.051, 2.268, 2.078, 2.231, 2.217, 2.291, 2.428, 2.219, and 2.210. (These and additional results for the Gini and Pareto coefficients across all models are provided in the annex.) This variation can be explained by the differential treatment of top-income households across models. Different models assign different weights to households with the highest incomes. By estimating households' survey-response probability as a function of their log-expenditure (or log-income), versus regular or squared expenditure, we assign very different weights to the highest-income households, while keeping weights of lower-income households similar. Figure 7 plots the alternative weights across households with different expenditures.[3] Clearly, the weights diverge for the richest households. Correspondingly, the estimated Lorenz curves differ particularly for highest-income households (as evident in Figure A2 in the annex).

Figure 7. Household weights across selected models
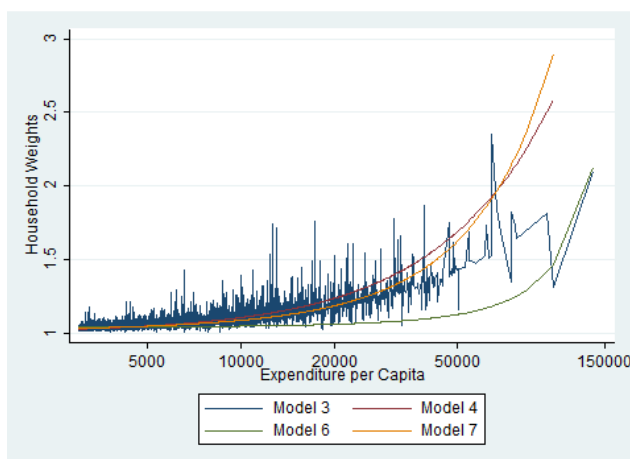
Extreme observations

In this section we test the sensitivity of the Gini coefficients to extreme observations on the right-hand side of the distribution (top incomes), in the raw data as well as in the income distribution corrected for

---

[3] Expenditure, rather than income, is shown for clarity of presentation, since most models use functions of expenditure. Note that the weights from Model 3 are a function of income, hence their plot against expenditure is not as smooth as for other models.

unit non-response. If top incomes turn out to be influential, we then correct for their presence using an estimated Pareto distribution as discussed in the methodological part.

In our data, the Gini is very sensitive to extreme observations irrespective of sample size. In the Figure below, we recalculated the Gini for the CAPMAS surveys by removing top-income observations one at a time, up to 100 observations and for each of the four years considered. This was done on the 25% subsample for 2000, 2009 and 2011 and for the full sample in 2005, which means that the sample size used is different for 2000 and 2009 (12,000 observations), 2005 (48,000 observations) and 2011 (4,000 observations). In this way, we can check how different sample sizes affect the sensitivity of Gini coefficients to top observations. 100 observations were chosen for removal in recognition of the finding by Neri et al. (2009) that up to 0.2% of income observations may represent outliers.

We can clearly see a tendency for the Gini to decline rapidly, and we can also see that the sensitivity to top observations is different for income and expenditure, and different for the four years considered. The scale of the sensitivity is related to sample size and the welfare aggregate. For both income and expenditure, the steepest curves are those for 2011 (the smallest sample) and the least steep are those for 2005 (the largest sample). This is perhaps expected as larger subsamples are likely to capture extreme observations more completely, and the Gini may be less sensitive to each one of them. It is also evident that the Gini on expenditure is more sensitive to extreme values than the Gini on incomes. This is less expected given that income has a higher Gini than expenditure and given that expenditure is less likely to have extreme observations.

**Figure 8 – Sensitivity tests of the Gini to the removal of the top 100 observations**
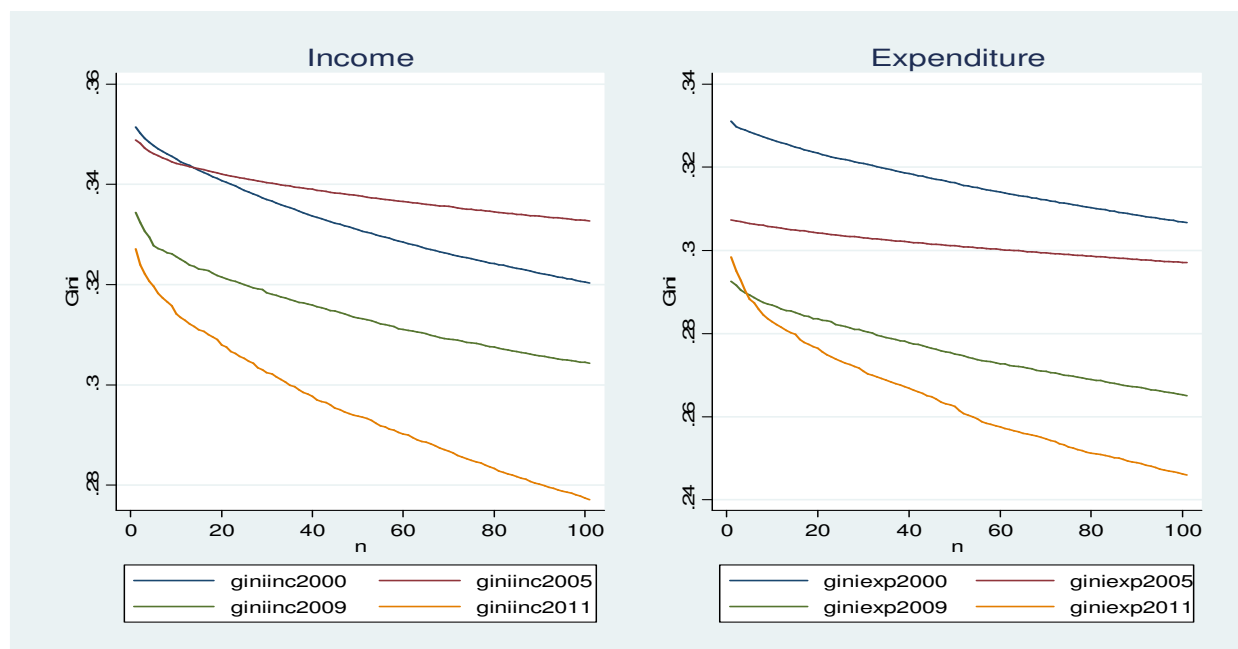
It is clear that the extreme values in the Egyptian distribution of income and expenditure cannot be ignored. On the one hand, removing some of the top observations may contribute to underestimation of inequality if these observations are accurate and representative of the underlying population. On the other hand, by keeping top observations that arise from data errors or those that do not represent the underlying

population well may lead to overestimation of inequality. In both cases, our inequality estimates would be biased, particularly in small sample extractions.

The sensitivity of the Gini to extreme observations persists when we correct for unit non-response. A sensitivity analysis reported in Figure 9 shows that inequality measures are very sensitive to the top 0.025% of observations. In this analysis, we recalculate the Gini and Pareto coefficients after removing 0.025%-0.2% households with the highest incomes (12-96 households in the 100% sample of the 2009 HIECS). A significant portion of the difference in Gini coefficients across models disappears as we remove the highest-earning 0.025%-0.05% of households (12-24 households). Exclusion of additional high-income households does not yield significant changes. The difference in statistics across models appears to converge to a particular level, which decreases at a much slower rate with exclusion of additional households.[4] (Figures A3 in the annex reports the same patterns for the Pareto and inverted Pareto coefficients.)

Figure 9. Gini coefficients for income and expenditure in trimmed distributions (Models 3-7)
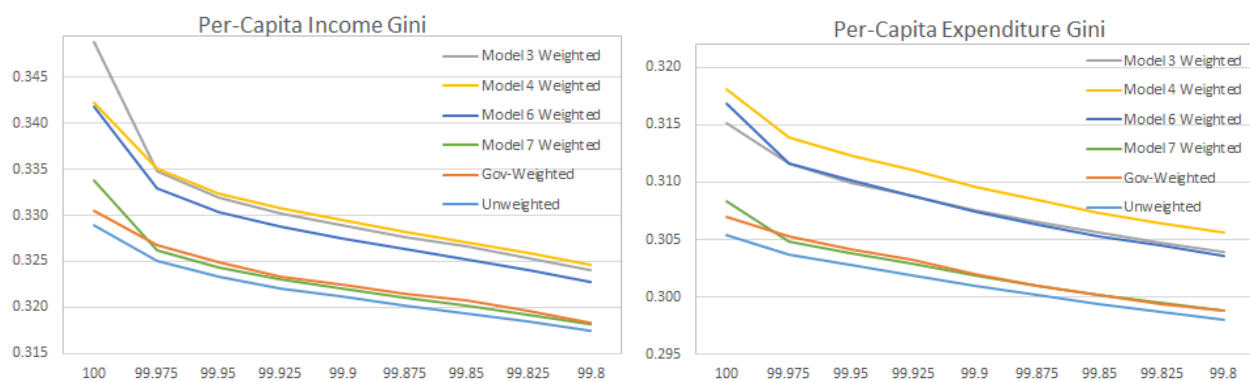
(a) Gini coefficient for income per capita   (b) Gini coefficient for expenditure per capita

---

[4] Not surprisingly, Model 3 yields a distribution of income that is more sensitive to the removal of highest-earning households than other models. (Refer to the left panels of Figures 9 and A3.) This is because household weights in Model 3 are functions of income, whereas weights in other models are functions of expenditure. The converse about lower sensitivity of the Model 3 Gini coefficient for expenditures does not hold, however. Because the distribution of expenditures suffers less from extreme observations than income, as Table 1 has suggested, expenditure Ginis across the alternative models vary less in the overall sample, and Model 3 Gini is no less sensitive to the top 0.025% of observations than Ginis from other models.

Gini coefficients from Models 3-6 are always substantially higher than the unweighted or the CAPMAS-weighted Gini coefficients, for both income and expenditure. On the other hand, Gini coefficients from Model 7, the linear model, converge to the uncorrected Ginis after the initial 12 top households are removed. This suggests that the imputed income (or expenditure) distribution from the linear model does not differ much from the uncorrected distribution, except for the influence of the topmost 0.025% observations. With the exception of the linear Model 7, the differences between all response-bias corrected and CAPMAS-corrected Gini coefficients for incomes are significant at the 1% level in the overall sample, but become significant even at the 0.1% level when the top 0.025-0.2% of households are removed. This is because the presence of the highest-earning households in our sample introduces noise that increases standard errors even more than it moves model Gini coefficients away from the CAPMAS weight-corrected Gini coefficients. Hence, as high-income households are excluded, the values of uncorrected and corrected Gini coefficients become closer to each other, but their differences retain their statistical significance.

Note: '100' indicates full, untrimmed income distribution. '99.975' indicates income distribution with the 0.025% households with the highest incomes trimmed (12 households in the 100% sample of the 2009 HIECS). Similarly, '99.8' indicates the trimming of 0.2% of highest-earning households (96).

The discussion above suggests that observations with the highest incomes affect the measurement of inequality. Excluding these observations from the sample yields lower and more homogeneous estimates of inequality across models. A question arises whether exclusion is appropriate theoretically, given that it reduces sample size and may result in the censoring of meaningful observations. Here we address these questions by comparing actual observations of top incomes with values imputed under their expected Pareto distribution, and estimating the effects on Gini coefficients. This provides an alternative way to evaluate robustness of our Gini coefficients to the presence of extreme income observations in the sample. In view of our results about survey non-response by top-income households, this also allows us to comment on the relative significance of the two statistical issues.

Table 3 presents semi-parametric estimates of Gini coefficients, obtained by replacing the highest top 10 percent of income observations (alternatively, 5% or 20%) with values imputed from the corresponding Pareto distribution as per Cowell and Flachaire (2007), and Davidson and Flachaire (2007). The first three rows show the benchmark non-parametric estimates – unweighted; corrected for sampling probability using CAPMAS weights; and corrected for non-response bias as per Model 4. These three rows again illustrate the importance of correcting for survey non-response. These rows serve as a benchmark to which the following semi-parametric estimates will be compared.

The next three rows present the main results – semi-parametric estimates when the top 10 percent of incomes are imputed from a corresponding Pareto distribution. The following six rows report on a robustness check, where such imputation is performed on top 5 percent, or top 20 percent of incomes.

The main finding is that the CAPMAS data do not appear to suffer from extreme income observations relative to what would be predicted if our top-income data followed the Pareto distribution exactly. The corrected Gini coefficients are essentially unchanged, falling or rising by a very small amount. This suggests that the exclusion of top incomes in the previous section is not warranted on the grounds that they are outliers, but simply as a robustness test of the Gini estimates to individual income observations. The size of the correction for extreme observations is trivial compared to the correction for unit non-response. The results for expenditure per capita are analogous, and are shown in the annex.

In the income distribution uncorrected for non-response bias, the semi-parametric Gini coefficient – corrected for the possible presence of extreme observations among the top 10% of incomes – is 0.3278 compared to the non-parametric value of 0.3289. When we increase the range of top incomes to be imputed, from 10% to 20% of households, the semi-parametric Gini falls to 0.3273. In the income distribution sampling-corrected using CAPMAS weights, the semi-parametric Gini coefficient is same as the non-parametric estimate, 0.3305. Finally, in the income distribution corrected for non-response bias using weights from Model 4, the corrected Gini coefficient is again the same as the uncorrected value, 0.3423. When we increase the range of top incomes to be imputed, from 10% to 20% of households, the semi-parametric Gini rises slightly, to 0.3425.

Table 3. Non-parametric and semi-parametric estimates of Gini coefficients

Modeling of top incomes	Correction for extreme observations	k	Sampling correction	Pareto coefficient a (s.e.)	Gini$_{n-k}$, (s.e.)	Gini$_k$, (s.e.)	Gini (s.e.)
Non-parametric	No	k=10%	No	2.4279 (.0309)	.2191 (.0007)	.2584 (.0069)	.3289 (.0023)
	No	k=10%	Yes, CAPMAS	2.3919 (.0326)	.2175 (.0007)	.2654 (.0070)	.3305 (.0024)
	No	k=10%	Yes, Model 4	2.2501 (.0329)	.2214 (.0007)	.2844 (.0112)	.3423 (.0035)
Semi-parametric	Yes	k=10%	No	2.4279 (.0309)	.2191 (.0007)	.2594	.3278
	Yes	k=10%	Yes, CAPMAS	2.3919 (.0326)	.2175 (.0007)	.2643	.3305
	Yes	k=10%	Yes, Model 4	2.2501 (.0329)	.2214 (.0007)	.2857	.3423
Semi-parametric	Yes	k=5%	No	2.4638 (.0937)	.2463 (.0008)	.2546	.3288
	Yes	k=5%	Yes, CAPMAS	2.4378 (.0969)	.2452 (.0008)	.2580	.3305
	Yes	k=5%	Yes, Model 4	2.2507 (.0961)	.2503 (.0008)	.2856	.3422
Semi-parametric	Yes	k=20%	No	2.4190 (.0223)	.1864 (.0007)	.2606	.3273
	Yes	k=20%	Yes, CAPMAS	2.3811 (.0234)	.1849 (.0007)	.2658	.3306
	Yes	k=20%	Yes, Model 4	2.2603 (.0232)	.1876 (.0007)	.2840	.3425

We can now come back to the question of within-j/between-j trade-off discussed in the methodological section. We argued that using a highly aggregated j would be likely to overshoot the Gini correction and would lead to results that are less consistent with the Pareto corrections proposed. Indeed, our non-response correction – of 1-2 percentage points – is smaller than that reported by Korinek et al. (2006, 2007) – of 4-5 percentage points.

To test the claims regarding appropriate geographic aggregation, we have re-estimated the models in Table 2 using governorates by urban and rural substrata (50 areas) rather than PSUs (see Table 4). If we compare the models with the best fit (model 4) we find that using governorates by urban and rural areas raises the corrected Gini (s.e.) for income from 34.23 (0.0035) to 37.14 (0.0129) and the corrected Gini for expenditure from 31.81 (0.0025) to 34.19 (0.0075). Across most models, the estimated Ginis rise by 3-5 percentage points for income, and by 1-4 percentage points for expenditure.

In our view, Table 2 provides more accurate estimates for the HIECS data than Table 4. First, Ginis estimated at the governorate by urban/rural areas are consistently higher than the semi-parametric Ginis estimated using the alternative Cowell and Flachaire (2007) and Davidson and Flachaire (2007) methodology proposed while the Ginis estimated with PSUs are very much in line with those estimates. Second, in Table 4, all Ginis show significantly higher standard errors. Third, the HIECS data has a much higher household response rate (96.3%) than the US Current Population Survey (91.7%), implying less bias. And fourth, inequality is much lower in the HIECS data, suggesting that the percentage-point

correction may be lower. The optimal tradeoff of the within-*j*/between-*j* number of bins depends on the nature of the model and on the nature of the data at hand. This paper has proposed a different approach and applied this approach to a different data set as compared to Korinek et al. (2006 and 2007). Clearly, the question of optimal within-*j*/between-*j* trade-off will require testing in a separate paper to be fully exhausted but this paper showed that an alternative path is possible and also preferable in the case of the HIECS data.

Table 4. Estimation Results for Various Logistic Models of Response Probability (by governorate and urban/ rural areas)

Specification of $g(X)$	$E(\theta_1)$ / s.e.	$E(\theta_2)$ / s.e.	Objective Value: Sum of Squared Weighted Errors	Factor of Proportio -nality (σ^2)	Akaike Informat. Criterion	Schwarz Informat. Criterion	Per-Capita Income Gini / s.e.	Per-Capita Expendit. Gini / s.e.
Household level								
1: $\theta_1+\theta_2 \log(\text{income})$	20.8870	-1.7686	780,896	.8543	486.81	484.19	.4411	.3398
	(.0088)	(.0008)					(.0389)	(.0070)
2: $\theta_1+\theta_2 \log(\text{expenditure})$	25.5496	-2.2284	299,122	.3321	438.83	436.21	.3798	.3625
	(.0073)	(.0007)					(.0151)	(.0181)
Per capita								
3: $\theta_1+\theta_2 \log(\text{income})$	15.8384	-1.4714	577,654	.6505	471.74	469.12	.4210	.3375
	(.0063)	(.0007)					(.0301)	(.0086)
4: $\theta_1+\theta_2 \log(\text{expenditure})$	18.6483	-1.7947	299,994	.3321	438.97	436.36	.3714	.3419
	(.0062)	(.0006)					(.0129)	(.0075)
5: $\theta_1+\theta_2 \log(\text{exp.})^2$	9.9506	-.0916	344,805	.3828	445.94	443.32	.3784	.3452
	(.0028)	(.0000)					(.0188)	(.0101)
6: $\theta_1 \log(\text{exp})+\theta_2 \log(\text{exp})^2$	2.0269	-.1934	450,540	.5036	459.31	456.70	.3862	.3481
	(.0005)	(.0001)					(.0266)	(.0134)
7: $\theta_1+\theta_2 10^{-3} \text{expenditure}$	3.1297	-.0344	2,189,226	2.2715	538.35	535.74	.3594	.3202
	(.0007)	(.0000)					(.0256)	(.0104)
8: $\theta_1+\theta_2 10^{-9} \text{expenditure}^2$	2.9787	-.1329	2,599,937	2.6735	546.95	544.34	.3375	.3089
	(.0008)	(.0005)					(.0089)	(.0037)
9: $\theta_1+\theta_2 \text{expenditure}^{1/2}$	4.3705	-.0195	1,107,645	1.2019	504.29	501.67	.3859	.3399
	(.0009)	(.0000)					(.0373)	(.0165)
10: $\theta_1+\theta_2 \text{expenditure}^{1/3}$	6.1436	-.1785	667,983	.7437	479.00	476.39	.3889	.3459
	(.0014)	(.0001)					(.0335)	(.0156)

Note: Sample size is 50 governorate-urban/rural strata containing 46,857 household observations. Standard errors on Gini coefficients are bootstrapped estimates.
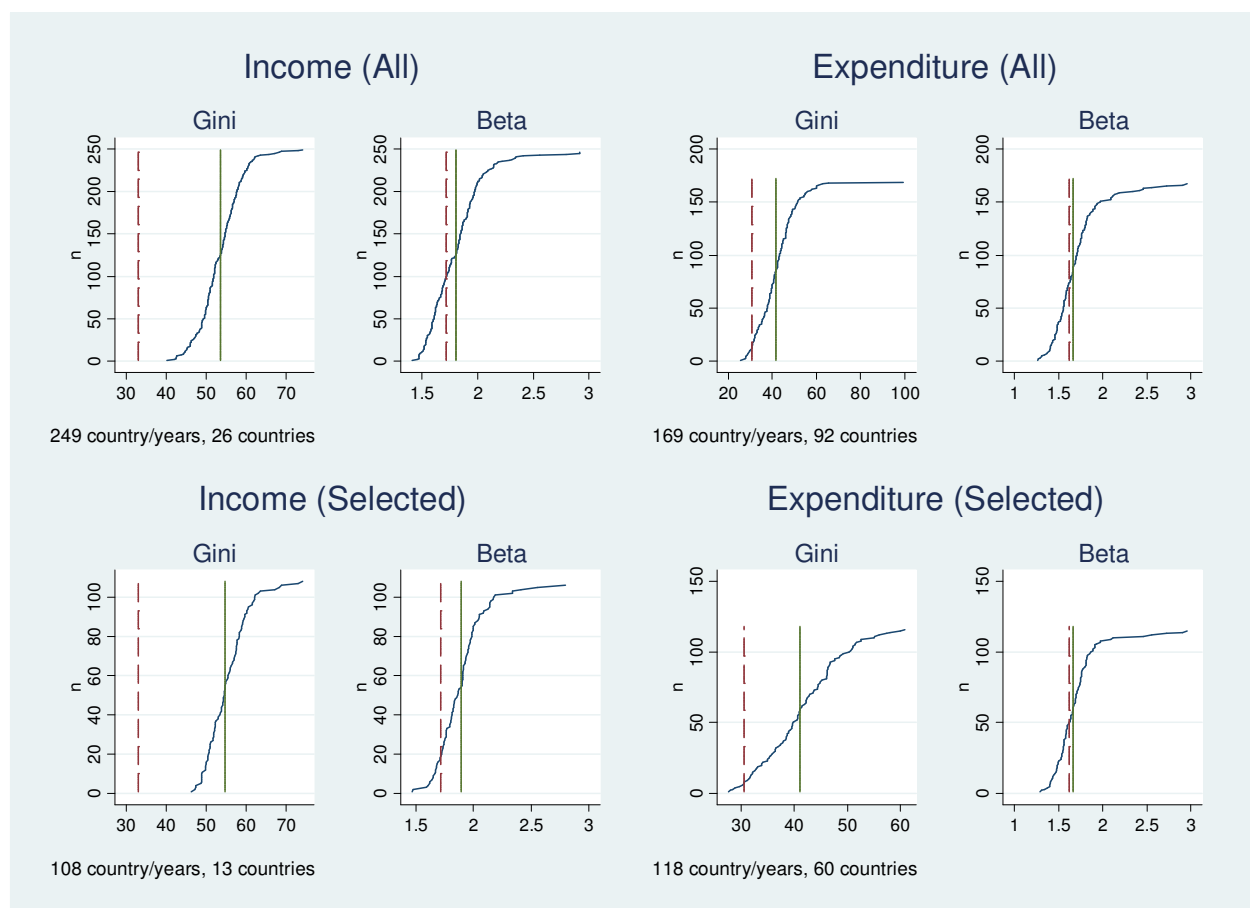
6. How different is Egypt from other countries?

In this section, we compare the Ginis and the inverted Pareto (beta) coefficients estimated for Egypt with a sample of world country/year Ginis and betas. The purpose is to put our results into the global context and understand whether our results pertain to an exceptional case-study or, rather, to an ordinary distribution of incomes. For these comparisons, we will use a sample of 107 countries and 418 country/year observations taken from the World Bank micro data repository. This database joins and standardizes several databases of household budget surveys for developing countries available at the World Bank. For each country and year it contains the full distribution of incomes, expenditures or both depending on the country and year considered.

The following figure plots the Gini and beta coefficients for both income and expenditure per capita across country surveys sorted in ascending order. The top panels use all 107 countries available in the database while the bottom panels use a selection of 65 countries that are the closest to Egypt in terms of GDP per capita (2008 USD). In the eight panels of the figure, we superimpose the Gini and the beta coefficients estimated for Egypt from the 2009 full sample (dashed line) and the median value for the full world distribution (solid line).

The Egyptian Ginis are clearly situated in the lower part of the world distribution for both income and expenditure. This is the case also if we restrict the analysis to countries at similar levels of GDP per capita. This confirms that the Gini in Egypt is low by world standards. Instead, if we consider the beta coefficient and all country/year points, Egypt falls very close to the median value for both income and expenditure. This is also the case with expenditure for the selected sample of similar countries while the beta coefficient is slightly to the left of the median value if we consider selected countries and income. This last result should be taken with caution because the income panel for selected countries includes only 13 countries which are all Latin American countries.

Figure 10 - Gini and inverted Pareto coefficients for Egypt and the rest of the world

In essence, the right-hand tails of the Egyptian distributions are not much different from other countries despite a very low income inequality. As we showed throughout the paper, the Gini is very sensitive to top incomes. The Egyptian beta being close to the world median value suggests that top incomes are well

represented as compared to world countries and yet inequality is still very low. This is rather robust evidence of the good quality of the Egyptian data, a finding consistent with the rest of the paper and with the World Bank (2013) report on inequality in Egypt.

7. Discussion

This paper has evaluated income inequality and the distribution of top incomes in Egypt in the presence of a variety of potential statistical issues. As a byproduct, it has evaluated the quality of data in the Egyptian Household Income, Expenditure and Consumption Survey (HIECS).

We discussed the problem of item non-response in household surveys, but finding no missing items in the HIECS data, we confirmed data quality on these grounds. We then tested and corrected for the problem of unit non-response by top income households. Correction for unit non-response increased the estimate of inequality by 1.3 percentage points. The estimated Gini coefficient for income per capita rose from 0.329 to 0.342, while the Gini for expenditure per capita rose from 0.305 to 0.318, statistically very significant.

Given the importance of representation of top incomes in the sample, we next evaluated how influential are individual income observations at the upper tail of the Egyptian distribution, and whether they present a measurement issue. We found, however, that the Egyptian distribution of top incomes follows rather closely the Pareto distribution, so the observed top incomes appear to be representative of the underlying population and need to be considered when measuring inequality. This analysis reinforces the case for assigning of greater weight to the observed top incomes to correct for the systematic non-response of top income households in the population.

Finally, we benchmarked the estimated inequality and top income distribution in Egypt vis-à-vis 418 household surveys drawn from 107 countries, and found that the Gini coefficient in Egypt is significantly below median values for other countries, while the distribution of top incomes is around the median. Income inequality in Egypt is thus confirmed to be low while the distribution of top incomes is not atypical as compared to other countries.

There are several policy implications of these results that are relevant for Egypt today. First, the paper has validated the quality of the Egyptian HIECS with respect to top observations, the income and expenditure aggregates and the measurement of income inequality. Also, in the world of household surveys, the Egyptian data stand out as particularly good data. There are many more issues that could be explored in relation to data quality that were not covered but the tests conducted in this paper show that the HIECS compares well to world standards. The HIECS data cannot be simply dismissed as "unreliable" because people have a different perception of income inequality.

Second, these findings motivate the search for factors that could explain popular perceptions about income inequality elsewhere. As the World Bank (2013) report has shown, there are many factors that could explain perceptions of inequality that are little related with the measurement of income inequality itself and that are little researched, including the role of expectations about the future, changes in the reference groups, the expansion and penetration of the social media or the lack of GDP trickle-down effects. The priority for Egypt today may not be the reduction of income inequality but the expansion of the growth base, providing more opportunities to economically marginalized groups such as the youth and

women, providing more voice to the media-excluded groups such as the poor and rural residents and others. Inequality of opportunities, inequality of rights, inequality of aspirations and inequality of values are some of the inequality dimensions that are easily confounded with income inequality but that should be carefully distinguished by the policy maker.

Third, the fact that GDP growth did not trickle down to households during the decade that preceded the 2011 Egyptian revolution is very consistent with the fact that income inequality was low and changed little during the period. Preliminary results of an on-going research on GDP in Egypt show that growth has been mostly captured and retained by corporations and not paid to households via wages, benefits or dividends. The overarching goals of the World Bank are poverty reduction and shared prosperity measured in terms of the income growth of the bottom 40% of the population. Achieving these objectives largely relies on making growth inclusive of the bottom 40%, something that has not been happening in Egypt over the past decade. This is another question that requires further attention and priority as compared to income inequality.

References

Atkinson, A, Piketty, T. and Saez, E. (2011) Top incomes in the long run of history, Journal of Economic Literature, 49, 3-71.

Cowell, F.A. and Victoria-Feser, M.-P. (1996) Poverty measurement with contaminated data: A robust approach, European Economic Review, 40, 1761-1771.

Cowell, F.A. and Victoria-Feser, M.-P. (1996) Robustness properties of inequality measures, Econometrica, 64, 77- 101.

Cowell, F.A. and Victoria-Feser, M.-P. (2007) Robust Lorenz curves: a semiparametric approach, Journal of Economic Inequality, 5, 21–35.

Cowell, F.A. and Flachaire, E. (2007) Income distribution and inequality measurement: The problem of extreme values, Journal of Econometrics, 141(2), 1044–1072.

Davidson, R. and Flachaire, E. (2007) Asymptotic and bootstrap inference for inequality and poverty measures, Journal of Econometrics, 141(1), 141-166.

Deaton, A. (2005) Measuring Poverty in a growing world (or measuring growth in a poor world), The Review of Economics and Statistics, LXXXVII (1), 1-19

Korinek, A., Mistiaen, J.A. and Ravallion, M. (2006) Survey nonresponse and the distribution of income, Journal of Economic Inequality, 4, 33-55.

Korinek, A., Mistiaen, J.A. and Ravallion, M. (2007) An econometric method of correcting for unit nonresponse bias in surveys, Journal of Econometrics, 136, 213-235.

Neri, L., Gagliardi, F., Ciampalini, G., Verma, V. and Betti, G. (2009) Outliers at upper end of income distribution (EU-SILC 2007), DMQ Working Paper n. 86, November 2009.

Pareto, V. (1896) La courbe de la repartition de la richesse, Ecrits sur la courbe de la repartition

31

de la richesse, (writings by Pareto collected by G. Busino, Librairie Droz, 1965), 1-15.

World Bank (2013) Inside Inequality in Egypt: Historical trends, recent facts, people's perceptions and the spatial dimension, mimeo.

**Figure A1 – Top incomes and expenditures (EGP/year/capita)**
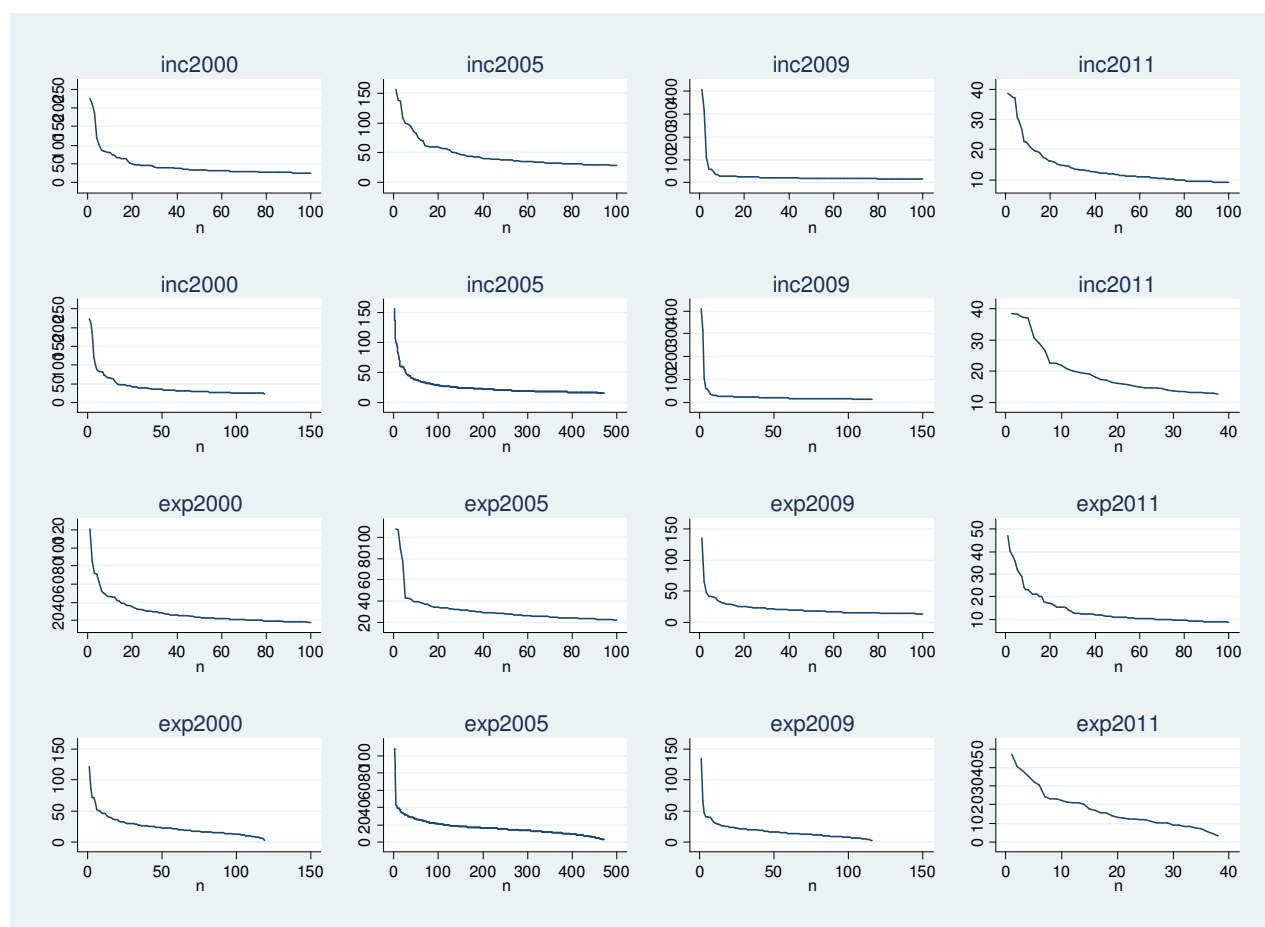
Note: x-axis=top 100 observations (top panels for income and expenditure) or top 1% of observations; y-axis=income per capita or expenditure per capita per year. The size of the four samples is different with 12,000 observations for 2000 and 2009, 48,000 observations for 2005 and 4,000 observations for 2011.

Figure A2. Differences between Lorenz curves (Model 4)
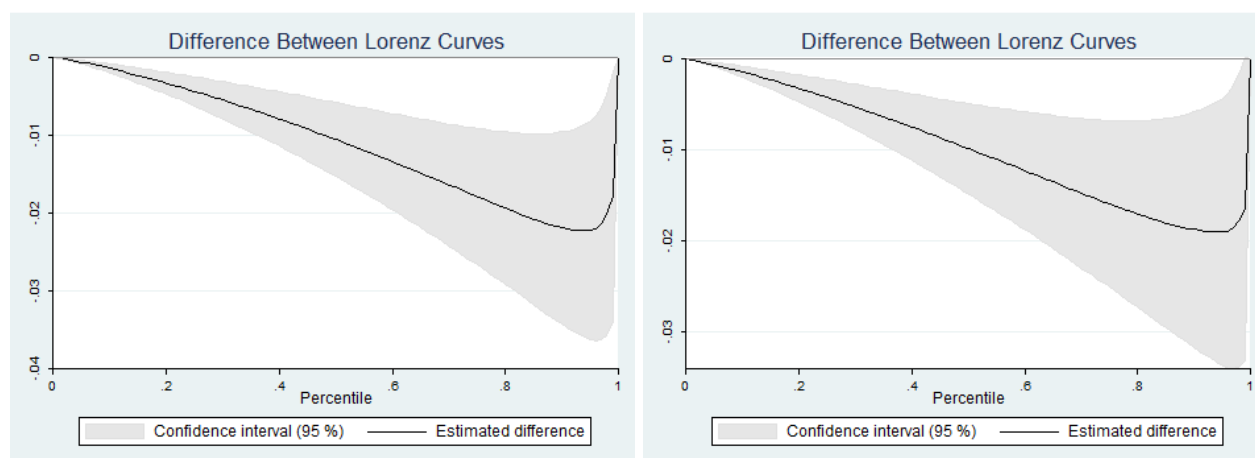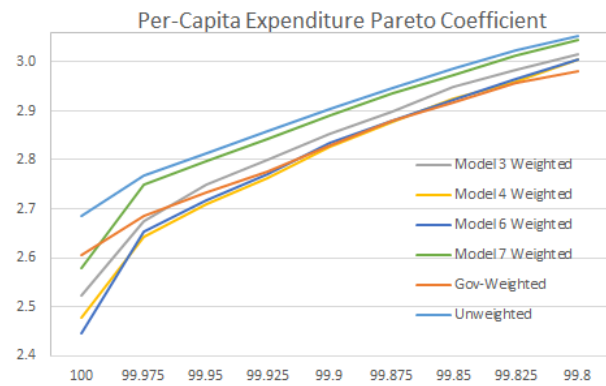

(a) Unweighted vs. Weighted (Model 4)   (b) CAPMAS-Weighted vs. Weighted (Model 4)
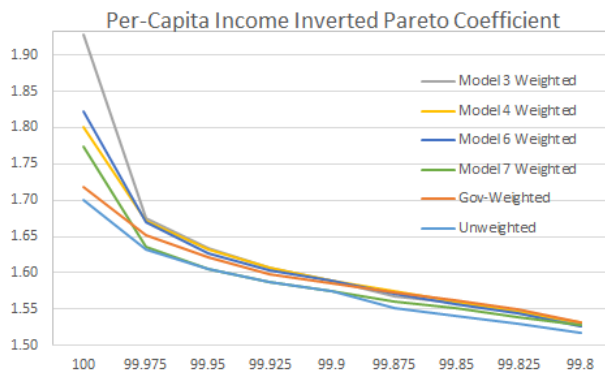
Figure A3. Pareto and inverted Pareto coefficients for income and expenditure per capita in trimmed distributions (Models 3-7)
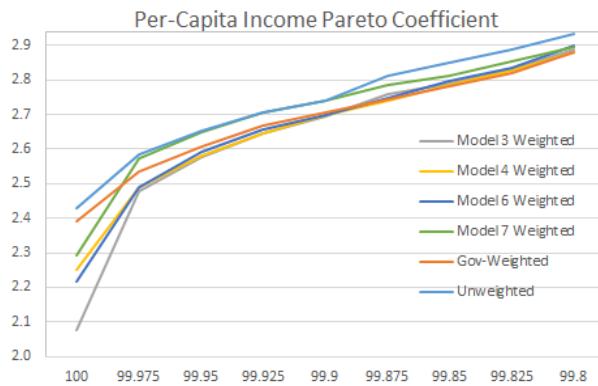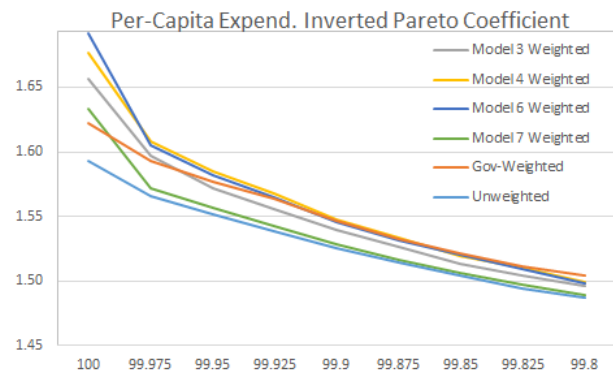

     (a) Pareto coefficient for income per capita   (b) Pareto coefficient for expend. per capita

 (c) Inverted Pareto coefficient for income per capita   (d) Inverted Pareto coef. for expend. per capita

Note: '100' indicates full, untrimmed income distribution. '99.975' indicates income distribution with the 0.025% households with the highest incomes trimmed (12 households in the 100% sample of the 2009 HIECS). Similarly, '99.8' indicates the trimming of 0.2% of highest-earning households (96).