



Missing Top Income Recipients

Martin Ravallion¹ 

Received: 2 June 2021 / Accepted: 19 February 2022 / Published online: 19 March 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Low response rates among rich households are thought to be a serious problem in many applications using household surveys. The paper discusses the various ways the problem can be dealt with, and makes some recommendations for practice, including in developing countries. Under certain conditions, income-selective non-compliance with an initially randomized assignment can be corrected by reweighting the data. This requires that the surveys pick up at least some top incomes. If not, then income tax records can help, including in estimating distributional national accounts. However, tax data come with their own concerns including tax avoidance/evasion, weak coverage of informal sectors and illicit incomes, and concerns about construct validity, given the limitations of taxable income as a basis for inter-personal comparisons of economic welfare. An appropriately weighted survey-based distribution of an acceptable measure of economic welfare need not be less reliable for most purposes of distributional analysis than income-tax records, including in combination with surveys. The choice will depend on the question to be addressed, and country-specific circumstances. These measurement issues warrant further research across multiple settings.

Keywords Inequality · Top incomes · Surveys · Nonresponse · National accounts

JEL Codes D63 · I32 · H31

1 Introduction

Household surveys have long been used for measuring poverty and inequality, and for related tasks in distributional analysis including policy evaluation. The sampling theory developed in

For helpful comments on an earlier draft the author thanks James Boyce, Gero Carletto, Lucas Chancel, Denis Cogneau, Benoît Decerf, Francisco Ferreira, Robert Groves, Stephen Jenkins, Dean Jolliffe, Anton Korinek, Michael Lokshin, Salvatore Morelli, Ercio Muñoz, Thomas Piketty and Dominique van de Walle.

✉ Martin Ravallion
mr1185@georgetown.edu

¹ Department of Economics, Georgetown University, and the NBER, Georgetown, Washington, DC, USA

the early C20th gave surveys a firmer statistical foundation for such purposes. Yet, in mid-C20th America, when Simon Kuznets wanted to measure income shares at the top end of the distribution he turned instead to income tax records and national accounts (Kuznets and Jenks 1953). In the new Millennium, tax records have been widely used this way, in what is now dubbed the “top incomes literature.”¹ Today we see the top end being studied with a very different data tool to that used for the bottom and middle.

The reliability of surveys for measuring top incomes is a longstanding concern. Those who agree to be interviewed may under-state their incomes or simply refuse to respond to a sensitive question on some income component. There are imputation/matching methods that can address such problems of “item nonresponse” (as it is referred to in statistics) by drawing on the questions that are in fact answered.² It is not obvious a priori that surveys face a bigger problem of income under-reporting by the rich than tax-records. While income misreporting to avoid taxation typically comes with penalties, the monetary incentive for the rich to hide incomes from the tax office is also greater than for surveys.

This paper focuses on another concern. Some proportion of sampled households do not participate in a survey; they may refuse to do so, or it may be difficult to find anyone at home. This is often referred to as “unit nonresponse” in statistics. Unit nonresponse rates have generally been rising over time and rates of 20% or more appear to be common.³ Surveys often try to avoid unit nonresponse, such as by using “call-backs” to nonresponding households and paying fees to those who agree to be interviewed. Nonetheless, some nonresponse is practically unavoidable. It is another matter whether it creates a bias in distributional analysis using surveys.

The specific influence of income on the probability of responding makes the topic especially relevant to the measurement of poverty and inequality, and distributional analysis in economics more broadly, including in studying fiscal incidence. Naturally, if the distribution of income is the same among responders and nonresponders then the measures will be unaffected. That would be the case in expectation if nonresponse were random. Then it is not a serious concern for bias. However, as textbooks on non-sampling errors warn, unit nonresponse is unlikely to be random.⁴ A number of observers have questioned the claims made about inequality, and how it evolves over time, based on household surveys with less than 100% response rates, also noting the discrepancies with other data sources.⁵

The literature has long recognizing that rich people are less likely to participate in surveys, raising concerns about bias in the initial sample-based estimates of the income shares. This is not to deny that there are also concerns about coverage of the poorest, some of whom are homeless and so less easily interviewed or even outside the sample frame. (Income under-reporting is also known to be a problem with regard to income sources that are important for poor people, notably farming.⁶) Nor should it be presumed that problems related to top incomes (whether survey compliance or under-reporting) imply that inequality is underestimated in surveys. The proportionate impacts of similar problems at lower income levels also need to be

¹ Piketty’s (2003) study of top income shares in France was influential in this literature, as was the review paper by Atkinson et al. (2011).

² The classic treatment of this topic is Little and Rubin (1987).

³ See the evidence for multiple countries in Luiten et al. (2020). Meyer et al. (2015) review response rates for major U.S. surveys and other threats to the quality of survey data.

⁴ For example, see the discussion in Groves and Couper (1998).

⁵ An early example, anticipating later issues in the literature, is Boyce (1993), with reference to inequality in the Philippines over 1960–1985.

⁶ See, for example, Carletto et al. (2021).

considered. While acknowledging these points, the focus here is on missing top income recipients.

There has been considerable effort and progress in addressing this issue over the last 20 years or so, and a review of where we are now seems warranted, especially given the interest in expanding applications to developing countries. The paper aims to assess prevailing practices in addressing unit nonresponse in the form of missing top income recipients. The context includes developing countries with weak statistical systems, as well as rich countries with “state-of-the-art” systems. The paper considers the obviously important concerns about internal validity for distributional analysis. The paper also discusses the (less familiar) issues of construct validity, which refers (in this context) to whether the correction method delivers a measure that is consistent with the concept of economic welfare—such as a comprehensive measure of real consumption per equivalent single adult—that one is interested in for making the inter-personal comparisons required for distributional analysis. Better measurement of the distribution of income is valuable, and the recent efforts in this direction are welcome. Nonetheless, there is undoubtedly scope for doing better, though that requires a critical perspective on current practices, both survey-based and tax-record based.

A theme of the paper is that thinking about behavioral responses helps to clarify the concerns about the non-behavioral accounting methods underlying current practices, and in thinking about better methods and measures. Modelling compliance requires identifying assumptions, and these are contestable, though certainly no more so than the assumptions made by ad hoc non-behavioral methods. Non-compliance can take the form of outright refusal or not being home for interviews. Either way, it is an outcome of choices made by those sampled.⁷ Thus, it is not unlike the activity choices that economists often model in other contexts.

To help motivate the discussion, consider the following behavioral model.⁸ The potential survey respondent must weigh the perceived benefits—the satisfaction of doing one’s civic duty, say—against the expected cost. Also, the longer you spend doing the survey, the more it starts to eat into other valued activities. Thus, we can expect that the marginal cost of survey participation rises with participation (such as measured by the time spent doing the survey). The cost (both total and marginal) is also likely to rise with respondent income; the opportunity cost of the time required to comply rises with income (due to higher wage rates); the time required for the interview probably also rises as it takes longer to interview people with more complex income streams. It seems reasonable to assume that the marginal benefit of participation is largely unaffected by participation, or at least does not rise with participation. Nor is the marginal benefit likely to rise with income. Under these assumptions, there is a unique optimal level of individual participation, equating marginal cost with marginal benefit. As income rises, desired participation falls. The rich will be less likely to comply than the poor, although this model does not tell us that the compliance problem is confined to the “rich.”

With this model in mind, the paper begins with correction methods during survey implementation. The paper then turns to *ex post* methods, both internal to the survey and using other external data sources, notably income tax records and national accounts. Along the way, the paper points to various ways that current practices might be improved.

⁷ It may sometimes be a choice by the interviewer; for example, if she does not want to approach people in some dwelling or area. A different behavioral model is needed for this case, though it is not considered further here.

⁸ This model can be extended in many ways, but a simple version suffices for this expositional purpose.

2 *Ex Ante* Correction Methods for Survey Data

There are two stages to the most common form of sample design. In the first stage, a listing is done of all the geographic areas where the population of interest is found, and a random sample of those areas is drawn—called the primary sampling units (PSUs). In the second stage, a listing is made of all the units (typically households) in each of the selected PSUs, and a random sample is drawn from each list. Some of those sampled cannot be interviewed—the occupants are never home at the right time, or they (implicitly or explicitly) refuse to be interviewed.⁹ (Refusal appears to be more common than noncontact.)

Statistics offices often try to correct for selective nonresponse in the second stage of the two-stage design. There are limits to how effectively this can be done. One approach is to encourage “call-backs,” i.e., interviewers returning to nonrespondents (sometimes in a structured way, called “double-sampling”). This requires that there is a non-selective desire to comply among those who did not do so before, which seems implausible.¹⁰ Another option is to look for “observationally similar” replacements for the non-compliers, but that is clearly challenging when the (observable or not) characteristics are covariates of compliance. A dual-frame design can also be used whereby the conventional street-address sample in the second stage is augmented by a sample of high-income households, for which the sample frame is another source, such as income-tax records.¹¹ Thus, the rich can be over-sampled, recognizing that they have a lower response rate in the single (initially random) sample.

Some major statistics offices have had a long-standing practice of paying survey respondents. This is invariably a fixed sum of money (given that tailoring the sum to the participants requires data one is unable to collect). However, once we recognize that survey participation is a matter of individual choice—nobody is obliged to comply with the statistician’s randomized assignment—such payments could well make the bias due to selective compliance even greater. A fixed fee paid to those who agree to participate will increase the probability of participation, but it can also increase the nonresponse bias. This happens under the (plausible) assumption that the payment provides a stronger incentive for the poor to participate. Payments may also undermine the perceived civic duty benefit of compliance. Thus, compensating survey participants may increase the sample size but reduce the researcher’s ability to draw valid inferences from that sample.

3 *Ex Post* Correction Internally to the Surveys

It will be obvious to any user of micro survey data for empirical analysis that it is desirable to correct for selective compliance internally to the survey data. Doing so can retain both the statistical integrity of the survey design and the great many applications for micro-data files in distributional analysis, including policy-related applications.

⁹ Here the focus is on selective compliance in surveys. Public access data sometimes come “top-coded” by the statistics office. Burkhauser et al. (2012) discuss this problem further and how it can be dealt with.

¹⁰ On reducing bias using call-backs see Groves (2006). Other methods found in practice are reviewed in Brick (2013), which includes references to the literature in statistics.

¹¹ For example, in estimating the distribution of wealth in the U.S., such a dual-frame method is used by the Federal Reserve’s Survey of Consumer Finances; see Kennickell (2008) for details.

Existing methods for *ex post* internal correction rely mainly on constructing a new set of weights, which are applied to the sample data with the aim of better representing the population.¹² These adjustments to the standard weights (reflecting sample design) are usually based on previously identified correlates of survey response rates, assuming uniform response rates within specified categories. A common method uses the empirically-observed response rates by categories of respondents defined by the observed data. Armed with group-specific predictor of the probability of compliance one can reweight the data to try to reduce bias.¹³

Naturally, these methods can only get us so far in dealing with income-based nonresponse, since income is unobserved for nonrespondents. Naturally, the reweighting is constrained by the available data on covariates of participation. A conditional independence assumption is made with regard to latent factors influencing participation. Given the importance of income as a (latent) determinant of compliance, the conditional independence assumption is implausible. Furthermore, nothing can be done if certain types of respondents never respond. (The discussion returns to this point.) Bias remains likely in existing reweighted estimates of the distribution of income based on surveys.

A micro-behavioral model—grounded in arguments or evidence about why there is non-compliance—can provide a useful extra tool for addressing selective response. The essential idea is that survey participation (like any activity choice) is an economic choice problem. As noted in the Introduction, there are benefits (such as doing one’s civic duty) and costs (the opportunity cost of time or concerns about revealing sensitive information) to survey participation. More research on these costs and benefits is needed, but here the discussion will rely solely on some seemingly defensible behavioral assumptions to illustrate the idea.

A Simple Expository Model Suppose that there are two income groups, “rich” and “the rest,” with means μ^R and μ^N (“N” for the non-rich), and two geographic areas, A and B with equal populations. A sample survey is done, with one goal being to determine how the population is split between the two income groups. To focus on the problem of selective participation, it is assumed that the survey gets μ^R and μ^N right. The challenge is then to get the correct population share of the rich (S), as required for both the overall mean, $\mu \equiv S\mu^R + (1 - S)\mu^N$, and for distributional analysis. The survey response rates by area are R_A and R_B respectively, which are data.¹⁴ The estimated shares of the populations of each area who are rich, based on the raw survey data, are denoted \widehat{S}_A and \widehat{S}_B , which can be calculated from the data. The corresponding true values (denoted S_A and S_B) are unobserved. Also, while the response rates by area (R_A and R_B) are data, that is not true of the response rates by income group (R^R and R^N).

One might guess that there is little or no hope of knowing the true values of S_A and S_B and hence $S \equiv (S_A + S_B)/2$. Yet, armed with only the data on R_A , R_B and \widehat{S}_A , \widehat{S}_B , one can infer the true values if we make two assumptions. The first is that all the rich have the same response rate (R^R) and all the non-rich have the same response rate (R^N) and that $R^N > R^R > 0$.¹⁵ The

¹² These are sometimes referred to as “nonresponse adjustment factors,” to be applied to the regular sample weights (such as allowing for stratification in survey design). Expositions and discussions of the methods used in practice can be found in Groves (2006) and Heeringa et al. (2017).

¹³ The literature on such re-weighting methods goes back to at least Deming (1953).

¹⁴ In other words, the share R_A of those who were randomly sampled in area A were actually interviewed.

¹⁵ A bias also arises if $R^N < R^R$, but the reverse inequality is more plausible, and it is the focus here.

last inequalities are key. There are two aspects: that the rich have a lower response rate, and that at least some of the rich do in fact respond.

Second, it is assumed that the probability of compliance does not vary between areas A and B independently of income. Call this geographic independence. This is a key identifying assumption, as it precludes any independent way in which location (specifically the geographic area for which response rates are measured) matters to compliance. That assumption can be questioned; for example, there may be local “social” effects on the propensity to contribute to public goods, including surveys. (This should be manifest in other observable differences, so there is scope for using control variables in practice.)

Together these two assumptions constitute a behavioral model of compliance. The model is the following system of four nonlinear equations in the endogenous variables, R^R, R^N, S_A and S_B (collected on the left side) as functions of the observed variables, $R_A, R_B, \widehat{S}_A, \widehat{S}_B$ ¹⁶:

$$S_i R^R + (1 - S_i) R^N = R_i \quad (i = A, B)$$

$$S_i R^R / [S_i R^R + (1 - S_i) R^N] = \widehat{S}_i \quad (i = A, B)$$

The first pair of equations give the observed overall response rate by area, while the second pair gives the estimated population shares of the rich prior to any correction for selective compliance. By solving for R^R, R^N, S_A and S_B we will have fully corrected for the errors due to selective compliance in the sample survey. Nothing else is needed to back out the correct numbers.

Consider the implications for summary statistics of the distribution. Given that the rich are less likely to participate, we will underestimate their share from the survey data.¹⁷ The proportionate bias in estimating the population share of the rich is simply the ratio of the overall response rate to that for the rich ($S_i / \widehat{S}_i = R_i / R^R$). With two income groups, the Gini index is $G = S \left(\frac{\mu^R}{\mu} - 1 \right)$ so the implications of underestimating S are ambiguous (noting that, while $\mu^R > \mu$, underestimating S will underestimate μ). Lower S and/or G makes it more likely that the Gini index is increasing in S .¹⁸

To illustrate the potential bias, consider the following stylized example.¹⁹ A simple random sample is drawn of (say) 1000 households across A and B. The empirically estimated share of the population who are rich is 2.3%, and its standard error, based on the raw sample data (ignoring the selection bias) is 0.5%, implying a 95% confidence interval (CI) of (1.3%, 3.3%). The estimated population shares for the rich by area are $\widehat{S}_A = 0.4\%$ and $\widehat{S}_B = 4.2\%$, and survey response rates by area are $R_A = 90\%$ and $R_B = 86\%$. Knowing only these data, but invoking the behavioral model above, it can be readily shown that the true (unbiased) population shares of the rich $S_A = 1\%$ and $S_B = 9\%$. The true overall population share of the rich of 5% is markedly higher than the estimated rate of 2.3%, and well outside the latter’s

¹⁶ The model is from Ravallion (2016, Box 3.14) though no solution method is given there. An Addendum is available from the author outlining a numerical algorithm for solving the model.

¹⁷ In the special case in which there is no selectivity in survey responses—that rich and non-rich are equally likely to respond ($R^R = R^N$)—the solution is $\widehat{S}_A = S_A$ and $\widehat{S}_B = S_B$.

¹⁸ It is readily verified that $\frac{\partial G}{\partial S} > (<) 0$ as $\frac{G+S}{1-S} < (>) 1$.

¹⁹ For computational convenience (and without any loss), the following examples are constructed by reversing the model, by solving for R_A, R_B and $\widehat{S}_A, \widehat{S}_B$ as functions of R^R, R^N, S_A and S_B .

95% CI, as obtained when ignoring selection bias. This reflects the economic difference in response rates, which (on solving) turn out to be $R^R = 40\%$ and $R^N = 90\%$. Since it is assumed here that selection bias due to unit-nonresponse is the only problem (so mean incomes are measured correctly for each of the two income groups), the behavioral model implies that the true income share of the rich is more than double the survey-based estimate. The implications for the Gini index depend on the value of μ^R/μ^N .²⁰ A typical value for even relatively low inequality countries would be 10, implying that the true Gini index is 0.29 while the estimated value (ignoring selective participation) is 0.17. In a relatively high inequality, with $\mu^R/\mu^N = 20$, the true index is 0.46, as compared to its estimated value of 0.30.

Of course, with less difference in response rates between the rich and the rest the extent of the bias is smaller. All this example illustrates is that there is the potential for large biases in estimated distributional shares and measures of inequality due to selective response.

While this expository model is too simple to be useful in practice, it does point to some more general observations of wide relevance. First, if the probability of survey compliance falls as incomes rise there will be first-order dominance, implying that all population shares of the “rich” are underestimated, however one defines “rich.”²¹ (Of course, by the same logic, standard absolute poverty measures are over-estimated given first-order dominance.)

Second, lower participation rates for richer people do not automatically imply that inequality measures such as the Gini index are underestimated (though one frequently reads unqualified claims to the contrary in the literature). Lorenz dominance is not assured. This is obvious enough in the 2×2 case as long as the incomes of the poor and non-poor are unaffected, i.e., it is only the shares of the population in the two groups that one gets wrong; then the true Lorenz curve must intersect the estimated one.²² Thus, the qualitative implications for an inequality measure will depend on the precise measure that is used—how it weights different segments of the distribution—and on the data.²³

Third, the standard errors calculated from raw sample survey data may be very deceptive about the true imprecision, given the existence of the selective nonresponse that has not been properly accounted for in analyzing the sample survey.

A More General Approach The above approach to correcting the data can be generalized for estimating the entire distribution of income and estimate its parameters (such as the mean and inequality indices), allowing for selective compliance. Korinek et al. (2007) propose a method for doing so with multiple income groups and geographic areas. In the spirit of the simple 2×2 example above, the idea is to use the geographic distribution of survey response rates (the proportion of the original random sample for that area that agreed to be interviewed) to infer how the probability of agreeing to be interviewed varies with income and other covariates. One postulates a micro-level compliance function, giving the probability of responding if sampled, as a function of household income (and potentially other covariates). The probability is assumed to be positive across the full support of the actual distribution, including the rich.

²⁰ Note that the Gini index for two groups can be written as $G = S(1 - S)\gamma/(S\gamma + 1)$ where $\gamma \equiv (\mu^R/\mu^N) - 1$.

²¹ This uses a result on stochastic dominance for poverty measures found in Atkinson (1987).

²² This is obviously unavoidable in the case of two income groups, since underestimation of the population share of the richer group also implies that the income of the poorer group relative to the mean is lower. The problem arises more generally, as shown in Korinek et al. (2006).

²³ For example, Atkinson (2007, Figure 2.1) reports co-movement over time between the Gini index for the UK and the income share of the top 1%, although there were still some time periods when they moved in opposite directions.

(This is the assumption that $R^R > 0$ in the 2×2 expository model of the previous section.) On top of this, it is assumed that the specific sample drawn includes at least some of the rich (even though very few may be found in the final sample). Thus, the sample distribution has the same support as the true distribution. We can call this common support.

One then asks what the compliance function would need to be to generate the observed response rates across areas. Under the identifying assumptions one can infer the compliance function. Notice that it is not the response rates by area that one uses to do the re-weighting; rather, one is finding the latent individual probability of compliance as a function of individual income (and other covariates) that best fits the observed aggregate data. Thus, one obtains a much finer correction for selective nonresponse by income.

The intuition for the estimation method for the micro-response function is as follows: the empirical relationship between the geographic distribution of response rates and the (measured) geographic differences in mean incomes is the first clue to the underlying relationship between individual response rates and individual incomes—the compliance function that one needs to reweight the data. It is only the first clue, however. Once one reweights the data using that clue, one has a new estimate of the probability function. One can then proceed iteratively to find the function most consistent with the data. Korinek et al. (2007) show that this problem can be solved in one step, using a GMM estimator for the parameters of the micro-level compliance function. Once one has this function, one can proceed as in past methods of correcting for selective compliance, by reweighting each observed individual by the probability of that type of individual responding.

Using the geographic spread of response rates in modelling compliance with randomized assignments in the CPS, Korinek et al. (2006) find that the probability of being interviewed falls steadily as income rises in the U.S., from 95% or higher for the poorest decile to only 50% for the richest, and under 20% among top incomes. Thus, the observations one has on rich households need to be weighted up quite a lot relative to those for poor ones.

Korinek et al. (2007) compare their weights to those used in the CPS to address nonresponse bias and find that the CPS does not adequately adjust for the (strong) income effect on survey nonresponse implied by the behavioral model. The CPS method does not deal with the (strong) income effect on response implied by the simulation method used by Korinek et al.²⁴

Korinek et al. (2006) show that their reweighting adds around 0.05 to the Gini index of income inequality in the U.S., when compared to the un-adjusted CPS data. More recent estimates by Morelli and Muñoz (2019) indicate larger upward revisions; for the U.S. over the period 2016–18, the mean CPS Gini index is 0.46 without correcting for selective compliance, but rises to 0.53 with the corrections (using the Korinek et al. method). The same study finds that the income shares of the top 1% are underestimated (8.4% on average over 2016–18 without correction, versus 15.2%) and the poverty rate is overestimated (12.9% versus 11.3%).

The bias in CPS inequality measures and top income shares implied by the Morelli-Muñoz estimates has tended to rise over time, though it has been more stable for the poverty rates. Thus, higher inequality tends to come with larger underestimation of the extent of inequality due to selective compliance with surveys.

²⁴ Prior to the mid-1990s, a contributing factor was the truncation methods used by the CPS to preserve the confidentiality of top-income respondents in the public access files. This practice further diminished our ability to accurately measure the high end. Since then, top incomes have been swapped above a threshold, so the marginal distribution should not change much.

Recommendations The data requirements for re-weighting are not especially demanding. The key ingredient is the distribution of response rates by geographic area. Any statistics office should be able to provide this data, ideally at the level of the PSU. However, it appears that such data are not routinely available, or only at a high level of geographic aggregation. For example, these are available for the U.S. Current Population Survey (CPS) but at state level. All surveys should provide an identifier at the unit-record level indicating whether that sample observation is from the first draw or a replacement.

Nor is the Korinek et al. (2007) method computationally difficult, especially given the introduction of the KMR command in Stata, as developed by Muñoz and Morelli (2021). The latter paper describes the program, how to use it, and provides an application.

Partial information on non-compliers can also help in further reducing bias due to selective compliance. For example, one can use data collected on the number of call-backs, or observations from the listing data from the second stage of the two-stage sampling can help in modelling compliance choice for those who are randomly sampled in the PSU.

The listing data can provide covariates for predicting compliance. Statistics offices could make more use of the listing data for this purpose, including in collecting data on observable dwelling attributes. Even more data might be available in some cases. For example, the samples for China's Urban Household Surveys (done by the National Bureau of Statistics) are drawn from a first, very short, survey with high response rates. This allows one to model compliance choice since the final sample is drawn for the full, much longer, survey.²⁵

4 *Ex Post* Correction Using Other Data Sources

Given the scope for re-weighting survey data to better represent the missing top-incomes, the case for turning to non-survey data on the rich must rest on a view that surveys exclude them—that common support fails. The top incomes literature has assumed this (at least implicitly). However, evidence to support that assumption is scarce.

An influential study by Székely and Hilgert (2007, Table 3) compared the highest reported incomes in surveys for 18 countries in Latin America with the known salaries of managers in medium to large firms (based on independent business surveys). For four countries, the highest recorded income in the survey was less than that of the managers. They also calculated the mean income of the 10 richest households in each survey, which they find to be less than that of the manager for 10 of the 18 countries. Many of these surveys in Latin America do not appear to be representing the rich adequately to permit re-weighting. However, there do not appear to be any other studies one can point to. The types of calculations done by Székely and Hilgert are relatively straightforward, and their study could easily be done in other regions and time periods.

It is surely surprising that the literature on top incomes using non-survey data has boomed with so little evidence on whether common support holds, such that survey data could be adequately re-weighted. This aspect of the foundations of this literature seems weak. More research is clearly needed on this issue, and on the differences between countries at different stages of economic development.

²⁵ Shaohua Chen and the author worked with NBS to develop a methodology for this purpose, which they are now implementing nationally.

Income-Tax Records The concern that under-reporting and selective compliance have led to an under-estimation of “top incomes” in surveys has re-activated interest in the use of external data drawn from income tax records, following Kuznets and Jenks (1953).²⁶ As already noted, such data can help *ex ante* in improving survey-based estimates of the distribution by allowing an over-sampling of the rich. More effort should go into further developing and applying such methods. Here we turn to *ex post* methods using data sources external to the survey.²⁷

The arguments for using tax records to supplement surveys are not as clear-cut as might be thought. As noted, from the evidence available, the common support assumption for earnings appears to be reasonable for the U.S. (where a lot of the work on using tax records at the top has been done), so re-weighting cannot be ruled out as an option for this reason, at least for labor earnings. (We do not have appear to have similar evidence for earnings from capital, however.) The comparisons by Yonzan et al. (2020) of survey data with tax records for the U.S. suggest that the two sources only start to diverge appreciably at the very top—around the top 1%. It is the top end where tax records have a role. When the question being addressed relates to the “bottom 99%,” there is no clear case for using tax records when good survey data are available.

Penalties for misreporting taxable income may encourage fuller and more accurate compliance among the rich than one would expect from a household survey. However, rich respondents will weigh such penalties against the much larger monetary incentives to understate taxable income than is the case with surveys. Anecdotally, any minimally adequate accounting of consumption in a household survey—or even relevant observations from the listing stage, such as on housing attributes—would probably provide a better indication of Donald Trump’s standard of living than his reported taxable income.

Access to income tax data—even as aggregate tabulations—is also a concern for much of the developing world. For Sub-Saharan Africa, such data are only currently available for a few countries (Côte d’Ivoire, Senegal, South Africa, to my knowledge).

While it is widely acknowledged that income tax coverage is generally weak for people with low incomes, it is often assumed, often implicitly, that income tax records (when available) adequately represent the top decile (say).²⁸ This has been questioned in the light of the revelations of the extent of off-shore tax evasion at the very top of the distribution, based on leaked records from facilitating financial institutions (such as the famous Panama Papers) (Alstadsæter et al. 2019). The extent of illicit external capital flows must give one pause in assessing top incomes from any source. Ndikumana and Boyce (2019) provide estimates of the extent of capital flight from selected countries in Africa, which imply a substantial loss of taxable income, with the loss occurring at both the time the income is earned, and from the (untaxed) return on the accumulated capital, presumably held in foreign bank accounts.

The reliability of tax records for this purpose also depends on the coverage of the income-tax system. This is not only about vertical coverage (between “rich” and “poor”) but also horizontal (among the “rich”). Even when the money stays in the country, it can be hidden from tax authorities. This is a special concern in developing countries with weak enforceability

²⁶ Examples include Feenberg and Poterba (1993), Piketty (2003), Banerjee and Piketty (2005), Atkinson et al. (2011), Piketty et al. (2018) and Bollinger et al. (2019).

²⁷ Though beyond the scope of this paper, it should be noted that interpolations are often required because public tax tabulations do not give the upper 1% (say). Pareto’s Law has been widely used for this purpose. Open top intervals in the tabulations mean that interpolation methods can matter. See Piketty (2003) and Jenkins (2017) on the use of Pareto’s Law for this purpose and Atkinson (2007), who discusses options. Burkhauser et al. (2018) use instead cell-mean imputations at the top.

²⁸ See, for example, the discussion in Piketty and Saez (2007, Section 5.2) writing about inequality in the U.S.

of tax laws and large informal sectors. For example, only 4% of India's population in 2018/19 filed an income tax return and only 60% of those paid any tax (Economic Times 2020). This would not be worrying for estimating the upper tail if we knew that the 2.4% accounted for by income taxpayers adequately represented that tail, but that is far from clear in a country where half or more of national income is generated by the informal sector. While credible evidence is (naturally) scant on this point, it would be hard to convince anyone who knows much about India that the incomes of the rich are captured well by the country's income tax records. Assuming that informal-sector incomes are adequately captured in national income aggregates (and that is an assumption) and that at least some of the rich are found in the informal sector (which seems plausible), the results of Banerjee and Piketty (2005) and Chancel and Piketty (2017) under-estimate the income shares of India's rich.

The use of income-tax records for addressing missing top incomes also raises issues of comparability across countries and over time. Tax policy changes can readily alter incentives to report income among the rich.²⁹ Not only do tax laws differ, but enforcement and compliance also differ in non-random ways. Withholding taxes can assure high coverage of labor incomes in the formal sector, but evasion is common in the informal sector and for other income sources in the formal sector, such as profits from enterprises.³⁰ With economic development, more of the economy becomes formalized, which entails greater administrative capability for raising revenues, including from direct taxation.³¹ The comparability problems are obvious.

It remains an open question how much the rise in the estimated top income shares based on tax records (as exemplified by Chancel and Piketty 2017, for India) is exaggerated by rising state capacity coming with economic development, also confounding inter-country comparisons. Suppose that the rich are found in both the formal sector (paying income taxes) and the informal sector. The national accounts pick up both sectors. Economic development comes with a rising share of the population in the formal sector. Suppose that the mean income of the rich is no different whether in the formal or informal sectors, and that economic growth is distribution-neutral, meaning that all incomes rise by the same proportion, whether rich or not. Then the estimated income share held by the rich, based on tax records, will automatically rise with economic development even if the actual share remains unchanged.

Furthermore, the degree of enforcement of tax laws may well reflect high-end inequality, as the rich use their power to make it easier for them to evade taxation by hiding (essentially informalizing) their activities. When inequality is high it is harder to know how high it really is.

Combining Surveys and Tax Records A promising new direction for research on this topic is found in the studies that have combined estimates of the upper tail based on income tax records with household surveys for the rest of the distribution. This can be thought of as an imputation problem, whereby top survey incomes are replaced with imputations consistent with tax records, as in Jenkins (2017) and Burkhauser et al. (2018) (using different imputation methods). One can also relax the common support assumption by using the tax records as a clue for revealing non-compliance among top-income recipients, so as to estimate the probabilities (conditional on income) that are required for reweighting the survey data; this approach is found in Medeiros et al. (2018) and Blanchett et al. (2019). A key issue is the

²⁹ For example, the 1986 Tax Reform in the U.S. appears to have encouraged high income tax payers to report higher incomes (Feenberg and Poterba 1993).

³⁰ See, for example, the evidence for the U.S. reported in Johnson and Rose (2019).

³¹ Indeed, following Besley and Persson (2011), one can think of the ability to collect income taxes as the main indicator of state fiscal capacity, which is itself a fundamental pillar of economic development.

choice of a merging point, above which one switches from the survey-based distribution to that from the tax records. Blanchett et al. (2019) set this point to assure a continuous function of the relative densities of observed and corrected incomes after reweighting.

The construct validity in using income tax records for distributional analysis, in combination with household survey data, requires more attention.³² A limitation of tax records is that “taxable income” need not accord well with the types of household consumption or disposable income measures normally preferred when using survey data for measuring poverty or inequality. A country’s tax laws need not treat everything that matters to a household’s economic welfare as taxable income.³³ Private and public transfers (including pensions) are typically excluded or only partly covered. Intra-household transfers are not, of course, included. (While we may have good reason to question the unitary model of the household, there can be little doubt that some income sharing occurs.) In countries that require individual tax filing, a high-income earner is treated as “rich” while her dependent spouse with no taxable income is deemed to be “poor.”³⁴ As noted, tax laws also change over time, and differ between countries, potentially confounding distributional comparisons.

Any measure of economic welfare has both an income numerator and a deflator (or a non-homothetic equivalent income function). The construction of that deflator has long been understood to matter to inter-personal comparisons. Tax records do not normally tell us much about households and their characteristics. This constrains the scope for adjusting for differences in family size or composition and differences in the local cost of living, as we would do with a household survey, let alone to calibrate more sophisticated equivalent income functions that aim to provide a monetary metric of utility. How far one can get in constructing defensible deflators depends on the tax system and the access to micro tax records, both of which vary across countries and over time.

When merging survey data for non-filers with commonly available summary tabulations of income tax records what one ends up doing is using the (micro) survey data to estimate an equivalent of individual taxable income for those with incomes that do not legally require income tax filing. Yet it would seem unlikely that an analyst using the survey would ever consider taxable income (as defined by local tax laws) to be a valid welfare indicator for measuring household poverty. Moreover, the merged distribution would probably overestimate the extent of economic inequality, notably because it ignores transfers.

Combining surveys with tax records can also throw new light on nonresponse patterns. Bollinger et al. (2019) were able to link CPS survey respondents to administrative data on taxable earnings; thus, they could calculate item nonresponse rates for earnings across the full support of earnings found in the administrative data. As expected, the nonresponse rate rises toward the top (and the bottom), but the rate does not go to zero; about 25% of the top percentile reported earnings in the survey.

National Accounts (NA) From what we know, it is evident that there are discrepancies between grossed-up survey-based aggregates (using standard weights) and NA data

³² Construct validity issues also arise when using surveys alone, with regard to the comparability of consumption and income measures across different surveys. These problems are reasonably well understood, however.

³³ In some cases, public reports on income taxes include “non-taxable income,” as reported in the tax filings; see, for example, Morgan (2017) using data for Brazil. This can help reduce one of the concerns about construct validity.

³⁴ Atkinson et al. (2011) discuss further the implications of individual versus joint filing for measuring the upper tail of the distribution of income.

(Ravallion 2003; Deaton 2005). This has led some researchers to scale up all survey-based income levels so that their mean accords with national income or private consumption per capita in the NA.³⁵

The obvious concern with this adjustment method is that those discrepancies need not be distribution-neutral (meaning that the relevant Lorenz curve is unaffected). As noted, the evidence we have suggests that selective compliance is decidedly non-neutral with regard to relative distribution, although the bias could go in either direction for standard inequality indices. The bulk of the discrepancy may well be at the top end of the distribution, such that survey-based measures underestimate measures of inequality but get poverty measures roughly right. For example, Lakner and Milanovic (2016) allocate the discrepancy entirely to the top decile (along with a Pareto interpolation). This is more plausible than assuming distribution-neutrality, although the “top decile only” assumption is hard to justify. As already noted, the problem of under-reporting and/or selective compliance does not start at some critical income level but reaches more deeply. Another approach is found in Chen and Ravallion (2010) who invoke a Bayesian argument to combine survey and NA-based measures in measuring global poverty, recognizing that there are errors in both data sources.

A new initiative under this heading is the idea of “distributional national accounts” (DINA) (Alvaredo et al. 2016; Piketty et al. 2018; Chancel et al. 2019). This is an ambitious effort in using various data sources, including income tax records but also survey data and other sources, to estimate distributions for national income as reported in the NA. The income tax records can be useful for this purpose, as demonstrated by Piketty et al. (2018) for the U.S..

DINA has forced a new effort to take account of non-taxable income sources, such as certain public transfers and non-taxable capital income. Of course, doing so requires extra assumptions. Here we can be talking about large shares of income that are not taxable, and this is not only for the non-filers that one expects to pick up adequately from surveys. Even in the U.S., about one third of labor income is not taxable, and nor is about two-thirds of capital income (Piketty et al. 2018). Thus, assumptions are required for distributing quite large shares of these income sources. For income from capital, which is highly unequally distributed, the assumptions made can undoubtedly make a non-negligible difference. The study by Piketty et al. (2018) is very clear about its assumptions, but it is also clear that they are contestable and in ways that have not yet been fully explored. For example, Piketty et al. scale up reported capital income in the tax records to match that in the NA, assuming that their relative distribution is the same. Yet (as they note, though dismiss), there are reasons to imagine that the discrepancy may not be distribution-neutral.

Assumptions are also required on tax incidence. Piketty et al. (2018) assume that taxes have no impact on national income or how it is allocated between capital and labor. For example, payroll taxes are borne entirely by workers. These are strong assumptions.

There are further uncertainties in how to distribute the benefits from public spending. Non-behavioral incidence analysis has long been popular, and is computationally straightforward (though not without its challenges).³⁶ A number of (longstanding) objections have been raised, especially (though not only) in the context of developing countries (van de Walle 1998). Behavioral responses to public spending can readily cloud inferences about incidence.

Studies of specific social programs have also revealed that the formal assignment mechanism (as indicated by administrative dictates) can differ greatly from what happens on the ground, as

³⁵ See, for example, Bhalla (2002) and Sala-i-Martin (2006).

³⁶ A useful compendium of papers on this method can be found in Lustig (2018).

revealed by surveys and qualitative observations. In the case of the largest cash transfer program in the world, China's *Di Bao* program, the formal incidence implies a 100% marginal tax rate on recipients, in that the transfers are intended for those living below a stipulated *Di Bao* line, and are intended to bring everyone up to that line. Yet micro-econometric analysis of survey-based evidence indicates much lower marginal tax rates in practice (Ravallion and Chen 2015). Assigning benefits according to existing administrative rules can give results that are far from the truth. On top of this, there are many categories of public spending that have no obvious assignment mechanism. It is a wild guess how those benefits are distributed.

How best to treat capital gains has also been an issue. In keeping with the concept of "national income" in the NA, Piketty et al. (2018) exclude capital gains (asset price changes). Larrimore et al. (2021) adopt a broader definition of "income" that includes capital gains and (under certain assumptions) find substantially larger gains in median income in the U.S. over recent decades. Much of this appears to be due to house-price inflation. The Larrimore et al. results also indicate substantially lower gains in top income shares than found in the work of Piketty and colleagues. Whether one accepts that unrealized capital gains should be included as income remains an open question. However, the contrast between the Larrimore et al. results and those of Piketty et al. must give us some pause to consider that question further.

All these methods for imputation using non-survey data require assumptions, some of which we have limited ability to assess at this stage. There is much appeal to the DINA idea, especially in the context of opening up the scope for more distributional analysis of longstanding macroeconomic and public finance issues. However, this must be considered a provisional exercise. Protocols will no doubt improve in the future, in the light of better information and more defensible assumptions, and a better understanding of the sensitivity of results to changes in those assumptions. The better information may well include better weighting methods for survey data recognizing the continuing problem of selective compliance.

The NA-based estimates of the distribution of income also beg the question as to whether the NAS aggregates are any more reliable than the survey-based aggregates. Even without errors in either data source, they are not measuring the same thing. In other words, there is also a concern about the construct validity of methods that try to "correct" surveys using the NA.

On top of this, errors are also a concern about the NA. There are the aforementioned problems of the hidden incomes, including capital held in offshore tax havens and so missing from the NA.³⁷ Even aside from this problem, many guesses go into the calculation of NA aggregates for many countries, not least poor countries with weak statistical systems, and poor compliance with international recommendations for the NA. Serious weaknesses have been identified in the NA for various countries in Sub-Saharan Africa.³⁸ At the same time, a number of countries in Africa have made great progress in the design and implementation of good quality household surveys, especially in the measurement of household consumption.³⁹

Recommendations External sources can usefully complement survey data on top incomes. Sources such as income-tax records can aid imputations and help in devising better weights,

³⁷ Zucman (2013) estimates that wealth hidden offshore accounts for about 10% of world GDP.

³⁸ In the context of Africa's NA, see Jerven (2013) and Devarajan (2013); the latter paper reports that only half of the countries in Sub-Saharan Africa have implemented the recommendations of the UN System of National Accounts for 1993, with many using methods dating back to the 1960s.

³⁹ Examples that stand out (based on experience, the literature and observations by past colleagues at the World Bank) are Ethiopia, Ghana, Kenya, Malawi, Nigeria, Rwanda, Tanzania and Uganda. Of course, many data problems remain, but the progress is undeniable.

especially when common support does not hold at the high end. However, there are limits to how much income tax records can help, especially in countries with large informal sectors (including top incomes). The construct validity issues in using external sources have received too little attention, especially in normative measurement applications.

Merging survey data with tax records would clearly be easier if survey respondents were asked for their unique tax identification number. This would be an important asset for distributional analysis, including policy applications. However, one cannot be optimistic about compliance with that question, especially given the risk of identity theft. (Personally, I would not be willing to give my Social Security Number to an interviewer at my front door, and I am sure I am not alone.) Progress in assuring confidentiality would clearly help here.

Unlike the methods that work internally with the surveys, there is a question begging about how to calculate the variance of estimates based on merging with external sources, such as data from income tax records. Further work is needed on this issue.

DINA is another promising step in the recent literature. However, the quality of NA data, especially in developing countries, warns us against treating those data as the benchmark. I would be good to see a complete inventory of compliance with the recommendations of the *System of National Accounts* (UN 2008), though we know enough now to warn for caution. There are developing countries where the survey data may well be the benchmark for better national accounts.

5 Conclusions

The literature on top incomes is full of claims that (to paraphrase) “household surveys are not representative of the population.” Such claims are made in a casual way, which largely ignores sampling theory and the implementation efforts of survey statisticians. Nor is it evident in what sense the various “stitch up” methods for combining survey and non-survey data found in the literature on top incomes can be considered representative of the relevant population.

That said, it is plausible that top income recipients are harder to get to participate in surveys—with a potential for bias due to selective compliance. This is well recognized in principle by survey statisticians, but practice in implementing corrections is another matter. Unless the survey weights properly reflect the likely income gradient in compliance, one expects an underestimation of top income shares. However, despite frequent claims to the contrary, selective compliance need not imply underestimation of the value of standard inequality measures; since one is moving densities, there will not (in general) be Lorenz dominance when comparing the estimated and true distributions. Empirical results for the U.S. do suggest that standard inequality measures are substantially underestimated. Indeed, the biases stemming from selective compliance in surveys (across all income levels) could well make nonsense of the degree of imprecision indicated by conventional calculations of standard errors that ignore the problem. It is hard to believe standard inequality measures and distributional shares from household survey data that have not been appropriately re-weighted. And the concern does appear to be greater among top income recipients.

An attraction of correcting for the problem of missing top incomes internally to the survey is that one can establish internal validity without confronting an issue of construct validity, i.e., that the concept of income or consumption considered appropriate for measuring poverty or inequality is not being compromised. Furthermore, we retain the survey data on a wide range of household characteristics for each household (alongside its income), thus retaining the many options for micro-empirical analysis.

Selective compliance does not on its own mean that we need to turn instead to non-survey data sources. A key issue is whether at least some of the rich participate in surveys, even if their response rate is lower. If this common support condition holds then (under other identifying assumptions), the survey data can be reweighted to eliminate the bias. While the top incomes literature has assumed (explicitly or otherwise) that common support fails, there is very little evidence one can point to at present to support that view.

While more evidence is needed on whether common support holds, if it does not then there is an important role for other data sources. The most promising source has been income tax records, in which the rich are supposed to comply on legal grounds, though still facing strong incentives for evasion/avoidance. The literature has found that tax records indicate higher top income shares than typically found in surveys. To some extent, this reflects deficiencies in how the survey weights are designed; the methods found in practice could probably be improved. A further reason for caution in making such claims is (again) their construct validity. There has been too little discussion of the limitations of individual taxable income as a monetary metric of welfare for distributional analysis, especially in normative applications.

Various methods of anchoring surveys to national accounts have also been used to try to improve survey-based estimates. The simplest method entails re-scaling all survey incomes to attain a mean income that is inferred from the national accounts. This is very hard to defend. More sophisticated approaches also take the national accounts data to be correct and use income-tax records to construct a distribution of national income as measured in the national accounts. One can question whether the national accounts provide more useful data for distributional analysis than found in surveys designed (in part) for that purpose. Even more importantly, however, there is ample reason to question any assumption that the discrepancies between the two data sources—in the aggregate or for components—would leave inequality unchanged.

The promise of credible distributional national accounts is great, but we have some way to go before prevailing methods can be considered credible in much of the world. The imputation methods require many assumptions, given that large shares of income are not legally taxable or are found in informal sectors, or foreign bank accounts, beyond the reach of the tax authorities. Users of these estimates do not, in my opinion, give sufficient attention to the underlying nitty gritty of measurement. The question is also left begging as to why we can ignore measurement errors in the national accounts, and this question must loom large as we take such methods to many developing countries, as is also a concern with regard to income tax records with limited coverage in such countries. There are some developing countries where I would be inclined to trust the available survey-based aggregates more than national accounts.

While such measurement issues need not prevent us making the best estimates possible from the data available, the validity of the data and methods can, and should, be challenged continually, and under public scrutiny. That is how we make progress in measurement. Surveys have both their strengths (such as in being able to cover the informal sector) and weaknesses (such as in finding weights that properly address selective compliance by the rich). New methods anchored to tax records offer much promise as a complementary data source, but also call for further critical attention, especially with respect to their construct validity and their credibility in developing country settings. The promise is great, but there is clearly much more work to do.

How much all this matters to policy remains to be seen. Just as top incomes are hard to reveal for measurement purposes—and we may never know just how unequal income and wealth are distributed—they can be hard to tax for financing and redistributive purposes. The hope remains that progress in advancing measurement will also help foster better policies.

References

- Alstadsæter, A., Johannesen, N., Zucman, G.: Tax evasion and inequality. *Am. Econ. Rev.* **109**(6), 2073–2103 (2019)
- Alvaredo F, Atkinson A, Chancel L, Piketty T, Saez E, Zucman G (2016) Distributional National Accounts (DINA) Guidelines: Concepts and Methods Used in the World Wealth and Income Database,” WID Working Paper 2016/1, Paris School of Economics. (An updated version, 2020, can be found here)
- Atkinson, A.B.: On the measurement of poverty. *Econometrica.* **55**, 749–764 (1987)
- Atkinson, A.B.: Measuring top incomes: methodological issues. In: Atkinson, A.B., Piketty, T. (eds.) *Top Incomes over the Twentieth Century*. Oxford University Press, Oxford (2007)
- Atkinson, A.B., Piketty, T., Saez, E.: Top incomes in the long run of history. *J. Econ. Lit.* **49**(1), 3–71 (2011)
- Banerjee, A., Piketty, T.: Top Indian incomes, 1922–2000. *World Bank Econ. Rev.* **19**(1), 1–20 (2005)
- Besley, T., Persson, T.: *Pillars of Prosperity: the Political Economy of Development Clusters*. Princeton University Press, Princeton (2011)
- Bhalla, S.: *Imagine There’s no Country: Poverty, Inequality and Growth in the Era of Globalization*. Institute for International Economics (Peterson Institute), Washington DC (2002)
- Blanchett, T., Flores, I., Morgan M.: “The Weight of the Rich: Improving Surveys Using Tax Data,” WID Working Paper 2018/12, World Inequality Lab, Paris School of Economics (2019)
- Bollinger, C.R., Hirsch, B.T., Hokayem, C.M., Ziliak, J.P.: Trouble in the tails? What we know about earnings nonresponse 30 years after Lillard, Smith, and Welch. *J. Polit. Econ.* **127**(5), 2143–2185 (2019)
- Boyce, J.K.: *The Philippines. The Political Economy of Growth and Impoverishment in the Marcos Era*, London, Macmillan and OECD (1993)
- Brick, M.: Unit nonresponse and weighting adjustments: a critical review. *J. Off. Stat.* **29**(3), 329–353 (2013)
- Burkhauser, R., Feng, S., Jenkins, S., Larimore, J.: Recent trends in top income shares in the United States: reconciling estimates from march CPS and IRS tax return data. *Rev. Econ. Stat.* **94**(2), 371–388 (2012)
- Burkhauser, R., Héroult, N., Jenkins, N.S., Wilkins, R.: Top incomes and inequality in the UK: reconciling estimates from household survey and tax return data. *Oxf. Econ. Pap.* **70**(2), 301–326 (2018)
- Carletto, G., Tiberti, M., Zezza A.: “Measure for Measure: Comparing Survey Based Estimates of Income and Consumption for Rural Households,” World Bank Research Observer, forthcoming (2021)
- Chancel, L., Piketty T.: “Indian Income Inequality, 1922–2014: from British Raj to Billionaire Raj?” Working Paper 2017/11, World Wealth and Income Database, Paris School of Economics (2017)
- Chancel, L., Cogneau, D., Gethin, A., Myczkowski A.: “How Large Are African Inequalities? Towards Distributional National Accounts in Africa, 1990–2017,” WID Working Paper 13/2019, Paris School of Economics (2019)
- Chen, S., Ravallion, M.: The developing world is poorer than we thought, but no less successful in the fight against poverty. *Q. J. Econ.* **125**(4), 1577–1625 (2010)
- Deaton, A.: Measuring poverty in a growing world (or measuring growth in a poor world). *Rev. Econ. Stat.* **87**(1), 1–19 (2005)
- Deming, W.E.: On a probability mechanism to attain an economic balance between the resultant error of response and the Bias of nonresponse. *J. Am. Stat. Assoc.* **48**, 743–772 (1953)
- Devarajan, S.: Africa’s statistical tragedy. *Rev. Income Wealth.* **59**(S1), 9–15 (2013)
- Economic Times: “The Budget 2020: the Deceptive Rise in India’s Income Tax Base,” The Economic Times 2020
- Feenberg, D., Poterba, J.: Income inequality and the incomes of very high-income taxpayers: evidence from tax returns. *Tax Policy Econ.* **7**, 145–177 (1993)
- Groves, R.: Nonresponse rates and nonresponse Bias in household surveys. *Public Opin. Q.* **70**(5), 646–675 (2006)
- Groves, R., Couper, M.: *Nonresponse in Household Interview Surveys*. John Wiley and Sons, New York (1998)
- Heeringa, S., West, B., Berglund, P.: *Applied Survey Data Analysis*, 2nd edn. Taylor and Francis, London (2017)
- Jenkins, S.: Pareto models, top incomes and recent trends in UK income inequality. *Economica.* **84**(334), 261–289 (2017)
- Jerven, M.: *Poor Numbers*. In: *How we Are Misled by African Development Statistics and What to Do about it*, vol. 46, pp. 336–340. Cornell University Press, Ithaca (2013)
- Johnson, B., Rose, P.: “Federal tax Compliance Research: Tax Gap Estimates for Tax Years 2011–2013,” Internal Revenue Service Research, Applied Analytics & Statistics, Washington DC (2019)
- Kennickell, AB.: “The Role of Over-sampling of the Wealthy in the Survey of Consumer Finances,” Paper presented at the [56th International Statistical Institute, Lisbon](#) (2008)
- Korinek, A., Mistiaen, J., Ravallion, M.: Survey nonresponse and the distribution of income. *J. Econ. Inequal.* **4**(2), 33–55 (2006)

- Korinek, A., Mistiaen, J., Ravallion, M.: An econometric method of correcting for unit nonresponse Bias in surveys. *J. Econ.* **136**, 213–235 (2007)
- Kuznets, S., Jenks, E.: *Shares of Upper Income Groups in Income and Savings*. National Bureau of Economic Research, Cambridge (1953)
- Lakner, C., Milanovic, B.: Global income distribution: from the fall of the Berlin Wall to the great recession. *World Bank Econ. Rev.* **30**(2), 203–232 (2016)
- Larrimore, J., Burkhauser, R.V., Auten, G., Armour, P.: Recent trends in U.S. income distributions in tax record data using more comprehensive measures of income including real accrued capital gains. *J. Polit. Econ.* **129**(5), 1319–1360 (2021)
- Little, R.J.A., Rubin, D.B.: *Statistical analysis with missing data*, 3rd edn. John Wiley, New York (1987)
- Luiten, A., Hox, J., de Leeuw, E.: Survey nonresponse trends and fieldwork effort in the 21st century: results of an international study across countries and surveys. *J. Off. Stat.* **36**(3), 469–487 (2020)
- Lustig, N.: *Commitment to Equity Handbook: Estimating the Impact of Fiscal Policy on Poverty and Inequality*. The Brookings Institution, Washington DC (2018)
- Medeiros, M., de Castro Galvao, J., de Azevedo Nazareno, L.: Correcting the underestimation of top incomes: combining data from income tax reports and the Brazilian 2010 census. *Soc. Indic. Res.* **135**(1), 233–244 (2018)
- Meyer, B.D., Mok, W.K.C., Sullivan, J.X.: Household surveys in crisis. *J. Econ. Perspect.* **29**(4), 199–226 (2015)
- Morelli, S., Muñoz, E.: “Unit Nonresponse Bias in the Current Population Survey,” mimeo, Stone Center, City University of New York (2019)
- Morgan, M.: “Extreme and persistent inequality: new evidence for Brazil combining National Accounts, surveys and fiscal data, 2001–2015.” World Inequality Lab Working Papers 2017/12, Paris School of Economics (2017)
- Muñoz, E., Morelli, S.: KMR: a command to correct survey weights for unit nonresponse using groups’ response rates. *Stata J.* **21**(1), 206–219 (2021)
- Ndikumana, L., Boyce J.K.: “Magnitudes and Mechanisms of Capital Flight From Angola, Côte D’Ivoire and South Africa.” PERI Working Paper 500, University of Massachusetts, Amherst (2019)
- Piketty, T.: Income inequality in France, 1901–1998. *J. Polit. Econ.* **111**(5), 1004–1042 (2003)
- Piketty, T., Saez, E.: Income and wage inequality in the United States, 1913–2002. In: Atkinson, A.B., Piketty, T. (eds.) *Top Incomes over the Twentieth Century*. Oxford University Press, Oxford (2007)
- Piketty, T., Saez, E., Zucman, G.: Distributional National Accounts: methods and estimates for the United States. *Q. J. Econ.* **133**(2), 553–609 (2018)
- Ravallion, M.: Measuring aggregate economic welfare in developing countries: how well do National Accounts and surveys agree? *Rev. Econ. Stat.* **85**, 645–652 (2003)
- Ravallion, M.: *The Economics of Poverty: History, Measurement and Policy*. Oxford University Press, New York (2016)
- Ravallion, M., Chen, S.: Benefit incidence with incentive effects, measurement errors and latent heterogeneity: a case study for China. *J. Public Econ.* **128**, 124–132 (2015)
- Sala-i-Martin, X.: The world distribution of income: falling poverty and... convergence. *Period. Q. J. Econ.* **121**(2), 351–397 (2006)
- Székely, M., Hilgert, M.: What’s behind the inequality we measure? An investigation using Latin American data. *Oxf. Dev. Stud.* **35**(2), 197–218 (2007)
- United Nations: *System of National Accounts*. United Nations (and Other International Agencies), New York (2008)
- van de Walle, D.: Assessing the welfare impacts of public spending. *World Dev.* **26**(3), 365–379 (1998)
- Yonzan, N., Milanovic, B., Morelli, S., Gornick, J.: “Drawing a line: comparing the estimation of top incomes between tax data and household survey data.” Stone Center Working Paper 27, City University of New York (2020)
- Zucman, G.: The missing wealth of nations: are Europe and the U.S. net debtors or net creditors? *Q. J. Econ.* **128**(3), 1321–1364 (2013)