

Indian income inequality, 1922-2014
From British Raj to Billionaire Raj ?

Lucas Chancel^{1,2}
Thomas Piketty^{1,3}

First draft: July 1st, 2017
Online publication: September 5, 2017
This version: September 7, 2017

Abstract. We combine household surveys and national accounts, as well as recently released tax data in a systematic way to track the dynamics of Indian income inequality from 1922 to 2014. According to our benchmark estimates, the share of national income accruing to the top 1% income earners is now at its highest level since the creation of the Indian Income tax in 1922. The top 1% of earners captured less than 21% of total income in the late 1930s, before dropping to 6% in the early 1980s and rising to 22% today. Over the 1951-1980 period, the bottom 50% group captured 28% of total growth and incomes of this group grew faster than the average, while the top 0.1% incomes decreased. Over the 1980-2014 period, the situation was reversed; the top 0.1% of earners captured a higher share of total growth than the bottom 50% (12% vs. 11%), while the top 1% received a higher share of total growth than the middle 40% (29% vs. 23%). These findings suggest that much can be done to promote more inclusive growth in India. Our results also appear to be robust to a range of alternative assumptions seeking to address data limitations. Most importantly, we stress the need for more democratic transparency on income and wealth statistics to avoid another "black decade" similar to the 2000s, during which India entered the digital age but stopped publishing tax statistics. Such data sources are key to track the long run evolution of inequality and to allow an informed democratic debate on inequality.

1 World Inequality Lab, Paris School of Economics

2 IDDRI - Sciences Po

3 Ecole des Hautes Etudes en Sciences Sociales

Corresponding author : lucas.chancel@psemail.eu

We thank Nitin Bharti for extremely valuable research assistance.

NSSO micro data for years 1983 and after obtained through CEPREMAP and PjSE, which we gratefully acknowledge for their support.

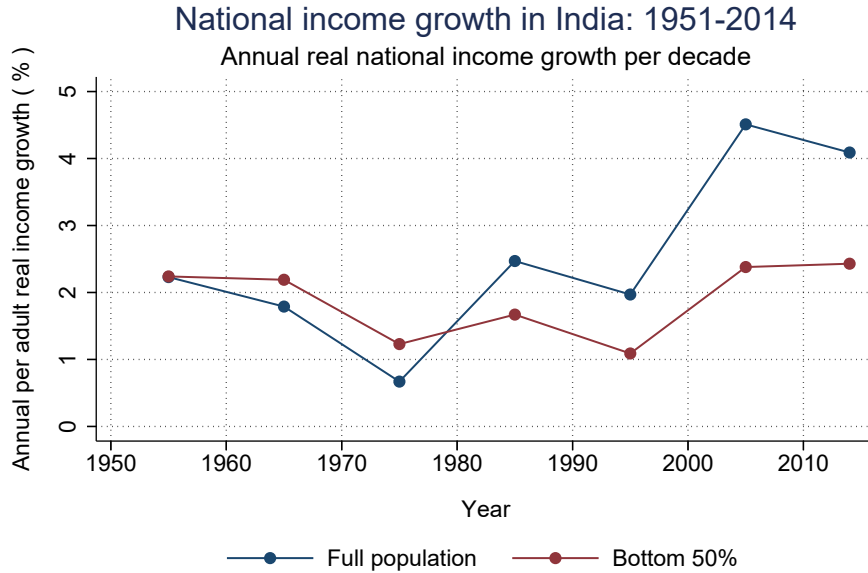
India introduced an individual income tax with the Income Tax Act of 1922, under the British colonial administration. From this date, up to the turn of the 20th century, the Indian Income Tax Department produced income tax tabulations, making it possible to track the long-run evolution of top incomes in a systematic manner. Using this data, Banerjee and Piketty (2005) showed that the share of fiscal income accruing to the top 1% earners shrank substantially from the mid-1950s to the mid-1980s, from about 13% of fiscal income, to less than 5% in the early 1980s. The trend was reversed in the mid-1980s, when pro-business, market deregulation policies were implemented. The share of fiscal held of the top 1% doubled from approximately 5% to 10% in 2000.

According to National Accounts estimates, post-2000 income growth has been substantially higher than in the previous decades. Average annual real income growth was below 2% in the 1960 and 1970s, it reached 2.5% in the 1980s and 2% in the 1990s¹. Since 2000s it is of 4.4% on average since 2000 (Figure 1). Little is known however on the distributional impacts of economic policies in India after 2000 in part because the Income Tax Department stopped publishing income tax statistics in 2000, and also because self-reported survey data does not provide adequate information concerning the top of the distribution (fiscal data is not perfect either, but it delivers higher and more plausible income levels for the top). In 2016, the Income Tax Department released tax tabulations for recent years (2011-12, 2012-13 and 2013-14), making it possible to revise and update previously published top income estimates and better inform public debates on growth and income inequality. We find that the bottom 50% group grew at a substantially lower rate than average growth (Figure 1a) since the 1980s. Middle 40% grew at a slower rate than the average (Figure 1b). On the contrary, top 10% and top 1% grew substantially faster than the average since 1980 (Figure 1c).

The first objective of this paper is to mobilize this newly released set of tax data in order to track the evolution of income inequality from 1922 to 2014. The second objective is to go beyond top income shares and produce estimates of income dynamics throughout the entire distribution using concepts that are consistent with National Accounts (following, as much as possible, the Distributional National Accounts Methodology, see Alvaredo et al., 2016).

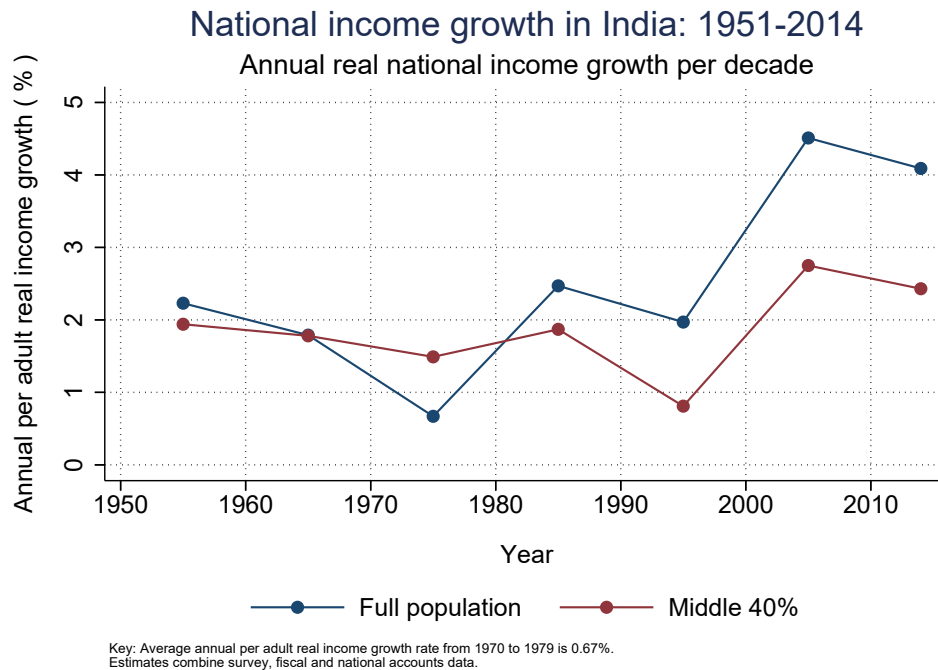
¹ Appendix 1 presents real per adult annual growth rates using GDP from United Nations National Accounts Database (used in this paper) and the World Bank Database.

Figure 1a - National income growth in India: full population vs. bottom 50% income group, 1951-2014



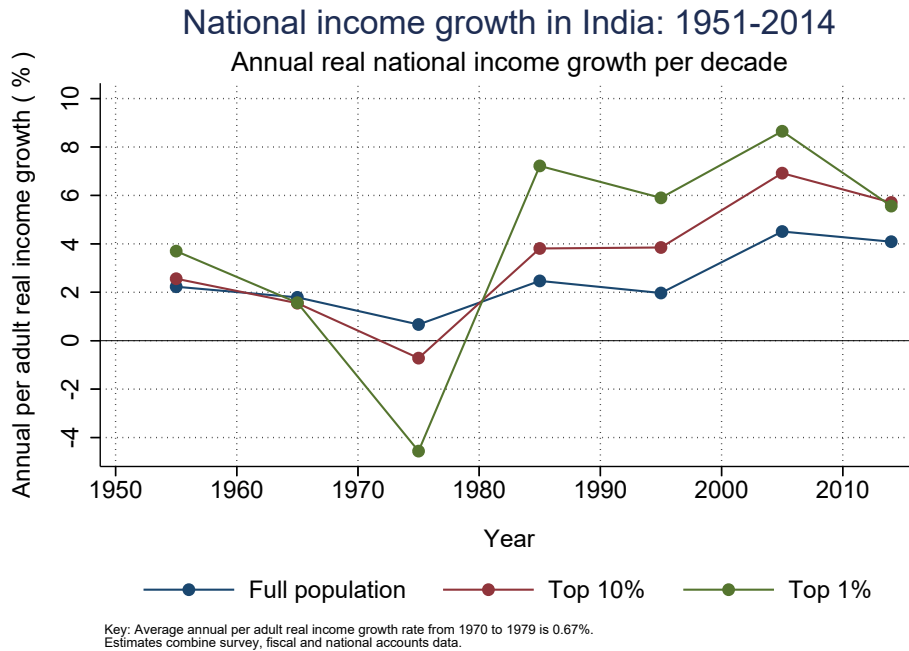
Source: Authors' computations using tax and survey data and national accounts.

Figure 1b - National income growth in India: full population vs. middle 40% income group, 1951-2014



Source: Authors' computations using tax and survey data and national accounts.

Figure 1c - National income growth in India: full population vs. top 1% and top 10% income groups, 1951-2014



Source: Authors' computations using tax and survey data and national accounts.

To do so, we combine in a systematic manner household survey, fiscal and national accounts data. Such an exercise is fraught with methodological and conceptual difficulties given the lack of consistent historical income inequality data in India. Indeed, the tax data available only covers the very top of the distribution of Indian earners (more than 6% of total population in 2014). In addition, the National Sample Survey Organization (NSSO) household surveys measure consumption rather than income. We repeatedly stress that there are strong limitations to available data sources, and that more democratic transparency on income and wealth statistics is highly needed in India. That said, we find that our key results are robust to a large set of alternative assumptions made to address data gaps. The present paper should be viewed as an exercise in transparency: we propose a method to combine the different available sources (in particular national accounts, tax and survey data) in the most possible transparent way, and we very much hope that new data sources will become available in the future so that more refined estimates can be constructed. All our computer codes are available on-line so that everybody can use them and contribute to improve the methods.

The rest of this paper is organized as follows. Section 1 discusses the Indian income inequality data gap of the past two decades, section 2 describes our

data sources and methodology, section 3 presents our key findings, section 4 briefly discusses their policy relevance and section 5 concludes.

1 ENTERING THE DIGITAL AGE WITHOUT INEQUALITY DATA

1.1 Economic policy shifts since the 1980s

Over the past thirty years, the Indian economy went through profound evolutions. In the late seventies, India was recognized as a highly regulated economy with socialist planning. From the 1980s onwards, a large set of liberalization and deregulation reforms were implemented. In this context, it is unfortunate that Indian authorities stopped in 2000 publishing income tax tabulations, which represent a key source of data to track consistently the evolution of top incomes.

Under Prime Minister Jawaharlal Nehru (in power from 1947 to 1964), India was a statist, centrally directed and regulated economy. Transport, agriculture and construction sectors were owned and administered by the Central Government, commodity prices were regulated and the country had important trade barriers. Nehru's followers, including Indira Gandhi's (1966-77 and 1980-1984) prolonged these policies and implemented a highly progressive tax system. In the early 1970s, the top marginal income tax rate reached record high levels (up to 97.5%).

From the mid 1980s onwards, liberalization and trade openness became recurrent themes among Indian policymakers. The Seventh Plan (1985-1990), led by Rajiv Gandhi (1984-1989), promoted the relaxation of market regulation, with increased external borrowing and increased imports. The tax system was also gradually transformed, with top marginal income tax rates falling to 50% in the mid-1980s. In the late 1980s, when India faced a balance of payment crisis, it called for International Monetary Fund assistance. Financial support was conditioned to structural reforms which pushed forward the deregulation and liberalization agenda.

What came to be known as the first set of economic reforms (1991-2000) placed the promotion of the private sector at the heart of economic policies, via denationalizations, disinvestment of the public sector, deregulation (dereservation and delicensing of public companies and industries)². These reforms were implemented both by the Congress government of N. Rao (1991-1996) and its successors, including the conservative Janata Party government of A. Vajpayee (1998-2004). The reforms were prolonged after 2000, under the 10th and

² Economic policies also sought to rationalize the public sector, its branches now had to pursue the objectives of profitability and efficiency. The opening of imports, exchange rate floating regime and banking, capital market opening were also implemented.

subsequent five-year plans. These plans ended government fixation of petrol, sugar or fertilizer prices and led to further privatizations, in the agricultural sector in particular.

The impacts of these reforms in terms of growth has been praised by public authorities. Real per adult national income growth, which has more sense from the point of view of individual incomes than commonly used GDP³, significantly increased after the reforms. It was 0.7% in the 1970s, 2.5% in the 1980s, 2.0% in the 1990s and 4.4% since 2000 (Figure 1). However, little is known on the distributional characteristics of post-2000 growth.

1.2 The income inequality data gap

Public debate over liberalization policies largely focused on their macroeconomic impacts (Ramaswami, Kotwal, Wadhwa, 2011) and on the impacts on poverty, with a substantial reduction in poverty rates⁴ (World Bank, 2017; Deaton & Dreze, 2002; Deaton & Kozel, 2005). How the Indian economy fared in terms of inequality has been arguably less discussed. This can partly be explained by a lack of consistent data on the distribution of incomes or wealth for the recent period.

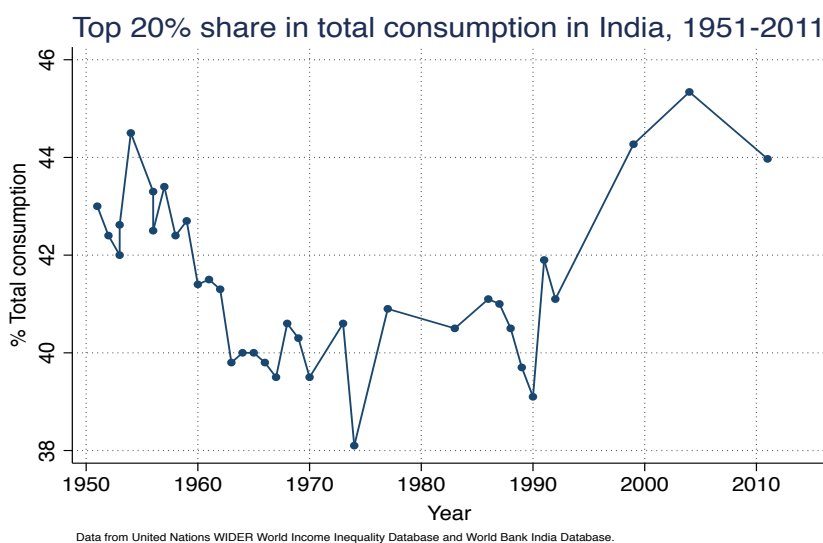
Some evidence suggesting a rise in income inequality in India after the turn of the century can however be found in NSSO surveys and other sources available in openly-available sources. Figure 2 presents the share of total consumption attributable to the top 20% of consumers, available online from the World Bank and United Nations WIDER World Income Inequality Database (UN-WIDER WIID). The data shows a decrease in top quintile consumption share from the fifties to the seventies from around 43% to 40% and an increase thereafter (in line with Banerjee and Piketty findings) to close to 44%. There are important irregularities with the data, but the overall "U-shape" trend seems relatively consistent⁵.

³ Net national income is equal to GDP minus depreciation of fixed capital plus net foreign incomes.

⁴ The share of Indians under the \$1.9 poverty line went from 45.9% in 1993 to 21.2% in 2011 (PovcalNet, 2017)

⁵ As discussed below, income surveys sources exist for 2005 and later years; in particular data from the National Council for Applied Economic Research (NCAER) and from the Inter University Consortium for Applied Political and Social Sciences Research (ICPSR). These data sources however do not enable comparison before and after 2000.

Figure 2 - Top 20% consumption share from NSSO surveys



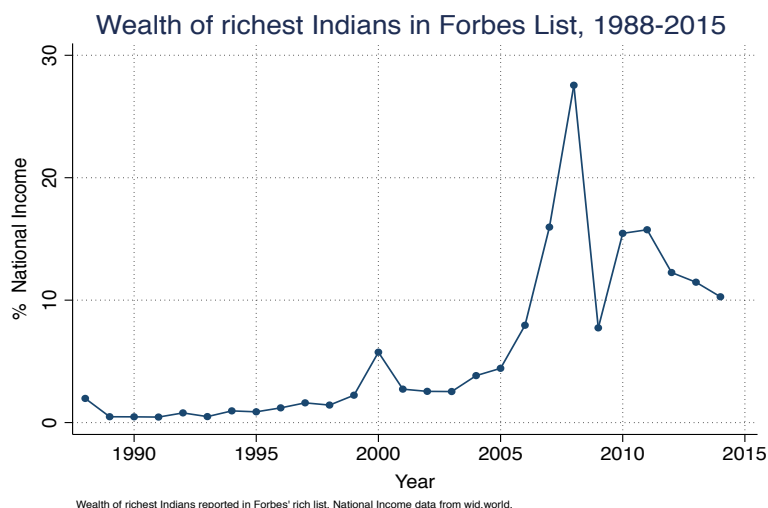
Source: Authors' computations using data from United Nations WIDER Income Inequality Database and World Bank India Database (based upon NSSO surveys)

The shortcomings of household survey data in monitoring the evolution of inequality are well known; because of underreporting and undersampling issues, surveys fail to properly capture inequality dynamics at the top of the distribution (Atkinson and Piketty, 2007, 2010). What is more, NSSO surveys only focus on consumption rather than income and the distributional dynamics of these two concepts can differ notably. In addition, the relatively limited magnitude of the changes observed in NSSO data calls for care in the interpretation of such results. Consumption data available through surveys constitutes part of the evidence, but are not sufficient to inform debates on Indian inequality.

Other data sources, such as Forbes' Indian Rich lists, suggest an important increase in the wealth of the richest Indians after 2000 (see

Figure 3). The wealth of the richest Indians reported in Forbes' India Rich List, amounted to less than 2% of National income in the 1990sn, but increased substantially throughout the 2000s, reaching 10% in 2015 and with a peak of 27% before the 2008-9 financial crisis. Such data suggests a rise in wealth inequality levels throughout the post-2000 period, but does not enable a consistent analysis of income inequality over the long run. This is confirmed by simple simulations using a fixed normalized wealth distribution and taking into account rising average nominal wealth over the period (unfortunately Indian wealth data is very limited so it is difficult to go further).

Figure 3 - Wealth of richest Indians in Forbes' Rich List



Source: Authors' computations based upon Forbes billionaire rankings and WID.world national income data.

The recent release of income tax tabulations by the Income Tax Department for the post 2011 period does, however, allow for a more consistent analysis of the dynamics of income in India since the turn of the century.

2 DATA SOURCES AND METHODOLOGY

We present the data used to produce series on the evolution of income for the entire distribution from 1951 to 2014 (period covered by both household surveys and tax data, as well as national accounts) and for the evolution of incomes of the top 1% share and above from 1922 to 2014 (period covered by tax data and national accounts only, with no survey data prior to 1951).

2.1 Description of the different data sources

2.1.1 Tax data

The Indian Income Tax Department released tax tabulations for the fiscal years 1922-1923 to 1998-1999, and interrupted the publication in 2000. After several public calls for more democratic transparency over Indian inequality data⁶, the ITA released tax tabulation for years 2011-12 to 2013-14. All these tabulations report the number of taxpayers and the gross and returned income for a large number of

⁶ See for instance <http://www.bbc.com/news/world-asia-india-36186116>

income brackets⁷. Gross income corresponds to pre-tax income before certain deductions are applied to compute returned income⁸. Tax units are defined as individuals or Hindu Undivided Families (HUF, family clusters allowed to file their income jointly). The number of HUF represented roughly 20 % of tax returns in the interwar period, 5% in 1990 and less than 2.5% in 2011.⁹

The exact reason why Indian tax administration stopped publishing data in 2000 remains unknown. One potential explanation for this is the change in the sampling method employed in the late 1990s, with a resulting loss in the precision of estimates. Indeed, official tax tabulations were based on the entire population until the early 1990s - or based on stratified samples with sampling rates close to 100 percent for top incomes as is the case in most OECD countries, but seem to be based on uniform samples of all tax returns after this period and up to 2000 (Banerjee and Piketty, 2005). The latter method led to less precise results¹⁰. Another potential explanation for the halt in tax reporting could just be the lack of interest in income statistics and inequality (which given the rise in top income shares observed from mid 1980s to 2000 seems rather surprising).

Interestingly enough, the number income tax payers in India has increased substantially over the past decades. Less than 0.5% of the population filing tax returns up to the 1950s, between 0.2 and 1% over the period between 1960 to 1990, before a substantial increase thereafter; from 1% to close to 3% in the late 1990s and more than 6% in the latest period (Figure 4)¹¹. This increase over twenty years is impressive, yet comparatively, the current figure is similar to the levels observed in France and in the USA in the late 1910s, and much lower than the levels observed in the interwar period (about 10-15%) and in the decades following World War 2 (50% or more) in these two countries (Piketty, 2001; Piketty and Saez, 2003). With revenues from income tax equivalent to approximately 2% of GDP, India receives more revenue than China (1%), but significantly less than other emerging countries

⁷ According to the Income Tax Department, a number of tax payers paid their taxes but did not file returns in fiscal years 2011-2013. In order to take into account these individuals, we assumed that they fell in the lower income tax brackets. We tested alternative assumptions: i) assuming they are fully representative of other income filers and ii) assuming they all fall in the lowest taxable bracket. These alternative assumptions have very limited impact on our final results. Minor corrections were also done to raw tax data and mainly pertain to the clubbing of brackets in some years as the average income was incompatible with the bracket they were categorized. In such rare cases, we club erroneous brackets in the lower bracket. Year 1997 was removed altogether, as data is erroneous.

⁸ Deductions are defined at chapter VI of the Income Tax Act. They include premiums of annuity plans, equity fund investments, medical or health insurance, certain forms of donations, etc. Focusing on gross income is more accurate in terms of pre-tax income and is also less impacted by changes in the definitions of deductions. Income losses (such as business income losses) have to be adjusted while computing Gross Total Income as per Income Tax law. Note that imputed rent for owner occupied dwellings were included in Income tax computations before 1986 and removed afterwards. More precisely, post 1986 tax data excludes imputed rent for first residence, but not for secondary residences.

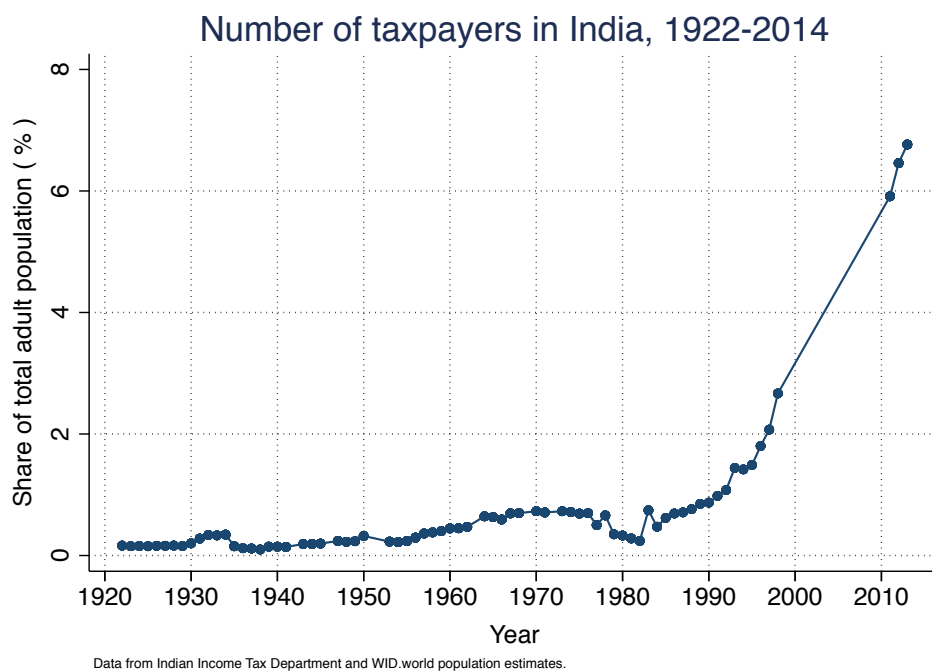
⁹ One should note that the Indian income tax data is entirely based upon individual income. This corresponds to equal-split income (ie. income shared among spouses) only if we assume that all tax-payers are either single or married to other tax-payers falling in the same bracket, which strictly speaking cannot be true. This implies that our estimates tend to over-estimate inequality as compared to the equal-split benchmark and to under-estimate inequality as compared to the individualistic benchmark. If and when we access to micro-level Indian tax data, we will be able to refine this analysis and compute separate equal-split and individualistic series.

¹⁰ For year 1997, see Appendix 2.

¹¹ This figure includes estimated tax payers who did not file returns post-2011. They represent approximately 30% of the number of tax payers, according to the Income Tax Department.

such as Brazil and Russia (4%), and South Africa and the OECD countries (9%) (OECD, 2017).

Figure 4 - Evolution of the proportion of income-tax taxpayers in India



Source: Authors' computations using data from Indian Income Tax Department and UN population data.

2.1.2 NSSO consumption data

The NSSO, led by the Ministry of Statistics and Program Implementation started an all-India consumer household expenditure survey (AIHS) after its independence in 1947. The first round of the AIHS was carried out in 1951 and surveys were then conducted on an annual basis. The size of rounds varies since the quinquennial AIHS has a larger sampling of about 120 000 households and five times less for smaller other rounds. The reach of the quinquennial survey is extensive in terms of consumption items (ranging from daily used food, clothing to durable goods and services such as construction, education and healthcare). NSSO surveys however do not measure individual or household incomes¹², in part because agricultural and business incomes are judged to be volatile and assumed to be much less reliably measured than consumption.

¹² The Employment Unemployment Surveys report wages for the working-age population, but other sources of income are not covered.

Since the first survey rounds, NSSO produced 30 days reference period estimates. This period is known as the Universal Reference Period. Post-1990, concerns were raised about the sensitivity of the reference period on the estimates and NSSO started publishing alternative reference periods (7 days and 365 days). As Deaton and Kozel (2005) note, shorter recall periods tend to lead to higher consumption estimates. However, experiments carried out with different reference periods by the NSSO working group concerned concluded that there is no clear superiority of a period over another. We thus use the Universal Reference Period. This choice is also motivated by the fact that the 30 days period is the only one that is consistent throughout the entire period of analysis (1951-2014).

For recent years (1983 to 2011) we use quinquennial rounds 38 (1983), 43 (1987-88), 50 (1993-94), 55 (1999-2000), 61(2004-05), 66 (2009-10). Micro data at the household level was obtained from the NSSO. For earlier rounds (rounds 3 to 32), for which we could not access micro data files, we use the Poverty and Growth in India Database of the World Bank (Ozler et al., 1996) which provides rural and urban per capita consumption tabulations for a dozen quantile groups for years 1951 to 1978. All rounds and corresponding years used are summarized in Appendix 3, along with the summary statistics of each round. We describe in section 2.2.2 the procedure used to infer the full distribution of income from these surveys and how we interpolate missing years.

2.1.3 National Accounts data

From 1950 to the present day, we use GDP data from WID.world, based on National Accounts Statistics (NAS) from 1971 to 2013, on World Bank (after 2013) and on Maddison (2007) from 1950 to 1970¹³. WID.world then performs its own computations to infer Net Foreign Income and Consumption of Fixed Capital (Blanchet and Chancel, 2016). Before 1950, we use historical National Income growth rates from Sivasubramonian (2000).

A well know puzzle in Indian statistics (Deaton and Kozel, 2005; CSO, 2008) pertains to the difference in survey consumption growth rates and national accounts growth rates, particularly during the recent period. Figure 5 shows the total growth rate of Net National Income and Household Final Consumption Expenditure from NAS and personal consumption from NSSO, from 1983 to 2011. According to NAS, national income grew at 475% and household consumption grew at slightly more than 300%, while NSSO data indicates that household consumption grew at 200%.

¹³ In the 1990s we observe noticeable differences between real GDP growth estimates obtained from UN SNA and those reported by the World Bank (see Appendix 1).

Figure 5 - Cumulated growth rates according to NAS and NSSO



Source: Authors' computations using national accounts and NSSO data.

Several reasons have been put forward to explain this gap, including (i) population coverage (it is different between NSSO and NAS, since Non Profit Institutions Serving Households and homeless individuals are not covered by NSSO surveys); (ii) valuation and integration of certain types of services in survey questionnaires (it was argued that the treatment of cooked meals served by employers to employees leads to underestimation of the total value of services consumed by households in the NSSO surveys (CSO, 2008) while other services such as financial intermediation that are particularly important among top earners, are not included in survey estimates (Sundaram and Tendulkar, 2005); (iii) imputed rents (while the NAS incorporates imputed rents, NSSO surveys do not¹⁴); (iv) consistency of National Accounts estimates (Kulshreshtha and Kar, 2005) ; (v) under-reporting and under-sampling of top incomes in survey data (Banerjee and Piketty, 2005). We should stress from the outset that we do not pretend to solve this complex issue. The divergence probably involves several, if not all of the factors above cited.

¹⁴ When correcting for imputed rents the Central Statistical Organization (2008) finds a large and growing share of total consumption remains unexplained.

What we seek here is to better estimate the fraction of the difference that can be explained by the absence of top earners in survey data. We do not think that this factor alone can explain the entire gap, as it has sometime been suggested (Lakner and Milanovic, 2015).

2.1.4 IHDS income and consumption survey

The Inter University Consortium for Applied Political and Social Sciences Research (ICPSR), based at the University of Michigan, provides access to the India Human Development Survey (IHDS), conducted in 2005 and 2011-12 among more than 40 000 households from rural and urban areas. The survey provides information at the household level on both income and consumption. Consumption related questions were designed so as to match the NSSO questionnaire, using similar item categories and similar referencing periods. The definition of income in the IHDS survey includes all sources of income: labour income (wages and pensions), capital income (rents, interests, dividends, capital gains) as well as mixed (or business) incomes¹⁵. Government benefits, reported in the survey, are excluded from the analysis for consistency with tax tabulations; our focus is pre-tax income.

The IHDS is one of the very few surveys estimating both consumption and income in India. This is particularly useful as it enables a tentative reconstruction of NSSO unobserved income levels, using IHDS information. We describe this methodology in section 2.2.2. IHDS micro data is also openly available via the ICPSR website, which makes it particularly convenient¹⁶.

2.1.5 UN statistics population data

We define the theoretical population of tax payers as the total number of adult individuals in India. We use adult population data from UN Population Prospects (2015) from 1950 to today. UN Population prospects provide 5-year age range annual population tables, based on national census and their own estimation procedures. The adult population is defined as the number of individuals over age 20. Before 1950, we use total population estimates from Sivasubramonian (2000) and reconstruct the adult population using total population growth rates given by the same author.

¹⁵ Imputed rents are not included in IHDS survey. They are not taken into account in NSSO data, nor in tax data after 1986.

¹⁶ We were not able to access the micro files of the National Council for Applied Economic Research's National Income and Expenditure Survey, done in 2004-5 and 2010-11.

2.2 Methodology

2.2.1 Estimation of top fiscal incomes

Following Banerjee and Piketty (2005), we first reconstruct top income thresholds and levels, using generalized Pareto interpolation techniques. The main methodological difference with Banerjee and Piketty lies in the use of generalized Pareto interpolation techniques (Blanchet, Fournier and Piketty, 2017) rather than standard Pareto distributions. Generalized Pareto interpolation¹⁷ allows for the recovery of the distribution based on tax tabulations without the need for parametric approximations. This method has demonstrated its ability to produce very precise results and also has the advantage of generating smooth estimates of the distribution, i.e. generating a differentiable quantile function and a continuous density, while other methods introduce kinks around the thresholds used as inputs for the tabulation.

The generalized Pareto interpolation procedure generates 127 generalized percentiles, namely p_0 to p_{100} , corresponding to 100 fractiles of the distribution. The top fractile is split into 10 deciles ($p_{99.0}$ to $p_{99.9}$), its top decile itself split in ten deciles ($p_{99.90}$ to $p_{99.99}$), the tenth decile again split in ten deciles ($p_{99.990}$ to $p_{99.999}$). The top generalized percentile thus corresponds to the top 0.001% of the population. As shown in Figure 4, tax data in India is only reliable above the p_{94} threshold for the recent period and above the $p_{99.9}$ threshold when we go backwards in time.

2.2.2 Estimation of bottom survey incomes

One of the main difficulties of our exercise is related to the fact that NSSO does not include questions on individual and/or household income. Our strategy consists of using observed income-consumption profiles in IHDS data to reconstruct income profiles from NSSO consumption data. We first estimate income and consumption levels for each generalized percentile of the distribution of income and consumption given by IHDS data. For each survey and each percentile of the distribution, we construct observed income-consumption ratios $\alpha_{1p} = y_p / c_p$, with y_p and c_p respectively with a mean income and consumption within quantile p . We call this strategy A1. To obtain a theoretical income-consumption profile over percentiles, we take average of years 2005 and 2011-12. In practice, the two profiles differ only marginally. We then construct two alternative ratios, α_{2p} and α_{0p} , referred to as strategies A2 and A0 respectively. In strategy A2, we assume

¹⁷ Available online at www.wid.world/gpinter

that $\alpha_{2p} = 1$ for $\alpha_{1p} \leq 1$ and $\alpha_{2p} = \alpha_{1p}$ otherwise. This second strategy is equal to assuming no negative savings rates among the poor. In strategy A0, we define $\alpha_{0p} = (\alpha_{1p} + \alpha_{2p})/2$ for $\alpha_{1p} \leq 1$. This strategy assumes that there can be negative savings rates, remittances or household transfers, but that the true α_p value lies between strategy A1 and strategy A2. Income consumption ratios for the different strategies are presented in Appendix 4. We find that these different strategies have no effect on the trends we observe and a limited impact on top share estimates, as we show in section 3.4.

The choice of these different strategies indeed impacts on the estimated share of total savings in the economy. In strategy A1 total savings are close to 0, which seems too low compared to the current rate of savings in India (about 30%). This figure is close to 5% in strategy A0 and approximately 10% in strategy A2. These values are more or less constant throughout the entire period covered whereas in National accounts they move from about 10% in the 1960s to 30% today. However, using strategy A0 and factoring in top incomes in the analysis allows us to find an aggregate savings rate of the same order of magnitude as those observed today (see Appendix 5).

2.2.3 Interpolating survey and tax data for missing years.

Our objective is to produce yearly estimate for the full distribution from 1951 to 2014. Given that survey or tax data is not available for all years, it is necessary to interpolate tax and/or survey data for a certain number of years. In order to do so, we interpolate missing years using a constant growth rate between known intervals t and $t+N$ ¹⁸.

As described in sections 2.1.2 and 2.1.4, we have two available sources for the estimation of survey income for the years 2005 and 20010-11, NSSO and IHDS. However, the trends observed in the surveys are somehow divergent. The ratio of reconstructed NSSO total income to total personal income from national accounts decreases, while the ratio of IHDS total income to total personal income from national accounts is stable. The choice of one or the other source of data has implications on our final inequality statistics: using IHDS means for the estimation of the bottom of the distribution (strategy B1) yields a lower rise in top income shares than when using the NSSO survey (strategy B2). However, using NSSO mechanically accentuates the rise in top shares over the period and the strategy B1

¹⁸ In practice, for each average income at percentile p of the survey (or tax) distribution, we define $y_{p,t+1} = y_{p,t} \times g$ where $g = (y_{p,t+N}/y_{p,t})^{1/N}$, with g the growth rate, $y_{p,t+1}$ the average income at percentile p and year $t+1$.

is therefore used as our benchmark, as it represents the conservative approach. That said, we cannot rule out strategy B2, if we believe NSSO surveys are consistent throughout the entire period covered. We provide results for strategy B2 in the data appendix.

Between 2000 and 2011, we do not observe any tax statistics, but we do observe survey data in 2005 and in 2011. Survey data is not satisfactory to track the dynamics of top incomes, but it is better than no data at all. We thus estimate the growth rates of each percentile between 1999 and 2005 on the basis of their evolution observed in the survey distribution. The resulting estimates show the top 10% share evolving in the same direction between 2005 and 2011 in our final results as in the survey. We see this strategy as the best we can have with the available data at hand.

2.2.4 Combination of tax and survey data

Several strategies can be used to correct for missing top incomes in survey data. These include the modification of the weights assigned to top earners in household surveys, the addition of extra observations of top earners or the multiplication of income levels at the top (Burkhauser et al, 2016), and each has its own strengths and weaknesses. We think that an acceptable method should be consistent, in producing distributions with plausible statistics, in particular, the shape of inverted Pareto beta coefficients curves should be relatively smooth. The method followed should also be transparent, in so-much as it should provide a statistical outcome that could be anticipated from an economic perspective; survey inequality should in principle increase when we factor in top fiscal incomes. Furthermore, a simple strategy would also be better than a complex one.

Our preferred strategy is to assume that surveys are reliable from the bottom of the distribution up to a certain percentile and that tax data is reliable after another. In practice, this amounts to multiplying income of the top percentiles in the survey by a certain factor, given by tax data. More precisely: we suppose that survey data is reliable from p_0 to p_1 - this means that between p_0 and p_1 , averages and thresholds are given by the distribution of interpolated (estimated) survey income. In our benchmark scenario, which we refer to as strategy C1, $p_1=p_{90}$. We also test alternative ranges: (i) $p_1=p_{95}$, which we refer to as strategy C2 and (ii) $p_1=p_{80}$, referred to as strategy C3. As shown in section 3.4, these different strategies have no impacts on the recent and long term income trends observed in India and have only a moderate impact on income concentration levels.

We then suppose that tax data is reliable from a certain percentile, p_2 , up to the top of the distribution. p_2 is given by the population share under the first taxable bracket observed in the tax data. This value varies from $p_2=99.9$ in the 1950s to $p_2=95.5$ in the 2010s¹⁹. Therefore, our strategy implies that averages and thresholds for all percentiles above p_2 are given by the distribution interpolated from observed tax data. Appendix 5 gives the precise value of p_2 for each year.

Between p_1 and p_2 , we test several strategies for the progression of income levels and thresholds at a given point of time. We define a convex junction profile (strategy D1), a linear profile (strategy D2) and a concave profile (strategy D3). We adopt D1 (convex profile) as our benchmark strategy as it corresponds to the profile observed for recent years, for which we have more observed fiscal data at the top; more than 6% of the population against 0.1% for the earlier period (see Appendix 6). We find that these different strategies have negligible impacts on top share results. In fact, the bulk of the correction we apply to survey incomes occurs above p_2 , not between p_1 and p_2 .

2.2.5 From total fiscal income to national income

Total fiscal income is the total personal income that would be reported by individuals or tax units, if all of them reported their revenues to the tax administration. In the case of India, we do not observe this value because of the limited tax base. One way to recover it, following Atkinson (2007), is to start from the sum of primary incomes obtained by households reported in national accounts and operate a series of deductions and additions towards a definition closer to taxable income. This is the approach followed by Banerjee and Piketty (2005) and appears appropriate given that their focus was restricted to top incomes only. By construction, total fiscal income evolves at the same rate as pre-tax national income under this approach

The other approach consists of reconstructing total fiscal income via the combination of top fiscal incomes and observed (or estimated) survey income, as we detailed in the previous section. This is equivalent to assuming that tax data give true fiscal incomes for individuals over p_2 and that estimated survey data gives the true fiscal incomes for individuals below p_1 . In this approach, reconstructed fiscal income and total national income can evolve at a different

¹⁹ In Indian tax files, there is a non-negligible proportion of filers falling below the first taxable bracket.

pace. Over the years, we observe a growing gap between reconstructed total income from surveys and total national income (see Appendix 7). This divergence is the repercussion of the gap between household consumption surveys and national accounts discussed in section 2.1.3. We show in section 3.4 that we can account for a non-negligible share of this gap after the combination of survey and tax data, but that a large part of the difference remains unexplained.

In order to produce income estimates comparable to other countries, we chose to rescale our fiscal income estimates to match total pre-tax national income from national accounts. In practice, we preserve the distribution obtained from the combination of tax and survey data and simply rescale average and threshold levels of all percentile groups by a yearly factor so that we match total national income.

In further work, we intend to distribute retained earnings to the top of the distribution following the DINA guidelines (Alvaredo et al, 2016). This would most likely increase the level of inequality in the recent period, since the growth of retained earnings is likely to be concentrated among top earners. The amount by which our results would vary presumably remains limited though. Indeed, assuming retained earnings are equal to 10% of national income, distributing half of them to the top 1% would increase its share by about 1 percentage point.

2.2.6 Definition of a benchmark scenario

The combination of our different strategies defines 54 scenarios (3 A scenarios x 2 B scenarios x 3 C scenarios x 3 D scenarios). We stress that most of the combinations of scenarios among these 54 possibilities can be *a priori* justified, and as such, we provide results for all corresponding series in our data appendix. We see our benchmark scenario (A0B1C1D1) as being at the same time plausible and conservative compared to most of the scenarios tested, as top income shares are lower than in most scenario and also increase at a slower rate over the recent decades. Robustness tests are presented in section 3.4.

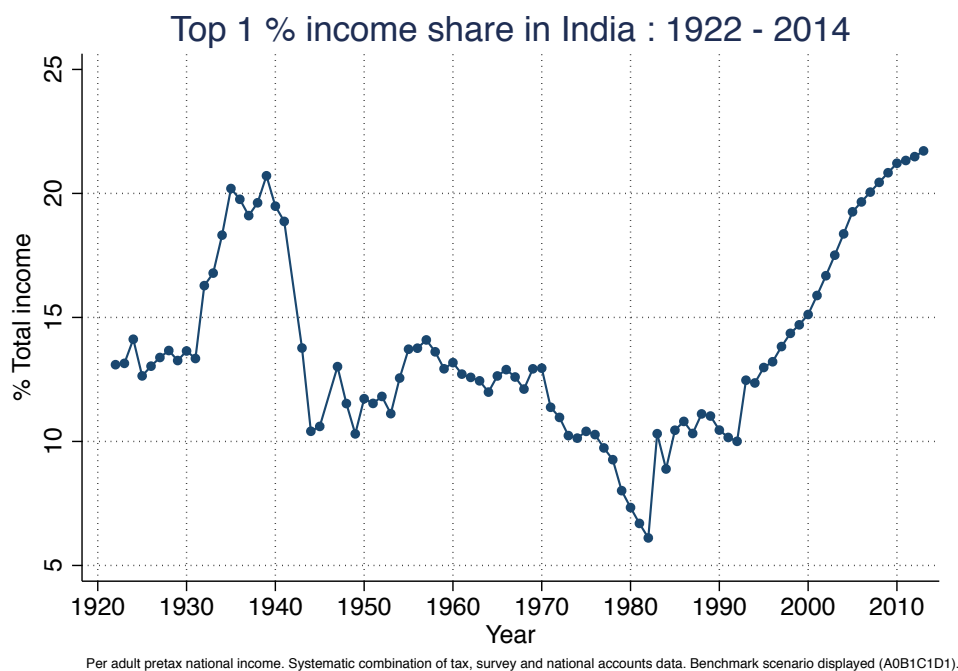
3 RESULTS

3.1 Sharp rise in top income shares since the mid-1980s

Our results exhibit a strong rise in top income shares since the mid-1980s. In our benchmark estimation scenario, the share of national income attributable to the top top 1% reached 21.7% of national income in 2013-14, up from 6.2% in 1982-1983 (see Figure 6). This is the highest level recorded since the establishment of the income tax in 1922. The top 1% share of national income was at 13% of national income in 1922-23 and increased to 20.7% in 1939-40, at the dawn of World War II.

It then dramatically decreased to 10.3% in 1949-50 and further decreased from the late 1960s to the early 1980s.

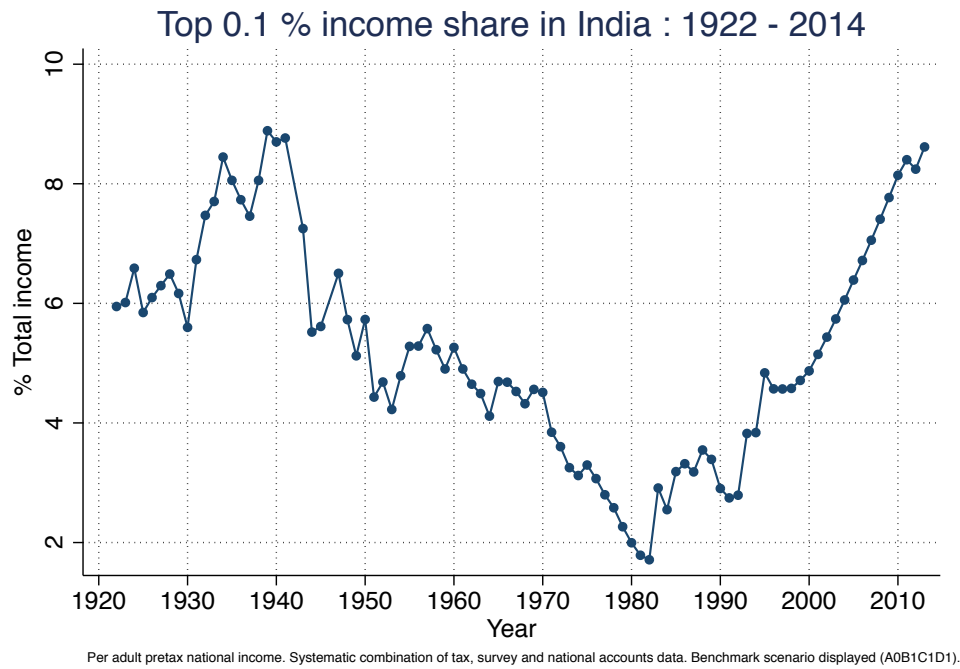
Figure 6 - Top 1% income share in India, 1922-2014



Source: Authors' computations using tax and survey data and national accounts.

As expected, the top 0.1% income share dynamics exhibit a similar pattern in our benchmark scenario (see Figure 7). Top 0.1% earners captured 8.6% of total income in 2013-2014. This only slightly below its pre-independence peak of 1939-40 (8.9%). The top 0.1% then saw a strong drop during World War II (down to 5.5% in 1944-45), followed by a continued reduction up to 1982-83 (when it reached 1.7%). From 1983-84 onwards, the share of national income accruing to the top 0.1% rose almost continuously.

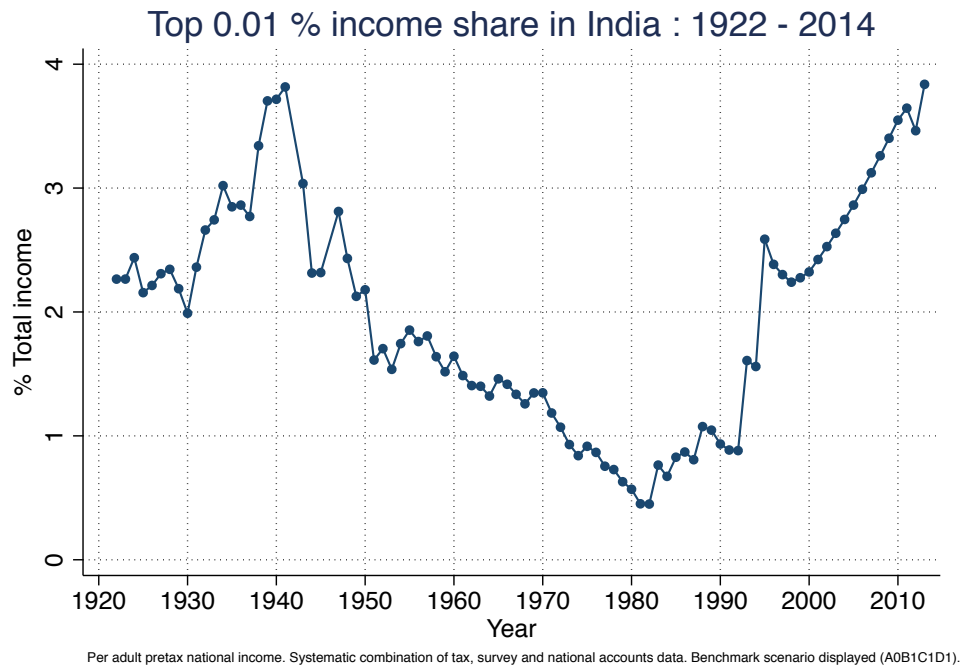
Figure 7 - Top 0.1% income share in India, 1922-2014



Source: Authors' computations using tax and survey data and national accounts.

Looking at the 0.01% earners (Figure 8), we also observe a strong increase in their share of national income since the mid 1980s, reaching 3.8% in 2013-2014, up from 0.4% in 1982-83. In 1941-42, the top 0.01% earned 3.8% of total income. The share of national income earned by the top 0.001% share is presented in the Appendix and also display a sharp since the mid-1980s.

Figure 8 - Top 0.01% income share in India, 1922-2014



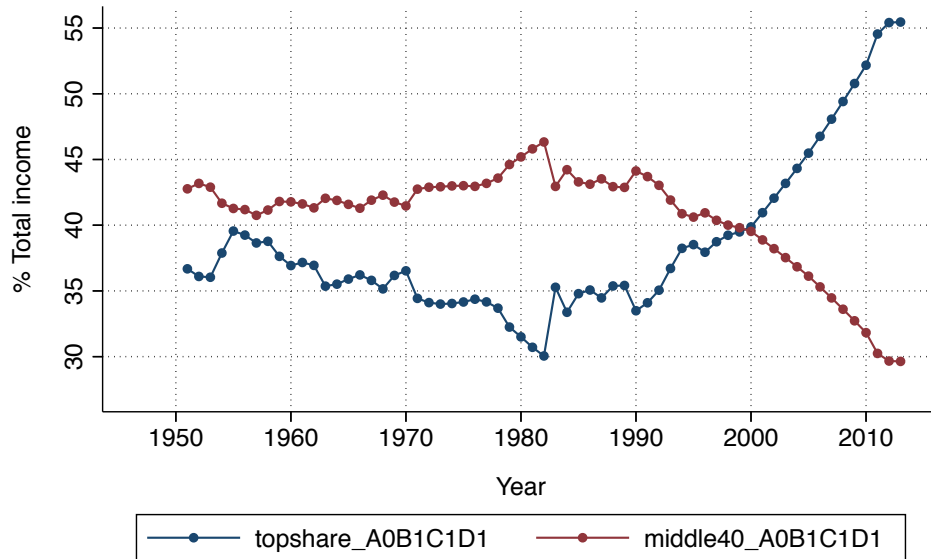
Source: Authors' computations using tax and survey data and national accounts.

3.2 Fall in Middle 40% and bottom 50% shares

We now turn to post-1951 results, which we have for the entire distribution of income. Figure 9 shows the mirror evolution of top 10% share in total income and middle 40% share (i.e. individuals above the bottom 50% earners and below the top 10%). In the mid-fifties, the top 10% and the middle 40% held about 40% of total income each, the share of the middle 40% progressively increased from the mid-fifties to 1982-83, reaching 46% of total income. It then decreased afterwards. At the turn of the Millenium, the top 10% and the middle 40% groups captured exactly the same amount, 40%. However, by 2013-14, the middle 40% share had fallen to an historically low level of 29.6%.

Figure 9 - Top 10% vs. Middle 40%: 1951-2014

Top 10 % vs. Middle 40 % income shares in India 1951-2014

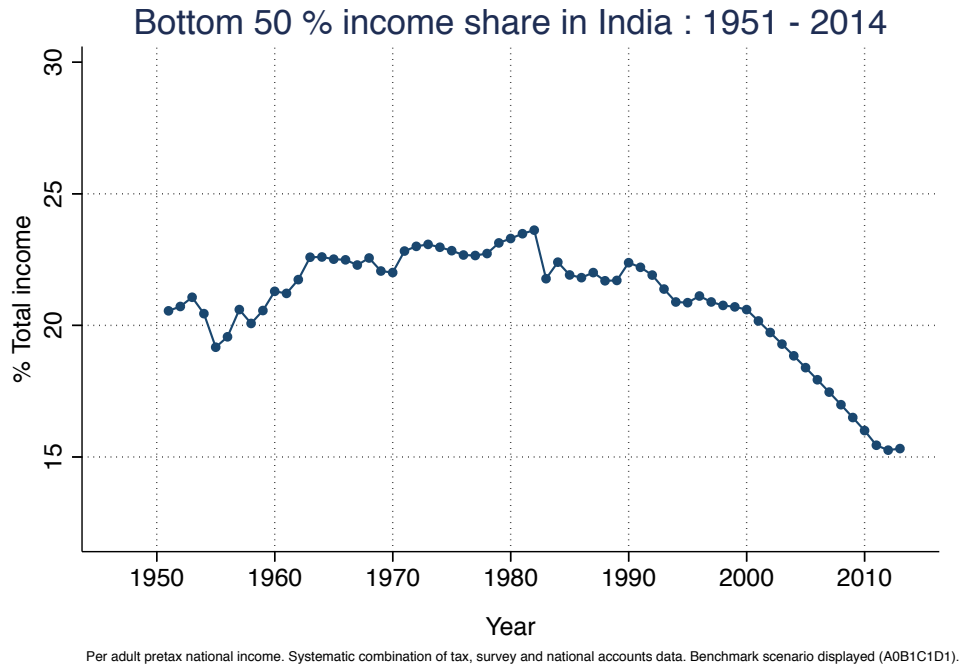


Per adult pretax national income. Systematic combination of tax, survey and national accounts data. Benchmark scenario displayed (A0B1C1D)

Source: Authors' computations using tax and survey data and national accounts.

The dynamics for the bottom 50% of the income distribution exhibit a similar pattern to that of the middle 40% (Figure 10). Bottom 50% share of national income increases from 19% in 1955-56 to 23.6% in 1982-1983, but then decreases sharply and almost continuously thereafter (20.6% in 2000-2001 and 14.9% in 2013-14).

Figure 10 - Bottom 50% income share: 1951-2014



Source: Authors' computations using tax and survey data and national accounts.

3.3 Total growth rates by income group

We compare total growth rates across the full distribution of incomes over the 1980-2014 period and compare these results in perspective to other countries available in the WID.world database, namely China, France and the USA. We also provide global growth estimates for the corresponding global groups.

Figure 11 - Total growth rates by percentile, 1980-2014

Income group (distribution of per-adult pre-tax national income)	Total cumulated per adult real growth (1980-2014)			
	India	China	France	USA
Full population	187 %	659 %	35 %	62 %
Bottom 50%	89 %	312 %	25 %	4 %
Middle 40%	93 %	615 %	32 %	45 %
Top 10%	394 %	1 074 %	47 %	119 %
<i>incl. Top 1%</i>	750 %	1 534 %	88 %	198 %
<i>incl. Top 0.1%</i>	1 138 %	1 825 %	161 %	306 %
<i>incl. Top 0.01%</i>	1 834 %	2 210 %	223 %	437 %
<i>incl. Top 0.001%</i>	2 726 %	2 546 %	261 %	621 %

Distribution of pre-tax income among adults. Estimates combine survey, fiscal and national accounts data.

Source: Authors' computations using tax and survey data and national accounts.

Figure 11 shows that income growth rates in India over the 1980-2014 period substantially increase as we progress upwards through the distribution of income. The bottom 50% of earners experiences a growth rate of 97% over the period, while the top 10% saw a 376% increase in their incomes. The equivalent figures for the top 0.01% and top 0.001% were 1834% and 2776%, respectively. Appendix 10 shows the same results on an annual growth rate basis.

Unequal growth dynamics over the period are not specific to India. Income growth rises the higher up the income distribution one proceeds in China, in the USA and in France as well. India's dynamics are, however, striking: it is the country with the highest gap between the growth of the top 1% and growth of the full population. It is also interesting to note that bottom 50% of earners grew three times more slowly in China than in India, the middle 40% six times more slowly than their Chinese counterparts, but that the incomes of those at the very top of the Indian have grown at a faster pace than in China.

While Figure 11 is particularly meaningful from the perspective of individual growth dynamics, it is also useful to balance this with information on the share of total growth captured by different income groups (Figure 12). Indeed, income growth per adult growth can appear high for certain groups, but can represent little from the perspective of total growth at the country level. Our results show that the top 0.1% earners captured more total growth than the bottom 50% (12% vs. 11% of total growth) over the period. The top 0.1% of earners represented less than 800 000 individuals in 2013-14, this is equivalent to a population smaller to Delhi's IT suburb, Gurgaon. It is a sharp contrast with the 389 million individuals that

made up the bottom half of the adult population in late 2013. At the opposite end of the distribution, the top 1% of Indian earners captured 29% of total growth, as much as the bottom 84%. The comparison of these figures with China and other countries is particularly noteworthy. Out of the four countries, India is the country where the middle 40% benefitted from the least from total growth over the period. The bottom 50% however captured a similar share of total growth in India and in China (respectively 11% and 13%).

Figure 12 - Share of total national growth captured by different income groups, 1980-2014

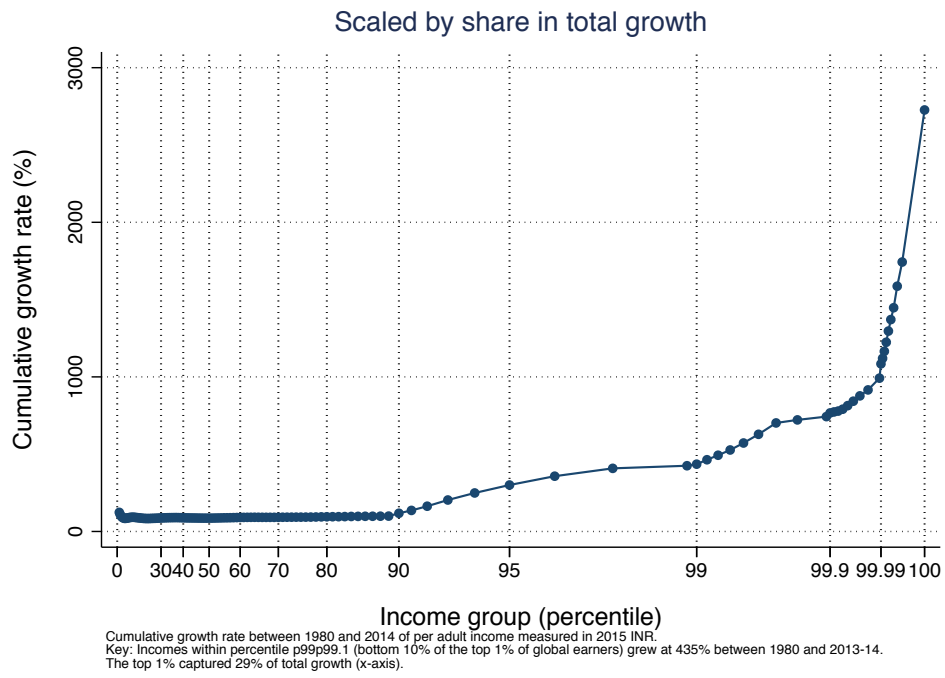
Income group (distribution of per-adult pre-tax national income)	India	China	France	USA
<i>Total</i>	100%	100%	100%	100%
Bottom 50%	11%	13 %	17 %	1 %
Middle 40%	23%	43 %	42 %	33 %
Next 9%	37%	29 %	20 %	32 %
Top 1%	29%	15 %	21 %	34 %
Top 0.1%	12%	6 %	12 %	18 %
Top 0.01%	6%	3 %	6 %	9 %
Top 0.001%	3%	1 %	2 %	4 %

Distribution of pre-tax income among adults. Estimates combine survey, fiscal and national accounts data.

Source: Authors' computations using tax and survey data and national accounts.

Figure 13 plots the information presented in Figure 11 and in Figure 12 for the full distribution of income.

Figure 13 - Total growth rate by percentile - 1980-2014



Source: Authors' computations using tax and survey data and national accounts.

Figure 14 shows income levels and income thresholds for different groups and corresponding adult population size in 2013-2014. Top 1% earners earn on average INR 3.1 million (22 times national average) versus INR 43 734 (0.3 times national average) for the bottom 50% and INR 86 841 (0.6 times national average) for the middle 40%/

Figure 14 - Income inequality in India, 2014

Income group <small>(distribution of per-adult pre-tax national income)</small>	Number of adults	Income threshold	Average income	Comparison to average (ratio)
Average	778 380 224	0 INR	142 550 INR	1
Bottom 50%	389 190 112	0 INR	43 734 INR	0.3
Middle 40%	389 190 112	63 419 INR	86 841 INR	0.6
Top 10%	77 838 024	191 713 INR	772 621 INR	5
<i>incl. Top 1%</i>	7 783 802	1 226 689 INR	3 093 335 INR	22
<i>incl. Top 0.1%</i>	778 380	4 303 757 INR	12 287 814 INR	86
<i>incl. Top 0.01%</i>	77 838	17 042 120 INR	54 596 666 INR	383
<i>incl. Top 0.001%</i>	7 784	70 223 784 INR	266 561 728 INR	1870

Distribution of pre-tax income among adults. Estimates combine survey, fiscal and national accounts data.

Source: Authors' computations using tax and survey data and national accounts

Figure 15²⁰ shows the growth rate over different income groups in India for the 1951-1980 period. The situation is reversed as compared to the 1980-2014 period: the higher the group in the distribution of income, the lower the growth rate over the period. Real per adult income of the bottom 50% middle 40% groups grew substantially faster (respectively 87% and 74%) than average income (65%). On the contrary, top 0.1%, top 0.01% and top 0.001% income groups experienced a severe decrease in their real incomes (-26%, -42% and -45% respectively).

Figure 16 reveals that bottom 50% group captured 28% of total growth over the 1951-1980 period, vs. 49% for the middle 40% and 24% for the top 10%.

²⁰ Appendix 11 presents this data into average annual growth rates and Appendix 12 plots the total growth rate curve by percentile for this period.

Figure 15 - Total growth rates by percentile, 1951-1980

Income group (distribution of per-adult pre-tax national income)	Total real per adult income growth (1951-1980)
Full population	65 %
Bottom 50%	87 %
Middle 40%	74 %
Top 10%	42 %
<i>incl. Top 1%</i>	5 %
<i>incl. Top 0.1%</i>	-26 %
<i>incl. Top 0.01%</i>	-42 %
<i>incl. Top 0.001%</i>	-45 %

Distribution of pre-tax income among adults. Estimates combine survey, fiscal and national accounts data.

Source: Authors' computations using tax and survey data and national accounts.

Figure 16 - Share of total growth captured by percentile groups, 1951-1980

Income group (distribution of per-adult pre-tax national income)	Share of growth captured by income group (1951-1980)
Full population	100%
Bottom 50%	28%
Middle 40%	49%
Top 10%	24%
<i>incl. Top 1%</i>	1%
<i>incl. Top 0.1%</i>	-2%
<i>incl. Top 0.01%</i>	-1%
<i>incl. Top 0.001%</i>	-0.4%

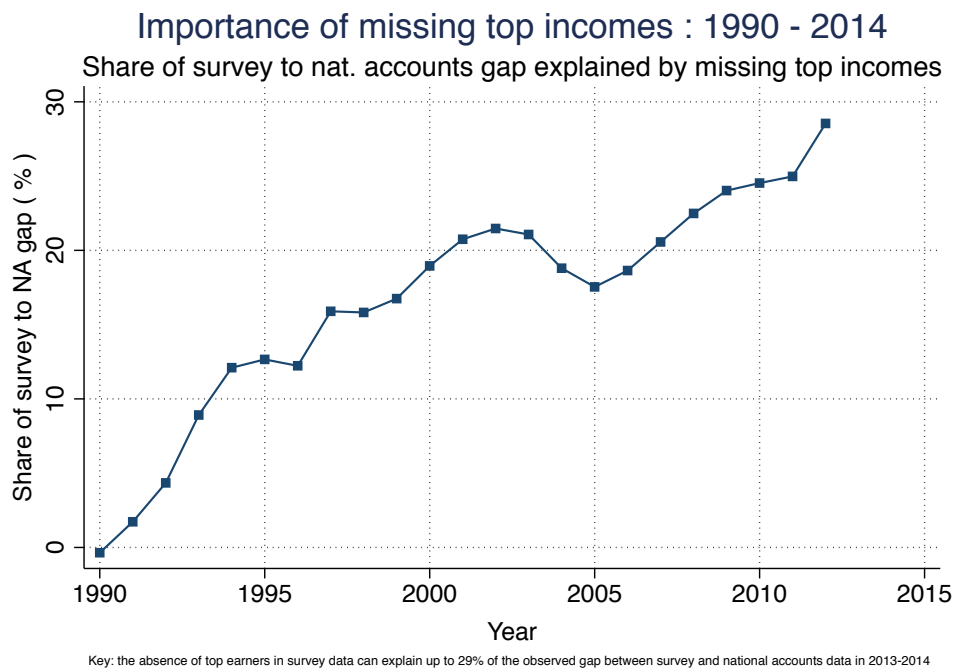
Distribution of pre-tax income among adults. Estimates combine survey, fiscal and national accounts data.

Source: Authors' computations using tax and survey data and national accounts.

3.4 Growing share of income gap explained by top incomes

We compare the theoretical fiscal income obtained from national accounts²¹ to our reconstructed fiscal income and the total income estimated from household surveys. This comparison reveals the share of survey and national accounts discrepancy discussed in section 2.1.3, that can be attributed to the absence of top earners in survey data. We find that our reconstructed fiscal income bridges a growing and non-negligible gap between national accounts surveys data. The share of the gap explained by our reconstructed fiscal income rises from about 0% in 1990 to more than 28% in 2014.

Figure 17 - Importance of missing top incomes



Source: Authors' computations using tax and survey data and national accounts.

²¹ Supposed to be 70% of net national income, following Banerjee and Piketty (2005)

3.5 Measurement issues and robustness tests

One of the main assumptions underlying our results is that tax data measures the actual income shares of the richest. There are a number of reasons why this may not entirely be true. A potential issue with tax data is that the surge in top incomes may reflect improvements in the income tax department's ability to measure and tax the incomes of the richest. The tax cuts in the early 1990s might have reduced the incentives among the wealthy for evading the income tax. Indeed, there were a number of innovations in tax collection in the 1990s, such as the 1998 introduction of the "one in six rule" that required everyone who satisfied at least one of six criteria (such as owning a car and travel abroad) to file a tax return. We note however that the overall decline in the top marginal rate was quite moderate: the top marginal tax rate dropped from 50% in 1987-1988 to less than 40% in 1999-2000 (and only minor evolutions after, see Figure 18). By comparison, the share of the top 0.01% was huge: it went up from 0.7% in 1987-88 to more than 2% in 1999-2000. If this entire change is to be explained by a shift in tax rates, the implied elasticity would have to be enormous. Another key limitation of the Indian tax series is the ten year break from 2000 to 2010. We did not find evidence of significant changes in the tax legislation, that could explain the rise in top shares post-2000. We also note that the post-2000 rise does not mark a discontinuity in the series, but comes more as the prolongation of rising top shares trend observed in the 1990s. The trend is also in line with the rise of inequality observed in consumption surveys and in wealth rich lists. The release of tax tabulations for the years 2000 to 2010 would allow us to better analyze year-on-year evolutions for this crucial period.

In order to test the robustness of our results to data limitations (including the tax data gap of the 2000s and the growing gap between national accounts and consumption surveys), we present our results along the 54 estimation strategies described in section 2.2.6. These 54 scenarios reflect a wide range of alternative assumptions to make up for the lack of consistent data for the entire distribution of income. We find that our main results are robust to all the strategies tested.

Appendix 12 shows the evolution of the top 1%, 0.1% and 0.01% shares from 1922 to 2014 across the 54 scenarios, along with our benchmark series (thick red line). The results only differ slightly between the different scenarios before 2005. In 1982-83, the top 1% share indicates 5.5% in the lower case scenario vs. 6.6% in the upper case. After 2005, the spread between scenarios is higher: top 1% income shares indicate 21.4% in the lower case scenario and 29.6% in the upper case scenario in 2013-2014. The higher spread after 2005 is essentially due to strategy B assumptions (ie. whether NSSO consumption surveys in 2005 and 2010 are

rescaled upwards). Our benchmark strategy consists in rescaling the income levels estimated from AIHS upwards - on the basis of IHDS data - to temper the rise in top shares at the end of the period. Considering these assumptions, the trends are remarkably similar across all scenarios, but the true top share values could be higher than what we obtain in our benchmark results.

Results for the middle 40% and the bottom 50% groups are relatively more sensitive to our sets of scenario assumptions, as Appendix 13 and Appendix 14. We find a 2.5 p.p. spread on lower case and upper case scenarios for middle 40% shares on average and an average 8 p.p. spread for bottom 50% income shares. This spread is essentially due to assumptions on the savings profiles of lower consumption groups (strategies A0, A1, A2). The A0 scenario reflects a mid-range position between the 0 negative assumption (scenario A2) and the profiles obtained from the IHDS dataset, with arguably excessive negative savings rates²². Long run results for bottom 40% and middle 50% groups are consistent across all scenarios: a slight increase from 1951-52 to 1983-84 and a significant decrease afterwards.

To sum up, we see our set of alternative scenario assumptions as a way to shed light on the gaps in our current knowledge of Indian income inequality. Our results are robust to a wide range of alternative assumptions but we do not pretend that these new series are definitive. More modestly, we hope they can encourage the publication of full series from 2000 to 2010. All computer codes are provided in the data appendix of the paper and can be used to produce alternative strategies, if novel data addressing current gaps were to be released.

4 DISCUSSION

4.1 The mid-1980s turnaround

Our findings confirm and amplify the conclusions of Banerjee and Piketty (2005) on Indian inequality in the long run, namely i) a marked decrease in inequality in the early fourties ii) an even stronger reduction in top income shares in the 1950-70s and iii) a significant increase from the mid-eighties onwards. Current income inequality in India is higher than during pre-independence period. This holds true from the creation of the Income Tax in 1922 to independence in

²² We note particular divergences around between 1978 and 1983 for both middle 40% and bottom 50% shares. This is explained by the fact that from 1978 to 1983, as shown in Appendix 3, we do not have survey distributional data and we interpolate them on the basis of 1978 and 1983 information. The combination of interpolated survey income levels for these specific years and certain of our strategies - in particular strategy C3 ($p_{0.1}=80$) and D3 (concave junction profile), tend to reduce "next 9%" income levels (ie. individuals above the bottom 90% but below the top 1%) and relatively increase levels of the bottom 90%. These 'extreme' scenarios are the less plausible of the set of assumptions in our view.

1947 when comparing the top 1% share of national income, but also for the pre-1922 period. Before 1922, the best available estimates show that the top 0.1% income share varied between 5 and 7% of national income vs. more than 8% today in our benchmark, conservative scenario.

We note that the reduction in top income shares was smaller during the interwar period than the reduction which occurred throughout the 1950-1970s. This seems consistent with the interpretation posited for industrialized countries' (Piketty, 2001; Piketty and Saez, 2003). The shock induced by the Great Depression of the 1930s and the War had relatively lesser impacts in India than in the USA and Europe. In India, strong government control along with an explicit goal to limit the power of the elite²³ seems to have played a key role in reducing top income inequality after independence in 1947. The set of "socialist" policies implemented up to the 1970s included nationalizations, strong market regulation and high tax progressivity.

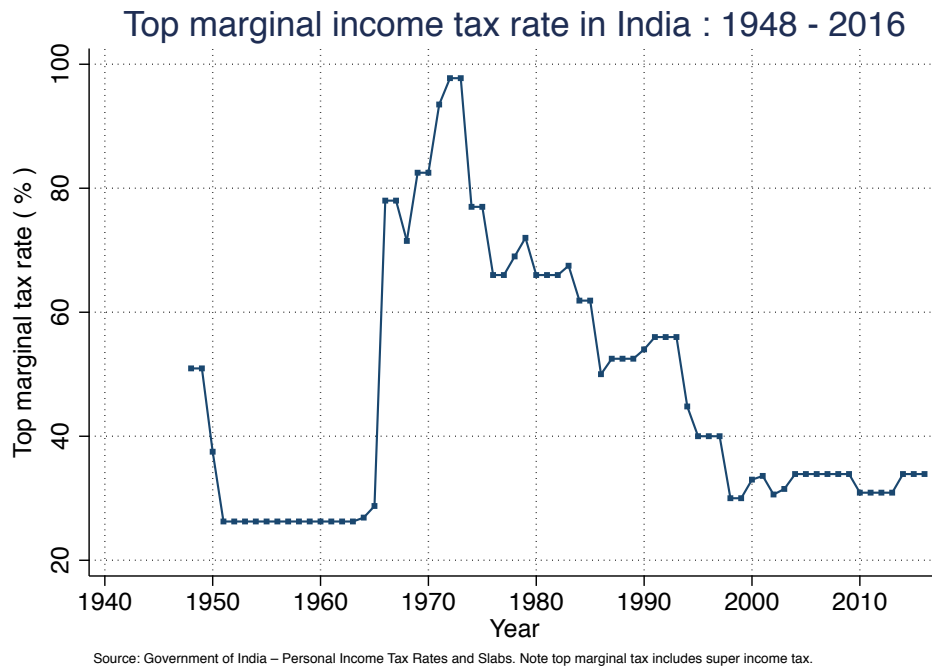
Railways were nationalized in 1951, air transport in 1953, banking in 1955, 1969²⁴ and 1980, oil industry in 1974 and 1976 to cite but a few. Along with the transfer of private to public wealth and reduction of capital incomes they implied, nationalizations came along with government setting over pay scales. In the private sector, incomes were constrained by extremely high tax rates: between 1965 and 1973, top marginal tax rates rose from 27% to 97.5%²⁵ (Figure 18). Such evolutions may have reduced rent-seeking behavior at the top of the distribution via a process of discouragement, which in presence of excessive bargaining power and rent-seeking is the efficient thing to do (Piketty, Saez, Stantcheva, 2014).

²³ An anecdote may reflect this view on fairness which prevailed in Nehruvian politics: when industrialist Tata asked then Prime Minister J. Nerhu about allowing profits in Stata-owned industries, J. Nehru answered, *"Never talk to me about profit, [...], it is a dirty word"* (Das, 2000).

²⁴ 14 banks were nationalized, representing 70% of the sector.

²⁵ These figures include the "super tax" on top incomes.

Figure 18 - Top marginal income tax rate in India, 1948-2016



Source: Authors' computations using Government of India data.

As discussed in section 1.1, from the early 1980s onwards, the Indian economy underwent reverse transformations. The turnaround of income inequality (in 1983-84, see Figure 6 to Figure 10) seems consistent with the implementation of a new economic policy agenda to disengage the public sector and to encourage entrepreneurship as well as foreign investments. The start of the process has been associated with the nomination of Rajiv Gandhi as Prime Minister in 1984.

In terms of tax progressivity, however, the downwards trend in fact started earlier - in the mid-1970s (Figure 18). That said, marginal income tax rate remained at fairly high levels until 1984-85 when Rajiv Gandhi's government reduced the rates from 62% to 50%. Why year 1983-84 marks so abruptly the turning point of our inequality series over the recent period remains a topic of enquiry. Several factors can be at play: anticipations in the 1984-85 change in the top marginal tax rate, and anticipations of a more pro-business environment, could have had a positive impact on top incomes, in line with the rent-seeking theory posited by Piketty, Saez and Stantcheva (2014). Other factors could include the combination of a strong recession in the agricultural sector the previous year (-5% agricultural production due to severe droughts in 1982-1983), which impacted income groups at the bottom. A surge in top earners filing tax returns, because of less stringent tax policies, is not to be excluded and could explain why the change is so abrupt this

year. However, the fact that the rise in inequality is prolonged throughout the 1990s and in the recent period shows that this factor is very unlikely to play decisive role in the observed trends.

Available macro series also show that the wage share in the private corporate sector has been declining in India since the early to mid-1980s (in contrast to the 1970s, when the profit share was declining; see Nagaraj (2000) and Tendulkar (2003), which is consistent with the time for the turnaround proposed here.

Our results are also consistent with the evolution of Indian wealth inequality according to All-India Debt and Investment survey data (Anand and Thampi, 2017). Recently released wealth inequality estimates indeed show a sharp increase in wealth concentration from 1991 to 2012, particularly after 2002. The increase in wealth inequality at the top of the distribution is a logical outcome of the highly unequal income growth we report in this paper over the recent period.

4.2 Shining India for the rich mostly?

Our results shed light on a particularly striking characteristic of Indian growth over the past three decades: the very moderate rise of the "middle class" - at least defined as individuals above median income and below the top 10% earners. Incomes of the middle 40% grew at 102% over the 1980-2014 period. Compared to industrialized countries' growth rates for this group, the figure is impressive. In the Indian context however, the middle 40% were notably below average growth (187%). Since 1980, the middle 40% group in India captured a much smaller share of total growth (25%) in than its counterparts did in China or Europe (more than 40%) or even the USA (33%). This result should help us better characterize what has been termed as "the rise of India's middle class". From the perspective of our newly income inequality dataset, "Shining India" corresponds to the top 10% of the population (approximately 80 million adult individuals in 2014) rather than the middle 40%. Relatively speaking, the shining decades for the middle 40% group corresponded to the 1951-1980 period, when this group captured a much higher share of total growth (49%) than it did over the past forty years. It is also important to stress that, since the early 1980s, growth has been highly unevenly distributed within the top 10% group. This further reveals the unequal nature of liberalization and deregulation processes. India in fact comes out as a country with one of the highest increase in top 1% income share concentration over the past thirty years.

5 CONCLUSION

We combine historical and novel tax data with household surveys and national accounts data in order to produce the novel estimates of the full distribution of adult pre-tax income in India, from 1951 to 2014 and for the top 1% of the distribution from 1922 to 2014.

We document a large increase in the level of inequality in India over the recent period and a large increase in the current level as compared to survey-based statistics generally used in public debates. We find that our results are robust to a large set of alternative estimation strategies addressing important data gaps. According to our benchmark estimates, the top 1% income share is at its highest level (22%) since the create of the Income Tax during the British Raj, in 1922. Top income shares and top income levels were sharply reduced in the 1950s to the 1970s at a time when strong market regulations and high fiscal progressivity are implemented. During this period, bottom 50% and middle 40% incomes grew faster than average. The trend reverted in the mid 1980s with the development of pro-business policies.

We certainly do not have the capacity to put an end to debates over the impact of economic reforms on inequality or poverty India. Our contribution is in fact relatively modest; better data series on the distribution of income inequality can and should lead to better informed democratic conversation on the state of the Indian economy. We stress the need for more research dedicated to reconcile micro and macro estimates of income and consumption inequality in India. Efforts following the Distributional National Accounts Guidelines (Alvaredo et al., 2016), published on the WID.world database, seek to go in this direction. Ultimately, meeting this objective will not be possible without the participation and expertise of official statistical agencies, in India and elsewhere.

References

- Anand I., Thampi, A., Recent Trends in Wealth Inequality in India, *Economic and political weekly*, 51(50), December 2016.
- Atkinson, A. B. (2007). Measuring top incomes: methodological issues. Atkinson, AB and T. Piketty: *Top Incomes over the Twentieth Century: A Contrast between Continental European and English-Speaking Countries*, 18-42.
- Alvaredo, F., Atkinson, A. B., Chancel, L., Piketty, T., Saez, E., & Zucman, G. (2016). *Distributional National Accounts Guidelines: Concepts and Methods used on WID.world*. WID.world Working Paper 2016/1
- Alvaredo, F., Bergeron, A., & Cassan, G. (2017). Income concentration in British India, 1885–1946. *Journal of Development Economics*.
- Atkinson, A. B., & Piketty, T. (Eds.). (2007). *Top incomes over the twentieth century: a contrast between continental european and english-speaking countries*. OUP Oxford.
- Atkinson, A. B., & Piketty, T. (Eds.). (2010). *Top incomes: A global perspective*. Oxford University Press.
- Banerjee, A., & Piketty, T. (2005). Top Indian Incomes, 1922–2000. *The World Bank Economic Review*, 19(1), 1-20.
- Blanchet, T., Chancel, L. (2016). World National Accounts Series Methodology, WID.world working paper series 2016/2
- Blanchet, T., Fournier, J., & Piketty, T. (2017). Generalized Pareto Curves: Theory and Applications. WID. world Working Paper. 2017/3
- Burkhauser, R. V., Héroult, N., Jenkins, S. P., & Wilkins, R. (2016). *What has been happening to UK income inequality since the mid-1990s? Answers from reconciled and combined household survey and tax return data* (No. w21991). National Bureau of Economic Research.
- CSO. (2008). Report of the Group for examining discrepancy In PFCE Estimates From NSSO Consumer Expenditure Data And Estimates Compiled By National Accounts, Division Central Statistical Organisation, Ministry Of Statistics & Programme Implementation, Government Of India, 2008.

Deaton, A., & Kozel, V. (2005). Data and dogma: the great Indian poverty debate. *The World Bank Research Observer*, 20(2), 177-199.

Deaton, A., & Dreze, J. (2002). Poverty and inequality in India: a re-examination. *Economic and political weekly*, 3729-3748.

Kotwal, A., Ramaswami, B., & Wadhwa, W. (2011). Economic liberalization and Indian economic growth: What's the evidence?. *Journal of Economic Literature*, 49(4), 1152-1199.

Kulshreshtha, A. C., & Kar, A. (2005). Estimates of Food Consumption Expenditure from Household Surveys and National Accounts: The Great Indian Poverty Debate.

Milanovic, B., & Lakner, C. (2015). Global Income Distribution: From the Fall of the Berlin Wall to the Great Recession. *World Bank Economic Review*.

Nagaraj, R. 2000. "Indian Economy since 1980? Virtuous Growth or Polarization?" *Economic and Political Weekly*, August 5.

Ozler, B., Datt, G., & Ravallion, M. (1996). A database on poverty and growth in India. *World Bank Development Research Group, Washington DC*.

OECD, 2017. 2017 OECD Economic Survey of India. OECD Publications, Paris.

PovcalNet (2017). Online database accessed at :
iresearch.worldbank.org/PovcalNet/povOnDemand.aspx

Piketty, T. (2001). *Les hauts revenus en France au XXème siècle*. Grasset.

Piketty, T., & Saez, E. (2003). Income inequality in the United States, 1913–1998. *The Quarterly journal of economics*, 118(1), 1-41.

Piketty, T., Saez, E., & Stantcheva, S. (2014). Optimal taxation of top labor incomes: A tale of three elasticities. *American economic journal: economic policy*, 6(1), 230-271.

Sivasubramonian, S. (Ed.). (2000). *The national income of India in the twentieth century*. Oxford University Press, USA.

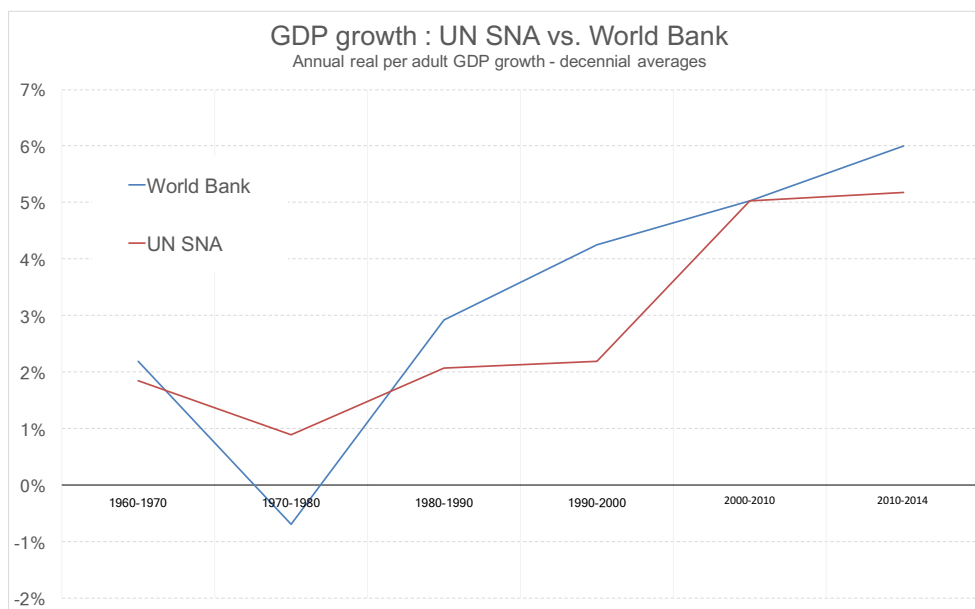
Sundaram, K., & Tendulkar, S. D. (2005). Poverty Outcomes in India in the 1990s. editors Angus Deaton & Valerie Kozal, The great Indian Poverty Debate, Macmillan India Ltd.

Tendulkar, Suresh D. 2003. "Organised Labour Market in India—Pre and Post Reform." University of Delhi, Delhi School of Economics

World Bank (2017). PovcalNet Database. <http://iresearch.worldbank.org/PovcalNet/povOnDemand.aspx> . Accessed on 15/06/2017.

Appendix

Appendix 1 - GDP growth estimates according to UN National Accounts and World Bank



Note: UN SNA and World Bank growth rates obtained from UN SNA detailed tables and World Bank online database. Figures are deflated using the GDP deflator. Adult population is taken from UN population data.

Appendix 2 - List of corrections done to raw tax files

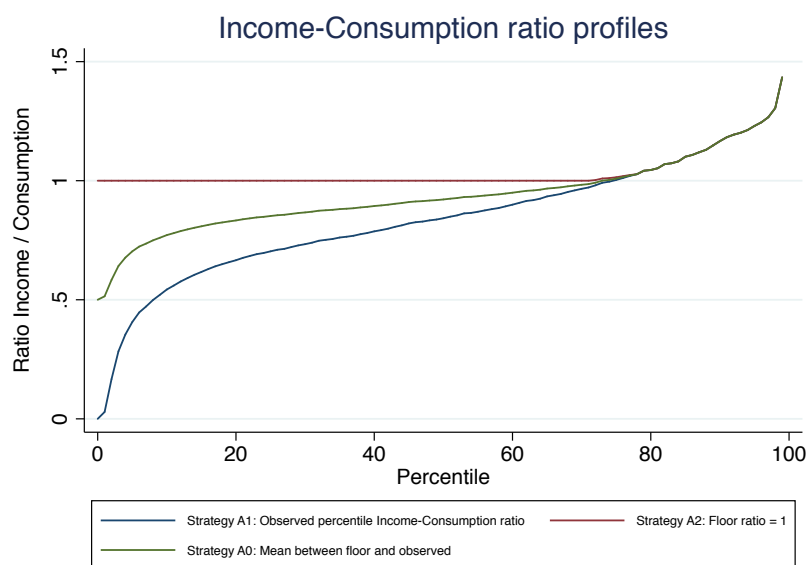
Year	Correction
1948-1951	The first bracket of 1000 is removed altogether
1951	Merging of 70k and 60k brackets into 50k to 100k brackets
1965	Merging of 15k and 17.5k brackets into 12.5k to 20k brackets
1979	Merging of 40k bracket into 30k to 50k brackets
1994	Merging of 400k and 500k brackets into into 300k
1997	Not used for analysis due to erroneous values
2013	Correction for assumed typo on very top bracket.

Appendix 3 - List of NSS consumption surveys and summary statistics

NSS Round	year	Mean consumption - survey	Mean income - strategy 1	Mean income - strategy 2	Mean income - strategy 3	Gini consumption - survey	Gini income - strategy 1	Gini income - strategy 2	Gini income - strategy 3	p90/p10 ratio consumption - survey	p90/p10 ratio income - strategy 1	p90/p10 ratio income - strategy 2	p90/p10 ratio income - strategy 3
3	1951	483.2	480.7	528.6	504.6	0.36	0.48	0.40	0.44	4.8	10.5	5.5	7.2
4	1952	420.6	417.7	459.8	438.8	0.35	0.47	0.40	0.44	5.0	10.7	5.8	7.5
6	1953	424.5	420.6	463.2	441.9	0.35	0.47	0.39	0.43	4.1	8.9	4.7	6.2
7	1954	341.1	338.5	373.1	355.8	0.35	0.47	0.40	0.44	4.5	9.7	5.2	6.8
9	1955	312.9	312	341.9	326.9	0.36	0.47	0.41	0.44	4.7	9.9	5.4	7.1
10	1956	357	355.8	391.3	373.6	0.36	0.48	0.41	0.45	4.7	10.3	5.5	7.2
12	1957	358.5	358.4	394.7	376.5	0.36	0.49	0.41	0.45	4.2	9.2	4.9	6.4
13	1958	377.3	373.9	412.3	393.1	0.35	0.47	0.40	0.44	4.6	9.9	5.3	6.9
14	1959	412.2	409.6	451.7	430.6	0.35	0.48	0.40	0.44	4.0	8.8	4.6	6.0
15	1960	413.4	406.9	451.2	429.1	0.33	0.46	0.38	0.42	3.9	8.4	4.5	5.9
16	1961	441	434.7	481.2	458	0.34	0.46	0.38	0.43	4.2	9.1	4.9	6.4
17	1962	453.7	447.1	495.7	471.4	0.34	0.46	0.38	0.42	4.0	8.7	4.6	6.1
18	1963	470.5	459.9	512.3	486.1	0.32	0.45	0.37	0.41	3.6	7.6	4.2	5.4
19	1965	554.5	541.8	603.7	572.7	0.32	0.45	0.36	0.41	3.7	8.0	4.3	5.6
20	1966	591.4	577.3	643.5	610.4	0.31	0.45	0.36	0.40	3.7	8.1	4.3	5.7
21	1967	649.3	632.7	704.8	668.7	0.31	0.44	0.36	0.40	3.9	8.3	4.5	5.8
22	1968	701.4	681.9	760.7	721.3	0.31	0.44	0.35	0.40	3.7	8.0	4.3	5.6
23	1969	701.9	688.8	765.7	727.3	0.32	0.46	0.37	0.41	3.8	8.1	4.4	5.7
24	1970	739.3	723.7	805.4	764.6	0.32	0.45	0.37	0.41	3.9	8.3	4.5	5.8
25	1971	756.8	738.2	823	780.6	0.31	0.44	0.36	0.40	3.8	8.0	4.3	5.7
27	1973	929.4	912.4	1013.9	963.2	0.33	0.46	0.37	0.42	3.9	8.4	4.5	5.9
32	1978	1443.6	1445.2	1594.3	1519.8	0.36	0.49	0.41	0.45	3.9	8.5	4.5	5.9
38	1983	2478.6	2440	2700.7	2570.4	0.33	0.46	0.38	0.42	4.2	9.0	4.8	6.3
43	1988	4157.4	4104	4542.2	4323.1	0.34	0.47	0.39	0.43	4.2	9.1	4.9	6.4
50	1994	7298.8	7185.3	7965.4	7575.3	0.33	0.46	0.38	0.42	4.1	8.9	4.8	6.3
55	2000	12804.1	12511.7	13921.6	13216.7	0.32	0.45	0.37	0.41	3.9	8.5	4.5	5.9
61	2005	12549.3	12481.3	13789.1	13135.2	0.35	0.48	0.40	0.44	4.2	9.0	4.8	6.3
66	2010	20322.3	20344.4	22414.7	21379.5	0.36	0.49	0.41	0.46	4.3	9.3	5.0	6.5

Source: Authors' computations using NSSO data

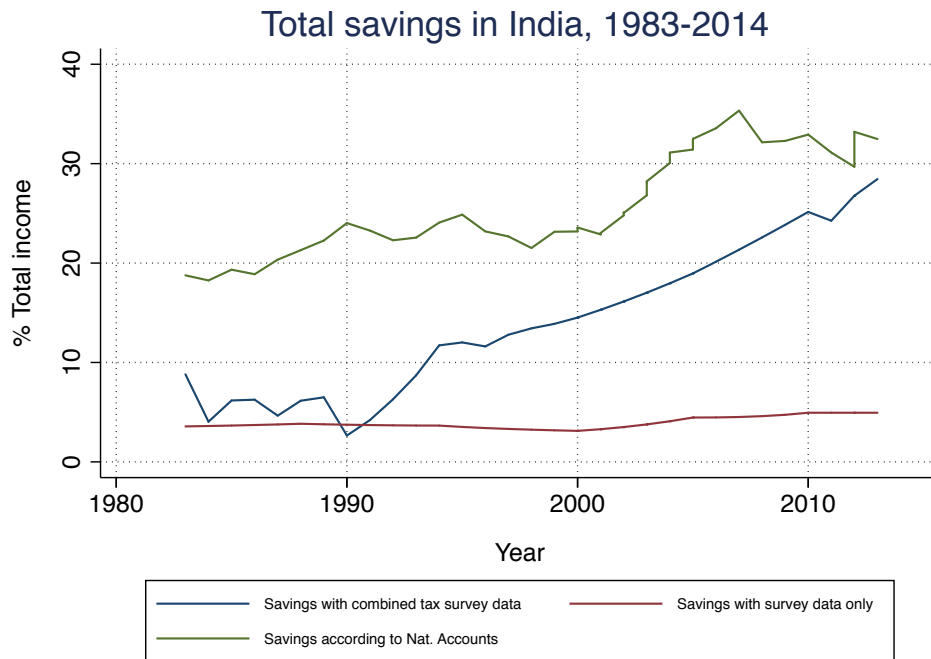
Appendix 4 - Income-Consumption ratio profiles



Source: Authors' computations using IHDS data

Note: Savings profile 1 corresponds to observed IHDS ratios, savings profile 2 corresponds to observed ratios, constrained to be superior to 1 and profile 3 to the mean between profile 1 and profile 1 when the observed ratios are inferior to 1.

Appendix 5 - Aggregate savings



Note: National accounts data shows gross savings as a share of gross disposable income.

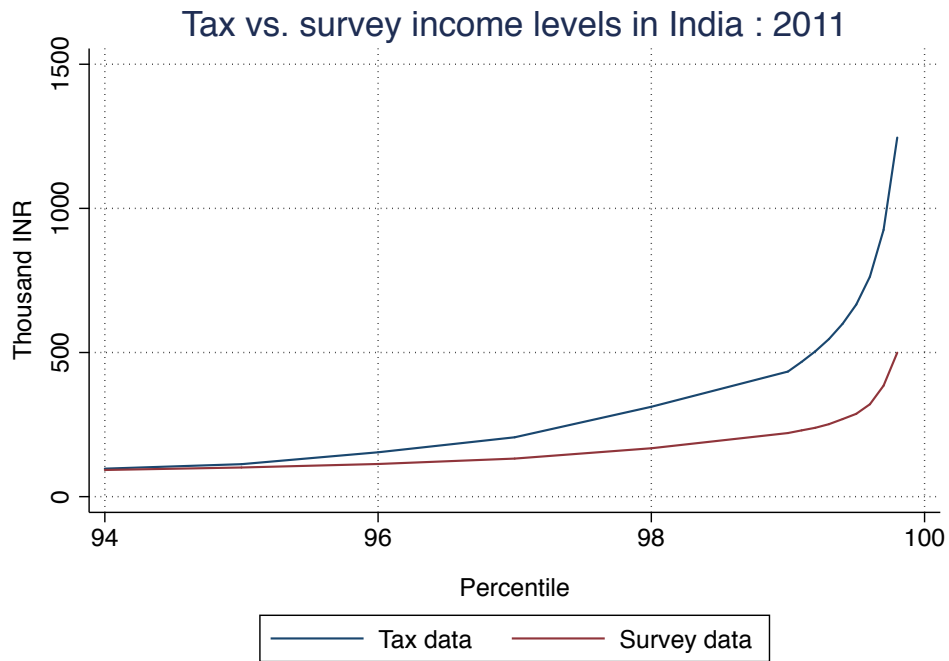
Source: Authors' computations using tax, survey and national accounts data.

Appendix 6 - Junction percentile

year	percentile	year	percentile	year	percentile	year	percentile
1922	99.9	1943	99.8	1967	99.4	1989	99.3
1923	99.9	1944	99.8	1968	99.4	1990	99.2
1924	99.9	1945	99.8	1970	99.4	1991	99.1
1925	99.9	1947	99.8	1971	99.4	1992	99
1926	99.9	1948	99.8	1973	99.4	1993	99
1927	99.9	1949	99.8	1974	99.4	1994	98.8
1928	99.9	1950	99.7	1975	99.4	1995	98.7
1929	99.9	1953	99.8	1976	99.4	1996	98.4
1930	99.8	1954	99.8	1977	99.6	1997	98.1
1931	99.8	1955	99.8	1978	99.5	1998	97.6
1932	99.7	1956	99.8	1979	99.7	2011	94.5
1933	99.7	1957	99.7	1980	99.7	2012	93.9
1934	99.7	1958	99.7	1981	99.8	2013	93.6
1935	99.9	1959	99.7	1982	99.8		
1936	99.9	1960	99.6	1983	99.7		
1937	99.9	1961	99.6	1984	99.7		
1938	99.9	1962	99.6	1985	99.6		
1939	99.9	1964	99.5	1986	99.5		
1940	99.9	1965	99.5	1987	99.5		
1941	99.9	1966	99.5	1988	99.4		

Source: Authors' computations using tax data.

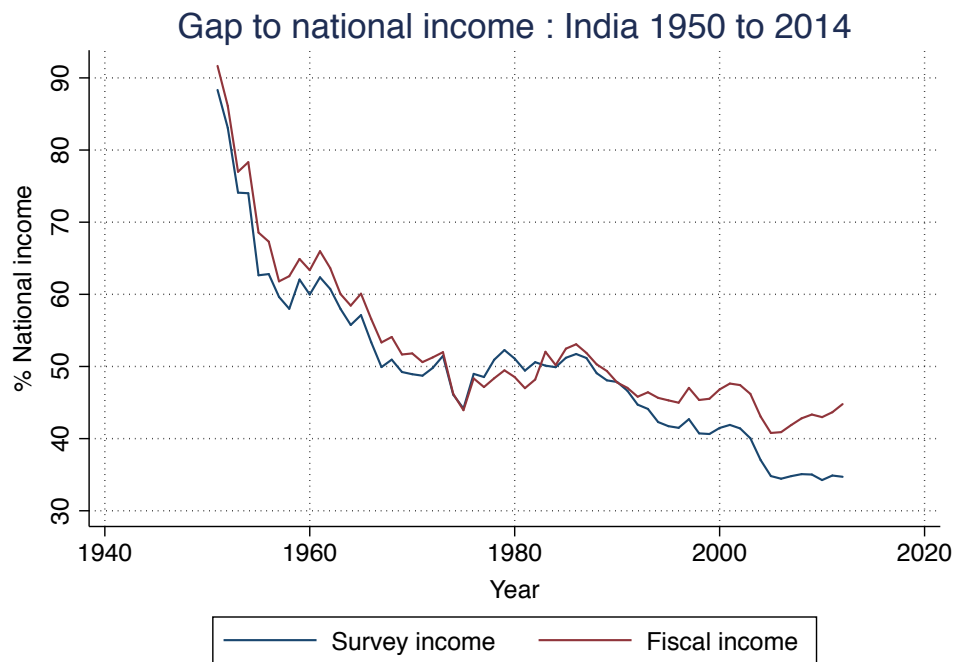
Appendix 7 - Survey vs. tax incomes at the top, 2011



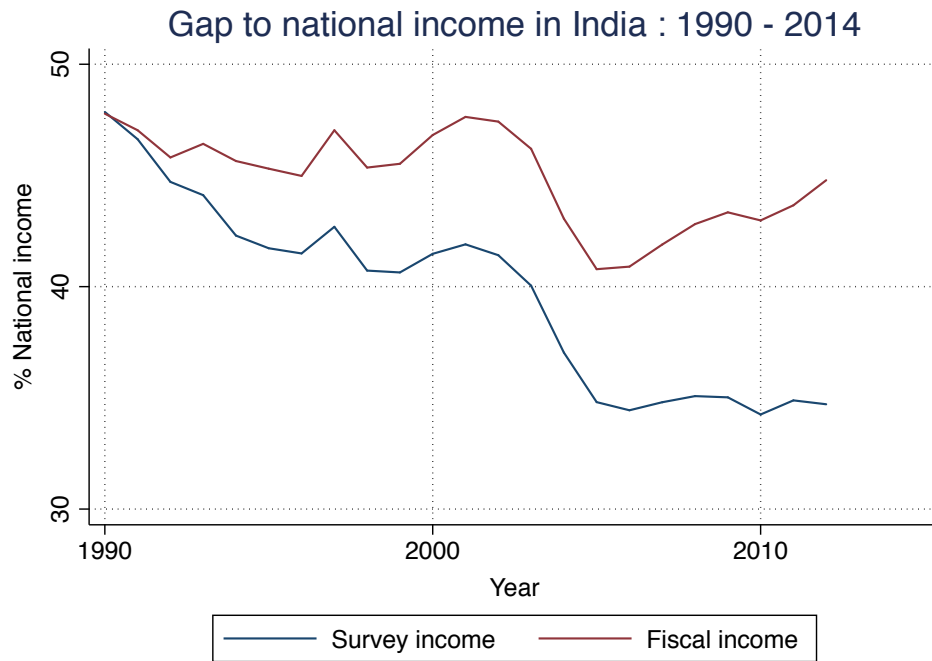
Source: Authors' computations using tax and survey data and national accounts.

Appendix 8 - Gap to National Income

Appendix 8-1 Gap to National income, 1950-2014

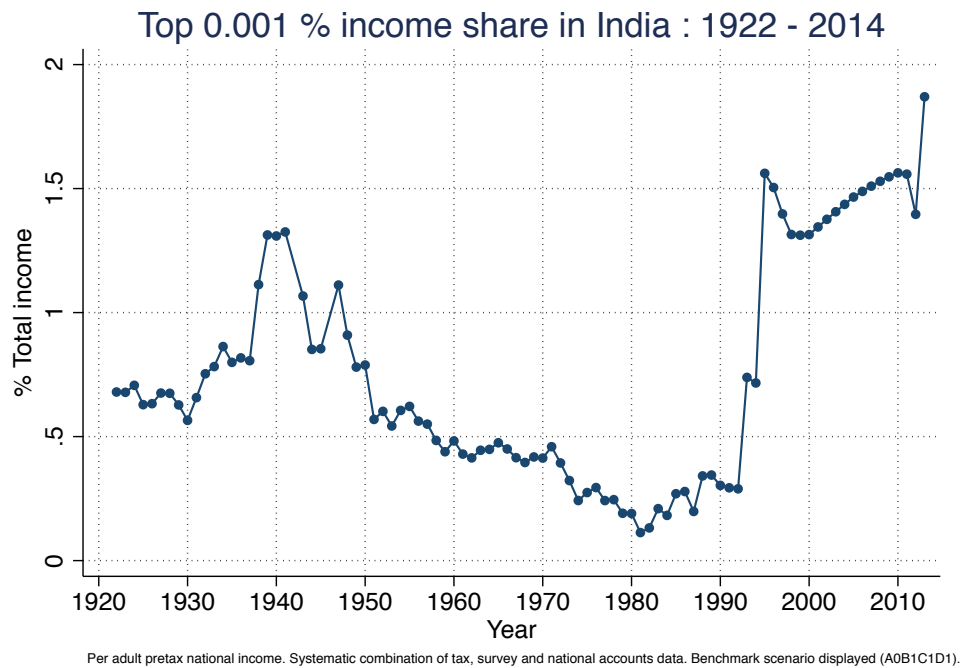


Appendix 8-2 Gap to national income: 1990-2014



Source: Authors' computations using tax and survey data and national accounts. Fiscal income is reconstructed for years 2000 to 2010, as explained in Section 2.

Appendix 9 - Top 0.001% income share in India: 1922 - 2014



Source: Authors' computations using tax and survey data and national accounts.

Appendix 10 - Average annual real per adult income growth by percentile (1980-2014)

Income group (distribution of per-adult pre-tax national income)	Average annual real per adult real growth (1980-2014)			
	India	China	France	USA
Full population	3.25 %	6.33 %	0.92 %	1.48 %
Bottom 50%	1.94 %	4.38 %	0.69 %	0.12 %
Middle 40%	2.02 %	6.14 %	0.85 %	1.13 %
Top 10%	4.96 %	7.75 %	1.17 %	2.41 %
<i>incl. Top 1%</i>	6.70 %	8.83 %	1.93 %	3.36 %
<i>incl. Top 0.1%</i>	7.92 %	9.38 %	2.95 %	4.34 %
<i>incl. Top 0.01%</i>	9.39 %	9.98 %	3.61 %	5.22 %
<i>incl. Top 0.001%</i>	10.66 %	10.44 %	3.97 %	6.17 %

Distribution of pre-tax income among adults. Estimates combine survey, fiscal and national accounts data.

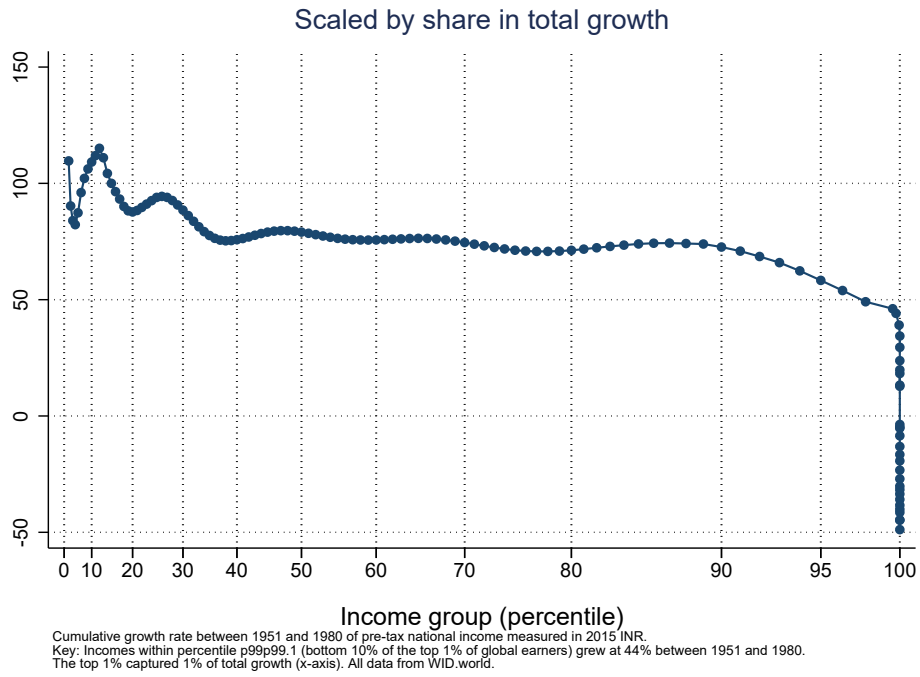
Source: Authors' computations using tax and survey data and national accounts.

Appendix 11 - Average annual real per adult income growth by percentile (1951-1980)

Income group (distribution of per-adult pre-tax national income)	Average annual real per adult income growth (1951-1980)
Full population	1.7%
Bottom 50%	2.2%
Middle 40%	1.9%
Top 10%	1.2%
<i>incl. Top 1%</i>	0.2%
<i>incl. Top 0.1%</i>	-1.0%
<i>incl. Top 0.01%</i>	-1.9%
<i>incl. Top 0.001%</i>	-2.0%

Distribution of pre-tax income among adults. Estimates combine survey, fiscal and national accounts data.

Appendix 12 - Total growth rate by percentile - 1951-1980

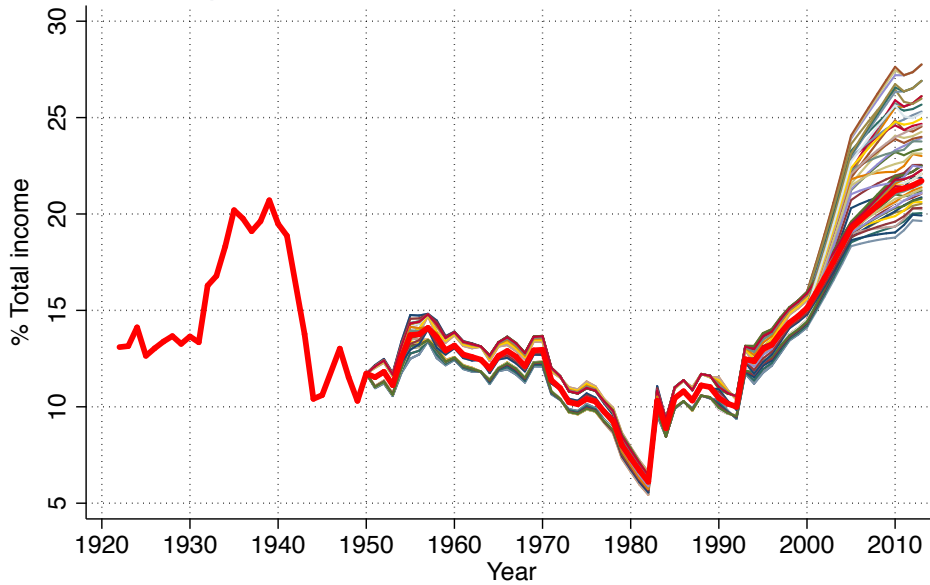


Source: Authors' computations using tax and survey data and national accounts.

Appendix 13 - Top income shares across 54 scenarios

Appendix 13-1 Top 1% income shares across 54 scenarios

Top 1 % income share in India : 1922 - 2014

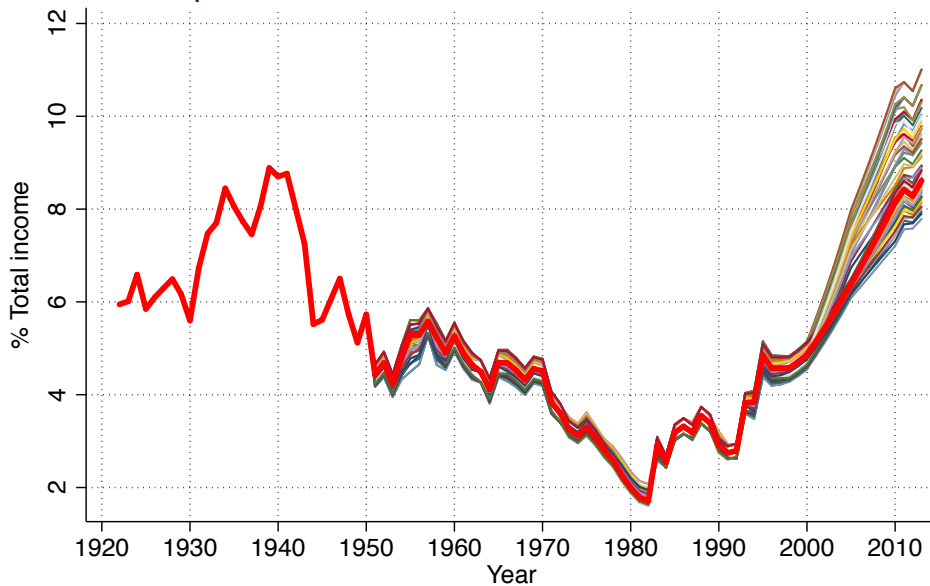


Per adult pretax national income. Systematic combination of tax, survey and national accounts data. All scenarios displayed. Thick red line corresponds to the benchmark scenario (A0B1C1D1).

Source: Authors' computations using tax and survey data and national accounts.

Appendix 13-2 Top 0.1% income shares across 54 scenarios

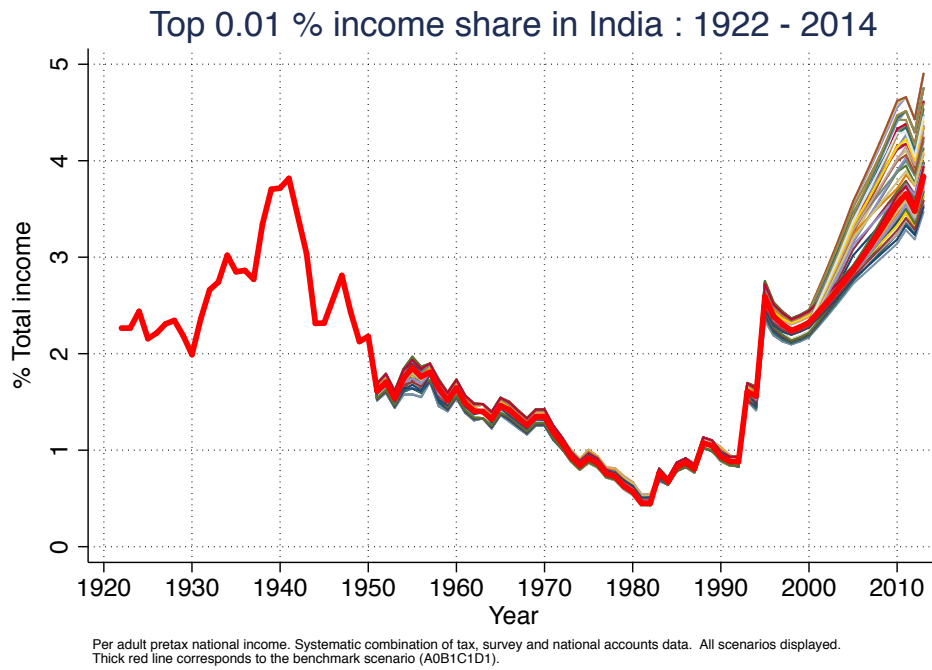
Top 0.1 % income share in India : 1922 - 2014



Per adult pretax national income. Systematic combination of tax, survey and national accounts data. All scenarios displayed. Thick red line corresponds to the benchmark scenario (A0B1C1D1).

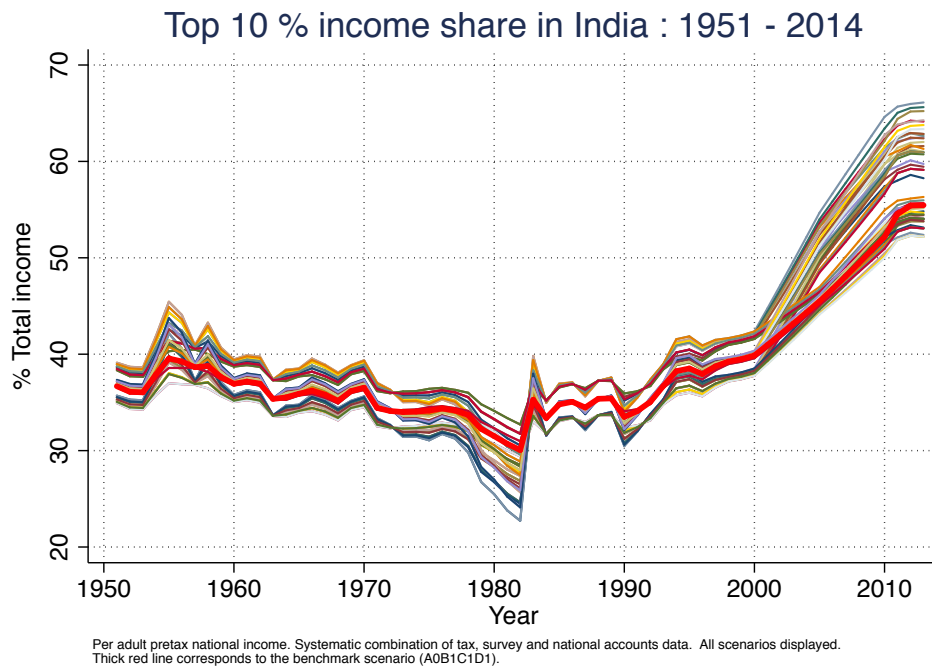
Source: Authors' computations using tax and survey data and national accounts.

Appendix 13-3 Top 0.01% income shares across 54 scenarios



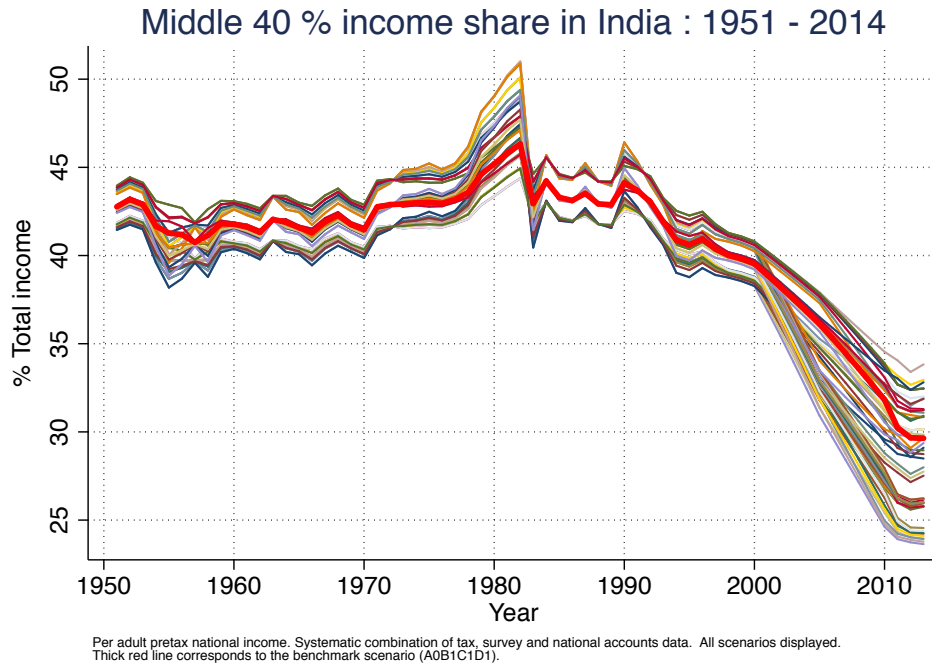
Source: Authors' computations using tax and survey data and national accounts.

Appendix 13-4 Top 10% income shares across 54 scenarios

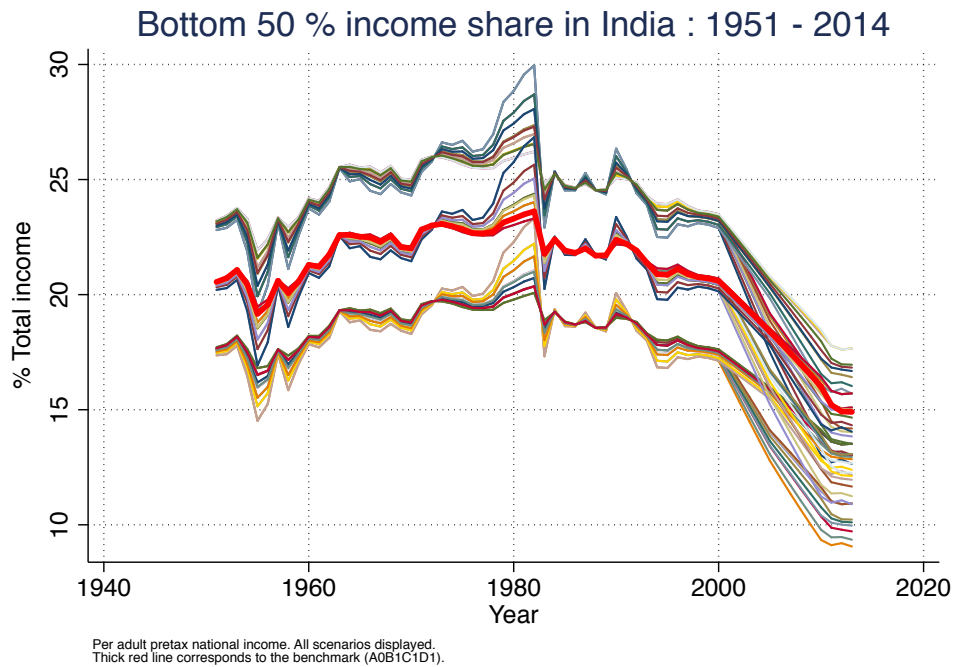


Source: Authors' computations using tax and survey data and national accounts.

Appendix 14 - Middle 40% income shares across 54 scenarios



Appendix 15 - Bottom 50% income shares across 54 scenarios



Source: Authors' computations using tax and survey data and national accounts.