ISSN No. 2454 - 1427

CDE
December 2022

Do the Wealthy Underreport their Income?
Using General Election Filings to Study the Income-Wealth Relationship in India

Ram Singh
Email: ramsingh@econdse.org
Department of Economics,
Delhi School of Economics

Working Paper No. 331

Centre for Development Economics
Delhi School of Economics
Delhi- 110007

Do the Wealthy Underreport their Income?
Using General Election Filings to Study the Income-Wealth Relationship in India

Ram Singh
Delhi School of Economics
University of Delhi, Delhi 110007
Email: ramsingh@econdse.org

ABSTRACT

Income reporting behaviour of individuals from different wealth groups remains under-researched, though media reports abound on how billionaires across countries pay very little by way of income tax. In this paper, we use new and publicly available data to examine the income-reporting behaviour of individuals and households across wealth groups in India. To this end, we use election filings by the contestants for the lower house of the Indian Parliament. These affidavits are the only simultaneous source of information on the wealth and income of individuals and households in the country. We supplement the affidavits data with Forbes' listing of billionaires and the statistics published by the income tax department, and several other sources. Our findings are novel on several counts. We find that the wealthier a household is, the smaller the income it reports relative to its wealth. On average, a 1% increase in family wealth is associated with a decrease of more than 0.5% in the reported income as a ratio of wealth. The total income reported by the bottom 10% of families amounts to more than 188% of their wealth; in contrast, the wealthiest 5% [respectively 0.1%] of families reported income equivalent to just 4% [respectively 2%] of their wealth. From another perspective, the total income reported by the wealthiest 0.1% of families is only about a fifth of the returns from their capital. For the Forbes-listed 100 families, an even higher fraction of capital income goes unreported in the income tax returns. The income-wealth ratios for individuals exhibit very similar patterns. While the decreasing income-wealth ratio is consistent with our model, we show that tax avoidance is an important factor behind the abnormally low values of the capital income reported by the top wealth groups. The income missing from the income tax returns has implications for estimates of income inequality and makes the tax regime regressive. In addition, we find evidence of tax evasion across wealth groups. Finally, we show that *ceteris paribus*, women report lower incomes than men and that individuals exposed to greater media and civil society scrutiny report relatively high incomes.

Key Words: Income, Wealth, Capital Income, Income-Wealth Ratio, Tax Avoidance, Tax Evasion

JEL Classification: D31, D63, H24, H26

Acknowledgement of Funding. I am thankful to Amit Bhaduri, Surjit Bhalla, Nitin Bharti, Keshav Choudhary, Sugato Dasgupta, Edward Gleaser, K.L. Krishna, J.V. Meenakshi, Subhash C Pandey, Thomas Piketty, Ajit Ranade, Andrei Shleife, Charan Singh and Gabriel Zucman for detailed comments and suggestions during the course of this study. I thank Soumyajit Ray for excellent research support, Saveri Sargam, and Nandini Kumar for useful inputs, and the CDE for institutional support. The research presented here has been facilitated by the IoE, University of Delhi grant no Ref. No. loE/062/Grants-Schools/2022-2023, and the Indian Council of Social Science Research (ICSSR) grant F.No. 02/68/GN/2021-22/ICSSR/RP /MJ.

Abbreviations and Notations

ADR:	Association for Democratic Reforms
AIDIS:	All-India Debt Investment Survey
AOP:	Association of Persons
AY:	Assessment Year
BOI:	Body of Individuals
CAG:	Comptroller and Auditor General of India
CBDT:	Central Board of Direct Taxes
CEO:	Chief Executive Officer
CSO:	Central Statistical Organisation
DDT:	Dividend Distribution Tax
ECI:	Election Commission of India
FL:	Forbes' List
GDP:	Gross Domestic Product
GE:	General Election
GPI:	Generalised Pareto Interpolations
GTI:	Gross Total Income
HH:	Household
ITA:	Income Tax Act
ITR:	Income Tax Return
LIC:	Life Insurance Corporation
LLP:	Limited Liability Partnership
LTCG:	Long-term capital gains
NSSO:	National Sample Survey Office
OECD:	Organisation for Economic Co-operation and Development
RBI:	Reserve Bank of India
SC:	Scheduled Castes
SNA:	System of National Accounts
ST:	Scheduled Tribes
STCG:	Short Term Capital Gain Tax
TI:	Total Income
UR:	Unreserved
USD:	United States Dollar
WM:	Wealthiest Member in a Household

Notations

Y_A:	Agricultural income
Y_H:	Imputed rent on housing property
Y_{Cg}:	Capital gains
Y_{Eq}:	Equity income
Y_{KD}:	Direct Capital Income
Y_L:	Labour income
Y_P:	Property (rental) income
Y_{PID}:	Personal Income Direct – labour and direct capital income
Y_R:	Total income reported to tax authorities
Y_T:	Income reported to tax authorities as taxable income
Y_{Td}:	Income that gets taxed by the tax authorities
W:	Wealth, i.e., market value of the all the assets owned minus all liabilities

1. Introduction

The income reporting behaviour of taxpayers is a central issue in public finance. Tax revenue depends on the income taxpayers report: the higher the reported income, the larger the tax revenue, and vice versa. The income reporting behaviour is also critical from an equity viewpoint. If wealthy groups can get away with paying tax on a relatively small part of their income, the outcome can be a regressive income tax regime, which in turn can exacerbate income and wealth inequalities. Of late, income levels reported by wealthy groups have attracted much attention from the media and think tanks. Media reports abound on how billionaires like Jeff Bezos, Elon Musk, and Warren Buffett pay very little by way of income tax. In India also, while movie stars such as Akshay Kumar, Amitabh Bachchan, and Salman Khan are reported to be among the top-income taxpayers, very few of the wealthiest Indians figure on the list.[1] The most of media reports are based on purported leaks of individual tax files or other unauthenticated sources.

As such there is very little empirical research on the relationship between wealth and reported income at the individual or household levels.[2] In this paper, we model the income-wealth relationship and empirically examine relationships using a new publicly available source along with several other data in public domain to examine income reporting behaviour of different wealth groups.

We find that, on average, the wealthier an individual is, the smaller the reported income relative to their wealth. Similar is the case with the income reported by households. These findings are consistent across the entire wealth spectrum: the larger the wealth, the smaller the reported income. Contrary to the common belief, we find that the Indian tax regime is not progressive with reference to the total income. We show that the income-wealth ratios reported by the top wealth groups are minuscule compared to the rate of return on their assets. We find evidence of tax avoidance by the wealthy groups, and the use of farmland and other properties for tax evasion across wealth groups. In addition, we find profession and gender-fixed effects. For instance, *ceteris paribus*, women report smaller incomes than men. Full-time agriculturists and politicians also report relatively low-income levels.

These novel findings are mainly attributable to the new dataset based on the election filings in the form of affidavits submitted by the contestants of elections to the Lok Sabha, the House of Representatives in the Parliament of India. The election filings are the only simultaneous source of information on the wealth and income levels of Indians. The affidavits examined by us provide information on the wealth and income of 7,596 households (HH) and their adult members. We examine the income-wealth relationship based on these election filings as a starting point. However, to cover the entire wealth and income spectrums in India, we supplement the affidavits data with the Forbes' List (FL) of billionaires and the statistics published by the Government of India's Central Board of Direct Taxes (CBDT).

The coverage of the affidavit data itself is extensive. The HH wealth covered by it ranges from a negative wealth (net liability) of ₹51 crores to a net worth of more than ₹8,911 crores — a figure not very far below what is seen at the left tail of the wealth distribution for the Forbes-listed families. Similarly, the annual family incomes it lists vary from a paltry sum of ₹178 to as high as ₹206 crores. Further, by its structure, this dataset is reasonably representative of the Indian context in terms of

[1] See ProPublica June 2021, Indian Express, and India Today, July 2022.

[2] A few studies do inform us about the income-wealth relationship but only for broad categories of wealth groups in the United States and Europe. See Dynan (2009), Piketty (2014), and Chancel, Piketty, Saez, Zucman, et al. (2022). These studies do not report results at unit level. The relationship between national wealth and income has been extensively examined for many countries For a review of the literature on the subject, see Piketty (2014). For the evolution of the income-wealth relationship in India, see Kumar (2019).

the regional and rural-urban distribution of the population. As discussed in the next section, the dataset includes all leading social, demographic, professional, and educational categories. We will refer to the affidavit data as the General Election (GE) data.

The GE data exhibits several properties consistent with what can be gleaned from the other independent sources; such as the concentration of wealth and income at the right tail of the distribution[3] and male dominance over family income and wealth, among others. Also, it exhibits properties known about the portfolio choices by different wealth groups. For instance, the share of financial assets such as company stocks and firm ownership increases with wealth levels. The asset portfolios of the wealthiest individuals in the data resemble the portfolios held by the most affluent non-politician Indians in the FL. Furthermore, asset holdings exhibited by the wealthy groups in the affidavit data are very similar to what is observed in international studies such as Piketty (2018, chapter 9), Wolff (2017), OECD (2018), and Chancel, Piketty, Saez, Zucman, et al. (2022).

The GE data, in addition to being the only simultaneous source of information on incomes and wealth, pass the 'smell' tests on several counts in addition to possessing the properties discussed above. For instance, as will be discussed in Section 5, the downward trend in the income-wealth ratio emerging from the data nests well with what can be inferred from other independent data sources such as the FL and the statistics published by the Indian Income Tax Department put together. The trends are also consistent with the inferences based on the available information for other countries.[4]

Still, there can be legitimate concerns regarding the representativeness of GE data for Indian society. Technically speaking, the income-wealth relationships emerging from the data might not hold in general. Section 3 discusses several aspects of this concern in detail and presents the robustness checks used in this study to address them, including using other sources to examine the veracity of the results emerging from the GE data.

Even though election filings are an invaluable source of information as far as information on wealth and its components are concerned, they are only partially informative as far as income is concerned. Affidavits provide information only on what can be described as the "net taxed income" reported by the candidates and their family members to the tax department. In particular, these documents do not offer information on the total income the candidates and their families reported to tax authorities. Specifically, they contain no information on the various forms of capital income declared under the head called "tax-exempt income".

To estimate the total income reported by individuals as taxable, we use statistics published by the CBDT. First, we examine the relationship between net taxed income and total taxable income in the statistics available from the CBDT; we then use this relationship to estimate the latter from the former for the GE dataset. In addition, we use CBDT data to calculate the top income levels reported in the ITRs. For this purpose, we derive the Generalised Pareto Interpolations (GPIs) for the annual statistics published by the Tax Department. These interpolations are used to estimate the top levels for net taxed income and taxable income.

To estimate the various forms of tax-exempt capital income — such as agricultural income, dividend income, long-term capital gains, etc. — we use the information on asset ownership along with the rate of returns for different asset classes. For example, the dividend income is estimated as the value of stocks owned multiplied by the average dividend yield rate for the top 100 private listed companies. We use these estimates along with the relevant tax rules to compute the quantum of

[3] See Lancet and Piketty (2018) and Sahasranaman and Jensen (2021), among others.
[4] See Dynan (2009) and Chancel, Piketty, Saez, Zucman, et al. (2022).

capital and labour incomes reported to tax authorities. Further details on this are provided in Sections 4 and 5.

We find that the reported income as a proportion of wealth decreases with the wealth. On average, the wealthier a household is, the smaller its reported income is relative to its wealth. This decreasing trend in the income-wealth ratio persists for all versions of income reported to tax authorities, namely net-taxed income, gross income reported as taxable, and total income declared, including the capital income reported under the category of tax "exempt income". The decreasing trend holds for individuals as well. The income-wealth ratios for HHs and individuals decrease continuously with wealth, taking very low values at the top wealth levels.

According to our estimates, for the bottom 10% of households, reported income is almost double their wealth. In contrast, for the top 1% of families, the total reported income amounts to just 3–4% of their wealth. For the wealthiest 0.1%, the total reported income adds up to less than 2% of their wealth. For the most affluent ten families in the FL, the reported income adds up to just about half a per cent of their wealth. The relationship between individual wealth and reported income exhibits a very similar pattern. Even if we ignore the labour income and consider only the capital income as a reference point, our estimates suggest that the income reported in ITRs of the wealthy and super-wealthy groups is a small fraction of the returns from their wealth.

It is worth emphasising that the wealthy groups' income missing from tax returns can lead to higher investment by these groups resulting in higher employment and growth rates for the entire economy. As macroeconomic analysis is beyond the scope of this paper, we focus on estimating the quantum of the reported income and the factors behind the relatively low levels reported by the wealth groups.

We show that the portfolio choices by different wealth groups and the tax rules account for much of the decreasing trend in the reported income-wealth ratio. Tax avoidance is an important factor behind relatively small values of the reported income-wealth ratio observed for the super-wealthy groups. In addition, we find evidence of tax evasion across wealth groups.

As the dynamics of capital income are similar across market economies, our findings should be relevant beyond the Indian context. Specifically, our study contributes to four kinds of literature. First, by examining the relationship between wealth and reported income for individuals and HHs across wealth groups, it contributes to an area that has remained under-researched in the Indian and international contexts. From Dynan (2009), Piketty (2014), OECD (2018), and Chancel, Piketty, Saez, Zucman, et al. (2022), one can infer only the aggregate income wealth ratios for select wealth groups. Our findings show that the broad patterns discernible from these studies hold at the individual and household levels.

Second, we contribute to the debate on the progressivity of a tax regime. We show that to reduce their tax liability, the affluent groups transfer only a fraction of the returns from the capital to their personal accounts. Consequently, a significant fraction of their capital income does not find its way to the income tax data. Our findings underscore the case for a wholesome assessment of income tax regimes. The standard approach considers a regime progressive if the applicable marginal tax rates increase with the reported income. See Nayak and Paul (1989), Piketty and Quin (2009), Besley and Persson (2014), CBGA-India (2015), Chancel and Piketty (2019), and Datt, Ray and Teh (2022), among others. However, when the tax rules allow the affluent to choose how much of their income gets reported in the income tax accounts, the question we should ask is this: Is the tax regime progressive with respect to the actual total income as opposed to merely the income reported to tax authorities?

We present evidence suggesting that the Indian tax regime is regressive vis-à-vis the total income as opposed to the income reported in the ITRs. Our most generous estimates suggest that the tax paid by the wealthiest 5% of individuals amounts to less than one-fifth of their capital income, and the tax liability of the wealthiest 0.1 percentile is just about one-tenth of their capital income. Super-wealthy Indians on the FL pay tax amounting to a mere 5% of their capital income.

The tax rules render the tax regime even more regressive with respect to wealth. We show that at top wealth levels, the wealthier a taxpayer, the smaller the income tax paid relative to wealth. For the wealthiest centile, the tax liability amounts to about 1% of their wealth. For the wealthiest 0.1% of individuals, the tax liability amounts to approximately 0.7% of the wealth. Super-wealthy Indians in the FL face income tax liability amounting to just 0.4% of their wealth. The relative tax liability of the ultra-wealthy groups is lower than that of the middle-wealth groups, even after considering the various exemptions granted to the latter under the tax law.

Third, our results are also relevant to the literature on income inequality in India. Several studies have estimated income inequality using the statistics published by the Income Tax Department and other sources such as the National Sample Survey Office (NSSO), the Central Statistics Office (CSO), and the Reserve Bank of India (RBI).[5] As discussed below, the income tax data used by these studies miss a part of the opulent group's income and thus underestimate inequality.

Fourth, by showing that the set of the top income-rich Indians differs from the country's wealthiest individuals, our results supplement similar findings in international contexts, as seen in Piketty (2014) and Chancel, Piketty, Saez, Zucman, et al. (2022). Finally, our results are also relevant for studies on the differential effect of transparency on reporting behaviour. In line with the broad findings reported in Djankov, La Porta, Lopez-de-Silanes, and Shleifer (2010), Bø, Slemrod, and Thoresen. (2015), Libman, Schultz and Graeber (2016), and Szakonyi (2022), we find that people exposed to media and civil society scrutiny have a stronger incentive to report their incomes truthfully.

The paper is organised into the following sections. Section 2 introduces a mathematical model that provides an analytical framework for the empirical analysis. Readers not keen on technical details may skip Section 2 and proceed directly to Section 3, which discusses the datasets and the summary statistics used in this study. Section 4 summarises the methodology used to estimate various types of reported income. Section 5 presents our findings on the relationship between the different types of income reported by the taxpayers on the one hand and their wealth on the other. Section 6 presents the regression results on the determinants of income-wealth ratios for individuals and households. Section 7 examines the proportion of total individual income missing from the tax returns filed to the tax authorities. It also discusses the mechanisms and tax rules that lead to partial income reporting by opulent groups. Section 8 discusses the implications of the missing income for the progressivity of the tax regime and the existing estimates of income inequality in India. In Section 9, we offer concluding remarks. The methodology used to estimate the various kinds of income reported to the tax authorities are provided as an online Appendix.

[5] See Ojha and Bhatt, (1964), Banerjee and Piketty (2005), Sarkar and Mehta (2010), Basole (2014), Ahmed and Bhattacharya (2017), Sinha et al. (2017), Assouad, Chancel and Morgan, M. (2018), Chancel and Piketty (2019), Sahasranaman and Jensen (2021), Ghatak (2021), and Datt, Ray and Teh (2022), among others.

2. Wealth, Income and Reporting Rules: A Model

In this section, we develop a conceptual framework to answer the question: What relationship should we expect between different forms of income and wealth? Readers not interested in these technical details may skip to the next section.

Following Djankov, La Porta, Lopez-de-Silanes, and Shleifer (2010), Piketty (2014), Asher and Novosad (2019), and Fisman et al (2019), we consider an individual's wealth to be the market value of all assets owned, net of all debts owed. A household's wealth is simply the sum of the wealth of its members. Our definition of wealth includes all assets (financial and non-financial) together with consumer durables and jewellery on which the ownership rights can be enforced, including the right to sell on the market. Following Piketty (2014), we use the terms wealth and capital interchangeably.

We denote wealth by W and income by Y. Personal income, Y_{PI}, consists of all the earnings by an individual in a given year. It has two components: labour and capital income. Let Y_L and Y_K denote the labour and capital income respectively. Thus, personal income $Y_{PI} = Y_L + Y_K$.

The labour income of an individual is the annual total remuneration received for services provided. It includes earnings in the form of salary or wages, commissions, honoraria, etc. The capital income, Y_K, on the other hand, is the total annual returns from the wealth owned by the individual. Specifically, capital income is the sum of economic returns from all assets combined. It includes dividend income from stocks, interests from deposits, rental income from property, equity income from stakes owned in estates and trusts, and profits from corporations, sole proprietorships, and partnerships. It also includes capital gains from assets owned. We use the terms *total capital income* and *returns from wealth* interchangeably.

The total capital income can be split into two categories. The first is what we call direct returns or *"direct capital income"*. This is the "regular" income from capital. For example, rent is a regular direct income from a commercial property. Similarly, a residential property generates direct income as rent if leased out, and as "imputed rent" in the case of self-occupied dwellings. Interest is a direct income from instruments such as bonds, bank deposits, and savings accounts. Profits are a direct income from the ownership of firms, sole proprietorships, and partnerships. Company stocks also provide direct payment in the form of dividends.

In addition, a capital asset provides economic returns in the form of *capital gains*, defined as the appreciation in the market value of the asset. Wealth assets, such as residential and commercial properties, stocks, and equities, tend to appreciate over time, leading to capital gains for owners. Capital gains from an asset remain unrealised unless the asset is exchanged or sold. If realised, capital gains must be reported to the tax authorities as capital income. We term the unrealised capital gains *"indirect capital income"*. The total income from an asset is the sum of the direct and indirect income.

Note that so far, we have not included the "realised" capital gains as part of capital income. This neglect is deliberate and temporary. We will revisit this issue later in this section.

To formalise the relationship between wealth and capital income, let us suppose that wealth consists of n assets; $1, \dots, n$. Let the market value of these assets be $A_1, A_2, \dots, A_n$, respectively. Let $A = A_1 + A_2 + \cdots + A_n$. Thus wealth $W = A - L$, where $L \geq 0$ denotes the liability of the individual or family. By definition, $\frac{W}{A} \leq 1$, and it is plausible to assume that $W'(A) > 0$. We define $s_i = \frac{A_i}{A}$, i.e., s_i is the share of the first asset in the asset portfolio. Let, y_{iD} and y_{iI} denote the direct and indirect

annual (income) from asset i respectively.

The total annual return from an asset is the sum of the direct and indirect income generated by it. So, the total (yearly) returns from an asset i is $y_i = y_{iD} + y_{iI}$. The total direct capital income from all assets combined is $Y_{KD} = y_{1D} + y_{2D} + \cdots + y_{nD}$. The total indirect income is $Y_{KI} = y_{1I} + y_{2I} + \cdots + y_{nI}$. The total capital income from all assets is $Y_K = Y_{KD} + Y_{KI}$. For simplicity, let us assume $y_{iD} > 0$ for all $i = 1, \dots, n$, so $Y_{KD}>0$.

Let $r_i = \frac{y_i}{A_i}$ denote the rate of annual returns on asset i. Now we can express Y_K as $Y_K = \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \mathrm{r_I}\, A_i$. Moreover, we can rewrite $r_i = \frac{y_{iD}}{A_i} + \frac{y_{iI}}{A_i} = r_{iD} + r_{iI}$, i.e., the rate of total returns is simply the sum of the rates of direct and indirect returns from the asset.

The available evidence suggests that, the riskier an asset is, the higher the rate of return on it, and vice-versa. For instance, stocks and shares are riskier assets than commercial properties, which are in turn riskier than fixed-term bank deposits. The rate of returns follows the same order even after factoring in applicable taxes. On average, rates of return on stocks and shares are higher than those on property investments, which are typically more rewarding than fixed-term deposits.[6]

Without loss of generality, let us assume that the riskiness of the asset increases with index $i = 1, \dots, n$; that is, asset k is riskier (more volatile) than asset j, if $j < k$. The higher-risk-higher-returns relationship implies that r_i increases with the index i. Formally, if $j < k$ then $r_j < r_k$. As to the relationship between W and A, we assume that the ratio $\frac{W}{A}$ increases with A. Thus, the higher the worth of the assets is, the smaller the liability will be as a ratio of assets.

2.1 The Direct and Indirect Capital Income

Now, consider two individuals at wealth levels W and $\widehat{W}$ with the corresponding asset levels A and $\hat{A}$ respectively. A and $\hat{A}$ may or may not be equal. Suppose the individual at wealth W has asset allocation as $A_1, A_2, \dots, A_n$. The share of ith asset in the first portfolio is $s_i = \frac{A_i}{A}$. Let the individual with asset $\hat{A}$ choose allocation $\hat{A}_1, \hat{A}_2, \dots, \hat{A}_n$, with $\hat{s}_i = \frac{\hat{A}_i}{\hat{A}}$. We consider the asset allocation $(\hat{A}_1, \hat{A}_2, \dots, \hat{A}_n)$ to be riskier relative to the allocation $(A_1, A_2, \dots, A_n)$ if the following holds: For all $k \leq 1, \dots, n$,

$$\sum_{i=1}^{k} \hat{s}_i \leq \sum_{i=1}^{k} s_i \qquad (2.1)$$

The above inequality is strict for at least some k. Simply put, portfolio $(\hat{A}_1, \hat{A}_2, \dots, \hat{A}_n)$ is riskier than $(A_1, A_2, \dots, A_n)$, if the former assigns a larger share of investment to the risky assets. Several studies, including this one, show that the wealthier an individual, the larger is the share of risky assets in their portfolio and vice versa.[7] Accordingly, we assume that individuals exhibit increasing appetite for risky assets as their wealth grows. Specifically, assume that for wealth levels W and $\widehat{W}$, whenever $W < \widehat{W}$ the relationship in (2.1) holds.[8] This assumption and that r_i is increasing in i,

[6] Campbell et al. (2019) shows that larger account holders diversify more effectively and thereby earn higher average. For a review of literature on relationship between riskiness of portfolio, returns, and wealth see Bach, Calvet, and Sodini (2018), Fagereng, et al (2020) and Wojciech and Zwick. (2020).

[7] See Guiso and Paiella (2008), OECD (2018), and Section 3 of this paper.

[8] Formally, we assume individuals are risk averse with von Neumann-Morgenstern utility function $u(W)$ such that $-u''(W)W/u'(W)$ is decreasing in W. For simplicity, assume that there is no discounting. Individual

leads to the following inference:

$$\widehat{W} > W \Rightarrow \sum_{i=1}^{n} r_i \hat{s}_i > \sum_{i=1}^{n} r_i s_i \tag{2.2}$$

In other words, the weighted rate of returns increases with wealth. In addition to the effect of decreasing risk aversion on portfolio choices, and the scale effects, the average rate of returns increases with W on account of several other factors. For instance, investment opportunities expand with wealth. The wealthy are better at spotting investment opportunities and can even afford to hire financial advisors to earn higher returns on their investment(s), especially from equities, bonds, and other financial assets. Moreover, wealthy individuals have more bargaining power vis-à-vis the lenders. Thus, their relative cost of borrowing — and hence their burden of debt servicing — is relatively low. This, in turn, implies increasing returns to wealth. On all of these counts too, capital income is expected to be an increasing and convex function of wealth.

To be clear, the above assumptions are not a logical necessity. They are motivated by what is observed in the data examined in this study and several others.[9] Formally put, the above assumptions imply $Y_K'(W) > 0$ and $Y_K''(W) > 0$. Of course, there can be no capital income without wealth, so $Y_K(0) = 0$. It can be seen that $Y_K(0) = 0$, $Y_K'(W) > 0$ and $Y_K''(W) > 0$ together imply:

$$\frac{\partial}{\partial W}\left(\frac{Y_K(W)}{W}\right) > 0. \tag{2.3}$$

In summary, due to the decrease in risk aversion with wealth, and on account of the other factors discussed above, an increase in wealth leads to more than proportionate increases in total capital income. However, this logic does not extend to the two components of capital income: direct and indirect capital income. This is because the relationship between wealth and the rate of direct returns, on one hand, and wealth and indirect returns, on the other, is very different.

To illustrate this point, let us revisit the case of company stocks, real estate, and fixed-term deposits. As mentioned above, risk and total returns go hand in hand. Accordingly, for any given amount of investment, on average, the total returns are the highest for stocks, followed by real-estate, which are in turn followed by instruments such as fixed-term deposits with banks. Now consider the direct income from these three assets. The direct income from company stocks is dividends. Direct returns from real estate are rents, and from fixed deposits are interest incomes. Save for some exceptions, the rate of direct returns is the highest for fixed-term deposits (upward of 7–8%), relatively low for real estate (2–4%), and the lowest for equity (1–2%).[10] In other words, there is an inverse relationship between the rate of direct returns and the riskiness of an asset. According to available evidence, the inverse relationship between the risk and the rate of direct returns holds for most assets and continues to hold even after we factor in taxes on the direct income.

In Section 6, we will examine the underlying causes behind the inverse association between

investors choose their asset portfolios to maximise the expected utility of the terminal wealth, including direct and indirect returns. With these assumptions, one can show that the wealthier an individual is, the greater their share of riskier assets is, and vice versa (see Cass and Stiglitz, 1972).

[9] See Bach, Calvet, and Sodini (2018), Fagereng, et al (2020) and Wojciech and Zwick. (2020).

[10] In most countries examined in empirical studies, while the average rental yields are in the range of 4% to 7%, the dividend yield rates range from 2-5%. See Jorda et al. (2017), Demers and Eisfeldt (2022), Eichholtz et al. (2021) and Jorda et al. (2017) for evidence across several countries. In India, interest rates have been low of late but fixed-term interest rates are in the range of 7–8%. On the other hand, rental incomes tend to be between 2–4% of the property value. In contrast, the average dividend income from stocks of the top 500 private listed companies is less than 2% of the market values of these assets.

riskiness and the rate of direct returns. For now, we take as given the observed relationships between risk and the direct rate of returns. Accordingly, we assume that r_{iD} is decreasing in i as the latter is an index of the asset's riskiness. It can be seen that when r_{iD} is decreasing in i and appetite for risk is increasing in wealth (i.e., whenever $W < \widehat{W}$, the relationship in (2.1) holds) we get the following result:

$$\widehat{W} > \mathrm{W} \rightarrow \sum_{i=1}^{n} r_{iD}\hat{s}_i < \sum_{i=1}^{n} r_{iD}s_i \tag{2.4}$$

i.e., $\widehat{W} > \mathrm{W}$ implies $\frac{\sum_{i=1}^{n} r_{iD}\hat{A}_i}{\hat{A}} < \frac{\sum_{i=1}^{n} r_{iD}A_i}{A}$, i.e., $\frac{\hat{Y}_{KD}}{\hat{A}} < \frac{Y_{KD}}{A}$. Since the ratio $\frac{A}{W}$ is decreasing in A and hence in W, we get $\frac{\hat{Y}_{KD}}{\widehat{W}} < \frac{Y_{KD}}{W}$. In other words, the average rate of direct returns decreases with wealth. Formally put,

$$\frac{\partial}{\partial W}\left(\frac{Y_{KD}(W)}{W}\right) < 0 \tag{2.5}$$

While the rate of direct returns decreases with the asset's riskiness, the share of risky assets increases with wealth. These two aspects of the capital income imply that the rate of direct returns, i.e., the ratio of the direct income to wealth, $\frac{Y_{KD}(W)}{W}$, is decreasing in W. However, from (2.3) we know that the ratio of the total capital income to wealth, $\frac{Y_K(W)}{W}$, increases with wealth. Now,

$$\left[\frac{\partial}{\partial W}\left(\frac{Y_K(W)}{W}\right) > 0 \text{ and } \frac{\partial}{\partial W}\left(\frac{Y_{KD}(W)}{W}\right) < 0\right] \Rightarrow \frac{\partial}{\partial W}\left(\frac{Y_{KI}(W)}{W}\right) > 0.$$

Simply put, the rate of indirect returns is increasing in wealth.[11]

Next, consider the labour income, Y_L. This depends on several factors such as the quality of individual health, education, work experience, and labour market conditions, among other things. Plausibly, the labour income depends on wealth — a great facilitator of access to quality healthcare and education, and hence an essential determinant of labour income. Moreover, for any given level of education, the wealthy enjoy better employment opportunities with remunerative wages, especially at low- and medium-wealth levels. A considerable body of evidence supports the positive relationship between wealth and labour market outcomes.[12] At very high wealth levels, though, the effect of wealth on wages is expected to be modest at best.

Accordingly, we take that controlling for other factors, Y_L is an increasing and concave function of W. Moreover, labour income can be positive even when the individual has no wealth at all. Formally, we assume $Y_L(0) > 0$, $Y_L'(W) > 0$ and $Y_L''(W) < 0$. Now, differentiating $\frac{Y_L(W)}{W}$ with W, we get $\frac{\partial}{\partial W}\left(\frac{Y_L(W)}{W}\right) = \frac{WY_L' - Y_L}{W^2} = \frac{g(W)}{W^2}$, where $g(W) = WY_L' - Y_L$. Since $Y_L'' < 0$, so $g'(W) < 0$, i.e., $g(W)$ is decreasing in W. Moreover, as $Y_L(0) > 0$ and $g(0) < 0$, $\frac{g(W)}{W^2}$ is negative for all $W \geq 0$. Therefore,

$$\frac{\partial}{\partial W}\left(\frac{Y_L(W)}{W}\right) < 0. \tag{2.6}$$

Summing up, we have the following result.

[11] The unrealised capital gains seem to be a significant component of the increasing returns on the wealth. Piketty (2014, Chapter 12), Saez and Zucman (2016), and Kaymak, Leung, Poschke (2020).

[12] See, for example, references in https://www.urban.org/sites/default/files/publication/49116/2000178-How-are-Income-and-Wealth-Linked-to-Health-and-Longevity.pdf

Proposition 1: The average rate of direct capital income, i.e., $\frac{Y_{KD}(W)}{W}$, and the labour income as a ratio of wealth, i.e., $\frac{Y_L(W)}{W}$, are decreasing in wealth.

2.2 Income Reporting Rules and Tax Evasion

Now, we consider the relationship between wealth and reported income. Like most other tax regimes, the Indian law require reporting of all forms of labour income, i.e., Y_L. As to the capital income, only the direct capital income, i.e., Y_{KD} must be reported; Y_{KI} is not required to be reported. In other words, out of the total personal income only $Y_L + Y_{KD}$ must be reported. In view of the above we can predict how the sum $Y_L + Y_{KD}$ will vary with wealth. We describe the term $Y_L + Y_{KD}$ as the direct personal income and denote it by Y_{PID}. That is,

$$Y_{PID} = Y_L + Y_{KD.}$$

Since $\frac{Y_L(W)}{W}$ and $\frac{Y_{KD}(W)}{W}$ are both decreasing in W,

$$\frac{\partial}{\partial W}\left(\frac{Y_{PID}(W)}{W}\right) < 0.$$

Assuming no tax evasion, we have the following claim.

Proposition 2: The personal income as a ratio of wealth decreases as wealth level increases.

As discussed above, risky assets tend dominate the portfolios held by the wealthy groups. While the wealth seems to be a major determinant of portfolio choices, the possible heterogeneity in the propensity towards risk and other factors have a bearing on portfolio choices.[13] Without going into underlying causes, let us consider the effect of portfolio choices on reported income, while holding the wealth level fixed. Since the rate of returns varies across assets, it is obvious that the direct capital income and hence the personal income will differ across portfolio choices, even when the total wealth remains the same. In terms of our notations, for a given W, portfolio choices affect the direct and indirect capital incomes, i.e., Y_{KD} and Y_{KI}.

Specifically, in the context of the relationship (2.1), assume $\widehat{W} = W$. That is, portfolio $(\hat{A}_1, \hat{A}_2, \dots, \hat{A}_n)$ is riskier than $(A_1, A_2, \dots, A_n)$, even though total wealth is the same. It can be seen that (2.1) in view of the assumptions that r_{iD} is decreasing in i, gives us $\sum_{i=1}^{n} r_{iD}\hat{s}_i < \sum_{i=1}^{n} r_{iD}s_i$. This, in turn, means that we must have $\hat{Y}_{KD} < Y_{KD}$. In simple terms, for any given level of wealth, the direct capital income decreases with the riskiness of the asset portfolio. Since, only direct capital income and labour income are reported, we have the following claim.

Proposition 3: Assuming no tax evasion, the ratio of reported income to wealth, $\frac{Y_{PID}(W)}{W}$, depends on the shares of various assets in the portfolio. Ceteris paribus, the ratio $\frac{Y_{PID}(W)}{W}$ decreases with riskiness of the portfolio.

For Propositions 2 and 3 it bears emphasising that our definition of direct capital income includes direct returns from all assets constituting the wealth. However, these propositions are equally valid if some specific asset, say residential property, is dropped from the definition of W as long as the capital income from the dropped asset (imputed rent) is also excluded from $Y_{KD.}$. It is crucial to keep the set of assets same for W and $Y_{KD.}$.

Further, note that we have not considered the unrealised capital, neither as a part of the direct capital income nor under the indirect income from capital. Moreover, as is shown in Section 4, the

[13] For a review of literature, see Curcuru, S. at el (2010).

realised capital gains are only a tiny fraction of the direct personal income. Therefore, we expect the prediction in Propositions 2 and 3 to hold both with and without factoring in realised capital gains as a part of the direct capital income.

Finally, consider the effect of tax evasion by underreporting of capital income. It is straightforward that underreporting of the direct capital income from any one or more of the assets will reduce the reported values of ratios $\frac{Y_{KD}(W)}{W}$ and $\frac{Y_{PID}(W)}{W}$ to less than their actual values.

Now, consider the effect of an increase in share an asset whose income can easily be underreported. Underreporting of capital income from an asset is tantamount to the reported rate of return being less than the actual rate of return. Further, an increase in the share of an underreported asset essentially means that among the reported rates of return a larger share (weight) is assigned to assets with lower rates of return. Specifically, let

r'_{iD} denote the reported rate of return.

Start from a situation where wealth levels is fixed, there is no underreporting, and asset shares are $s_i, i = 1, \dots, n$. Now consider a change in reported returns such that $r'_{jD} < r_{jD}$ for some j, $1 < j < n$, and $r'_{iD} = r_{iD}$ for all $i \neq j$. Further, let the share of the underreported asset increase to s'_j, i.e., $s'_j > s_j$. This increase in the share of underreported asset will entail adjustments in other asset shares. Let the adjusted asset shares be denoted by s'_i for $i = 1, \dots, n$. If $\sum_{i=1}^{k} s'_i \leq \sum_{i=1}^{k} s_i$ holds for all $k = 1, \dots, n$, then we get $\sum_{i=1}^{n} r'_{iD} s'_i < \sum_{i=1}^{n} r_{iD} s_i$. This, in turn, leads to $\frac{Y'_{KD}(W)}{W} < \frac{Y_{KD}(W)}{W}$, where $Y'_{KD}(W)$ and $Y_{KD}(W)$, respectively, are the reported direct capital income. This immediately gives us: Cateris paribus, $\frac{Y'_{PID}(W)}{W} < \frac{Y_{PID}(W)}{W}$, where $Y'_{PID}(W) = Y'_{KD}(W) + Y_L(W)$. In simplistic terms, we get the following proposition.

Proposition 4: An increase in share of an underreported asset associated with a decrease in the share of relative safe assets,[14] leads to a reduction in the reported values of ratios $\frac{Y_{KD}(W)}{W}$ and $\frac{Y_{PID}(W)}{W}$ to less than the actual values.

3 Data Sources and Preliminary Findings

3.1 Data Sources

We work with several data sources including ProwessIQ, the Forbes List of billionaires, data on Income Tax Returns, and annual accounts of listed companies managed by the wealthiest families in India. Below, we provide a brief description of these data sources and summary statistics relevant to our study.

General Election (GE) Data: The data is based on the sworn affidavits submitted by contestants for elections to the Lok Sabha, the lower house of Parliament of India. Specially appointed returning officers scrutinise the affidavits for accuracy, correctness, and completeness. The GE data is the only source that simultaneously provides information on both income and wealth for individuals and households. *"Myneta"*,[15] an online platform run by the Association for Democratic Reforms (ADR), offers easy access to the information contained in these affidavits in the form of digitised records. These records are the primary source of our GE dataset. We have verified the accuracy of the digital

[14] In our dataset, we find that increase in share of farmland and commercial property is highly correlated with a decrease is share of less risky assets. See online Appendix-I.
[15] https://www.myneta.info

records for a small sample of randomly selected affidavits directly taken from the Election Commission of India (ECI) website.[16]

Even though there have been 17 General Elections since the independence of India, only in 2011 did the ECI mandate the declaration of wealth and income for election contestants. Therefore, only affidavits filed in the last two GEs — 2014 and 2019 —provide information on wealth and incomes of the candidates, their spouses, and dependents (i.e., the candidates' family or household).

Table 3.1 below describes the assets reported in the affidavits and their broad categories. The liabilities comprise all types of loans and dues owed to government agencies. Wealth is defined as the value of all assets owned minus the liabilities.

Table 3.1: Categorisation of GE asset types

Assets Category	GE Assets Type
Land	Agricultural Land + Non-Agricultural Land
Durables	Motor Vehicles + Other assets, such as values of claims/interests
Buildings	Commercial Buildings + Residential Buildings + Other Immovable Assets
Shares	Bonds, Debentures and Shares in companies and firms
Deposits	Cash + Deposits in Banks, Financial Institutions and Non-Banking Financial Companies + NSS (National Savings Schemes), Postal Savings, etc. + LIC or other Insurance Policies
Jewellery	Gold and Jewellery
Receivable	Personal loans/advances given

We combine observations on wealth and income from GE 2014 and GE 2019 to present an overall picture. As they pertain to two different points in time, the GDP deflator is used to convert wealth and income levels to March 2019 prices.[17] Our main results, however, hold without merging the data from the two GEs. The HHs from the two GEs have very similar demographic attributes such as genders, castes, educational qualifications, and professions.

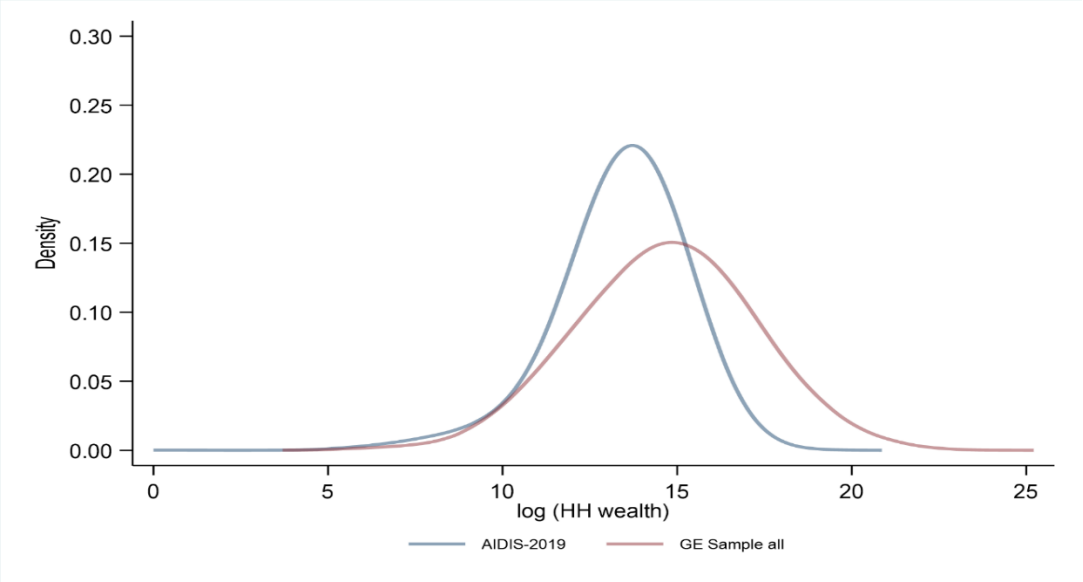
Put together, the affidavits from the two GEs provide wealth information for 16097 contestants and families members. Figure 3.1 also shows the wealth distribution for the households covered in the GE data vis-à-vis the wealth distribution for the Indian population emerging from the All-India Debt Investment Survey (AIDIS) of 2019, at constant prices.
As is evident from the plots, an average politician is wealthier than the average Indian. The elected politicians are even wealthier. In 2019, the average family wealth of Lok Sabha members was USD 998,311, whereas the average household wealth (as per NSSO 77th round) was just USD 26,867. This is in line with what has been observed in other countries. In general, politicians are wealthier than the rest.[18] Of course, the wealth differences between politicians and non-politicians vary across countries. For instance, Indian politicians are much more affluent than the general population. The gap between the average family wealth of politicians and the average wealth of the population is 1 to 37 in India versus 1 to 4 in the case of the US[19], and possibly even higher in China.

Comparing elected politicians in our dataset with their counterparts in other countries paints a

[16] https://affidavit.eci.gov.in/
[17] For this purpose, we take the value of the adjusted deflator in January–March 2014 (around 97.28) and during January–March 2019 (around 112.35).
[18] For discussion, see Piketty (2014). For a comparison of politicians' wealth vis-à-vis the country averages, for Sweden see Bó et al. (2017); for United Kingdom see Eggers and Hainmueller (2009); for the USA see Lenz and Lim (2009), for Italy see Gagliarducci, Nannicini, and Naticchioni (2008).
[19] Calculations are based on the data available at the online forum Open Secretes for the US, and the Harun list for China.

picture of sharp contrast. Though the Indian politicians are much wealthier than the rest, even the most opulent among Indian politicians are a no match to their American and Chinese counterparts. In 2019, the ten most affluent members of the Lok Sabha had a total wealth of USD 0.3 billion. In contrast, according to Open Secretes, in 2018, the combined wealth of the ten wealthiest members of the House of Representatives was USD 1.2 billion. The combined wealth of the wealthiest ten members of China's National People's Congress (NPC) in 2018 was USD 132.3 billion, much higher than Indian and American members.[20]

Figure 3.1: Household Wealth distribution: GEs vs AIDIS



In addition to wealth, each contestant must disclose the amount reported as the "total income" in the income tax return (ITR) forms filed for themselves, their spouse, and dependents. In the ITR forms, the term "total income" refers to the net taxed income calculated as the taxable income reported by the taxpayer minus the "deductions" available to them under the tax rules. In other words, this is the net income taxed in the taxpayer's "hands". By definition, the net taxed income is less than the total amount reported taxable. For example, assume a taxpayer who reports a taxable income of ₹10 lakhs but is eligible for tax deductions amounting to ₹2.6 lakhs. So, his taxed in-hand income is ₹10-2.6 lakhs (i.e., ₹7.4 lakhs). If this taxpayer were to contest an election, ₹7.4 lakhs would be the income reported in his affidavit. In other terms, the income reported in an affidavit does not include the part of the reported income that qualifies as deductions under the tax rules. We thus cannot know the total taxable income reported by the candidate to the tax department from an affidavit.

Additionally, the affidavits do not cover income declared under the category called "tax-exempt income", which includes incomes such as agricultural income. In other words, the affidavits do not provide information on the entire income reported by the candidates and their families in the ITRs. Section 4 discusses this in greater detail.

For as many as 8,501 candidates, we found income information to be entirely missing, as these candidates did not report their income. We dropped all households (HHs) with missing income from our analysis. According to the Election Commission, the punishment for inaccurate and inaccurate reporting includes fines, imprisonment for up to six months, and disqualification from the contest. While one cannot rule out misreporting altogether, it seems plausible to assume that most cases

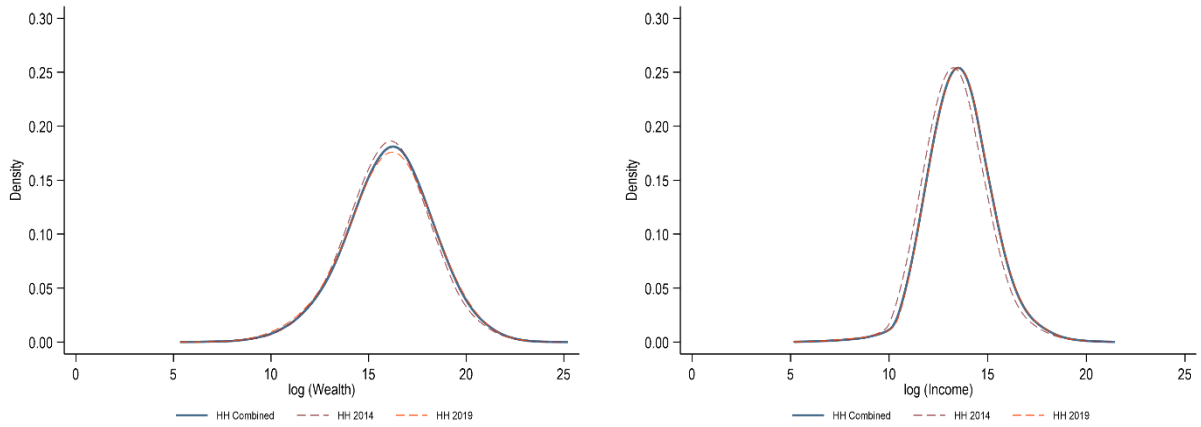
[20] Figures of China are computed by matching Hurun list of wealthiest for 2018 with NPC delegate list to identify the wealthiest members of the latter. Of course, if the fraction hidden wealth is higher for the Indian politicians than their foreign counterparts, the actual differences will be smaller.

with missing income pertain to candidates whose families earn less than the ₹2.5 lakh threshold for filing ITRs. This inference is supported by a comparison of the wealth distribution in the sample we used in the study vis-a-vis the wealth distribution for the data points dropped because of missing income information. See online Appendix_I. The most observations in the *dropped sample* (that is, units with missing income information) are low-wealth households who plausibly are also low-income households.[21] A hypothetical inclusion of the dropped sample would only marginally affect our claims related to middle-wealth groups. In particular, it is unlikely to make material difference to our findings on the wealthy and super-wealthy groups.

For the 7,596 households in the used sample, we have information on wealth as well as income. The General Elections of 2014 and 2019 each account for roughly half of these observations. Below we present density plots for the distribution of the log of wealth for GE 2014 and GE 2019 separately as well as the two years combined (i.e., the entire sample used).[22] The average income and wealth are higher for 2019 than for 2014. Figure 3.2 shows wealth and income distributions of HHs in the used sample.

Figure 3.2: Wealth and income distribution of HHs for GE 2014, GE 2019, and combined

(a) Wealth (b) Income



Note: (a) Density plots for log (Household Wealth) and log (Household Income) are based on the Epanechnikov kernel function; (b) Half-width of the kernel is set at 0.9 for log (Household Income) density plot; (c) log (Household Income) is based on the returned income as disclosed in the election affidavit.

Table 3.2 shows the average HH wealth and income across wealth percentiles. As expected, average income increases with wealth. The share of riskier assets such as equity and commercial properties also increases in tandem with wealth (also see Figure 3.3 below). The GE data also exhibits several properties known of the wealth distribution and, separately, the income distribution in India; such as, the concentration of wealth and income in a few hands,[23] and male dominance over family income and wealth, among others. As can be seen in Appendix-I, income in the GE data is less

[21] Most likely, these households do not have to file tax returns as their income is below the taxable threshold, so the income requirement in GE affidavits is irrelevant for these households. It is pertinent to mention that some of the dropped observations belong to the middle and upper-middle wealth levels. An examination of such cases suggests that in such cases, most of the wealth is farmland and residential property. As farm income is not taxable, such families do not file tax returns and hence must not report their income in the affidavits.

[22] In principle, we should fit a distribution that is left truncated at Rs. 2.5 lakhs. But given the focus of this paper, we estimate log normal distributions based on the observed data alone.

[23] See Lancet and Piketty (2018) and Sahasranaman and Jensen (2021), among others.

concentrated than wealth — a pattern that has been observed in international contexts too.[24] All Appendices are available online.

Even though the average household in the GE data is wealthier than the average Indian family, the dataset does not cover the top wealth levels. The highest wealth reported in the GE data was US$1.3 billion at 2019 prices. On the other hand, the least wealthy family in the 2019 FL had a total wealth of US$1.4 billion. To cover the entire spectrum of wealth distribution in India, we thus supplement the GE data with data from the FL. Together, the GE and FL data cover the entire range of wealth spectrum in India in that the pooled data has HHs with negative, zero and very little wealth to those figuring at the very top of the wealth ladder. However, its coverage of the right tail is even better.

Table 3.2: Household wealth and income (GE data) across percentiles

Wealth Percentile	No. of HH	Avg. HH Wealth	Avg. HH Income	'Safe Assets' % Total Assets	'Risky Assets' % Total Assets
p0 - p5	380	-39,51,493	6,41,146	56.2	5.1
p5 - p10	380	2,90,253	3,51,043	39.8	2.8
p10 - p15	379	7,09,479	4,48,498	30.6	6.7
p15 - p20	381	13,14,795	4,88,736	25.2	9.0
p20 - p25	380	20,54,144	5,20,649	22.5	9.5
p25 - p30	379	30,02,482	5,42,202	19.9	14.3
p30 - p35	381	40,18,420	5,73,599	17.4	12.3
p35 - p40	380	52,78,640	7,33,610	17.2	13.8
p40 - p45	380	67,79,390	7,81,518	15.7	13.2
p45 - p50	380	86,68,396	8,97,422	15.9	14.2
p50 - p55	380	1,10,24,009	9,41,918	15.6	16.9
p55 - p60	379	1,39,62,219	11,72,538	15.4	17.1
p60 - p65	381	1,81,36,096	12,75,765	15.0	19.3
p65 - p70	380	2,39,60,570	14,54,769	15.6	21.2
p70 - p75	380	3,22,21,960	17,82,296	12.6	22.7
p75 - p80	380	4,28,45,496	25,05,900	12.4	24.0
p80 - p85	380	6,05,33,676	26,57,566	10.1	27.5
p85 - p90	380	9,39,46,888	42,70,143	9.3	29.9
p90 - p95	380	18,36,24,992	78,24,278	9.0	32.8
p95 - p100	380	1,13,16,48,128	3,50,99,180	6.0	42.2
p99 - p100	76	3,57,65,61,408	9,86,49,136	3.7	47.3
p99.90 - p100	8	18,06,27,49,696	35,18,48,448	1.1	68.9

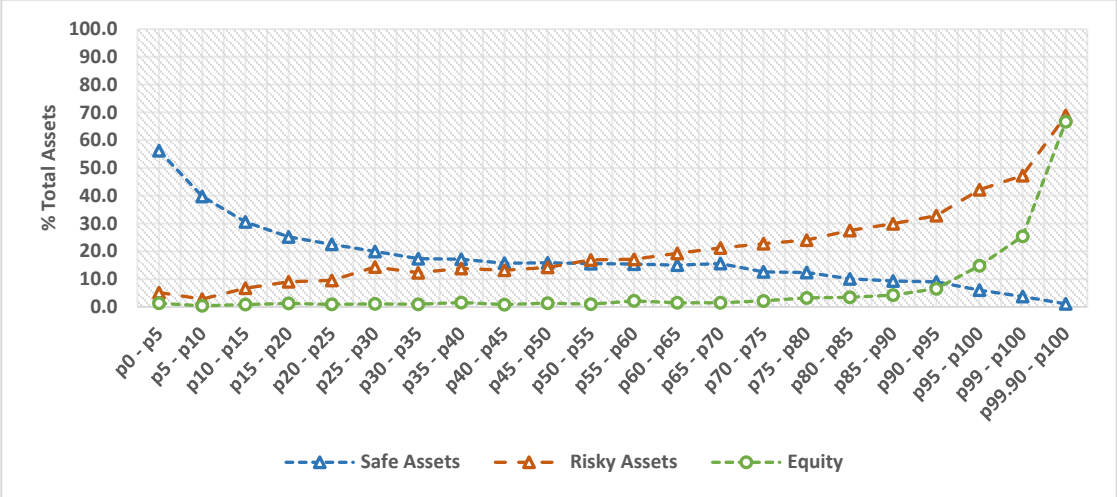
Note: a) Safe assets consist of cash, deposits in banks, financial institutions and non-banking financial companies and investments in National Savings Schemes and postal savings; b) Risky assets consist of equity, non-agricultural land, and commercial buildings; c) the two classes are not exhaustive.

Forbes List (FL) is an annual listing of the 100 wealthiest Indian families. The List comprise families of business tycoons and some CEOs at the top of the wealth pyramid. In recent years, a few promoters of start-ups and unicorns have also made it to the list. Only six of these families are headed by women. The FL provides information on family wealth. Notably, the definition of wealth used on the FL is the same as that used by us for the GE data.

[24] See Dynan (2009), Piketty (2014, Chapter 12), Saez and Zucman (2016), and Chancel, Piketty, Saez, Zucman, et al. (2022).

The concentration of wealth in the hands of the wealthiest Indians on the FL is several orders of magnitude higher than the concentration exhibited by the individuals in the GE data.[25] In recent years, the wealth held by billionaires in FL has come to account for an increasing fraction of the national income.

Figure 3.3: Different assets as a % of total assets across the wealth percentiles



Note: a) Safe and risky assets are as defined for Table 2.2 above, b) 'Equity', a part of risky assets, comprises bonds and debentures, shares and units in companies/mutual funds, and firm shares.

Income Tax Returns (ITR) data: For our analysis, we need to estimate the total income reported by different wealth groups. However, as mentioned above, the income disclosed in GE affidavits is only a part of the total income reported in the ITRs. We thus use statistics published by the Central Board of Direct Taxes (CBDT) and other sources to estimate the total income reported by the individuals and families in the GE dataset and those on the FL.

The CBDT statistics used by us are for the category of "individuals".[26] These statistics provide information on the number of ITRs, and the average income reported under various income brackets. They cover incomes ranging from zero to more than ₹500 crore. The two types of income covered by this data are what we have described above as the taxed-in-hand income and the reported taxable income.

The statistics on the taxed-in-hand income are extracted by the CBDT from the ITRs and clubbed together for the various income groups. This data is published as tables listing the number of taxpayers and the average incomes for different income brackets. The tax data provide similar information on the taxable income reported by the taxpayers from various income groups. Within an income bracket, the difference between the two types of income arises from multiple deductions and exemptions allowed on the declared value of the taxable income. For instance, in the example cited previously, a candidate's reported taxable income was ₹10 lakhs but the net taxed-in-hand income was ₹7.4 lakhs, with the difference between the two stemming from the deductions worth ₹2.6 lakhs availed by the candidate. Overall, deductions amounting to ₹2.5–₹4.5 lakhs can be availed depending on the investment decisions of the taxpayer. Summing up:

[25] According to Karmali (2021), In 2021, wealthiest family on the FL had a net worth of $92.7 billion and the least wealthy family wealth was $1.94 billion.

[26] We use the tax data for Assessment Year (AY) 2013-14 to AY 2018-19. This period covers the two GEs studied by us. Statistics for after this period have not been released as of January 2022.

$$TAXED - IN - HAND\ INCOME\ =\ TAXABLE\ INCOME\ -\ DEDUCTIONS$$

We use the CBDT statistics to estimate the relationship between the taxed-in-hand income and the reported taxable income for various income groups. This estimated relationship, in turn, is used to compute the reported taxable income for HHs and individuals covered by our study.

In addition, we use the CBDT statistics to estimate the top levels of the taxed-in-hand income and the top levels of taxable income reported in the ITRs, as described in the next section.

Prowess and Annual Accounts of Companies: The CBDT data does not offer any information regarding the income reported under the head "exempt income". This category includes agricultural income and dividends, among several other types of incomes. To estimate the dividend and equity income, we use details of the wealth portfolio available in GE and the dividend yield rates using "Prowess", a database of the financial performance of over 40,000 Indian companies.[27] For the top families on the FL, the equity income is computed directly from the annual accounts of their group companies. We use a similar approach to estimate other types of exempted income. Section 4 and Appendix-II contain further details on this approach.

Before concluding this section, we need to address the following important question.

3.2 Is GE data representative?

A study of the GE data is of interest in its own right. Moreover, as is discussed below, several attributes of this dataset enable us to make an educated guess about the nature of the income-wealth relationship for the general population.

The coverage of the GE data is extensive. By virtue of its structure, the data is reasonably representative of the Indian context in terms of the regional and rural-urban distribution of the population. It also includes all the leading social categories. Reservation of Lok Sabha seats for the members of Scheduled Castes (SCs) and Scheduled Tribes (STs) means that these disadvantaged sections of the society are proportionately represented in the sample. Moreover, the dataset covers a wide range of educational levels (candidates range from being illiterate to holding PhD degrees) and professions that range from landless labourers, farmers, and artisans to landlords in rural areas; from wage earners and self-employed businesspersons from urban centres to professionals, CEOs, and promoters of big companies.

The data also exhibits several known properties of wealth and income distributions in India, such as the concentration of wealth and income in the hands of the male members of households.[28]

The average wealth holding in the used GE sample is higher than the national average but we do not use the data to estimate wealth and income distributions. Instead, we use it to examine the relationship between the reported wealth and income levels. From this viewpoint, in addition to being the only simultaneous source of information on wealth and income, the GE data passes the 'smell' test on several counts.

For instance, shares of the financial assets and commercial property increase with the wealth level. See Figure 3.3. For the top 0.1% in the GE data, the share of equity assets in the total wealth is 83%, which is comparable to the corresponding share for families on the FL.[29] In other words, the asset holdings of the wealthiest in the GE data resemble what can be seen from alternative data sources about the most affluent non-politician Indians. The asset holding trends exhibited in the GE data are

[27] The list includes all listed companies and most unlisted public companies, and private companies of all ownership groups.
[28] In 72% of HHs, the male share of wealth is greater than that of the female; only in 28% HHs do women have a larger share of wealth.
[29] Top income levels reported in GE (up to ₹194 crores). See Section 3.

also in line with patterns observed in many international studies on the composition of wealth at the top.

As we will see in Section 4, the income-wealth ratio emanating from the GE data is very high for low-wealth groups but takes relatively small values for the wealthiest groups. Moreover, the income-wealth ratios observed in the GE data are decreasing in wealth. (See sections 4 and 5 below). These findings are consistent with the whatever evidence is available on this issue in the context of other countries.[30] Moreover, the decreasing trend observed in the GE data squares well with what can be inferred from other independent data sources such as the FL and ITR data put together.

If the decreasing trend in the income-wealth ratios emanating from the GE data hold for the entire population, we would expect the ratios to be the lowest for families on the FL. In particular, the income-wealth ratio for these families should be smaller than the wealthiest in the GE data, as the wealth owned by the former groups is relatively large. This is precisely what we find. The average ratios for FL families are significantly smaller than that for the wealthiest group in the GE data. In keeping with these decreasing trends, the income-wealth ratios for the top 10 families on the FL are smaller than the rest of the list.

Table 3.3: Income rank of the wealthiest HHs and individuals in GE data

Income Rank	Top 100 Wealthiest Households			Top 100 Wealthiest Individuals		
	2014	2019	Overall	2014	2019	Overall
Top 100	47%	48%	34%	42%	42%	35%
101 – 200	18%	17%	22%	18%	17%	19%
201 – 300	8%	7%	10%	8%	8%	6%
Greater than 300	27%	28%	34%	32%	33%	40%

The income-wealth relationship in the GE data is also consistent with the evidence available from media reports. Periodic media reports suggest that the wealthiest Indians are not among the top income tax-paying individuals. As discussed earlier, the list of top income tax filers is dominated by movie stars, cricketers, etc., while the most affluent members of the FL do not figure on the list. This anomaly between the top wealth and the top reported income levels is also evident in the GE data. Table 3.3 shows the income ranks of the wealthiest 100 HHs and individuals in the GE data. Out of the 100 wealthiest individuals in the GE data, only 35% have reported income levels belonging to the top 100 income levels in the dataset. This partly explains why the top 10% of HH in GE data account for 80% of the total wealth, but only 66% of total taxed income. The wealthiest individuals and families are not the same as those who report the highest incomes.

Notwithstanding these properties of the GE data, there can be concerns about its representativeness for the Indian society. Many people consider the election contestants a breed different from the rest of society. Thus, technically, the concern is that the income-wealth relationships emerging from the GE data might not hold in general.

Several studies argue that the politicians are subjected to greater scrutiny and thus have stronger incentives to report their finances more truthfully than the general public.[31] In India, income reported by the non-politician is a private information between them and the tax department. In contrast, income declared by politicians in their affidavits can easily be accessed by the media and the other third parties. Still, there is a perception that politicians report a relatively small share of their total income and wealth. In any case, it is possible that the reported income-wealth is different

[30] For a discussion, see Piketty (2018, chapter 2). To our knowledge, the overall patterns of the income-wealth ratio for individuals and household levels have not been examined comprehensively.
[31] For a review of this literature see Libman, Schultz and Graeber (2016) and Szakonyi (2022).

for politicians just because they have political abilities.

One way to account for this possibility is to use a measure of the political ability of candidates and check if it has a bearing on the income-wealth ratio. To this end, we use *vote share* as a proxy for political ability of a candidate. Controlling for wealth and other correlates, we find that the vote share has a statistically significant bearing on the income-wealth ratio.

4. The Estimation Methodology

In this section, we summarize the methodology used to estimate the various types of income examined in this study. First, we describe the leading forms of income reported by taxpayers under the Indian tax law.

4.1 Reported Taxable Income

All Indians with taxable income must report their earnings to the tax authorities under two leading categories: the income taxable in the hands of the recipient (i.e., the filer of the returns), and the income legally treated as tax-exempt in the hands of the receiver. For concreteness, let:

Y_T denote the income taxable in the hands of the recipient,
Y_{Ex} denote the income treated as tax-exempt in the hands of the recipient, and
Y_R denote the total reported income. By definition, $Y_R = Y_{Ex} + Y_{T.}$

The taxable income, Y_T , is the sum of all types of income a taxpayer reports as taxable in their hands, i.e., the income on which the recipient themselves are liable to pay tax. It includes salary and all other forms of labour income. Besides, it consists of some forms of capital income - interest income, rentals, and capital gains. It also has 'business income' defined as the profits and gains of a business or profession. Simply put, business income includes profits for non-corporate entities, including mixed income of self-owned enterprises and income of the professionals such as lawyers and doctors from private practices.

In effect, a part of Y_T becomes tax-free due to the various tax deductions and exemptions available to taxpayers. We have defined the taxed-in-hand income, Y_{Td}, as that part of the Y_T on which taxpayers themselves pay tax. In other words, for a tax unit (an individual or a household), Y_{Td} is equal to Y_T *minus* the tax deductions availed by the unit.[32] The Y_T and Y_{Td} do not cover several types of capital income reported under the head called the (tax) "exempt income," such as agriculture income, and profits from firms and partnerships, for which the recipient is not liable to pay tax. Therefore, the exempt income is a class separate from Y_T and hence Y_{Td}.

Simply put, the total income reported by a taxpayer can be defined as the *sum* of the income reported as taxable, and the income reported as tax-exempt. By definition, the taxable income, Y_T, is only a part of the total income reported by the taxpayers in their ITRs, i.e., Y_R. The exact relationship among the different types of income in ITRs can be expressed as: $Y_R = Y_{Ex} + Y_{T.} = Y_{Ex} + Y_{Td} + Y_{Dd}$. Therefore, $Y_R > Y_T > Y_{Td}$.

Our estimation methodology follows the following order: We start with using the direct and precise information we have for Y_{Td} and use it to estimate Y_T. We obtain exact information on Y_{Td} for individuals and HHs in the GE dataset directly from the affidavits. To estimate Y_{Td} for the FL individuals, first we use the Generalised Pareto Interpolations (GPIs) to estimate the right tail of the

[32] In the terminology of the income tax returns (ITR) forms used by the Indian Tax Department, Y_T is called the "gross total income" (GTI). Y_{Td} is called the "total income" (TI), and is commonly referred to as the "returned income" by professionals such as accountants.

distribution of taxed-in-hand incomes reported to the Tax Department.[33] The GPIs are then used to precisely isolate the group averages for the top income levels: the top 10, the next 11-20, the top 100, the next 101–200, 201–300, and 301–400, and so on.

To estimate Y_T for individuals and HHs in the GE data, we use the statistical relation between Y_T and Y_{Td} derived from the statistics published by the Income Tax Department. To estimate Y_T for FL individuals, we use the GPIs derived from the tax data.

4.2 Tax-exempt Capital Income

While all forms of labour income must be reported as a part of Y_T, several types of capital income are tax-exempt. The reporting rules for capital income depend on the type and size of the income. So, as a first step, here we summarize the methods used for estimating various forms of the direct capital income Y_{KD} from different types of assets. By definition:[34]

$$Y_{KD} = Y_D + Y_{Eq} + Y_{Cg} + Y_A + Y_P + Y_H,$$

where
Y_D is the interest earned from assets under categories 'Deposits' and 'Receivable' in Table 3.1;
Y_{Eq} denotes the equity income;
Y_{Cg} represents the capital gain;
Y_A is the income from agricultural land;
Y_P is the rental income from commercial properties;
Y_H is the imputed rent on residential properties used for housing.

Each of these forms of capital income, except Y_H, must be reported under Y_T or under category Y_{Ex}. The imputed rent, Y_H, is not required to be reported under the tax rules.

We estimate various forms of direct capital income using the value of the underlying assets and their rates of return. For instance, the dividend income of a household is estimated as the value of stocks owned multiplied by the average dividend yield rate for the top 100 private listed companies; rental income from commercial properties is calculated as the value of the property *times* the average rents (as a proportion of the property value). Similarly, we estimate the gross value of the various other types of capital income - profit shares, agricultural income, rentals from non-agricultural land and buildings, capital gains, and the imputed rent. Given variations in the rate of returns, we estimate capital income for each class of assets, for a range of returns admissible under the available evidence; for details, see the online Appendix II.

Most Indians, except the super and ultra-wealthy groups, hold their assets directly in their own names. So, estimating the tax-exempt capital income by the bottom 95% HHs and individuals in the GE data is easy. As to the interest and rental incomes, Y_D and Y_P, these incomes are taxable in the hands of the recipient and, therefore, already included in Y_T. Of the other forms of capital income, agricultural income is the dominant form of the capital income reported as exempt by the bottom 95% of HHs and individuals. Their entire equity income (from partnership firms and companies) also qualifies as tax-exempt. The reason for this is that all profits from firms and partnerships are tax-exempt in the hands of the recipient. Besides, their dividend income falls well below the ₹10 lakhs threshold for the taxability of dividends.

[33] Compared to the alternatives available, the GPIs are better suited for estimating the right tail of income distribution. See Blanchet, Fournier, and Piketty (2017).
[34] As can be seen from the definition, it is (realistically) assumed the jewelry, and personally owned durable assets like cars do not yield any direct income or imputed gains.

Specifically, for the bottom 0–95 percentiles of HHs and individuals, $Y_{Ex} = Y_{Eq} + Y_A$. In the absence of any direct source of information, profit rates from the equity in firms and partnerships are taken to be the same as the dividend yield rates. Simply put, for all units in the GE data, the entire equity income, Y_{Eq}, is estimated as the value of equity multiplied by the average dividend yield rates for the top 100 private listed companies. In view of the evidence presented in Section 2 on returns from various assets, we take the agricultural income to be 0.08–4% of the land value. Accordingly, we estimate farm income corresponding to the three leading rates: 2% being the most plausible case, 4% the absolute upper bound, and 0.08% the absolute lower bound on agriculture income.

As far as the wealthiest groups (the top 5% of HHs and individuals in the GE data and those on the FL) are concerned, their interest income, Y_D, is part of their reported taxable income, just like the other wealth groups. However, it is challenging to estimate the Y_{Eq}, Y_A, Y_P, and Y_{Cg} reported by the wealthy groups. As elaborated upon in Appendix-II, these groups receive a significant share of their capital income in the name of financial intermediaries such as limited liability partnerships (LLPs), the association of persons (AOPs), and the body of individuals (BOIs). To the extent that the capital income received in the intermediaries' accounts is distributed to the partners as partnership shares, it must be reported in the ITRs under the category of exempt income. However, part of the income is retained in the account of the intermediaries and is not required to be reported in the ITRs of the partners.

The point is that there is uncertainty about the proportions of Y_{Eq}, Y_A, Y_P, and Y_{Cg} received by the wealthy groups in their own name vis-à-vis the amount received in the account of financial intermediaries used by them. So, we consider a range of possibilities regarding reporting of the direct capital income by these wealthy groups. This range includes scenarios where most of their Y_{Eq}, Y_A, Y_P, and Y_{Cg} get reported in the ITRs. It also includes the case where most of these capital incomes remain unreported in the tax returns.

Going by the evidence discussed in the Appendix, the most plausible assumption about the reporting of capital income by the wealthiest groups (the top 5% of HHs and individuals in the GE data, and the FL) can be summarised as follows:

Most Plausible Scenario takes that only half of the direct capital income from equity, farmland, rented properties and capital gains are reported in the ITRs. That is, half of $[Y_{Eq} + Y_{Cg} + Y_A + Y_P]$ (the sum of direct capital income received in individual accounts or in the accounts of financial intermediaries) is reported. Out of this, one-fourth of the total $[Y_{Eq} + Y_A + Y_P + Y_{Cg}]$ is reported as tax-exempt capital income, and one-fourth under Y_T.[35] Therefore, $Y_{Ex} = 0.25\left[Y_{Eq} + Y_A + Y_P + Y_{Cg}\right]$.

Though this estimate of the wealthy groups' total reported income appears to be the most plausible, given the uncertainty over the direct and indirect ownership of assets by wealthy groups and for the sake of completeness, we also estimate the absolute upper and the absolute lower bounds on the capital income reported as tax-exempt by the wealthiest 5% of units in GE data and the FL.

Absolute Lower-bound estimates are derived from the following assumption: A 5% of the sum $[Y_{Eq} + Y_A + Y_P + Y_{Cg}]$ is reported as exempt income, and an equal amount gets reported as a part of Y_T. This means that an absolute lower bound on Y_{Ex} is estimated as: $Y_{Ex} = 0.05\left[Y_{Eq} + Y_A + Y_P + Y_{Cg}\right]$.

[35] Under this assumption, the unreported half of the amount $[Y_{Eq} + Y_{Cg} + Y_A + Y_P]$ gets retained in the accounts of the intermediaries.

Absolute Upper-bound estimates are derived from the following assumption: The wealthiest groups report 95% of the sum $[Y_{Eq} + Y_A + Y_P + Y_{Cg}]$ as exempt income; and, the remaining 5% as a part of Y_T. That is, $Y_{Ex} = 0.95[Y_{Eq} + Y_A + Y_P + Y_{Cg}]$. Simply put, under the absolute upper bound, Y_{Ex} is equal to 95% of their gross capital income, (whether received in individual accounts or through intermediaries).

4.3 The Total Reported Income

The total reported income is the sum of taxable income reported to tax authorities *plus* the income declared as exempt income, i.e., $Y_R = Y_T + Y_{Ex}$.

Estimating the total reported income by the bottom 95% HHs and individuals in the GE data is straightforward. From Section 4.1, we already have a precise estimate of the taxable income, Y_T, reported by these groups. To compute their Y_R, we just need to estimate their Y_{Ex}. As is explained in the previous subsection, for the bottom 0–95 percentile of HHs and individuals, $Y_{Ex} = Y_{Eq} + Y_A$. Therefore, their total reported income is $Y_R = Y_T + Y_{Eq} + Y_A$.

The top 5% of HHs and individuals in the GE data and those on the FL receive a major share of their capital income in the name of financial intermediaries such as LLPs, AOPs, and BOIs. Given the uncertainty about the share of capital income reported by these groups in their ITRs, we estimate their total reported income for each of the above-discussed three leading scenarios for Y_{Ex}.

Recall that our most plausible scenario assumes that the wealthiest 5% of HHs and individuals in the GE data and the FL, we have $Y_{Ex} = 0.25[Y_{Eq} + Y_A + Y_P + Y_{Cg}]$. Accordingly, our most plausible estimate of the total reported income is:

$$Y_T + 0.25[Y_{Eq} + Y_A + Y_P + Y_{Cg}].$$

Absolute Lower-bound estimates $Y_{Ex} = 0.05[Y_{Eq} + Y_A + Y_P + Y_{Cg}]$. This means that an absolute lower bound on the total reported income is estimated as:

$$Y_T + 0.05[Y_{Eq} + Y_A + Y_P + Y_{Cg}].$$

Absolute Upper-bound estimates $Y_{Ex} = 0.95[Y_{Eq} + Y_A + Y_P + Y_{Cg}]$. Therefore, the upper bound on the total income reported is estimated as:

$$Y_T + 0.95[Y_{Eq} + Y_A + Y_P + Y_{Cg}.]$$

Simply put, under the absolute upper bound, the total reported income of the wealthiest groups is estimated as the sum of their taxable income *plus* 95% of their gross capital income (whether received in individual accounts or through intermediaries).

Our absolute upper bound is very likely an overestimation of the reported exempt income and hence the total income reported by the wealthy groups on at least two counts: First, the capital income itself is overestimated due to the unrealistically high assumption about the rental income from all types of land and commercial properties (assumed to be 4% of the property value). Second, the assumed share of the capital income reported as exempt income (95% of the total) is much above what is supported by the available evidence and common sense. The upper bound scenario assumes that the wealthiest group transfers almost all the income received in intermediaries' accounts to their individual accounts. If they did so, it would defeat the very purpose of using financial intermediaries. Besides, it would increase the tax obligation compared to a situation where the capital income is directly received in individual accounts.

4.4 Direct personal income

The direct personal income, Y_{PID}, as defined in Section 2, is the sum of the labour income and the direct capital income, including the imputed rent on self-occupied dwellings. There is a direct relationship between Y_{PID} and Y_R. The latter includes the entire labour and the direct capital income except the imputed rent on self-occupied dwellings. Formally, $Y_{PID} = Y_R + Y_H$, where Y_H is the imputed rent from buildings used for self-housing.[36]

Summing up, $Y_{PID} > Y_R > Y_T > Y_{Td}$. Moreover, we have the following relationship between the Y_{PID} and the other types of income examined by us:

$$Y_{PID} = Y_R + Y_H = Y_T + Y_{Ex} + Y_H = Y_{Td} + Y_{Dd} + Y_{Ex} + Y_H.$$

To estimate the direct personal income, we add the value of the imputed rent to our estimates of Y_R. The imputed rent is calculated as rental inequivalent for the residential properties.

4.5 Capital and Labour Incomes

We can use our estimates of Y_T and Y_{KD}, to estimate capital and labour income values reported in tax returns. In terms of notations, let

Y_{KD}^* denote the total capital income reported in the ITRs.

Specifically, Y_{KD}^* is part of the direct capital income, Y_{KD}, that gets reported in the income tax returns either as taxable or as tax-exempt income. One component of Y_{KD} is not reported. This is the imputed rent on residential properties, Y_H. As explained in the previous subsections, all other forms of direct capital income of the bottom 95% of HHs and individuals must be reported in the ITRs –as taxable or tax-exempt income. This means that for them, $Y_{KD}^* = Y_{KD} - Y_H$.

Given the uncertainty about the capital income reported by the wealthiest 5% units in the GE data and the FL groups, we estimate a range of possible values of the direct capital income reported by these groups. Corresponding to the most plausible scenario discussed in the previous subsection, we estimate the reported direct capital income as:

$$Y_{KD}^* = Y_D + 0.5\left[Y_{Eq} + Y_{Cg} + Y_A + Y_P\right].$$

Corresponding to the absolute lowest bond scenario, we estimate Y_{KD}^* as:

$$Y_{KD}^* = Y_D + 0.1\left[Y_{Eq} + Y_{Cg} + Y_A + Y_P\right].$$

The absolute upper bound on Y_{KD}^* is estimated as:

$$Y_{KD}^* = Y_D + Y_{Eq} + Y_{Cg} + Y_A + Y_P.$$

Notably, under the absolute upper bound, direct capital income reported by the wealthiest 5% units in the GE data and the FL groups is estimated as the sum of the gross values of their direct capital income in all its forms, regardless of whether the income is received directly in individual accounts or through intermediaries.

We estimate the labour income, Y_L, as a residual category. As discussed above, all forms of labour income are included in Y_T. Besides, interest income, capital gains, and rental income are the only form of capital income included in Y_T for the bottom 95% of GE units. Therefore, their labour income is estimated as:

$$Y_L = Y_T - Y_D - Y_{Cg} - Y_P.$$

For the wealthiest 5% units in the GE data and the FL units, we estimate a range of their labour income corresponding to the most plausible scenario, the absolute lower bound and the absolute upper bound

[36] Y_R also includes realised capital gains, which are a negligible share of the total reported income and hence of Y_{PID}.

scenarios discussed above.

Recall that the entire interest income is reported under Y_T. Also, under the most plausible scenario, one-fourth of the capital income, $Y_{Eq} + Y_{Cg} + Y_A + Y_P$, is reported under Y_T. Therefore, Y_L is estimated as:

$$Y_L = Y_T - Y_D - 0.25[Y_{Eq} + Y_{Cg} + Y_A + Y_P].$$

Guided by similar reasoning, we compute the absolute lower bound on Y_L of the wealthiest 5% units in the GE data and the FL units as:

$$Y_L = Y_T - Y_D - 0.95[Y_{Eq} + Y_{Cg} + Y_A + Y_P],$$

The absolute upper bound on Y_L of the wealthy groups is taken as:

$$Y_L = Y_T - Y_D - 0.05[Y_{Eq} + Y_{Cg} + Y_A + Y_P].$$

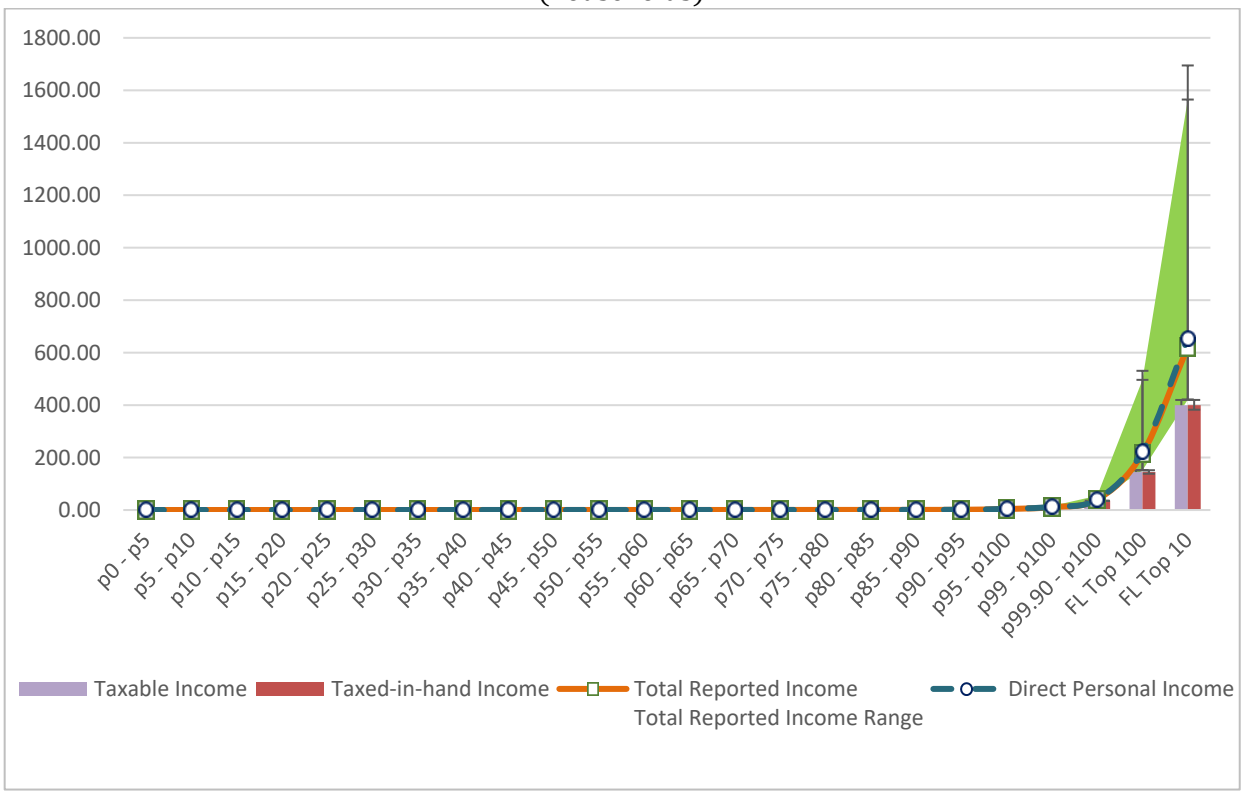
5. The Income-Wealth Ratios

In this section, we present our empirical findings on the relationship among the various types of income discussed and wealth. First, we present some plots showing the relationship among the leading types income reported by taxpayers, i.e., Y_{Td}, Y_T and Y_R, along with Y_{PID}.

5.1 The Cobra Curves

To start with, we present relationship among Y_{Td}, Y_T and Y_R reported by different wealth groups along with their Y_{PID}. In view of our focus on the reported income, plots below show the range-estimates of the Y_R. For other versions of income, we depict only the point estimates. See Figure 5.1.

Figure 5.1: Average Y_{Td}, Y_T, Y_R and Y_{PID} income reported by different wealth groups (households)

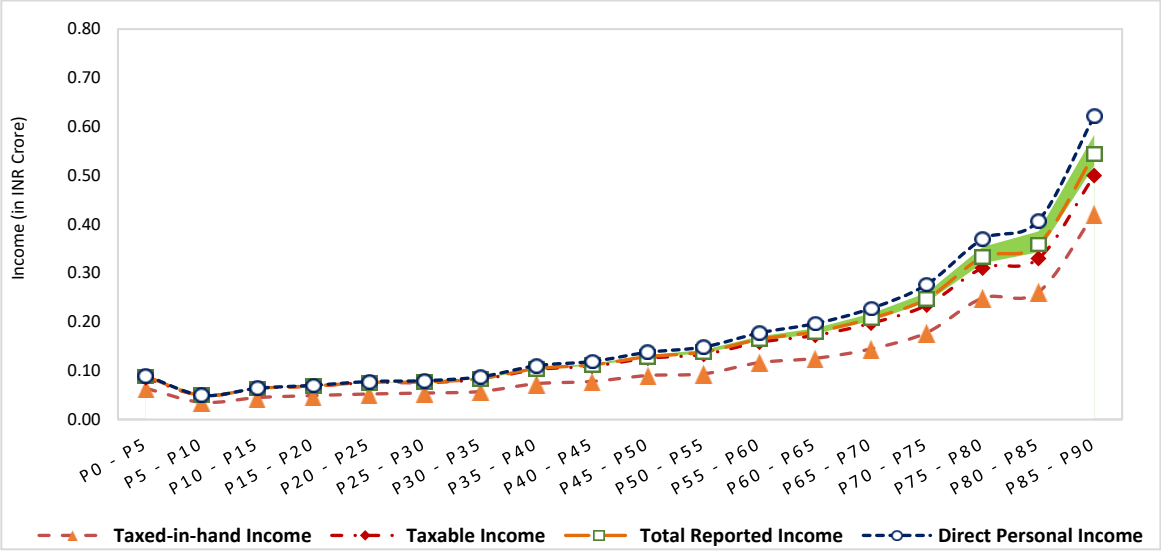


The income plots appear to be flat for the bottom 99% due to a relatively massive increase in the income levels for the top wealth groups — the top 1% in the GE data and the FL. For the ease of illustrating increasing trends in income at all wealth levels, below, we present separate plots for the

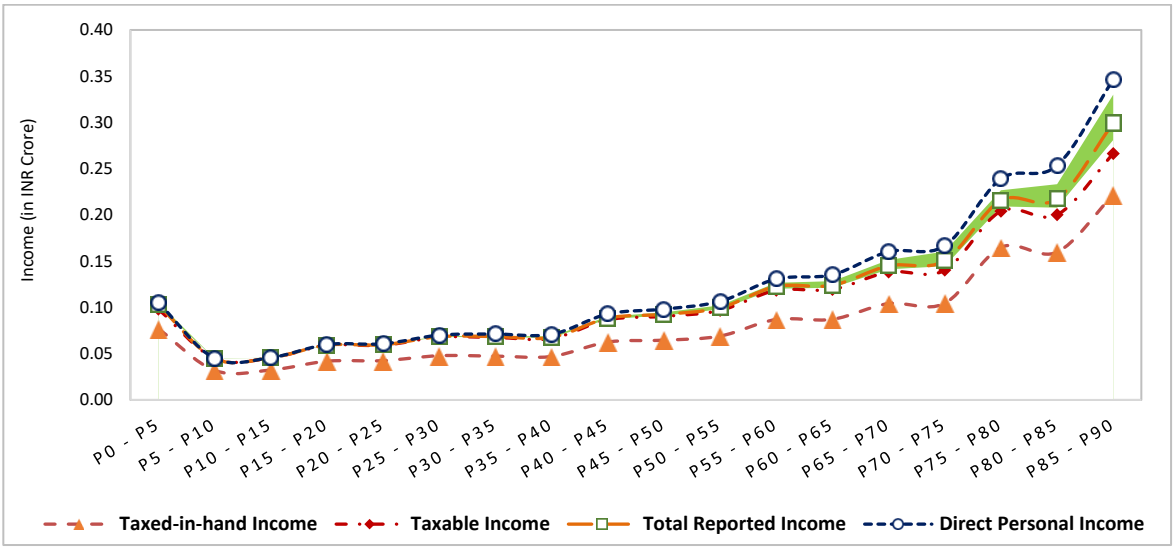
$p_5 - p_{10}$ to $p_{85} - p_{90}$ of household and candidates. The average income increases with wealth. See plots in Figure 5.2. Income wealth relationship for the wealthiest members shows similar patterns.

Figure 5.2: Average Income Y_{Td}, Y_T, Y_R and Y_{PID} reported across wealth groups

Plot A: Household



Plot B: Candidate



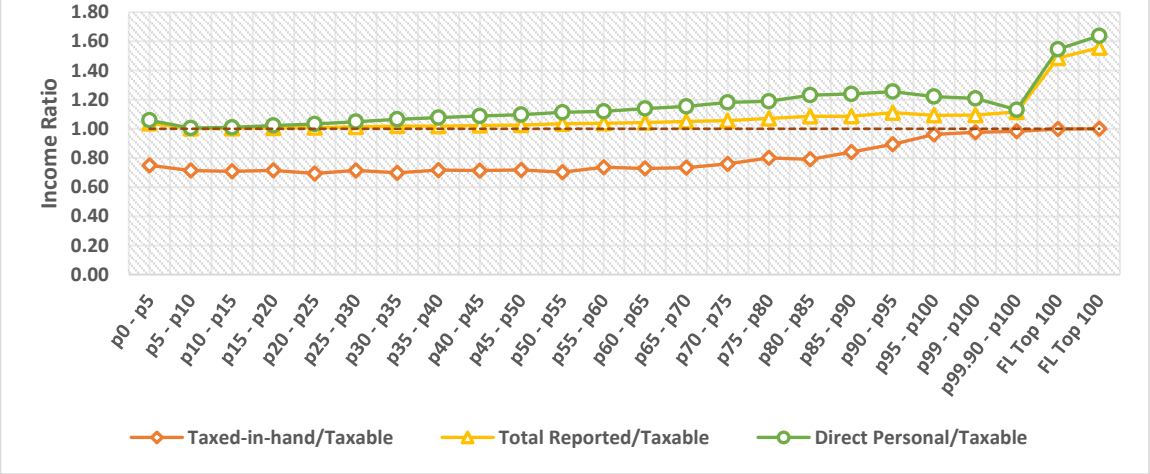
Plots in Figure 5.3 show values of Y_{Td}, Y_T and Y_{PID} for individuals and HHs from various wealth groups, relative to their Y_R. As explained earlier, at low- and medium-income levels, the deductions and exemptions are a significant proportion of the Y_T. In contrast, they are a small fraction of the taxable income for the richest taxpayers. Since all types on income increase with wealth, for the low and middle-wealth groups Y_{Td} is significantly smaller than their Y_T. The difference between the Y_T and Y_{Td} decreases with wealth. For the wealthy, the two are approximately equal.

However, the relative difference between Y_T and Y_R increases with wealth. It is instructive to note that for the wealthy and super-wealthy groups, our estimates of Y_R reported by them are significantly higher than their Y_T. In other words, a significant part of the income reported by the wealthy groups falls under the category of tax-exempt income. For the top 1% HH in the GE data,

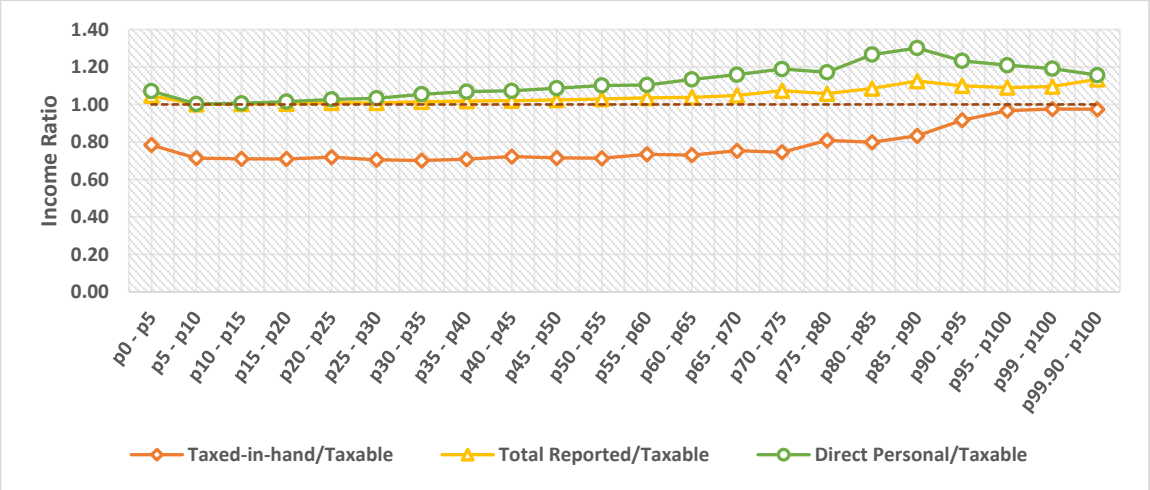
the estimated Y_R is about 112% of their Y_T. For the FL, Y_R is nearly 150% of the Y_T. The relationship between Y_T and Y_R is also very similar for individuals.

Figure 5.3: Ratio of Y_{Td}/Y_T, Y_R/Y_T and Y_{PID}/Y_T reported by different wealth groups

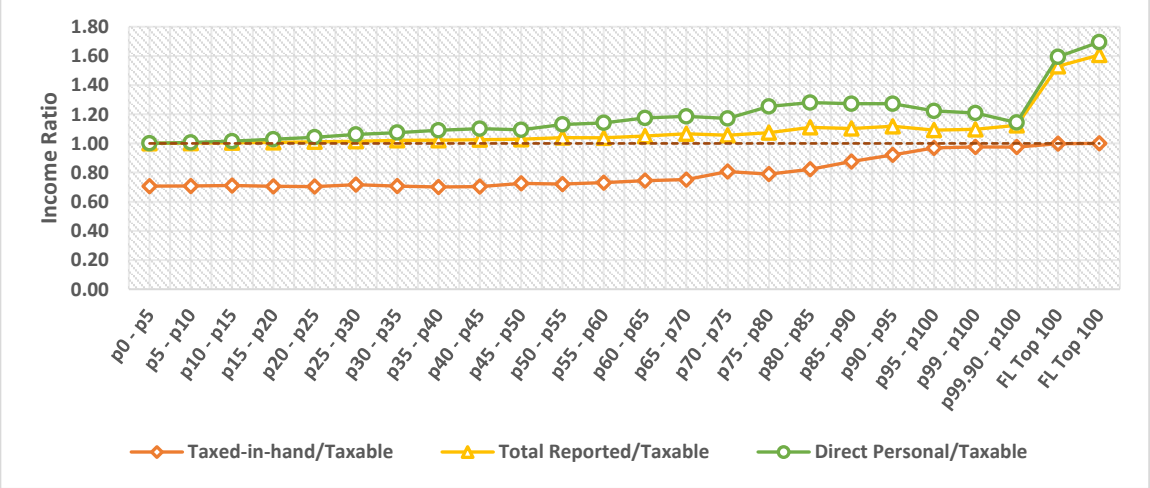
Plot A: Household



Plot B: Candidate



Plot C: Wealthiest member



By contrast, the relationships between Y_{PID} and Y_R do not follow consistent patterns. As wealth increases, initially, Y_{PID} increases relative to Y_R. It reaches its peak for the wealth groups at p90-p95; thereafter, it decreases continuously. The initial increase in Y_{PID} vis-à-vis Y_R is because of the dominance of residential property in the asset portfolio of the middle-wealth groups. At very high wealth levels, the share of residential property is comparatively tiny, and hence Y_{PID} is only marginally higher than Y_R. A similar relationship holds between Y_{PID} and Y_R for individuals.

Summing up, for all categories of income, the average income reported by the wealthiest Indians is significantly higher than that reported by other groups. However, the trends turn out to be very different when we compare the reported income with the corresponding wealth.

From Section 2 we expect the ratio of personal income to wealth to be decreasing in the latter. The ratio of the reported income to wealth is a matter of empirical investigation. The same is the case with the ratio of wealth and the income reported as taxable. Below we show that the income-wealth ratios decrease with family wealth.

To give a sense of the magnitude of reported income relative to wealth, first we present the income-wealth ratios for various groups. To this end, we employ two approaches. Under Approach 1, we first compute the income-wealth ratio at the unit level, i.e., for each HH and individual separately. For example, assume there are two individuals in a wealth group. Let the income and wealth reported by the first individual be 100 and 150 respectively. Let the corresponding figures reported by the second individual be 200 and 250. Under Approach 1, the average of individual income-wealth ratios is computed in the group. Accordingly, in this example of two individuals, the ratio is computed as: $1/2\left(\frac{100}{150}+\frac{200}{250}=\frac{2}{3}+\frac{4}{5}\right) = 11/15 \approx 0.733$.

Under Approach 2, on the other hand, the income-wealth ratio is computed for different wealth groups, say for the wealthiest 1%, the bottom 5%, etc. Under this approach, the ratio is computed as the total income of all units in a given group divided by their total wealth. In the context of the above example, the income wealth ratio will be $\frac{100+200}{150+250}=\frac{300}{400}=0.75$. It is easy to see that with a large enough set of individuals, the two approaches are expected to lead to very similar results.

In the rest of this section, we show that the income-wealth ratios are decreasing in wealth, regardless of the method used to compute the ratio. This decreasing trend persists for all types of income discussed above, i.e., $Y_{Td}, Y_T, Y_R, Y_{PID}, Y_{KD}$ and Y_L. The ratios reported by the super and ultra-wealthy groups are tiny, even if we use the most generous estimates of the income reported by them.

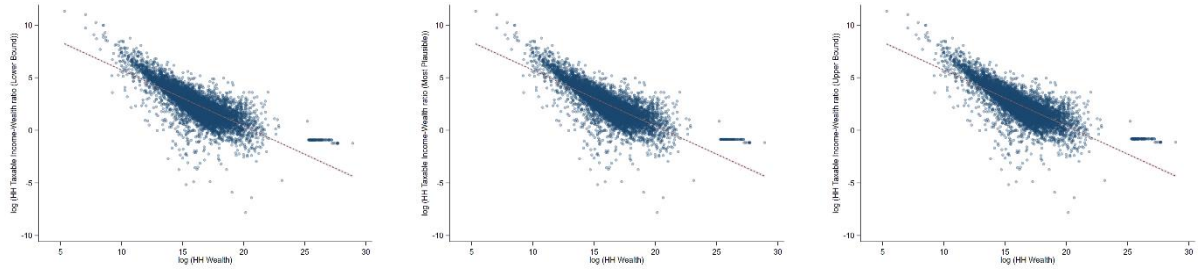
5.2 The Taxable Income–Wealth Ratio (Y_T/W)

Figure 5.4 shows the log-log scatter plots of Y_T/W ratios reported by households versus their wealth. Plots A, B and C, respectively, correspond to what we consider as the most plausible, the lowest bound and the upper bound on the total taxable income reported by the FL families. Plots for individuals are very similar and omitted from presentation.

Group levels income-wealth ratios are presented in Table 5.1. While reading the tables, it will help to remember that the wealth percentiles have been calculated separately for the households, candidates, and wealthiest members. Furthermore, we have dropped individuals with zero income or wealth from the analysis. This has resulted in the number of individuals (candidates and wealthiest members) to be smaller than the number of HHs. For the FL, the candidates' income-wealth ratios are not relevant.

Figure 5.4: Household income-wealth ratio ($\frac{Y_T}{W}$) vs wealth scatter plots

(a) Lower Bound (b) Most Plausible (c) Upper Bound



Note: Linear prediction curve is fitted into the scatter plot with 95% confidence interval.

Table 5.1: Reported taxable income–wealth ratio across wealth groups, $\frac{Y_T}{W} * 100$

Wealth Percentiles	*Approach 1*		*Approach 2*	
	Households	Individuals	Households	Individuals
p0 - p5	1038.1	1167.6	-21.7	-17.8
p5 - p10	187.5	509.7	169.4	419.5
p10 - p15	91.7	144.4	89.3	134.6
p15 - p20	52.6	86.8	52.0	85.5
p20 - p25	36.9	51.0	36.5	49.9
p25 - p30	25.3	38.2	25.3	37.8
p30 - p35	20.5	26.6	20.5	26.5
p35 - p40	19.5	19.7	19.4	19.7
p40 - p45	16.2	20.0	16.2	19.7
p45 - p50	14.5	16.1	14.4	16.0
p50 - p55	12.2	13.5	12.2	13.5
p55 - p60	11.5	12.9	11.4	12.8
p60 - p65	9.7	10.0	9.7	10.0
p65 - p70	8.2	8.9	8.3	8.9
p70 - p75	7.4	7.3	7.3	7.4
p75 - p80	7.3	6.9	7.3	6.9
p80 - p85	5.6	5.2	5.6	5.2
p85 - p90	5.3	4.5	5.4	4.4
p90 - p95	5.0	5.5	4.8	5.3
p95 - p100	3.8	3.9	3.2	3.4
p99 - p100	3.2	3.4	2.8	3.0
p99.90 - p100	1.6	1.4	2.0	1.9
FL Top 100	0.40 – 0.44		0.40 – 0.44	
FL Top 10	0.29 – 0.32		0.29 – 0.32	

Note: (a) Units in top 1% (p99 – p100) and 0.1% (p99.90 – p100) are a subset of observation in top 5% (p95 – p100). The second approach is used for the FL; (b) In several instances, while the family wealth and income are positive, candidates has reported zero income or wealth.

Also, the negative income-wealth ratios in Approach 1 are due to the negative aggregate wealth i.e., the aggregate liabilities are greater than the assets. We have thus dropped the bottom p0–p5 units

from the plots in the interest of visual clarity. Moreover, for the FL, our plots show income-wealth ratios corresponding to the leading scenario — the range of the estimated ratios can be seen in the relevant tables.

The Y_T/W ratios in columns 2 and 3 of Table 5.1 are based on the first approach for computing the income-wealth ratio; ratios in columns 4 and 5 are based on Approach 2. As can be seen from the plots in Figure 5.5, the income-wealth ratios for the wealthiest members (one member from each HH) are very similar to the ones for HHs and candidates. To avoid the clutter, from the tables we have dropped the ratios for the wealthiest members.

The two approaches for computing Y_T/W ratios produce very similar results, except for the bottom 5% of HHs and individuals. Income heterogeneity is relatively much higher for low-wealth groups. This is because the HH wealth levels in these groups remain low, but the family income can vary significantly. When the HH wealth is negligible, the income-wealth ratio can jump violently depending on the income. This explains the very high ratio generated by the second method. Otherwise, under both methods the income-wealth ratio decreases consistently with wealth. On average, the taxable income reported by a HH as a proportion of its wealth falls continuously with wealth. In other words, the wealthier a HH, the relatively small the taxable income it reports.

The magnitudes of the ratios produced by the two approaches are also comparable. The taxable income reported by low-wealth households in the bottom 5–10 percentiles is more than 170%, i.e., more than 1.7 times their wealth. In contrast, for the top 5% HHs, the ratio reduces to merely 3.2%, i.e., the reported taxable income amounts to only 3.2% of their wealth. The ratio drops to less than 2% for the wealthiest 10% of the top percentile of the HHs in the GE data.

In line with these overall trends, the FL families have reported the most diminutive taxable income relative to their wealth. For the wealthiest 100 families on the FL, the estimated ratio is in the range of [0.4, 0.6] %, i.e., the reported taxable income is at most 0.6% of their wealth.[37] For the wealthiest 10 Indian families, the reported taxable income is at most 0.4% of their wealth.

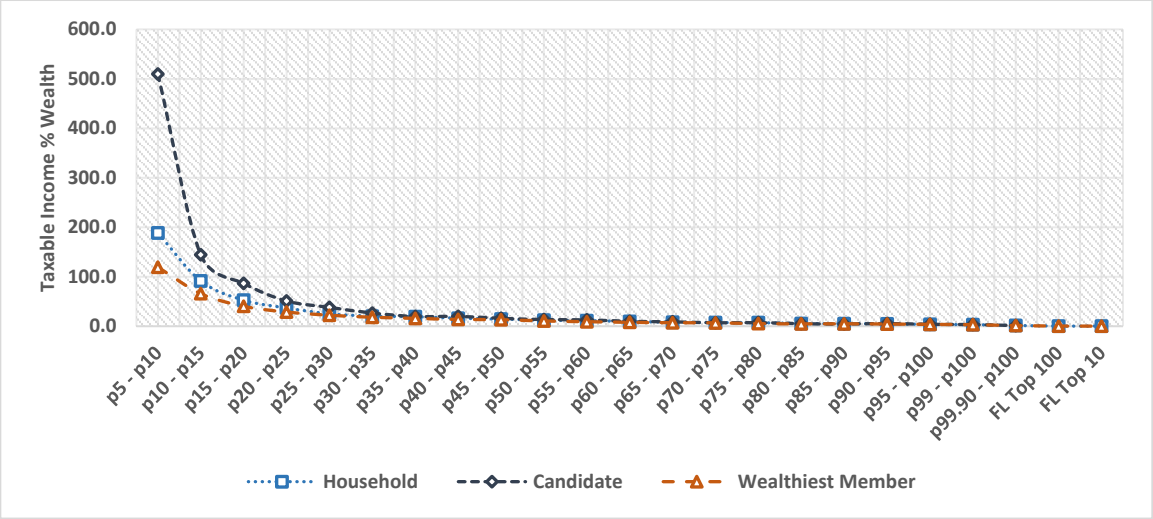
The income-wealth ratios for individuals (wealthiest members of households and the candidates) follow very similar patterns. On average, the more affluent the individual is, the smaller the reported value of their taxable income tends to be. See Figure 5.5.

Given the similarity of the ratios emerging from the two approaches, to optimise space, the rest of tables in this section are generated using Approach 1. To provide a sense of the numbers produced by the two approaches, all plots presented hereafter are based on Approach 2. Additional details can be found in the Appendix-I.

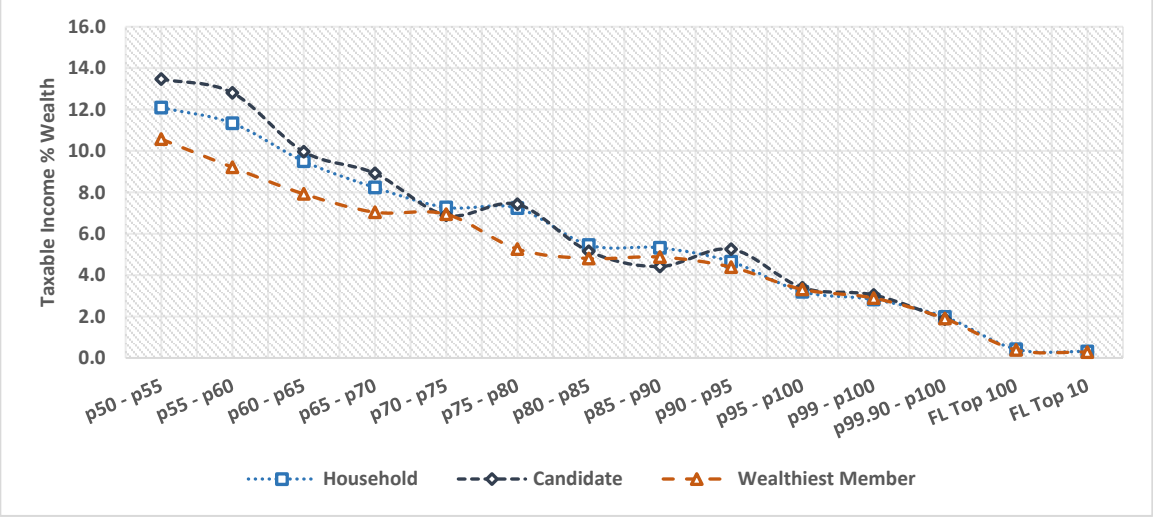
[37] As explained in Section 3, for the FL families, we have estimated a range of the income-wealth ratios on account of uncertainty related to the household income, and for the FL individuals on account of uncertainty related to their share in the family wealth.

Figure 5.5: Reported taxable income as a percentage of wealth, across groups

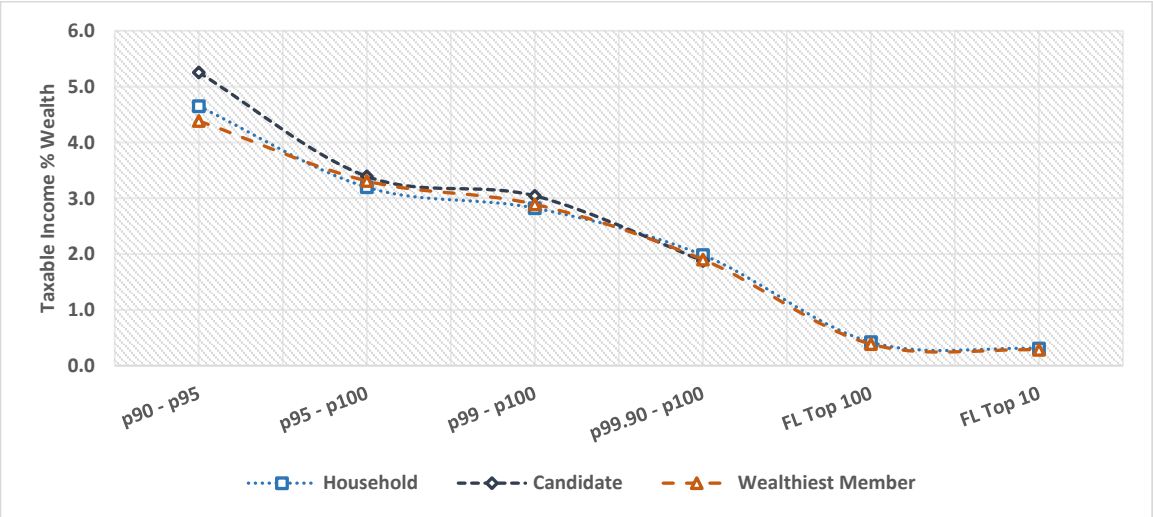
Plot A: p5 – p10 to the top 10 families on the FL



Plot B: p50 – p55 to the top 10 families on the FL



Plot C: p90 – p95 to the top 10 families on the FL

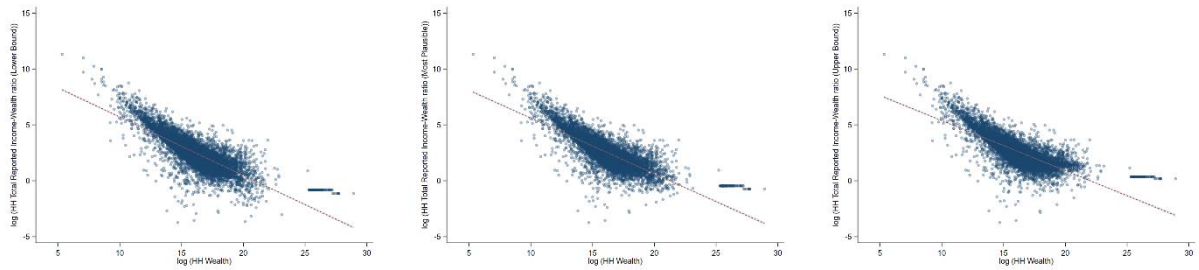


5.3 The Total Reported Income-Wealth Ratio (Y_R/W)

Now we consider the total reported income as a ratio of the reported wealth. Figure 5.6 show the scatter plots of the Y_R/W ratios reported by households versus their wealth. Plots A, B and C, respectively, correspond to what we have described above as the most plausible, the lowest bound and the upper bound on the total income reported by households. Plots for individuals are very similar.

Figure 5.6: Household total reported income-wealth ($\frac{Y_R}{W}$) vs. wealth scatter plots

(a) Lower Bound (b) Most Plausible (c) Upper Bound



Note: Linear prediction curve is fitted into the scatter plot with 95% confidence interval.

Table 5.2 presents the group level income-wealth ratios for what we have described as the most plausible estimate of the total reported income. It also shows the range generated by the lower and the upper bounds on the total reported income.

Overall, the total reported income as a ratio of the wealth decreases continuously for the HHs as well as the individuals. As shown in Table 5.2, according to our most plausible estimates, the reported income is more than 187% of the wealth for the HHs in the lowest decile. In contrast, for the top 1% of HHs in the GE data, the total reported income, including the labour income, amounts to just 3–4% of their wealth. For the top 0.1%, the ratio drops to less than 2%. For the wealthiest 10 families on the FL, the reported income adds up to just about half a percent of their wealth! The income-wealth ratio for the individuals follows a similar pattern: the wealthier an individual is, the smaller is their reported income. Even if we consider the absolute upper bound on reported income, i.e., even if we assume that the super and ultra-wealthy report most of their capital income, their income-wealth ratios turn out to be the lowest for the FL families and individuals.

5.4 The Direct Personal Income-Wealth Ratio (Y_{PID}/W)

The decreasing $\frac{Y_R}{W}$ ratios in the previous subsection give us a good sense of the total income reported by different strata relative to their wealth. However, these ratios do not serve as a basis for empirically testing Proposition 2 proved above. The reason is that while W comprises all assets, Y_R does not include the imputed rent. Therefore, to empirically examine Proposition 2, we consider $Y_{PID}(= Y_R + Y_H)$. It includes income all forms labour income and direct income from all assets, including residential properties. We expect the ratio $\frac{Y_{PID}}{W}$ to decrease with W.[38] This indeed is the case, as can be seen from Table 5.3 below. As expected, the effect of the imputed rent on the total income is noticeable only for the middle wealth groups for whom the residential property is a significant component of the total wealth. The share of residential properties is relatively small at the high wealth levels and reduces to a negligible level for those on the FL. Consequently, for the wealthy groups, the income-wealth ratios with and without imputed rent are comparable. The

[38] As explained above, Y_R, and hence Y_{PID}, includes realised capital gains but this amount is a negligible fraction of the total reported income (approximately 2.6% Y_T).

income-wealth ratios for the individuals follow a pattern similar to the HHs.

Table 5.2: Total reported income–wealth ratio ($\frac{Y_R}{W} * 100$) across wealth groups

Wealth Percentiles	Households			Individuals		
	Lower Bound	Most Plausible	Upper Bound	Lower Bound	Most Plausible	Upper Bound
p0 - p5	1037.8	1037.5	1037.0	1167.0	1166.2	1164.8
p5 - p10	187.6	187.7	187.9	509.7	509.8	510.0
p10 - p15	91.8	91.9	92.1	144.5	144.7	144.9
p15 - p20	52.7	52.9	53.1	86.9	87.1	87.4
p20 - p25	37.0	37.2	37.4	51.2	51.4	51.8
p25 - p30	25.4	25.6	25.9	38.3	38.5	38.8
p30 - p35	20.7	20.9	21.3	26.8	27.0	27.4
p35 - p40	19.7	19.9	20.2	19.9	20.1	20.4
p40 - p45	16.3	16.6	16.9	20.1	20.4	20.7
p45 - p50	14.7	14.9	15.2	16.2	16.5	16.8
p50 - p55	12.4	12.7	13.1	13.7	13.9	14.2
p55 - p60	11.7	11.9	12.3	13.1	13.4	13.8
p60 - p65	9.9	10.1	10.5	10.1	10.3	10.7
p65 - p70	8.4	8.7	9.1	9.1	9.3	9.7
p70 - p75	7.6	7.9	8.4	7.5	7.8	8.2
p75 - p80	7.5	7.7	8.1	7.1	7.4	7.9
p80 - p85	5.8	6.1	6.5	5.4	5.6	6.0
p85 - p90	5.5	5.8	6.2	4.7	5.0	5.5
p90 - p95	5.2	5.5	5.9	5.7	6.0	6.4
p95 - p100	3.8	4.2	5.9	3.9	4.2	5.9
p99 - p100	3.2	3.5	5.0	3.4	3.7	5.3
p99.90 - p100	1.6	1.8	2.8	1.4	1.7	3.0
FL Top 100	0.44-0.48	0.61-0.65	1.40-1.44			
FL Top 10	0.32-0.35	0.47-0.49	1.17-1.20			

Note: Range of income-wealth ratio for the FL is attributed to (a) range of values in Y_T and (b) scenarios of assets bifurcation and their different yield. Most Plausible Estimates assumes $\delta_A = 2\%, \delta_P = 2.5\%$; Absolute Upper Bound assumes $\delta_A = 4\%, \delta_P = 4\%$, Lower Bound assumes; $\delta_A = 0.08\%, \delta_P = 0.08\%$.

Table 5.3: The ratio of personal income to wealth ($\frac{Y_{PID}}{W} * 100$)

Wealth Percentiles	Households			Individuals		
	Lower Bound	Most Plausible	Upper Bound	Lower Bound	Most Plausible	Upper Bound
p0 - p5	1037.5	1036.4	1035.3	1166.7	1165.1	1163.1
p5 - p10	187.7	188.2	188.8	510.2	511.2	512.2
p10 - p15	92.0	92.5	93.1	144.7	145.1	145.7
p15 - p20	53.0	53.8	54.6	87.2	88.0	88.9
p20 - p25	37.3	38.1	38.9	51.4	52.3	53.2
p25 - p30	25.7	26.5	27.4	38.6	39.4	40.3
p30 - p35	21.0	21.9	22.8	27.1	28.1	29.1
p35 - p40	20.0	21.0	22.0	20.2	21.1	22.0
p40 - p45	16.7	17.6	18.6	20.5	21.4	22.4
p45 - p50	15.0	15.9	16.9	16.6	17.5	18.5
p50 - p55	12.7	13.6	14.6	14.0	14.9	15.8
p55 - p60	12.0	12.8	13.8	13.4	14.2	15.2
p60 - p65	10.2	11.0	12.0	10.4	11.3	12.2
p65 - p70	8.7	9.5	10.4	9.4	10.3	11.3
p70 - p75	7.9	8.7	9.7	7.8	8.6	9.5
p75 - p80	7.8	8.7	9.6	7.4	8.2	9.2
p80 - p85	6.1	6.9	7.8	5.7	6.5	7.5
p85 - p90	5.8	6.6	7.5	5.0	5.8	6.8
p90 - p95	5.4	6.2	7.0	5.9	6.7	7.6
p95 - p100	4.0	4.8	6.9	4.1	4.8	6.9
p99 - p100	3.4	4.1	5.9	3.6	4.2	6.1
p99.90-p100	1.7	1.9	2.8	1.4	1.7	3.1
FL Top 100	0.44-0.48	0.63-0.67	1.50-1.54			
FL Top 10	0.33-0.36	0.59-0.62	1.27-1.30			

Note: As in Table 5.2

5.5 The Capital and Labour Income Ratios ($\frac{Y_{KD}}{W}$ and $\frac{Y_L}{W}$)

Tables 5.4 and 5.5 present the $\frac{Y_{KD}}{W}$ and $\frac{Y_L}{W}$ ratios for different wealth groups of individuals and HHs. As predicted by Proposition 1, both ratios decrease with wealth. As can be seen in Appendix I, $\frac{Y_{KD}^*}{W}$ ratios are also decreasing in wealth. Though our estimates of direct capital income and labour income put separately are less precise than the total reported income, the consistency of the findings across categories and range lend credence to our estimates of these ratio.

Table 5.4: Reported Capital income–wealth ratio ($\frac{Y^*_{KD}}{W} * 100$) across wealth groups

Wealth Percentiles	Households			Individuals		
	Lower Bound	Most Plausible	Upper Bound	Lower Bound	Most Plausible	Upper Bound
p0 - p5	-17.6	-23.0	-28.8	-28.6	-33.6	-39.1
p5 - p10	11.0	16.8	23.3	23.3	28.9	34.8
p10 - p15	6.3	10.2	14.2	8.1	12.2	16.6
p15 - p20	4.9	7.5	10.4	6.1	9.1	12.3
p20 - p25	4.2	6.6	9.1	5.1	8.0	11.3
p25 - p30	3.4	5.7	8.2	4.0	6.4	9.0
p30 - p35	2.9	4.8	6.9	3.1	5.1	7.3
p35 - p40	2.9	4.8	7.0	2.7	4.5	6.5
p40 - p45	2.4	4.2	6.1	2.8	4.7	6.7
p45 - p50	2.5	4.3	6.2	2.5	4.2	6.1
p50 - p55	2.2	4.0	5.9	2.6	4.4	6.3
p55 - p60	2.3	3.9	5.7	2.3	4.0	5.9
p60 - p65	2.0	3.6	5.3	2.1	3.7	5.3
p65 - p70	2.1	3.7	5.5	2.0	3.6	5.3
p70 - p75	1.8	3.3	4.9	2.0	3.9	5.9
p75 - p80	2.4	3.9	5.6	1.9	3.3	4.9
p80 - p85	1.7	3.0	4.5	1.6	2.9	4.4
p85 - p90	1.6	3.0	4.5	1.6	3.0	4.5
p90 - p95	1.6	2.9	4.3	1.6	3.0	4.5
p95 - p100	0.5	1.2	2.8	0.5	1.2	2.9
p99 - p100	0.3	0.9	2.1	0.3	1.0	2.4
p99.90 - p100	0.2	0.6	1.3	0.2	0.7	1.7
FL Top 100	0.1	0.5	1.1			

Before concluding this section, a few remarks are in order. We have worked with three possibilities regarding reporting of capital income by the wealthy groups (the top 5% of the GE units and those on the FL). Even if we go by the absolute upper bound on the income reported by super-wealthy group (the wealthiest 1% and the FL units) and use the point estimates of the income reported by the rest of the unit, the reported ratios are smallest for the super-wealthy group. Therefore, the uncertainty over the part of the capital income that gets reported in the ITRs does not change the fact that the wealthiest group reports the lowest income in relative terms.

Besides, our estimates of the exempt income do not include the tax-exempt long-term capital gains under Section 54 of the Indian Income Tax Act (ITA), which mainly include capital gains from the sale of housing property. We do not have any source of information on this form of income. However, at any point, such capital gains can accrue to only a minuscule fraction of the individuals. Therefore, inclusion or exclusion of these gains is not expected to significantly affect the income-wealth ratios presented here. At any rate, this form of capital income can be a significant fraction of the total

income only for the middle wealth groups; for the wealthy and the super-wealthy, the residential property itself is a tiny fraction of their wealth holding. This means that by not including the capital gains under Section 54, we might have slightly underestimated total income at the middle wealth levels. By implication, the inclusion of these gains will add to the sharpness of the fall in the ratio. The other forms of tax-exempt incomes accrue to the taxpayers rather infrequently. We expect these incomes to be a negligible fraction of the taxable income and not bias our results.

Table 5.5: Labour income–wealth ratio ($\frac{Y_L}{W} * 100$) across wealth groups

Wealth Percentiles	Households			Individuals		
	Lower Bound	Most Plausible	Upper Bound	Lower Bound	Most Plausible	Upper Bound
p0 - p5	-1.6	-3.8	-6.2	14.5	12.0	9.1
p5 - p10	155.1	157.6	160.2	394.0	396.0	398.1
p10 - p15	79.1	81.4	83.8	123.3	125.3	127.6
p15 - p20	45.4	46.6	47.8	77.3	78.7	80.2
p20 - p25	30.4	31.6	32.9	42.9	44.2	45.7
p25 - p30	20.4	21.4	22.6	32.1	33.3	34.5
p30 - p35	16.4	17.2	18.1	22.2	23.1	24.1
p35 - p40	15.2	16.1	17.1	15.9	16.7	17.5
p40 - p45	12.6	13.4	14.2	15.7	16.5	17.4
p45 - p50	10.7	11.5	12.4	12.4	13.2	14.1
p50 - p55	8.7	9.6	10.5	9.7	10.6	11.4
p55 - p60	7.9	8.7	9.6	9.2	10.0	11.0
p60 - p65	6.6	7.3	8.1	6.8	7.5	8.3
p65 - p70	5.1	5.8	6.6	5.9	6.6	7.4
p70 - p75	4.4	5.1	5.9	3.5	4.4	5.4
p75 - p80	4.1	4.7	5.4	4.6	5.3	6.0
p80 - p85	3.1	3.7	4.3	2.8	3.4	4.0
p85 - p90	2.8	3.5	4.2	2.1	2.6	3.3
p90 - p95	2.5	3.1	3.6	2.9	3.5	4.1
p95 - p100	2.1	2.4	2.5	2.2	2.5	2.7
p99 - p100	1.8	2.2	2.4	2.1	2.4	2.6
p99.90 - p100	1.1	1.6	1.8	1.0	1.5	1.6
FL Top 100						

Finally, we want to note that the ratios reported here are more reliable for the relatively wealthy groups. The value of ratios reported for the low-wealth groups could change, especially for p1-p30, if we had income information for the sample dropped from analysis, as the dropped sample comprises mainly of low-wealth groups.

6. The Correlates of Income-Wealth Ratio

In this section, we use regression analysis to identify the main determinants of the income-wealth ratios, presented in the previous section. To this end, we consider all leading versions of income discussed above, though our focus is on total income reported to tax authorities, Y_R.

Our set of explanatory variables is partitioned into two categories. The first category, referred to as *wealth and asset share controls*, comprises the individual's (or household's) wealth level and the composition of assets held. The second category, referred to as *other controls,* includes characteristics (e.g., education level, gender) that are likely to independently influence the income-wealth ratios.

6.1 Wealth and Asset Share Controls

Given the predictions emanating from our model in Section 2, wealth is expected to be a significant predictor of the ratio of the reported income and the total personal income relative to wealth (Proposition 2). Indeed, from Section 5, we can see that all versions of the income-wealth ratio decrease with wealth. So, every empirical specification that we consider includes the (natural) log of wealth as an explanatory variable.

Proposition 3 demonstrates that assets' shares in the wealth portfolio have a bearing on the income-wealth ratio. Accordingly, we view asset shares as correlates of the income-wealth ratios. This gives us five regressors -- *Equity, Banking, Advances, Agri_Land,* and *Com_Prop,* measured as share of total assets in equity, bank deposits, personal advances, farmland, and commercial property, respectively. These and other correlates are defined in Table 6.1 below.

The issue of tax evasion is pertinent for our analysis. Official reports and other studies show taxpayers misuse farmland to underreport their income.[39] Another set of studies suggests that people underreport rental incomes from commercial properties. This issue also motivates us to use the shares of farmland and commercial properties — the primary channels of underreporting — as controls. Observe that an increase in shares of these assets is associated with a decrease in safe assets' share (see Appendix I). Given this relationship among asset shares, Proposition 4 expects the income-wealth ratios to decrease with share of farmland and commercial properites if there is underreporting of farm and rental incomes.

Summing up, the following hypothesis is derived from the discussions in Sections 2 and 5:

H1: The income-wealth ratio is: decreasing in wealth; decreasing in the share of agricultural land and commercial property; and decreasing in the share of risky assets such as equity.

Table 6.1: Description of variables used in the regression analysis and their data sources

Variable	Description	Source
log W	Natural log of wealth	GE Affidavit Data
Banking	Share of the banking assets in total assets, defined as the "value of cash + deposits in bank + NSS + postal savings" divided by the "value of all assets combined"	GE Affidavit Data

[39]See Compliance Audit of Union Government Department of Revenue Direct Taxes by the Comptroller and Auditor General of India (2019).

$Equity$	Share of equity in total assets. Equity comprises bonds, debentures, and shares/stocks owned	GE Affidavit Data
$Advances$	Share of personal advances in total assets. Personal advances are private loans given out to others	GE Affidavit Data
$Agri\_Land$	Share of agricultural land in total assets, defined as the "value of agricultural land" divided by the "value of all assets combined"	GE Affidavit Data
Com_Prop	Share of commercial property in total assets. Commercial property comprises commercial building and non-agriculture land.	GE Affidavit Data
$Vote$	When candidate i contests in constituency j, $Vote =$ (votes received by candidate i)/(votes received by winner of constituency j)	ECI Results Data
$Criminal$	Number of criminal cases registered against the candidate	GE Affidavit Data
$Education$	Variable measures the highest educational degree attained by the candidate. The higher the degree, the larger the value taken by the variable.	GE Affidavit Data
$Winner$	$Winner$ dummy takes value 1 if the candidate won in both the 2014 and the 2019 General Elections; 0 otherwise	ECI Results Data
$Party$	Party dummy takes value 1 if the candidate contested elections as a registered state or national party nominee; 0 otherwise	ECI Results Data
$Agriculture$	$Agriculture = 1$ if the candidate's profession is Agriculture and Allied Activities; 0 otherwise	GE Affidavit Data
$Social$	$Social$= 1 if the candidate's claimed profession is Social Work and/or Politician;[40] 0 otherwise	GE Affidavit Data
$Male$	Male dummy takes value 1 if the candidate's gender is male; 0 otherwise	ECI Results Data
$Unresev$	Unresev dummy is 1 if the (social) category is "General", i.e., Unreserved (UR); 0 otherwise	ECI Results Data
2019	2019 is year dummy. It takes value 1 if W and Y are reported for the year 2019; 0 otherwise	GE Affidavit Data

6.2 Other Controls

The paper's formal theory does not address a host of individual-specific factors that work together to generate income and wealth for individuals and households. For HHs, data limitations do *not* allow us to model these factors of interest. However, for the category of individuals (candidates), we have assembled a comprehensive set of such variables (see Table 6.1 for details).
First, we have data for three demographic variables: dummy variable D_{Male} (which takes the value one if the candidate is male), dummy variable $D_{Unresev}$ (which takes the value one if the candidate's social category is "General"), and *Education* (which measures the educational attainment of the candidate). Some of our empirical specifications examine the determinants of the income-wealth ratios controlling for these demographic characteristics.

Second, the frequency and size of cash-based transactions differ across vocations, which means that the ability to underreport income varies across occupations. We use two profession dummies – *Agriculture* and *Social (worker)* – to allow for profession-fixed effects in our empirical analysis of reported income-wealth ratios.

[40] Most of candidates claiming to be full time politician have also mentioned their occupation as social worker.

Third, we examine if the reported income-wealth ratios are different for candidates with a criminal history. To this end, we use the variable *Criminal* (which measures the number of criminal cases against a candidate) as an explanatory variable in our analysis.

Fourth, a concern for our empirical analysis is that politicians may exhibit reporting behaviours that are markedly different from the rest of the population, i.e., the income-wealth ratios candidates report may differ from the rest of the population because they have political abilities. To examine if political ability affects the reported ratios, we use the variable *Vote* (which measures the candidate's vote share) as a proxy for political ability. We assume that the larger a candidate's vote share, the greater is her political ability. Further, we examine the reported ratios for candidates with "extraordinary" political abilities. This set consists of winners of both GEs - dummy variable D_{Winner} takes the value one in this case.

Summing up, we propose the following hypothesis for individuals (candidates):

H2: The income-wealth ratio is: decreasing in the vote share of the candidate; different across professions; increasing in education; and falling in the degree of criminality.

6.3 The Empirical Strategy

Our paper makes use of two data sets: the data set of households (including FL families) and the data set of individuals (candidates). For each data set, we estimate, as a baseline specification, the following model of the income-wealth ratio:[41]

$$log\left(\frac{Y_i}{W_i}\right) = \alpha_0 + \alpha_1 log W_i + \beta_S S_i + \beta_Y 2019_i + \beta_D X_i + \epsilon_i \qquad (6.1)$$

where (Y_i / W_i) is the income-wealth ratio; W_i is wealth; S_i is the vector of regressors representing shares of income yielding assets (i.e., deposits in banks, personal advances (loans), equity, agricultural land, and commercial properties);[42] 2019 is a year-specific dummy (which takes the value 1 if income and wealth are reported in 2019; and is 0 otherwise); X_i are other controls (see footnote for details);[43] and ϵ_i is a mean-zero error term, presumed to be orthogonal to the regressors. The model is estimated using ordinary least squares.

To show that our main results are robust with respect to model specification, we also consider two versions of the baseline model (6.1). In the first version, we estimate the relationship between income-wealth ratio and wealth without including the asset shares (i.e., S_i) as regressors. In the second version, asset shares are retained as regressors but "other" controls (i.e., X_i) are excluded. Further, model (6.1) and its two variants are estimated for different versions of income, i.e., by

[41] The distribution of W is skewed towards large values. As seen from Figure 3.1 and Figure A5.2 in Appendix-I, W and all versions of the (Y/W) ratio follow a log-normal distribution. Given this, and following the approach in Asher and Novosad (2019) and Fisman et al (2019), we use $log\ W$ instead of W as an explanatory variable, and log of the (Y/W) ratios as dependent variables.

[42] The residual category of assets includes assets that do not yield income directly, such as gold, jewellery, durables, and properties used for housing.

[43] For the data set of households, data limitations allow us to include just one variable in X_i -- this is the dummy variable $Unresev$ (which takes the value 0 for households classified as SC or ST, and is 1 otherwise). For the data set of individuals, we are in a position to include many additional variables in X_i (refer to Section 6.2). Some of these variables are the individual's educational attainment, the individual's gender, number of criminal cases against the individual, and so on.

taking $\underline{Y_i}$ to be Y_{Td}, Y_T, Y_R and Y_{PID}.[44]

Finally, how do we interpret the model in (6.1)? We do not claim that our results have a causal interpretation. Instead, we view the regression as a procedure that *describes* patterns in our data set – nonetheless, the estimated coefficients on the regressors provide *suggestive* evidence for our theories explaining the income-wealth ratios.

6.4 Results

Tables 6.2 and 6.3 report the regression results for households (HHs), including the FL families. Tables 6.4 and 6.5 report the regression results for individuals.

Consider, first, the estimated link between wealth and the income-wealth ratio. As can be seen from Tables 6.2-6.5, *all* versions of the income-wealth ratio are decreasing in wealth. On average, a 1% increase in wealth is associated with approximately 0.5% (respectively, 0.6%) decrease in $\frac{Y_R}{W}$ for HHs (respectively, individuals).

Consider, now, the impact of asset shares on the income-wealth ratio. The income-wealth ratios are, as expected, increasing in the share of bank deposits and personal advances. But, the coefficients on *Equity, Agri_Land* and *Com_Prop* are puzzling at first and merit scrutiny. Notice that a one percentage point increase in equity share leads to approximately 2% increase in the $\frac{Y_R}{W}$ ratio reported by HHs (refer to Table 6.3). The effect equity on individuals' income is smaller but significant (refer to Table 6.5). For every percentage point increase in farmland share, there is about 0.3% (respectively, 0.2%) decrease in the ratio for HHs (respectively, individuals). A one percentage point increase in commercial property share is associated with 0.2% decrease in the ratio for HHs and individuals.

Absent tax evasion, these results would seem counterintuitive. The rental yield is generally higher than the dividend yields.[45] As discussed in Section 2, the rate of direct returns tends to be the lowest for equity compared to the other income-yielding assets, including commercial properties. Thus, at any given level of wealth, direct capital income, and hence $\frac{Y_R}{W}$, is expected to decrease with the equity share. However, we find the coefficient of the equity share to be positive. In other words, holding constant wealth levels and other factors, the more significant the share of equity in the asset portfolio, the higher is the reported income. On the other hand, larger shares of farmland and commercial property are correlated with lower reported incomes, and vice versa.

Our findings provide three pieces of evidence for underreporting of income across wealth groups. The first is the negative and statistically significant coefficients for the shares of farmland and commercial properties. If there is underreporting on account of farmland or rental properties, the income-wealth ratios would decrease with an increase in shares of these assets, as is predicted by Proposition 4. This is exactly what we find. An underreporting of rental incomes reduces the tax burden on recipients, and in the process pulls down the reported values of taxable and the total

[44] We have also estimated model (6.1) with the added term $(\log W)^2$. See Appendix-I. Under this version of the model, coefficients on $\log W$ and $(\log W)^2$ are both significant, with positive and negative sign, respectively. This suggests a flattening of the decreasing pattern in the income-wealth ratios; otherwise, the results are very similar to the ones presented above. However, the effect of $(\log W)^2$ on the income-wealth ratios becomes comparable to the effect of $\log W$ at values of W three-times the maximum wealth level observed in the data. Moreover, values of the R^2 do not change much.

[45] For a review of literature on this issue see Jorda et al. (2017).

income declared. Given that the coefficient of equity is positive, the reported rental income can even be smaller than the dividend income, holding fixed the value of the underlying asset.

The agricultural income share's negative and statistically significant coefficient has somewhat different implications. Farm income is tax exempt; so, taxpayers have very little incentive to underreport it. However, farmland offers its owners the opportunity to disguise a part of their taxable income as (non-taxable) agricultural income, thus pulling down the declared value of the taxable income. This in turn pulls down reported value of the total income. As can be seen from regression results, all versions of the $\frac{Y}{W}$ ratio are decreasing in the farmland share. This result is supported by the above-cited audit reports showing such misuse of farmlands by taxpayers.

The second piece of evidence is provided by the positive and statistically significant coefficient on the dummy variable *2019.* Thus, *ceteris paribus*, average reported income was higher in 2019 compared to 2014. Our analysis suggests that taxpayers from all groups have reported higher income in 2019 than before,[46] implying that in the past all groups have managed to evade tax. The increased reporting in 2019 is likely due to a dramatic increase in searches and surveys conducted by the income tax department. According to a Comptroller and Auditor General of India report (2021), the number of operations to penalise tax evasion increased from 5,896 in 2013-14 to approximately 14,000 in 2019.

The third piece of evidence for general evasion comes from the positive and statistically significant coefficient on the variable *Votes*. *Ceteris paribus*, the larger a candidate's vote share, the higher is their reported ratio, and vice versa. In other words, the stronger the political abilities, relatively high is the reported income. To make sense of this result, we should remember that the media and civil society scrutiny are stricter for candidates with a serious chance of winning an election.[47] To avoid falling foul of the Election Commission and risking their chance at the hustings, these candidates have a stronger incentive to report their incomes truthfully. That the reported ratios increase with vote share also implies that media and official scrutiny have a more pronounced effect on reporting of income relative to wealth.

This result suggests that non-politician citizens (presumably with little or no political abilities) report smaller incomes than similarly placed politicians – it indicates towards a general tendency to underreport income across wealth groups, except for those who cannot avoid the media glare and official scrutiny. Technically, our results likely have an upward bias; we overestimate the reported income across all wealth groups. By implication, the income reported by ordinary citizens is likely smaller than what our results indicate.

Controlling for the vote share, we find that the income-wealth ratios for "super-politicians" — i.e., candidates who won both GEs — are not different from the rest (the coefficient on the variable $Winner$ is statistically insignificant, and not shown here). The positive (and significant) coefficient on the variable $Party$ shows that the average reported income is relatively high for candidates belonging to national- and state-level parties, even though they tend to be wealthier than the rest. This finding further underscores the role of scrutiny and enforced transparency on income reporting behaviours.

We briefly draw out five conclusions regarding the impact of various individual characteristics on the reported income-wealth ratios. First, the dummy variable $Unresev$ is not consistently

[46]The interaction of 2019 with wealth turns out to be insignificant.
[47]It seems plausible to assume that the candidates have a good sense of their electoral prospects. So, candidates with better prospects end up with relatively high vote shares *ex-post*.

significant, suggesting that social identity does not systematically affect the reported income-wealth ratios. Second, the effect of candidates' age (not shown here) is not significant. Third, since the coefficients on variables *Agriculture* and *Social* worker are negative, c*eteris-paribus*, full time agriculturists and politicians report relatively low incomes. Fourth, women, on average, report smaller incomes than men (refer to the coefficient on variable $Male$). This finding appears to be a consequence of two factors: at any given wealth level, labour market outcomes, including wages, are worse for women. In all, they receive less than one-fifth of the national labour income. Also, as women own a larger share of non-income yielding assets like gold and jewellery, the income-wealth ratios are expected to be relatively low for them. Fifth, candidates' educational qualification positively correlates with the $\frac{Y}{W}$ratios, while the degree of criminality has a negative association.

Table 6.2: $\frac{Y_{Td}}{W}$ and $\frac{Y_T}{W}$ income-wealth ratios for households (including the FL)

Ratios/ Variables	$\log\left(\frac{Y_{Td}}{W}\right)$			$\log\left(\frac{Y_T}{W}\right)$		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
$log\ W$	-0.513*** (0.008)	-0.524*** (0.008)	-0.524*** (0.008)	-0.533*** (0.007)	-0.541*** (0.007)	-0.541*** (0.007)
$Banking$		0.010*** (0.001)	0.010*** (0.001)		0.009*** (0.001)	0.009*** (0.001)
$Equity$		0.021*** (0.001)	0.020*** (0.001)		0.019*** (0.001)	0.019*** (0.001)
$Advances$		0.012*** (0.002)	0.012*** (0.002)		0.011*** (0.002)	0.011*** (0.002)
$Agri\_Land$		-0.009*** (0.001)	-0.009*** (0.001)		-0.008*** (0.001)	-0.009*** (0.001)
Com_Prop		-0.002*** (0.001)	-0.002*** (0.001)		-0.002*** (0.001)	-0.002*** (0.001)
2019	0.166*** (0.024)		0.120*** (0.022)	0.158*** (0.023)		0.116*** (0.022)
$Unresev$	0.049 (0.027)		-0.001 (0.025)	0.039 (0.027)		-0.007 (0.025)
$Constant$	10.293*** (0.121)	10.515*** (0.123)	10.453*** (0.124)	10.945*** (0.115)	11.123*** (0.118)	11.065*** (0.118)
R^2	0.574	0.646	0.648	0.609	0.671	0.672
$Observation$	7533	7533	7533	7533	7533	7533

*Note: Robust standard errors in parentheses. Significance level (p-value): *0.05 **0.01 ***0.001*

Our finding on Y_{Td}, Y_T and Y_{PID} are very similar to Y_R. In sum, our results are robust to the inclusion/exclusion of various dummies and individual characteristics. Moreover, we get similar results if we restrict the regression analysis to the wealthiest 50% of units. This exercise is of

interest since the sample used by us closely resembles the population of relatively wealthy contestants and the super wealthy in the FL.

However, we would like to emphasise that some of our results are sensitive to the specification of the dependent variable. To avoid the effect of skewed distribution of the variables $\frac{Y_i}{W_i}$, recall that we have chosen to work with $log\left(\frac{Y_i}{W_i}\right)$as the dependent variable. If we substitute $\frac{Y_i}{W_i}$ for $log\left(\frac{Y_i}{W_i}\right)$ as the dependent variable and restrict the analysis to the top 50% of the GE data points, our results still remain very similar to the ones presented above in terms of sign and significance levels of the coefficients. However, some results change if the substitution of $\frac{Y_i}{W_i}$ for $log\left(\frac{Y_i}{W_i}\right)$ is applied to the entire dataset (see Appendix-I), but wealth remains the most important determinant of $\frac{Y_i}{W_i}$ ratios. In view of the skewed distributions of $\frac{Y_i}{W_i}$ ratios for the bottom 50% of the data points, such changes in the results are not surprising.

Table 6.3: $\frac{Y_R}{W}$ and $\frac{Y_{PID}}{W}$ income-wealth ratios for households (including the FL)

Ratio/Variable	$\log\left(\frac{Y_R}{W}\right)$			$\log\left(\frac{Y_{PID}}{W}\right)$		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
$log\ W$	-0.498*** (0.007)	-0.517*** (0.007)	-0.516*** (0.007)	-0.475*** (0.006)	-0.474*** (0.007)	-0.474*** (0.007)
$Banking$		0.009*** (0.001)	0.009*** (0.001)		0.008*** (0.001)	0.008*** (0.001)
$Equity$		0.021*** (0.001)	0.020*** (0.001)		0.015*** (0.001)	0.015*** (0.001)
$Advances$		0.011*** (0.001)	0.010*** (0.001)		0.008*** (0.001)	0.007*** (0.001)
$Agri\_Land$		-0.003*** (0)	-0.003*** (0)		-0.006*** (0)	-0.006*** (0)
Com_Prop		-0.002*** (0)	-0.002*** (0)		-0.005*** (0)	-0.005*** (0)
2019	0.166*** (0.021)		0.117*** (0.02)	0.148*** (0.019)		0.110*** (0.018)
$Unresev$	0.041 (0.024)		0 (0.023)	0.041 (0.022)		0.005 (0.021)
$Constant$	10.480*** (0.111)	10.722*** (0.111)	10.661*** (0.11)	10.232*** (0.102)	10.278*** (0.105)	10.220*** (0.105)
R^2	0.627	0.682	0.683	0.647	0.701	0.702
$Observation$	7533	7533	7533	7533	7533	7533

*Note: Robust standard errors in parentheses. Significance level (p-value): *0.05 **0.01 ***0.001*

Table 6.4: $\frac{Y_{Td}}{W}$ and $\frac{Y_T}{W}$ income-wealth ratios for individuals

Ratio/ Variable	$\log\left(\frac{Y_{Td}}{W}\right)$			$\log\left(\frac{Y_T}{W}\right)$		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
log W	-0.718***	-0.616***	-0.677***	-0.732***	-0.632***	-0.692***
	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)
Banking		0.008***	0.006***		0.007***	0.006***
		(0.001)	(0.001)		(0.001)	(0.001)
Equity		0.013***	0.012***		0.012***	0.011***
		(0.002)	(0.002)		(0.002)	(0.002)
Advances		0.012***	0.010***		0.011***	0.009***
		(0.001)	(0.001)		(0.001)	(0.001)
Agri_Land		-0.007***	-0.006***		-0.007***	-0.006***
		(0.001)	(0.001)		(0.001)	(0.001)
Com_Prop		-0.002***	-0.002***		-0.002***	-0.002***
		(0.001)	(0.001)		(0.001)	(0.001)
Vote	0.567***		0.484***	0.541***		0.465***
	(0.041)		(0.039)	(0.039)		(0.037)
Criminal	-0.005**		-0.005***	-0.005**		-0.004***
	(0.002)		(0.001)	(0.002)		(0.001)
Education	0.029***		0.021***	0.030***		0.022***
	(0.006)		(0.005)	(0.005)		(0.005)
2019	0.136***		0.131***	0.135***		0.131***
	(0.024)		(0.023)	(0.023)		(0.022)
Male	0.049		0.129**	0.048		0.124**
	(0.041)		(0.04)	(0.039)		(0.038)
Unresev	0.142***		0.088***	0.131***		0.082**
	(0.026)		(0.026)	(0.026)		(0.025)
Party	0.137***		0.137***	0.146***		0.146***
	(0.03)		(0.028)	(0.029)		(0.027)
Agriculture			-0.211***			-0.208***
			(0.04)			(0.039)
Politicians			-0.190***			-0.169***
			(0.038)			(0.037)
Constant	12.859***	11.844***	12.269***	13.406***	12.434***	12.852***
	(0.131)	(0.124)	(0.145)	(0.124)	(0.117)	(0.138)
R^2	0.679	0.693	0.708	0.705	0.715	0.73
Observation	6956	6956	6956	6956	6956	6956

*Note: Robust standard errors in parentheses. Significance level (p-value): *0.05 **0.01 ***0.001*

Table 6.5: $\frac{Y_R}{W}$ and $\frac{Y_{PID}}{W}$ income-wealth ratios for households

Individuals	$\log\left(\frac{Y_R}{W}\right)$			$\log\left(\frac{Y_{PID}}{W}\right)$		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
log W	-0.687***	-0.605***	-0.660***	-0.644***	-0.555***	-0.603***
	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.008)
Banking		0.008***	0.006***		0.006***	0.005***
		(0.001)	(0.001)		(0.001)	(0.001)
Equity		0.014***	0.013***		0.010***	0.009***
		(0.002)	(0.002)		(0.001)	(0.001)
Advances		0.011***	0.009***		0.008***	0.006***
		(0.001)	(0.001)		(0.001)	(0.001)
Agri_Land		-0.002***	-0.001**		-0.005***	-0.004***
		(0)	(0)		(0)	(0)
Com_Prop		-0.002**	-0.002***		-0.004***	-0.004***
		(0.001)	(0)		(0)	(0)
Vote	0.492***		0.426***	0.437***		0.377***
	(0.034)		(0.033)	(0.031)		(0.03)
Criminal	-0.003***		-0.002***	-0.003***		-0.002***
	(0.001)		(0.001)	(0.001)		(0.001)
Education	0.025***		0.019***	0.026***		0.019***
	(0.005)		(0.005)	(0.005)		(0.005)
2019	0.135***		0.124***	0.123***		0.117***
	(0.021)		(0.02)	(0.019)		(0.018)
Male	0.072*		0.119***	0.046		0.110***
	(0.035)		(0.035)	(0.033)		(0.032)
Unresev	0.125***		0.081***	0.120***		0.080***
	(0.024)		(0.023)	(0.022)		(0.021)
Party	0.125***		0.127***	0.107***		0.109***
	(0.026)		(0.026)	(0.026)		(0.026)
Agriculture			-0.187***			-0.164***
			(0.034)			(0.03)
Politicians			-0.162***			-0.145***
			(0.032)			(0.03)
Constant	12.841***	12.015***	12.385***	12.340***	11.475***	11.779***
	(0.115)	(0.112)	(0.13)	(0.112)	(0.108)	(0.125)
R^2	0.721	0.727	0.741	0.728	0.738	0.751
Observation	6956	6956	6956	6956	6956	6956

*Note: Robust standard errors in parentheses. Significance level (p-value): *0.05 **0.01 ***0.001*

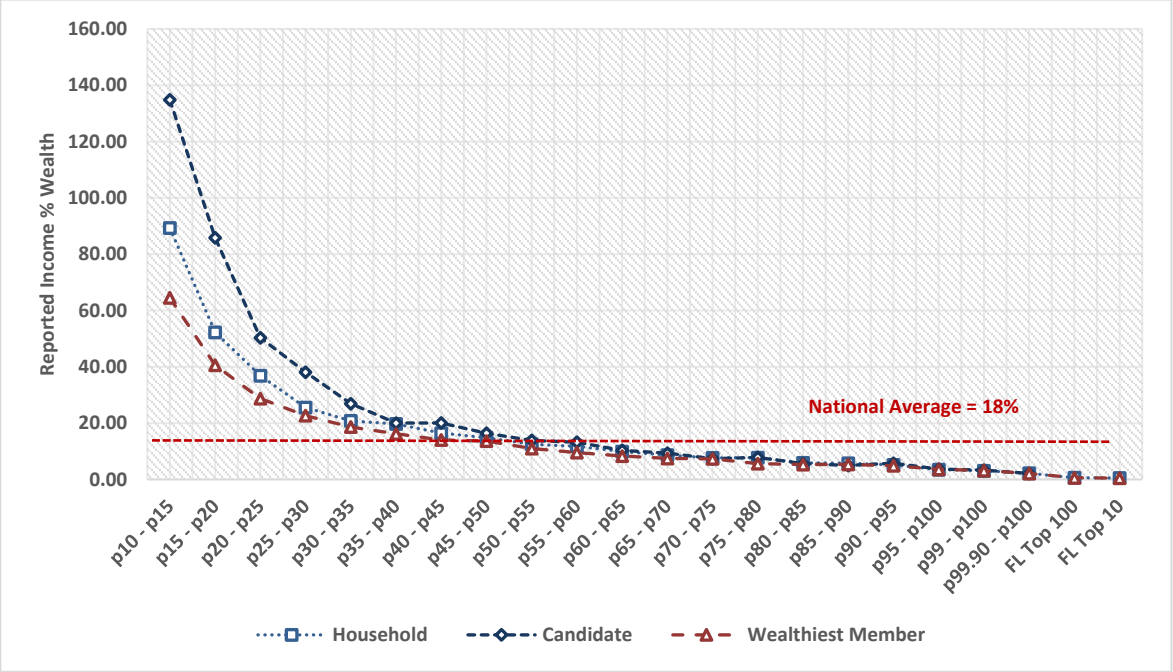
7. The Missing Income at the Top: *How much and how come?*

As is clear from our findings in Sections 5 and 6, incomes reported by the bottom 10% of wealth groups are several times their wealth. In contrast, incomes reported by the groups at the top of the wealth pyramid are a minuscule fraction of their wealth. The same is true for the other forms of income reported by the different groups vis-à-vis their wealth. The reported income relative to wealth decreases continuously until it is reduced to a negligible fraction of wealth for super-wealthy groups.

On the face of it, these findings do not seem surprising. In view of Proposition 2 and the related discussion in Sections 2 and 3, we expect the personal income-to-wealth ratio, and reported income-to-wealth ratio to be decreasing in wealth. Moreover, we expect the income-wealth ratio to be relatively low for wealthy groups because of their supersized wealth holdings.

Yet, the income-wealth ratios reported by wealthy Indians seem to be inexplicably low. For one, these ratios are far below the national average. In the decade of the two GEs studied by us (i.e., during 2010-20), the average national income was 18-20% of the average private wealth.[48] The income levels reported by the wealthy groups pertain to the same period but are much small by comparison. Figure 7.1 depicts the $\frac{Y_R}{W}$ ratios for wealthy groups compared to the ratio of national income to wealth. The ratios reported by wealthy groups are far below the national average. The total income-wealth ratio reported by the wealthiest 20% is less than a third of the national average. The estimated ratio for the wealthiest 0.1% is just 12% of the national average. For families in the FL, it is merely one-twentieth of the national average!

Figure 7.1: Ratio of total reported income to wealth, i.e., $\frac{Y_R}{W}$, versus the national average of $\frac{Y}{W}$



Another helpful perspective is provided by the average rate of returns on capital, i.e., the capital income expressed as a percentage of the value of the capital stock. The national average rate of returns on the aggregate stock of capital can be estimated as the capital share of the national income

[48] In other words, during this period, the ratio of private wealth to the national income hovered in the range of 5-5.6. See GIR (2022, page 78).

times the ratio of national income to national wealth. For the decade relevant to this study (2010-20)), the capital share of the national income has been upward of 40%. In the same period, the national income has been 18-20% of the national wealth, most of which is private wealth. [49]

Therefore, even by a conservative approach, the national average of the rate of returns on private capital turns out to be at least 7.2% ($= 0.4 \times 0.18 \times 100$). Formally put, for the country as a whole, the average ratio of the capital income to wealth was upward of 7.2% in the decade covered by our study. During that period, one could easily get this kind of return even from fixed deposit accounts with commercial banks. The returns from mutual funds and direct equity investments were much higher.

As the rate of returns on capital is increasing in wealth, the rate of returns for the wealthy should be greater than the national average. Thus, for wealthy groups, capital income should be significantly higher than 7.2% of their wealth. Using corporate debt rates as a reference point, the net rate of returns on the capital owned by the top wealth groups, say, for the wealthiest 20%, should be upward of 10%. This assumption is additionally justified given that during the last two decades, the average Indian growth rate has been upward of 6%, and historically, the rate of returns on capital has been several percentage points higher than the economic growth rate.

In simpler terms, even if we disregard the labour income earned by the super and ultra-wealthy, their total income is expected to be greater than 10% of their wealth simply on account of their capital income. However, the income levels they reported present a strikingly different picture.

Figure 7.2: Total reported income, Y_R, as a percentage of the capital income (7.2% of wealth).

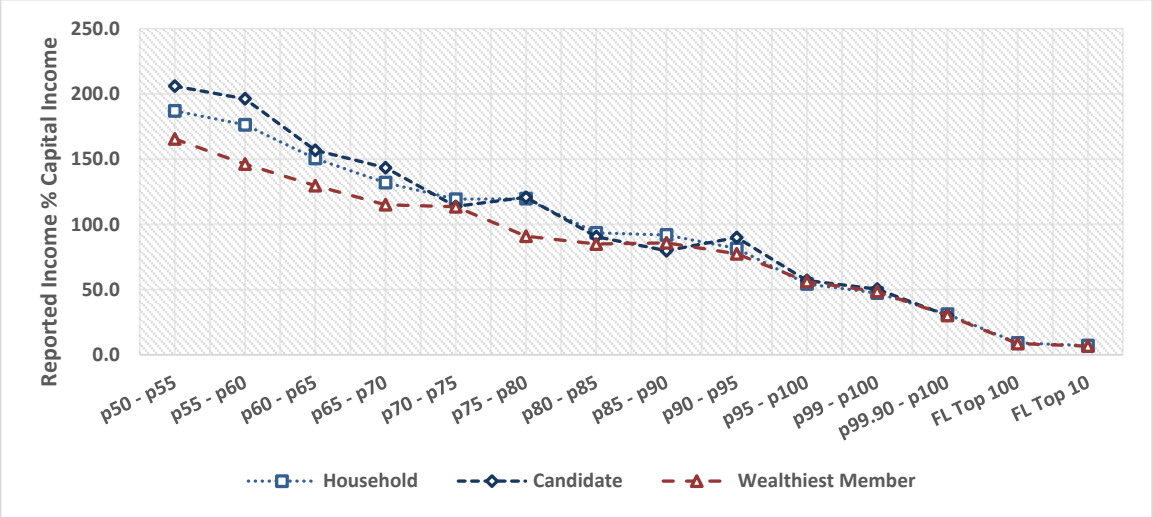


Figure 7.2 shows the reported income, Y_R, as a ratio of the capital income, taken to be 7.2% of wealth. As is evident from the figure, for the top 15% of families and individuals, the reported income is less than the return from their capital. The total income reported by the top 5% HHs and individuals is about half their capital income. The total income reported by the top 0.1% adds up to less than one-third of the returns from the capital owned by this group. If we assume the rate of return on capital of the wealthy groups to be 10% (a very plausible assumption), the total income reported by the top 5% HHs and individuals is about a third of their capital income. The total income reported by the top 0.1% adds up to only a fifth of the returns from the capital owned by this group. The FL families' total reported income is less than 10% of their capital income.

[49] On the share of national income, see FRED Economic Data (2022). The ILO (2018) estimates the wage share at 35.4% in 2013. For the national income wealth ratio, see Chancel and Piketty (2019), and Saez, E., Zucman, G. et al. (2022).

In other words, even after factoring in all types of income declared by the top 0.1% of families and individuals, their reported income amounts to just one-fifth of what they earn from capital alone. Since the reported capital income is less than the total income reported, this means that the capital income reported by this group is less than 20% of the returns from their capital; by implication, at least 80% of their capital income goes unreported in the ITRs. By similar logic, more than 95% of the capital income of families in the FL goes unreported! These numbers suggest that the share of the unreported capital income increases with wealth.

Furthermore, the difference between the total income reported by wealthy groups and their actual total income is more significant than what gets captured through the above figures. There are two reasons for this. First, we have quantified only the capital income of wealthy groups, but their actual total income also includes labour income, and hence is greater than the capital income. Second, the available evidence suggests that the rate of returns on capital owned by the wealthy groups is much higher than the 7.2% assumed by us.

At the same, it should be noted that except for the bottom 30%, for all other wealth groups, the ratio of capital income to wealth is less than the national average (Table 5.4). In principle, this can be due to several factors, such as, a large share of low-return assets in the wealth portfolio, underreporting of capital income, etc. What makes the wealthy groups' income-wealth ratios particularly interesting is the dominance of high-return assets in their portfolios, and there is minimal scope for them to hide their income illegally. Yet, the reported income-wealth ratios are the lowest for these groups.

Therefore, we must ask: What explains the vast proportions of the missing income at the top? The answer to this question lies in the types of assets owned by wealthy groups, the forms of capital income received, and the tax rule applicable to various kinds of capital income.

Tax Avoidance and Evasion

From Sections 2 and 3, we know that the wealthy groups in GE data hold most of their wealth as equity, non-agricultural land, and commercial properties. This class of assets enables owners to influence the split of the capital income between what is required to be reported and what can go legally unreported. To understand this, it is helpful to remember an essential consequence of the dominance of equity and commercial property in the asset portfolio of wealthy groups. It means that *capital gains*, i.e., the appreciation in the market value of the assets, is a dominant form of capital income for the wealthy groups. For accounting purposes, the capital gains from an asset are treated as "unrealised" unless exchanged or sold.

Under the Indian tax law, only realised capital gains from a sale or a transfer of an asset are taxable. Unrealised capital gains are thus neither taxable nor required to be reported in the ITRs. This means that as long as an investment is not sold out, there is no tax liability on unrealised gains, regardless of the quantum of appreciation in the asset's value. Even when the asset is finally sold or transferred to the next generation, the effective tax rate on the accumulated capital gains is much lower than the tax on other forms of realised income. Therefore, wealthy groups have a strong incentive to avoid realising capital gains to reduce their tax liability. They do so by staying invested in equity and commercial properties. Their motivation to stay invested is matched by their ability to do so. This is the primary reason realised capital gains are a tiny fraction of the capital income of the wealthy.

The tax rules applicable to the other forms of capital income are also fungible. Take, for instance, the case of direct income from equity assets in the form of dividends, i.e., profits distributed among a company's stockholders. Profits are taxed in the accounts of the company. Additionally, profits distributed as dividends are taxed in the form of Dividend Distribution Tax (DDT) or the tax liability

for the recipient. On the other hand, reinvested profits not only do not invite any additional tax but also boost the market value of company stocks, leading to hefty capital gains for the owners that remain unrealised and untaxed for the most part. Therefore, the reinvested profits lead to two benefits for the stockholders: they reduce the tax burden while propelling the value of equity capital. Eying these gains, wealthy groups want to reinvest most of their profits and keep the dividends payouts as low as possible. Such tax avoidance is a rational response by wealthy individuals who, in their capacity as CEOs, board members, or promotors of group companies, decide whether and how much of the profits will be distributed as dividends.

While such practices are an international phenomenon,[50] dividend pay-outs by Indian companies are meagre compared to most other countries - the dividend yields in developed countries other than the US range from 2% to 5%; for India, these rates are in the range of 1.1%-1.5%.[51] The average dividend yield of the top 100 private listed companies amounts to a dividend income of just 0.85% of the value of their equity assets; the dividend yield for companies controlled by the top 10 families on the FL is even smaller.

Interestingly, the Indian dividend yield rates are comparable to the US, as the incentives created by the tax rules are also similar. In both countries, the dividend income rates are lowest for the wealthiest individuals,[52] and deliberately suppressed dividend yields seem to be one of the leading reasons why the reported equity income is a small fraction of the total equity income of the wealthy groups. On top of this, in India, a part of the dividend income is received indirectly and thus retained in the accounts of entities like LLPs. Consequently, a significant fraction of equity income can go unreported in the income tax accounts.

This logic also applies to the income from non-agricultural land and commercial properties. Only rentals are required to be reported; capital gains are not. Moreover, our regression analysis suggests that the rental income is underreported. As rental transactions often do not create a verifiable record trail, they can be manipulated easily.[53]

As a result of the tax avoidance discussed above, only a tiny proportion of the capital income of the wealthy groups gets reported, while a more significant fraction of the returns from their capital goes missing from the income tax data. The ability of wealthy groups to choose their income levels extends to their labour income. A case in point is the labour income of the wealthiest Indian, who has kept his labour income fixed at ₹15 crore annually since 2008-09. The amount includes salary, perquisites, allowances, and commission from his entire business empire.[54]

Before concluding this section, we must highlight that the relatively high income reported by the

[50] For a discussion on tax avoidance and evasion in other contexts, see Chetty and Saez (2005), Kari and Karikallio (2007), Feldman and Slemrod (2007), Boissel and Matray (2022), and Lobel, Scot and Zúniga (2023).
[51] Dividend yields in developed countries range from 2% to 6%. In contrast, for India these rates are in the range of 1.1%-1.5%. See Monteiro et al (2020), and Bespoke Investment Group report (2019). Also see Kanojia and Singh (2013), and Labhane (2019).
[52] Saez and Zucman (2022) find the dividend yield for the top 120 wealthiest persons excluding the top 10 to be 1.4% in the USA.
[53] Perhaps, it is for this reason that rental yields are reported in India are low (2-4%) by international comparison. The average rental yields are in the range of 4% to 7% in many countries. See Demers and Eisfeldt (2022) for USA, Eichholtz et al. (2021) for France and Netherlands and Jorda et al. (2017) for evidence across several countries.
[54] This amount is just 1% of the family dividend income, which, in turn, is not even half a percent of their family wealth. See https://economictimes.indiatimes.com/news/company/corporate-trends/mukesh-ambani-keeps-salary-capped-at-rs-15-cr-for-12th-yr-in-a-row/articleshow/76533898.cms?

middle and low-wealth groups does not mean that these groups report all returns from their capital. Our results in Section 6 suggest that ipso facto reported income decreases with the share of agricultural land and commercial properties, across wealth groups. This finding and the available evidence[55] indicate that all groups misreport a part of their taxable labour income as tax-free agricultural income and underreport their rental income. On this count, the problem of tax evasion is acuter for the middle and above average wealth levels, given the dominance of farmland and property in asset portfolios.

8. Two Implications: Tax Regressivity and Under-estimated Inequality

This section discusses two implications of the decreasing income-wealth ratio and the income missing from the top. The first relates to the Indian tax regime, which is considered progressive in that the marginal tax rate (the rate applicable on each additional unit of income) increases with the reported income.[56] However, as we have seen above, income levels reported by individuals and HHs in their ITRs are less than their total income. The difference between the income declared in income returns, on the one hand, and the actual total income, on the other hand, can be huge, especially for the high-wealth groups. This calls for a re-examination of the tax regime to see if it is progressive with respect to the total income as opposed to the reported income typically used as a reference point. Moreover, as wealth is an essential determinant of capital income, labour income, and social status, it is meaningful to ask: How does the tax liability of different groups compare to their wealth? Below we explore these issues in brief. We examine the tax liabilities for the wealthiest members of the HHs — one member from each HH, including the families on the FL.[57]

First, consider the tax liability on the income reported as taxable, i.e., Y_T. Tax liability on Y_T is essentially the tax liability for Y_{Td}; recall, out of Y_T, only Y_{Td} is taxed, and we have the precise information on the latter. The tax liability for Y_{Td} is computed using the online calculator provided by the income tax department for the assessment year 2019-20.[58] The taxpayer is assumed to be a male below 65 with a "resident" status. Moreover, the source of taxed income is taken to be salary. These assumptions mean we have estimated each group's highest direct tax liability.

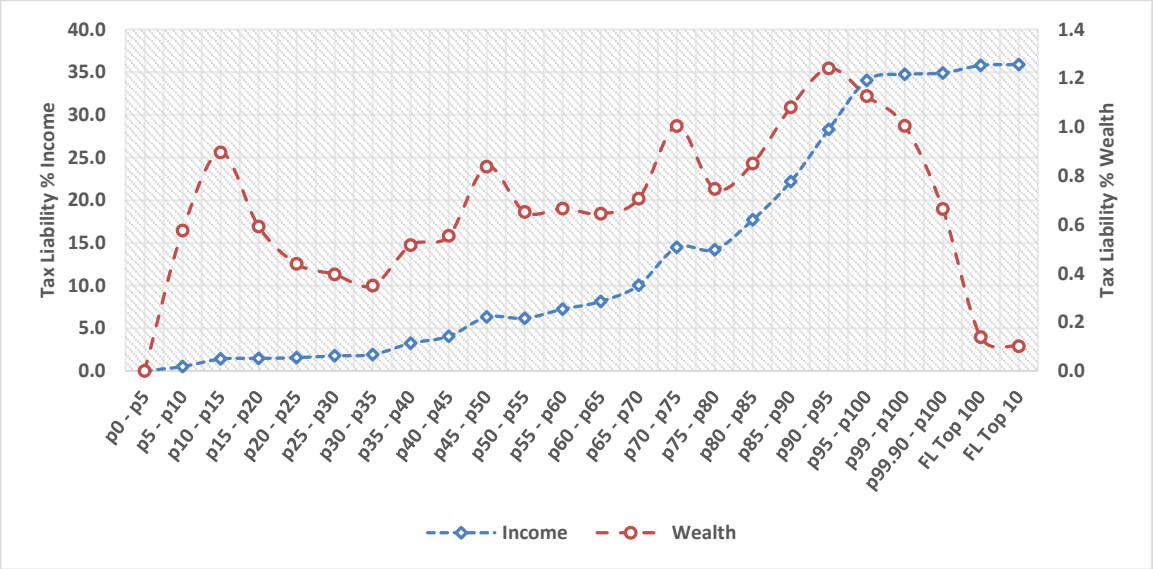
Figure 8.1 shows the estimated tax liability as a ratio of Y_T, and the tax liability as a ratio of wealth. As can be seen, the tax regime is progressive for the reported taxable income but not with respect to wealth. At the top wealth levels, the wealthier the taxpayer, the smaller the tax liability relative to wealth.

Next, consider the ratio of tax liability of different wealthy groups as a ratio of their capital income. We take the capital income to be 7.2% of the wealth — an underestimation of the capital income and hence the total income of the wealthy groups. As can be seen from Figure 8.1, there is an inverted U-shaped relationship between tax liability and capital income. Tax paid by the wealthy groups in the 95–99 percentiles amounts to less than one-fifth of their capital income, hence their total income. By a similar argument, the average tax liability of the top 0.1 centiles in the GE data is

[55] See Compliance Audit of Union Government Department of Revenue Direct Taxes by the Comptroller and Auditor General of India (2019) on use of agricultural land to exaggerate the reported exempt income.
[56] Under the old regime, the effective marginal tax rate including the surcharge increased up to ₹1 crore. Under the new regime, the effective rates increase up to ₹5 crore. See https://cleartax.in/s/income-tax-slabs
[57] The reason for choosing the wealthiest members is that in the ITR files, a tax unit is generally an individual. Only a few families file joint returns, and we do not have the data to estimate tax liability at the household level. Tax liability for the candidates can easily be computed, but this exercise cannot cover the individuals on FL; otherwise, we get results very similar to the ones presented here for the wealthiest members.
[58] Income tax calculations are done at https://www.incometaxindia.gov.in/pages/tools/income-tax-calculator.aspx. Computations are based on the tax rate applicable for an adult resident Indian.

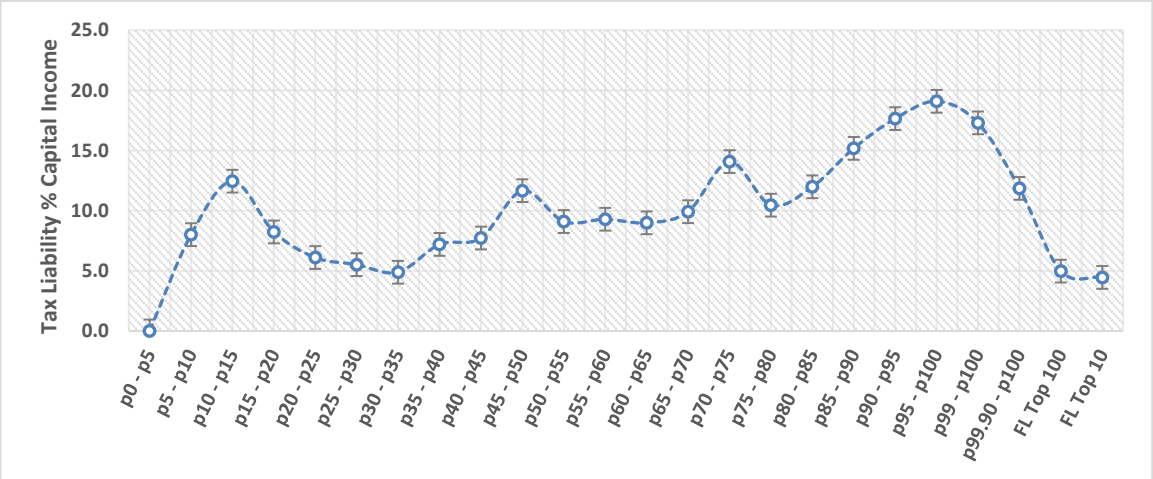
less than one-tenth of their income. This is smaller than the liability for individuals in the 80-85 percentiles. The tax liability for the super-wealthy Indians on the FL is not even 5% of their income!

Figure 8.1: Tax liability as a percentage of reported taxable income (Y_T) and wealth



For the tax regime to be progressive with reference to wealth, the reported taxable income by the wealthiest 0.1% has to go up by 100%. The value of the income reported by the FL has to be 12 times what is observed in the data. For the tax regime to be progressive with respect to income, as a rough estimate, the taxable income reported by the wealthiest 0.1% has to go up at least by 60%. The declared income value for the FL has to be at least four times what can be observed in the data.

Figure 8.2: Tax liability as a percentage of capital income (7.2% of wealth)



We should point out that the tax liability discussed here does not factor in all of the tax paid on the income received by an individual. The reason is that the income reported as taxable in ITRs, i.e., Y_T, does not include all types of individual income subjected to taxation. It leaves out the individual income not taxed in the hands of the recipient, such as the dividend income amounting to less than ₹10 lakhs. However, the ratios presented in Figure 8.2, along with the fact that not-taxed-hand dividend income is only a tiny fraction of the total income of the wealthiest groups, provide persuasive evidence to prove that the tax liability as a ratio of the total income decreases with wealth at the right tail of the distribution. Since the absolute value of the total income is expected to increase in wealth, the tax liability as a ratio of the total income decreases with the latter, making the effective tax regime regressive, at least at the top.

In addition, our results highlight two serious issues with the existing estimates of income inequality in India. As discussed in the introduction to this study, most existing estimates of income inequality rely on taxable income reported in the ITRs, i.e., Y_T, in terms of our notations.

We have shown that Y_T is less than the total income reported by the taxpayers in their ITRs, i.e., Y_R. We have also shown that the difference between Y_T and Y_R increases with wealth and also with income levels. As shown in Figure 4.3, for the rich and the super-rich, Y_T is quite small compared to the total income reported by these groups in their ITRs. For instance, for the top 10% of candidates in the GE data, the total reported income is 10–11% larger than their Y_T. For those on the FL, Y_R is 60-70% larger than their Y_T. Given these findings, it is clear that existing studies on income inequality have missed accounting for a substantial part of the income reported by the wealthy groups in their ITRs.

On top of this, as discussed in Section 6, the total reported income, Y_R, itself, is a small fraction of the total income of the rich and the super-rich groups. Our study points to a staggering difference between the income metrics that feed into existing studies on inequality and the actual income of the most prosperous Indians. According to our estimates, the total income reported by the wealthiest 5% of individuals and households is less than a third of their capital income. It is an even smaller fraction of their total income. The income reported by the top 0.1 centiles adds up to less than one-tenth of their actual total income. For the individuals and families in the FL, the total reported income is not even 5% of their total income. By capturing only a small fraction of the total income at the top, the existing studies have underestimated inequality in the country.[59]

The second issue pertains to using income tax data to identify the financially elite groups. The (income) richest groups are commonly also considered the wealthiest. Our study shows that the top income earners identified by the income tax data are not necessarily the wealthiest; neither are the wealthiest the highest income reporters. As shown in Section 3, most of the 100 (income) richest individuals do not feature among the wealthiest 100 individuals, and vice versa.

9. Conclusions and Remarks

Our analysis shows that the wealthier a household, the lesser the income reported by it relative to its wealth. Formally, the reported income-wealth ratio decreases with family wealth. This decreasing trend persists whether we consider the income reported to tax authorities as taxable or take the total income declared to them. According to our estimates, for the bottom 10% of families, the total reported income amounts to more than 188% of the family wealth. In contrast, for the top 0.1%, this ratio drops to about 2%. For the top 100 families in the FL, the total reported income is less than 0.6% of family wealth. The income-wealth ratios for individuals also exhibit very similar patterns.

We have shown that the tax rules account for much of the decreasing trend in the reported income-wealth ratio. Tax avoidance is an important factor behind relatively small values of the reported income-wealth ratio observed for the super-wealthy groups. On this account, a large share of the affluent groups' total income goes unreported in the income tax returns. Moreover, our empirical analysis suggests that people across the wealth spectrum underreport their rental income and misreport part of their taxable income by (presumably) disguising it as tax-free farm income.

The net result is that the total income reported by the wealthiest 5% of individuals is only a third of the returns from their capital. The total income reported by the top one-tenth of the top centile adds

[59] Since our focus is not inequality per-se, we refrain from estimating the magnitude of underestimation.

up to just about one-fifth of the returns from their capital. In other words, their total reported income amounts to less than 20% of their capital income; at least 80% of returns from their capital go unreported in the income tax files. For the families in the FL, more than 90% of returns from their capital do not figure in their reported income.

We find profession and gender-fixed effects. *Ceteris paribus*, women have reported lower incomes than men, and full-time agriculturists and politicians reported relatively low-income levels. Further, holding other factors constant, people with criminal records also reported relatively low incomes. In contrast, individuals exposed to greater media and civil society scrutiny have reported relatively high-income levels.

The missing income of the affluent groups underscores the case for re-assessing the progressivity of taxation. There is a case for considering the total income, and not just the reported income, for this purpose. Moreover, we have shown that the effective tax rate is progressive neither with reference to wealth nor the total income. At the top wealth levels, the wealthier an individual is, the smaller their relative tax liability tends to be. Going by the available evidence,[60] it will not be surprising if future research in other countries also leads to similar findings.

The missing income at the top has implications for the existing estimates of income inequality.[61] We have shown that the existing studies on the subject fail to capture a non-negligible fraction of the top total income levels; thereby, they most likely have underestimated income inequality.

In conclusion, we note some limitations of this study. Our regression analyses show that the income-wealth ratio is increasing in the vote share of political candidates. This means that our results have an upward bias. In other words, for any given level of wealth, except at the very top, the income reported by an average Indian is probably even smaller than what is seen in our estimates. We cannot verify whether this is the case. Moreover, our categorisation of individuals among different professions and educational qualifications is not precise due to a lack of detailed information. A study based on a more extensive database with more granular information might produce different results.

While this study focuses on income reporting behaviour of different wealth groups, discussing the implications of possible under-reporting of wealth by various wealth groups is pertinent. At the top wealth levels, this is not a serious issue. It is a concern only with respect to assets like land and buildings. Accordingly, we have revisited the income-wealth ratios presented in Section 5 by simply inflating the declared values of land and buildings by 25%. The income-wealth ratios still decrease sharply and continuously with wealth, but the fall is slightly less steep now. Future studies may lead to results somewhat different from ours.

Finally, we have highlighted only the adverse implications of underreporting the top income levels. Tax avoidance can have some positive implications, as well. In principle, it can lead to higher investment by the wealthy groups resulting in greater employment and growth for the entire economy. Macroeconomic analysis is beyond the scope of this paper.

[60] See ProPublica June 2021.

[61] See Ojha and Bhatt, (1964), Banerjee and Piketty (2005), Basole (2014), Ahmed and Bhattacharya (2017), Sinha et al. (2017), Assouad, L., Chancel, L., and Morgan, M. (2018), Chancel and Piketty (2019), Ghatak (2021), and Sahasranaman and Jensen (2021) for an overview of these findings.

References

Ahmed, M., & Bhattacharya, N. (1972). Size Distribution of Per Capita Personal Income in India: 1955-56, 1960-61 and 1963-64. *Economic and Political Weekly*, 1581-1588.

Aivazian, V., Booth, L., & Cleary, S. (2001). Do Firms in Emerging Markets Follow Different Dividend Policies from Those in the US? Evidence from Firms in Eight Emergent Countries. *Journal of Financial Research*, *26*(3), 175-188.

Anand, I. & A. Thampi 2016. "Recent Trends in Wealth Inequality in India," *Economic and Political Weekly*, 51, 59–67.

Asher, S., & Novosad, P. (2019). Rent-seeking and criminal politicians: Evidence from mining booms. *The Review of Economics and Statistics*, 1-44.

Assouad, L., Chancel, L., & Morgan, M. (2018, May). Extreme Inequality: Evidence from Brazil, India, the Middle East, and South Africa. In *AEA Papers and Proceedings* (Vol. 108, pp. 119-23).

Bach, Calvet, and Sodini (2018). "Rich Pickings? Risk, Return, and Skill in the Portfolios of the Wealthy," CEPR Discussion paper 11734, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2706207&download=yes

Banerjee, A., & Piketty, T. (2005). Top Indian incomes, 1922–2000. *The World Bank Economic Review*, *19*(1), 1-20.

Bardhan, P. (2017). The pattern of income distribution in India: A review. In B. S. Banerjee, *Poverty and income distribution in India* (pp. 69–112). Juggernaut.

Basole, A. (2014). Dynamics of income inequality in India: Insights from world top incomes database. *Economic and Political Weekly*, 14-17.

Besley, T., & Persson, T. (2014). Why do developing countries tax so little?. *Journal of Economic Perspectives*, *28*(4), 99-120.

Bespoke Investment Group (2019). Dividend Yields Around the World. Dividend Yields Around the World.

Blanchet, T., J. Fournier, and T. Piketty (2017). "Generalized Pareto Curves: Theory and Applications," WID. World Working Paper. 2017/3, 2017.

Bø, E. E., J. Slemrod, and T. O. Thoresen. (2015). "Taxes on the Internet: Deterrence Effects of Public Disclosure." *American Economic Journal: Economic Policy*, 7 (1): 36-62.

Boissel, C., & Matray, A. (2022). *Dividend taxes and the allocation of capital* (No. w30099). National Bureau of Economic Research.

Campbell, J. Y., Ramadorai, T., & Ranish, B. (2019). Do the rich get richer in the stock market? Evidence from India. *American Economic Review: Insights*, *1*(2), 225-40.

Comptroller and Auditor General (CAG) of India (2019). Compliance Audit of Union Government Department of Revenue Direct Taxes (Chapter V)

Cass, D., & Stiglitz, J. E. (1972). Risk aversion and wealth effects on portfolios with many assets. *The Review of Economic Studies*, *39*(3), 331-354.

CBGA (2016). Centre for Budget and Governance Accountability, Annual Report 2015-16. Chancel, Lucas, Piketty, Thomas., Saez, E., Zucman, G. et al. (2022) *World Inequality Report* 2022.

Chancel, L., & Piketty, T. (2019). Indian Income Inequality, 1922-2015: From British Raj to

Billionaire Raj?. *Review of Income and Wealth*, *65*, S33-S62.

Chetty, R., & Saez, E. (2005). Dividend taxes and corporate behavior: Evidence from the 2003 dividend tax cut. *The Quarterly Journal of Economics*, *120*(3), 791-833.

Curcuru, Stephanie, John Heaton, Deborah Lucas, Damien Moore (2010). Heterogeneity and Portfolio Choice: Theory and Evidence, Ch. 6, in *Handbook of Financial Econometrics: Tools and Techniques.* (Eds: Yacine Ait-Sahalia, Lars Peter-Hansen) North-Holland. Volume 1. Pp. 337-382.

Dal Bó, Ernesto, Frederico Finan, Olle Folke, Torsten Persson and Johanna Rickne (2017). "Who Becomes A Politician?". Quarterly Journal of Economics, 132(4), 1877–1914.

Datt, G., Ray, R., & Teh, C. (2022). Progressivity and redistributive effects of income taxes: evidence from India. *Empirical Economics*, *63*(1), 141-178.

Demers, A., & Eisfeldt, A. L. (2022). Total returns to single-family rentals. *Real Estate Economics*, *50*(1), 7-32.

Dhongde, S. (2007). Measuring the impact of growth and income distribution on poverty in India. *Journal of Income Distribution*, *16*(2), 25-48.

Djankov, S., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2010). Disclosure by politicians. *American Economic Journal: Applied Economics*, *2*(2), 179-209.

Dynan, K. E. (2009). Changing household financial opportunities and economic security. *Journal of Economic Perspectives*, *23*(4), 49-68.

Eggers, A., & Hainmueller, J. (2009). MPs for Sale? Returns to Office in Postwar British Politics. American Political Science Review, 103(4), 513-533.

Eichholtz, P., Korevaar, M., Lindenthal, T., & Tallec, R. (2021). The total return and risk to residential real estate. *The Review of Financial Studies*, *34*(8), 3608-3646.

Fagereng, A, L. Guiso, D. Malacrino, L. Pistaferri (2020). Heterogeneity and persistence in returns to wealth, *Econometrica*, Vol 88(1), pp 115-170.

Feldman, N.E. and Slemrod, J. (2007). Estimating tax noncompliance with evidence from unaudited tax returns. *The Economic Journal*, 117: 327-352.

Fisman, R., Schulz, F., & Vig, V. (2014). The private returns to public office. *Journal of Political Economy*, *122*(4), 806-862.

FRED Economic Data (2022). Share of Labour Compensation in GDP at Current National Prices for India. (Accessed on August 21, 2022).

Gagliarducci, S., Nannicini, T., & Naticchioni, P. (2008). Outside income and moral hazard: the elusive quest for good politicians, IZA Discussion Papers, No. 3295, Institute for the Study of Labor (IZA), Bonn.

Ghatak, M. (2022). India's Inequality Problem, *The India Forum*, July 2, 2021.

International Labour Organization (2018). India Wage Report. CH1211 Geneva 22, Switzerland.

Jordà, Ò., Knoll, K., Kuvshinov, D., Schularick, M., & Taylor, A. M. (2019). The rate of return on everything, 1870–2015. *The Quarterly Journal of Economics*, *134*(3), 1225-1298.

Kanojia, S., & Singh, B. (2013). Dividend and Payout Policy: An Empirical Analysis of Listed Indian Companies. *Available at SSRN 2345423*.

Kari, S., & Karikallio, H. (2007). Tax treatment of dividends and capital gains and the dividend decision under dual income tax. *International Tax and Public Finance*, *14*(4), 427-456.

Karmali, N. (2021). India's 100 Richest 2021: Collective Wealth Skyrockets To Record High Of $775 Billion. Forbes.com.

Kaymak, B., Leung, D., & Poschke, M. (2020). Accounting for Wealth Concentration in the US.

Kopczuk Wojciech and E Zwick (2020) Business Incomes at the Top. *Journal of Economic Perspectives*, 34 (4): 27-51.

Kumar, R. (2019). The evolution of wealth-income ratios in India 1860-2012. *Available at SSRN 3111846*.

Labhane, N. B. (2019). Dividend policy decisions in India: Standalone versus business group-affiliated firms. *Global Business Review*, *20*(1), 133-150.

Lenz, G., and Lim, K., Getting Rich(er) in Office? Corruption and Wealth Accumulation in Congress (2009). In APSA 2009 Toronto Meeting Paper.

Libman, A., Schultz, A., & Graeber, T. (2016). Tax return as a political statement. *Review of Law & Economics*, *12*(2), 377-445.

Lobel, F., T. Scot and P. Zúniga (2023). Corporate Taxation and Evasion Responses: Evidence from a Minimum Tax in Honduras, *American Economic Journal: Economic Policy,* forthcoming.

Lundberg, J., & Waldenström, D. (2018). Wealth inequality in Sweden: What can we learn from capitalized income tax data?. *Review of Income and Wealth*, *64*(3), 517-541.

Monteiro, A., Sebastião, H., & Silva, N. (2020). International evidence on stock returns and dividend growth predictability using dividend yields. *Revista Contabilidade & Finanças*, *31*, 473-489.

Nayak, P. B., & Paul, S. (1989). Personal income tax in India: alternative structures and their redistributive effects. *Economic and Political Weekly*, 2779-2783.

OECD. Publishing. (2018). *The Role and Design of Net Wealth Taxes in the OECD*. Chapter 2. OECD.

Ojha, P. D., & Bhatt, V. V. (1964). Pattern of income distribution in an underdeveloped economy: A case study of India. *The American Economic Review*, *54*(5), 711-720.

Paiella, M., & Guiso, L. (2004). Risk aversion, wealth and background risk. In *2004 Meeting Papers* (No. 525). Society for Economic Dynamics.

Piketty, T. (2014). Capital in the twenty-first century. In *Capital in the twenty-first century*. Harvard University Press.

Piketty, T., & Zucman, G. (2014). Capital is back: Wealth-income ratios in rich countries 1700–2010. *The Quarterly Journal of Economics*, *129*(3), 1255-1310.

Piketty, T., & Qian, N. (2009). Income inequality and progressive income taxation in China and India, 1986-2015. *American Economic Journal: Applied Economics*, *1*(2), 53-63.

Saez, E., & Zucman, G. (2022). *Top Wealth in America: A Reexamination*. Working Paper: 30396, National Bureau of Economic Research.

Sarkar, S., & Mehta, B. S. (2010). Income inequality in India: pre-and post-reform periods. *Economic and Political Weekly*, 45-55.

Siblis Research. https://siblisresearch.com/data/global-dividend-yields/.

Singh, R. (2012). Inefficiency and Abuse of Compulsory Land Acquisition. *Economic and Political Weekly*, 14-17. 46-53.

Singh, R. (2013). Litigation over Eminent Domain Compensation Authors, *International Growth Centre*, London School of Economics, WP 2013/9.

Sinha, R., Pearson, P., Kadekodi, G., & Gregory, M. (2017). *Income distribution, growth and basic needs in India*. Routledge.

Sahasranaman, A., & Jensen, H. J. (2021). Dynamics of reallocation within India's income distribution. *Indian Economic Review*, *56*(1), 1-23.

Szakonyi, D. (2020). Indecent Disclosures: Anticorruption Reforms and Political Selection. *American Journal of Political Science*.

Wolff, Edward N. (2020). *The Wealth Project.* Institute for New Economic Thinking.