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

Ram Singh

January 2023

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

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

ABSTRACT

The income reporting behaviour of different wealth groups is a critical public finance issue that has remained under-researched in the Indian and international contexts. We model and estimate the relationship between wealth and reported income for individuals and families across different wealth groups. We use a new dataset based on affidavits filed by election contestants, the Forbes List of billionaires, and the statistics published by the Indian Tax Department. We show that the wealthier the individual or the family, the lesser is reported income relative to 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 in the data amounts to more than 188% of their wealth; in contrast, the wealthiest 5% [respectively 0.1%] of families reported incomes that were just 4% [respectively 2%] of their wealth. The total income reported by the wealthiest Forbes list families is less than 0.6% 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, and at least 80% of their capital income goes unreported in the income tax returns. For the Forbes-listed 100 families, more than 90% of the capital returns do not figure in their reported incomes! The income-wealth ratios for affluent individuals exhibit very similar patterns. We discuss the processes responsible for the "missing" income of the wealthy groups and show that this "missing" income leads to an underestimation of income inequality. Furthermore, it reduces the tax liability of the wealthiest percentile group to a mere 1% of their wealth. The tax liability of the wealthiest 0.1 centiles and the Forbes-listed families is less than one-tenth of their capital income. Tax paid by these groups relative to their wealth is smaller than the relative tax liability for middle-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. Our analysis suggests that recent measures taken by the Indian central government against illicit income and wealth hoarding have delivered the intended results.

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_{Cg}:   Capital gains
Y_{Eq}:   Equity income
Y_P: Property (rental) 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 reported by taxpayers: the higher the reported income, the larger the tax revenue, and vice versa. The taxman, faced with increasingly ambitious targets of tax collection, thus wants to know what fraction of the total income is being reported by the citizens. The issue is also of concern to other government agencies whose ability to sponsor and execute welfare schemes and programmes depends on tax revenue, which is a dominant component of government finances. The income reporting behaviour of different wealth groups 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, these issues 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, while movie stars such as Akshay Kumar, Amitabh Bachchan, and Salman Khan are among the top income taxpayers, very few of the wealthiest Indians figure on the list.[1]

Yet, the income reporting behaviour of different wealth groups has remained under-researched in Indian and other contexts even though the relationship between national wealth and income has been extensively examined for many countries.[2] There is very little empirical research on the relationship between wealth and reported income at the individual or household levels.[3] The most plausible reason for the absence of empirical microeconomic research on the subject is the lack of data required for this purpose. Data sources that provide information on both individual wealth and income levels are hard to come by.

In this paper, we compile and use a new dataset to examine incomes reported to tax authorities by different wealth groups. This dataset is based on affidavits submitted by the contestants of elections to the Lok Sabha, the house of representatives in Parliament of India. These affidavits are the only simultaneous source of information on the wealth and income levels of a large number of Indians. These documents provide information on the wealth and income of 7,596 households (HH) and their adult members. To our knowledge, this study is the first to use election affidavit data to examine the relationship between wealth and the income reported by individuals and households from across wealth groups. We supplement this source of information with the Forbes' List (FL) of billionaires, and statistics published by the Government of India's Central Board of Direct Taxes (CBDT). These data sources enable us to cover India's entire range of wealth and income distributions.

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 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 virtue of its structure, this dataset is reasonably representative of the Indian context in terms of the regional and rural-urban distribution of the population. As we will discuss 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.

---

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

[2] 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).

[3] 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).

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,[4] and male dominance over family income and wealth, among others. As one would expect, the share of the financial assets such as company stocks and firm ownership increases with wealth levels. The patterns emerging from the affidavit data are also consistent with what can be gleaned from the other independent sources. For instance, the asset portfolios of the wealthiest individuals in the data resemble the portfolios held by the most affluent non-politician Indians on 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).

It is worth emphasising that we do not use GE data to estimate wealth distribution or income distribution. We use it only to examine the relationship between wealth and corresponding income levels reported to tax authorities. From this viewpoint, besides being the only simultaneous source of information on incomes and wealth, the GE data pass the 'smell' tests on several counts in addition to possessing the properties discussed above. For instance, as will be discussed in Section 4, the downward trend in the income-wealth ratio emerging from the data nests well within 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 gleanable about the leading wealth groups in other countries.[5]

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.

As such, the affidavits are an invaluable source of information as far as information on family wealth and its components are concerned. However, these documents are only partially informative as far as income is concerned. Affidavits provide information only on what can be described as the net "taxed in hand 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 reported as "taxable income" in the candidates' income tax returns (ITRs). Additionally, they contain no information on the income reported by the candidates and their families 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 reported to tax authorities; we then use this relationship to estimate the latter from the former. In addition, we use CBDT data to estimate 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 exempt income — such as agricultural income, dividend income, long-term capital gains, etc. — we use the affidavit data on asset ownership along with various sources on the rate of returns for different classes of assets. Further details on this are provided in Sections 4 and 5.

We show that the reported income as a proportion of wealth decreases with the wealth. On average, the wealthier a household is, the smaller its income is relative to its wealth. This decreasing trend

---

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

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 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 and fall precipitously 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 on 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, the income reported in ITRs of the wealthy and super-wealthy groups is a small fraction of the returns from their wealth.

As the dynamics of capital income modelled and empirically examined by us are similar across market economies, our findings should be of interest and relevance beyond the Indian context. Specifically, our study contributes to three kinds of literature. First, by examining the relationship between wealth and reported income for individuals and HHs from across various 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 of 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, the affluent Indians, much like their counterparts in other market economies, can choose what fraction of their capital income gets transferred to their individual accounts and in what forms. To minimise tax liability, they transfer only a tiny fraction of the returns from capital to their personal accounts. We show that only a tiny proportion of their capital income gets accounted for in their tax reports a more significant fraction of these wealthy groups' capital income goes missing from tax data and therefore remains untaxed.

Third, our findings underscore the case for a wholesome assessment of the income tax regimes. The standard approach considers a regime to be progressive if the applicable marginal tax rates increase with the reported income. Accordingly, 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, conclude that the Indian income tax regime is progressive. However, when the affluent can choose how much of their income gets taxed, 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 regime is even more regressive with respect to wealth. We show that at top wealth levels, the wealthier a taxpayer, the smaller is the tax paid relative to wealth. For the wealthiest centile, the tax liability amounts to about 1% of their wealth. For the wealthiest 0.1% individuals, the tax liability amounts to approximately 0.7% of the wealth. Super-wealthy Indians on the FL face effective 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.

Our results also make relevant contributions 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).[6] As will be discussed below, the income tax data used by these studies miss a substantial share of the opulent group's income and thus underestimate inequality.

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). 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), Libman, Schultz and Graeber (2016) and Szakonyi (2022), our results suggest that people exposed to media and civil society scrutiny have a stronger incentive to report their incomes truthfully. 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.

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 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 5 presents the regression results on the determinants of income-wealth ratios for individuals and households. Section 6 examines the proportion of total individual income that goes missing from the reports filed to the tax authorities. It also discusses the mechanisms that facilitate partial income reporting by opulent groups. Section 7 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 8, we offer concluding remarks. The details of methodology used to estimate the various kinds of income reported to the tax authorities are in the Appendix.

2. Wealth and Income: A conceptual framework

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

---

[6] 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), and Datt, Ray and Teh (2022), among others.

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 we have not included the "realised" capital gains as part of capital income. The neglect of unrealised capital gains in the model 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 $A = A_1 + A_2 + \dots + A_n$. 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 $Y_{KD} = y_{1D} + y_{2D} + \dots + y_{nD}$. The total indirect income is $Y_{KI} = y_{1I} + y_{2I} + \dots + y_{nI}$. 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 $Y_K = \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} r_I A_i$. 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.[7]

Without loss of generality, let us assume that the riskiness of the asset increases with index $i = 1, \ldots, n$; 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.

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 $A_1, A_2, \ldots, A_n$. 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 $\hat{A}_1, \hat{A}_2, \ldots, \hat{A}_n$, with $\hat{s}_i = \frac{\hat{A}_i}{\hat{A}}$. We consider the asset allocation $(\hat{A}_1, \hat{A}_2, \ldots, \hat{A}_n)$ to be riskier relative to the allocation $(A_1, A_2, \ldots, A_n)$ if the following holds: For all $k \leq 1, \ldots, n$

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

The above inequality is strict for at least some k. Simply put, portfolio $(\hat{A}_1, \hat{A}_2, \ldots, \hat{A}_n)$ is riskier than $(A_1, A_2, \ldots, A_n)$, if the former assigns a larger share of investment to the risky assets. Several studies, including this one, show that the wealthier an individual is, the larger is the share of risky assets in their portfolio and vice versa.[8] 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.[9] This assumption and that r_i is increasing in i, 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.

---

[7] See, for instance, Campbell et al. (2019), who show that larger account holders diversify more effectively and thereby earn higher average than holders of smaller accounts.

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

[9] 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 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).

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.[10] 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 $Y_K'(W) > 0$ and $Y_K''(W) > 0$ 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 each of 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%).[11] 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 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:

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

i.e., $\widehat{W} > W$ implies $\frac{\hat{Y}_{KD}}{\hat{A}} < \frac{Y_{KD}}{A}$. 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

---

[10] 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).

[11] Recently, interest rates have been low but fixed-term interest rates are still 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.

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.[12] Thus we have the following result.

Proposition 1: The average rate of direct income, $\frac{Y_{KD}(W)}{W}$, decreases with wealth. The average rate of indirect income, $\frac{Y_{KI}(W)}{W}$, increases with wealth.

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.[13] 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,

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

That is, the ratio $\frac{Y_L(W)}{W}$ is decreasing in W. As both components of personal income (capital and labour income) increase with wealth, the total personal income, Y, is an increasing function of W. The question now is: What can we say about the income-wealth ratio $\frac{Y}{W}$? Given the above, Y_K is an increasing and convex function of W but Y_L is an increasing and concave function of W. Therefore, how the ratio $\frac{Y}{W}$ varies with W cannot be predicted *a priori*.

At the top wealth levels, however, the $\frac{Y}{W}$ ratio is expected to be increasing in W. For the ultra-wealthy groups, the share of Y_L is relatively small; Y_K accounts for most of their total income. That is, $Y \approx Y_K$. As the latter increases more than proportionately with wealth, total income for the wealthy is expected to follow suit. Consequently, their income-wealth ratio is also likely to increase with wealth.

Moreover, one can predict how the sum of labour income and the direct capital income, i.e., $[Y_L + Y_{KD}]$ 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,

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

---

[12] Several studies show that the unrealised capital gains (indirect returns) are a significant component of the increasing returns on the wealth. For details, see Piketty (2014, Chapter 12), Saez and Zucman (2016), and Kaymak, Leung, Poschke (2020).

[13] 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

Since $\frac{Y_L(W)}{W}$ and $\frac{Y_{KD}(W)}{W}$ are both decreasing in W, we have the following result.

Proposition 2: The direct personal income as a ratio of wealth decreases as wealth level increases, i.e., $\frac{\partial}{\partial W}\left(\frac{Y_{PID}(W)}{W}\right) < 0$.

Proposition 2 provides the basis for framing hypotheses for our empirical analysis. 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. However, as is shown in Section 4, the realised capital gains are only a tiny fraction of the direct personal income. Therefore, we expect the prediction in Proposition 2 to hold both with and without factoring in realised capital gains as a part of the direct capital income.

Finally, consider the effect of portfolio churning while holding fixed wealth levels. Holding the wealth constant, we do not expect portfolio choice to exert any significant effects on labour income. However, the rate of direct and indirect returns varies across assets. Therefore, different allocations of a given amount of wealth across asset classes will result in different values of direct and indirect returns. In terms of notations used, for any given wealth W, different asset portfolios will lead to varying values of Y_{KD}, Y_{KI}, and hence Y_K in general. Specifically, we make the following claim.

Proposition 3 For any given level of wealth, the ratio of direct personal income to wealth depends on the shares of various assets in the portfolio.

While concluding this section, it bears emphasising that our definition of direct capital income includes direct returns from all assets constituting the wealth. However, the above predictions will hold even if some of assets are dropped from the definition of W and the corresponding income is excluded from $Y_{KD.}$. It is crucial to keep the set of assets constant for W and $Y_{KD.}$.

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"*,[14] 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 records for a small sample of randomly selected affidavits directly taken from the Election Commission of India (ECI) website.[15]

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

---

[14] https://www.myneta.info
[15] https://affidavit.eci.gov.in/

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

In addition, 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. 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 with missing income pertain to candidates whose families earn less than the ₹2.5 lakh threshold for filing ITRs. We dropped all households (HHs) with missing income from our analysis.

We were able to gather information on HH income for the remaining 7,596 candidates and their households. The General Elections of 2014 and 2019 each account for roughly half of these observations. The HHs from the two GEs have very similar demographic attributes such as genders, castes, educational qualifications, and professions.
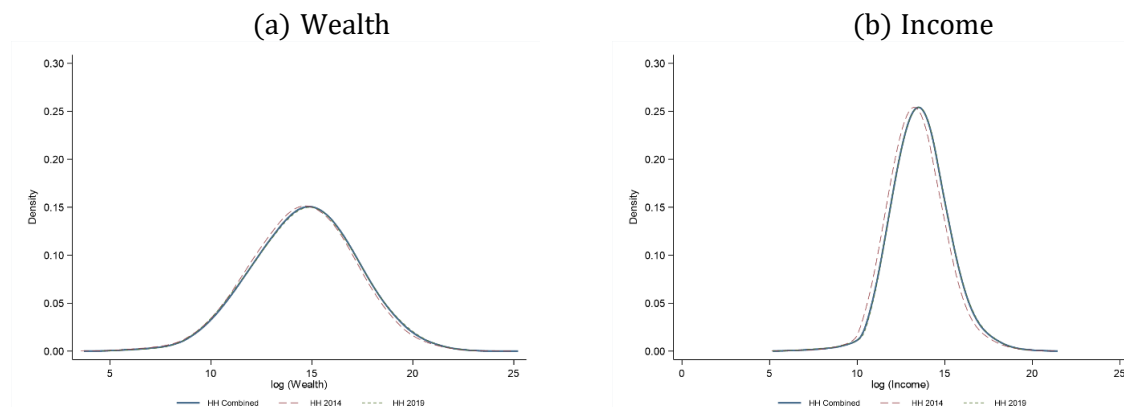
Occasionally, we combine observations on wealth and income from GE 2014 and GE 2019 to present an overall picture. As they pertain to two different points in time, the GDP deflator is used to convert wealth and income levels to March 2019 prices.[16] Our main results, however, hold without merging

---

[16] 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).

the data from the two GEs.

Sources have been mentioned for all the tables and figures. Tables and figures without reference cited have been generated using our own computations on the relevant data.

Figure 3.1: 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 returned income as disclosed in the election affidavit.

As distributions of wealth and income are both highly skewed towards large values, the density plots are not very revealing. So, in Figure 3.1 we present density plots for the distribution of the log of wealth for GE 2014 and GE 2019 separately as well as combined.[17] The average income and wealth are higher for 2019 than for 2014.

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). As can be seen in the Appendix, income in the GE data is less concentrated than wealth — a pattern that has been observed in international contexts too.[18] All Appendices are available online.

Figure 3.2 presents comparison of the wealth distribution in our sample via-a-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.[19] 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
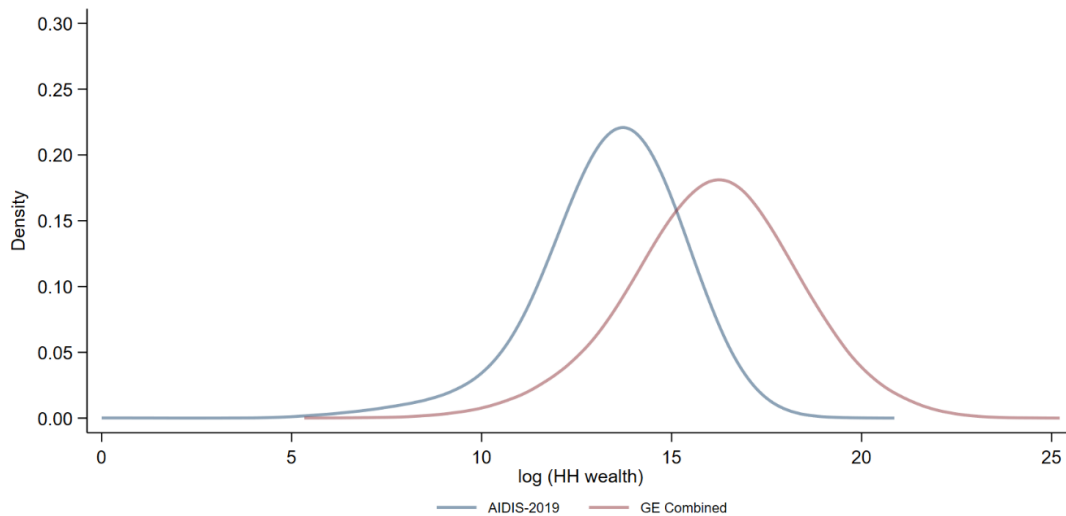
---

[17] 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.

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

[19] 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).

to 37 in India versus 1 to 4 in the case of the US[20], and possibly even higher in China.

Figure 3.2: Household Wealth distribution GE vs AIDIS

Comparing elected politicians in our dataset with their counterparts in other countries paints a 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.[21]

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.

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.

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.[22] In recent years, the wealth held by billionaires in FL has come to account for an increasing fraction of the

---

[20] Calculations are based on the data available at the online forum Open Secretes for the US, and the Harun list for China.
[21] 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.
[22] 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.
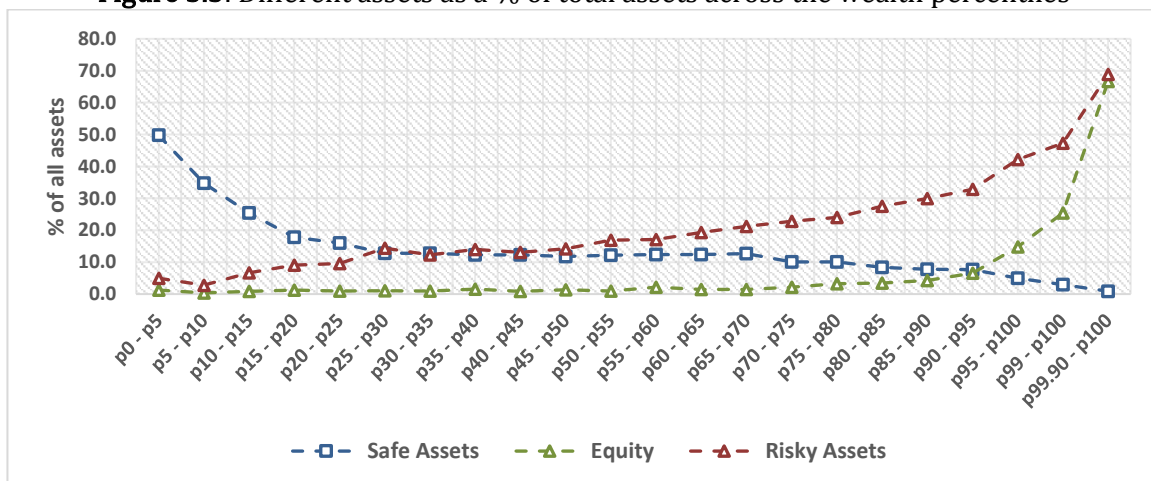
national income.

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	-3,951,493	641,146	56.2	5.1
p5 - p10	380	290,253	351,043	39.8	2.8
p10 - p15	379	709,479	448,498	30.6	6.7
p15 - p20	381	1,314,795	488,736	25.2	9.0
p20 - p25	380	2,054,144	520,649	22.5	9.5
p25 - p30	379	3,002,482	542,202	19.9	14.3
p30 - p35	381	4,018,420	573,599	17.4	12.3
p35 - p40	380	5,278,640	733,610	17.2	13.8
p40 - p45	380	6,779,390	781,518	15.7	13.2
p45 - p50	380	8,668,396	897,422	15.9	14.2
p50 - p55	380	11,024,009	941,918	15.6	16.9
p55 - p60	379	13,962,219	1,172,538	15.4	17.1
p60 - p65	381	18,136,096	1,275,765	15.0	19.3
p65 - p70	380	23,960,570	1,454,769	15.6	21.2
p70 - p75	380	32,221,960	1,782,296	12.6	22.7
p75 - p80	380	42,845,496	2,505,900	12.4	24.0
p80 - p85	380	60,533,676	2,657,566	10.1	27.5
p85 - p90	380	93,946,888	4,270,143	9.3	29.9
p90 - p95	380	183,624,992	7,824,278	9.0	32.8
p95 - p100	380	1,131,648,128	35,099,180	6.0	42.2
p99 - p100	76	3,576,561,408	98,649,136	3.7	47.3
p99.90 - p100	8	18,062,749,696	351,848,448	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.

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".[23] 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:

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

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.[24] 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 the Appendix contain further details on this approach.

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

3.2 Is GE data representative?

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;

---

[23] 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.

[24] The list includes all listed companies and most unlisted public companies, and private companies of all ownership groups.

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.[25]

Moreover, it is worth emphasising that we do not use the GE 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 figure for families on the FL.[26] 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 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.[27] 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

---

[25] 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.

[26] Top income levels reported in GE (up to ₹194 crores). See Section 3.

[27] 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.

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.[28] 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 relationship between reported income and wealth is different for the politicians qua politicians, i.e., the reported income-wealth is different 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. However, our results presented in Section 5 are quite counterintuitive.

4. The Income-Wealth Ratios: *The Cobra curves*

In this section, we present our empirical findings on the relationship among the various types of income reported by taxpayers and their wealth. Our focus is the total reported income and the income reported as taxable. However, to present the overall picture we also consider the net taxed-in-hand income and the total personal income. From Section 2 we expect the ratio of personal income to wealth to be decreasing in the latter. However, the ratio of the total 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.

For our empirical analysis, we include all assets including consumer durables and jewellery as part of wealth. We decided to include durables and jewellery for two reasons. First, durables increase productivity of labour and therefore have an indirect bearing on income. Second, ownership of gold and jewellery can not only help ease credit constraints but actually help earn direct capital income.[29] Therefore, these assets are relevant for our study. Additionally, for our empirical analysis, it is difficult to get disaggregated information on the distribution of different types of assets within a family, especially for the families on the FL. To keep the definition of wealth consistent across individuals and HHs, we define wealth as the value of total assets *minus* the total liabilities. Moreover, our definition of wealth is consistent with several empirical works based on the affidavits filed by election contestants.[30] Given the small share of consumer durables and jewellery, there exclusion from the definition of wealth is not expected to make much difference to the relationship between income and wealth, especially for the affluent groups for who share of these assets is negligible.

---

[28] For a review of this literature see Libman, Schultz and Graeber (2016) and Szakonyi (2022).
[29] The Gold Monetization Scheme 2015 enables gold owners to earn an tax-free interest income up to 2.5% of the value by depositing gold with the government.
[30] See literature cited in Fisman et al (2014) and Bhavnani (2012).

As to the income, all Indians with taxable income 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 reported by a taxpayer as taxable in their hands, i.e., the income on which the recipient themselves are liable to pay tax. It includes salary and other forms of labour income, professional income, interest income, rentals, capital gains, as well as capital income from businesses and other sources not included in the exempt category. 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 the taxpayer actually pays 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.[31]

The Y_T and Y_{Td} do not cover several types of 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: $Y_R = Y_{Ex} + Y_T = Y_{Ex} + Y_{Td} + Y_{Dd}$.

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 income and all of 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.[32]

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

Our estimation methodology follows the following order: We start with Y_{Td} and use it to estimate Y_T, which along with estimates of Y_{Ex}, is used to estimate Y_R and Y_{PID}.

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 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

---

[31] 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.
[32] Y_R also includes realised capital gains, which are a negligible share of the total reported income and hence of Y_{PID}.
[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).

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. We have estimated the total reported income, Y_R, by supplementing the Y_T of the concerned unit (individual or household) with the estimated value of their Y_{Ex}. To estimate the imputed rent based on the value of the residential property, we use average rental rates. The methodological details of the estimation process have been discussed in Appendix I.

Here, we present some plots showing the relationship between the above-discussed four versions of income reported by different wealth groups. Plots below show the entire range of the estimated Y_R. For other versions of income, we depict only the point estimates. See Figure 4.1.
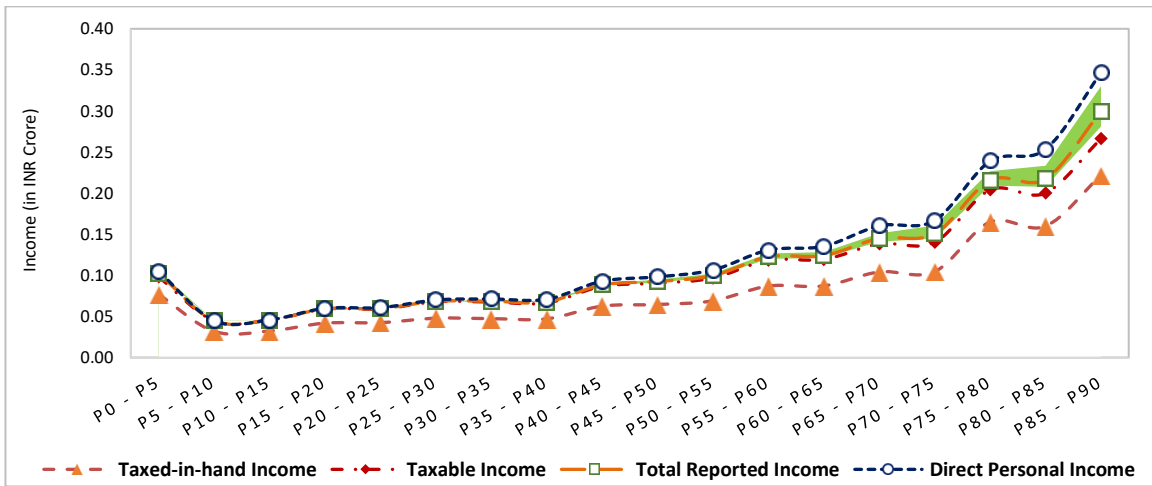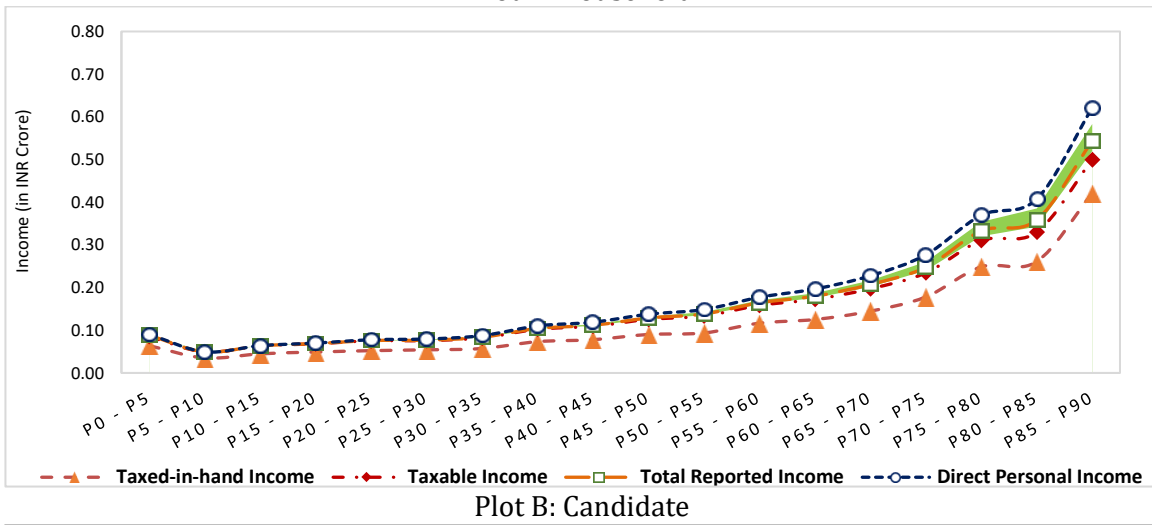
The income plots appear to be flat for the bottom 99%, even though in reality they are not. The apparent flatness is due to a relatively massive increase in the income levels for the top wealth groups — the top 1% in the GE data and the FL. For the ease of illustrating increasing trends in income at all wealth levels, below, we present plots for the $p_5 - p_{10}$ to $p_{85} - p_{90}$ of household and candidates. Income wealth relationship for the wealthiest members shows similar patterns.

Figure 4.1: Average Y_{Td}, Y_T, Y_R and Y_{PID} income reported by different wealth groups (households)

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. However, as can be seen from the Figure 4.2, the average income is increases with wealth. Therefore, for the low and middle-wealth groups Y_{Td} is expected to be significantly smaller than their Y_T. For the wealthy, the two should be approximately equal.

Figure 4.2: Average Income $Y_{Td}, Y_T,\ Y_R$ and Y_{PID} reported across wealth groups
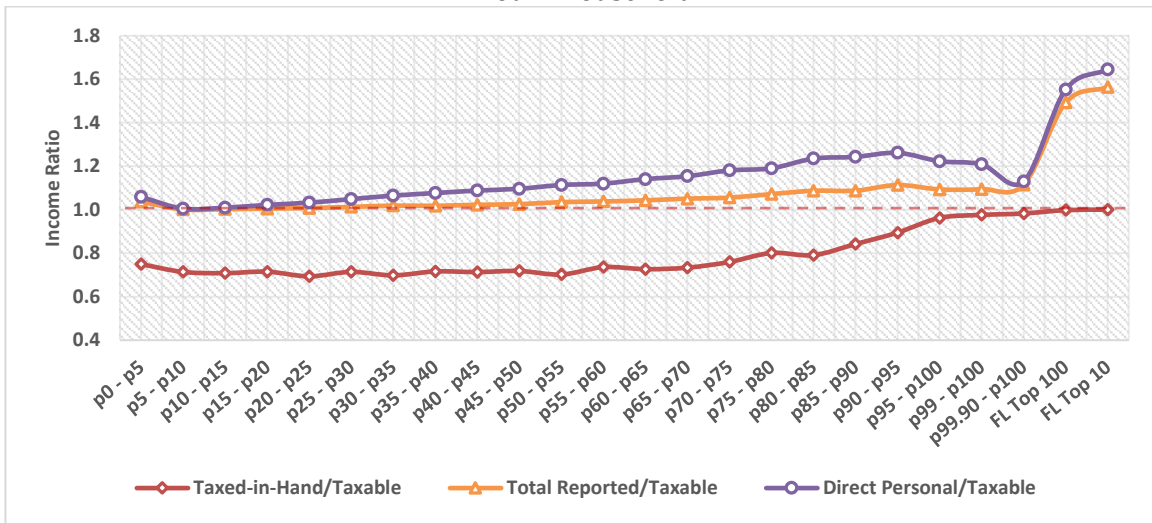
Plot A: Household

Plot B: Candidate

Plots in Figure 4.3 show the relationship among the various income categories for individuals and HHs from various wealth groups. Expectedly, the difference between the Y_T and Y_{Td} decreases with wealth, but 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 the total income reported by them, i.e., Y_R, are significantly higher than their taxable income, 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.
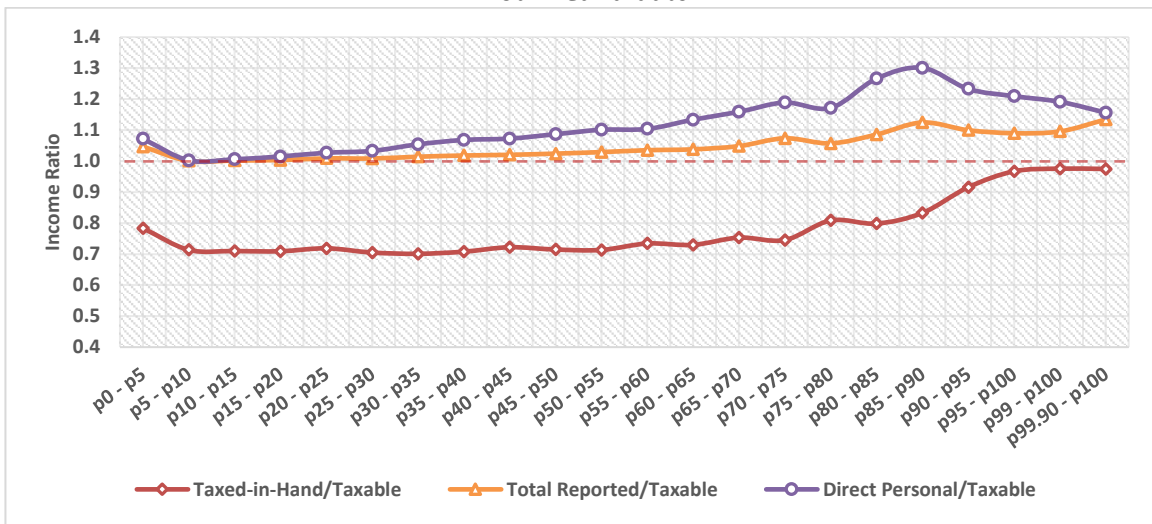
By contrast, the relationships between Y_{PID} including the imputed rent and the Y_T do not follow consistent patterns. As wealth increases, initially, the rent increases relative to Y_R. It reaches its peak for the wealth groups at p90-p95; thereafter, it decreases continuously. The initial increase 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. As is shown in Plots B and C in Figure 4.3, a similar relationship holds between Y_{PID} and Y_R for individuals.

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 are different when we compare the
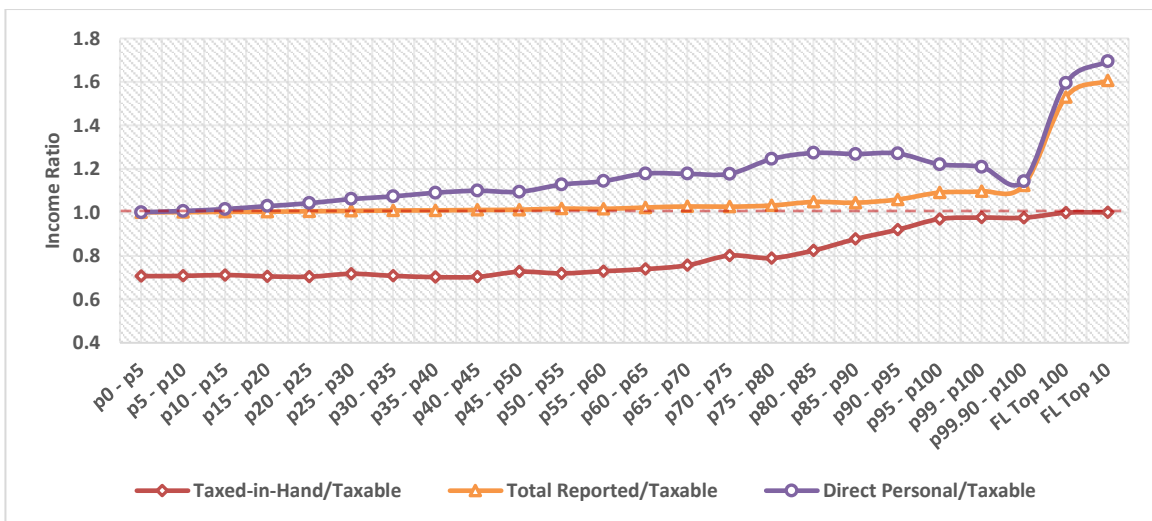
Figure 4.3: Ratio of Y_{Td}/Y_T, Y_R/Y_T and Y_{PID}/Y_T reported by different wealth groups

Plot A: Household

Plot B: Candidate

Plot C: Wealthiest member

reported income with the corresponding wealth. Below we show that the income-wealth ratio (i.e.,

the HH income expressed as a percentage of the HH wealth) decreases with family wealth. This decreasing trend persists for all categories of income: Y_{Td}, Y_T, Y_R and Y_{PID}. The income-wealth ratios for individuals exhibit similar patterns as well. Wealthier an individual, the lesser is the reported income relative to 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.
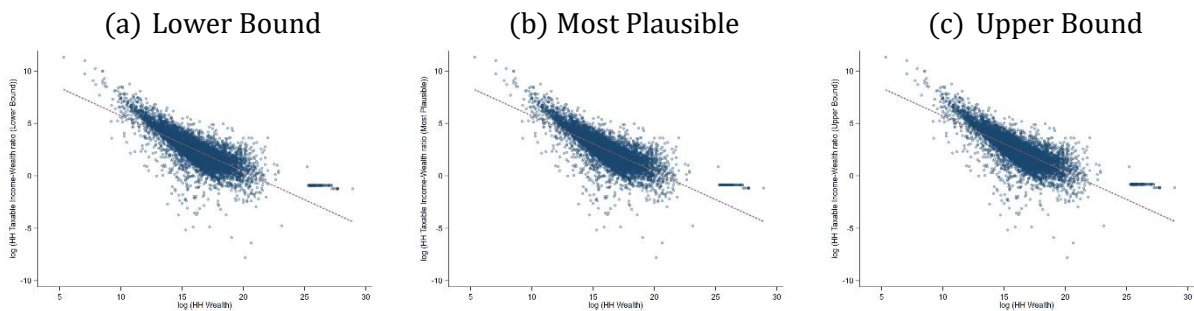
Given the similarity of the ratios emerging from the two approaches, to optimise space, all tables in this section other than Table 4.1 are generated using Approach 1. To provide a sense of the numbers produced by the two approaches, all plots presented hereafter are based on Approach 2.

In the rest of this section, we show that the income-wealth ratios are decreasing in wealth for all three versions of the reported income, regardless of the method used to compute the ratio. The downward trend is very pronounced at the super-high and ultra-large wealth levels, even if we use the most generous estimates of the income reported by wealthy groups.

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

Now, we present the reported taxable income, Y_T, as a proportion of the wealth. Figure 4.4 show the log-log scatter plots of Y_T/W ratios reported by households 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.

Figure 4.4: Household income (Y_T) 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.

Group levels income-wealth ratios are presented in Table 4.1. 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. 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 to preserve the scale and 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.

Table 4.1 shows the Y_T/W ratios for individuals and households across the wealth spectrum. The ratios in columns 2 and 3 of the table 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 below, the income-wealth ratios for the wealthiest 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 of HHs The Appendix has all the details.

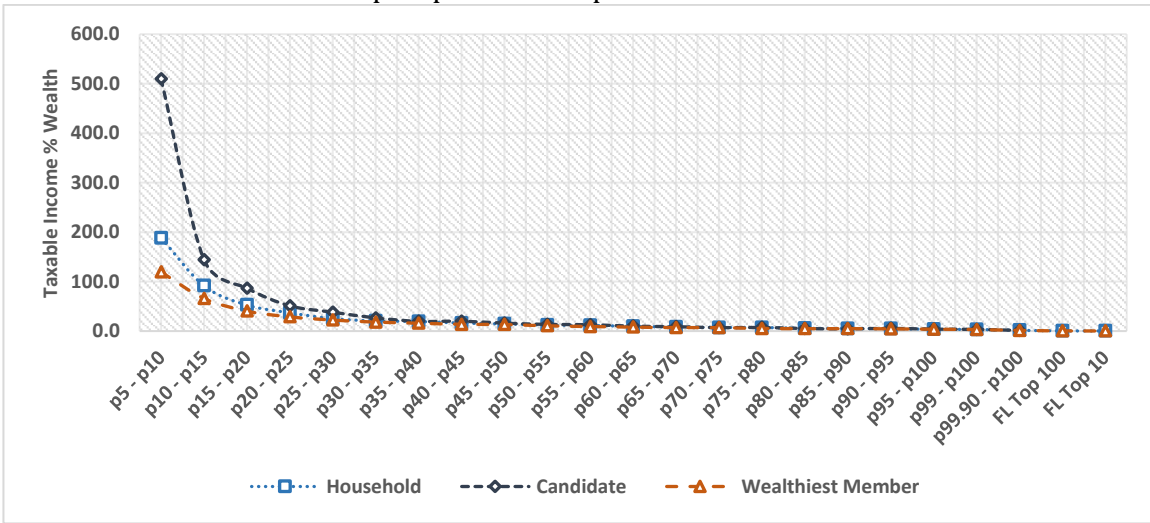**Table 4.1:** Reported taxable income–wealth ratio across wealth groups, $\frac{Y_T}{W} * 100$

Wealth Percentiles	Approach 1		Approach 2	
	Households	Candidates	Households	Candidates
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[34]	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.
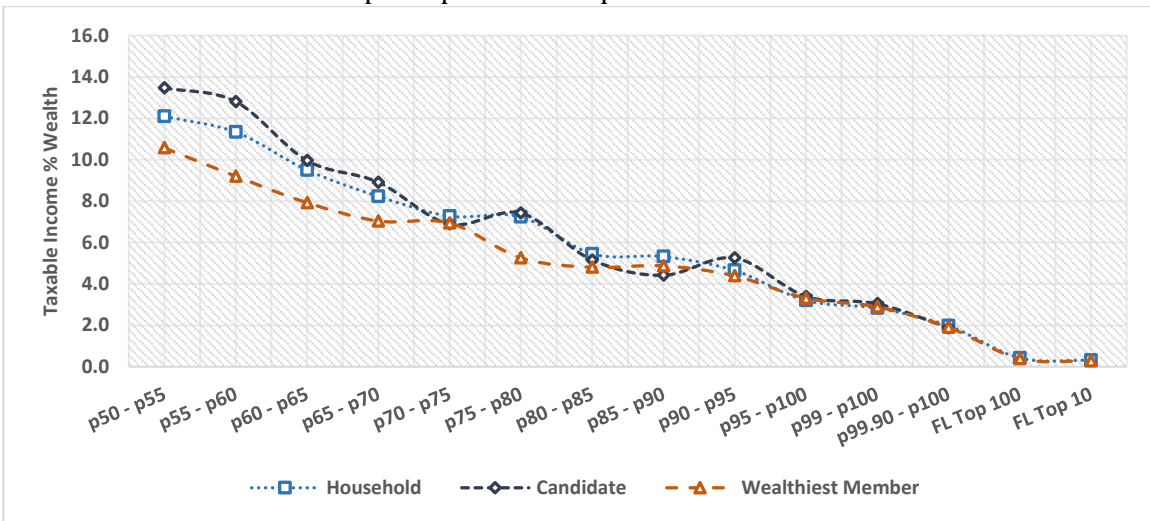
---

[34] For the FL individuals and households, we do not have unit-level information on the reported income. So we have worked with the average wealth and income for the top 100 and the top 10, thus rendering methods 1 and 2 estimates of the income-wealth ratio for the FL the same.

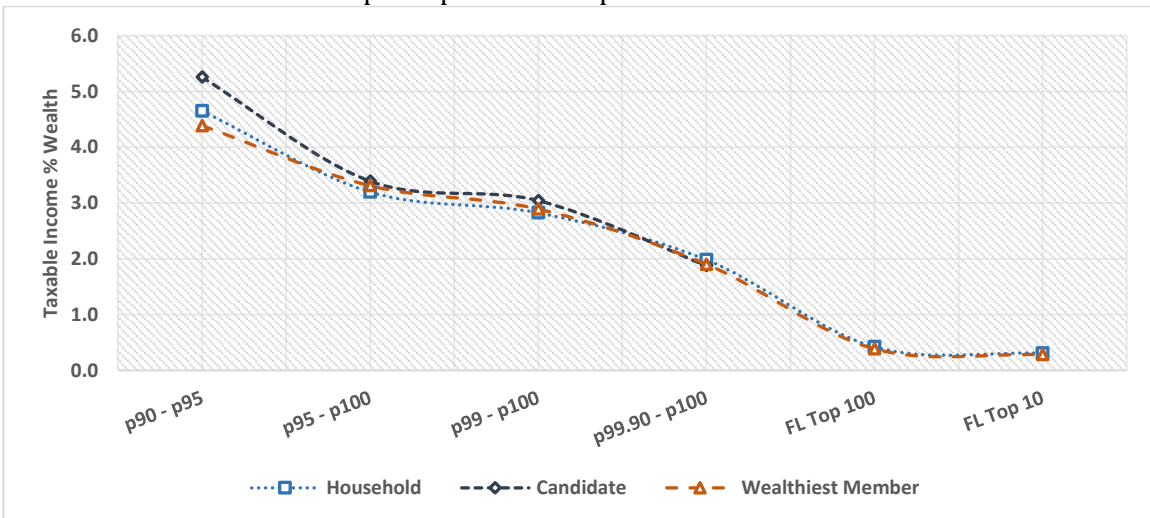**Figure 4.5:** Reported taxable income as a percentage of wealth, across groups (Approach 2)

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

As can be seen from Table 4.1, the two approaches 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.

Overall, the two methods used to compute the income-wealth ratios produce consistently downward trends. Under both methods, the income-wealth ratio decreases with wealth. On average, the taxable income reported by a HH as a proportion of its wealth falls continuously with household 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.[35] 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.

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

Now we consider the total reported income as a ratio of the reported wealth. As discussed above, the total reported income, Y_R, is the sum of taxable income reported to tax authorities *plus* the income declared as "exempt income". In terms of notations, $Y_R = Y_T + Y_{Ex}$. The latter category includes agricultural income and a part of capital income in the form of dividends and profits from firms and partnerships. As detailed in the Appendix, we have estimated capital income using the value of the underlying assets and their rates of returns. 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 non-agricultural properties is calculated as the value of the property *times* the average rents (as a proportion of the property value), and so on. Based on these estimates of the capital income and the reporting rules under the tax law, we have estimated the capital income reported as tax exempt.

It is easy to estimate the total reported income by the bottom 95% HHs and individuals in the GE data. From Section 4.1, we already have a precise estimate of the taxable income, Y_T, reported by these groups. To compute the total income reported by these groups, we just need to estimate their total tax-exempt capital income. As to the rental income, it is taxable in the hands of the recipient and therefore already included in Y_T. Profits from self-owned enterprises are also part of 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 partnership are tax exempt in the hands of the recipient. Besides, their dividend income falls well below the ₹10 lakhs threshold for taxability of dividends.

---

[35] 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.

Specifically, for the bottom 0–95 percentile of HHs and individuals, $Y_{Ex} = Y_{Eq} + Y_A$. Therefore, their total reported income is the sum of the taxable income, the entire equity income, and the agricultural income, i.e., $Y_R = Y_T + Y_{Eq} + Y_A$. In the absence of any direct source of information, profit rates from the equity in 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.

It is challenging to estimate the exempt income reported by the wealthiest groups (say, the top 5% of HHs and individuals in the GE data and those on the FL). As elaborated upon in the Appendix I, these groups receive a major share of their capital income in the name of financial intermediaries such as the limited liability partnerships (LLPs), association of persons (AOPs), and body of individuals (BOIs). To the extent that the 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, the part of income is retained in the account of the intermediaries and does not get reported at all in the ITRs of the partners.

The point is that there is uncertainty about the fraction of the direct capital income reported by super-wealthy groups, and also the capital gains realised by them. In terms of notations, we are not sure about the proportions of Y_{Eq}, Y_A, Y_P, and Y_{Cg} received by the wealthy groups in their own name and in the account of financial intermediaries used by them. Given the high degree of uncertainty, we consider a range of possibilities around the reporting of 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 un-reported in the tax returns.

Going by the evidence discussed in Appendix I, 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:

One-fourth of the total $[Y_{Eq} + Y_A + Y_P + Y_{Cg}]$ (the sum of direct capital income received in individual accounts or in the accounts of financial intermediaries) is reported as exempt income.

Under this assumption the total reported income for the household and individuals in the top 5% of the GE data and the FL is estimated as: $Y_T + 0.25[Y_{Eq} + Y_A + Y_P + Y_{Cg}]$.

As noted earlier, this estimate of the wealthy groups' total reported income appears to be the most plausible. However, given the uncertainty over the income reporting behaviour of these wealthy groups and for the sake of completeness, we also estimate what we consider as the absolute upper and the absolute lower bounds on the total income reported by the wealthiest 5% of units in GE data and the FL.
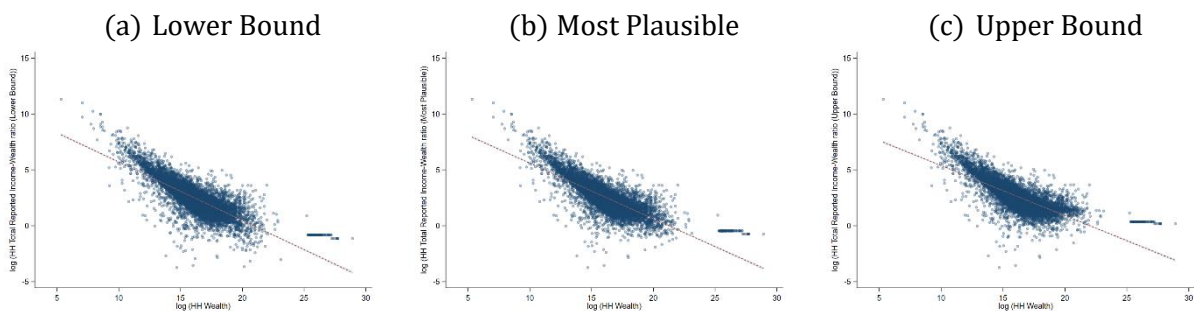
Absolute Lower-bound estimates are derived from the following assumption: Only 5% of the sum $[Y_{Eq} + Y_A + Y_P + Y_{Cg}]$ is reported as exempt income. This means that an absolute lower bound on the total reported income is estimated as: $Y_T + 0.05[Y_{Eq} + Y_A + Y_P + Y_{Cg}]$.

Absolute Upper-bound estimates 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. That is, 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 of the income received in intermediaries' accounts to their individual accounts. If they did so, it would defeat the very purpose of using the financial intermediaries. Besides, it would increase the tax obligation compared to a situation where the capital income is directly received in individual accounts.

Figure 4.6 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 4.6: Household total reported income (Y_R) 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 4.2 and Figure 4.7 presents the group level income-wealth ratios. The dark plots in Figure 4.6 are based on what we have described as the most plausible estimate of the total reported income. The plots also show the range generated by the lower and the upper bounds on the total reported income. Even if we consider the absolute upper bound of the range presented in these tables and plots, 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. Within the GE data, the ratio is the lowest for the top 0.1% units. The range takes lowest value for the wealthiest 10 families on the FL. Therefore, the uncertainty over the part of the capital income between that gets reported in the ITRs does not change the fact that the wealthiest group in the country reports the lowest income in relative terms.

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 4.2, in our most likely scenario, 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
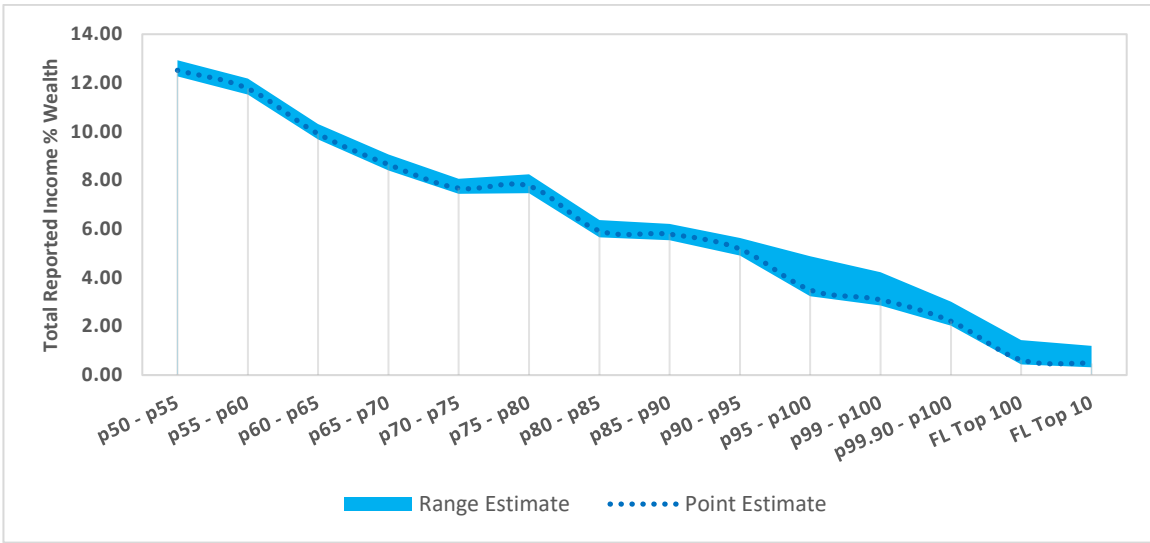
28

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.

Table 4.2: Total reported income–wealth ratio ($\frac{Y_R}{W} * 100$) across wealth groups (Approach 1)
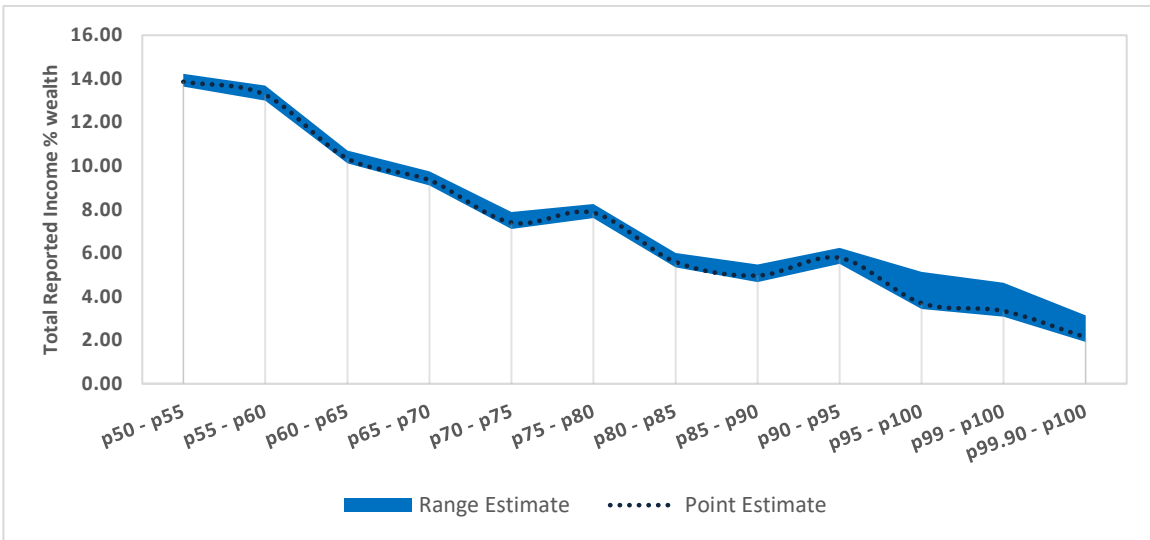
Wealth Percentiles	Households			Candidate		
	Lower Bound	Most Plausible	Upper Bound	Lower Bound	Most Plausible	Upper Bound
p0 - p5	1037.8	1037.5	1037.0	1167.0	1166.2	1164.8
p5 - p10	187.6	187.7	187.9	509.7	509.8	510.0
p10 - p15	91.8	91.9	92.1	144.5	144.7	144.9
p15 - p20	52.7	52.9	53.1	86.9	87.1	87.4
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; $\delta_A = 0.08\%, \delta_P = 0.08\%$.
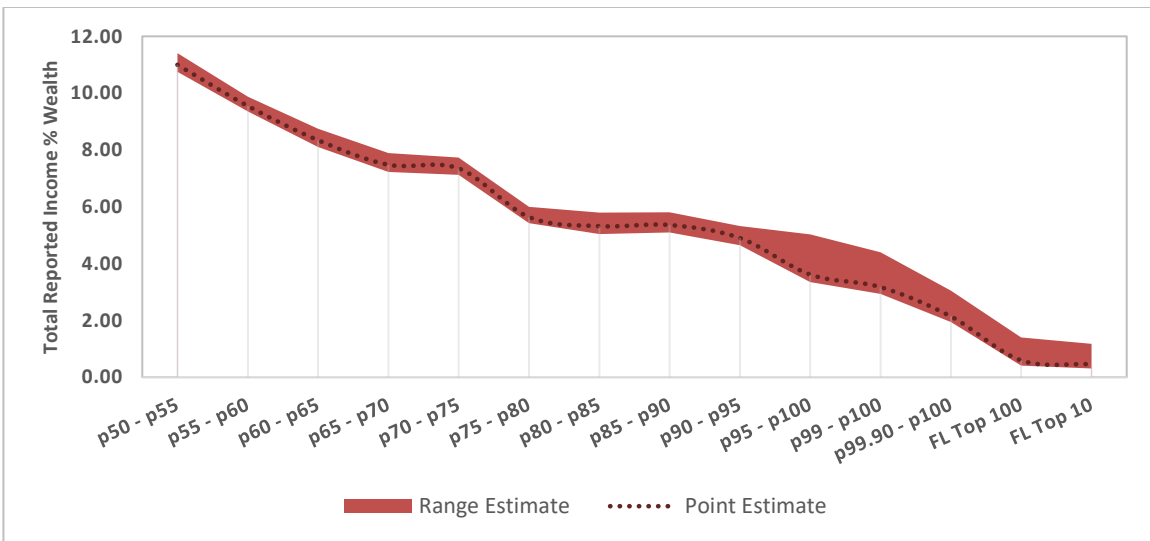
Figure 4.7: Total reported income as a %age of wealth, across wealth groups (Approach 2)

Plot A: Household

Plot B: Candidates

Plot C: Wealthiest Member

4.3 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 the inconsistency in the types of assets covered by the definition of Y_R on the one hand, and of W on the other. By definition, W comprises all assets. However, Y_R does not include (imputed) income received from the self-occupied residential property (imputed rent enjoyed by the taxpayers is not reported in their ITRs).

Therefore, to empirically examine Proposition 2, we consider $Y_{PID}(= Y_R + Y_H)$. It includes income from labour and all forms of assets that contribute to the direct income from wealth as defined in Section 2, including residential properties. Specifically, we test the $\frac{Y_{PID}}{W}$ ratio.[36] 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). So, we expect the ratio $\frac{Y_{PID}}{W}$ to decrease with W. This indeed is the case, as can be seen from Table 4.3 and Figure 4.9 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, the income-wealth ratios with and without imputed rent are comparable. For the top 1% of HHs in the GE data, the reported personal income, including labour income and imputed rent, amounts to about 3.5% of the wealth. For the wealthiest 10 families on the FL, the reported personal income is slightly above half a percent of the wealth. The income-wealth ratio for the individuals follows a similar pattern.

Before concluding this section, a few remarks are in order. 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.

---

[36] Alternatively, we could work with the refined $\frac{Y_R}{W}$, where value of the residential property is subtracted from W. However, this approach makes the wealth negative for a significant number of observations at low, middle and upper-middle wealth levels, generating large negative ratios even for upper-middle groups.
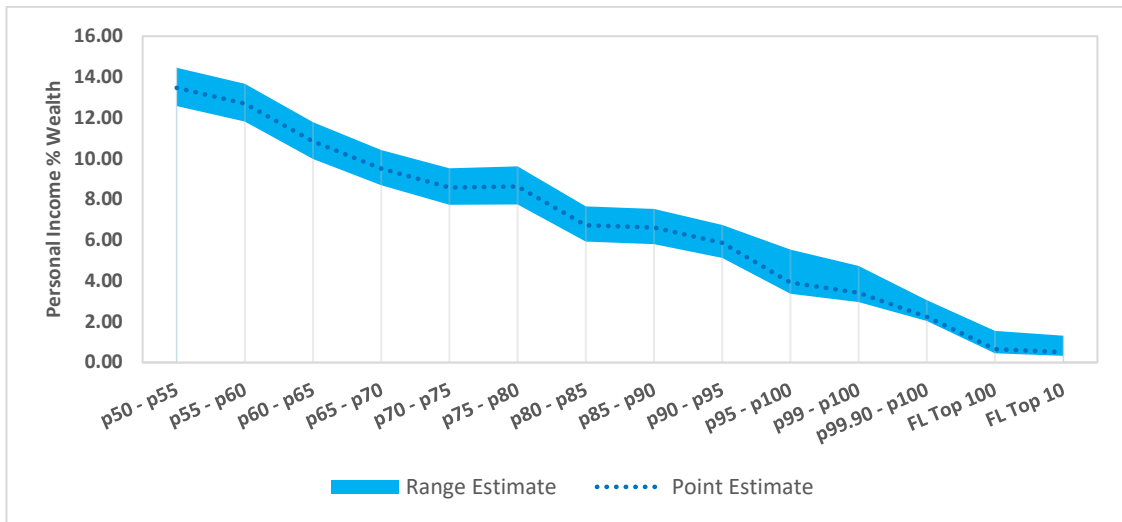
Table 4.3: The ratio of personal income to wealth ($\frac{Y_{PID}}{W} * 100$) (Approach 1)

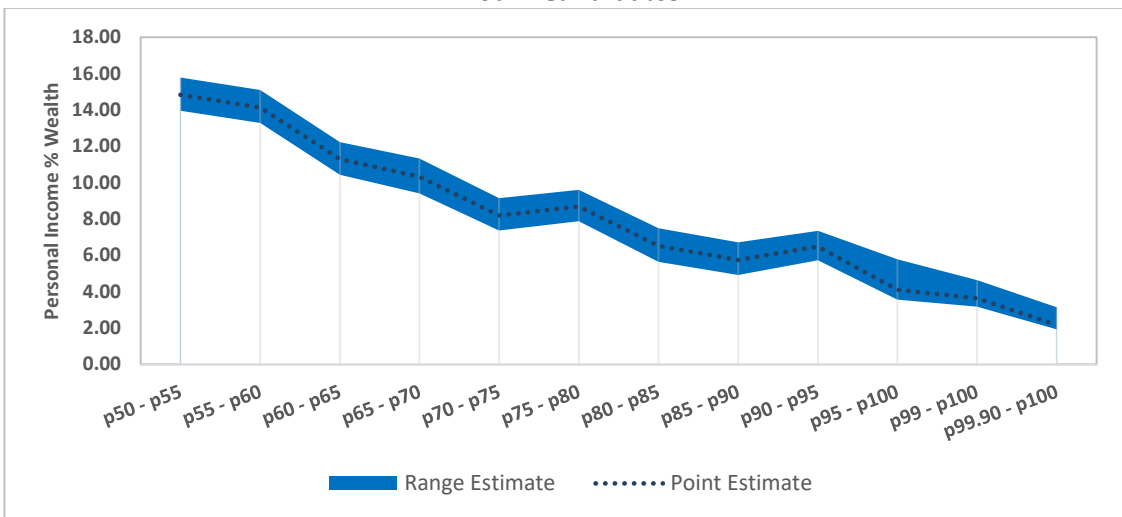| Wealth Percentiles	Households			Candidate		
	Lower Bound	Most Plausible	Upper Bound	Lower Bound	Most Plausible	Upper Bound
p0 - p5	1037.5	1036.4	1035.3	1166.7	1165.1	1163.1
p5 - p10	187.7	188.2	188.8	510.2	511.2	512.2
p10 - p15	92.0	92.5	93.1	144.7	145.1	145.7
p15 - p20	53.0	53.8	54.6	87.2	88.0	88.9
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: 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; $\delta_A = 0.08\%, \delta_P = 0.08\%$.

Figure 4.9: Personal income as a %age of wealth, across wealth groups (Approach 2)
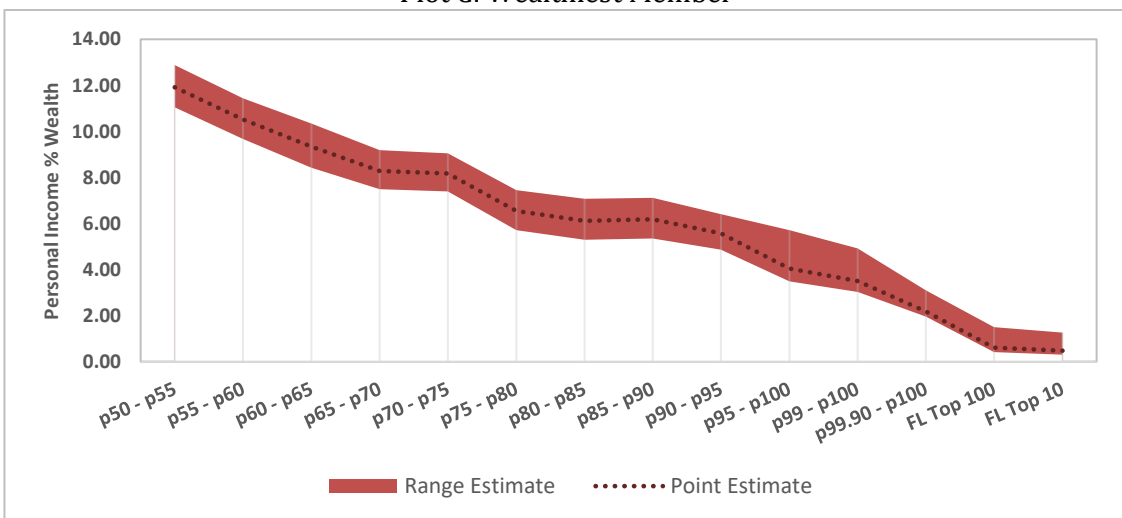
Plot A: Household

Plot B: Candidates

Plot C: Wealthiest Member

5. The Correlates of Income-Wealth Ratio

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

Given the predictions emanating from our model in Section 2, wealth is expected to be a significant predictor of the ratio of the personal income relative to wealth. Indeed, our findings in Section 4 show that all versions of income-wealth ratios decrease with wealth. Besides, in Section 2, we discussed why capital income varies across assets. Specifically, for any given level of wealth, from Proposition 3, we know that the shares of different assets in the wealth portfolio have a bearing on the ratio of the direct personal income to wealth. Accordingly, we consider asset shares as possible determinants of the income-wealth ratios. This gives us regressors *Equity, Banking, Advances, Agri_land,* and *Com_prop* as shares of equity, bank deposits, personal advances, farmland, and commercial property respectively. These are defined in Table 5.1 below.

Table 5.1: 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
Equity	Share of equity in total assets. Equity comprises bonds, debentures, and share/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 the total assets, i.e., "value of the agricultural land" as a ratio of the value of all assets.	GE Affidavit Data
Com_Prop	Share of commercial property defined as commercial building + non-agriculture land.	GE Affidavit Data
D_{2019}	$D_{2019} = 1$ if W and Y are reported for the year 2019 $D_{2019} = 0$ otherwise	GE Affidavit Data
$D_{Unresev}$	$D_{Unresev} = 1$ if the (social) category is "General", i.e., Unreserved (UR); $D_{Unresev} = 0$ otherwise	ECI Results Data
Vote	Variable *Vote* for candidate *i* contesting in constituency *j* $Vote_{ij} = $ (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 capturing the highest educational degree attained by the candidate. The higher the degree, the larger the value taken by this variable.	GE Affidavit Data
D_{Winner}	$D_{Winner} = 1$ if the candidate won both 2014 and 2019 GE elections; 0 otherwise	ECI Results Data
D_{Party}	$D_{Party} = 1$ if the candidate contested elections as a registered state or national party nominee; 0 otherwise	ECI Results Data
Profession *Agriculture Politicians*	Profession =1 if the candidate's profession is *Agriculture and allied activities*.	GE Affidavit Data

	Profession =2 if the candidate's profession is *Politicians and social workers*; 0 for any other profession	
D_{Male}	$D_{Male} = 1$ if candidate gender is male	ECI Results
	$D_{Male} = 0$ if candidate gender is female or other	Data

Summing up, the discussion in Sections 2 and 4 offers the following hypothesis for households.

H1: The income-wealth ratio is: decreasing in wealth; increasing in the share of bank deposits and personal advances; decreasing in the share of agricultural land and commercial property; decreasing in the share of equity assets; and relatively low for the general category.

The distribution of W is skewed towards large values. As can be seen from Figure 3.1 and Figure A5.2 in Appendix I, W and all versions of $\frac{Y}{W}$ follow a log-normal distribution. So, following the approach in Asher and Novosad (2019) and Fisman et al (2019), we use $log\ W$ instead of W as an explanatory variable. For the same reason, we use log of the income-wealth ratio as the dependent variable. Specifically, for the households in our datasets, including the FL families, we use the following specification to test the above hypothesis:

$$log\left(\frac{Y_i}{W_i}\right) = \alpha_0 + \alpha_1 log\ W_i\ +\ \beta.S_i + \beta_{2019}D_{2019i} + \beta_G D_{Unreser} + \epsilon_i \qquad (5.1)$$

where S is the vector of regressors representing shares of income yielding assets such as deposits in banks, personal advances (loans), equity, agricultural land, and commercial properties. The residual category of assets includes assets that do not yield income directly, such as gold, jewellery, durables, and properties used for housing. We use the dummy $D_{Unresev}$ for the social category as a possible explanatory variable. $D_{Unreser}$ takes value 0 for the SCs and STs, and 1 for the rest. The year dummy, D_{2019} is used to examine the year fixed-effects. $D_{2019}=1$ for the income and wealth reported in 2019; and, 0 otherwise.

The above specification is estimated for different versions of the income, i.e., by taking $\underline{Y_i}$ to be Y_{Td}, Y_T, Y_R and Y_{PID}. Results are for $\underline{Y_i} = Y_{Td}$ are omitted from the main text.

5.1 The Empirical Strategy

A host of factors work together to generate any given level of income and wealth for individuals and households. For HHs, data limitations do not allow us to model several factors of interest, and endogeneity thus remains a serious concern.

To reduce the likelihood of omitting a variable of interest, we consider a comprehensive list of regressors for the category of individuals (candidates). We have information on several demographic and other characteristics such as the age, educational attainment, gender, and profession of candidates.[37] Thus, we examine the income-wealth ratio controlling for demographic characteristics in addition to the variables used in (5.1). As evident in literature, when holding constant other factors, age and education have a favourable effect on income.[38] The variables used to capture these individual characteristics are described in Table 5.1 above. In addition, we factor in individuals' profession, as the frequency and size of cash-based transactions differ across professions, which in turn means that the ability to underreport income can vary across

---

[37] We do not have such detailed information for non-candidate members of the household in the dataset. Therefore, our empirical analysis of individuals is restricted to the candidates.
[38] See Piketty (2018, chapter 7).

occupations. We use profession dummies to examine profession-fixed effects on the reported income.

In addition to these, we examine if the reported ratios are different for candidates with a criminal history. To check this, we use the number of criminal cases against a candidate as an explanatory variable.

Given this large set of regressors for candidates, endogeneity becomes a smaller concern for the regression results presented in Tables 5.3–5.4. Still, it is conceivable that we may have omitted some relevant factors. One approach to address this concern and bolster inferences of causality would have been to use an instrumental variable (IV); however, identification of a suitable IV for our main regressors has proved to be challenging given the available data sources. In view of this limitation, we rely on additional evidence (presented in Sections 2, 3 and 4 above) to support our empirical finding that the income-wealth ratio decreases with wealth. In the next section, we will introduce more evidence to support and contextualise our inferences.

Another concern for our empirical analysis is that politicians may exhibit reporting behaviours that are very 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 the political ability affects the reported ratios, we use a candidate's vote share (*Vote*) as a regressor to proxy for political ability assuming that the larger a candidate's vote share is, the greater is their political ability. Further, we examine the ratio for candidates with "extraordinary" political abilities. This set consists of winners of both General Elections and is represented by the dummy variable D_{Winner}.

Yet another concern arises on account of possible underreporting of income and wealth. As to the underreporting of wealth, its scope exists mainly for tangible assets such as land and buildings. Several studies have established that people tend to underreport land and property values[39] Income from these assets can be manipulated even with greater ease. For instance, it is widely believed that people use farmlands to (mis)report their income from non-farm sources under the disguise of tax-exempt agricultural income. Another strand of literature suggests that people underreport rental incomes from commercial properties. In contrast, it is relatively difficult to underreport financial assets such as bank deposits and equity, and the incomes arising from them. To capture the effect of such underreporting, we use the shares of farmland and commercial properties — the primary channels of underreporting — as control variables. Given that underreporting of income is easier than that of wealth, we expect these shares to affect the income-wealth ratios.

Still, in the absence of the suitable information, the exact extent of underreporting is hard to estimate. As a robustness check, we revisit the ratios presented in Section 4 by simply inflating the declared values of land and buildings by 25%. We find that the income-wealth ratios are still decreasing in wealth.

Summing up, for individuals, in addition to Hypothesis 1, we propose the following hypothesis.

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

To test the above hypothesis for individuals, we use the following specification:

---

[39] See Singh (2012 and 2013).

$$log\left(\frac{Y_i}{W_i}\right) = \alpha_0 + \alpha_1 log\, W_i + \beta.S_i + \beta_{2019}D_{2019i} + \beta_G D_{Unreseri} + \beta_M D_{Mi} + \gamma.X_i + \epsilon_i \quad (5.2)$$

where X is the vector of individual characteristics such as age, education, number of criminal cases, and whether the candidate contested elections as a nominee of a national party. The dummy $D_{Male} = 1$ for male candidates, and $=0$ otherwise.[40]

As a background to our empirical analysis, t-tests reveal that the average wealth levels are relatively low for the HHs and individuals belonging to SC and ST categories Candidates of from national/state political parties are wealthier than the rest. We do not find statistically significant differences in the wealth levels of candidates from different professions, gender, and elections years. Further details are provided in Table A5.7 in Appendix II.

5.2 Results

Now we present our econometric results. In the main text, we have presented results for the three versions of the income discussed above; namely, the taxable income (Y_T), the total income reported to the tax department (Y_R), and the direct personal income (Y_{PID}). To save the space, results for Y_{Td} are very similar and are presented in Appendix II.

Tables 5.2–5.3 present results using ordinary least squares (OLS) estimates for HHs with and without including the FL families. Tables 5.4 shows the regression results for individuals. Model 1 uses specifications (5.2) and (5.3) without fixed effects. Model 2 uses the full specification with fixed effects. The figures reported in brackets are use heteroscedasticity-corrected robust standard errors.

As can be seen from the regression tables, for households and also for individuals, all versions of the income-wealth ratio are decreasing functions of W. As expected, the ratios are increasing in the share of bank deposits and personal advances, when other factors are held fixed.

Specifically, on average, a 1% increase in wealth is associated with approximately 1.5% [respectively 1.6%] decrease in $\frac{Y_R}{W}$ for HHs [respectively individuals] [respectively $\frac{Y_{PID}}{W}$]. From Section 4, we know that the steepness of the fall in the ratio declines at the top wealth levels, indicating non-linear relationship between income and wealth. A 1% increase in equity share leads to 0.06-0.08% increase in the $\frac{Y_R}{W}$ ratio reported by the HHs. The corresponding figure for individuals is 0.07%. For every 1% increase in farmland share, there is 0.03% [respectively 0.02%] decrease in the ratio for HHs [respectively individuals]. A 1% increase in commercial property share is associated with 0.03% decrease in the ratio for HHs and individuals.

---

[40] We have also estimated models (5.1) and (5.2) with an added term $(\log W)^2$. See Appendix. Under this version of the models, coefficients of $\log W$ and $(\log W)^2$ are both significant with positive and negative sign, respectively, thereby suggesting 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. So, the term $(\log W)^2$ is not included the main model.

Table 5.2: Households without the FL income-wealth ratio

Households without FL	$\log\left(\frac{Y_T}{W}\right)$		$\log\left(\frac{Y_R}{W}\right)$		$\log\left(\frac{Y_{PID}}{W}\right)$	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
$\log W$	-0.554*** (0.007)	-0.553*** (0.007)	-0.530*** (0.007)	-0.529*** (0.007)	-0.484*** (0.007)	-0.483*** (0.007)
Banking	0.009*** (0.001)	0.009*** (0.001)	0.009*** (0.001)	0.009*** (0.001)	0.007*** (0.001)	0.007*** (0.001)
Equity	0.013*** (0.002)	0.013*** (0.002)	0.014*** (0.001)	0.014*** (0.001)	0.010*** (0.001)	0.010*** (0.001)
Advances	0.012*** (0.001)	0.012*** (0.001)	0.012*** (0.001)	0.012*** (0.001)	0.009*** (0.001)	0.008*** (0.001)
Agri_Land	-0.008*** (0.001)	-0.008*** (0.001)	-0.003*** (0)	-0.003*** (0)	-0.005*** (0)	-0.006*** (0)
Com_Prop	-0.002*** (0.001)	-0.002*** (0.001)	-0.002** (0.001)	-0.002** (0.001)	-0.004*** (0)	-0.004*** (0)
D_{2019}		0.104*** (0.022)		0.105*** (0.02)		0.100*** (0.018)
$D_{Unresev}$		0.002 (0.025)		0.009 (0.023)		0.012 (0.021)
Constant	11.331*** (0.118)	11.268*** (0.118)	10.938*** (0.111)	10.874*** (0.111)	10.440*** (0.107)	10.378*** (0.107)
R^2	0.655	0.656	0.668	0.669	0.683	0.684
Observation	7433	7433	7433	7433	7433	7433

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

On the face of it, these results seem counterintuitive. The rental yield is generally higher than the dividend yields. As discussed in Section 2, the rate of direct returns tends to be the lowest for equity compared to the other income-yielding assets. Specifically, the ratio of dividend yields to the value of equity is less than the ratio of rental income to the property value, which is lower than the ratio of interest income to the value of the corresponding instrument (e.g., bank deposits, etc.). Thus, at any given level of wealth, direct capital income, and hence the $\frac{Y_R}{W}$, 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, the higher is the reported income. On the other hand, larger percentages of farmland and commercial property are correlated with lower reported incomes, and vice versa.

Households with FL	$\log\left(\frac{Y_T}{W}\right)$		$\log\left(\frac{Y_R}{W}\right)$		$\log\left(\frac{Y_{PID}}{W}\right)$	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
$log\ W$	-0.541*** (0.007)	-0.541*** (0.007)	-0.517*** (0.007)	-0.516*** (0.007)	-0.474*** (0.007)	-0.474*** (0.007)
Banking	0.009*** (0.001)	0.009*** (0.001)	0.009*** (0.001)	0.009*** (0.001)	0.008*** (0.001)	0.008*** (0.001)
Equity	0.019*** (0.001)	0.019*** (0.001)	0.021*** (0.001)	0.021*** (0.001)	0.015*** (0.001)	0.015*** (0.001)
Advances	0.011*** (0.002)	0.011*** (0.002)	0.011*** (0.001)	0.010*** (0.001)	0.008*** (0.001)	0.007*** (0.001)
Agri_Land	-0.008*** (0.001)	-0.009*** (0.001)	-0.003*** (0)	-0.003*** (0)	-0.006*** (0)	-0.006*** (0)
Com_Prop	-0.002*** (0.001)	-0.002*** (0.001)	-0.002*** (0)	-0.002*** (0)	-0.005*** (0)	-0.005*** (0)
D_{2019}		0.106*** (0.022)		0.117*** (0.022)		0.110*** (0.018)
$D_{Unresev}$		-0.007 (0.025)		0 (0.023)		0.005 (0.021)
Constant	11.123*** (0.118)	11.065*** (0.118)	10.722*** (0.111)	10.661*** (0.111)	10.278*** (0.105)	10.220*** (0.105)
R^2	0.671	0.672	0.682	0.683	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*

This result supports the widely held belief that people underreport agricultural and rental income. Most agricultural incomes and a significant fraction of rentals are received in cash. These incomes do not create a verifiable trail of transactions and can easily be misreported. An underreporting of rental income reduces the tax burden on recipients, and in the process pulls down the reported values of taxable as well as the total income declared.

Table 5.4: Individuals' income-wealth ratio

Individuals	$\log\left(\frac{Y_T}{W}\right)$		$\log\left(\frac{Y_R}{W}\right)$		$\log\left(\frac{Y_{PID}}{W}\right)$	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
$\log W$	-0.632*** (0.008)	-0.687*** (0.009)	-0.605*** (0.007)	-0.654*** (0.008)	-0.555*** (0.007)	-0.600*** (0.008)
Banking	0.007*** (0.001)	0.006*** (0.001)	0.008*** (0.001)	0.006*** (0.001)	0.006*** (0.001)	0.005*** (0.001)
Equity	0.012*** (0.002)	0.011*** (0.002)	0.014*** (0.002)	0.013*** (0.002)	0.010*** (0.001)	0.009*** (0.001)
Advances	0.011*** (0.001)	0.009*** (0.001)	0.011*** (0.001)	0.009*** (0.001)	0.008*** (0.001)	0.006*** (0.001)
Agri_Land	-0.007*** (0.001)	-0.006*** (0.001)	-0.002*** (0)	-0.001** (0)	-0.005*** (0)	-0.004*** (0)
Com_Prop	-0.002*** (0.001)	-0.002*** (0.001)	-0.002** (0.001)	-0.002*** (0)	-0.004*** (0)	-0.004*** (0)
Vote		0.462*** (0.041)		0.425*** (0.036)		0.375*** (0.033)
Criminal		-0.004*** (0.001)		-0.002*** (0.001)		-0.002*** (0.001)
Education		0.022*** (0.005)		0.020*** (0.005)		0.019*** (0.005)
Age		-0.002* (0.001)		-0.003** (0.001)		-0.002 (0.001)
D_{Winner}		0.024 (0.05)		0.023 (0.045)		0.022 (0.041)
D_{2019}		0.128*** (0.022)		0.121*** (0.02)		0.113*** (0.018)
D_{Male}		0.134*** (0.039)		0.132*** (0.035)		0.118*** (0.033)
$D_{Unresev}$		0.054* (0.025)		0.061** (0.024)		0.067** (0.022)
D_{Party}		0.147*** (0.027)		0.130*** (0.025)		0.111*** (0.023)
Agriculture		-0.205*** (0.039)		-0.184*** (0.033)		-0.161*** (0.03)
Politicians		-0.162*** (0.037)		-0.151*** (0.033)		-0.139*** (0.03)

Constant	12.434*** (0.117)	12.882*** (0.137)	12.015*** (0.112)	12.421*** (0.13)	11.475*** (0.108)	11.799*** (0.124)
R^2	0.715	0.73	0.727	0.741	0.738	0.751
Observation	6953	6953	6953	6953	6953	6953

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

The negative and significant coefficient of the commercial property share for all versions of $\frac{Y}{W}$ supports this inference. Given that the coefficient of equity is positive, it follows that the reported rental income is even smaller than the dividend income at a corresponding level of wealth. In the absence of underreporting, we would not expect this result.

The negative and statistically significant coefficient for the agricultural income share has somewhat different implications. Farm income is tax exempt; so, taxpayers have very little incentive to underreport it. However, farmland may offer 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 pull 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 government audit reports showing such misuse of farmlands by taxpayers to reduce their tax burden.[41]

The dummy variable $D_{Unresev}$ is insignificant, showing indicating that the social identity of a candidate and HH does not consistently affect the reported ratios. Age, similarly, does not appear to have a significant bearing on the reported ratios. Among other individual characteristics, educational qualification has a positive correlation with the $\frac{Y}{W}$ ratios, while the degree of criminality has a negative correlation. Ceteris-paribus, the larger the number of criminal cases, the smaller the reported income.

Our results are interesting with respect to candidates' political ability measured as the vote share. *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 bear in mind that the media and civil society scrutiny are stricter for candidates with a serious chance of winning an election.[42] To avoid falling foul of the Election Commission and risking their chance at the hustings, these candidates have a stronger incentive to truthfully report their incomes as well as wealth. That the ratios are increasing in vote share also implies that media and official scrutiny have a more pronounced effect on reported incomes relative to reported wealth.

This result corroborates an earlier inference that underreporting incomes is easier than misreporting wealth. Besides, it suggests that the non-politician citizens (who presumably have no or very little political abilities) report smaller incomes than similarly placed politicians. 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. This result points toward a general tendency to underreport income across wealth

---

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

[42] 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*.

groups, except for those who cannot avoid the media glare and official scrutiny.

Controlling for the vote share, the income-wealth ratios for "super-politicians" — i.e., candidates who won both GEs — are not different from the rest. The positive (and significant) coefficient of the variable D_{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 other candidates. This finding further underscores the role of scrutiny and enforced transparency on income reporting behaviours.

We find the year, profession and gender fixed-effects. *Ceteris-paribus*, full time agriculturists and politicians report relatively low-income. On average, women tend to report smaller incomes than men. This latter finding appears to be a consequence of the 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. Women also own a larger share of non-income yielding assets like gold and jewellery. The income-wealth ratios are thus expected to be relatively low for women. Our finding on Y_R and, Y_{PID}, are very similar to Y_R.

In conclusion, we would like to add that our results are robust to the inclusion/exclusion of dummies, fixed effects, and individual characteristics. However, we would like to emphasise that some of our results are sensitive to the specification of the dependent variable. As discussed above, to avoid the effect of skewed distribution of the variables $\frac{Y_i}{W_i}$, and in the interest of fitness of the model, 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 Tables A5.3-A5.6 in Appendix II), 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.

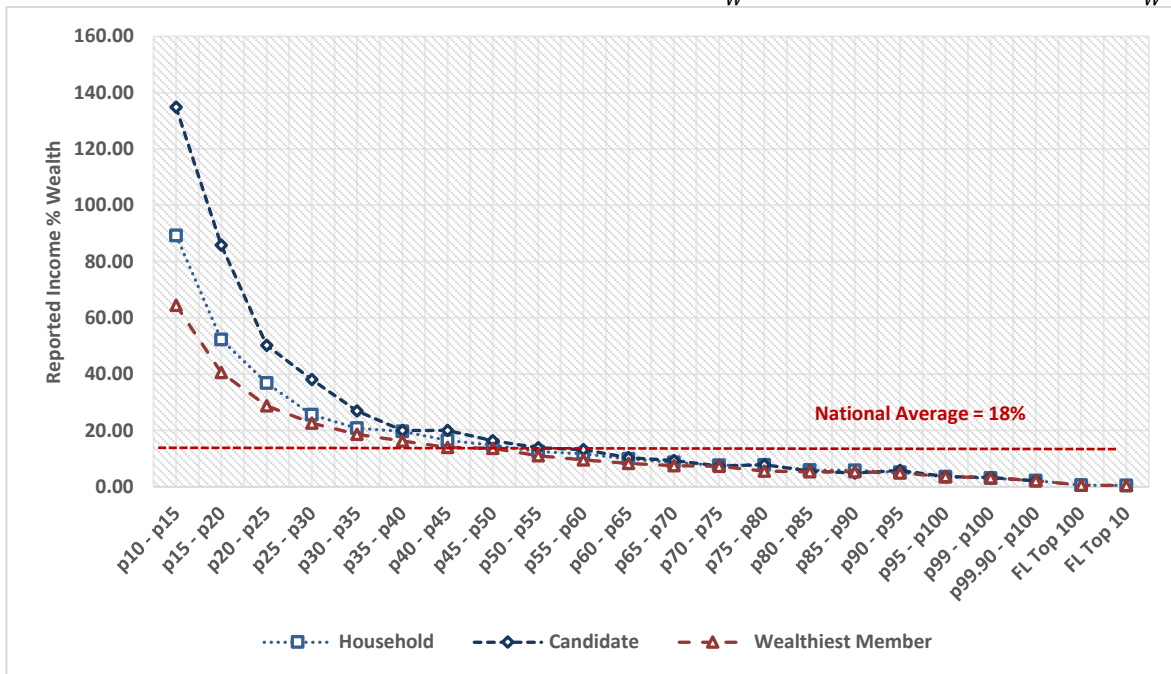## 6. The Missing Income at the Top: *How much and how come?*

As is clear from our findings in Sections 4 and 5, the reported income relative to the corresponding wealth decreases continuously and sharply with the latter. Incomes reported by the bottom 25% 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 ratio of the estimated personal income reported by the different groups vis-à-vis their wealth. The personal 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 very high for the poor. Consider a rural landless household living off an annual wage income of ₹1.20 lakhs. Assume the only asset owned by the household is a tiny house worth ₹40,000 and that it owes a debt of ₹10,000. In this case, the income is 400% of the family wealth. The ratio can be even higher for households with comparable income but lesser or no wealth. In contrast, consider a Forbes list family with a net equity wealth of, say, ₹10,000 crores. Assume the rate of total returns on its capital is 15% (a high rate of returns by all means). Even if this household earns another 500 crores as

labour income, its cumulative income-wealth ratio will be 0.2= (1500+500)/10,000, i.e., the income will be just 20% of the family wealth. More generally, 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 the two GEs studied by us (i.e., during 2010-20), the average national income was 18-20% of the average private wealth.[43] The income levels reported by the wealthy groups pertain to the same period but are much small by comparison. Figure 6.1 depicts the $\frac{Y_R}{W}$ ratios for wealthy groups compared to the average national income as a ratio of national wealth (18%). It is evident that the income-wealth 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 on the FL, it is merely one-twentieth of the national average!

Figure 6.1: Ratio of total reported income to wealth, i.e., $\frac{Y_R}{W}$, versus the national average of $\frac{Y}{W}$

Even considering the average returns on capital, incomes reported by wealthy groups are far below the expected levels. The rate of returns on capital is the capital income expressed as a percentage of the value of the capital stock. At the national level, the average rate of returns on the aggregate stock of capital can be estimated as the capital share of the national income *times* the ratio of national income to national wealth. The larger the share of capital in the national income or higher the income-wealth ratio is, the higher will be the rate of returns on wealth, and vice versa. For the decade relevant 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 in the range of 18-20% of the national wealth,

---

[43] 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).

most of which is private wealth. [44]

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. In other words, one could ensure returns comparable to the national average by liquidating their assets and putting the proceeds in fixed deposits. 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 returns from equity-oriented mutual funds as a reference point, the 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 the fact 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. The massive fortunes enjoyed by the wealthiest groups in recent years also suggest a higher than 10% rate of returns on their capital. Moreover, total income also includes labour income in addition to the capital income.

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 presents a strikingly different picture.

Figure 6.2: Total reported income, Y_R, as a percentage of the capital income (7.2% of wealth).
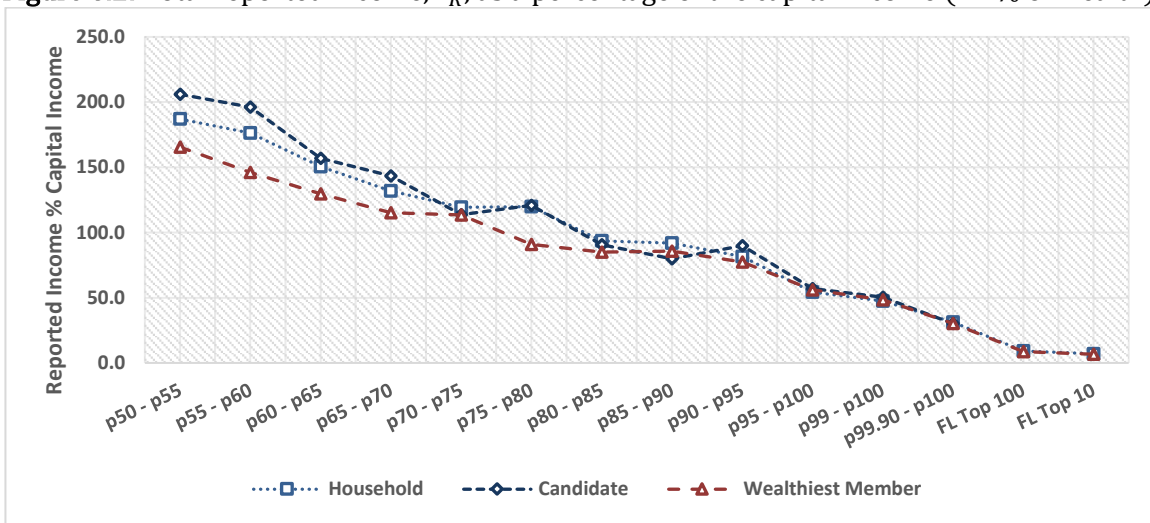

Figure 6.2 shows the reported income, Y_R, as a ratio of the capital income, taken to be 7.2% of wealth. As is evident from the figure, for the top 15% of families and individuals, the reported income is less than the return from their capital. The total income reported by the top 5% HHs and individuals 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

---

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

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.

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 reported total income, 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 on the FL goes unreported!

Given the lion's share of capital income in the total income of wealthy groups, the difference between the reported and the actual income is mostly on account of the capital income that goes unreported. Therefore, the above numbers show that the share of the unreported 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 rate of returns on capital owned by the wealthy groups seems to be much higher than the 10% rate assumed by us.

In view of the above-discussed large proportions of the income of wealthy groups going unreported in the individual tax accounts, we have to 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 the wealthy groups, the forms of capital income received, and the reporting requirements for various kinds of capital income.

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 manipulate 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 bear in mind an essential consequence of the dominance of equity and commercial property in the asset portfolio of wealthy groups. It means that the *capital gains*, i.e., the appreciation in the market value of the assets, is a dominant form of capital income for the wealthy groups. The market value of commercial properties, stocks, and equities tend to appreciate over time, leading to the accumulation of massive capital gains for the owners. For accounting purposes, the capital gains from an asset are treated as "unrealised" unless they are exchanged or sold.[45]

Under 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, it is not a tax liability regardless of the quantum of appreciation in the asset's value on account of the unrealised capital gains. 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, to reduce their tax

---

[45] In the parlance of commerce, unrealised capital gains are a part of the economic income from capital but not the accounting income. The latter includes only the actually realised income.

liability, wealthy groups have a strong incentive to avoid realising capital gains. They do so by staying invested in the equity and commercial properties. Their motivation to stay invested is matched by their ability to do so.[46] This is the primary reason for realised capital gains often being a tiny fraction of the capital income of the wealthy. Appendix II contains a deeper discussion of this point.

Guided by similar considerations, wealthy groups manipulate other forms of capital income. Take, for instance, the case of direct income from equity assets in the form of dividends, i.e., profits distributed among the stockholders of a company. 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. As discussed above, the capital gains 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 into the group companies by keeping their dividends pay-outs as low as possible. Wealthy individuals, such as CEOs, board members, or promotors of group companies, decide whether and how much of the profits will be distributed as dividends. Again, their incentive to reinvest profits is backed by the authority they enjoy in the hierarchy of corporate governance. Such manipulations of capital income in response to the dividend tax are an international phenomenon.[47]

However, compared to the scenario in several other countries, dividend pay-outs by the Indian companies are meagre.[48] 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. Deliberately suppressed dividend yields are 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, a part of the dividend income might be received indirectly and thus retained in the accounts of entities like LLPs. Consequently, most of the equity income of the wealthy groups goes unreported in the form of undistributed profits and unrealised capital gains.[49]

This logic also applies to the income from non-agricultural land and commercial properties. Only rentals and farm income are required to be reported; capital gains are not. The rental yields are generally higher than the dividend yields.[50] Thus, direct capital income is expected to increase with the share of commercial property in the wealth. However, rental transactions often do not create a verifiable record trail and hence can be manipulated and underreported easily. As can be seen from our regression analysis, at any given level of wealth, the reported income decreases as the share of the land and property increases. Our results suggest that the reported rental income is even smaller than the dividend income at a corresponding level of wealth.

---

[46] In contrast, middle-wealth groups have to sell their assets to meet other financial needs and pay capital gains tax in the process.

[47] For a discussion, see Chetty and Saez (2005), Kari and Karikallio (2007), and Boissel and Matray (2022).

[48] Compared to the assets of Indian companies, their declared profits are also low. For more information, see Kanojia and Singh (2013). Also see Labhane (2019).

[49] As an aside, since most of the profits remain undistributed, one can understand why top Indian corporations want to reduce debt on their books: Given the huge cash in their accounts, they do not need to borrow much.

[50] The rents tend to be in the range of 2–4% of the property value. While the rental income is only a fraction of the total returns from the property, it is still more than three times the rate of realised income from the equity assets.

As a result of the above-discussed manipulations and underreporting, 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 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 all his business empire.[51]

Before concluding this section, we find it pertinent to highlight that the relatively high income reported by the middle and low-wealth groups does not mean that these groups report all returns from their capital. Our results in Section 5 suggest that ipso facto reported income decreases with the share of agricultural land. Our results and the available evidence[52] suggest that people across wealth groups report a part of the taxable labour income as agricultural income to avoid paying tax. Even then, income-wealth ratios are relatively very high for the low and middle wealth groups due to two interrelated factors. Given the small amounts of wealth held by these groups, their capital income is relatively small compared to their labour income. Thus, the scope of manipulating the former is restricted. Besides, their wealth is smaller than their income, leading to high income wealth ratios.

7. Two Implications of the Decreasing Income-Wealth Ratio

This section discusses two implications of the decreasing income-wealth ratio and the income missing from the top.

7.1 Tax regressivity

The Indian tax regime is considered to be progressive in that the marginal tax rate (the rate applicable on each additional unit of income) increases with the reported income.[53] However, as we have seen above, income levels reported by individuals and HHs in their ITRs are only a fraction of their total income. The difference between the declared and actual total income 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, which is typically used as a reference point. Moreover, as wealth is an important 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. 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.

First, consider the tax on the income taxable in the hands of the individual receiving it, 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
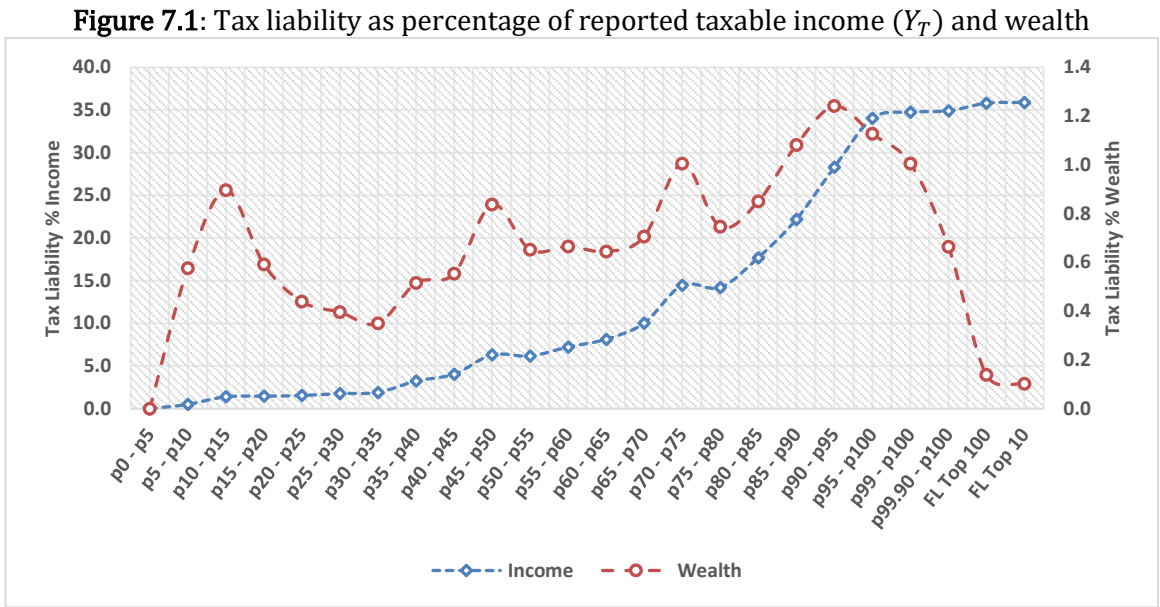
---

[51] 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?

[52] 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.

[53] 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

the precise information on the latter. The computations are done using the online calculator provided by the income tax department for the assessment year 2019-20.[54] It is assumed that the taxpayer is a male aged below 65 with a "resident" status. Moreover, the source of taxed income is taken to be salary. These assumptions mean that we have estimated the highest direct tax liability applicable to each group.

Figure 7.1 shows the estimated tax liability as a ratio of the income reported as taxable, i.e., Y_T, and the ratio of 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.

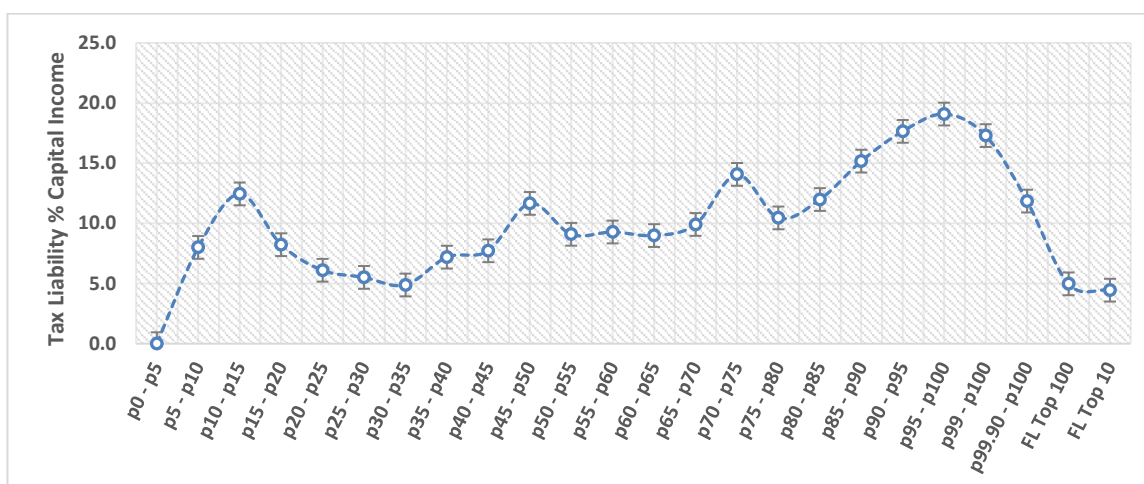**Figure 7.1**: Tax liability as percentage of reported taxable income (Y_T) and wealth

Next, consider the ratio of total tax liability of the 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 7.2, there is an inverted U-shaped relationship between the 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, and hence their total income. By a similar argument, the average tax liability of the top 0.1 centiles in the GE data is less than one-tenth of their income. This is even 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!

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%. Reported value of the income reported by the FL has to be 12 times of what can be observed in the data.

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

---

[54] 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.

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. An examination of the progressivity of the tax regime with respect to the total income requires an estimation of the total tax liability for various income groups. It is a complex exercise and a requires a separate study.

However, the ratios presented in Figure 7.2, along with the fact that most of the income of the wealthiest groups is their capital income, 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 average total income is increasing in wealth, this means that the tax liability as a ratio of the total income is decreasing in the latter, making the effective tax regime regressive, at least at the top.

Even if we go by a conservative estimate of the capital income in Figure 7.2, for the tax regime to be progressive, the taxable income reported by the wealthiest 0.1% has to go up at least by 60%. Reported value of the income reported by the FL has to be at least four-time of what can be observed in the data.

7.2 Underestimated Inequality

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 income levels. Further, as is evident from 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 families 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 level of 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 on 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, these existing studies have underestimated the levels of income inequality in the country.[55]

The second issue pertains to identifying financially elite groups using income tax data. 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 in India do not feature among the wealthiest 100 individuals, and vice versa.

8. Conclusions and Remarks

In this paper, we have modelled and estimated the relationship between wealth and reported income for more than 7,600 families and their adult members. We have used several data sources, including affidavits filed by election contestants, the ProwessIQ dataset, the Forbes List of billionaires, the annual statistics published by the Income Tax Department, as well as the annual accounts of listed companies managed by the wealthiest Indian families. 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 as taxable by households, or the total income declared by them, including the income reported as tax-exempt. These decreasing income-wealth ratios are consistent with what is predicted by our model. However, the magnitude of ratios is strikingly small particularly for the affluent groups.

According to our estimates, the average income reported as taxable by the bottom 10% of households is equivalent to more than 170% of the family wealth. In contrast, for the top 5% of HHs, the reported taxable income amounts to less than 4% of their wealth. For the top 0.1 percentile of HHs, the reported taxable income is less than 2% of their family wealth. For the wealthiest ten families on the FL, the reported taxable income is less than 0.6% of their wealth.

The results are very similar for the total income reported to tax authorities. 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 on 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 — the wealthier an individual, the smaller is the reported income relative to their wealth. On average, the total income reported by the bottom 10% of individuals is more than 120% of their wealth; for the wealthiest 5% of individuals, it is just about 3.7% of their wealth. For the top 0.1% of the most affluent, the total reported income is only about 2% of their wealth. The ultra-wealthy individuals on the FL report the lowest income — about 0.5% of their wealth.

We have shown that the low values of the reported income-wealth ratios for affluent groups are primarily because a large share of their total income goes unreported. The problem of missing

---

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

income is most pronounced for the capital gains enjoyed by them. According to our estimates, the total income reported by the wealthiest 5% of individuals is approximately only a third of the returns from their capital. The total income reported by the top one-tenth of the top centile adds up to just about one-fifth of the returns from their capital. In other words, even after factoring in all types of declared income, their total reported income amounts to less than 20% of their capital income; at least 80% of returns from their capital go unreported. For the families on the FL, more than 90% returns from their capital do not figure in their reported income. The proportions of the missing income are much higher if we compare the reported income with the total (labour plus capital) income.

The missing income of the affluent groups underscore the case for going beyond the standard approach for assessing the progressivity of taxation. There is a case for considering the total income, and not just the reported income, for this purpose. We have shown that the tax paid by the wealthiest 5% amounts to less than one-fifth of their capital income. The average tax liability of the wealthiest 0.1 centiles is just one-tenth of the returns from their capital. The tax liability of the super-wealthy Indians on the FL is less than 5% of their capital income!

We also find profession-, year-, and the gender-fixed effects. Moreover, our empirical analysis suggests that people across the wealth spectrum underreport their rental income, and misreport part of their taxable income by disguising it as tax-free farm income. *Ceteris paribus*, women tend to report lower incomes than men, and that full-time agriculturists and politicians report relatively low levels of income. Further, holding other factors constant, people with criminal records also report relatively low incomes. In contrast, individuals exposed to higher levels of media and civil society scrutiny report relatively high levels of income.

Moreover, we have shown that the effective tax rate is not progressive with respect to wealth. At the top wealth levels, the wealthier an individual is, the smaller their relative tax liability tends to be. Even with the most generous estimates, the tax liability of the top centile amounts to 1% of their wealth. For the top one-tenth of the top centile, the total tax liability amounts to less than 0.8% of their wealth. The super-wealthy Indians on the FL pay tax that is less than 0.2% of their wealth — much smaller than the tax liability for individuals at middle wealth levels.

These findings should be of interest beyond just the Indian context since the dynamics of capital income modelled and empirically examined by us are similar across market economies. Indeed, going by the available evidence,[56] it will not be surprising if similar research in other countries leads to similar findings, with some variations, of course, in the proportions of missing incomes.

Specific to the Indian context, the missing income at the top has implications for the existing estimates of income inequality.[57] Studies on the subject typically rely on the statistics on taxable income as published by the Indian Tax Department. Our study points to a significant difference between the income levels that are fed into these studies and the actual income of the affluent Indians. By failing to capture a non-negligible fraction of the top total income levels, these studies may have severely under-estimated inequality levels in the country.

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

---

[56] See ProPublica June 2021.
[57] 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), and Sahasranaman and Jensen (2021) for an overview of these findings.

reported by an average Indian is probably smaller than what is seen in our estimates. We cannot verify whether this is the case. Moreover, due to a lack of more detailed data, our categorisation of individuals among different professions and educational qualifications is not very precise. Further, endogeneity is also a concern. A study based on a bigger database with more granular information might produce different results.

Finally, it is pertinent to discuss the implications of possible under-reporting of wealth by different wealth groups. In the absence of the available information, we have revisited the income-wealth ratios presented in Section 4 by simply inflating the declared values of land and buildings by 25%. We find that the income-wealth ratios still decrease sharply and continuously with wealth, but the fall is less steep now. Future studies based on objective estimates of asset values may lead to results somewhat different from ours.

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.

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 Portfo lios 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.

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

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.

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.

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.

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

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.

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

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.

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.

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*. 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.

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

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.

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