

Evaluating Tax Return Data as a Source of Income Statistics*

UNAYAMA Takashi

Professor at Institute of Economic Research, Kyoto University

SANO Shinpei

Professor, Graduate School of Economics, Kobe University

YUGAMI Kazufumi

Professor, Graduate School of Economics, Kobe University

INABA Kazuhiro

National Tax College

Abstract

This study evaluates the usefulness of income tax return data, which is newly available in Japan, as a source of income statistics, focusing on its characteristics, advantages, and limitations. The large scale of the data and its accuracy constitute major advantages, but they exclude individuals who are not required to file tax returns as well as income exempt from taxation, which limit their ability to capture the whole distribution of income.

Approximately 23 million individuals file tax returns each year, with about 5 million additional individuals reported as dependents. Together, these groups account for only about 30 percent of the total population aged 15 and over. In particular, most wage earners and pensioners do not file returns, necessitating the use of supplementary information.

Among wage earners and pensioners who do file returns, many do so to claim deductions such as the medical expense deduction or the tax credits for mortgage loan interest. Given the nature of these institutions, filers are supposed to be in distinct economic conditions—such as facing illness or having recently purchased a home—which indicates a potential selection bias within this group.

By contrast, for self-employed individuals, the tax return data provide comprehensive information. The number of business income filers closely aligns with the self-employed

* We are grateful to Yuta Takahashi and participants of seminar at the Ministry of Finance for their valuable comments. This study has also benefited from financial support provided by the Grant-in-Aid for Scientific Research (KAKENHI, Category A, Grant No. 23H00053) and from the Joint Usage and Research Center Project at the Institute of Economic Research, Kyoto University. This article is based on a study first published in the Financial Review 160, pp. 150-181, Unayama Takashi, Sano Shinpei, Yugami Kazufumi, and Inaba Kazuhiro, 2025, "Tax return data as income data" written in Japanese. This study is a part of "Joint Statistical Research Program" conducted in collaboration with the National Tax College (NTC), under the approval of the National Tax Agency (NTA) (in February 2023), in accordance with "Guideline on the Utilization of National Tax Data in the Joint Statistical Research Program." The views expressed herein are those of the authors and do not necessarily reflect the views of NTC or NTA.

population reported in statistical surveys, and the total reported income is broadly consistent with the national accounts data. Since more than half of self-employed filers submit returns almost every year, the tax return data are highly valuable as a panel dataset for studying the self-employed.

This study also evaluates the tax return data as a source about information on top-income individuals. For the super-rich—those in the top 0.1 percent of the population—coverage is almost exhaustive. However, when extending the focus to the top 1 percent, the presence of individuals who do not file returns becomes non-negligible. Moreover, due to the Separate Withholding Taxation system on interest and dividend income, the coverage would be less complete, which likely results in an underestimation of top-income individuals.

Future challenges include the integration of tax return data with other datasets on wage earners and pensioners to construct a more comprehensive database of incomes. To understand the entire distribution of income, comparison between tax return data and existing statistical surveys is necessary.

Keywords: Tax Return Data, Income Data, High Income Individuals

JEL Classification: D31, E01, H2

I. Introduction

This paper investigates the characteristics of newly available income tax return data, which has recently become accessible for academic research, as a source of income statistics. Income represents rewards for contributing to production process through labor or asset ownership and is regarded as the right to access economic sources. Since income varies greatly across individuals, understanding who earns what kind of and how much income is indispensable for understanding the macroeconomy and for implementing appropriate policies. The purpose of this paper is to assess whether tax return data can serve as a reliable foundation for such information.

Traditionally, the analysis of income in economics has relied mainly on official statistics, which are derived from government surveys. At the macro level, the System of National Accounts provides a comprehensive picture of income generation and its distribution across institutional sectors, published as the Income and Outlay Accounts. However, these data are limited to sector-level information and provide little insight into individuals. At the household level, microdata from surveys such as the Family Income and Expenditure Survey, the National Survey of Family Income and Expenditure, the Comprehensive Survey of Living Conditions, the Basic Survey on Wage Structure, and the Survey on Private Sector Wages have been used.

Those data are collected under the Statistics Act, which imposes certain legal obliga-

tions, including penalties. Yet given the nature of information collection, enforcement cannot completely prevent misreporting or refusal to respond. In practice, survey implementation depends heavily on voluntary cooperation. In recent years, rising privacy concerns under the Act on the Protection of Personal Information have led to increased survey refusals, making data collection increasingly difficult not only for income but across official statistics more broadly.

To address this issue, administrative record data—particularly income tax return data—have attracted attention as a complement to, or substitute for, official statistics. Administrative records are collected and maintained by government agencies in the course of daily operations such as applications, registrations, and mandatory reporting. Because these records are closely linked to administrative processes, they provide rich information, are difficult to falsify or conceal, and are far less prone to non-response. Consequently, they may serve as more reliable data sources than surveys (Unayama, 2024).

One of the largest sources of “administrative record information” is tax return data. A tax return is a declaration filed by an individual to calculate and pay income tax on income. By submitting a tax return, individuals provide detailed information on their income, allowing for a comprehensive understanding of income levels.

Tax return records are regarded as a valuable source of information for analyzing the distribution of income in other countries too, and have already been widely used in research (Slemrod, 2016). For instance, Piketty (2003) and Garbinti et al. (2018) analyze changes in income inequality in France, while Piketty and Saez (2003) conduct a similar analysis for the United States. Their work demonstrates that the incomes of the very top earners constitute a large share of national income, thereby laying the foundation for subsequent studies of inequality.

In line with these international developments, researchers have tried to use tax return data in Japan. In the past, aggregated statistics based on such returns have been published through sources such as the *National Tax Agency Statistical Yearbook* and the *Sample Survey of Income Tax Returns*. Recently, however, individual-level tax return data have become available. This newly available data provide detailed information about sources of income and taxation, cover the top-income earners, and trace the same individuals for multiple years.

Despite its many advantages, tax return data also have several limitations. While individuals who are legally required to file are comprehensively covered, information on those without filing obligations is excluded. For example, wage earners employed by a single employer with annual incomes below 20 million yen settle their taxes through year-end adjustments and are not required to file a tax return (Article 121 of the Income Tax Act). Such individuals constitute a non-negligible share of the total population.

In addition, the information collected is limited to what is necessary for tax administration, which means that certain economically significant details, such as household composition, are missing. For instance, since the abolition of the dependent exemption for children under sixteen in 2011, it has no longer been possible to observe whether taxpayers have

children; and due to the abolition of the widower deduction and special widow deduction in 2020 (accompanied by the introduction of the single-parent deduction), the reporting of gender on tax forms was discontinued, making it impossible even to distinguish by gender.

In the following sections, this paper aims to clarify the boundaries between the information included in the tax return data and that which is excluded. Whether an individual is required to file a tax return is determined by the tax system and can be clearly delineated from the perspective of tax administration. However, the attributes of individuals who file returns are shaped by their income circumstances, making this an empirical issue.

In Section II, we organize the statutory prerequisites related to tax filing and confirm the types of individuals and sources of income included in the tax return data. We also present the concept of “fiscal income” as a measure of income that enables international comparisons along with its method of calculation. Section III analyzes which types of individuals file tax returns, using the return data. We highlight that many filers are wage earners subject to year-end adjustment who submit returns primarily to claim deductions such as medical expenses and the special mortgage interest deduction. Section IV examines the extent to which the tax return data capture the income of the household sector as measured in the System of National Accounts, and we conduct a comparison with macroeconomic data. Section V focuses on high-income individuals, analyzing the specific characteristics of the tax return data for this group and identifying the limits of coverage. Finally, Section VI summarizes the discussions in each section and outlines future research challenges.

II. Income and Tax Return

II-1. Tax Return Data as Income data

Ideally, income statistics should capture the overall picture of income. This requires identifying who receives income, when it is received, what type of income it is, and in what amount. Accordingly, to evaluate tax return data as a source of income information, we must first define what constitutes income. However, as Brooks (2018) notes, defining income poses a considerable challenge both theoretically and empirically.

The definition of income may vary depending on the intended purpose of the data. Even when income can be defined conceptually, whether it can be measured in practice remains a distinct issue. In this study, we adopt the definition of income provided by the System of National Accounts (SNA), the most widely used framework in economics, which is based on international standards established by the United Nations.

In the SNA, income is defined as what “accrue to institutional units as a consequence of their involvement in processes of production or ownership of assets that may be needed for purposes of production” (2008 SNA Manual, Chapter 7). In other words, it represents the right to dispose of newly created value added. A classical definition, attributed to Simons (1938), conceives income as “the sum of consumption and the change in net worth between the beginning and end of the period.” Since income should equal the sum of consumption

and saving, this appears consistent with the SNA definition. However, because some assets, such as equities, fluctuate in price independently of production, the two definitions diverge. Put differently, the adoption of the SNA definition here implies that capital gains and losses arising from asset price changes are excluded from income.

In the SNA, the aggregate income of a nation is referred to as Net National Income (NNI). It is calculated as gross domestic product (GDP)—the total value added produced domestically—minus capital consumption (depreciation), plus net income received from abroad. The practical methods of calculation are specified in the SNA manual, which provides the international standards for compiling GDP and related statistics, and classifies income according to the mode of contribution to production, such as compensation of employees, mixed income, and operating surplus. In Japan, these statistics are published by the Cabinet Office as the National Accounts (GDP statistics), and the total amount and distribution of NNI can be examined through a set of statistical tables known as the Income and Outlay Accounts.

The SNA assumes that NNI is initially distributed through the market mechanism according to contributions to production and then is re-distributed through government transfers and taxation, which determine the disposable income. The recipients of income are not necessarily individuals; corporations and other institutional units also earn income. One might argue that corporate income ultimately accrues to individuals. Nonetheless, in this paper we focus on the accounting distribution of income without imputing corporate income to individuals.

The SNA provides detailed information on the distribution of income across institutional sectors—households, corporations, government, non-profit organisation, and the rest of the world—thereby serving as an excellent source of macro-level income data. At the same time, it provides little insight into the distribution of income within sectors. For example, it does not capture disparities between high- and low-income households or between older and younger cohorts within the household sector. Thus, while the SNA excels at the macroeconomic level, it is insufficient for analyzing intra-sectoral income distributions.

To capture the economic conditions within each sector, data commonly referred to as micro statistics are typically employed. While the SNA is a “processed statistic,” constructed by compiling and adjusting information collected from various surveys, micro statistics are generated through direct surveys of economic agents. Examples include the *Family Income and Expenditure Survey* and the *Financial Statements Statistics of Corporations by Industry*. These surveys investigate the actual financial flows as recognized by individual economic units, using randomly sampled individuals or households to ensure representativeness at the national level.

The tax return data employed in this study are limited solely to individual income, and in that sense resemble micro statistics, as they allow for the observation of each filer’s income at the individual level. However, both the coverage of individuals and the scope of income differ between the two. The tax return data encompass only those individuals with a legal obligation to file, and the information captured is confined to what is necessary for tax-

ation purposes.

Because of this similarity in perspective, it is more straightforward to compare the tax return data with micro statistics than with the SNA. Nonetheless, both micro statistics and the tax return data have their own strengths and weaknesses, and a direct comparison does not necessarily yield a clear answer as to which provides the more desirable source of information. Accordingly, this paper focuses primarily on assessing the extent to which the tax return data reflect the macroeconomic situation.

II-2. Taxable Income and Tax Return

In Japan, taxpayers are required to calculate their own income and file a return for tax payment (Article 120 of the Income Tax Act). This filing constitutes the *Final Tax Return* (*Kakutei-Shinkoku*, in Japanese), and the aggregated information from submitted returns constitutes the tax return data. Since only taxable income can be observed in these data, understanding their nature requires an understanding of the mechanism of tax filing itself.

Under the Income Tax Act, in order to reflect differences in the nature of income and in tax-paying ability depending on how income is earned, income is classified into ten categories, as summarized in Table 1. Except for certain tax-exempt items, these categories of income are subject to taxation, and income tax is levied according to established rules. In principle, tax liability is determined based on an individual's *aggregate taxable income*—the total across all income categories for that person. This framework ensures fairness in tax burden and enables redistribution through the application of progressive tax rates. Taxing total income on a consolidated basis is referred to as the *Comprehensive Taxation system*.

There are, however, exceptions to this principle. To account for differences in tax-paying capacity by income type, certain portions of income are excluded from aggregation and taxed separately. This system, called *separate taxation*, applies different tax rates from those under Comprehensive Taxation. Separate taxation is further divided into two forms: *self-assessed separate taxation*, which requires a tax return, and *withholding separate taxation*, in which the payer withholds tax at source, thereby completing the tax obligation. Self-assessed separate taxation applies, for example, to retirement income, forest income, part of dividend income, part of capital gains, and some miscellaneous income. By contrast, a typical case of withholding separate taxation is interest income. For instance, interest on bank deposits is subject to withholding by the payer, which completes the tax procedure (Article 3 of the Special Taxation Measures Act). Income subject to withholding separate taxation is not reported on the Final Tax Return and thus cannot be observed in the tax return data.

Furthermore, even for income that falls under Comprehensive Taxation and is in principle subject to mandatory filing, certain conditions exempt the taxpayer from filing a final return (Article 121 of the Income Tax Act). Such exemptions apply to wage income, retirement income, and certain miscellaneous income related to public pensions. For example, wage earners who meet specified conditions are not required to file, since their employer (the withholding agent) completes the tax settlement through the year-end adjustment process.

Table 1. Categories of Income under the Income Tax Act

Type of Income	Article	Summary
Interest Income	Article 23	Interest from public bonds and deposits, distributions from joint investment trusts, bond investment trusts, and publicly offered bond-type investment trusts.
Dividend Income	Article 24	Dividends of surplus or profits received from corporations, distributions of surplus, monetary distributions from investment trusts and investment corporations, fund interest, and distributions from specified beneficiary certificate-issuing trusts.
Real Estate Income	Article 26	Income from real estate, rights pertaining to real estate, or the leasing of ships or aircraft (including the establishment of superficies or perpetual leases, and other arrangements allowing others to use real estate, etc.).
Business Income	Article 27	Income derived from agriculture, fisheries, manufacturing, wholesale, retail, services, and other businesses prescribed by government ordinance (excluding forest income or transfer income).
Employment Income	Article 28	Salaries, wages, remuneration, allowances, bonuses, and other similar forms of compensation.
Retirement Income	Article 30	Lump-sum payments such as retirement allowances, lump-sum pensions, and other forms of compensation received at the time of retirement.
Forest Income	Article 32	Income from logging or transferring forests (excluding cases where the forest is logged or transferred within five years of acquisition, which are not classified as forest income).
Transfer Income	Article 33	Income from the transfer of assets, excluding transfers of inventory assets and income from logging or transferring forests.
Occasional Income	Article 34	Temporary income other than that arising from continuous profit-making activities, and not in the nature of remuneration for labor, services, or consideration for the transfer of assets.
Miscellaneous Income	Article 35	Income not falling under any of the above categories.

Source: Author's compilation based on provisions of the Income Tax Act.

The rationale behind exempting these categories from filing is that the employer or payer already has full knowledge of the annual payment amount, enabling the application of progressive tax rates, while reliable tax collection can be secured through withholding at source. As a result, a substantial portion of income subject to these “no-return-required” provisions is not captured in the tax return data.

Based on the taxation and filing mechanisms stipulated in the Income Tax Act, incomes can be classified into four categories, as summarized in Table 2. The rows distinguish between income that is subject to mandatory filing and income that is not, while the columns distinguish between Comprehensive Taxation and separate taxation. However, depending on the detailed type of income, the taxpayer's choices, and the overall income situation, certain cases cannot be unambiguously assigned to a single category. In such instances, income is listed in multiple categories in parentheses to indicate possible classifications.

Broadly speaking, the category of income subject to Comprehensive Taxation with filing

Table 2. Classification of Income by Taxation System and Filing Obligations

	Comprehensive Taxation	Separate Taxation
Mandatory Filing	<ul style="list-style-type: none"> • Business Income • Real Estate Income • Dividend Income • Occasional Income • Transfer Income (excl. those from real estate or equity) • Employment Income (multiple employer or over 20 million yen) • Interest income • Miscellaneous 	[separate self-assessed taxation] <ul style="list-style-type: none"> • Transfer Income from real estate or equity, securities in general account, securities in Specified account without withholding tax • Forest Income • Interest Income • Dividend Income
No Filing	[year-end adjustment] <ul style="list-style-type: none"> • Employment Income (year-end adjustment applied) [no-return-required] <ul style="list-style-type: none"> • Miscellaneous Income (public pension under 4 million yen) 	[separate self-assessed taxation] <ul style="list-style-type: none"> • Transfer Income (in Specified Account with withholding tax) • Retirement Income [withholding separate taxation] <ul style="list-style-type: none"> • Interest Income • Dividend Income

Source: Author's compilation based on provisions of the Income Tax Act.

obligations (upper left) corresponds to the portion recorded in the “Form 1” of the income tax return, while income subject to Separate Self-Assessment Taxation (upper right) corresponds to the portion recorded in “Form 3.” By contrast, the income categories shown in the lower section represent income that cannot be observed in the tax return data.

Some of the income not captured in the tax return data can nonetheless be partially observed through information reported by the paying entities, such as withholding agents. For example, wage income is covered by the *Survey on Private Sector Wages* conducted by the National Tax Agency. This survey, designated as a Fundamental Statistical Survey under the Statistics Act, aims to capture the actual conditions of wage payments in the private sector. The survey collects information from withholding agents (e.g., firms) that pay wages, covering about 300,000 employees across roughly 20,000 establishments, and produces national estimates of wage conditions. Since it is based on employer-side data, however, it does not capture income at the individual level; for instance, individuals employed at multiple establishments are, in principle, double-counted. Accordingly, any joint analysis using the tax return data and this survey requires appropriate adjustments.

II-3. Taxable Income and Fiscal Income

The contents of tax return data consist primarily of information provided for taxation purpose. In contrast, micro statistics such as household surveys are designed to collect information that is important from social, economic, and administrative perspectives. While tax return data do not include general information irrelevant to taxation, they yield more accurate and detailed information since tax returns are subject to significant penalties for misre-

porting or non-filing.

In the tax return, the most important measure of income is taxable income, which forms the basis for calculating income tax. Taxable income is calculated by subtracting necessary expenses and various income deductions from gross receipts. Gross receipts and necessary expenses are calculated separately for each type of income, and the totals are aggregated to determine taxable income.

Gross receipts include all economic gains received by an individual within a given period, though their scope varies depending on the type of income. For example, in the case of employment income, they correspond to the “amount paid” listed on the withholding statement; for business income, they are the gross revenues from sales or services; and for real estate income, they consist of rental income such as rents. In what follows, the amounts received before subtracting necessary expenses will be referred to simply as “gross receipts.”

Necessary expenses consist of expenditures incurred in the process of earning income. For example, in the case of business income, these include costs of goods purchased, personnel expenses, taxes and dues, and depreciation. For real estate income, property taxes and repair/maintenance costs are included. The residual after deducting such expenses constitutes the “income” for each income category. In the available Forms 1 and 3 of the tax return, both gross receipts and income amounts are recorded, but information on necessary expenses is not directly observable.

For income subject to Comprehensive Taxation, the sum of income amounts across different categories is reported as the “total income amount” in the “Income Amounts, etc.” column of Form 1. From this total, income deductions such as the basic deduction, spousal deduction, dependent deduction, social insurance premium deduction, medical expense deduction, and charitable contribution deduction are subtracted to yield taxable income. For income subject to Separate Self-Assessment Taxation, taxable income is reported for each income category under the “Calculation of Tax” section of Form 3.

Although taxable income is an important measure for tax administration, it is not necessarily suitable as a general measure of income. The system of deductions reflects social circumstances and economic conditions, aiming to align tax burdens with taxpayers’ ability to pay, and is also shaped by historical and political factors. Moreover, deduction systems differ substantially across countries, making taxable income a difficult variable for international comparison.

For this reason, previous studies using tax return data often focus on the concept of fiscal income, defined as “the total amount of income subject to taxation” before applying any deductions. Fiscal income is considered preferable not only because it serves as a measure of “market-determined income,” reflecting contributions to production, but also because it provides a more internationally comparable measure. In the international project Distributional National Accounts (DINA), which compares income inequality across countries using tax return data, fiscal income has been adopted as the baseline measure.

From this perspective, the “total income amount” reported in Form 1 might seem to serve as a measure of income before the application of deductions. However, it is insuffi-

cient as a measure of fiscal income. One reason is that, in certain categories of income, deductions beyond necessary expenses are already applied in the calculation of income. The most notable examples are the employment income deduction and the public pension deduction. Similarly, for occasional income, the reported income is calculated as half the amount remaining after subtracting 500,000 yen (and necessary expenses) from the gross receipt. To construct internationally comparable fiscal income, it is necessary to use income amounts prior to the application of such deductions.

Even with these adjustments, however, only a portion of “the total amount of income subject to taxation” can be captured. The “total income amount” in Form 1 aggregates only the income subject to Comprehensive Taxation—that is, only the upper-left portion of Table 2. To broaden coverage, this study also incorporates income subject to Separate Self-Assessment Taxation. Nevertheless, income not requiring tax returns, as well as income subject to Separate Withholding Taxation, cannot be observed. In this sense, fiscal income is systematically underestimated in the tax return data.

By definition, fiscal income includes all types of income recognized under the Income Tax Act. However, this definition also encompasses Transfer Income, which is generally excluded from the economic concept of income. Transfer Income, which is generated by the sale of assets such as land or stocks, corresponding to what economics calls capital gains.

In economics, income is usually understood as a flow derived from economic activity, and capital gains are therefore not included. In principle, capital gains should accrue annually through market valuation, but under the tax system they are recognized as income only at the time of asset transfer. Taking these differences into account, this paper defines fiscal income as the total taxable income excluding transfer income (both Comprehensive and Separate Self-Assessment Taxation categories).

III. Population Coverage of the Tax Return Data

III-1. Share for Tax Return Filers

As noted above, individuals who earn income are not necessarily required to file a tax return in Japan. For most of employees, tax obligations are fully settled through procedures such as Year-End Tax Adjustment (Nenmatsu Chosei, in Japanese), making tax filing unnecessary. Therefore, before examining the nature of incomes reported through tax returns, it is useful to observe how common tax filing is in Japan.

Table 3 presents the number of tax return filings from 2014 to 2020. According to “A. Number of Tax Filers,” the total number of returns increased steadily from 22.69 million in 2014 to 23.54 million in 2019, but declined slightly to 22.96 million in 2020, when the COVID-19 pandemic began. This figure amounts to only about one-sixth of the total population each year shown in “D. Total Population.” According to the Labor Force Survey, the labor force during this period was around 69 million, with 64 to 67 million employed individuals. Since most of these individuals are assumed to have earned income, the data imply

that only a fraction of income earners file tax returns.

The filings included in tax return data are not limited to the ordinary Final Tax Returns, which are submitted by March each year to report annual income of the previous year. Other categories also exist, such as “Quasi-Final Tax Returns” (Jun Kakutei Shinkoku) or “Retirement Income Tax Return” (Taishoku Shotoku Shinkoku). A Quasi-Final Tax Returns is required when the taxpayer is absent from Japan at the time of filing, due to reasons such as emigration or death. A Retirement Income Tax Return applies when one receives retirement allowances¹. According to Table 3, however, such non-ordinary returns are negligible in number compared to ordinary filings.

The situation regarding tax filing in Japan differs substantially from that in the United States and France, which have been used in previous studies such as Piketty (2003), Piketty and Saez (2003), and Auten and Splinter (2024), in two important respects.

First, the proportion of individuals who file tax returns is significantly lower in Japan. In the United States and France, almost everyone with income above a certain threshold are, in principle, required to file tax returns. As noted above, however, Japan operates a Year-End Tax Adjustment system, under which tax filing is unnecessary in many cases; this institutional feature accounts for much of the observed difference. For example, the share of households that filed tax returns in the United States in 2020 is estimated to have been 82 percent, which is approximately four times the corresponding figure in Japan. In addition, Piketty (2003) reports that the proportion of taxpayers filing returns in France is in the range of 50 to 60 percent.

The United Kingdom, which has a system similar to Year-End Tax Adjustment called Pay As You Earn (PAYE), exhibits a filing rate of approximately 20 percent of the adult population, indicating that Japan is not necessarily an exceptional case in this respect. Nevertheless, because the coverage of data based on tax filings is strongly determined by the structure of the tax system, it is important to note that, even with the tax return data here, the methodologies employed in the previous studies cannot be directly applicable to the Japanese context.

The second major difference from previous studies using tax return data lies in the unit of taxation, or the tax filing unit. In Japan, income tax returns are filed on an individual basis. By contrast, in many countries—such as the United States, where couples may file jointly under the Married Filing Jointly system—tax returns are filed on a household basis, combining spouses and dependents into a single unit. This means that information is also available for spouses and dependent children; even if they have no earnings, they can be explicitly identified as having zero income. As a result, the overall distribution of income across the economy is more comprehensively observable.

¹ Retirement Income Tax Returns are one of optional tax files. When a non-resident individual receives retirement allowances in Japan that constitute Japan-source income, which is subject to withholding tax at a rate of 20.42%, he/she can elect this optional filing to compute the income tax liability based on the same rules with a resident individual.

² The dependent deduction can be applicable when the individual's total annual income does not exceed 480,000 yen (or 380,000 yen for tax years prior to 2019).

On the other hand, in Japan, all tax returns are filed on an individual basis. If a spouse or child has taxable income, they must file a separate individual return. However, for relatives without income, it is possible to estimate the number of individuals with very low income by examining the application status of the dependent deduction.²

In the tax return data, the number of dependents is not directly available, but, since the deduction amount per dependent is fixed (though varying by age and relationship), the total number of dependents can be calculated from the aggregate deduction claimed. In addition, because the dependent deduction does not apply to relatives under the age of 15, the under-15 population can also be counted as part of the dependents.

This estimate is reported in Table 3, row “B. dependents,” which averages around 5 million. Since 2014, the beginning of the data period, the number of dependents has shown a downward trend, reflecting factors such as declining household size and rising female labor force participation. Meanwhile, the estimated population under age 15, based on Population Estimates, is shown in row “C. Estimated population under 15,” amounting to between 15 and 16 million.

If we count both dependents and this under 15 population as individuals with “very low income,” then, combined with the tax filers themselves, the data cover roughly 44 million individuals. This corresponds to about 35 percent of the “D. Total Population,” meaning that the incomes of two-thirds of the population are not observed in the tax return data. From a population perspective, then, unlike in countries such as the U.S. or France used in earlier studies, the Japanese data are not well suited for capturing the entire national income distribution.

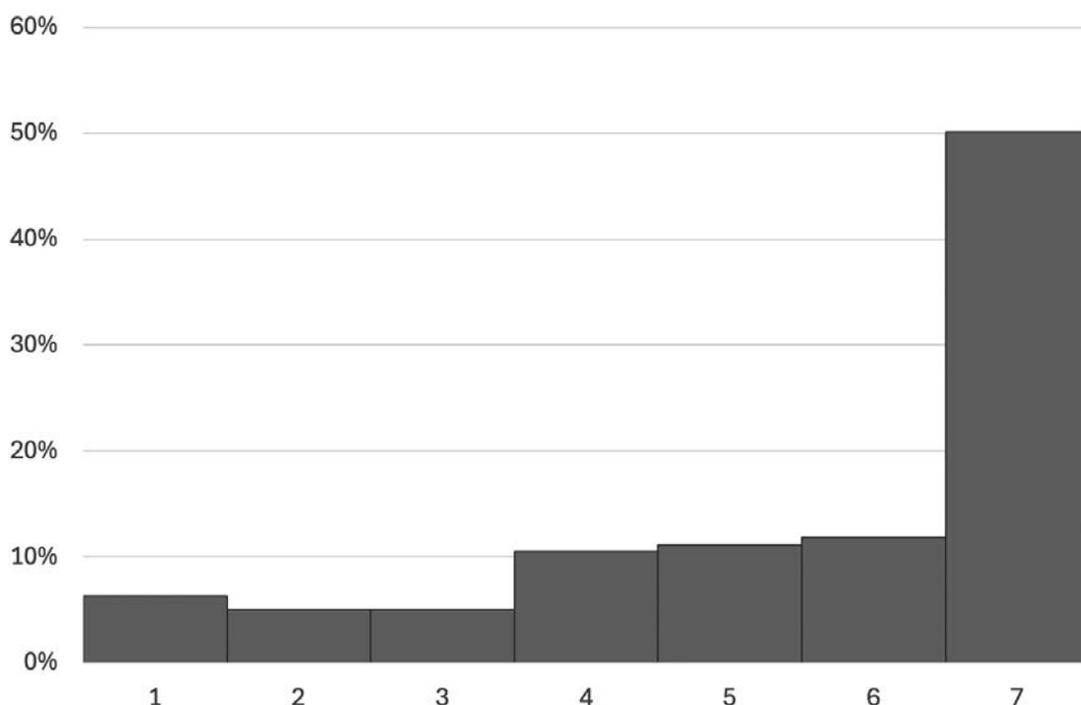
With a panel context, the fact that only about 15 percent of the population files a tax return in any given year does not mean that the remaining 85 percent never appear in the data. For example, if different individuals file in different years, then pooling data across seven years might result in coverage of nearly the entire population at least once.

Table 3. Number of Tax Filers (1,000 people)

	2014	2015	2016	2017	2018	2019	2020
A. Number of Tax Filers	22,688	22,837	23,007	23,254	23,398	23,538	22,959
Quasi-Final Tax Returns							
Due to Emigration	4	4	3	4	4	4	3
Due to Death	155	154	152	151	152	148	145
Retirement Income Tax Returns	57	51	51	52	59	76	86
B. Dependents	5,654	5,483	5,161	5,023	4,879	4,716	4,510
C. Population under age 15+	16,233	15,945	15,780	15,592	15,414	15,210	15,031
D. Total Population	127,083	127,095	126,933	126,706	126,443	126,167	126,146
E. Coverage (= (A+B+C) / D)	35.1%	34.8%	34.6%	34.6%	34.6%	34.4%	33.7%

Source: Authors' calculation based on the Tax Return Data (A-B), and Population Estimates compiled by Statistics Bureau of Japan (C-D).

Figure 1. Distribution of Number of Tax Returns per Individual



Source: Authors' calculation based on the tax return data. Among those who file in 2017.

To investigate this possibility, Figure 1 counts how many times each filed during the 2014-2020 period. Each tax return record contains a unique identifier for the filer, allowing linkage across years. Using this panel structure, we construct a histogram of the number of years filed within the seven-year span.

Among the 23.25 million individuals who filed in 2017 (the central year of the period), the distribution shows that about half filed in all seven years, while about 10 percent filed in four, five, or six years. By contrast, only about 7 percent filed in just one year. This pattern indicates that those who file tax returns generally do so almost every year, while those who do not file are largely absent across the entire period. That is, even by accumulating tax return data over time, the data do not represent a situation in which the entire population appears in the data at least once at some point; rather, approximately 10 percent of the population is repeatedly observed in the tax filing data. This finding implies that the strength of tax return data as income data lies not in providing comprehensive coverage of the entire population, but rather in enabling panel analysis of the same individuals over time. For more effective use as a panel data, it is essential to clarify which segment of the Japanese population the dataset represents.

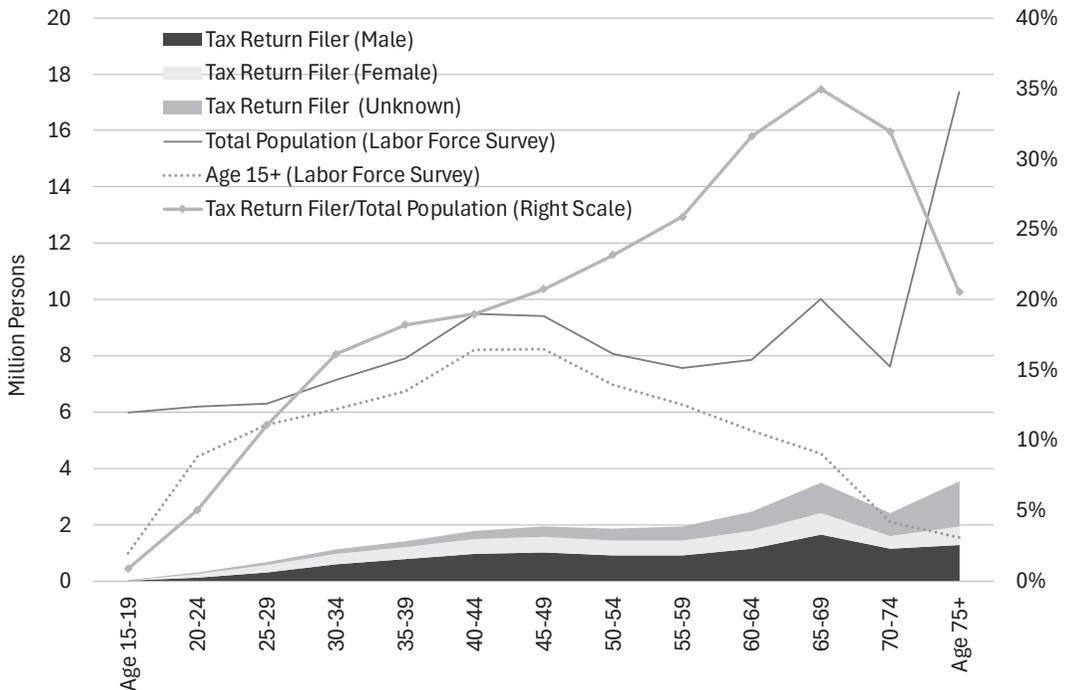
III-2. Characteristics of Tax Return Filers

Up to this point, we have seen that tax return data can reasonably be regarded as panel data on the incomes of roughly 10 percent of the population. Whether individuals file a tax return depends on their income structure and on institutional factors such as the Year-End Tax Adjustment. Unlike statistical surveys based on random sampling, the individuals included in this dataset constitute a selective group, and the data do not represent the Japanese population. In what follows, we evaluate tax return data by examining the attributes of the population captured within it.

Figure 2 plots the number of tax filers by sex and age, using 2017 data as in Figure 1. Age is calculated from the date of birth field on the tax return, which is almost universally completed, leaving virtually no missing cases. By contrast, although tax returns included a field for sex through 2019, around 30 percent of filers are coded as “unknown,” mainly due to missing entries. Since income levels are known to differ substantially by sex, this large share of missing data—and the fact that the sex field has since been discontinued—represents a serious limitation of the tax return dataset.

By age, a disproportionately large share of filers are elderly, aged 65 and above. Figure 2 also shows the total population from the Labor Force Survey. While Japan’s aging demo-

Figure 2. Demographics of Tax Return Filers



Source: Authors’ calculation based on the tax return data. Among those who file in 2017. Total Population and Labor Force Population are based on the *Labor Force Survey*.

graphics mean that older cohorts are relatively large, the differences across age groups in the total population are not as pronounced as in the filer distribution, where the elderly are clearly overrepresented. To account for demographic structure, the figure also reports filing rates by age group relative to population. These rates peak at ages 65-69, where nearly 35 percent of individuals file tax returns.

Even though labor is the primary source of income for most individuals, the bulk of tax filers are concentrated in their 50s and older, when labor force participation has already peaked or begun to decline. In this sense, the tax return data are better understood not as a dataset for analyzing labor income, but rather as one that is particularly well suited for capturing the income dynamics of the elderly.

Looking at the age distribution by sex, among filers whose sex is identifiable, the share of men is almost constant across age groups. Overall, excluding those with unknown sex, men account for 66 percent of filers. Given that in Japan roughly 80 percent of household heads are male (rising to about 90 percent among multi-person households), this implies that a non-negligible portion of tax filers are not household heads. If multiple members of the same household file returns, then interpreting the distribution of individual incomes as equivalent to household income distributions becomes problematic.

Moreover, the proportion of individuals with unknown sex rises with age, precisely in the older cohorts where filers are most prevalent, making it especially difficult to determine whether elderly filers are household heads. For these reasons, linking tax return data with household-based surveys such as the Family Income and Expenditure Survey or the National Survey of Family Income and Expenditure is far from straightforward.

In addition to age and sex, another attribute available in tax return data is “occupation.”³ The occupation field is intended primarily for reporting the type of business in the case of the self-employed, and while it is recorded in free-text form, it is subject to some degree of categorization. However, this classification system differs from the standard occupational classifications used in statistical surveys, and occupational categories are not mutually exclusive (e.g., “university professor” can be “civil servant” or “company employee”). While it provides some information on industry, it is difficult to use comprehensively as a standardized attribute.

As an alternative, we classify filers here according to their primary source of income as “income source category.” Classification by sources of income is already employed in the National Tax Agency’s *Sample Survey of Income Tax Returns*. Panel (a) of Table 4 shows the classification method used there: filers whose business income accounts for the majority of their total income are designated “Business Income Earners” (roughly corresponding to the self-employed), while the remainder are further classified as “Real Estate Income Earners,” “Wage Earners,” “Miscellaneous Income Earners,” and “Others.” However, the *Sample Survey of Income Tax Returns* is limited to individuals with positive tax liabilities, thereby

³ The tax return form includes fields for personal information such as name, address, and telephone number; however, such personal information is not included in the tax return data that are provided to researchers.

Table 4. Income Source Group Classification

(a) Classifications used in the Sample Survey of Income Tax Returns	
Business Income Earner	Those whose Business Income is greater than the sum of all other income
Non-Business Income Earner	Others who have tax liabilities
Real Estate Income Earner	Real Estate Income is greater than other income
Wage Earner	Employment Income is greater than other income
Miscellaneous Income Earner	Miscellaneous Income is greater than other income
Others	All others
(b) Classifications based on the Fiscal Income	
Business Income Earner	Those whose Modified Business Income is greater than the sum of all other income
Real Estate Income Earner	Real Estate Income is greater than other income except for business income
Capital Income Earner	Capital Income is greater than other income except for business income
Wage Earner	Modified Employment Income is greater than other income except for business income
Pensioner	Modified Pension Income is greater than other income except for business income
Miscellaneous Income Earner	Modified Miscellaneous Income is greater than other income except for business income
Others	All others

- Modified Business Income = Business Income + Agriculture Income + Forest Income
- Real Estate Income = Real Estate Income
- Capital Income = Gross Interest Income + Gross Dividend Income (including separate taxation part)
- Modified Employment Income = Gross Employment Income + Gross Retirement Income
- Modified Pension Income = Gross Pension Income
- Modified Miscellaneous Income = Miscellaneous Income - Pension Income

Source: Panel (a) Sample Survey of Income Tax Returns, Panel (b) authors' definition.

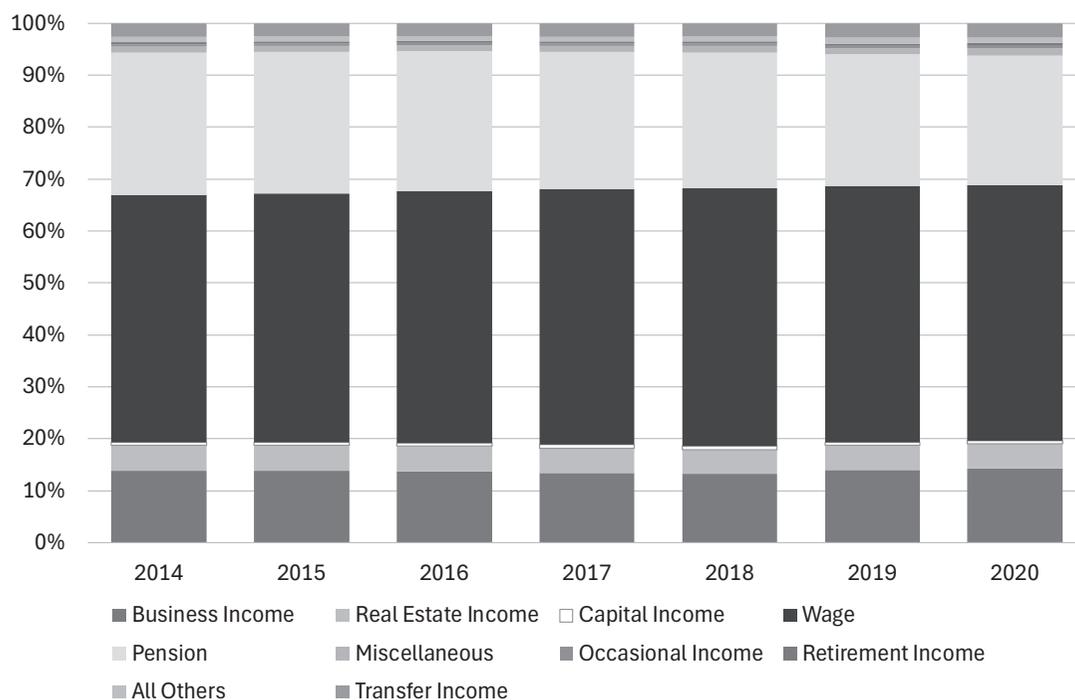
excluding those with income but no net tax liability (e.g., refund filers).

In this study, we extend the method to cover all tax filers. Drawing on the classification in the survey as a reference, we develop an independent set of criteria, shown in Panel (b) of Table 4. Like Panel (a), our approach relies on the share of each income source in total income, but it differs in two respects: (1) in the definition of total income and of income by source, and (2) in the degree of classification detail.

For income, we adopt the concept of fiscal income, defined on the basis of gross receipts before deductions. Specifically: (1) we include both the “total income amount” subject to Comprehensive Taxation and the income subject to Separate Withholding Taxation; (2) we use gross receipts before applying the wage income deduction or the public pension deduction for Employment Income and Pension Income; and (3) we exclude capital gains (Transfer Income) from total income. On this basis, we refine the classification used in the *Sample Survey of Income Tax Returns*, adding categories such as “Capital Income Earners,” for those whose largest share of income is interest or dividends, and “Pensioners,” for those whose main income source is public pensions.

Figure 3 shows the classification of filers by year. The largest group consists of Wage Earners, who account for about 50 percent of filers, or roughly 11 million individuals. The

Figure 3. Shares of Tax Return Filers by Main Income Source



Source: Authors' calculation based on the tax return data. Among those who file in 2017.

next largest group is Pensioners, at nearly 30 percent, or about 6 million individuals. Business Income Earners follow at around 15 percent (3.2 million), and real estate income earners at about 5 percent (1.1 million). Almost all filers fall into one of these four categories, making the classification essentially the same as in the Sample Survey of Income Tax Returns.

According to the *Labor Force Survey*, the number of employees ranged from about 55 to 60 million between 2014 and 2020, while the number of wage earners covered in the *Survey on Private Sector Wages* is at a similar level. The wage earners identified here are defined somewhat more narrowly as those whose primary income source is wages, but even so, a simple comparison suggests that fewer than 20 percent of employees file tax returns.

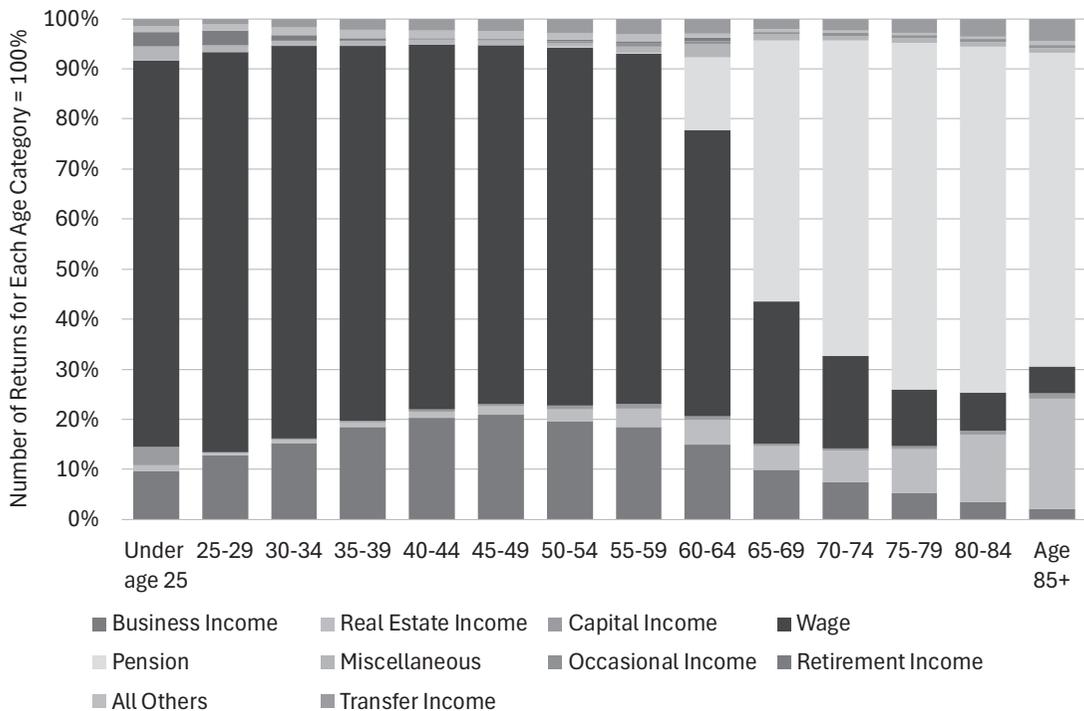
Similarly, according to the *Overview of the Employees' Pension Insurance and National Pension Programs*, the number of public pension recipients during the same period was about 40 million, whereas the number of pension income earners identified here is around 6 million (slightly declining to 5.6 million in 2020). Thus, only about 15 percent of pension recipients file tax returns.

Individuals whose sole income sources are wage or public pension typically do not have tax return obligations. Their tax processes are settled through Year-End Adjustment (or similar system for pensioners). In that sense, it is natural that the share of filers is limited to around 20 percent.

By contrast, individuals whose primary source of income is business income—where filing a tax return is in principle mandatory—should exhibit relatively high filing rates. According to the *Labor Force Survey*, the population of “self-employed workers” was 5.56 million in 2014 and 5.26 million in 2020. On the other hand, the “self-employed” category in our classification, which is defined as the sum of Business Income Earners and Real Estate Income Earners are about 4.3-4.4 million (since it would be possible that Real Estate Income Earners are regarded as a self-employed in the *Labor Force Survey*). This implies that roughly 80 percent of the self-employed files tax returns. Even among the self-employed, there are cases in which filing is not required—for example, when businesses operate at a loss. This suggests that self-employed individuals who file and those who do not may differ systematically in important ways, potentially limiting comparability between the two groups.

These income source group classifications also vary significantly by age. Figure 4, based on 2017 tax return data, shows the share of each group by age. Wage Earners, who make up half of all filers, are concentrated overwhelmingly among those under 65, accounting for about 70 percent of filers in the younger cohorts. By contrast, Pensioners are concentrated among those 65 and older, comprising about 70 percent of filers aged 70 and above. Given that many firms in Japan adopt a mandatory retirement age of 65 and that public pension benefits typically begin at age 65, this distribution is quite understandable. Business Income

Figure 4. Income Source Groups by Age



Source: Authors' calculation based on the tax return data. Among those who file in 2017.

Earners display a hump-shaped profile, peaking around age 50, which is consistent with findings from the *Labor Force Survey* and other statistics.

Another notable feature is the rising share of Real Estate Income Earners among older age groups. Earning Real Estate Income requires ownership of property, and it is well established that asset holdings increase with age. This observation, therefore, is a natural consequence of demographic and wealth patterns. While studies such as Piketty (2003) and Piketty and Saez (2013) have emphasized the significant role of real estate income in driving inequality in other countries, the evidence here suggests that its influence in Japan is relatively limited.

III-3. Reasons for Tax Return

We now turn to the contents of the returns themselves to examine the reasons why individuals file tax returns, thereby clarifying the nature of tax return data as income data. As already noted, the obligation to file is determined under the Income Tax Act. Thus, by observing the types of income reported and the claims for income and tax deductions, we can infer

Table 5. Tax Return Items and Income Source Group

	All Filers (a)	Business (b)	Real Estate (c)	Wage (d)	Pensioners (e)
Business Income > 0	19%	100%	6%	4%	9%
Real Estate Income > 0	14%	3%	100%	10%	13%
Capital Income > 0	8%	3%	10%	6%	9%
Employment Income > 0	61%	14%	28%	100%	29%
Other Income > 200,000 yen	---	---	---	23%	---
Empl. Income > 20 mil yen	1%	0%	0%	3%	0%
Pension Income > 0	27%	12%	43%	10%	100%
Other Income > 200,000 yen	---	---	---	---	33%
Pension Income > 4 mil yen	2%	0%	0%	0%	6%
Other Misc. Income > 0	17%	7%	21%	11%	31%
Any Personal Deduction > 0	11%	9%	29%	8%	16%
Medical Expense Ded. > 0	33%	19%	42%	34%	42%
Charitable Contrib. Ded. > 0	10%	6%	9%	14%	5%
Any Tax Credit for Housing/Mortgage > 0	8%	7%	2%	14%	1%
Tax Credit for Political Contributions > 0	2%	1%	3%	2%	2%

Source: Authors' calculation based on the tax return data. Average for 2014-2020. Personal Deduction includes those for Spouse and Dependents. Housing/Mortgage related tax credits are the special tax credit for housing loan and earthquake-resistant housing improvements.

the reasons for filing.

Table 5 summarizes the items reported in tax returns. For example, among all filers (column a), 61 percent reported employment income. As shown in Figure 3, those whose primary source of income is Employment Income (i.e., Wage Earners) accounted for about 50 percent, but an additional 10 percent reported Employment Income as a secondary source. Even taking these cases into consideration, our conclusion is not materially changed that about 20 percent of employees in the *Labor Force Survey* file returns.

By income sources, Business Income and Real Estate Income generally require tax filing. Across all filers, the shares who report them are 19 percent and 14 percent, respectively—figures not very different from the overall proportions of the self-employed. Among Wage Earners, only 4 percent also report business income, indicating that multiple sources of income are relatively uncommon.

Focusing on Wage Earners (column d), 23 percent report additional income exceeding 200,000 yen, and only 3 percent report employment income exceeding 20 million, the threshold that triggers mandatory filing. While some may have Employment Income from multiple employers, the reported income alone does not make clear why filing was necessary.

Among Pensioners (column e), 33 percent report additional income, and 6 percent report pension income exceeding 4 million yen, which requires filing. As noted earlier, many elderly filers also have real estate income, suggesting that they are more likely than Wage Earners to be in situations requiring tax filing. Nonetheless, for more than 60 percent of Pensioners, their income alone does not explain the need to file.

Overall, about 70 percent of filers are Wage Earners or Pensioners, groups for whom the need to file seems relatively weak. The content of the returns suggests that, in many such cases, filing is driven not by income conditions per se, but by the use of deductions. Tax liability is determined by total income minus income deductions, and the final tax due is further reduced by tax credits, many of which can only be applied through filing. Typical examples include the Medical Expense Deduction, the Special Deduction for Housing Loan Interest, and the Charitable Contribution Deduction.

The Medical Expense Deduction can be applicable when medical expenditures paid for oneself, a spouse, or other dependents exceed a certain threshold, but a tax return filing is always required regardless of the source of income. According to Table 5, 33 percent of all filers, 34 percent of Wage Earners, and 42 percent of Pensioners claimed this deduction. In other words, at least one-third of Wage Earners and two-fifths of Pensioners file tax returns for the purpose of claiming the Medical Expense Deduction.

Similarly, the proportion claiming housing-related tax credits such as the Special Tax Credits for Housing Loan Interest is 8 percent among all filers and 14 percent among Wage Earners. This credit, commonly known as the mortgage deduction, allows taxpayers who purchase, newly construct, or renovate a home and occupy it to deduct from their tax liability an amount proportional to the outstanding loan balance. In the first year of claiming the credit, a tax return must be filed. Thus, for Wage Earners in particular, eligibility for hous-

ing-related tax credits constitutes a non-negligible reason for filing.

These findings suggest that a substantial share of returns are filed in order to claim deductions. Evidence from actual tax outcomes supports this interpretation. Based on the filing results, taxpayers fall into one of three categories: those with additional tax due (“tax payable”), those entitled to refunds (“refund”), and those with neither. The “tax payable” cases arise when withholding falls short of total liability, while the “refund” cases reflect overpayment, often due to withholding. Roughly 60 percent of all returns result in refunds; and among Wage Earners and Pensioners, the proportion is closer to 70 percent. Although refunds cannot be attributed solely to deductions, the high incidence suggests that many filers in these groups are in special situations—such as incurring large medical expenses or taking out a new mortgage—that necessitate filing to claim deductions and obtain refunds.

IV. Income Coverage of the Tax Return Data

IV-1. National Income and Fiscal Income

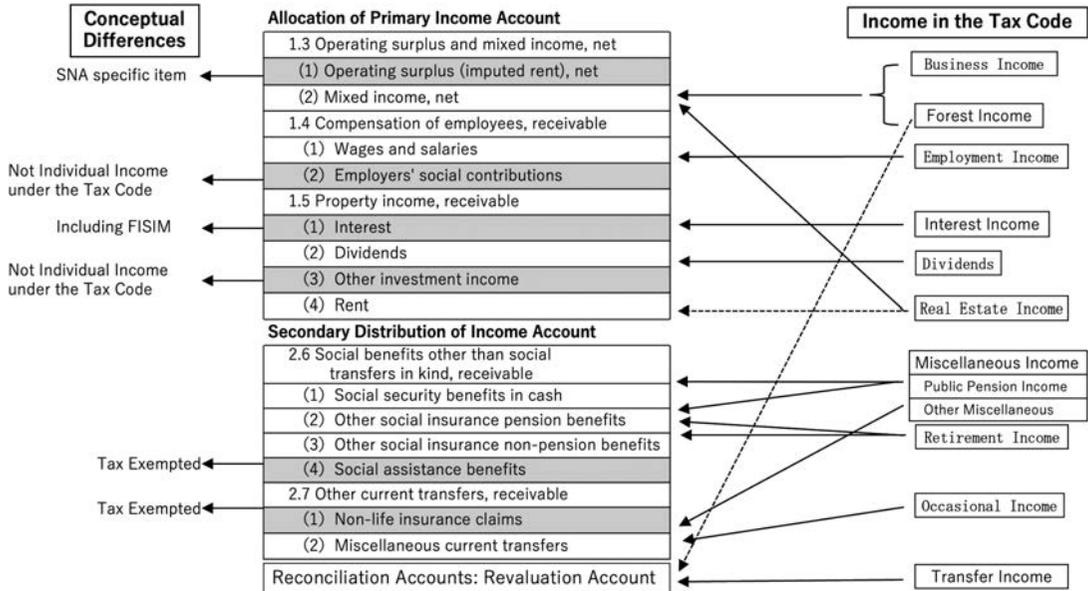
So far, we have examined what type of individuals are included in tax return data, and we have seen that coverage of the data is relatively limited in terms of number of persons. We have also confirmed that, unlike statistical surveys, the individuals captured in the dataset are systematically selective. However, this does not necessarily mean that tax return data cannot be used to study income at the national level. Precisely because the dataset is skewed toward higher-income individuals, it may still cover a substantial share of aggregate income. This section therefore investigates the share of national income that is captured in the tax return data.

To clarify this coverage, we must first establish a benchmark measure of “total income.” Here, we focus on national income as reported in the System of National Accounts (SNA), which comprehensively measures overall economic activities. In addition, to limit the scope to incomes relevant to the individuals included in the tax return data, we focus on receipts in the household sector as the relevant institutional unit.

Although the economic value received by corporations ultimately accrues to individuals, from the perspective of constructing nationwide income statistics, retained earnings held within corporations are not observed in individual tax returns and are therefore excluded from tax return data in general. Restricting the scope to household income makes it possible to compare incomes on a consistent basis with the tax return data.

The Income and Outlay accounts in the SNA track flows from the stage at which income is generated to the stage at which it is spent, and are disaggregated into sub-accounts that describe the economic activity of each institutional sector. In this paper, we define household income as the sum of “receipts” recorded in two sub-accounts of Income and Outlay accounts—the Allocation of Primary Income Account and the Secondary Distribution of Income Account. The former records income distributed as a result of participation in production processes or the ownership of assets needed for production purposes, while the lat-

Figure 5. Income Source and the SNA Classifications



Source: Authors' classification based on the SNA manual and the Tax

ter records redistribution through taxes, social security, and other transfer mechanisms.

Figure 5 summarizes the items recorded in the income and outlay accounts of the SNA and their correspondence to tax return data. While the SNA captures, in principle, all flows that can be regarded as income in the economic sense, some of these are not included in tax return data. We therefore begin by clarifying the scope of income covered in each dataset.

A typical example of personal income recorded in the SNA but not recognized under the tax code is the Imputed Rent from owner-occupied housing. This is recorded under “1.3 Operating Surplus and Mixed Income” as “(1) Operating Surplus (Owner-Occupied Dwellings)” in the Allocation of Primary Income Account. It represents the implicit return from owning one’s own home and can be viewed as a form of property income. Although it is unquestionably an economic form of income, it is only an imputation of hypothetical rent payments and receipts. Thus, it does not appear in tax return data.

Similarly, in the case of interest income, the SNA incorporates adjustments for Financial Intermediation Services Indirectly Measured (FISIM).⁴ In the Income and Outlay accounts, “interest” within property income (receipts) includes these FISIM amounts, which do not correspond to actual cash flows. Since the tax return data do not contain such adjustments, we use pre-FISIM values for comparability.

Differences also arise from accounting conventions rather than imputed items. A prominent example is Employers’ Social Contributions. In Japan, social insurance premiums are split equally between employers and employees. The employer’s portion is included in

⁴ See https://www.esri.cao.go.jp/jp/sna/data/data_list/kakuhou/files/h21/pdf/fisim.pdf, for the discussion about FISIM.

payslips and is not treated as taxable income. In the SNA, however, it is recorded under “1.4 Compensation of Employees (Receipts)” as “(2) Employers’ Social Contributions” to the household sector. From an economic perspective, this is appropriate, as the employer is effectively prepaying part of the employee’s obligation. For comparability with the tax return data, however, these amounts are excluded in our calculations.

Another exclusion is “(3) Other Investment Income” under “1.5 Property Income,” which records property income arising from assets managed by financial institutions on behalf of households. For tax purposes, this is recognized only when it is actually paid out (e.g., at maturity), creating timing differences, so we exclude it from the comparison.

Some categories are excluded entirely because they are non-taxable and therefore never observed in the tax return data. These include: “(3) Other Non-Pension Social Insurance Benefits” and “(4) Social Assistance Benefits” under “2.6 Social Benefits Other Than Social Transfers in Kind,” as well as “(1) Non-Life Insurance Claims” under “2.7 Other Current Transfers (Receipts).” These correspond to benefits such as maternity allowances and child-care leave benefits, Social Assistance Benefits (public assistance), and insurance payouts from fire or automobile insurance. Since they are tax-exempt, they do not appear in the tax return data.

For the tax return data to serve as comprehensive national income statistics, it would be desirable to supplement these “invisible incomes” with additional sources. However, methods to integrate such data at the micro level have not been firmly established, even in international projects such as DINA, and remain a subject of debate (See, for example, Auten and Splinter, 2024). Thus, a priority is first to assess to what extent the tax return data covers national income in quantitative terms.

As discussed above, this paper defines fiscal income as excluding capital gains (Transfer Income). This aligns with the treatment in the SNA, where capital gains are recorded in the balance sheet adjustment account as valuation changes, rather than as income, which is limited to returns to production or transfers across institutional sectors.

In principle, one could compare the two datasets including capital gains. However, since the SNA records unrealized gains on a mark-to-market basis when they occur, while the tax return data record realized gains only upon asset transfer, harmonization is not feasible. Accordingly, we exclude capital gains when comparing coverage.

When comparing tax return data with the SNA, it is preferable not only to look at aggregates but also to match categories where possible. Although the classifications differ because the datasets are constructed for different purposes, some mapping is feasible. On the right side of Figure 5, we illustrate how items in tax returns correspond to categories in the SNA. For example, “Business Income” in the tax returns corresponds broadly to “Mixed Income” in the SNA. Mixed Income refers to earnings of the self-employed, combining returns to both labor and capital, which cannot be easily separated. All Business Income can be considered part of Mixed Income, though Mixed Income may also include portions of Forest Income under the tax code.

Under the Income Tax Act, Forest Income arises from sales of timber harvested from

one's own forest or from transfers of forestland. It includes both elements similar to Business Income and elements akin to capital gains. In the SNA, these are distinguished: income from selling timber grown and harvested by the owner (net of costs) is treated as Mixed Income, while sales of forestland are treated as capital transactions, not income. Here, we treat Forest Income as part of Mixed Income, even though this may include the capital gains part.

For "Employment Income," "Interest Income," and "Dividend Income," the correspondence is straightforward. Employment Income maps to Wages and Salaries. Although Employers' Social Contributions are also included in Compensation of Employees in the SNA, they are separately identified and can be excluded. Interest and dividends likewise correspond directly to Property Income categories in the SNA.

More complex are "Real Estate Income," which consists of rent for land, buildings and other estate. If rental activities are conducted at a scale considered a business, they are classified as business income under the tax code. In the SNA, however, rental income is generally treated as mixed income (e.g., "housing rental services"), while land rent (payments for leasehold rights) is classified as property income.⁵ Since it is difficult to separate these components, we compare them by treating Real Estate Income in the tax return data as part of Business Income, while in the SNA we combine Mixed Income with Rent.

"Retirement Income," which includes lump-sum severance payments, one-time benefits upon retirement, and lump-sum old-age benefits under defined-benefit or defined-contribution corporate pension plans, is also a complicated item. In the SNA, retirement lump sums are classified differently depending on funding sources: unfunded, accrual-based benefits are treated as "Other Non-Pension Social Insurance Benefits," while benefits paid from funded schemes are classified as "Other Social Insurance Pension Benefits." However, this category also includes voluntary pension schemes, so retirement lump sums cannot be isolated cleanly.

Conversely, under the tax code, corporate pensions (both defined-benefit and defined-contribution) are grouped with public pensions and are classified as "Miscellaneous Income Related to Public Pensions." Public pensions, in turn, correspond to "Cash Social Security Benefits" under "Social Benefits Other than Social Transfers in Kind." Accordingly, in tax return data, both "Miscellaneous Income Related to Public Pensions" and "Retirement Income" correspond broadly to sub-items of social benefits in the SNA.

Finally, other income categories in the tax law—such as Miscellaneous Income (excluding public pension-related items) and Occasional Income—cannot easily be mapped to the SNA, since they cover a wide variety of receipts.

IV-2. Income Observed in the Tax Return Data

Based on the categorization of Figure 5, we can compare the two datasets, the SNA and the tax return data. Table 6 presents the national total income in the SNA and the tax return

⁵ The rental income in the SNA includes "copyright royalties," which are treated as miscellaneous income for tax purposes, but this has not been adjusted here.

data, in. Here, we report averages over the seven calendar years from 2014 to 2020.

According to the SNA, the total income is 431 trillion yen. Of the total, 86 trillion yen in items unique to the SNA—such as imputed rents and FISIM—as well as tax-exempt income

Table 6. National Total Income in the SNA and Tax Return Data (average of 2014–2020)

Classification			Total Amount (Trillion Yen)	
SNA		Tax Return Data	SNA	Tax Return
1.3 Operating surplus and mixed income, net	(1) Operating surplus (imputed rent), net		22	
	(2) Mixed income, net	Business Income Forest Income Real Estate Income	16	16
1.4 Compensation of employees, receivable	(1) Wages and salaries	Employment Income	231	67
	(2) Employers' social contributions		41	
1.5 Property income, receivable	(1) Interest	Interest Income	4	0
	(2) Dividends	Dividends	8	2
	(3) Other investment income		11	
	(4) Rent			
2.6 Social benefits other than social transfers in kind, receivable	(1) Social security benefits in cash	Public Pension Income Retirement Income	71	19
	(2) Other social insurance pension benefits			
	(3) Other social insurance non-pension benefits			
	(4) Social assistance benefits		8	
2.7 Other current transfers, receivable	(1) Non-life insurance claims		3	
	(2) Miscellaneous current transfers		15	
Unclassifiable		Miscellaneous Income Other than Public Pension Temporary Income		2
Capital Gains		Transfer Income		7
Total			431	112
Fiscal Income (excl. Transfer Income, incl. Unclassifiable)			345	106
Total of Imputed Rent, FISIM, Tax Exempted, Non-individual income			86	

Source: Authors' calculation based on the tax return data and the SNA.

are, by definition, excluded from the tax return data. On the other hand, the total reported income, or fiscal income (including capital gains) in the tax return data is 112 trillion yen. This total includes 7 trillion yen in capital gains, which are not considered income in the SNA.

Thus, the portion of income comparable across the two datasets—fiscal income excluding capital gains—amounts to 345 trillion yen in the SNA and 106 trillion yen in tax return data. Although Table 6 reports averages for 2014–2020, the coverage ratio is stable over time. Fiscal income rises modestly in both sources, from 329 trillion yen to 364 trillion yen in the SNA and from 100 trillion yen to 107 trillion yen in the tax return data, but the coverage ratio remains close to 30 percent throughout.

While these discrepancies may partly reflect statistical error, they primarily result from institutional features of Japan’s tax system. For instance, although the concepts of “Wages and Salaries” in the SNA and “Employment Income” in tax return data are very similar, the majority of wage earners are taxed solely through Year-End Adjustment and therefore do not appear in the tax return data. In what follows, we refer to the share of household income in the SNA that is captured in tax return data as the coverage ratio. Under this definition, the coverage ratio of fiscal income is 31 percent.

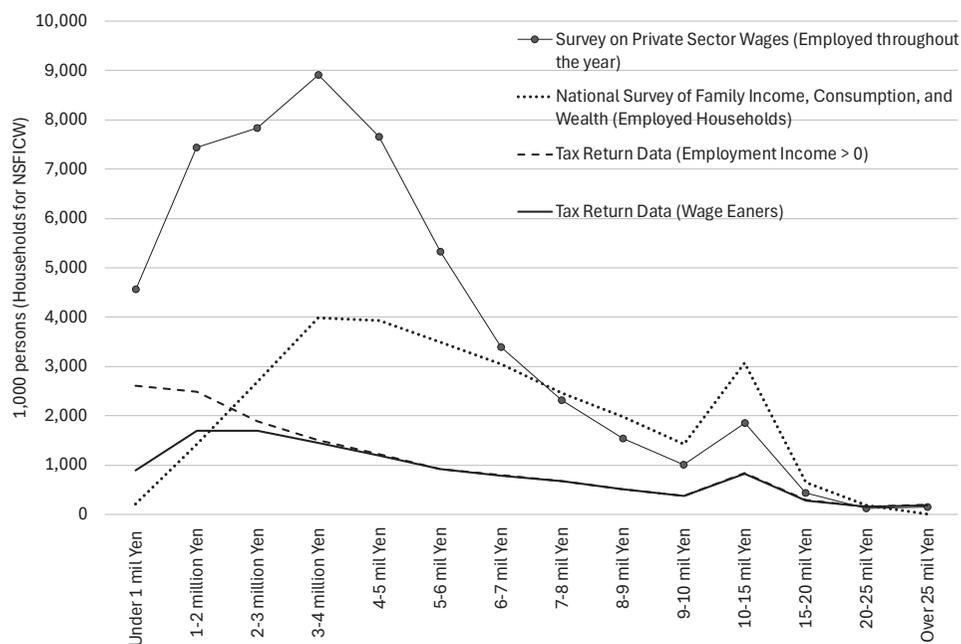
This level of coverage is considerably lower than in previous studies for the United States or France. For example, Auten and Splinter (2024) report that the U.S. tax return data capture 60–66 percent of national income, with the remaining gap attributable to items such as imputed rents, corporate retained earnings, and tax-exempt interest. In our case, the focus is restricted to household income, excluding retained earnings, and imputed rents are already excluded as an SNA-specific item. On a comparable basis, then, the U.S. coverage rate is even higher, and the gap relative to Japan is larger.

The single largest reason for Japan’s low coverage ratio is labor income—called “Wages and Salaries” in the SNA and “Employment Income” in the tax return data. In the SNA, this amounts to 231 trillion yen, whereas the tax return data capture only 67 trillion yen (29 percent). Social Transfers such as public pensions also show a large gap: 71 trillion yen in the SNA compared with 19 trillion yen (27 percent) in the tax return data.

As noted in the previous section, about 20 percent of employees and 15 percent of pension recipients file tax returns. Relative to these figures, the coverage ratios for wages and pensions are somewhat higher, suggesting that average income of employees who do file are higher than that of those who don’t. To explore this further, Figure 6 presents the distribution of employment income by income class in the tax return data. Two definitions of wage earners are used: those who are classified as “Wage Earners” above, and all individuals who report non-zero Employment Income on their tax returns (that is, including those for whom wages are not the primary source of income). For comparison, the figure also shows the distribution of employees by income class as estimated in the *Survey on Private Sector Wages*,

⁶ Since the *Survey on Private Sector Wages* is based on information collected from withholding agents, the survey unit is not individuals but withholding slips. Accordingly, it cannot be compared on a per-person basis. To address this issue, the sample here is limited to withholding slips for those who were employed throughout the year, so as to be closer to the individual basis.

Figure 6. Distribution of Employment Income in the Tax Return Data (Year 2019)



Source: Authors' Calculation based on the tax return data, *National Survey of Family Income, Consumption, and Wealth* (NSFICW), and *Survey on Private Sector Wages*.

and the distribution of worker households by annual income class from the 2019 *National Survey of Family Income and Expenditure*.⁶

According to this figure, the distribution of employment income is fairly flat, with the highest frequency observed at the low range of 2-3 million yen. A comparison with the distribution from the *Survey on Private Sector Wages* shows that the modal class differs: among all wage earners, the peak occurs below 4 million yen, encompassing roughly 9 million individuals, whereas only about 1.5 million appear in the tax return data. This implies that the filing rate among the middle-income group is on the order of 15 percent. Individuals who are not classified as wage earners but nevertheless report some Employment Income are concentrated below 2 million yen, consistent with side-job earnings. For high-income wage earners with employment income of 10 million yen or more, the gap between the tax return data and the survey narrows, indicating a higher filing rate; in particular, for the class above 20 million yen—where filing is mandatory—the two sources nearly coincide.

These observations suggest that focusing solely on wage earners who file tax returns reveals very little about the overall income distribution of wage earners or employees. As noted, many wage earners who do file claim the Medical Expense Deduction or the Special Tax Credits for Housing Loan Interest, implying that they are more likely to be in atypical economic circumstances. In this sense, tax return data are not an ideal source for measuring employment income.

By contrast, for the self-employed—whose income generally requires filing—the data

Table 7. Business Income in the SNA and the Tax Return Data (Trillion Yen)

	SNA			Tax Return Income			
	Total	Mixed Income	Rents	Total	Business Income	Real Estate Income	Forest Income
2014	16.7	13.9	2.8	15.0	8.5	6.5	0
2015	17.8	15.0	2.8	15.1	8.6	6.5	0
2016	17.5	14.4	3.0	15.8	9.2	6.6	0
2017	17.7	14.5	3.2	15.8	9.1	6.6	0
2018	14.3	11.1	3.2	15.6	9.0	6.6	0
2019	14.1	10.9	3.2	15.8	9.3	6.6	0
2020	12.9	9.6	3.3	16.1	9.6	6.5	0

Source: Authors' calculation based on the tax return data. For the SNA, Income and Outlay Accounts for Households.

appear reasonably reliable. In Table 6, the SNA and tax return figures are very close, yielding a coverage ratio near 100 percent. Although there are minor conceptual differences in the treatment of Real Estate Income and Forest Income, their impact is limited. That said, the time-series evidence presents interpretive challenges. Table 7 reports trends in business income as measured in the SNA and in the tax return data, disaggregated by component to account for definitional differences. While the seven-year averages are nearly identical at around 16 trillion yen, the annual movements diverge. From 2014 to 2017, the SNA exceeds the tax return data and coverage ratio is roughly 90 percent. Thereafter, however, the SNA level falls by about 3 trillion yen and drops below the tax return series. Even when narrowing the comparison to Mixed Income versus Business Income, the divergence is evident: mixed income in the SNA declines sharply after 2015, whereas Business Income in the tax return data trends upward. Pinpointing the source of this discrepancy requires more rigorous analysis, including scrutiny of possible measurement error in the SNA.⁷

For the self-employed, the *Sample Survey of Income Tax Returns* has long served as an important information source. However, because that survey covers only those with a positive tax liability, it has not provided a complete picture of overall income. By contrast, the tabulations presented here, which include refund filers and others excluded from the survey, confirm that tax return data can broadly capture the income situation of the self-employed in Japan. Compared with employees, income data on the self-employed are scarce. *Survey on Private Unincorporated Enterprises* is virtually the only alternative. Especially as income data at the individual level (rather than the firm level), the tax return data constitute a uniquely valuable source.

⁷ Tateoka (2016) pointed out the accuracy of Mixed Income as a measure of income of the self-employed is not so high.

Relative to wage, business, and pension income, interest and dividend income is much less comprehensively captured. Given Japan's prolonged low interest rate environment, household receipts of property income amount to only 12 trillion yen even in the SNA (4 trillion yen in interest and 8 trillion yen in dividends). Yet in the tax return data the figures are even smaller, with less than 1 trillion yen in interest and only 2 trillion in dividends, for an overall coverage rate of about 15 percent.

Much of this property income is subject to Separate Withholding Taxation or to exemption from filing. For interest, only certain categories—such as interest on specific public bonds or foreign assets—are subject to Comprehensive Taxation and therefore appear in the tax return data; the remainder falls under Separate Withholding Taxation and is thus absent from the data. Dividend income is in principle comprehensively taxable, and dividends from listed corporations can also be declared under Separate Self-Assessment Taxation, which raises coverage somewhat. At the same time, Japan has a “no-return system” under which dividend income meeting certain conditions can be fully settled through withholding alone. Although this precludes the use of dividend credits and entails other disadvantages, the system is widely used and lowers coverage of dividend income in the data.

While property income is frequently taxed separately or at source in other advanced economies as well, cases where no return is required at all are less common.⁸ Consequently, the share of property income captured in tax return data varies across countries according to institutional arrangements. In the U.S. and France, where property income is generally subject to filing, coverage is relatively high (Piketty, 2003; Piketty and Saez, 2003; Piketty et al., 2018). By contrast, in Germany—where separate withholding taxation applies—coverage is lower (Bach et al., 2022). In this sense, too, Japanese tax return data are not necessarily suitable for comprehensively capturing property income at the national level.

V. Tax Return Data as Income Data for the Super-Rich

V-1. *Super-Rich and Tax Return*

We have examined the extent to which the tax return data can serve as “comprehensive income data” for Japan. To capture the entire economy, ideally one would have data that are both exhaustive in terms of income sources and representative of the entire population. However, the tax return data only capture income sources that require filing and exclude individuals who are covered by other tax systems such as Year-End Adjustment. In this sense, the tax return data alone are not well suited for constructing nationwide income statistics. The difficulty of using tax records alone to measure the entire population depends on institutions, but it is a challenge shared across many countries.

By contrast, pioneering research beginning with Piketty (2003) has utilized tax return data as a source for measuring the incomes of the super-rich. Piketty (2003) and Piketty and

⁸ See https://www.mof.go.jp/tax_policy/summary/income/kinyuushotoku.pdf.

Saez (2003) focused on the top 1 percent of the population, comparing them against the remaining 99 percent. In their studies, national income measured in macro statistics such as the SNA are taken as the benchmark for “the whole economy.” In statistical surveys, by design, the sample of the super-rich is extremely small, making stable analysis difficult. Tax return data, however, allow for near-complete coverage of this group, making reliable analysis feasible.

In this section, therefore, we consider the characteristics of Japanese tax return data from the perspective of their use for analyzing super-rich individuals. When the focus is restricted to this group, the coverage of the tax return data becomes much better. As noted above, only 15-20 percent of the population appears in the tax return data, largely because most of wage earners have their tax obligations settled entirely through Year-End Adjustment. Yet even wage earners are required to file if they receive more than 20 million yen in annual income from a single employer, ensuring their inclusion in the data. Similarly, public pension recipients must file if their pension income exceeds 4 million yen. In short, for super-rich individuals, exemption systems do not present a major obstacle, and tax return data can provide broad coverage.

V-2. *The Super-Rich in the Tax Return Data*

To clarify what level of income qualifies as “high income” in Japan, Table 8 presents the income percentiles among all tax filers, along with the average income within each bracket. The table uses data from 2017, the mid-point of the available period, and ranks individuals by fiscal income excluding capital gains. The results, however, are stable across years and do not change materially if other years are used.

Table 8. Average Income by Percentile (10,000 Yen)

Percentile (among Filers)	Threshold	Fiscal Income	Capital Gains	Total Income
-25percentile		79	48	127
25-50 percentile	167	241	15	257
50-75 percentile	296	406	18	424
75-90 percentile	515	692	29	722
90-95 percentile	865	1,058	51	1,110
95-99 percentile	1,224	1,760	110	1,870
99-99.5 percentile	2,671	3,123	276	3,399
99.5-99.9 percentile	3,585	4,918	722	5,639
99.9 and more	7,335	16,359	3,377	19,736

Source: Authors' calculation based on the tax return data for 2017. Percentiles are calculated based on the Fiscal Income excluding Capital Gains.

Following Piketty (2003) and subsequent research, the top income groups are generally defined as a share of the total population (or labor population). In Japan, however, because the proportion of individuals filing tax returns is relatively low, comparing the same percentiles would yield artificially high observed incomes. Here, the percentiles are calculated from the 23.25 million filers in 2017, so the top 0.1 percent corresponds to about 23,000 individuals, and the top 0.5 percent to about 116,000. In relation to Japan's total population aged 15 and above (about 100 million), focusing on the top 0.5 percent and top 5 percent of filers allows for comparisons with international research using consistent criteria.

From a policy perspective, in recent years attention has been drawn to the so-called "100-million-yen barrier," whereby the effective tax rate declines for taxpayers whose total income exceeds 100 million yen (Oka, 2024). According to Table 8, however, the threshold income at the 99.9th percentile among tax filers is around 75 million yen, indicating that those with incomes above 100 million yen are relatively rare even within the top 0.1 percent of filers. Relative to the 15-and-over population, this corresponds to less than 0.002 percent. Thus, focusing on those above 100 million yen would be too narrow for comparisons in the line of research following Piketty (2003). Instead, this paper focuses on the top 0.1 percent and top 1 percent of the population aged 15 and above.

First, we consider whether the rich group as the top 1 percent can be comprehensively observed in the tax return data. Table 8 shows that the threshold for the top 5 percent of filers is 12.24 million yen. This is well below the 20-million-threshold that makes filing mandatory for wage earners. In other words, some individuals with incomes equivalent to the top 1 percent may not file.

According to the *Survey on Private Sector Wages*, there are about 2.3 million wage earners with annual incomes between 10 million yen and 20 million yen. By contrast, only about 1.1 million such wage earners appear in the tax return data with fiscal income in this range. Thus, roughly half of the individuals who may belong to the top 1 percent of the total population do not file tax returns. In this sense, tax return data do not provide sufficiently comprehensive coverage of the top 1 percent of income earners.

On the other hand, at the level of the top 0.1 percent, nearly all relevant individuals are assumed to file. According to Table 8, the fiscal income threshold for the top 0.1 percent of filers is 73.35 million yen, and for the top 0.5 percent (that is, the top 0.1 percent of the total population) it is 35.85 million yen. These levels are unattainable for wage earners or pensioners who do not file and thus are unaffected by institutional exemptions from filing.

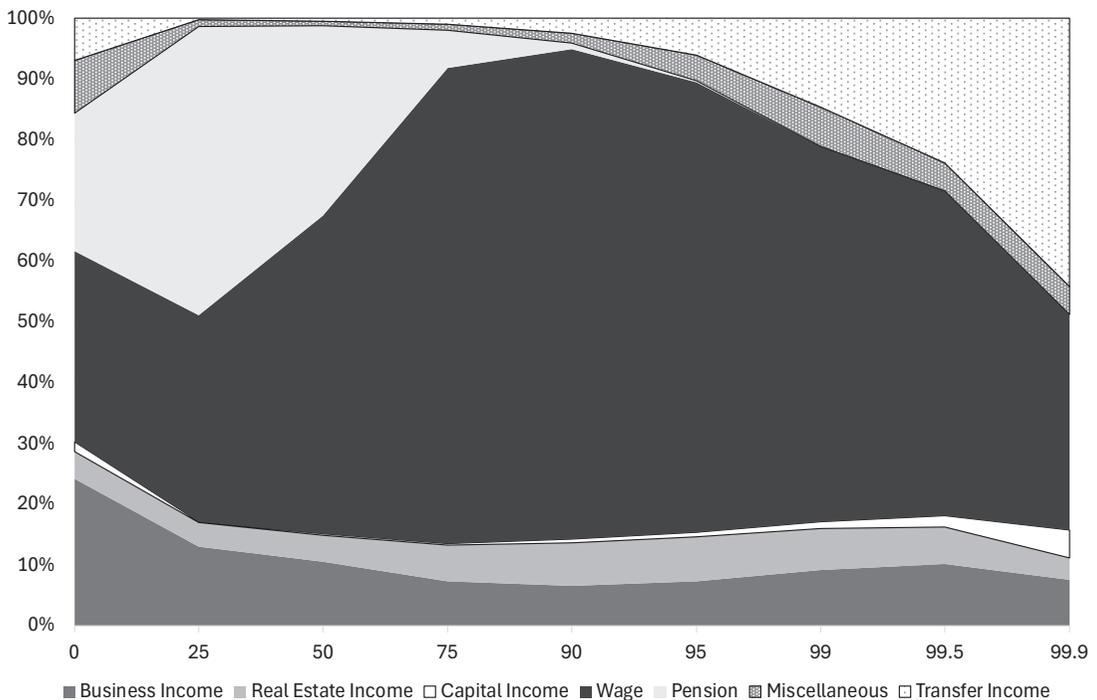
Nevertheless, these thresholds are strikingly low by international standards. Piketty et al. (2018) report that in the United States in 2014, the top 0.1 percent threshold was 1.96 million dollars (about 200 million yen at the time), and the top 1 percent threshold was 458,000 dollar (about 50 million yen). Similarly, Garbinti et al. (2018) report that in France in 2014, the top 0.1 percent threshold was 564,000 euro (about 80 million yen). Of course, per capita GDP (converted at market exchange rates) of Japan is lower than that of the United States, so lower income levels per se are not surprising. Yet in 2015, per capita GDP in dollar terms was \$57,000 for the United States, \$38,000 for France, and \$35,000 for Japan. These differ-

ences cannot fully account for the much lower thresholds observed in Japan.

These findings have often been interpreted to mean that income inequality in Japan is relatively modest. For example, Moriguchi and Saez (2008) and Atkinson et al. (2011) point out that in the postwar period, the share of national income accruing to the top groups in Japan was lower than in Western countries, a pattern attributed to highly progressive taxation and the institutional structure of wage determination. While some studies suggest that inequality has widened since the 1990s, this trend is generally attributed to an increase in the low-income population, and clear evidence of changes in the income share of the top groups has not been established.

At least for the period since 2014, for which the tax return data are available, the relatively low income levels observed among high-income individuals in Japan may be attributable to the Separate Withholding Taxation of interest and dividend income. As noted above, property income such as interest and dividends is often subject to Separate Withholding Taxation or exemption from filing. Thus, even when recipients of such income do file returns, these amounts may not be included in reported taxable income. In fact, property income reported in tax return data amounts to only about 10 percent of that recorded in the SNA, confirming that it is not comprehensively captured. Since interest and dividend income tends to rise with income, this exclusion is likely to lead to systematic underestimation

Figure 7. Share of Income Source Groups by Income Percentile



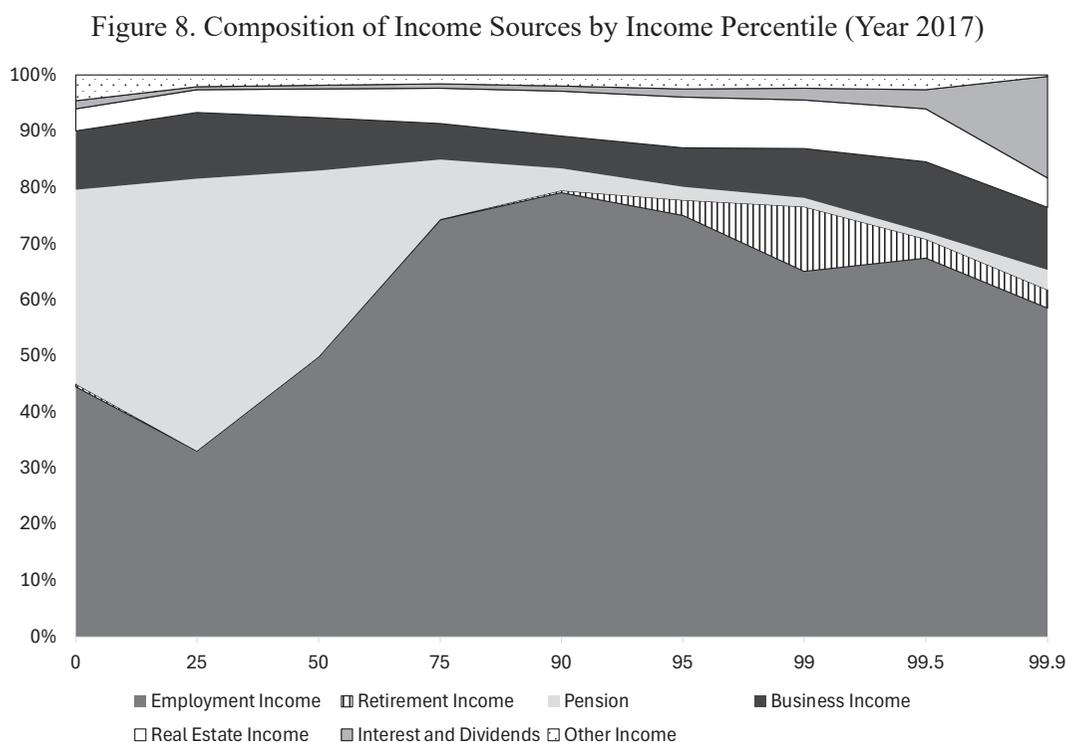
Source: Authors' calculation based on the tax return data. As for the definition of income source groups, see Table 3.

of high-income individuals' earnings. This interpretation is indirectly supported by the following evidence.

The first piece of evidence is that, even among the top 0.1 percent or 0.5 percent income groups, most filers are classified as wage earners. Figure 7 shows the distribution of filers by their primary source of income, as defined in Table 4, across income brackets ranked by fiscal income. Among all filers, about half are wage earners (Figure 3), and by age group, about 70 percent of those in the working-age population are wage earners (Figure 4). By contrast, Figure 7 indicates that 70 percent of filers in the top 25 percent income group are wage earners. Above the 90th percentile, the share of business income earners rises, and above the 95th percentile, the share of property income earners increases slightly. Yet even in the top 0.1 percent group, as many as 60 percent are wage earners.

According to Piketty, Saez, and Zucman (2018), in the United States the majority of income among households in the top 1 percent derives from capital income—such as dividends, interest, and capital gains on stock sales—a trend that has intensified since the 1970s. Even if capital gains are excluded here, thereby reducing the measured share of capital income, the fact that as many as 60 percent of the super-rich in Japan appear with wages as their primary income source seems implausible.

This point is further confirmed by a second piece of evidence, namely the composition of income sources. Figure 8 illustrates the composition of income sources by income per-



Source: Authors' Calculation based on the tax return data. Total income does not include capital gains.

centile. Unlike Figure 7, which classified taxpayers according to their predominant type of income, Figure 8 examines the aggregate composition of income within each percentile class. For example, in the lower income groups, pension recipients account for about 20 percent of tax return filer, yet public pension income makes up roughly 40 percent of total income. This reflects the fact that many low-income taxpayers other than those classified as Pensioners are also elderly individuals receiving public pensions.

What is particularly noteworthy here is that even among high-income groups, such as those above the 90th percentile, the bulk of income consists of labor income. Including retirement income—which accounts for about 10 percent around the 99th percentile—roughly 70 percent of income takes the form of compensation for labor. Even within the top 0.1 percent, wage income alone makes up more than 60 percent, indicating not only that few ultra-high-income individuals rely primarily on non-wage sources, but also that their non-wage income is relatively limited.

Although it is well known that richer people receive larger property income, such as interest, dividends and rents, the share of property income does not increase dramatically. Rents (Real Estate Income) maintains a roughly 10 percent regardless of total income. Interest and dividend are also modest even for the super-rich; about 3 percent even among the top 0.5 percent, and around 18 percent for the top 0.1 percent.

Garbinti et al. (2018) report that in France in 2012, the share of property income (capital income, including rent) exceeded 50 percent for the 99-99.5th percentile group, and rose above 80 percent for the top 0.1 percent. Auten and Splinter (2024) also find that in the United States since the 2000s, the share of property income (excluding capital gains) in fiscal income amounts to roughly 35 percent for the top 1 percent and about 45 percent for the top 0.1 percent. These comparisons suggest that the share of property income among the super-rich in Japan is strikingly low.

V-3. Impacts of Non-Reported Interest and Dividend

Evidence so far shows that, even within the top 0.1%, Wage Earners constitute the majority, and wages accounts for more than half of their income in Japan. This may suggest that there are only few people who earn enough capital income to be a super-rich. However, for reference, the 2024 edition of Forbes magazine's World's Billionaires List included 41 Japanese and 43 French individuals. It is therefore difficult to conclude that Japan alone lacks wealthy individuals who earn substantial amounts of interest and dividend income. A plausible explanation for this gap lies in interest and dividend income that goes unreported due to Separate Withholding Taxation and other mechanisms.

Suppose that the top 0.5% of income tax filers (corresponding to the top 0.1% of the population) earn approximately 35% of their taxable income from capital income, which is comparable to that observed among the top 1% in France and the United States. Then, since the reported interest and dividend income is only about 8 million yen, unreported capital income should be roughly 30 million yen (the true total income is the sum of average fiscal

income, 75 million yen, and this unreported capital income, 30 million yen. 35% of the true total income is equal to 40 million yen).

This level is plausible when viewed against the SNA. If the top 0.1% or 100,000 individuals each received 40 million yen in interest and dividends, the national total would amount to 4 trillion yen, or about 30 percent of the total interest and dividend observed in the SNA, which is not implausible in other countries. For example, Garbinti et al. (2018) show that in France since 2000, the top 1 percent account for 35 percent of property income. Moreover, if the estimated unreported interest and dividend income is added to the threshold for the top 0.1% (the top 0.5% in Table 8), the threshold rises to about 70 million yen substantially narrowing the observed gap with French case.

There is also some micro-level plausibility. According to estimates published annually by the Nomura Research Institute, there were about 90,000 “super-wealthy” households in Japan in 2021, defined as those with net financial assets of 500 million yen or more. With 500 million yen in assets earning a 6 percent return, interest and dividend income would amount to about 30 million yen, although achieving a 6 percent yield under current low interest rate situation is not easy.

These findings suggest that the incomes of the super-rich in Japan may not be fully captured. While earlier analysis already showed that the coverage rate of tax return data for interest and dividend income relative to the SNA is low due to separate withholding taxation, the impact appears particularly significant for top earners. For the calculation of income shares among very rich groups in Japan, developing methods to estimate unreported interest and dividend income emerges as an essential task.

VI. Conclusion

This paper has examined the characteristics of Japanese tax return data as a source of income information. The analysis shows that while tax return data provide valuable individual-level records of income, they also have important limitations. It has long been recognized that tax return data alone cannot comprehensively capture the national distribution of income. However, less attention has been paid to analyzing the attributes of the individuals who file tax returns.

Under Japan’s tax system, wage and pension income are often subject to withholding and year-end adjustment, eliminating the need for filing. As a result, most wage earners and pension recipients are not included in the dataset. Moreover, many of those who do file tax returns appear to do so for the purpose of claiming deductions such as the medical expense deduction or the Special Tax Credits for Housing Loan Interest. This suggests that among wage earners and pension recipients, those in exceptional social or economic circumstances are more likely to file. In this sense, tax return data cannot provide a representative picture of these groups.

By contrast, for individuals whose primary source of income is business income, the dataset appears to capture most of the relevant population. The number of business income

filers is close to the number of self-employed individuals in the Labor Force Survey, and their reported income is similar in magnitude to mixed income in the SNA. Moreover, more than half of these individuals file almost every year, making the data suitable for panel analysis. Since comprehensive data on the self-employed are limited, tax return data hold considerable promise for generating new insights into the income dynamics of this group.

This study also considered whether tax return data can be used to study high-income individuals. For the top 0.1 percent of the population, the income threshold is close to 80 million yen, a level unaffected by institutional features such as year-end adjustment. Thus, this group can be assumed to file returns without exception, making tax return data comprehensive for the ultra-high-income segment.

For the top 1 percent (about one million individuals), however, the threshold is roughly 13 million yen—well below the 20-million-yen threshold. This implies that even within the top 1%, there are likely to be high-income individuals who do not file, suggesting that tax return data are insufficient for analyzing this broader group. Even when focusing only on the super-rich, coverage is not complete. In particular, most interest and dividend income is subject to separate withholding taxation or exemption from filing, leading to likely underestimation of income among top earners. The relatively modest income share of high-income individuals in Japan, compared to other advanced economies, may partly reflect this difficulty in capturing property income accurately.

The findings of this study offer several important implications for the future use of tax return data and for policymaking. First, to establish a comprehensive income dataset at the national level, tax return data must be integrated with sources that cover non-filers, such as corporate tax records on wage payments and information from social insurance systems like public pensions. Second, integration of data on income subject to separate withholding taxation is also essential. While the impact is currently limited under low interest rates for most households, accurate measurement of property income such as interest and dividends is indispensable for analyzing high-income groups.

This study has compared tax return data with macro statistics to assess which types of income and which individuals are captured, and where major gaps remain. However, little analysis has yet been conducted on the distribution of individual incomes. For future research, it is important to explore how tax return data relate to survey-based statistics and how the two can be effectively combined. This would allow consistent observation of ordinary people outside the super-rich.

References

- Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez (2011), “Top Incomes in the Long Run of History.” *Journal of Economic Literature* 49(1): 3-71.
- Auten, Gerald, and David Splinter (2024), “Income Inequality in the United States: Using Tax Data to Measure Long-Term Trends.” *Journal of Political Economy* 132(7): 2179-2227.

- Bach, Stefan, Charlotte Bartels, and Theresa Neef (2022), “Distributional National Accounts: A Macro-Micro Approach to Inequality in Germany.” In *Measuring Distribution and Mobility of Income and Wealth*, edited by Raj Chetty, John C. Gornick, Barry Johnson, and Arthur Kennickell, 625-640. Chicago: University of Chicago Press.
- Brooks, John R (2018), “The Definition of Income.” *Tax Law Review* 71: 253-309.
- Garbinti, Bertrand, Jonathan Goupille-Lebret, and Thomas Piketty. 2018. “Income Inequality in France, 1900-2014: Evidence from Distributional National Accounts (DINA).” *Journal of Public Economics* 162: 63-77.
- Moriguchi, Chiaki, and Emmanuel Saez (2008), “Income Concentration in Japan, 1886-2005: Evidence from Income Tax Statistics.” *Review of Economics and Statistics* 90(4): 713-734.
- Oka, Naoki (2024), “New Developments in the Taxation of Financial Income and High-Income Individuals,” *Financial Review* No.157: pp. 133-168 (In Japanese).
- Piketty, Thomas (2003), “Income Inequality in France, 1901-1998.” *Journal of Political Economy* 111(5): 1004-1042.
- Piketty, Thomas, and Emmanuel Saez (2003), “Income Inequality in the United States, 1913-1998.” *Quarterly Journal of Economics* 118(1): 1-41.
- Piketty, Thomas, and Emmanuel Saez (2013), “Optimal Labor Income Taxation.” In Alan J. Auerbach, Raj Chetty, Martin Feldstein, and Emmanuel Saez eds. *Handbook of Public Economics*, Vol. 5, 391-474.
- Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman (2018), “Distributional National Accounts: Methods and Estimates for the United States.” *Quarterly Journal of Economics* 133(2): 553-609.
- Simons, Henry Calvert (1938), *Personal Income Taxation: The Definition of Income as a Problem of Fiscal Policy*. Chicago: University of Chicago Press.
- Slemrod, Joel (2016), “Caveats to the Research Use of Tax-Return Administrative Data.” *National Tax Journal* 69(4): 1003-1020.
- Tateoka, Kenjiro (2016), “Estimating the Capture Rate of Business Income: The Gap in Reporting Accuracy between Wage and Business Income Persists,” *JRI Review*, No. 5(35) (In Japanese).
- Unayama, Takashi (2024), “Economics and DX in the Taxation Practice,” Chapter 7, in Junichiro Maeda ed. *Digital Transformation in Tax Administration and Japan’s Future*. Kinyu-Zaisei Jijou Kenkyukai: Tokyo (In Japanese).
- World Inequality Lab (2024), *Distributional National Accounts Guidelines: Methods and Concepts Used in the World Inequality Database*.