Comparison of Taxes and Social Insurance Premium Burdens in House-
hold Accounts*

Taro Ohno
Associate Professor, Faculty of Economics, Management and Information Science, Onomichi City University; Senior Visiting Scholar, Policy Research Institute, Ministry of Finance

Masahiko Nakazawa
Former Professor, Research Center for Advanced Policy Studies, Institute of Economic Research, Kyoto University

Kazuaki Kikuta
Former Researcher, Policy Research Institute, Ministry of Finance

Manabu Yamamoto
Former Researcher, Policy Research Institute, Ministry of Finance

Abstract
This paper aims (1) to examine the characteristics of various statistics by comparing Indicated Values in response to questionnaires concerning various taxes and social insurance premiums and (2) to verify the validity of fitted values of the household tax burden (in terms of income and residential taxes) by comparing the Indicated Values with the Fitted Values.

First, the comparison of Indicated Values in different statistics showed that Indicated Values of tax and insurance premiums were underestimated in the National Survey of Family Income and Expenditure and the Family Income and Expenditure Survey. In addition, as a result of the comparison between the Indicated Values and Fitted Values and the examination of the distribution of disparities between them, it was found that the mean disparity between the Indicated and Fitted Values of income and residential taxes was zero and the dispersion was around 3%. It was also confirmed that the disparity occurred frequently because of entry errors for the tax amount and that entry errors for values in questionnaires that affected the disparity included “business income entry error” and “tax amount entry error due to errors in digits.”

The implication of the examination is that Fitted Values obtained through microsimulation analysis are estimates with little bias regarding the macro values obtained through collection of data and that there is little margin of error. In that sense, these Fitted Values have sufficient precision to be used for the evaluation of policies.

Keywords: taxes, insurance premium, indicated values, fitted values, microsimulation analysis

JEL Classification: C15, H24
I. Introduction

In recent years, household micro data has been streamlined in Japan. Based on this environmental development, the number of analyses using household micro data (questionnaire information) in the fields of taxation and social securities, such as the measurement of burdens on household accounts by taxes and insurance premiums, assessment of redistribution policy under the taxation and social security system, has increased. With regard to taxes and insurance premiums for each household account in these activities, there are two cases: (1) using burdens indicated in the questionnaire without making any changes (Indicated Value), and (2) using burdens calculated by applying information, such as family unit, income, etc. indicated in the questionnaire to the actual system (Fitted Value).

Examples of prior research using Indicated Values include Abe (2000), Ohishi (2006), Fukawa (2006), Oshio and Urakawa (2008), Oshio (2009), Tanaka (2010), Ohno et al. (2013, 2014). Examples of prior research using the Fitted Value include Yada (2010), Tanaka and Shikata (2012), Tanaka et al. (2013), Kitamura and Miyazaki (2013), Miyazaki and Kitamura (2014). Recently, microsimulation analysis is often implemented to examine the effects of policies before their establishment and policies scheduled to be established on household accounts. This method also corresponds to the case of using the Fitted Value. For example, prior research which performed microsimulation analysis of tax reform related to the income tax system includes Tajika and Furutani (2003, 2005), Tajika and Yashio (2006a, 2006b, 2008, 2010), Shiraishi (2010), and Doi and Park (2011).

Despite much prior research, there are many issues remaining to be examined and verified when using Indicated Values and Fitted Values respectively, such as “what characteristics do Indicated Values have for each of the statistics to be used?” or “Is the Fitted Value appropriate?”

* This research is supported by the Grants-in-Aid for Scientific Research (Series of Single-year Grants; Grant-in-Aid for Young Scientists (B); Research No.26780176), and was independently tallied using the information from questionnaires of the Comprehensive Survey of Living Conditions conducted by the Ministry of Health, Labour and Welfare. We hereby express our deep appreciation for the cooperation of the persons involved. We would like to express our thanks to Mr. Takashi Unayama (Hitotsubashi University), Mr. Saisuke Sakai (Ministry of Finance), Mr. Shunji Tada (Ministry of Finance), Mr. Keishi Hachisuka (Nippon Express Co., Ltd.), Mr. Byota Hara (Ministry of Finance), Ms. Tomoko Masuda (NTT Data Corporation), Mr. Takeshi Miyazaki (Kyushu University), Mr. Koyo Miyoshi (Aichi Gakuin University), Mr. Yasutaka Yoneda (Ministry of Finance), and participants of paper review committee of the Financial Review for their valuable feedback. Please note that this report is based on the personal opinions of the authors and does not indicate the official opinions of the institutions to which the authors belong.

1 In this report, burden amounts from household accounts are divided as follows into three categories: (1) “Actual Value”: the amount actually paid; (2) “Indicated Value”: the amount indicated by respondents in the questionnaire; and (3) “Fitted Value: the amount calculated by applying information, including family unit, income, etc., indicated in the questionnaire in the actual system.

2 In addition, examples of prior research which measured consumption tax burdens in household accounts include Ohtake and Kohara (2005), Takayama and Shiraishi (2010), and Shiraishi (2011). Examples of prior research which implemented microsimulation analysis of tax reform related to the consumption tax system include Yashio and Hasegawa (2009), Takayama and Shiraishi (2011), and Tanaka (2014). With regard to consumption tax burdens in household accounts, the burden amount (Indicated Value) was not often indicated in the questionnaire. Therefore, the burden amount (Fitted Value) that is calculated by multiplying a specified ratio in consideration of tax-free items by information on consumption indicated in the questionnaire (from the National Survey of Family Income and Expenditure, the Family Income and Expenditure Survey, etc.) is used.
As for Indicated Values, their measurement results receive no small impact from the difference in survey methods even in cases where the same content is surveyed by multiple statistics. They can be a cause of different conclusions drawn from statistics used in cases of measuring the burden of taxes and insurance premiums in household accounts. Therefore, it is important to understand the characteristics of each statistic used.

Regarding Fitted Values, there are two reasons to use Fitted Values although Indicated Values exist. First, they supplement data. As indicated by Tanaka et al. (2003), the National Survey of Family Income and Expenditure (surveys the burden in workers’ households and non-workers’ households, but) does not survey the burden of self-employed households. Therefore, there are no Indicated Values for self-employed households. For this reason, it is necessary to use Fitted Values in order to include self-employed households in the analysis samples for examination. Second, it is necessary for the analysis of the effects of policies based on microsimulation. In this case, the main issue is the comparison between the times before and after the establishment of policies with respect to the burden paid by households; however, it is usually impossible to understand conditions after establishment of the policies based on statistics from “before their establishment.” Therefore, it is necessary to use the Fitted Values in order to see the effects of policies before establishment.

In this way, it is meaningful to use the Fitted Values from the perspectives of the characteristics of the statistics to be used and for analytical purposes. The use of the Indicated Values and Fitted Values should not have an influence on the content of analysis. In this regard, it is based on Indicated Values and Fitted Values being consistent. For example, in cases of examining whether measurement results after the establishment of policies by microsimulation analysis are appropriate, examining how realistically the conditions before establishment of policies were reproduced can be one of the criteria at least.

This awareness of issues related to the appropriateness of Fitted Values has occurred before. For example, Tajika and Yashio (2006a) compared the Indicated Value and Fitted Value of income tax amounts by focusing on the mean value by each income class. In other words, the issue of appropriateness of the Fitted Value has received a certain amount of interest from analysts who have implemented microsimulation analysis.

The first purpose of this paper is to examine the statistical characteristics of the Indicated Values. In this paper, we examine the characteristics of each statistic and the differences between statistics by using the data of the National Survey of Family Income and Expenditure and the Family Income and Expenditure Survey of the Ministry of Internal Affairs and Communications and data from the Comprehensive Survey of Living Conditions of the Ministry of Health, Labour and Welfare and by comparing Indicated Values of various taxes and social insurance premiums between statistics. The second purpose is to verify the appropriateness of Fitted Values. In this paper, we use individual questionnaire data of the Comprehensive Survey of Living Conditions, compare the Indicated Values and Fitted Values of tax burdens (income tax and residential tax) in household accounts, and thereby examine their approximations and disparities and the causes of disparities.

Disparities are caused by both Indicated Values and Fitted Values. The reason why Indi-
cated Values become too small or too large is entry errors in the tax amount. On the other hand, the reason why Fitted Values become too small or too large is the impact of the model (for calculation of Fitted Value) or entry errors in income. We examine which cause has the highest frequency among these causes. We also examine specific cases of entry errors in questionnaires.

The outline of the paper will be as follows. First, in Section II, we compare Indicated Values between statistics and examine the characteristics of each statistic and the differences between statistics. In Section III, we compare Indicated Values and Fitted Values and then examine the appropriateness of Fitted Value. Then, in Section IV, we review the causes of disparities and examine the characteristics of the disparity and their causes. In closing, the conclusion is stated in Section V.

II. Statistical Characteristics of Indicated Values

II-1. Data

The analysis in this Section uses data from the National Survey of Family Income and Expenditure and the Family Income and Expenditure Survey of the Ministry of Internal Affairs and Communications (both sets of data have been published) and data from the Comprehensive Survey of Living Conditions of the Ministry of Health, Labour and Welfare (individual questionnaire data).

First, the survey method of each statistic was checked in relation to the Indicated Values of taxes and social insurance premiums. The National Survey of Family Income and Expenditure is conducted every five years and the survey targets approximately 57,000 family units throughout Japan (including 4,400 single-person households). There are four types of questionnaire: Household Questionnaire, Yearly Income and Savings Questionnaire, Durable Goods Questionnaire, and Family Account Book; and each of these questionnaires includes detailed survey items. In cases of the Family Account Book, workers’ households and non-workers’ households describe deductions based on income (in other words, the amounts of tax and social insurance premiums) in addition to income and expenditures during the survey period. It is necessary to note that only earned income tax (income tax pertaining to employment income) is subject to the income tax. Meanwhile, non-workers’ households, such as private practice household, etc., describe only expenditures during the survey period. The survey period of the Family Account Book is three months of the survey year (September through November) in cases of family units with two or more persons and two months of the survey year (October and November) in cases of single-person households.

The Family Income and Expenditure Survey is implemented every month and the survey targets approximately 9,000 family units throughout Japan. There are four types of questionnaire: Household Questionnaire, Yearly Income Questionnaire, Savings Questionnaire, and Family Account Book and each of these questionnaires includes detailed survey items. Among these questionnaires, the Family Account Book requires a description of income and
expenditures in the survey period for workers’ households and non-workers’ households. It is necessary to note that only earned income tax (income tax pertaining to employment income) is subject to the income tax. Meanwhile, non-workers’ households, such as private practice households, etc., describe only expenditures during the survey period. The survey period of the Family Account Book is six months in cases of family units with two or more persons and three months in cases of single-person households, and survey targets are replaced sequentially with newly selected family units.

The Large-Scale Survey of the Comprehensive Survey of Living Conditions is conducted every three years and the Small-Scale Survey is conducted in each interim year of the Large-Scale Survey. The Survey targets approximately 290,000 family units and household members throughout Japan for the Household Questionnaire and Health Questionnaire, approximately 7,000 persons throughout Japan for the Long-Term Care Questionnaire, and approximately 40,000 households and their household members throughout Japan for the Income Questionnaire and Savings Questionnaire. There are five types of questionnaires: Household Questionnaire, Health Questionnaire, Long-Term Care Questionnaire, Income Questionnaire, and Savings Questionnaire, and each of these questionnaires includes detailed survey items. In the Income Questionnaire out of these, annual income is described for one year from January through December of the previous year of the survey, income tax and social insurance premiums (health insurance premiums, pension premiums, long-term care insurance premiums, and other social insurance premiums) are described for one year of the previous year of the survey, and residential tax is described for one year of the survey year. As indicated above, it is necessary to note that the target year of the residential tax is different from other taxes and social insurance premiums.

Summarizing differences between statistics related to published data, the first difference is between target periods. For example, the Comprehensive Survey of Living Conditions can find the annual burden and the Family Income and Expenditure Survey can find the monthly burden as a yearly mean. In contrast, the National Survey of Family Income and Expenditure only gives the burden for two to three months; and there is a possibility that the burden paid by income other than regular income, which are obtained outside the survey period, such as bonuses, are not given, and therefore seasonal issues may occur easily. The second difference is between target family units. The National Survey of Family Income and Expenditure and the Family Income and Expenditure Survey can only give the burden of workers’ households. A workers’ household means a household where the household works at a company, government administration office, school, plant, store, etc.; provided, however, that it excludes households where the householder is an officer for a company or group, such as the president, director, board member, etc. In contrast, the Comprehensive Survey of Living Conditions is conducted for all households, including workers’ households and non-workers’ households. The third difference is who is defined as

---

3 To be precise, the workers’ household is categorized into “workers’ households among households with two or more persons,” “workers’ households among single-person households,” and “workers’ households among all households.”
a householder. It indicates the person who gains the majority of income for the household account in the National Survey of Family Income and Expenditure and Family Income and Expenditure Survey; however it indicates a person who is described as such in the Household Questionnaire in the Comprehensive Survey of Living Conditions. The fourth difference is the fact that family units for which the burden is nil are included in the National Survey of Family Income and Expenditure and Family Income and Expenditure Survey; however they are not included in the Comprehensive Survey of Living Conditions. The fifth difference is related to the scope of income tax that is limited to earned income tax in the National Survey of Family Income and Expenditure and Family Income and Expenditure Survey; however, it is not limited to earned income tax in the Comprehensive Survey of Living Conditions.

Due to the aforementioned differences between statistics, it is necessary to bridge these differences as much as possible. This paper uses the published data (total value) of the National Survey of Family Income and Expenditure and Family Income and Expenditure Survey for its comparison of statistics, while using the individual questionnaire data of the Comprehensive Survey of Living Conditions, in order to adjust for differences, such as target family units, definitions of a family unit, and summary treatment. In other words, we examine the data by limiting target family units to “workers’ households among family units with two or more persons,” householders to “the largest income earner,” and summary treatment to “include family units for which burden amount is nil.”

II-2. Comparison of Indicated Values between Statistics

In this part, we compare Indicated Values between statistics and examine the characteristics of each statistic and the differences between statistics. Figure 1 indicates the changes in income tax in each statistic. The amount unit is yen (monthly amount). The Comprehensive Survey of Living Conditions measured both cases where a householder is deemed to be the person who is described in the questionnaire and cases where it is deemed to be the largest income earner (in order to conform to the National Survey of Family Income and Expenditure and the Family Income and Expenditure Survey). It is observed with respect to income tax that the level of the Comprehensive Survey of Living Conditions is basically the largest, the second largest is the Family Income and Expenditure Survey, and it is the smallest is the National Survey of Family Income and Expenditure. Comparing them with the level of the Comprehensive Survey of Living Conditions, the income tax under the Family Income and Expenditure Survey in the same year corresponds to 60 to 70 percent of the Comprehensive Survey of Living Conditions and, the income tax under the National Survey of Family Income and Expenditure is the smallest. Figure 2 shows changes in residential taxes. In descending order of amount, the residential tax amounts in the Comprehensive Survey of Living Conditions are the largest, then Family Income and Expenditure Survey, and National Survey of Family Income and Expenditure. Compared with the level of the Comprehensive Survey of Living Conditions, the amounts in the Family Income and Expendi-
Figure 1
Changes in Burden: Earned Income Tax

Figure 2
Changes in Burden: Individual Residential Tax
diture Survey and National Survey of Family Income and Expenditure in the same year correspond to 80 to 90 percent of the amount of the Comprehensive Survey of Living Conditions and the differences between statistics are relatively small.

As mentioned above, differences in the definitions of target family units and householders are, at least, not causes of differences between income tax statistics. This is because these elements are adjusted before comparison of statistics. The major causes of the aforementioned differences are differences in the scope of income tax and impacts of survey methods under the National Survey of Family Income and Expenditure and Family Income and Expenditure Survey. Tada and Miyoshi (2015) indicate that entries of regular income and bonuses are inadequate in the Family Account Book (questionnaire) under the Family Income and Expenditure Survey. This underestimation of income may result in an underestimation of income tax amount since tax related to income that is not included in the questionnaire is not reflected completely in the statistics. The fact that differences between statistics with respect to residential tax are smaller than for income tax may support the aforementioned possibility. Residential tax is imposed on the income of the previous year and collected in monthly instalments under this system. Therefore, even if the entries of incomes in the National Survey of Family Income and Expenditure or Family Income and Expenditure Survey were inadequate, it is difficult to consider that it leads to an underestimation of the residential tax amount.

Next, let’s look at social insurance premiums and various insurance premiums as details of the social insurance premiums. Figure 3 shows changes in social insurance premiums in each statistic. With regard to social insurance premiums, the differences between statistics are relatively small by 2000 and the level of the Comprehensive Survey of Living Conditions is slightly larger than those of the Family Income and Expenditure Survey and National Survey of Family Income and Expenditure in 2000 and after. Figure 4 shows changes in pension premiums and Figure 5 shows changes in health insurance premiums. These figures are also characterized by the fact that the level of the Comprehensive Survey of Living Conditions is slightly larger than those of the Family Income and Expenditure Survey and National Survey of Family Income and Expenditure, like for social insurance premiums in 2000 and after. Figure 6 shows changes in long-term care insurance premiums. The major difference between the levels of the Comprehensive Survey of Living Conditions and of the Family Income and Expenditure Survey and National Survey of Family Income and Expenditure is confirmed in this Figure. Compared with the level of the Comprehensive Survey of Living Conditions, the levels of Family Income and Expenditure Survey and National Survey of Family Income and Expenditure in the same year are almost 50% of that of the Comprehensive Survey of Living Conditions.

As mentioned above, the size of the differences in social insurance premiums between statistics varies depending on the season. As shown in Figure 3, the level of the Family Income and Expenditure Survey is almost the same as the Comprehensive Survey of Living Conditions around 2000: however, it decreased to approximately 80% in 2003 and after. One of the causes is inadequate entry of incomes in the Family Account Book (question-
Figure 3
Changes in Burdens: Social Insurance Premiums

Figure 4
Changes in Burdens: Pension Premiums
Figure 5
Changes in Burdens: Health Insurance Premiums

Figure 6
Changes in Burdens: Long-Term Care Insurance Premiums
naire) under the National Survey of Family Income and Expenditure and Family Income and Expenditure Survey. In particular, underestimation of income, such as bonuses, may also have resulted in underestimation of the social insurance premium amount, which relates to income that is not reflected. These characteristics may have affected to the Indicated Values of social insurance premiums since the total remuneration system (total of salary and bonuses) was introduced for employee pension insurance and health insurance in 2003 and social insurance premiums also started to be collected from bonuses, etc.

III. Microsimulation and Indicated Values

III-1. Data and Calculation Method

In this Section, we use the microsimulation method and compare tax and social insurance premiums that are calculated based on the information of family unit properties and the income of family units, with the amount entered by family units. The following analysis is based on the individual questionnaire data (Household Questionnaire and Income Questionnaire) of the Comprehensive Survey of Living Conditions (surveyed in 2010) of the Ministry of Health, Labour and Welfare. These questionnaires described income information in 2009.

With regard to the selection of target family units for analysis (samples), first, we eliminated family units that did not enter an income tax amount and residential tax amount in the questionnaire, in order to compare Indicated Values and Fitted Values. Second, we eliminated family units for which the ages and social insurance premiums are unknown, since these items are necessary for calculating Fitted values. Third, we eliminated family units of employees transferred without their family members since they cannot specify dependents and affect the calculation of exemptions and deductions from income. Fourth, we eliminated family units for which members have been transferred by the company and are living separately from the family units (employees transferred without their family members), since allowances may not be included appropriately and details of household income are unclear. In the case of income tax, the sample size is a maximum of 19,926 family units and a maximum of 20,295 family units in the case of residential tax.

Next, the types of income used in this model are indicated below. Due to the inconsistency between the types of income under the Japanese Income Tax Act and categories of income obtained from income information under the Comprehensive Survey of Living Conditions, retirement income, timber income, occasional income, and capital gains under the Income Tax Act cannot be included in the considerations. Interest income, dividend income, and real property income under the Income Tax Act are given as the total value of these three types of income under the Comprehensive Survey of Living Conditions as property in-

---

4 Currently, the Japanese Income Tax Act categorizes income into the following ten categories based on their characteristics: (1) interest income; (2) dividend income; (3) business income; (4) real property income; (5) employment income; (6) retirement income; (7) capital gains; (8) timber income; (9) occasional income; and (10) miscellaneous income.
come. Therefore, taxpayers can select a separate withholding taxation for interest income and a separate withholding taxation for dividend income; however, these are treated as consolidated taxation in this model. All corporate pensions, individual pensions, etc. under the Comprehensive Survey of Living Conditions are treated as pension income.

Other income under the Comprehensive Survey of Living Conditions is not used in this model since it is defined as temporary allowances, celebration money or condolence money for ceremonial occasions, and it is not highly necessary to consider them concerning policies related to taxation and the social security system.

In this model, Fitted Values of income tax amounts and residential tax amounts are obtained by the following calculation method. In principle, calculate taxable income based on individual income information and family unit information after consideration of various exemptions, and then calculate the Fitted Values of tax amount. First, calculate the total income of all individuals by the following method. The terms placed between [TERM] in the following formula indicate the variable name for individual questionnaire data.

Employment income = [Compensation of employment] – Employment income exemption

Pension income = [Public pension or Governmental pension] + [Corporate pension or Individual pension, etc.] – Public pension, etc. exemption

Employer income = [Business income] + [Agriculture or Livestock income] + [Industrial homework income]

Total income = Employment income + Pension income + Employer income + [Property income]

The employment income exemption and public pension, etc. exemption can be calculated by applying the compensation amount of employment and the pension amount to systems respectively.

Next, apply the income exemption to all individuals and calculate taxable income.

Temporary taxable income 1 = Total income – Basic exemption – Social insurance premiums exemption

Temporary taxable income 2 = Temporary taxable income 1 – Exemption for spouse – Special exemption for spouse

Taxable income = Temporary taxable income 2 – Exemption for dependents

The income exemptions to be considered in this model are the basic exemption, exemption for spouse, exemption for dependents, and social insurance premiums exemption. When
calculating social insurance premiums exemptions, use the total value of pension insurance premiums, health insurance premiums, long-term care insurance premiums, and other insurance premiums (mainly employment insurance premiums) entered in the questionnaire. First, deduct the basic exemption and social insurance premiums exemption from total income and treat the result as temporary taxable income 1. If an individual has a spouse, apply the exemption for spouses. In this model, apply exemption to one of a married couple who has higher temporary taxable income based on the assumption where family units reasonably act to minimize the total taxable income amount of the family units. In the same way, deduct exemptions for spouse from temporary taxable income 1 and treat the result as temporary taxable income 2. If there is a dependent(s), apply the exemption for dependents. Apply the exemption to the largest income earner of the family unit (or the person who has the largest temporary taxable income 2).

Use the income after the exemption for dependents is applied as the taxable income and calculate the income tax payment amount (Fitted Value) by applying 2009 taxation and the residential tax payment amount (Fitted Value) by applying 2010 taxation. The income tax payment amount (Indicated Value) in 2009 and the residential tax payment amount (Indicated Value) in 2010 are entered in the questionnaires of the Comprehensive Survey of Living Conditions and the calculation method of the Fitted Values corresponds to them. Payment amounts of income tax and residential tax are treated by each household.

III-2. Comparison of Indicated Values and Fitted Values

We will compare Indicated Values and Fitted Values here and examine the approximation and disparities between them. Figure 7 is a scatter diagram of Indicated Values and Fitted Values. The unit amount is 10,000 yen (annual amount). Both income tax and residential tax are distributed approximately on the 45-degree line, which shows that the number of family units where the Indicated Value and Fitted Value are consistent is relatively large. However, both income tax and residential tax have disparities between Indicated Values and Fitted Values in family units with small payment amounts. This is observed in both cases where the Indicated Value is larger and where the Fitted Value is larger respectively.

In Figure 8, Kernel density estimation is implemented with distributions of Indicated Values and Fitted Values and the results are compared. In Panel (1) Income Tax, distributions of Indicated Values and Fitted Values are almost similar; however, some disparities are found where the payment amount is close to zero. This is detected also in Panel (2) Residential Tax.

Next, disparities between Indicated Values and Fitted Values are the focus of examination. The disparities will be indicated against income (hereinafter referred to as “Disparity Rate”). Figure 9 shows the distribution of the Disparity Rate. Looking at the graph, they are almost symmetrical with a center at zero; however, the frequency is slightly higher on the left end. These characteristics are detected in both income tax and residential tax. Descriptive statistics of Figure 9 is Column (1) of Table 1. The mean Disparity Rate of income tax
is 0.28% over income; 5th percentile is negative 2.04%; and 95th percentile is 2.90%. The mean Disparity Rate of residential income is negative 0.31% over income; 5th percentile is negative 3.40%; and 95th percentile is 2.45%. Therefore, the mean Disparity Rates of income tax and residential tax are almost zero and dispersion (based on 5th percentile and 95th percentile) is approximately 3% (2.04% to 3.40%).

These results will not be affected even if eliminating family units where Indicated Values and Fitted Values of the burden amount are zero, from samples. It may not be so difficult to lead also to zero Fitted Value for family units for which the indicated Value of the burden

Figure 7
Scatter Diagram of Indicated Values and Fitted Values

(1) Income Tax
(2) Residential Tax

(Note) Vertical axis indicates Indicated Values and horizontal axis indicates Fitted Values. (Unit amount is 10,000 yen.)

Figure 8
Kernel Density Estimator of Indicated Values and Fitted Values

(1) Income Tax
(2) Residential Tax

(Note 1) Vertical axis indicates densities and horizontal axis indicates payment amounts. (Unit amount is 10,000 yen.)
(Note 2) Solid line indicates Indicated Values and dotted line indicates Fitted Values.
amount is zero. It is rather important as a simulation model how close the Fitted Value level can be led to the Indicated Value with respect to family units for which the Indicated Value is positive. As indicated in Column (2) of Table 1, the mean Disparity Rate of income tax is 0.37% over income, 5th percentile is negative 2.35%, and 95th percentile is 3.83%. Similarly, the mean Disparity Rate of residential tax is negative 0.38% over income, 5th percentile is negative 3.74%, and 95th percentile is 2.81%. Therefore, there are no major changes in distribution of the Disparity Rate of both income tax and residential tax when compared with the aforementioned cases where all samples are used.

In cases where the Disparity Rate (over income) is used for the size of the disparity, even if the Disparity Rate is the same, the disparity amount will increase as the income of the family units in question increases (based on amount). In other words, using the Disparity Rate results in a relative underestimation of the disparity of high-income family units. Then, we use all samples and check distribution where the Disparity Rate of each family unit is

Figure 9
Distribution on Disparity Rate: Histogram (In Cases of All Observation Values)

(Note 1) Vertical axis indicates frequency and horizontal axis indicates Disparity Rate.
(Note 2) Disparity Rate = (Indicated Value – Fitted Value) / Household income

Table 1
Distribution of Disparity Rate: Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>(1) In cases of all observation values</th>
<th>(2) In cases of excluding family units with nil burden amount</th>
<th>(3) In cases of being weighted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Income tax</td>
<td>Residential tax</td>
<td>Income tax</td>
</tr>
<tr>
<td>Average</td>
<td>0.28%</td>
<td>-0.31%</td>
<td>0.37%</td>
</tr>
<tr>
<td>5th percentile</td>
<td>-2.04%</td>
<td>-3.40%</td>
<td>-2.35%</td>
</tr>
<tr>
<td>95th percentile</td>
<td>2.90%</td>
<td>2.45%</td>
<td>3.83%</td>
</tr>
<tr>
<td>Amount of observation values</td>
<td>19,926</td>
<td>20,295</td>
<td>15,185</td>
</tr>
</tbody>
</table>
weighted by the income of the family unit in question. Distribution of weighted Disparity Rates of both income tax and residential tax gives results that are almost the same as cases of using all samples. According to Column (3) of Table 1, which shows descriptive statistics, the mean Disparity Rate of income tax is 0.02% over income, 5th percentile is negative 3.19%, and 95th percentile is 3.34%. Similarly, the mean Disparity Rate of residential tax is negative 0.46% over income, 5th percentile is negative 3.73%, and 95th percentile is 2.20%.

As mentioned above, it is indicated that the mean Disparity Rates of income tax and residential tax are almost zero and that dispersion (in cases of using all samples) is approximately 3% (2.04% to 3.40%). The sample size of this model is approximately 20,000 family units and therefore it is valued that the mean of the sample mean of Disparity Rates is zero and dispersion is almost zero (= 0.03/√20000). Therefore, Fitted Values of microsimulation analysis, etc. have almost no bias in relation to tallied macro values, such as total tax revenues, and it results in an estimated value with almost no errors. In this context, its accuracy can be used sufficiently for the assessment of policies.

IV. Causes of Disparities between Fitted Values and Indicated Values

IV-1. Causes of Disparities

The causes of disparities are examined here. First, disparities are classified into three categories based on the size of the Disparity Rates: negative disparity is large; disparity is small; and positive disparity is large. Then, family units are classified into nine classes in total, including three classes in income tax and three classes in residential tax (3*3=9). According to the distribution of Disparity Rates (Section III), the dispersion of the Disparity Rate (in cases of using all samples) is approximately 2.04% to 3.40%. Then, when categorizing Disparity Rates, cases where the Disparity Rate of income tax is smaller than negative 0.04 are categorized as “negative disparity is large;” cases where the Disparity Rate is negative 0.04 or above and 0.04 or less as “disparity is small;” and cases where the Disparity is more than 0.04 as “positive disparity is large.” Similarly, cases where the Disparity Rate of residential tax is smaller than negative 0.04 are categorized as “negative disparity is large;” cases where the Disparity Rate is negative 0.04 or above and 0.04 or less as “disparity is small;” and cases where the Disparity is more than 0.04 as “positive disparity is large.”

Assumed causes of disparities are divided into three patterns. First, it is the case where Disparity Rates of both income tax and residential tax are large and disparity directions (a sign of the Disparity Rate) are the same. In this case, the cause of disparity is that the Fitted Values of tax amount may not be calculated correctly due mainly to the impact of the model or entry errors with income in the questionnaire. Second is the case where the Disparity Rate of either one of income tax or residential tax is large. In this case, since either income

---

5 This research examined three cases with respect to category thresholds: (1) in cases of +/-0.03; (2) in cases of +/-0.04; and (3) +/-0.05 respectively and the same trend is detected in all cases. This paper introduces measurement results of “in cases of +/-0.04.”
tax or residential tax has no disparity, there is a small possibility that the Fitted Value of the tax amount is not calculated correctly due to the impact of the model or entry errors with the income in the questionnaire. Therefore, the cause of the disparity is that the Indicated Value of the tax amount was not entered correctly due mainly to entry errors with the tax amount in the questionnaire. Third is the case where Disparity Rates of both income tax and residential tax are large, but disparity directions (a sign of the Disparity Rate) are different. In this case, the cause of the disparity is considered to be the incorrectness of the overall entry.

Table 2 shows the percentage of family units for each of the nine categories. The percentage of family units with a small Disparity Rate in both income tax and residential tax is 88.73% and it shows even from the micro level perspective that most of family units have almost no disparity.

Table 2
Categories of Disparity Rate: Percentage of Family Units

<table>
<thead>
<tr>
<th>Income tax</th>
<th>Residential tax</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Negative disparity is large.</td>
</tr>
<tr>
<td></td>
<td>1.04%</td>
</tr>
<tr>
<td>Disparity is small.</td>
<td>3.30%</td>
</tr>
<tr>
<td>Positive disparity is large.</td>
<td>0.06%</td>
</tr>
<tr>
<td>Total</td>
<td>4.40%</td>
</tr>
</tbody>
</table>

On the contrary, 11.27% of family units have a large Disparity Rate. From among them, the percentage of family units for which the cause of disparity seems to be an entry error of the tax amount is 9.56% (= 0.83% + 3.30% + 3.97% + 1.46%) and it represents a large percentage of said family units. Meanwhile, the percentage of family units for which the overall entry seems to be incorrect is 0.07% (= 0.01% + 0.06%) and it is extremely rare. Consequently, it is found that the impact of entry errors on the tax amount is high.

In the following, we focus on frequent cases from among cases where the cause of disparity seems to be an entry error of the tax amount, and look at their characteristics and the cause of disparity.

First, we focus on cases of small disparity in income tax and large negative disparity in residential tax. According to Table 2, 3.30% of family units correspond to this case. The Disparity Rate in residential tax is negative and the Indicated Value is smaller than the Fitted Value. This indicates that the Indicated Value of the residential tax amount is too small due to entry errors of the tax amount. Causes of the disparity are considered to be errors in entering the residential tax amount as too small due to incorrect entry in the previous fiscal, in addition to errors in entering the tax amount as too small due to errors in digits. Therefore, in cases where the incorrect residential tax amount was entered in the previous fiscal year and income in the previous year increased significantly, there is a possibility that disparities
will occur between the Indicated Value of the tax amount and the Indicated Value of income.

Next, we focus on cases of small disparity in income tax and large positive disparity in residential tax. According to Table 2, 1.46% of family units corresponds to this case. The Disparity Rate in residential tax is positive and the Indicated Value is larger than the Fitted Value. This indicates that the Indicated Value of the residential tax amount is too large due to entry errors of the tax amount. The causes of the disparity are considered to be errors in entering the tax amount as too large due to incorrect entry in the previous fiscal year with residential tax, in addition to errors in entering the tax amount as too large due to an error in digits. Therefore, in cases where the incorrect residential tax amount was entered in the previous fiscal year and income in the previous year decreased sharply due to the loss of a job, retirement, etc., there is the possibility that a disparity will occur between the Indicated Value of the tax amount and the Indicated Value of income.

Finally, we focus on cases of large positive disparity in income tax and small disparity in residential tax. According to Table 2, 3.97% of family units corresponds to this case. The Disparity Rate in income tax is positive and the Indicated Value is larger than the Fitted Value. This indicates that the Indicated Value of the income tax amount is too large due to entry errors of the tax amount. The Cause of the disparity is considered to be errors in entering the tax amount as too large due to an error in digits, etc.

Table 3
Distribution of Disparity Rate: In cases of eliminating family units which have a person(s) who earns business income and agriculture or livestock income

<table>
<thead>
<tr>
<th></th>
<th>(Posted again) In cases of excluding family units with nil burden</th>
<th>In cases of eliminating family units with business income, etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Income tax</td>
<td>Residential tax</td>
</tr>
<tr>
<td>Mean</td>
<td>0.37%</td>
<td>-0.38%</td>
</tr>
<tr>
<td>5th percentile</td>
<td>-2.35%</td>
<td>-3.74%</td>
</tr>
<tr>
<td>95th percentile</td>
<td>3.83%</td>
<td>2.81%</td>
</tr>
<tr>
<td>Amount of observation value</td>
<td>15,185</td>
<td>16,437</td>
</tr>
</tbody>
</table>

(Note) Family units for which the Indicated Value and Fitted Value are zero are eliminated here.

IV-2. Verification of Impact of Entry Errors

IV-2-1. Impact of Entry Errors of Business Income and Agricultural or Livestock Income

The impact of individual entry errors will be verified from among causes of disparity in the following. First, we focus on the possibility of entering (deducted) income as too large due to entry errors with business income or agricultural or livestock incomes, and verify whether it causes the disparity. If there is this type of entry error, the Fitted Value of the tax amount is calculated to be too large and it may cause a negative disparity.
In order to verify this possibility, we measure the distribution of Disparity Rates after excluding family units which have a person(s) earning business income or agricultural or livestock income. (In this paper, family units for which the Indicated Value and Fitted Value are zero are also excluded. The results will be compared later with the results of Column (2) of Table 1 which also excluded family units for which the Indicated Value and Fitted Value are zero.) According to Table 3, the average Disparity Rate of income tax is 0.53%, 5th percentile is negative 2.02%, and 95th percentile is 4.07%. Comparing these results with the results of Column (2) of Table 1, it is found that the distribution of Disparity Rates shifted to the positive direction and that the number of family units for which the Disparity Rate is negative (family units for which the Indicated Value is smaller than the Fitted Value) decreased. The mean Disparity Rate of residential tax is negative 0.24%, 5th percentile is negative 3.20%, and 95th percentile is 2.89%. Comparing these results with the results of Column (2) of Table 1, the distribution of Disparity Rates shifted to the positive direction, which is the same as the case of income tax. Table 4 shows classes of Disparity Rates. Compared with the results in Table 2, the percentage of family units where negative disparities with both income tax and residential tax are large, decreased, and this shows the same results as the aforementioned cases.

Based on the above, the fact that (deducted) income is entered as too large due to a typo with business income and agricultural or livestock income may be the cause of the disparity.

IV-2-2. Impact of Entry Errors with Digits

Next, we focus on the possibility of entering the tax amount as too large due to entry errors with digits and verify whether this fact caused the disparity. In cases of these recording errors, the Indicated Value of the tax amount becomes too large and it may cause a positive disparity.

In order to verify this possibility, first, we limit family units for which the Disparity Rate of income tax is 0.1 or larger and the Disparity Rate of residential tax is 0.05 or more, and then measure the distribution of Disparity Rates. (In this case, family units for which the Indicated Value and Fitted Value are zero are excluded.) According to Column (1) of Table 5, the mean Disparity Rate of income tax is 24.45%, 5th percentile is 10.40%, and 95th percentile is 51.44%. Compared with the results of Column (2) of Table 1, it is found that the distribution of Disparity Rates shifted to the positive direction. The mean Disparity Rate of residential tax is 9.76%, 5th percentile is 5.17%, and 95th percentile is 23.99%. Compared with the results of Column (2) of Table 1, distribution of Disparity Rate shifted to positive direction. These results are obtained naturally since the data is limited to family units with a high Disparity Rate.

Next, with regard to income tax, limit family units for which the Disparity Rate is 0.1 and above like the previous case, divide the Indicated Value of the tax amount of said family units by 10, and measure the distribution of Disparity Rates. With respect to the residential tax, also limit family units for which the Disparity Rate is 0.05 or larger, divide the Indicated Value of the tax amount of said households by 10, and measure the distribution of Dis-
parity Rates. (Family units for which the Indicated Value and Fitted Value are zero are eliminated here.) According to Column (2) of Table 5, the mean Disparity Rate of income tax is 0.41%, 5th percentile is negative 4.66%, and 95th percentile is 4.14%. Compared with the results of Column (2) of Table 1, the distribution of Disparity Rates becomes very close. The mean Disparity Rate of income tax is negative 0.70%, 5th percentile is negative 3.80%, and 95th percentile is 1.40%. Compared with the results of Column (2) of Table 1, the distribution of Disparity Rates becomes very close.

Based on the above, the fact that the tax amount is entered as too large due to an entry error with digits may also be the cause of the disparity.

IV-3. Consideration of Impact of the Model: Focusing on Tax Relief for Housing Loans

As mentioned above in Section III, most of the disparities between Indicated Values and Fitted Values of both income tax and residential tax are small. On the other hand, one of groups with a prominent disparity is the group of family units with a large loan amount. Fig-

Table 4

<table>
<thead>
<tr>
<th>Classes on Disparity Rate: In cases of eliminating family units which have a person(s) who earns business income and agriculture or livestock income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income tax (Post again) In cases of excluding family units with nil burden</td>
</tr>
<tr>
<td>Income tax</td>
</tr>
<tr>
<td>Average</td>
</tr>
<tr>
<td>5th percentile</td>
</tr>
<tr>
<td>95th percentile</td>
</tr>
<tr>
<td>Amount of observation values</td>
</tr>
</tbody>
</table>

(Note) Family units for which the Indicated Value and Fitted Value are zero are also eliminated here.
Figure 10 shows the distribution of disparities (amount) between Indicated Values and Fitted Values. In this Figure, the solid line indicates the distribution of family units with a loan amount of over 10 million yen and the dotted line indicates the distribution of family units with a loan amount of less than one million yen. The group of family units with a loan amount of over 10 million yen has a high frequency of family units with large disparities; in particular, a high frequency is found at the left end. This means that Indicated Values tend to be smaller than Fitted Values particularly in this type of group and this trend can be found in both income tax and residential tax.

One of the backgrounds of this trend can be the impact of tax relief for housing loans, which is not considered in the calculation of Fitted Values. This model is therefore partially modified to reflect the housing loan deduction. In the questionnaire of the Comprehensive Survey of Living Conditions, the floor area of residences and loan balance can be used. Based on the current system, if a family unit satisfies the following two requirements - (Requirement 1) floor area is $50m^2$ or more, and (Requirement 2) total income amount is 30 million yen or less—the specified rate of the loan amount will be deducted from the income tax amount. In this paper, we demonstrate the following two cases: (1) in cases of deducting 1% of the loan amount; and (2) in cases of deducting 0.5% of the loan amount.

Figure 11 shows the distribution of disparities (amount) after reflecting the tax relief for housing loans. Like Figure 10, the solid line indicates the distribution of family units with a loan amount of over 10 million yen and the dotted line indicates the distribution of family units with a loan amount of less than one million yen. In this case, the mountain of the distribution of the group of family units with loan amount of over 10 million yen moved to right side and frequency is high at the right end. This means that reflecting the tax relief for
housing loans in the model resolved the trend where Indicated Values become smaller than the Fitted Value; however, Indicated Values then tend to become larger than Fitted Values. This is detected in both cases (1) in cases of deducting 1% of loan amount and (2) in cases of deducting 0.5% of the loan amount. This is because the aforementioned simple method is adopted for calculation of Fitted Values and it has an impact on reducing the Fitted Value of the income tax amount of family units that are not actually using tax relief for housing loans. For example, actual tax relief for housing loans has a deduction period (ten years in principle) and a deduction amount limit exists depending on the number of years of using the system. In this study, these points cannot be reflected due to limited questionnaire information, such as the time of acquisition of a residence, etc. Another reason is that Indicated Values may have problems. If a family unit that is actually using tax relief for housing loans, such as a workers’ family unit, uses the tax relief for housing loans when filing an income tax return, the family unit may forget to reflect the tax reduction in the Indicated Value of the income tax amount. In this case, the Indicated Value becomes too large due to a recording error.

Figure 11
Distribution of Disparities between Indicated Values and Fitted Values:
In cases of considering tax relief for housing loans

(1) Income tax (in cases of deducting 1% of loan amount)  (2) Income tax (in cases of deducting 0.5% of loan amount)

(Note 1) Horizontal axis indicates disparity amount between Indicated Values and Fitted Values. (If the Indicated Value is larger, it is a positive value.)
(Note 2) The solid line indicates the group of family units with 10 million yen or more of loan amount and the dotted line indicates the group of family units with less than one million yen of loan amount.

As stated above, it may be important for calculating Fitted Values to consider tax relief for housing loans, while in order to reflect it more adequately, it is necessary to work additionally to carefully select target family units to apply this system. However, in order to implement a careful selection, more detailed information than the current questionnaire of the Comprehensive Survey of Living Conditions may be necessary.
V. Conclusion

This paper aims (1) to examine the characteristics of various statistics by comparing Indicated Values in response to questionnaires concerning various taxes and social insurance premiums and (2) to verify the validity of fitted values of the household tax burden (in terms of income and residential taxes) by comparing the Indicated Values with the Fitted Values.

First, the comparison of Indicated Values in different statistics showed that Indicated Values of tax and insurance premiums were underestimated in the National Survey of Family Income and Expenditure and the Family Income and Expenditure Survey. In addition, as a result of the comparison between the Indicated Values and Fitted Values and the examination of the distribution of disparities between them, it was found that the mean disparity between the Indicated and Fitted Values of income and residential taxes was zero and the dispersion was around 3%. It was also confirmed that the disparity occurred frequently because of entry errors for the tax amount and that entry errors for values in questionnaires that affected the disparity included “business income entry error” and “tax amount entry error due to errors in digits.”

The implication of the examination is that Fitted Values obtained through microsimulation analysis are estimates with little bias regarding the macro values obtained through collection of data and that there is little margin of error. In that sense, these Fitted Values have sufficient precision to be used for the evaluation of policies.

On the other hand, it is indicated that it is important for improving the appropriateness of Fitted Values to consider tax relief for housing loans. This paper has tried to include tax relief for housing loans when calculating Fitted Values; however, in order to reflect it more adequately, it is necessary to work additionally to carefully select target households to apply this system. In order to implement it, it may be necessary to have more detailed information than the current questionnaire of the Comprehensive Survey of Living Conditions.

References


Measures, University of Tokyo Press (in Japanese)


