

## SURVEY SYSTEMS AND SAMPLING DESIGNS OF CHINESE HOUSEHOLD SURVEYS, 1952-87

YOSHIRO MATSUDA

### I. INTRODUCTION

**F**OLLOWING the establishment of the People's Republic of China in 1949, the Chinese government built up statistical survey systems that cover the whole of mainland China. The government relies mainly on two systems of complete enumeration surveys used to carry out the central planning of the economy. One is the census type surveys such as the population census (first conducted in 1953) and the manufacturing census (first conducted in 1950). The other is the regular reporting system made up of the individual survey units. The survey units are closely related to the accounting units which differ somewhat from regular firms or organizations. Sometimes a unit may correspond to a conglomerate of firms and organizations, or even to an entire local government. In addition to these two types of complete enumeration survey systems, two types of sampling surveys have sometimes been carried out to get detailed information for planning. One is a crop-cutting survey to estimate the harvest and the other is a household expenditure survey to check on the living conditions of the people. In the past, however, sampling techniques were in most cases based on a so-called typical sampling due to Mao Zedong's philosophy and his survey experience. Moreover, during the Great Leap Forward movement and Cultural Revolution even such typical sampling surveys were abandoned. (See [11] [12].)

The following sections of this paper provide a brief history of household expenditure surveys in China based on articles and tables gleaned from official documents. The survey methods were disclosed in statistical journals only before 1958; thereafter came a long period of silence. During the Cultural Revolution not only information on survey techniques but no sort of statistical figures were made public. Following the end of the Cultural Revolution, statistical figures were again released, at first individually, but later large numbers of publications on the current survey systems and statistical figures began appearing. However, even in these papers detailed technical information was difficult to obtain. Recently three

---

This research was supported in part by a grant from the Ministry of Education and Cultural Affairs during the 1988-89 academic years (Scientific research on priority areas: comparative studies on East Asia. Grant number: 01605006). The first English version of this paper was read at the Third Japan-China Symposium on Statistics in 1989 [13]. The author is also greatly indebted to Professor Reetsu Kojima of Daito Bunka University for his collaboration and wishes to thank the anonymous references for their comments on the earlier version.

important books on the history of statistical surveys were published [6] [16] [17].<sup>1</sup> Based on these documents, the writer will try to reconstruct a tentative history of the survey systems and techniques used for the household expenditure surveys in China. The purpose for such a reconstruction is to evaluate the quality of these statistics.

Before discussing the Chinese situation, some words are needed about household expenditure surveys in general. In most countries doing household income and expenditure surveys, there is a difficulty overcoming the tendency of respondents to hide their income. On the other hand people are not so hesitant to answer about their expenditures in principle. But expenditure surveys require minute daily accounts keeping by respondents, and as a result the refusal ratio to be surveyed can also be very high. In order to improve the accuracy of an income survey, it must be supplemented by an expenditure survey because the difference between total income and total expenditures should result in a savings or an income deficit. Thus the revision of an income survey is possible after completing the record-keeping expenditure survey. However, there can be serious problems maintaining the interest of the record-keeping households during the whole survey period, and efforts have to be made to overcome people's boredom. Thus in most countries statistical bureaus adopt a rotation sampling method which requires each household to keep daily accounts for only a short time, usually a month or a few months. In addition, if the same sample households are used for more than two years, in most cases, they lose their representativeness because they became more careful about their household management through their experiences in keeping daily accounts.

Another important matter is the method of preparing a sampling frame for selecting respondent households. There are at least two ways of preparing a sampling frame. One is to carry out a large-scale income survey. This is the method the Chinese State Statistical Bureau has adopted. The other is to use area data obtained from the population census, housing census, etc. which supply comprehensive and exhaustive information about households in a specified area. This method is preferred in many developed countries.<sup>2</sup> With these general remarks, we will examine the history of household expenditure surveys in China.

The People's Republic of China organized its State Statistical Bureau (SSB) in

<sup>1</sup> A short bibliography of articles and monographs published before the Cultural Revolution is annexed to Matsuda [12]. Reading between the lines of the controversial publications, the three books mentioned in the text are quite informative. The State Statistical Bureau [6] describes the development of various statistical activities as viewed by the current SSB. Wang [16] follows the same approach as the State Statistical Bureau [6] but contains much more detailed information about the controversies during the Great Leap Forward movement and Cultural Revolution quoting the various documents which were not opened to the public. Yu [17] provides a more authoritative presentation of the history of statistical surveying in China than does the State Statistical Bureau [6] or Wang [16], but Yu puts more emphasis on describing the current situation.

<sup>2</sup> Some developing countries which have a good deal of experience in statistical surveying but do not have reliable population census results use area maps or village maps instead of population-census results as a sampling frame.

1952, basing it on the provincial statistical bureaus which had been established in three Northeastern provinces soon after the end of the civil war. Before organizing the SSB, the government carried out the first manufacturing census of its nationalized factories in 1950. However, to conduct nation-wide surveys, a national central bureau for statistics was needed, and the SSB was set up for this purpose. The first example of a nation-wide statistical survey was the Manufacturing and Agricultural Products Survey in 1952 which was sometimes referred to as the "Two Terms Survey" following the terminology of Mao Zedong's original instructions to the government. This survey listed all of the townships (*xiangs*) in the whole of mainland China, and was followed in 1953 by the population census. With these surveys China obtained a list of the townships, households, and nationalized factories throughout the nation, and as part of the preparations to carry out sampling surveys, the SSB had started a household income and expenditure survey.

The household income and expenditure survey in China has two series. One is for farm households which make up the majority of rural dwellers which is intended to survey these households, and the other is for worker households which was later extended to cover all kinds of city dwellers.<sup>3</sup> The former has been carried out on a nation-wide scale using data since 1954, and the latter using data since 1956.

## II. RURAL HOUSEHOLD INCOME AND EXPENDITURE SURVEY

In the Northeastern provinces where the provincial statistical bureaus had been established, household income and expenditure surveys started in the late 1940s. The establishment of the SSB in 1952 enabled the government to carry out new nation-wide rural household (*nongjia*) income and expenditure surveys based on the experiences in the Northeastern provinces and using the results of national census type surveys. The first trial was the 1954 Rural Household Income and Expenditure Survey (1954 nian nongjia shouzhi diaocha) for which a pilot survey had been made and preparation started in 1953 in collaboration with statisticians from the USSR. A point to be noted is the discrepancy between the year when the survey was taken and the date of the data. The year 1954 used in the title, the 1954 Rural Household Income and Expenditure Survey, means that the date of the data is for 1954 and does not mean the date of survey. In this case planning and preparation of the survey was started in 1953 and field operation was carried out in 1955.

The purpose of the survey was to compare rural household conditions in 1954 with those before the land reform; therefore comparable facts from before the land reform were also gathered. Because of this the dates of facts varied from province to province because the land reform was carried out across the country over a period of years. The survey items contained data not only about household

<sup>3</sup> Since 1984 the farm household survey has been transformed into rural household survey [17]. The non-farming income is now taken into account as an income source, however, it is not explicitly mentioned that non-farming households in rural areas have been selected.

expenditures for daily family consumption and production (flow data as a consumer and producer), but also about capital assets such as household land-holdings, draft animals for cultivation, and production equipment (stock data for production).

The sampling technique adopted was the systematic sampling procedure which can be regarded as a substitute of random sampling techniques. The SSB faced a variety of opposition from statisticians who were not accustomed to such a sampling method and who stuck to Mao's survey theory because the random sampling technique required a comparatively large number of samples to get a reasonable national average. The sample of the 1954 survey was to be 15,000–20,000 households at the time of designing the survey. The actual sample size surveyed was 15,432.

The sampling procedure the SSB adopted was to require that each province chose 20 households from selected sample township (*xiangs*) for a total of 500–1,000 households per province. The number of sample townships was derived from the fraction of assigned households in each province divided by twenty. The State Statistical Bureau's initial design was to choose sample townships systematically from the list of townships innumeraled in each counties (*xians*). However, most provinces used a mixture of a systematic sampling and a representative sampling. Provincial statistical bureaus first prepared area maps of the townships zoned by the rural conditions and terrain, such as being irrigated plains, on non-irrigated plains, in mountaineous areas, etc. Then they chose townships using the representative selection method. After choosing the representative townships, all the households were listed and numbered. From the list about 20 households were selected ranging from high to low income.

Being China's first experience for a nation-wide sampling survey, it took a very long time to process all the survey results. In 1956, before the results of the 1954 Rural Household Income and Expenditure Survey were completely released, they started a new household expenditure survey for the 1955 data. This new survey excluded expenditures for household production which had now been collectivized and whose production figures were amalgamated into the statistics of the collectives.

This survey done in 1956 for the 1955 data became a model for rural household surveys used subsequently. The survey was divided into two types of surveys. One was comparable to the 1954 survey where the data covered household accounts which could be remembered. The other required that specific households keep a selected daily account of their expenditures. The former corresponded to a survey for compiling a sampling frame. The sampling procedures were much the same as those used in the 1954 survey.

In working up the technical aspects of the two sampling survey systems as a whole, the SSB adopted a two-stage sampling. They first selected townships and compiled a list of households arranged in descending order of income level within a given township. They then selected households which would keep a daily record of income and expenditures. These households were selected at intervals from the top to the bottom of the list. The first-stage survey was supposed to be carried out in 1956. The second-stage sampling survey based on the results of the first-stage survey in 1956 was carried out from January to December 1956 for

collecting monthly data. The households keeping daily records numbered 14,172. In this new survey the SSB endeavored to draw up more detailed area maps in order to overcome defects in the 1954 survey. This time they selected counties (*xians*) and then chose three communes or townships (*xiangs*) from each county; and 15 households were then selected from each commune or township for a total of 45 households which were used as record-keeping households within the county selected.

During the Great Leap Forward movement household income and expenditure surveys were severely attacked, and following a statistical work conference in June 1958 at Baoting, these were terminated from first half of 1958. The left-wing statisticians opposed the sampling surveys claiming these suppressed the autonomous will of the people because they were centrally controlled by the specialists and intellectuals in SSB. The 1958 statistical work conference was different from statistical conference that had been held regularly before. The left-wing mobilized non-specialists in statistics as a pressure group. The official reasoning for suspending surveys was that the communization system like common dining rooms and kitchens made it difficult to keep records of every household account. In 1962 after the turmoil of the Great Leap Forward, the SSB was reorganized. The sample size of the 1962 survey was modest and amounted to 4,958 households. But the next year this doubled, and by 1965 it had reached the 1954 level. However surveying was again abandoned with the outbreak of the Cultural Revolution in 1966.

During the Cultural Revolution the nation-wide statistical survey system collapsed. The SSB was reorganized in 1969 as a small section of the State Planning Commission with only 15 members. By comparison in 1957 at its most active time, SSB had 675 members including 611 trained statisticians. Provincial statistical bureaus ceased to exist except for the one in Beijing, and statistical data and documents maintained in statistical offices became scattered during these years.

Comprehensive national figures from 1962 to 1965 were not fully released even after the Cultural Revolution. We only know the rough outline of the survey scale and quite abridged figures for 1965. Recently the data for the rural household survey which were dispersed during the Cultural Revolution were regathered and published under the title *Gasheng zizhiqu zhixiashi nongmin shouru xiaofei diaocha yanjiu huibian* [Selected materials on the rural household income and expenditure survey] [5]. It contains time-series data by province; however these are far from comprehensive and cross-comparison of data between provinces is difficult. (As to intertemporal comparison between regions using variance analysis, see [14]). In addition these materials lack sufficient detailed information on sampling design and size. Table I shows the size of samples from 1954 to 1988 for the entire nation.

With the passing of the Cultural Revolution, the SSB was set up once again and the SSB requested provinces to conduct surveys, if possible, in 1977. Only seventeen provinces submitted results from a total of 3,646 households surveyed. The rural household income and expenditure survey was restored in abridged form in 1978. The sample size was only 6,095 households and covered twenty provinces.

TABLE I  
SAMPLE SIZE OF RURAL HOUSEHOLD INCOME AND EXPENDITURE SURVEY

	Number of Households Surveyed	Notes on Coverage
1954	15,432 <sup>a</sup>	23 provinces
1955	—	(Preliminary survey)
1956	14,172	
1957	17,378	
1958 first half second half	18,000	770 counties, 1,205 cooperatives (suspended due to Great Leap Forward movement)
1962	4,958	27 provinces
1963	9,724	
1964	12,095	
1965	11,632 <sup>b</sup>	
1966-76	(suspended)	
1977	3,646	17 provinces (tentative)
1978	6,095	20 provinces (partial)
1979	10,282	23 provinces (partial)
1980	15,914	27 provinces (partial)
1981	18,529	28 provinces (partial)
1982	22,775	
1983	30,427	0.17% of all rural households
1984	31,375	29 provinces (846 counties)
1985	66,642	
1986	66,836	
1987	66,912	
1988	67,186	
1989	66,906	

Sources: [12] based on various documents published in China; and State Statistical Bureau, *Zhongguo tongji nianjian* [Statistical yearbook of China], various editions.

<sup>a</sup> Figures in [2]. In Yu [17] the number of households was 16,486 in 23 provinces and autonomous regions. The differences of 1,054 may be attributed to the later arrival of two provinces. See lines 1 and 2 of Table V.

<sup>b</sup> In Yu [17] the number was 11,683.

The survey system was basically the same as before. The major difference from before was the method of constructing the sampling frame and the size of the sample. These grew from 10,282 households in 23 of the provinces, autonomous regions, and municipalities in 1979 to 30,427 samples in all 28 provinces, autonomous regions, and municipalities in 1983. It covered about 0.17 per cent of all rural households. Thereafter the SSB endeavored to expand the number of sample households, and by 1985 these had increased by more than twice. This increase made it possible to compare regional differences and to analyze the patterns of farming undertaken by individual farming households which had farmed private plots and rented cultivated areas from communes.

### III. WORKER AND URBAN HOUSEHOLD INCOME AND EXPENDITURE SURVEYS

As mentioned in Section II, various regional household expenditure surveys for worker households in addition to rural households were done in the Northeastern provinces before 1954. The survey experience in these provinces was mainly related to the compilation of a consumer price index and a retail sales price index. To compile a general price index, weights for aggregating each commodity price index are needed. Such weights are usually obtained from consumer expenditure surveys. The importance of a consumer expenditure survey was stressed at the first national statistical conference in July 1951. As with the 1954 Rural Household Income and Expenditure Survey, statisticians from the USSR also cooperated with the SSB to carry out the worker household income survey. In 1955 the SSB issued a decree for the national survey of employed worker household income and expenditures in order to carry out the 1956 survey.

This 1956 survey was carried out in three-stage samplings. The first was the selection of twenty-seven cities by means of a typical sampling method. The second stage was a systematic sampling of state-owned manufacturing enterprises, and the third stage was a systematic sampling of employees from the list of employees working at the selected enterprises. The employees selected were asked to answer about his or her household as a whole. The sample size was around 6,000 households and not big enough to get regional figures. In 1957 the sampling scale was expanded. Thirty-two cities were chosen and the industries covered were manufacturing, commerce, education, and the government sector. In total 7,000 households were surveyed.

During the ensuing years the worker household income survey suffered much the same fate as the rural household income and expenditure survey. It was greatly reduced in 1958, then entirely stopped in 1960, but restored in 1961. From 1961 to 1966 the survey went through what could be called its second phase. The 1961 survey was a compromise between left-wing and right-wing statisticians at the SSB, and the survey itself was organized by the office of finance and trade under the State Council. The sample size was only 1,000 households selected from twenty-eight cities. In 1964 the survey was restored to the scale attained in the 1950s. In the 1964 survey, the size of the preliminary survey to measure income levels was 139,000 households from fifty-nine cities and twenty-four county towns (*xianchengs*) and that for the survey of households keeping accounts was 4,000 households from forty cities. With the coming of the Cultural Revolution, this survey, along with the rural household income and expenditure survey, was suspended. Of the second phase surveys, only the 1964 survey was released after the Cultural Revolution.

The worker household income survey was revived in 1978. The survey method was much like that of the second phase. It was a mixture of stratified typical sampling and systematic sampling, but then it gradually changed into and entirely stratified systematic sampling. The selection of households was done in two stages.

TABLE II  
SAMPLE SIZE OF WORKER AND URBAN HOUSEHOLD INCOME AND EXPENDITURE SURVEY

	Number of Households in Preliminary Survey <sup>a</sup>	Number of Households Surveyed (Daily Record <sup>b</sup> )	Effective Number of Samples <sup>c</sup>
1956	6,000		
1957 first half	5,900	5,900	5,350
second half	7,000	4,800 <sup>d</sup>	
1958 second half	(shrunk)		
1960 second quarter	(stopped)		
1961	1,000	—	
1964	139,000	4,000	3,537
1965	—		
1966	(stopped)		
1967-77			
1978	88,282		
1980	100,000	8,000	7,962
1981	200,000	8,000 <sup>e</sup>	8,715
1982	—		9,020
1983	—		9,060
1984	182,344		12,500
1985	258,795		17,143
1986			27,024
1987			32,855
1988			34,945
1989			35,235

Sources: [12] based on the various documents published in China and [8] [9]; and State Statistical Bureau, *Zhongguo tongji nianjian* [Statistical yearbook of China], various editions.

<sup>a</sup> Income survey.

<sup>b</sup> Income and expenditure survey.

<sup>c</sup> Number of samples used for tabulation and shown in various statistical papers.

<sup>d</sup> Some provinces selected their own sample households, totally 20,000.

<sup>e</sup> Provincial statistical bureaus selected 15,000 households and SSB chose about 8,000 among them.

The first was a large-scale preliminary survey to obtain the level of annual incomes; the second stage was selecting households to keep daily accounts. Once chosen, these households were asked to maintain records for the ensuing years until the next preliminary income survey.

The sample size of the 1978 survey was 88,282 households selected from state-owned enterprises, and it covered 67 cities and 93 county towns in sixteen provinces. This might have been the preliminary income survey only. The next survey in 1980 was expanded to cover households from state-owned and collective



TABLE III  
SCALE OF WORKER AND URBAN HOUSEHOLD INCOME AND EXPENDITURE SURVEYS

	Urban Population (1) (Million)	Sampling Ratio Guessed <sup>a</sup> (2) (%)	Kinds of the Survey <sup>b</sup> (3)
1957 first half	99.49	0.23	daily records
second half		0.31	preliminary
1964	129.50	5.08	preliminary
		0.14	daily records
1980	191.40	4.43	preliminary
1981	201.71	0.18	daily records
1984	330.06	2.14	preliminary
1985	382.44	0.17	daily records

Source: [12] based on the various documents published in China.

<sup>a</sup> Estimation of the sampling ratio is as follows: (Effective number of samples cited in Table II) / [Urban population in column (1)].

<sup>b</sup> Kinds of the surveys are as follows: Daily records—Daily record-keeping survey of income and expenditures; Preliminary—Income survey to determine the samples for keeping daily records.

enterprises in 47 cities. The preliminary income survey was done for 100,000 households and 8,000 households were selected for keeping daily accounts. Another feature of the 1980 survey was that every provincial statistical bureau was permitted to expand the survey scale and add more samples to collect additional local information. As a result, the total sample households reached 200,000 (including samples of the SSB) from 126 cities and 26 county towns for the preliminary income survey households, and 15,000 households for keeping accounts (including those of the SSB) from 102 cities and 77 county towns. The sample sizes and the sampling ratios conjectured from 1956 to 1985 are shown in Tables II and III respectively.

In 1984 the preliminary income survey was greatly expanded to cover all the differing households in urban areas. The intention was to get a sampling list of households in urban areas chosen by stratified sampling. This was done through systematic sampling. The households surveyed were required to report their annual income and other aspects of their households such as the number of household members and workers. In this the 1984 preliminary survey corresponded to the income surveys of 1964, 1978, and 1980. The noteworthy advancement over the three earlier surveys was that the results of the 1984 preliminary survey were made public for the first time. It was published only in English under the title *A Survey of Income and Household Conditions in China* [4], and was financially supported by the World Bank. It contained a total of 258,795 sample households, 182,344 of these selected by the SSB from 106 cities and 77 county towns; the remaining samples were selected by the provincial statistical bureaus.

Based on this 1984 preliminary income survey, the recording of expenditure data was extended to cover all urban households after 1985. The sample size was

about 23,000 households selected from 106 cities and 77 county towns as shown in Table VII (p. 346).

#### IV. TECHNICAL PROCEDURES FOR SURVEYS IN THE 1950s

##### A. *Survey Procedures for the 1954 Rural Household Income and Expenditure Survey*

The rural and worker household surveys conducted by the SSB in the 1950s had in common the practice of revising survey procedures by trial and error or through learning by doing. Along with the 1954 Rural Household Income and Expenditure Survey, various articles were published in official statistical journals of the SSB. Also a special anthology of articles mainly contributed by provincial statistical bureaus was published as *1954 nian nongjia shouzhi diaocha jingyan huibian* [Selected articles on the experiences of the 1954 rural household income and expenditure survey] [1], and accompanied the survey report entitled: *1954 nian woguo nongjia shouzhi diaocha baogao* [Report of the 1954 rural household income and expenditure survey of our nation] [2]. These documents made it possible for us to analyze the actual survey procedures in detail.

The 1954 survey was the first nation-wide sampling survey carried out in China. Prior to this the SSB had carried out a pilot survey in Shanxi Province during 1953–54. After this pilot survey the SSB held a national workshop on agricultural statistics between November 29 and December 15, 1954. After this workshop the proclamation for the preliminary survey on national farming household income and expenditure was decreed in February 1955. Sanctioned by this proclamation, the SSB carried out its survey from May to August 1955 asking for data about the 1954 harvest year (approximately from February 1954 to January 1955). The word “preliminary” was used because respondents were asked about their income and expenditures without requiring data based on daily kept accounts during the period concerned. Thus respondents provided data largely based on memory and in round figures. Field surveyors were urged to confirm respondents’ answers from various other data known to the surveyors and/or local authorities. For this survey many local cadres were mobilized as surveyors. The scale of the survey and cadres mobilized in Shanxi and Qinghai provinces are shown in Table IV.

The fieldwork for the survey was conducted by the provincial statistical bureaus. The compilation of the survey results was also entrusted to the local authorities. Consequently the SSB was unable to compile consistent national totals at one time. The final figures made public in official statistical journals sometimes contained those obtained from the samples of different provincial coverages. From our study we confirmed that the report in SSB [2] contained eight different kinds of coverage of the provinces tabulated. These are shown in Table V.

It is reported in SSB [1] that provincial statistical authorities had called for a systematic sampling based on lists of households in descending order of income level without mentioning the way of selecting the starting household. In most cases local authorities selected the first household from the top of the list which caused an upward bias in the estimations obtained.

TABLE IV  
 SURVEY PROCEDURES AND NUMBER OF PEOPLE AND CADRES MOBILIZED FOR THE 1954  
 RURAL HOUSEHOLD INCOME AND EXPENDITURE SURVEY

Stage of Survey Operation	Shanxi Province		Qinghai Province	
	Date or Day Spent Conducting Survey	Manpower Mobilized	Date or Day Spent Conducting Survey	Manpower Mobilized
Pilot survey	Feb. 2-Mar. 6, 1955		Mar. 10-Apr. 12, 1955	1,500 man-days
Chang'an		200 households by 103 surveyors		
Xianyang		100 households by 90 surveyors		
Field operations	Mar. 12-May 5, 1955	186 cadres	Apr. 16-May 27, 1955	3,500 man-days
Guanzhong	Mar. 12-Apr. 20, 1955		Visits to townships in 6 days	
Shanxi=Shanbei	Mar. 5-May 5, 1955		Visits to households in 30 days	
			Revision 5 days	
Error check of original questionnaire	May-Aug., 1955	89 cadres	June 1-Sep. 16, 1955	2,300 man-days
Rewriting	One month		June 1-18, 1955	
First checking	n.a.		June 20-25, 1955	
Second checking	n.a.		June 27-September 16, 1955	

Source: [1].

TABLE V  
DIFFERING PROVINCIAL COVERAGE USED IN TABULATION  
OF THE 1954 RURAL HOUSEHOLD INCOME  
AND EXPENDITURE SURVEY

	Number of Provinces	Number of Households
1.	25	16,468
2.	23	15,432
3.	22	15,292
4.	21	14,334
5.	18	12,175
6.	14	11,106
7.	13	10,718
8.	6	4,308

Sources: [1] [2].

B. *Survey Procedures for Worker Household Surveys in the 1950s*

Documents on worker household surveys in the 1950s are unfortunately not abundant. But several cross-classified tables newly published in Lin's *Zhongguo xiaofei jiegou xue* [Chinese consumption economics] [10] provide a glimpse of the survey procedures behind the figures when combined with other fragmentary information recently made public. The basic way of sampling was to select an enterprise, then rearrange the names on its payroll list according to worker wage level. A systematic sampling technique was then applied to the new list until enough employees were selected. In the 1956 survey the SSB initially tried to list up the state-owned enterprises by the order of average wage rate and the number of workers and then to select enterprises systematically. The result was a wide dispersion of enterprises all over China. The SSB reorganized the list to concentrate enterprises in twenty-seven selected cities. Geographical location was now taken into consideration when selecting the cities and the selection of an enterprise became dependent on its geographical location in the nation. Next the wage rate of workers in the enterprise was chosen as a selection criteria. From the list of workers arranged by the descending order of wage rate workers were systematically selected and asked to reply about their household income and expenditures. Table VI shows the kinds of industries selected from 1956 to 1958. However, the size of the samples were too small to classify enterprises by both industries and cities. The tabulated results were classified by industry only. Some tables in Lin [10] are classified by the occupational status of employees such as manual workers, technicians, and office workers. In order to judge the adequateness of the sampling procedure, the most conventional way is to compare the weight of each cell of tabulated income and/or expenditure classes with the composition in the population. But we have no such data. As is shown in Figure 1 the levels of per household income differ greatly between industries. However differences between occupation and status are also great. Thus the difference among industries may be the result of differences in the wage rates between indus-

TABLE VI  
COVERAGE OF INDUSTRIES SURVEYED

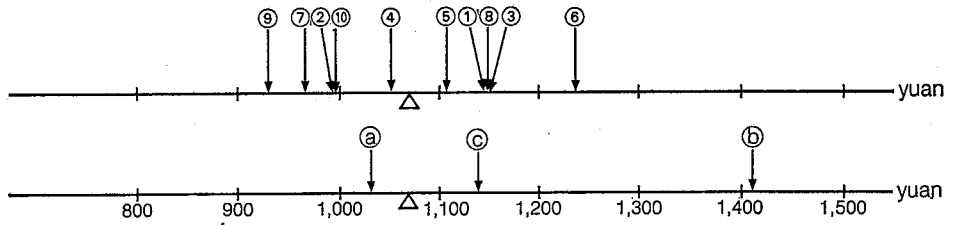
Industries Covered	Year of the Survey		
	1956	1957	1958
Manufacturing	*	*	*
Commerce			*
Education (Elementary and middle school)			*
Administrative organizations of the governments			*

Source: Constructed from tables in [10].

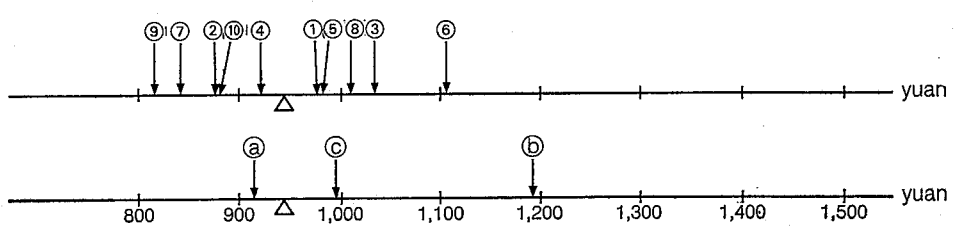
\* Industries used for the selection of sample households.

Fig. 1. Income and Expenditures per Household in 1956

A. Income per household



B. Expenditure per household



Source: For industry codes 1-10, tables in [10].

Notes: 1. a: Manual workers, b: Technicians, c: Clerical workers.

2. 1: Electricity, 2: Coal mining, 3: Chemicals and petroleum refining, 4: Steel and iron, 5: Machinery, 6: Rubber, 7: Pulp and paper, 8: Textile, 9: Food processing, 10: Printing.

3.  $\Delta$ : Average of entire samples.

tries and in the composition of occupational status. We have no further evidence about the adequateness.

Unfortunately all the tables newly published in Lin [10] do not contain the weights permitting calculation of the average per household and/or per person income and expenditures. The average income per household given in the table showing occupational status is 1,069 yuans in 1956. Our estimated average using the weights of occupational status for employees in the whole nation is 1,061.13 yuans. The latter is only 0.74 per cent lower than the former, showing that the composition of occupational status among sample factories was not much different from the composition of industries nation-wide that it affected to the results. One reservation is that we have no information about the family structure and the number of wage earners which might have had an effect on the average.

## V. CONSOLIDATED SURVEY PROCEDURES IN THE 1980s

### A. *Rural Household Survey in the 1980s*

Although the worker household survey has been reorganized into the urban household survey since 1984, the rural household survey is still restricted in scope to farm households. But even in rural townships there are many inhabitants occupied in various kinds of business other than farming. However these households do not make up a large percentage of the rural population. For example, one rural county in Shangtung Province has a total population of 547,800, of which the rural population makes up 93.86 per cent, or 514,200 people. Of these 514,200, 33,000 are nonfarmers. This 6.4 per cent of non-farming households are excluded from the survey.

However recent economic changes have brought a dramatic increase in the number of nonfarmers. In 1987 the number of rural households nationally and the size of the rural population came to 201.6 million and 857 million respectively. At the same time those pursuing family businesses grew from 961,000 households in 1981 to 10.3 million in 1987. This implies that 12 per cent of rural households are nonfarmers. Moreover, nonagricultural cooperative workers are also increasing. This means that the rural household surveys in the 1980s failed to reflect the increasingly large segment of rural nonagricultural households.

A more serious problem is that we do not know about the expenditures for production under the cooperative management system, especially that of the people's commune system. These expenditures were excluded from the surveys. The rural household survey was reconstituted in 1978 while the people's commune system still existed. This system covered the expenditures for production which were surveyed by the regular reporting system on the "basic accounting units" of the people's communes. After 1982 some of the people's communes functions were entrusted to individual farmers, and from 1983 management by individual farmers became dominant in the township. The 6.9 million production brigades and 48.0 million production teams, the basic accounting units in 1978, were changed into 4.85 million new cooperatives and 31.77 million farmer households.

As shown in Table I, the sample size of the rural household survey almost

doubled in 1985. The reason for this expansion in the scale of the survey might be attributable to the need for estimating the variance in the values of the survey items within the townships due to the reform of the management system. At the same time the survey items were expanded to cover all expenditures including production costs and investments of rural households. The released figures, however, were only the national average per person until 1986. Neither size distribution nor regional figures were published (except for those of consumption) until 1986. Some items in the 1986 survey were divided into three general economic regions: east, middle, and west. The net revenue per capita of the respective regions in 1987 was 122.5, 93.6, and 77.2 per cent of the national average; production expenditure per capita was 113.9, 96.7, and 84.9 per cent of the national average. At the same time consumption expenditure per capita was 120.5, 94.1, and 79.5 per cent respectively. This implies that the variance of expenditures may be smaller than that of income, something which is shown in the surveys of many other countries. The variance of expenditures for production seems smaller than that of consumption. This might have been the reason why the SSB did not make public the cross-sectional tables on the expenditures for production. The surveys in the 1980s also did not restore the survey items concerning basic production conditions such as the acreage of cultivated land and fixed production assets including cattle which were surveyed in 1954.

#### B. *Survey Procedures for the Urban Household Survey in the 1980s*

Reorganizing the worker household survey into the urban household survey in 1984 entailed a further expansion of the scale of the survey. The 1984 survey mentioned in Section III surveyed ownership of consumer durables and dwellings in addition to income. Although this acted as a preliminary income survey to serve as a frame for the further sampling of record-keeping households after 1985, it is not clear whether the ownership of a dwelling and/or consumer durables was used for stratification. The urban areas covered in the survey were divided into two groups; one was cities and the other was county towns. The selection of urban areas was done using a two-stage stratified systematic sampling. In the first stage all cities were listed according to the size of their nonagricultural population and grouped into three categories. Those having more than one million nonagricultural people were categorized as large; those with 500,000 to a million were medium-sized; and those having under 500,000 nonagricultural people were small. These three groups were later reclassified into six groups according to geography and climate, and levels of prices, production, and consumption. The listings of cities within each group were based on the average income class and in descending order of annual wage rate. Using this list of each group, selection of the cities was done systematically. Selected were those cities where worker population came across to each point of one million in the cumulation of workers on the list of cities.

The second-stage sampling was of the households that kept records of expenditures for at least one year. The State Statistical Bureau set up unified rule for the size of the samples of daily records. At least 200 households from large city,

100 households from each medium-size city and 50 households from each small town were selected to keep daily records. If a household was unable to continue keeping its records, a substitute household was selected from those in a similar income and expenditure class and having a similar number of household members and workers.

## VI. PROBLEMS HIDDEN IN TABULATION TECHNIQUES

Today most of the tabulations of a statistical survey in advanced countries are carried out centrally by survey organizations which collect the original questionnaires from the surveying units. When the local tabulation system is adopted, the local governments which are responsible for the surveying and computational procedures must convert the questionnaires into machine readable data to be sent to the central statistical office of the national government. This required conversion has been brought about because of the development of electronic data processing. However, the Chinese system has had to adopt the local tabulation system in order to get the tabulated results quickly under the severely restricted computing facilities that existed until the late 1980s without sending the original questionnaire data to the SSB. Also, the Chinese have not used the source-table compilation technique.<sup>4</sup> As a result the SSB receives only the fixed format tables tabulated by the provincial statistical bureaus, and it has not been able to recompile a new cross tabulation of its own. The situation has been the same for the provincial statistical bureaus which have had to be satisfied with receiving the tabulated tables from the local authorities who carry out the surveys in the field.

The tabulations required of the local authorities have been to sum up the original daily household data after normalizing the data in per capita terms. This requirement has remained true except for the earliest period of surveying in the 1950s. Although the tabulation procedures of this early period were never clearly mentioned in official documents, the 1954 Rural Household Income and Expenditure Survey and the worker household survey during 1956–58 seem to have been tabulated as averages per household.

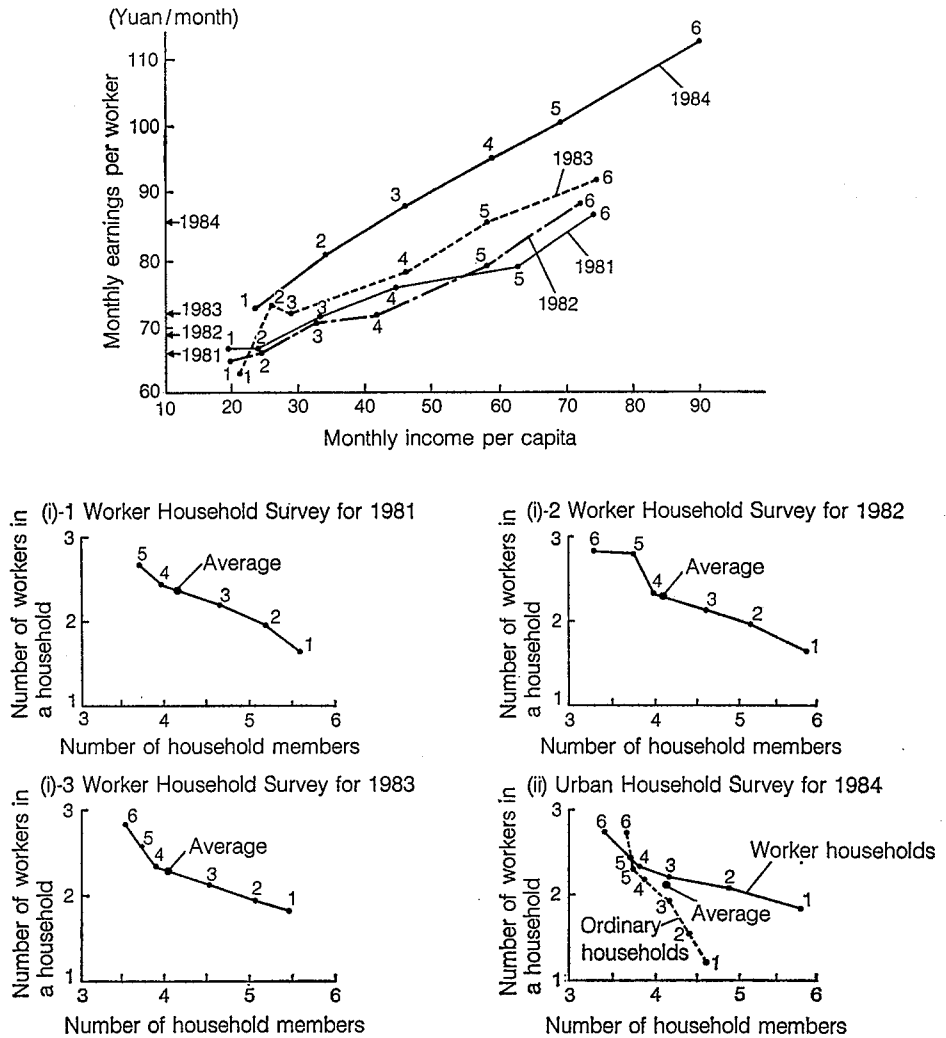
The normalization of every survey item per capita adopted after the 1950s entails the classifying of households by per capita income and then aggregating each questionnaires to get the average of the survey items per capita. The per capita income of a household is closely related to both the number of income earners and the number of household members. This situation is enhanced by the fact that women in most cases receive equal wages with men in socialist countries. Thus the level of per capita income increases with the number of workers in the households. Figure 2 clearly shows this relation.

The total effect of tabulation by per capita income may act to cancel out the income difference between income groups. Thus the detailed tables by income

<sup>4</sup> The source-table compilation technique is used in survey statistics to produce various two-way classified tables. To do this the compiler of the survey results produces a multi-way classified table, source table, which contains all the axes used in the two-way classified tables [11].



Fig. 2. The Relation between the Number of Household Members and Wage-earners



Note: The number for the dots is the code number for income classes based on monthly income per capita: 1 for the lowest, 6 for the highest.

group which have been released since the 1980s are difficult to analyze. Our supposition is based on the fact that each vector of the data by income group in the matrix tabulated by per capita income does not seem to be orthogonal. The values of the determinants of the matrix were nearly zero and the weights for households by per capita income group estimated from the tabulated data were different from

TABLE VII  
NUMBER OF SAMPLES USED FOR TABULATION

	1981	1982	1983	1984	1985	1986	1987
(1) Number of samples used for average value	8,715	9,020	9,060	12,500	17,143	27,024 (17,046) <sup>a</sup> (9,978) <sup>b</sup>	32,855 (25,265) <sup>a</sup> (7,590) <sup>b</sup>
(2) Number of samples used for quintile data		2,350		4,420	17,145		32,855 (25,265) <sup>a</sup> (7,590) <sup>b</sup>
(3) Number of samples gathered for the PSB <sup>c</sup>	11,267	14,070	n.a.	n.a.	23,158		32,855
(4) Number of samples of major cities used for average data	2,120	n.a.	n.a.	6,140	8,533		8,655

Sources: For 1981-85, [8]; and for 1986-87, [9].

<sup>a</sup> Cities.

<sup>b</sup> Local towns.

<sup>c</sup> Provincial statistical bureaus.

the published weights in the official reports and sometimes even provided unreasonable negative values for certain income groups.<sup>5</sup>

The tabulation practices peculiar to the Chinese system to overcome the shortage of computer facilities have gradually changed due to the adoption of electronic computer processing since the third population census in 1982. Since 1983 some of the provincial statistical bureaus began to send in their files containing the original questionnaire data to the SSB. However we cannot fully confirm the results of the changes that have occurred in the computation procedures. Table VII shows that the number of samples used to show the national average or the average by per capita income group is different from the number of samples used for quintile data. The table shows that different computational procedures are used to get quintile data. The implications are that the relisting of each household's income data to get quintile data may use fewer samples and not all of the samples obtained in order to lessen the burden of listing procedures than would be the case if using all of the samples gathered. Moreover tabulation of quintile data is still dependent on the local tabulation of households at each county

<sup>5</sup> If the vector  $\bar{x} = (\bar{x}_1, \dots, \bar{x}_k)$  is the mean of the vectors  $(x_{11}, \dots, x_{1i})'$ ,  $(x_{21}, \dots, x_{2i})'$ ,  $\dots$ ,  $(x_{k1}, \dots, x_{ki})'$ , then the weights  $w$  are obtained from

$$w = \bar{x} \cdot X^{-1}$$

in so far as  $X$  is a non-singular square matrix, and should be non-negative.

without using the original questionnaires. If our supposition is right, the selection criteria is absolutely necessary to judge the validity of quintile data. The only pertinent data which allow us to analyze the relation of income data per household and income data normalized per capita are shown in [5]. At present we have to be satisfied with the results based on income classes normalized per capita in order to estimate the distribution of income and expenditure. (Tentative results for income distribution based on an analysis of these data are shown in [14].)

Another problem is the ambiguity of the samples used in each table. The relation between the two groups of lines (3) and (4) in Table VII is not certain. From scattered information in various documents, there seems to be the possibility that two tables exist; one is the results from only the data sent to the SSB, and the other is the results from the many more regional samples collected in excess of the provincial quotas stipulated by the SSB.

## VII. VALIDITY OF THE SURVEY RESULTS

There are two series of data on the level of income and consumption per capita in time series. One is the recompiled time series shown in the *Zhongguo tongji nianjian* [Statistical yearbook of China], and the other is the sampling survey results which we have discussed above. The discrepancies of these two series have not been fully disclosed by the SSB. However, this does not imply that the survey data is unreliable. On the contrary we are inclined to accept the cross-tabulated data as more reliable than the time-series data.

In Table VIII we show the discrepancy in rural household consumption expenditures per capita in money terms between the recompiled time-series data and the national average data gathered from scattered cross-sectional data contained in various sources. The figures in columns (2)–(5) were obtained from the data published after 1980. Those in column (5) seemed to be estimated values and differ much from those in column (4) which were obtained from the cross-tabulated data. This clearly shows differences between the two series. Column (1) is based on our assumption about the weights of the number of household members and income, and this assumption is that net income is equal to total revenue minus agricultural costs and, as the figures are defined by household, deflated by the number of household members. To the extent that this assumption holds, the figures in column (2) are too low. One possible adjustment is a revision to cover the smallness of the sample size. However this procedure is difficult to support insofar as the sampling procedure examined. Another discrepancy is between the figures for 1962 and 1963 shown in columns (4) and (5). Although the figures in column (4) are computed by the weights of the distribution for farm household categories, all average incomes and expenditures of each are much higher than the national average in column (5). This implies that the figures in column (5) were computed using information outside of the survey and results obtained from various sources which have not been made public.

Table IX is the summary of income and expenditure data for urban households. Column (1) shows the national average of household income while column (2) to

TABLE VIII  
RURAL HOUSEHOLD INCOME AND EXPENDITURES  
(PER CAPITA IN A YEAR)

	(Yuan)				
	Net Income		Consumption Expenditures		
	(1)	(2)	(3)	(4)	(5)
1952					62
1953					69
1954	99.8	64.14		59.57	70
1955					76
1956		72.92		66.63	78
1957		72.95	70.86	70.86	79
1958					83
1959					65
1960					68
1961					82
1962		99.09		92.21	88
1963		101.32		93.86	90
1964		102.28		93.59	95
1965		107.20	95.11	95.11	100
1966					106
1967					110
1968					106
1969					108
1970					114
1971					116
1972					116
1973					123
1974					123
1975					124
1976		113.05			125
1977		117.09			125
1978		133.57	116.06	—	132
1979		160.17		134.51	152
1980		191.33	162.21	—	173
1981		223.44		190.81	194
1982		270.11		200.23	212
1983		309.77		248.39	234
1984		355.33	273.80	—	266
1985		397.60	317.42	—	323
1986		423.76	356.95		352
1987		462.55	398.29		388

Sources: For (1), author's own estimates based on the scattered tables in [2]; for (2), survey results in State Statistical Bureau, *Zhongguo tongji nianjian* [Statistical yearbook of China], various editions and [10]; for (3), survey results in *Zhongguo tongji nianjian*, various editions; for (4), [10]; and for (5), *Zhongguo tongji nianjian*, various editions.

Note: "—" means the same value as the figures in column (3).

TABLE IX  
URBAN HOUSEHOLD INCOME AND EXPENDITURES

	Household Average (1)	Income per Capita			Expenditures per Capita			
		(2)	(3)	(4)	(5)	(6)	(7)	(8)
1956	1,069.17	242.88			214.45			197
1957		253.56	—	—	221.94	222.00	—	205
1958					218.10			195
1959								206
1960								214
1961								225
1962								226
1963		232.25			213.90			222
1964		243.48	--	—	220.72	220.68	—	234
1965								227
1966								244
1967								251
1968								250
1969								255
1970								261
1971								268
1972								294
1973								306
1974								314
1975								324
1976								340
1977								360
1978								383
1979								406
1980								468
1981		500.40	—	—	456.84	—	—	487
1982		535.32	—	—	471.00	—	—	500
1983		572.88	—	—	505.92	—	—	531
1984		660.12	—	—	559.44	—	—	599
1985			821.40	748.92		732.24	673.20	747
1986			988.08	909.96		865.20	798.96	851
1987			1,012.20	—		884.40	884.40	979
1988						1,103.98	1,103.98	
1989				1,387.81			1,210.95	

Sources: For columns (1), (2), and (5), [10]; for column (3), [8] [9]; and for columns (4), (7), and (8), State Statistical Bureau, *Zhongguo tongji nianjian* [Statistical yearbook of China], various editions.

Note: State Statistical Bureau [5] has 1984 data where the average income per head is 520.32 yuan for ordinary household and 748.08 yuan for one-person household.

column (8) are per capita figures for other time-series data. With the exception of column (8) these are the survey-data estimates published after 1980 and differ much to each other.

As discussed in the previous sections, the survey system and sampling design

TABLE X  
ANNUAL INCOME PER WORKER HOUSEHOLD AND ANNUAL WAGE PAYMENT PER WORKER

	1957	1964	1981	1982	1983	1984
(A) Annual income per income earner in household survey (yuan) <sup>a</sup>	836	826	887	927	977	1,130
(B) Annual wage payment per worker at state-owned enterprises (yuan) <sup>b</sup>	637	652	812	836	865	1,031
(A)/(B) (%)	130	126	109	118	112	109
Sampling ratio (‰) <sup>c</sup>	2.29	1.11	1.90	2.09	1.87	2.48

Sources: [12] using the data in State Statistical Bureau, *Zhongguo tongji nianjian* [Statistical yearbook of China], various editions.

<sup>a</sup> (Monthly average income) × (12 months) × (Number of household members) / (Number of wage earners in a household).

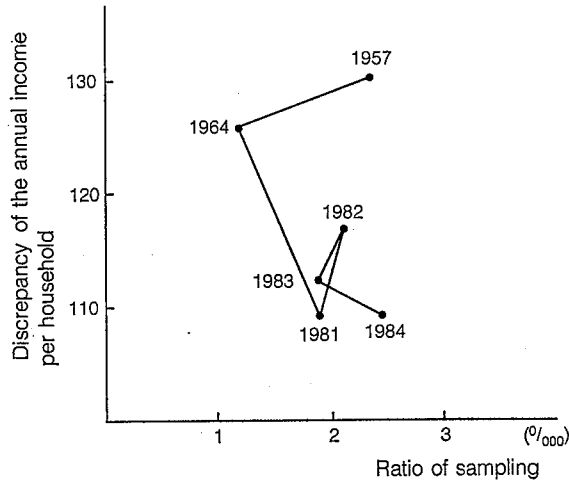
<sup>b</sup> (Wage bill of state-owned enterprises) / (Number of employees in state-owned enterprises).

<sup>c</sup> (Number of wage earners in a household) × (Number of households surveyed) / (Number of employees in state-owned enterprises).

often changed until the 1980s. To check the comparability of survey data for the different years, we estimated the annual income per wage earner in a household and annual wage payment per worker at state-owned enterprises which was based on the regular reporting system, as shown in Table X. This implies that the latter survey covered all enterprises concerned and was not affected by the sampling design. However this does not mean there was no sampling error. The discrepancies are not big during the 1980s but they are in 1957 and 1964. Figure 3 illustrates the relation between the rates of discrepancy and the sampling ratios, expressed in the percentage of the samples to the population. The high rate of discrepancy of 1964 might be attributed to the low sampling ratio. The high discrepancy figures for 1957, although its sampling ratio was high, seem to have been caused by the difference in the coverage of industries. As shown in Table VI, the coverage of industries in 1957 was restricted to the manufacturing sector only for the household survey, but the wage survey covered all industries. This may have widened the discrepancy rate between the household survey data and the wage survey data due to the differences of wage rates among industries.

The survey data are quite informative because of the existence of cross-section data. But this does not guarantee the intertemporal comparability of the survey data when the sampling design suffers drastic changes between survey times. At present we have no basis for denying the credibility of the series in column (8). At the same time however we also have no reason to reject the series in columns (2) through (7). The most difficult problem is that the level of accuracy of the household survey data might have changed with the increase of coverage and sampling ratios even during the 1980s. As noted for Table VII, the number of

Fig. 3. The Rate of Discrepancy between the Household Survey and Wage Survey and the Sampling Ratio



samples used for tabulation changed even among the tables in the same report without mentioning the system of tabulation. This entails the differences among the figures from columns (2) to (7) and each table could rightly reflect one aspect of the Chinese situation. More explicit presentation of the bias in the survey results is required, even in the case of the cross-section data, and further analysis is required to access the consistency between tables of the same survey. There are at least two types of regional figures among the tables after 1980. The first is the tables based on the reports to the SSB in fixed tabulation formats from the provincial statistical bureaus; the second is the tables which the SSB newly tabulated from the original machine-readable questionnaire data provided by the provincial statistical bureaus. It is not certain that the first type of data contain the samples added by the provincial statistical bureaus. Although the SSB recently published detailed reports [8] [9] containing regional figures, it is not clearly mentioned whether these regional figures belong to the first or second categories discussed above.

After analyzing the consistency between the tables of the same survey report, the comparability of the data in time series can be analyzed with the various other economic data explaining the economic situations as a whole. These problems should be entrusted to economists studying the Chinese economy and need to be taken up in a future article.<sup>8</sup>

<sup>8</sup> The work [7] being done by the State Statistical Bureau is one example of efforts by Chinese economists, but the bureau does not discuss the validity of its survey results.

## REFERENCES

1. China, State Statistical Bureau. *1954 nian nongjia shouzhi diaocha jingyan huibian* [Selected articles on the experiences of the 1954 rural household income and expenditure survey] (Beijing: Zhongguo-tongji-chubanshe, 1956).
2. ————. *1954 nian woguo nongjia shouzhi diaocha baogao* [Report of the 1954 rural household income and expenditure survey] (Beijing: Zhongguo-tongji-chubanshe, 1957).
3. ————. *Woguo nongmin shenghuo de juda bianhua* [Dramatic changes of rural people in China] (Beijing: Zhongguo-tongji-chubanshe, 1984).
4. ————. *A Survey of Income and Household Conditions in China* (Beijing: New World Press, 1985).
5. ————. *Gasheng zizhiqu zhixiashi nongmin shouru xiaofei diaocha yanjiu huibian* [Selected materials on the rural household income and expenditure survey] (Beijing: Zhongguo-tongji-chubanshe, 1985).
6. ————. *Zhongguo shehui zhuyi tongji gongzuo de jianli yu fazhan* [Development of Chinese statistical activities] (Beijing: Zhongguo-tongji-chubanshe, 1985).
7. ————. *Zhongguo nongmin shouru hanjiu* [Research on the revenue of Chinese farmers] (Taiyuan: Shanxi-renmin-chubanshe, 1987).
8. ————. "*Liuwu*" *qijian woguo chengzhen jumin jiating shouzhi diaocha ziliao* [Research materials on the national urban household income and expenditure survey during the Sixth Five-Year Plan period] (Beijing: Zhongguo-tongji-chubanshe, 1988).
9. ————. *1987 quanguo chengzhen jumin jiating shouzhi diaocha ziliao* [Research materials on the national urban household income and expenditure survey for 1987] (Beijing: Zhongguo-tongji-chubanshe, 1988).
10. LIN BAIPENG. *Zhongguo xiaofei jiegouxue* [Chinese consumption economics] (Beijing: Jingji-kexue-chubanshe, 1987).
11. MATSUDA, Y. "Similarity and Difference to Statistical Survey System between Japan and China," *Proceedings of the 2nd Japan-China Symposium on Statistics* (Fukuoka: University of Kyushu, 1986).
12. ————. *Chūgoku keizai-tōkei hōhōron* [A treatise on the methods of Chinese economic statistics], Kenkyū sōsho No. 361 (Tokyo: Institute of Developing Economies, 1987).
13. ————. "Survey System and Sampling Design of Chinese Household Surveys: 1952–1985," *Proceedings of the 3rd Japan-China Symposium on Statistics* (Tokyo: Soka University, 1989).
14. MIZOGUCHI, T., and MATSUDA, Y. "A Comparative Study on Income Distribution in People's Republic of China and India," Paper presented at the International Symposium on Making Economies More Efficient and More Equitable, Tokyo, November 27–29, 1989.
15. Tongji-gongzuo-bianjibu. "1954 nian nongjia shouzhi diaocha jianyao ziliao" [Essential materials on the 1954 rural household income and expenditure survey], *Tongji gongzuo*, 1957, No. 10.
16. WANG YIFU, ed. *Zhongguo tongji shigao* [History of statistics of China] (Beijing: Zhongguo-tongji-chubanshe, 1986).
17. YU WEI, ed. *Dangdai Zhongguo de tongji shiye* [Statistical activities of China today] (Beijing: Zhongguo-shehuikexue-chubanshe, 1990).