VALIDATION OF LARGE SCALE SAMPLE SURVEY DATA*
CASE OF NSS ESTIMATES OF HOUSEHOLD CONSUMPTION EXPENDITURE

By B. S. MINHAS**

Indian Statistical Institute

SUMMARY. Many simplistic comparisons have recently been made between the NSS estimates of household expenditure and the independent data on private consumption from the national accounting system (NAS). This paper attempts to (a) investigate the scientific rationale behind those comparisons, (b) indicate the weaknesses and strengths of the external validator data set as a touchstone to assess the reliability of the NSS estimates, (c) assess the degree of cross validation, or the lack of it, between the two data sets relating to two years in the 1970s, and (d) draw some lessons for improving the two data sets as well as caution against their incorrect use. Two supplementary notes, one dealing with standard errors of the NSS estimates of consumer expenditure and the other with the possibilities of non-sampling biases in the NSS budget estimates of cereal consumption, are provided in the appendices.

1. INTRODUCTION

1.1 Validation of the NSS estimates of household consumption expenditure, or for that matter any other primary data set generated by a large scale sample survey, involves many complicated issues. For expositional convenience, most of these issues can be discussed under two broad categories. One class of issues are more germane to the common textbook variety assessments of the internal validity of NSS estimates themselves. The other category of issues go much beyond the more easily tractable area of the internal validity of the survey estimates, and relate more to the examination

B 3–1

of whether or not the sample survey estimates are in agreement with "comparable" external data sets.

1.2 As strictly comparable external data sets are indeed rare to find, the comparisons with external data have to contend with inherent differences between concepts, definitions, coverage, time periods, methods of data collection, estimation procedures, etc. In view of all this, the question of external validation of sample survey data has to be approached with fairness and scientific detachment. The results of such comparisons are difficult to foresee : scientific scrutiny may or may not, lead to mutual validation of both data sets under examination. It may, nevertheless, provide important clues for improving the reliability of either one or both data sets.

1.3 Out of the wide variety of primary data sets generated by the NSS over the past thirty five years, the data on household consumption expenditure have been the most extensive and made available to the public quite regularly. These data sets on consumption expenditure have also been most widely used both by the research workers and policy makers. A number of instances (see the concluding section) of abuse of this data have also come to light. These abuses and misuses have arisen partly from certain misunderstandings arising out of lack of adequate familiarity with the NSS estimates, but more importantly on account of the fact that external validations have been attempted on the basis of superficial comparisons.

1.4 Although we do not intend to ignore altogether the question of inherent variability and internal validity of the NSS estimates of household consumption expenditure, the main focus of this paper is on the issues relating to external validation. The external data sets for purposes of this comparison are the CSO's estimates of total private consumption, which are indirectly derived by adjusting production and income flows of consumer goods and services in the framework of the National Accounting System (NAS). Our emphasis shall mainly be on the methodology of external validation, and less on detailed numerical comparisons of the two sets of estimates, which are available elsewhere.[1] The questions relating to scrutiny and editing of raw data are not the concern of this paper. Further this investigation is confined to the analysis of only the central sample data.

---

[1]Minhas B. S., S. M. Kansal, J. Kumar and P. D. Joshi (1986) : "On the reliability of the available estimates of private consumption expenditure in India," read at the *IARNIW Seminar on Tangible Wealth and Consumption*, Bhopal.

2. INHERENT LIMITS TO THE COMPARABILITY OF THE SO-CALLED "COMPARABLE" ESTIMATES

2.1 The basic rationale of comparisons between the estimates of house-hold consumption expenditure obtained from the NSS and the estimates of total private consumption from the NAS would seem to run along the following lines : The estimates of the NSS are subject to both sampling and non-sampling errors ; each of which in turn is composed of a systematic component (bias) and a random component (variance). As the estimation procedures adopted in the NSS are unbiased for large sample sizes, the sampling bias can be neglected and the sampling errors and non-sampling variance would be rather small. One could, therefore, hope that but for the non-sampling bias of the survey results, the NSS estimates of household consumption expenditure for a particular year should be broadly comparable with the accounting esti-mates of total private consumption of NAS for the same year. In other words such comparisons, among other things, are expected to throw some light on the non-sampling bias of the NSS estimates. However, this reasonable expectation of comparability could hold only if the external (NAS) estimates of private consumption were either based fully on "directly observed" data in the national accounting system, or the indirectly derived components of the NAS data—derived from old benchmark estimates and ad-hoc sample studies—did not suffer from sampling and non-sampling errors which were not negligible. This basic rationale of comparisons of NSS estimates with external data has to be kept firmly in mind.

A. Differences in Coverage

2.2 The coverage of the NSS and the NAS estimates is not identical. The NSS covers only the private households and tends to exclude houseless population and population residing in the institutional households such as prisons, orphanges, barracks and hospitals. The NAS estimates include the latter along with the private households. In addition, the NAS also includes the consumption expenditure of non-profit and charitable institutions which are engaged in social welfare and religious activities and provide free or subsidized education, medical, religious and other welfare services to the households. In view of this larger coverage, the NAS estimates would be expected to be higher than the NSS estimates.

2.3 In order to compare them with the estimates of private consumption of the NAS, the itemwise NSS per capita expenditure data can be used in conjunction with the total population of the country. This procedure is

expected to give us detailed NSS based estimates of total private consumption expenditure, reasonably comparable (in terms of coverage) with the NAS estimates, after certain minor adjustments to the latter. One such adjustment relates to government expenditure on several consumer items such as foodgrains, fruits and vegetables and clothing, which are provided by the government to individuals (e.g. defence personnel), either free or at subsidized rates, and are included under government expenditure. These should be added to the NAS estimates of private consumption expenditure before comparing them with the NSS.

2.4 In consequence of the foregoing adjustments for differences in coverage, the NSS total consumption could be slightly overestimated (under-estimated) if the actual per capita consumption of the defence and security personnel in barracks and the houseless and institutional population is lower (higher) than the average per capita consumption of the NSS. On balance, the two estimates might nevertheless be regarded as reasonably comparable (from the point of view of coverage alone), especially for the food group items.

B. Differences in Time Periods

2.5 The NSS was never expected or specifically asked to produce esti-mates of household consumption expenditure on a regular, annual or financial year basis. Its own anticipation of the periodicity of users demands for any particular set of data, or need ·for synchronization with similar periodic activities of outside agencies, has determined the frequency of the reappearance of certain subject matters in the cycle of NSS work. Selection of time periods for different rounds by the NSS (both in regard to choice of dates and length of period for field work) has generally been guided by certain technical as well as logistic considerations for the deployment of field staff all over the country.

2.6 The NAS estimates of total private consumption are presented every year and relate to the financial year, April to March. As far as the national accounts are concerned, the agricultural production year, July to June, is accounted for in the twelve month period ending with 31 March—3 months prior to end of June when the agricultural year ends. NSS, on the other hand uses variable time periods, varying in length as well as timing for different rounds. The detailed NSS estimates of household consumption expenditure are now presented only once in a quinquennium. The latest three such estimates pertain to 1972-73 (27th round from October 1972 to September 1973), 1977-78 (32nd round from July 1977 to June 1978) and 1983 (38th round from January to December 1983). It should be evident

that it is not easy to establish any straightforward time correspondence between the NSS estimates and the external data sets of the NAS for different financial years.

2.7 Whatever is reported to the NSS as current consumption by the households is unlikely to have been concurrently produced by the production units. Although the NSS estimates of consumption expenditure are generally free from seasonality effects when the survey period spans a year, the available external data (annual product flow estimates of CSO) may contain a substantial portion of goods and services produced during intervals of time outside the survey period of the NSS.   As the production of goods does not, by and large, have the characteristics of a steady-state flow in time, the product flow data of any particular year is likely to be impacted by the cyclical and seasonal elements in a manner different from the effects captured in the NSS estimates based on direct observations in the nearest survey period.   This task of adjusting for time period differences between the NSS and NAS estimates of consumption is difficult, particularly, for the products originating in agriculture— the sector most vulnerable to the vagaries of weather.

2.8 The importance of differences in time periods outlined above can be best illustrated by comparing the NSS 38th round estimates of household consumption with the NAS estimates for the financial year 1983-84. The latter estimate was 22% higher than the estimate derived from the 38th round (1983) of the NSS. It can nevertheless be shown that this large magnitude of the discrepancy (22%) is largely illusory. While the NAS estimates of private consumption in 1983-84 are based on the production of the agricultural year July 1983—June 1984, the consumption of agricultural products reported to the NSS in the 38th round January to December 1983, mostly pertain to agricultural production realised during the period July 1982 to June 1983. First crop of 1983-84 (Kharif) would have been harvested from October 1983 onwards.   Therefore, during the first ten months of the 38th round survey (i.e. from January to October 1983), the reported consumption of foodgrains by the households would entirely be from the Kharif and Rabi crops of 1982-83.   Only in the last two months of the survey period (November and December), the reported foodgrain consumption could have been from the Kharif crop of 1983-84.   For wheat, gram and barley, household consumption during the whole survey year (1983) would be from the Rabi crops of 1982-83.   In fact in the first sub-round of the survey (January to March 1983), reported consumption of rabi grains would have been from the Rabi crops of 1981-82. For all practical purposes, therefore, one can assume that the NSS estimates

of expenditure on foodgrains in 1983 were out of the harvests realised in
1982-83 (July 1982 to June 1983).

2.9   Since there was a big jump in foodgrain production in 1983-84 over
1982-83 (of about 22 million tonnes), and all this would have been included
in the NAS consumption estimates of 1983-84, the latter is naturally higher
than the NSS (1983) estimates by about Rs. 5,000 crores (accounting for 5
percentage points out of the apparent difference of 22% in total consumption)
due to foodgrains alone.   The same considerations would also apply in com-
paring the consumption estimates of oilseeds and their products from the
two data sources, as the production of oilseeds recorded a big increase of
27% from 10 million tonnes in 1982-83 to 12.7 million tonnes in 1983-84.
As most of the consumption of agricultural commodities and related products
reported by the households to the NSS in 1983 was from 1982-83 crops, a
rough adjustment of the time period effects for the agricultural sector alone
would make the apparent discrepancy between the NAS and NSS estimates
of total private consumption come down to about 12% in comparison with the
unadjusted difference of 22% in 1983-84.   The remaining difference of 12%
should narrow down further when the effects of the big harvests of 1983-84
on the other sectors of the economy are also suitably adjusted.

2.10   Similar adjustments for differences in time periods have also been
made in case of a large number of agricultural products to bring the NAS
estimates of private consumption for 1972-73 and 1977-78 (financial years)
in line for comparability with the NSS estimates of the 27th (October 1972—
September 1973) and 32nd (July 1977—June 1978) round of the survey.
Detailed results are presented elsewhere.[a]   It is sufficient to note here that
these adjustments for time period differences between the NAS and the NSS
do not always affect the NAS estimates of private consumption equally or
in the same direction.   The extent of adjustment is indicated by the time
discordance between the survey period of a particular NSS round and the
corresponding agricultural year ; and the direction of adjustment (positive
or negative) depends on whether the previous agriculutral year/season was a
year/season of significantly different (larger or smaller) production compared
to the agricultural year corresponding to a specific NSS round.

C.   Differences in Methods of Data Collection and Estimation Procedures

2.11   In addition to differences in coverage and time periods, the other
major cause of lack of comparability between the NSS and NAS estimates of

---

[a] Minhas, et al. (1986).

total private consumption may reside in the shortcomings in their respective methods of data collection and estimation procedures. An understanding of these methods is a necessary pre-requisite for negotiating comparability between the two data sets.

C.1   NSS Procedures

2.12   The NSS collects detailed itemwise consumption data in value and quantity (wherever possible) terms for the last 30 days[a] preceding the date of inquiry from the sample households by interviewing the head of the house-holds (and other members). The survey period of a round, normally of one year duration, is generally subdivided into four sub-rounds. The sample villages and urban blocks are distributed over the four sub-rounds in equal numbers for either of the sub-samples in which the total sample is divided. The sample households are canvassed in a staggered fashion all through the survey period and the estimates based on all the four sub-rounds taken together are free from seasonal variations.

2.13   The sampling scheme of the NSS consumer expenditure surveys is generally based on a stratified two stage sampling design. The rural and urban sectors of the country are separately divided into hundreds of strata. Villages and urban blocks are the first stage units (f.s.u's) in rural and urban sectors, respectively, and households comprise the second stage units (s.s.u's). The f.s.u's are usually selected from within the strata in the form of two in-dependent sub-samples. The f.s.u's, when selected with varying probabilities, are chosen either with replacement (PPSWR), or by systematic selection with probability proportional to size (PPS systematic). The s.s.u's (Households) are selected, either linear or circular, systematically. It must also be mentioned that various methods of sub-stratification and/or arrangements of households are also used by the NSS for selection of the s.s.u's from each selected f.s.u.

2.14   In the latest three full rounds of the NSS dealing with consumer expenditure namely : the 27th (1972-73), 32nd (1977-78) and 38th (1983), the sample size has been very large, comprising of about 13,000 to 14,000 f.s.u's and 121,000 to 158,000 s.s.u's. The sampling errors of the NSS esti-mates of private consumption, based as they are on very large samples, are indeed very small at the all-India level.

2.15   Even at the state level, the relative standard errors of total expen-diture in the bigger status vary between 2 and 3 percent ; whereas for expen-

---

[a]In recent rounds, consumption data have also been collected for the last 365 days for durable and semi-durable goods.

diture on food the relative standard errors are less than 2 percent. Further, although the precision of NSS estimates[4] of expenditure on individual items of most common consumption is satisfactory, nevertheless, for some items, such as quality foods, clothing, footwear, durable goods, medicines and consumer services, the standard errors of the estimates of sub-groupwise and item-wise expenditure at the state level tend to be high. At the all-India level, the sampling errors of these item-wise estimates of expenditure would be far less than at the state level ; nevertheless, for certain individual items they might turn out to be quite large.

2.16   Although an in-depth discussion of sampling errors will have to · be deferred to a later occasion, a somewhat detailed treatment of the non-sampling errors, which have been considered likely to affect the NSS estimates of consumption, belongs here and is given below :

2.16(a)   *Failure to recall precise information on the consumption of indivi-dual items within a broad group.* It may not be possible for the household members to recall correctly the quantity and value, say, of individual fruits and vegetables, or each of the grains in the foodgrains group, or each spice and condiment in the spices group, or each individual oil and fat consumed during the last thirty days. In view of this, the detailed item-wise household consumption expenditure within the group is not likely to be correctly apportio-ned by the reporting households. In other words, there is likelihood of some misclassification within the group, although the total expenditure on the item-group might be more correctly reported. NSS appear to be conscious of this difficulty. Nevertheless, very detailed listing of individual items comprising a group is adhered to in the schedules of inquiry to facilitate recall and minimize the chances of total omission of minor items by the respondents.

2.16(b)   *Wilful supression or under-reporting of certain consumption expenditures.* The consumption of liquor and tobacco is frowned upon, or is considered unacceptable, by certain sections of society. Also, quite often the head of the household or the informant may not be aware even if some members of the household are casual or regular consumers of liquor and tobacco. There are also certain sensitive items, such as gold and jewellary, whose purchases may not be fully reported by the households to the investigators. The NSS consumption estimates of these sensitive or socially unacceptable items are therefore likely to be gross underestimates.

---

[4]On all this see, V. R. R. Sarma and G. D. Rao (1980), "Standard errors of estimates from the NSS 28th round consumer expenditure survey", *Sarvekshana*, Vol. III, No. 4, pp. 87-106 and also the Supplementary Note in Appendix B of this paper.

2.16(c)   *Under-estimation of rents on dwellings.*   The NSS collects data only on the actual rent paid by households.   All imputed rentals from owner occupied dwellings are excluded from the NSS rental estimates, whereas the NAS covers both rented as well as owner occupied dwellings for the estimation of rent.

2.16(d)   *NSS design and estimates of certain items of expenditure associated with affluent sections of society.*   Some scholars have argued that the current sampling design of the NSS might be leading to systematic under-estimation of certain aggregates, like consumption of consumer durables, which are specifically associated with the rich.   Murthy (1977), for instance, observed that "since the NSS estimates are based on general purpose design with emphasis on point parameters, the estimates of tails, in which the users are specifically interested, are subject to larger sampling errors and hence not amenable to deeper analysis."[5]   Nevertheless these estimates are approximately unbiased.   It is also known that the larger sampling errors in the upper tail does not vitiate the usual estimates of incidence of poverty, nevertheless the precision of the estimates of aggregate expenditure on modern consumer durables,   important consumer services and certain other luxury goods might be affected.   In order to ensure adequate representation of the consumption patterns of the more affluent sections of society, the sample selection procedures, both for f.s.u's (particularly in large cities) and s.s.u's, have been more carefully designed for the 43rd round (1987-88), for which field work is currently in progress.   It is hoped that this would ensure better estimation of certain aggregates of expenditure which are specifically associated with the households in the upper tail of the distribution.

2.16(e)   *Possibility of duplication of expenditure on certain items, particularly foodgrains.*   The NSS estimates of aggregate consumer expenditure on foodgrains, especially cereals, obtained by multiplying the total population of India with the NSS per capita consumption estimates, have been found to be consistently higher than the corresponding estimates of the NAS.   As the NAS estimates are derived from independent estimates of production of cereals and other foodgrains released by the Ministry of Food and Agriculture (MFA), a number of possibilities of duplication and non-sampling biases in the NSS budget estimates of foodgrain consumption have been surmised by different scholars over the past 30 years.   Since foodgrains constitute a

[5]Murthy, M. N. (1977) : Use of empirical studies in evaluating sample design for estimating frequency distribution, *Bulletin of the International Statistical Institute*, XL VII, 3.

very substantial proportion of the average household budget in India, it is
very essential that in this paper, where the primary concern is validation of
consumer survey data, we seek a satisfactory resolution of the doubts and
suspicions raised about the possibilities of duplication in the NSS budget
estimates. The subject, being extremely important, has been dealt with in
Appendix C[4]. We find that most of these doubts and suspicions are based
on a priori reasoning, or on facts whose relevance to the issues under examina-
tion is extremely far fetched. Notwithstanding the lack of firm supportive
evidence, the suspicions might nevertheless linger on. The situation demands
now methodological studies to settle this question. The National Sample
Survey Organization has shown some renewed vigour in recent years in under-
taking methodological work : One methodological study, bearing on the
question of long vs. short consumer expenditure schedule, has just been
completed ; and another study on the suitability of different reference periods,
etc., is under progress in collaboration with the Indian Statistical Institute.
One hopes that the NSS would also undertake methodological work on the
question of duplication, in household budgets, of expenditure on cereals. In
the meanwhile, considerable light on this question can also be thrown by
computing standard errors (s.e.s) of the estimates of expenditure on cereals,
based on sub-samples X states (the old one degree of freedom estimates), and
comparing them with those derived directly from the household data. As
both these estimates of s.e.s are known to be unbiased (although quite different
in regard to their efficiency), the direction of change in the ratios of s.e.
estimates from the two procedures over the states could tell us a lot about the
presence/absence of non-sampling bias in the NSS estimates. No such com-
parisons, however, seem to have been reported in the literature thus far, but
we intend to take up this exercise on a later occassion.

C.2   CSO's Procedures for NAS Estimates of Private Consumption

2.17   The NAS estimates of private consumption are based on production
data of all consumer goods and services obtained from various agencies outside
the control of the CSO. These data are processed and adjusted by the CSO
by deducting exports, intermediate uses and net increases in stocks and adding
in imports to arrive at the availability of various goods and services for domes-
tic consumption and capital formation. This volume and value of all goods
and services available for domestic consumption is apportioned by the CSO
among government, business and private consumption. It is this last use-

---

[4] See Appendix C : A note on the possibilities of duplication and non-sampling biases in the
NSS budget estimates of cereal consumption.

category, i.e. private consumption, which is supposed to be comparable with the NSS estimates of household consumer expenditure.

2.18 The quantities of various consumer items (especially the food products are sub-divided into home-produced consumption and purchases from markets and these are separately evaluated at producer's prices and retail prices, respectively. Private consumption expenditure estimates are derived by deducting government consumption expendtiture from the total final consumption expenditure.[7]

2.19 For many consumer services, such as recreation, personal services, etc., CSO follows the income approach to estimate the value of output and then makes allocation between intermediate, private and government uses. Under this approach the census data on the number of persons engaged in different services and the average earnings per person (taken from ad-hoc survey reports) are utilized. Since the average earnings per person are not available on current basis, the base year figures are updated with the help of consumer price indices of urban non-manual employees for urban areas and wages of rural skilled workers for rural areas. The above procedure provides the value added figures which are inflated to obtain output figures by using output to value added ratios taken from the old survey reports.

2.20 Although the CSO makes sincere efforts to use all available information on production, intermediate uses, etc., still the NAS estimates suffer from a number of shortcomings which basically are due to non-availability of reliable and currently observed information to make various adjustments. Some of these shortcomings, which affect the NAS estimates of private consumption, are discussed below.

2.20(a) Data on changes in stocks are not only partial but also of poor quality. Very little information is available about the stocks held by traders, producers and households (especially the producer households). With the availability of 1978-79 ASI data on changes in stocks, and similar data available in ASI summary reports from 1979-80 onward, the reliability of the NAS estimates for some sectors might improve a little. Nevertheless, the prospects are not encouraging for improvements in data relating to food stock with traders and farmers. However, the currently available NAS estimates, against which validation of NSS estimates is being attempted suffer from this weakness.

---

[7] See, CSO (1980) : *National Accounts Statistics—Sources and Methods*, Department of Statistics, Government of India.

2.20(b) Data on marketable surplus (i.e. allocation between home-grown consumption and market purchases) of various consumer products are based on the old reports of Directorate of Marketing and Inspection (DMI) issued in the fifties and early sixties. Adjustment made on the basis of stale information (collected three decades ago) do not inspire much confidence in the end product. However, efforts are now being made through the Ministry of Agriculture to collect current data on marketable surplus ratios and the CSO proposes to use them as and when such data become available.

2.20(c) Data on intermediate uses are very weak and partial. For most foodgrains the base year ratios (like animal feed, wastages, etc.), which are 20 or more years old, are used in the absence of current information. There are problems also in adjusting the production of major foodgrains for amounts used as seed.

2.20(d) In a few cases the data and the estimation procedure of the NAS are revised in the light of current data. All these revisions are not always carried backward ; however, for major revisions the CSO adjusts the series backward for earlier years. Examples of revisions carried backward only for a limited number of years are the consumption estimates of other sugars, gram, salt and medical services, which have undergone substantial revisions at different points of time. Although the estimates for individual items are affected, the effects are insignificant at the overall level. In other words, the NAS estimates of private consumption for different years may not be strictly comparable among themselves. Their comparability with the NSS estimates for different (specific) years is likely to be affected in differential ways for different items depending on whether the consequences of the revisions in NAS procedures are equally reflected in years selected for comparison.

2.20(e) Data based on production of consumer goods in the unorganised. sector is terribly weak. This is likely to affect seriously the NAS consumption estimates of several consumer products like transistor radios, matches, non-alcoholic beverages, biscuits, confectionary etc. However, efforts are now being made to incorporate such data from the Directory and Non-directory Establishment Survey results for 1979-80. Nevertheless, the current NAS estimates are without the benefit of these data.

2.20(f) The data on trade and transport margins, wherever used, appear to be rather inadequate. CSO has initiated studies in this regard and has

approached a number of dealers for the purpose. The pay off from these
efforts would take time in arriving. In the meanwhile, the attempted adjust-
ments for trade and transport margins in the NAS estimates would continue
to inspire only limited confidence in their reliability.

2.20(g) For some items such as match boxes and alcoholic beverages,
the expenditure appears to be underestimated as the excise duty collections
alone in respect of such items are only marginally different from the
corresponding NAS estimate of private expenditure (inclusive of excise) on
these items.

2.20(h) For certain categories of expenditure, such as education, medi-
cines and medical services, the CSO use a combination of survey (NSS)
and administrative data. This practice has given rise to some duplica-
tion and overstatement of expenditure in NAS data, particularly for the
education sector.

2.20(i) CSO's estimates of expenditure on communications and railway
services are based on firm accounting data on gross receipts. Nevertheless
the difficulty arises in determining the share of the households in these receipts.
Currently these shares seem to be based on subjective judgements. It is,
therefore, necessary for the NAS to undertake/sponsor studies to work out
the share of expenditure attributable to households in these sectors. In the
meanwhile, sectoral aggregates of expenditure estimated by the NSS have to
be judged against external evidence which is no better than informed/
ill-informed guess work.

2.20(j) The allocation of total production of consumer durables between
households and industry (for the NAS estimates of private consumption of
durables) is another murky area. The CSO assumes certain fixed proportions
(identical for many years) of production of durables in the economy as con-
sumption by the households. This assumption is based on subjective judge-
ments. This is indeed an unhappy situation, with implications also for the
estimates of capital formation. Comparison with NSS data for this sector
therefore have to be taken with caution.

D. Incomparability Caused by Unrecorded Data in Official Records
versus Directly Reported Consumption

2.21 The NAS consumption estimates for various consumer goods are
based on official data on production, exports and imports. Circumstances
are known to exist which lead to deliberate under-recording of production of

certain items. For example the manufactures might under-report production of many items to evade excise duty. Such un-recorded production would also be missing from the corresponding private consumption estimates of the CSO. On the other hand, consumers are likely to report the purchases of all such items irrespective of whether or not their production was entered into the records of the enterprises.

2.22 Some exports and imports also take place without being recorded in the official data. Besides organized smuggling, petty trade across the long open borders is not an uncommon phenomenon. Examples of such unrecorded imports relate to items, such as watches, electronic consumer goods, gold and many kinds of textiles. Similarly there are many items, such as sugar, gur, liquor, tabacco and certain varieties of textiles, which are carried across the borders to the neighbouring countries without entry into official records.

2.23 All such transactions, missing from official records cannot be taken into account by the CSO in their estimates of private consumption. On the other hand, the NSS data can reflect all such transactions, as it does not depend on any official data records for its estimates of consumer expenditure. It is, therefore, natural to expect that the comparability of CSO and NSS consumption estimates for certain items would be severely affected.

E.   Sampling and Non-Sampling Errors : NSS vs. the National Accounts

2.24 Many discussions of sampling errors seem to imply as if only the NSS estimates suffer from these errors. This is a gross misconception. The national accounts data get their copious share of sampling errors, not from one but many sample surveys from which the production data base of the national accounts gets built up. Some of these surveys are conducted regularly every year, and others are of ad-hoc nature. The data on per hectare yields (and therefore total output) of principal crops and a large part of industrial production, for instance, are based on regular sample surveys. Where the data on production are not being collected regularly, the value of output is often estimated by the CSO from the results of old surveys conducted at different points of time. All these diverse production surveys, which form the building blocks of the National Accounts, carry their sampling errors into the edifice of the NAS in the most confounded manner. This confounding of sampling errors in the NAS is so intractable that many scholars even fail to recognize them. No estimates of sampling errors of the NAS estimates are known to exist and no demand for such estimates of errors seems to have

been made by the users of the NAS data. On the other hand, the sampling errors of the consumption estimates from NSS, even when these errors are known to be small, are expected to be routinely estimated without due appreciation of the costs involved in these computations.

2.25 In the discussion of differences in the methods of data collection and estimation procedures, we have provided fairly full details of the incidence of non-sampling errors in both data sets. Nevertheless, it is worth noting once again that the process of indirect derivation of numbers (not directly observed), by making all manner of adjustments on the basis of auxiliary data (which are stale, partial in coverage and often of very poor quality) and informed guess-work, leads to fast breeding of non-sampling errors which get cumulated (or get cancelled ?) in unknown ways in the NAS estimates of private consumption. On the other hand, the non-sampling biases (if any) in the NSS data are likely to be small and uniform or subject to very slow change over time ; and similarly over space and socio-economic groups (on all this see, Appendix C). To put it differently, the control of non-sampling errors should be a matter for much deeper concern in the NAS data than in the case of consumption estimates obtained from any single consumer expenditure survey. In the context of the latter, it is much easier to improve concepts, tighten up methods of data collection and field supervision to control suspected non-sampling errors.

2.26 It is quite obvious that the data base of NAS estimates of private consumption would need to be improved in a number of directions. Many of these improvements are planned for the future. At this stage it would however be fair to conclude that the currently available NAS estimates of private consumption (with which NSS estimates are being compared) fall a long way short of the touchstone standard expected of an external validator data set. Matter relating to the degree of precision and biases in the NAS estimates of many items and item-groups are yet much too much dependent on guess work and subjective judgements to permit assessment of the limits within which scientifically valid comparisons can be undertaken with the corresponding survey (NSS) estimates.

F. Differences due to Different Price Sets

2.27 An examination of consumer prices, derived from the value and quantity estimates of different cereals and pulses, revealed that NSS (implicit) prices were consistently higher for almost all the foodgrains than the CSO prices in 1972-73 and 1977-78. The relevant data are presented in Appendix

A-2. When the NSS quantities of different cereals and pulses are evaluated at CSO (NAS) prices, the apparent difference between the two sets of expenditure estimates for the foodgrains group gets reduced by about one-third in 1072-73 and by two thirds in 1977-78.

2.27(a) In terms of its conceptual relevance and comprehensiveness of fully representative coverage in space and time, the implicit prices set of the NSS is the ideal a welfare economist could wish for. They should be much better than the NAS prices for measuring cost of living and consumer welfare. Nevertheless, it is not obvious why NSS prices of foodgrains might be, more or less, consistently higher than the NAS prices. A firm answer to this question can be provided only on the basis of an empirical study, which is outside the scope of this paper.

G. Differences due to Unmatched Classification Schemes

2.28 There are many differences in the classification schemes of the two data sets. This should have little significance for the comparison of total consumption expenditure estimates, nevertheless, the itemwise/groupwise comparisons often get vitiated. In the national system of accounts, for instance, the expenditure in hotels and restaurants is classified under non-food in consumer services. In the NSS, on the other hand, the expenditure incurred by households in hotels and restaurants is included in the food group.

3. APPROXIMATE ADJUSTMENTS FOR NEGOTIATING COMPARABILITY

3.1 To adjust for all the seven major causes of apparent differences between the two sets of estimates is an almost impossible task. Nevertheless, some approximate adjustments can be made and their effects can be quantified. Some of the differences (for want of relevant information, or because of conceptual difficulties) on the other hand cannot be quantified, although their significance is capable of being assessed in qualitative terms.

3.2 The first crucial step in this process is to work out the itemwise details of consumer expenditure independently reported (without adjustments) by the two agencies for some recent years. One such exercise[3] was undertaken to disaggregate and reclassify total private consumption expenditure into about fifty comparable commodity/service groups. For the purposes of this paper, however, these details of expenditure have been aggregated into nine broad groups. The NSS as well as NAS estimates (unadjusted) for these groups are given in Appendix A.1. Prior to giving some answers to the question of

---

[3]Minhas, et al. (1986).

external validation, it is this raw material which has to be processed and adjusted to negotiate comparability.

3.3   As can be seen from Table 1, the unadjusted CSO (NAS) estimates of total private consumption were higher about 6 percent in 1972-73 and about 12 percent in 1977-78 than the corresponding NSS estimates. However, the differences were much larger in total non-food expenditure—NAS estimates were higher by 37 and 28 percent respectively in 1972-73 and 1977-78. As regards the total expenditure on food, the differences were much smaller ; and also they did not go in the same direction in the two years under study. The NAS estimates was about 7 percent lower in 1972-73, whereas in 1977-78 was about 2 percent higher than the NSS estimate.

TABLE 1. AGGREGATE PRIVATE CONSUMPTION EXPENDITURE EFFECT
OF ADJUSTMENT FOR DIFFERENCES IN CLASSIFICATION

(Rs. crores)

	1972-73			1977-78		
	CSO	NSS	CSO/NSS%	CSO	NSS	CSO/NSS%
A. food	21770 +444 (22214)	23420	93 (95)	37400 +757 (38157)	36500	102 (105)
B. non-food	13390 −444 (12946)	9790	137 (132) (132)	25680 −757 (24923)	20030	128 (124)
C. total consumption	35160	33210	106	63080	56530	112

3.4   At this level of aggregation, the only adjustment that can be easily made is the one relating to differences in classification. As indicated earlier, the expenditure in hotels and restaurants is booked under consumer services in the NAS and is classified in the non-food group. Upon inclusion of this item in the food group in the NAS data, the difference between the two estimates narrows down from 7 to 5 percent in 1972-73, but increases from 2 to 5 percent in 1977-78. In other words, in consequence of this adjustment, the NAS estimate of total expenditure on food is lower by five percent (−5%) in 1972-73 and higher by five percent (+5%) than the NSS estimate for 1977-78. For the non-food group, the apparent excess of the NAS over the NSS estimate is reduced from 37 to 32 percent in 1972-73 and from 28 to 24 percent in 1977-78.

3.5   At this stage, one might be tempted to draw some facile conclusions :

First, as the NSS estimate of expenditure on food in India is (5%) higher in 1972-73 and (5%) lower in 1977-78 than the NAS estimates (partially adjusted

for differences in classification only), one might conclude that there is no evidence of systematic non-sampling bias in the household consumer expenditure survey estimates (NSS) for the food group.

Secondly, Although the NSS has not computed standard errors of their estimates of food expenditure for the 27th (1972-73) or the 32nd (1977-78) round, nevertheless, on the basis of results relating to the 28th round (1973-74) data one can safely assume that relative standard errors of the NSS estimates of food expenditure are no larger than one percent at the all-India level[*]. However, the apparent differences of the order of 5 percent between the NAS and NSS estimates, although not insignificant, might not be considered as a cause for serious alarm.

Thirdly, as the differences in the estimates of expenditure on the non-food group between the NAS and NSS are indeed large and as the NAS estimates are higher in both the years, the NSS estimates might be regarded as biased in the downward direction.

3.6   These conclusions are not warranted at this stage. The story of validation has just begun and can brook no haste in jumping to premature conclusions. We need to decompose the total expenditure on food and non-food into smaller sub-groups of items to enable us to sort out at least those genuine differences (on account of factors enumerated in section II of this paper) which are amenable to approximate adjustments.

3.7   *Cereals and pulses.* Foodgrains (cereals+pulses) account for a major part of the total expenditure on food in India. They constitute about 47.5 and 47.0 percent of total food expenditure in CSO estimates and about 57 and 53 percent in the NSS estimates, respectively, for 1972-73 and 1977-78. The actual expenditure also shows (Table 2) that the NAS estimates were substantially lower than the NSS estimates in both the years—lower by about 23 percent in 1972-73 and 9 percent in 1977-78. These apparent differences can be investigated in quantitative terms with regard to the inherent discrepancies in implicit prices and reference periods of the two sets of data.

3.8   *Adjustment for consumer prices.* By evaluating the NSS quantities of different cereals and pulses at the NAS prices and then comparing them with the reported NAS estimates of expenditure, we find that the apparent differences were reduced from 23 to 16 percent in 1972-73 and from 9 to about 3 percent in 1977-78. Details are provided in Table 2.

---

[*]See the Supplementary Note (Appendix B) at the end of this paper.

TABLE 2. CONSUMPTION EXPENDITURE ON CEREALS AND PULSES

(Rs. crores)

items	1972–73					1977–78				
	NSS		CSO	CSO/NSS%		NSS		CSO	CSO/NSS%	
	I	II		I	II	I	II		I	II
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
cereals	11911	11175	9331	78.3	83.5	16919	15964	15271	90.2	95.6
gram	259	243	226	87.2	93.0	418	533	876	209.6	164.6
pulses	1249	894	806	64.5	90.2	1964	1675	1414	72.0	84.4
foodgrains	13419	12312	10363	77.2	84.2	19301	18172	17561	91.0	96.6

Notes : I. gives the consumption expenditure as reported by the NSS.

II. gives the consumption expenditure by evaluating the NSS quantities of different items at CSO prices.

3.9 As noted earlier, the CSO prices for all cereals and (especially) pulses where higher than the NSS prices in both the years. The only exception is the case of gram in 1977-78, where the NAS implicit price was 27 percent higher than the corresponding NSS price. The reason for this appears to be that the CSO booked more than 90 percent of total gram consumption under gram *dal* (dehusked split gram) and evaluated it at gram dal prices which are substantially higher than the whole gram price. In the NSS, on the other hand the directly observed proportion of gram consumed as gram dal comes to only about 50 percent. Another reason for the big jump in gram consumption in 1977-78 in the NAS estimates is a purely procedural quirk. In 1972-73, 35 percent of reported gram output was deducted as animal feed in the NAS, whereas in 1977-78 only about 8 percent was assumed as animal feed. The big jump in consumption expenditure on gram in 1977-78, therefore, is a consequence of the change in assumptions in regard to the proportion of gram deducted as animal feed from the official estimates of gram production. This revision in assumptions, although made on the basis of a (then recent) survey report on the estimates of concentrates fed to cattle in household dairy establishments, was inappropriately extended to the entire cattle population of India. The case of gram is an illustration of how assumption-driven estimates of NAS private consumption can go awry.

3.10 It is worth repeating that deduction for feed and wastages for all cereals and pulses are still based on the old marketing reports of the fifties and these deductions are made as fixed proportions of the reported production figures. These proportions need to be revised on the basis of fresh observations ; the deductions, as made at present, do not inspire confidence about their reliability.

3.11 We might also note that the aggregate seed requirements for various crops do not seem to fall in line with the increases in cereal productivity during the 1970s. The area under high yielding varieties of rice and wheat increased but the yields per hectare for these two crops increased far more rapidly. Nevertheless the deductions for seed requirements (in the NAS framework) as a proportion of total output of rice and wheat hardly show any fall.

3.12 *Time period adjustments.* The reference period differences can be roughly adjusted, particularly, in the case of important crops. This exercise[10] was undertaken in details by us. I do not wish to repeat it here, except for recalling our results for the estimates of consumer expenditure on foodgrains. Adjusting for difference in the reference periods between the NAS and NSS estimates, the apparent discrepancy in 1972-73 gets reduced by 4 percentage points, whereas in 1977-78 the discrepancy gets further elevated by about 3 percentage points.

3.13 The estimates of expenditure on foodgrains alone, comprising about half of the total expenditure on the food group, works out to be substantially higher (23% in 1972-73 and 9% in 1977-78) in the NSS as compared with the unadjusted estimates of the NAS. However, adjusting for differences in prices and time periods the picture comes out to be as follows :

DISCREPANCY IN FOODGRAIN CONSUMPTION
(VALUE) ESTIMATES

	(CSO less than NSS by %)	
	1972-73	1977-78
observed (unadjusted)	22.8%	9.0
due to implicit prices	7.0	5.6
due to time period	3.8	−3.4
due to other factors	12.0	6.8

In conclusion, adjustments for differences in prices and time periods do close the gaps between the two sets of estimates for both the years. Nevertheless substantial differences still remain—12% in 1973-74 and 6.8% in 1977-78.

3.14 We have been unable to fathom any sound basis to suspect the presence of systematic bias in the NSS estimates of expenditure on foodgrains Also the magnitudes of the relative standard errors of the estimates of expenditure on foodgrains are far too small to bridge the gap of 12 percent in 1972-73 and 7 percent in 1977-78 between the NSS and CSO estimates of

---

[10]Minhas, *et al.* (1986).

foodgrains consumption (say, at ± 2.0 s.e. level) On the other hand, it is strongly suspected that adjustments for seeds, feeds, wastages (and other assumptions about certain ratios in the NAS framework) may have led to systematic underestimation of private (human) consumption of foodgrains in the CSO data set in 1972-73 and 1977-78. The official estimates of area reported for different foodgrain crops might also deserve to be looked into with some care.

3.15 *Other food (other than foodgrains).* The unadjusted NAS estimates of expenditure on other food (total food minus foodgrains) were much higher than the NSS—by 14 percent in 1972-73 and 15 peercent in 1977-78. On reclassifying and including the NAS expenditure in hotels and restaurants in the other food groups, this discrepency increased further to 18 percent and 20 percent respectively in 1972-73 and 1977-78. The causes of this large one-sided difference in each of the two years needs to be investigated. This investigation is all the more necessary as the respective differences between the two data sets at the sub-group level (Appendix A-3) do not go in the same direction for all the sub-groups, or in both years for the same sub-group.

3.16 The NAS estimates in both the years were higher for milk and milk products, edible oils and vanaspati, fruits and vegetables and sugar and gur ; whereas the estimates for salt and spices and miscellaneous foods were higher in the NSS data. For meat, fish and eggs group, the NAS and NSS estimates were only marginally different.

3.17 Since milk and milk products, edible oils, sugar and gur are also used as inputs in commercial establishments, and these intermediate uses are not separately netted out in the NAS procedure for, dealing with hotels restaurants and sweets shops, the NAS and NSS estimates for these commodities are not therefore comparable. Also appropriate deductions for unrecorded exports of sugar and gur just cannot be made. In case of estimates of consumption of edible oils, the NAS includes expenditure on oil seeds under edible oils, whereas the NSS does not book any expenditure on oil seeds under this head. The reliability of NAS estimates for milk products and edible oils is also affected by the use of old ad hoc ratios in ear-marking a certain proportion of milk output for conversion into curds, cheese, butter and ghee and the proportion of different oil seeds taken as available for oil extraction. The time period differences also vitiate comparability between the two data sets for sugars as well as edible oils. For instances, the production of oil seeds in the agricultural year 1977-78 was about 8 percent higher than in the previous year. Bulk of the edible oils consumption observed by the NSS

in 1977-78, however, would have come from the oils extracted from the oil seeds produced in the previous year. The product flow estimates of edible oils consumption in 1977-78, based on the current year output, should therefore be higher than the NSS estimates of household expenditure for the survey year 1977-78.

3.18 *Milk and milk products.* The unadjusted NAS estimate of expenditure on this group was higher by 6 percent in 1972-73 and 10 percent in 1977-78. Nevertheless, estimated expenditure on liquid milk in the NAS data was lower than the NSS in both years (−11% in 1972-73 and −4% in 1977-78). Since some liquid milk is also used in hotels and restaurants and it is not separately accounted for in the NAS estimates of private consumption of milk and milk products, there is an element of incomparability in the two data sets. One should nevertheless expect the NAS estimates to be somewhat higher than the NSS. We can nevertheless adjust for difference in implicit prices which were higher in the NSS data for both the years. After carrying out this adjustment, the NAS estimates remain higher by only about 5 percent in both the years. Disregarding the difference on account of the known element of incomparability, the NSS estimates of household expenditure on milk and milk products, which are quite precise, should be taken as validated.

3.19 *Edible oils and vanaspati.* The NAS estimates for this group were considerably higher (+14% in 1972-73 and +37% in 1977-78) than NSS. However for vanaspati which, though a smaller fraction of this sub-group, is the most popular cooking medium in sweets shops, hotels and restaurants, the differences in the two sets of data were even higher. After removing the expenditure on oil seeds from the CSO estimates, the expenditure on edible oils (other than vanaspati) in 1972-73 was almost exactly equal in the two data sets. Upon making similar deduction for expenditure on oil seeds, the CSO estimates of edible oils for 1977-78 still remain higher (+24 percent) than the NSS. We must nevertheless remember that this large difference should come down appreciably when account is taken of the large increase (8%) in oil seeds output in 1977-78 as compared with the previous agricultural year, and due regard is taken of the edible oils contained in sweets, etc. bought by the households from outside. If proper account of edible oils and vanaspati entering purchased food items could be taken in the NSS estimates, the two sets of estimates may not differ in the same direction in the two years.

3.20 *Sugar, gur and other sugars.* The NAS estimates of expenditure on crystal sugar were higher by 12 and 16 percent respectively in the two years. As indicated earlier the NAS estimates are expected to be higher than

the NSS estimates.   Also one need not be unduly disturbed on account of this difference in crystal sugar.   However, one does need to worry about the staggering differences in gur and other sugars which are shown below :

TABLE 3.   CONSUMPTION EXPENDITURE ON SUGAR, GUR AND OTHER SUGARS
(Rs. crores)

	1972–73			1977–78		
	CSO	NSS	CSO/NSS %	CSO	NSS	CSO/NSS %
Sugar	743	661	112	1017	877	116
Gur	1316	529	249	1411	593	238
Other sugars	200	44	454	49	58	84
Total	2259	1234	183	2477	1528	162

3.21 We may be inclined to regard the NSS estimates of household expenditure on gur to be on the lower side. However the CSO estimates of gur (and other sugars) consumption do not seem credible. The production estimates of gur, on the basis of cane utilization data, were placed around 8.7 and 9.0 million tonnes respectively in 1972-73 and 1977-78. Whereas the CSO estimates of consumption (which are supposed to be smaller than production) come to 13.7 and 9.4 million tonnes for the same two years. While there might have been some unrecorded exports, there were no imports of gur in these two years. The story for other sugars is also difficult to trust : The NAS estimates of expenditure on other sugar was Rs. 200 crores in 1972-73, which fell down to Rs. 49 crores five years later in 1977-78 ; whereas the NSS estimate was Rs. 44 crores in 1972-73 and Rs. 58 crores in 1977-78. The NAS estimates of gur and other sugars seem to be too unreliable (both in 1972-73 and 1977-78) to serve as the validator data set for NSS expenditure on this sub-group.

3.22 *Fruits and vegetables.* For the group as a whole, the CSO estimates were higher than the NSS by about 70 percent in each of the two years.   Part of this big difference is due to classification problems ; some fruit products and processed fruits are included in this group by the CSO, whereas they appear under spices and miscellaneous food in the NSS.   Nevertheless, more serious problem in assessing the reliability of the NAS estimates of consumption of fruits and vegetables arises from the fact that we have little solid data available to us on a regular basis relating to the production of a vast multitude of items composing this sub-group.   Informed guesswork and judgemental projections are difficult to accept as validators of data obtained through

other methods. The NSS estimates of household expenditure on many items in this sub-group seriously lack in precision. Nevertheless for the total of fruits (fruits and nuts) the estimates are reasonably precise.

3.22(a)   In the vegetables group, potato is the only item for which regular estimates of production are available annually. The NAS estimates of potato consumption were lower by 33 percent in 1972-73 but were higher by 19 percent in 1977-78 as compared to NSS. A major part of this difference was due to differences in implicit prices in the two sources. In 1972-73 the CSO's implicit prices were 14 percent lower than the NSS prices, but in 1977-78 the CSO prices were 12 percent higher. Also the 1977-78 potato crop was 14 percent larger than the 1976-77 crop. Adjusting for the price and time period differences the NAS and NSS estimates of household expenditure on potato come close to each other, but the remaining discrepancies do not go in the same direction in the two years. Also, the NSS estimate of potato consumption is characterised by low relative standard errors.

3.23   *Salt and spices.* Since 1975-76 the CSO has been using NSS estimates of salt consumption and the estimates for 1977-78 are identical. Nevertheless, CSO estimates of consumption of spices were considerably lower than NSS—by about 53 percent in 1972-73 and 45 percent in 1977-78. The under-estimation by the CSO seems to be resulting from its non-inclusion of many processed spices (which are far more expensive than the original ingredients). Also some of the spices, such as green chillies, are classified in the vegetables group by the CSO ; in NSS they are included in the spices group. In the NSS, dried mango powder (amchur) is classified with spices ; whereas the CSO seems to account for it in mangoes under fruits. There are also considerable differences in prices between the two data sets. The CSO estimates of expenditure on salt and spices seem to be underestimates. On the other hand, the NSS estimates are not only precise but are likely to be bias-free.

3.24   *Miscellaneous food group.* This group includes tea (leaf), coffee (powder), other non-alcoholic beverages, biscuits and confectionery, tea and coffee cups (taken outside home), hotels and restaurants, sweets, refreshments and other miscellaneous foods. There are all manner of classification problems in this sundry-group— including exclusions and inclusions in one or the other data set. Nevertheless, in our earlier study[11] we did try to carry out all the adjustments that seemed reasonable as well as practical. The NSS estimates of consumer expenditure on this miscellaneous group were Rs. 1273 crores and Rs. 2082 crores, respectively, in 1972-73 and 1977-78. The CSO

---

[11] Minhas, *et al.* (1986).

estimates, on a comparable basis, were lower than NSS in both the years (being Rs. 884 crores in 1972-73 and Rs. 1630 crores in 1977-78). It is difficult to decide which set of estimates is better. Some of the excess of the NAS estimates for milk, edible oils and sugars, for instance, might be offset against the relative excess of the NSS estimates over the NAS in this miscellaneous group.

3.25 *Non-food expenditure.* Estimates of aggregate non-food expenditure, broken down into 8 sub-groups, are shown in Appendix A.I. For detailed comparisons between the NAS and NSS estimates for these 8 sub-groups and their components, the reader may refer to our earlier paper[13]. We shall address brief comments to the question of validation here. The first step in this regard is to adjust the two data sets for the obvious differences in classification and coverage.

3.26 As indicated earlier, the expenditure in hotels and restaurants is booked under consumer services in the NAS whereas the NSS estimates include this item under other food. Also the NSS estimates of expenditure on house rent do not include imputed rents of owner-occupied houses, whereas the NAS estimates are inclusive of imputed rents. Further, the NSS estimates of expenditure on education relate only to the household expendtiture on education, whereas the NAS estimates also include expenditures undertaken by the non-profit educational institutions (as the latter are in the nature of transfers to the household sector). Taking account of the above-said differences in classification and coverage, the comparative picture of aggregate non-food expenditure is presented in Table 4.

3.26(a) In other words, in consequence of these adjustments, the gap between the NAS and NSS estimates of aggregate expenditure on non-food items narrows down from 37 to 14 percent in 1972-73 and from 28 to 13 percent in 1977-78.

TABLE 4. AGGREGATE NON-FOOD EXPENDITURE ADJUSTED FOR
COVERAGE AND CLASSIFICATION
(Rs. crores)

	1972–73			1977–78		
	NAS	NSS	Ratio	NAS	NSS	Ratio
total non-food (unadjusted)	13300	9790	137	25080	20030	128
(i) hotel and restaurants	−444			−757		
(ii) imputed rents		+807			+1040	
(iii) education		+745			+1023	
	12040	11342	114	24923	22093	113

[13] Minhas, *et al.* (1086).

B 3–4

3.27 *Pan, tobacco and liquor.* As argued earlier, the NSS is likely to underestimate by a big margin the household expenditure on tobacco and intoxicants : it is difficult to get correct consumption data on these two items through the interview method in our social setting where the consumption of these items is frowned upon. However, the CSO estimates of consumption of these two items, although substantially higher than the NSS estimates, also appear to be downward biased, particularly in case of liquor. The CSO's estimates of expenditure on liquor are only marginally higher than the actual collection of excise duty on this item in both the years.

3.27(a) In other words, the sample survey estimates of household expenditure on tobacco and liquor are better ignored because of the inherent social inhibitions of the respondents in reporting the consumption of these two items. While in theory, it would be possible to derive reliable estimates for tobacco and intoxicants from the production data by making proper adjustments for exports, imports, trade and transport margins, particularly for excise and custom duties, nevertheless the available NAS estimates (following this method) seem to have serious flaws. Under the circumstances it might not be inappropriate to drop these two items from both the data sets and compare only the expenditure on all non-food (ANF) minus the expenditure on tobacco and intoxicants. Dropping tobacco and intoxicants from both data sets, the comparative picture of aggregate expenditure on all non-food (ANF-tobacco and intoxicants) comes out as under :

TABLE 5. AGGREGATE NON-FOOD EXPENDITURE (ANF)
MINUS TOBACCO AND INTOXICANTS

(Rs. crores)

	1972–73			1977–78		
	NAS	NSS	ratio	NAS	NSS	ratio
ANF	12946	11342	114	24923	22093	113
(i) tobacco	1117	612		1533	1000	
(ii) intoxicants	364	195		528	288	
ANF (i+ii)	11465	10535	109	22862	20805	110

3.27(b) As there is no tobacco against pan, supari, etc., the NSS estimates are quite precise and may be bias-free. The NSS estimates were 30 percent higher than NAS in 1972-73 but were lower by 8 percent in 1977-78. The wide variation between the two year is explained by the sudden rise in supari prices towards the end of 1977, which was fully reflected in the NAS (one shot) computations, but only slowly in the last two sub-rounds of the 32nd round survey.

3.28 After negotiating comparability in classifications, coverage differences in the treatment of imputed rents and expenditure of non-profit educational institutions, and excluding from both data sets the expenditure on tobacco and intoxicants, the aggregate estimates of expenditure on all non-food in the NAS data set were higher by 9 and 10 percent, respectively, in 1972-73 and 1977-78. This remaining discrepancy between the two data sets pertains to other sub-groups in the non-food sector and these sub-groups are discussed below :

3.29 *Clothing and footwear.* The NAS estimates for this sub-group are higher than the NSS by 11 percent in 1972-73 and 14 percent in 1977-78. Unlike certain other items of daily consumption clothing and footwear are not bought by households in a regular fashion. The NSS estimates of expenditure on this sub-group are characterized by large sampling errors. Taking into account the standard errors of the point estimates of the NSS for this sub-group, one might conclude that the aggregate household expenditure on clothing and footwear in the two sets of data is not too different from each other. The differences in time periods between the NSS and NAS may not be important here. Nevertheless the NAS estimates of expenditure on this sub-group, taken by themselves, might not be free from under-estimation.

3.30 *Gross rent, fuel and light.* We have dealt with imputed rents already and adjusted the NSS data for incomplete coverage. The NSS estimates of expenditure on fuel and light are appreciably higher than the NAS ; and the NAS estimates, particularly for non-commercial fuels, seem to be gross underestimates. The CSO might be advised to use the NSS estimates of expenditure on fuel and light in its estimation of private consumption and other national income aggregates. The basis for this observation was fully elaborated in our earlier paper.

3.31 *Medical care and health services.* The CSO estimates of private expenditure on this sub-group are based on the combined use of NSS estimates (for medicines) and administrative data on medical services. The two estimates were almost equal in 1972-73. However, the NAS estimate was 24 percent lower than the NSS in 1977-78. It is instructive to note that the NAS estimate for 1977-78 was not at all comparable with its own estimate for 1972-73. The NAS procedure for estimation of expenditure on medical services was very different in 1977-78 as compared with 1972-73. Also the NAS estimate of expenditure on medicines was grossly underestimated in 1977-78. This happened because the NSS estimates for 1977-78 were not available to the CSO when the national income estimates for 1977-78 were

released. The CSO took the NSS per capita expenditure on medicines in 1972-73 and carried it forward to 1977-78 by inflating it with the wholesale price index for drugs. This procedure took the price increases into account but totally ignored the quantitative increase in consumption of medicines during the intervening five year period. The CSO's projected per capita consumption of medicines in 1977-78 was Rs. 1.48 as compared with the NSS observed figures of Rs. 2.26. In consequence, the NAS estimate of consumption of medicines was about 25 percent lower than the NSS estimates of household expenditure in 1977-78. To conclude, the NSS estimates for this sub-group were comparable for both the years, whereas the CSO estimates not only lacked inter-se comparability in time but were also lower than the NSS estimates in both the years. The CSO could adopt both medicines as well as medical services. Estimates of NSS, as its experience with independently worked out estimates of medical services has been unsatisfactory.

3.32 *Recreation, entertainment, cultural services and education.* Assuming that the excess of expenditure on education in the NAS is the correct estimate of institutional expenditure on education (not covered by the NSS as it covers only the direct expenditure by households), and adding this difference in each of the two years to the NSS estimates for this sub-group, the comparative figures, adjusted for differences in coverage, are as follows :

(Rs. crores)

1972-73			1977-78		
NAS	NSS	NAS/NSS	NAS	NSS	NAS/NSS
1402	1384	104	2084	2323	90

Seen this way, the expenditure on entertainment, sports goods, etc., is higher by 4 percent in 1972-73 in the NAS but lower than NSS by about 10% in 1977-78. In fact the CSO estimate of entertainment expenditure for 1977-78 is a gross underestimate : the receipts from entertainment tax levies by states were by themselves more than double the total entertainment expenditure reported in the NAS. Despite the apparently close correspondence between the two sets of estimates in consequence of due adjustments, there is room to believe that both sets of estimates might be on the low side.

3.33 *Miscellaneous goods and services.* This sub-group consists of a large number of goods and services and the figures of expenditure shown against the group (line 9, Appendix A.I) are not easy to sort out for negotiating

comparability between the two sets. However, the expenditure in hotels and restaurants (Rs. 444 crores in 1972-73 and Rs. 757 crores in 1977-78), which is classified with miscellaneous services by the CSO, does not belong here. After this deduction, as these amounts have been taken account of in the other food group, the comparable figures are shown in parentheses (below line 9, Appendix A.I). The aggregate expenditure on this group works out to be less in CSO estimates as compared with the NSS—by about 14 and 6 percent respectively, in 1972-73 and 1977-78.

3.34 *Over-view for non-food.* Adjusting for differences in coverage and classification, and dropping out tobacco and intoxicants from both data sets the NAS estimate of private consumption expenditure on the non-food group was higher than the NSS estimates by 9 percent (Rs. 935 crores) in 1972-73 and by 10 percent (Rs. 2057 crores) in 1977-78. The NSS point estimate of expenditure on clothing and footwear, allowing for sampling errors which are rather large for this sub-group, are not significantly different from the corresponding NAS estimates. The NAS estimates for fuel and light, medical care and health services, entertainment and sports goods and miscellaneous goods and services sub-group, on the other hand, seem to be underestimates and are lower than the point estimates of household expenditure provided by the NSS for the same sub-groups. However, for the remaining two sub-groups, namely, consumer durables and transport and communications, the CSO estimates of expenditure are substantially higher than the NSS in both the years. The main reasons for the overall discrepency in the aggregate estimates of expenditure on all non-food between the NAS and NSS would have to be located in the respective estimates of expenditure for these two sub-groups. We must deal with them in some detail.

3.35 *Furniture, furnishing, household appliances (operations and repairs).* The NAS estimates of expenditure on this sub-group was higher than the NSS by 54 percent (+Rs. 351 crores) in 1972-73 and by 34 percent in 1977-78 (+Rs. 504 crores). One major drawback in estimating the expenditure on consumer durables (furniture, radio, TV, musical instruments, bicycles, scooters, motor cycles, passenger cars, household appliances, ornaments etc.) from product flow method is the absence of reliable information of their distribution by use—categories. Most of these goods are used by households as consumer durables and also be establishment as capital goods. The current NAS practice of taking certain proportion of their total supplies as private consumption is essentially based on subjective judgements. Another difficulty that stares in the face of product flow methods is that a good many of the

consumer durables are produced in the unorganized sector and reliable estimates of their production are not directly available and the derived data suffer from all manner of subjective correctives applied to incomplete and partial (often false and deliberately falsified) facts.

3.35(a) One can accept the proposition that the production based estimates of consumption of consumer durables could be quite realistic provided their total supplies were estimated from current production data with adequate coverage and the total availability thus estimated was distributed between capital goods and consumer durables on some sound basis. However, this does not seem to be the case at present. The NSS estimates of household expenditure on consumer durables, on the other hand, are based on current data. Nevertheless, they are characterized by very large standard errors. In other words, there are defficiencies in both the data sets as far as the estimates of expenditure for this sub-group are concerned. The validator data set is far short of the touchstone quality and the survey estimates are not precise. Both need improvements and one can suspend final judgement on the relative merits of the two till the planned improvements are executed both by the CSO and NSS.

3.36 *Transport and communications.* The estimates of expenditure on the modern consumer services—post, telephones, railway, taxi, bus, other conveyance etc., are widely at variance between the NSS and CSO. The NAS estimates are nearly three times the NSS in 1972-73 and well over three times in 1977-78. In money terms the difference amounts to Rs. 1134 crores in 1972-73 and Rs. 2921 crores in 1977-78. These are staggeringly large differences. Are these true difference ? What is the degree of underestimation in the NSS estimates ? Have the NAS judgements gone astray and produced gross overestimates of private expenditure on consumer services ? These are the questions that need answers, but precise quantitative answers do not seem possible at this stage.

3.37 Reliable accounting data are available on gross receipts of post and communication department as well as the railways. However, the problems arise in distributing these receipts among households, businesses and government. The NAS assume 40 percent of total receipts from postal and communication services as private expenditure, while NSS (observed) household share comes only to 10 percent. Similarly, NAS assumes 80 percent of total passenger receipts as private expenditure, whereas the NSS estimate of household expenditure on rail services is much less. The NAS estimates of expenditure on other modes of conveyance (including owner-driven vehicles) are also nearly twice (or more) as large as the NSS estimates,

3.38 Once again one may be tempted to regard the NSS estimates of household expenditure on communications and transport services as under-estimates, nevertheless, the NAS ratios of use of these services among house-holds, business enterprises and government also need to be revised on the basis of fresh studies. The relevant blocks in the consumer expenditure schedule of the NSS for the 43rd round have been redesigned to capture adequately the household expenditure on these services. The CSO are also taking steps to improve the data base of their judgements regarding the various ratios.

4. VALIDATION OF NSS ESTIMATES : AN OVER-VIEW

4.1 We are now in a position to sum-up our findings on the question of validation of the NSS estimates of household consumption expenditure against external evidence.

4.1(a) Disregarding those differences which might be genuine but cannot be adequately adjusted, there does not seem to be any straightforward evidence of non-sampling bias in the NSS estimate of total expenditure on food consumption in the two years. NSS estimates are higher (+5%) in 1972-73 and lower (−5%) in 1977-78 in comparison with the NAS estimates.

4.1(b) Considering the expenditure on foodgrains alone, the NSS esti-mates are higher than NAS in both years (being+12% and +6.8% respectively in 1972-73 and 1977-78), even after making the adjustments for differences in implicit prices and time periods. Although our search (Appendix C) has failed to turn up any convincing evidence to suspect the presence of systematic bias in the NSS procedures for collecting data on foodgrain (cereal) consump-tion, some suspicions would linger on unless the issue is resolved through appropriate methodological studies. On the other hand, our search has led us to the strong suspicion that the total consumption of foodgrains is under-estimated in the NAS. Many suggestions have been offered for improving the data base of the NAS estimates particularly in regard to various deductions (for seed, feed, wastages and other important ratios) which are so crucial in calculating the unduplicated net output of foodgrains available for human consumption.

4.1(c) The NAS estimates of private expenditure on milk and milk products, edible oils and vanaspati, and sugar and gur are expected to be higher as the CSO is unable to take separate account of intermediate uses of these commodities. The discrepancy between the NAS and NSS estimates

for these commodities was also caused by differences in implicit prices time periods and wrong derivations of production and/or availability for private consumption.

4.1(d) Adjusting for differences in implicit prices and keeping in mind the upward bias in the NAS estimates, the NSS estimates of household expenditure on milk and milk products in both the years stand fully validated.

4.1(e) After making adjustments in the NAS estimates of expenditure on edible oils (for seeds and appropriate time periods), the NSS estimates seem to pass the external validation test. However, the NAS estimates of vanaspati seem to be gross over-estimates as they are not appropriately netted for its use in commercial establishments.

4.1(f) The NAS estimates of private consumption of gur and other sugars are so severely flawed that they are not worthy of being considered as an external validator set. The NAS estimate of other sugars for 1972-73 is about 5 times the NSS estimate, whereas the NAS estimate for 1977-78 is less than one fourth of its own estimate for 1972-73 and also slightly lower than the corresponding estimate of the NSS for 1977-78. The NAS estimates of private consumption of gur in both the years are even higher than gur production derived from cane utilization data.

4.1(g) The NSS estimates of household expenditure on salt and spices and miscellaneous foods are appreciably higher than the corresponding NAS estimates. A part of this difference is attributable to unadjustable differences in classification between the two data sets and partly due to underestimation of expenditure on spices in the NAS. On the whole the NSS estimates for this sub-group are quite precise and appear to be reliable.

4.1(h) NSS estimates of household expenditure on meat, fish and eggs stand completely validated ; ignoring, of course, the possibility that both might be bad.

4.1(i) The NAS estimates of private consumption of vegetables and fruits are about 1.7 times the corresponding estimates of the NSS for 1972-73 as well as 1977-78—larger than the NSS by Rs. 1202 crores in 1972-73 and by Rs. 2289 crores in 1977-78. These are very large differences for this one sub-group and account for about two-thirds of the total difference in aggregate expenditure between the NAS and NSS (on all the seven sub-groups of other food taken together) in 1972-73 as well as in 1977-78. Although there are no

regular official estimates of production of fruits and vegetables (except for bananas, papaya and potato), which could serve as a reliable data base for deriving the NAS private consumption estimate of fruits and vegetables in the national accounting framework, one may nevertheless be inclined to consider that the point estimates of the household expenditure on many individual fruits and vegetables provided by the NSS (except for potato) might not be fully reliable—besides their large sampling errors, one could also suspect them to be underestimates. This would of course, be a subjective judgement, difficult to maintain, lest the NAS estimates of private consumption got anchored to regular production estimates of fruits and vegetables over time.

4.1(j) The unadjusted estimates of aggregate expenditure on other foods, dropping gur and other sugars from both data sets (as the later were not worthy of being used in a validation exercise), were as under :

(Rs. crores)

	1972-73			1977-78		
	CSO	NSS	CSO/NSS %	CSO	NSS	CSO/NSS %
(i) other foods	11852	10002		20594	17198	
(ii) gur & other sugars	1516	573		1460	652	
(i)—(ii)	10336	9429	110	19137	16546	116

Notice that on eliminating the expenditure on gur and other sugars from both data sets, the point estimates of NSS aggregate household expenditure on all six sub-groups of other food (other than foodgrains) differ from the corresponding NAS estimates only by Rs.907 crores (10%) in 1972-73 and Rs.2591 crores (16%) in 1977-78. This, of course, is the situation without making any allowance whatever for the known (and shown) upward bias of the NAS estimates of private consumption of milk and milk products, vanaspati and edible oils and crystal sugar in 1972-73 as well as 1977-78.

4.1(k) Comparing these differences in aggregate expenditure (NAS > NSS by Rs.907 crores (+10%) in 1972-73 and Rs.2591 crores (+16%) in 1977-78) on the six sub-groups of other food with the apparent differences in just one of the six sub-groups, viz. fruits and vegetables (NAS > NSS by Rs.1262 crores in 1972-73 and Rs.2289 crores in 1977-78), one gets a clear perspective on the validity of the NSS estimates vis-a-vis the external data set. Leaving aside gur and other sugars, if one took the point estimates of the NSS as the only correct estimates of expenditure on each of the sub-groups which have passed

B 3—5

the validation test, the overall excess of the NAS estimates of aggregate private consumption of other food in 1972-73 (Rs. 907 crores) will be less than Rs.1262 crores, the amount by which the fruits and vegetables estimate alone exceeds the NSS estimate for this sub-group in 1972-73. Similarly the overall excess of the NAS estimate of aggregate consumption on all other food (Rs.2591 crores) will be only slightly larger than the excess of Rs.2289 crores by which the NAS estimate of private consumption of fruits and vegetables (alone) exceeds the NSS point estimate of household consumption for this sub-group in 1977-78.

4.1(l)   Adjusting for classification and coverage differences in the non-food group and dropping tobacco and intoxicants from both the data sets (Table 5), the NAS estimates of aggregate private consumption (value) of all non-food items was larger by 9% (Rs. 930 crores) in 1972-73 and by 10% (Rs. 2057 crores) in 1977-78. We have shown that these apparent differences between the two aggregates are traceable mainly to just two sub-groups, consumer durables and modern services.

4.1(m)   Taking account of the relative standard errors of the NSS point estimates of household expenditure on clothing and footwear, the NSS estimates for this group may pass the validation test although the fact of NSS point estimates of clothing (not footwear) being lower than NAS in both the years is less than reassuring.

4.1(n)   The NSS point estimates of household expenditure on fuel and light are quite robust and appreciably higher than the assumption-driven NAS estimates of private consumption for this group. The CSO might use the NSS estimates for this sub-group in the construction of national accounts.

4.1(o)   The NSS estimates of household expenditure on medical care and health services are better both in terms of consistency in procedure in the two year as well as in terms of reliability. The NAS estimates, on the other hand, seem to suffer both from inconsistency and underestimation.

4.1(p)   The aggregate household expenditure on miscellaneous goods and services sub-groups, based on NSS point estimates, works out to be higher than the NAS in both the years. The NSS estimates for this sundry group of items, based as they are on current observations are more reliable than the corresponding miscellany of items and indirect procedures of the NAS.

4.1(q)   The NAS estimates of private consumption of consumer services (i.e. post, telephone and conveyance) although based on unverified assumptions regarding the distribution of gross receipts among household business and government, were much higher than the NSS in both the years—by

Rs. 1485 crores (351+134) in 1972-73 and by Rs. 3425 crores (504+2921) in 1977-78. In other words, the discrepancy between the expenditure estimates of the NAS and NSS for consumer durables and services is more than the overall discrepancy (of Rs. 930 crores in 1972-73 and Rs. 2057 crores in 1977-78) for all non-food sub-groups put together, including consumer durables and consumer services. The whole question of the validation of the total non-food sector estimates therefore, revolves around the views/judgements one takes in regard to underestimation/overestimation of expenditure on consumer durables and services—just these two sub-groups only. However, it is better if such judgements are suspended till more and better data are collected and properly analysed.

4.1(r) It is true that the NSS point estimates of household expenditure on consumer durables and consumer services are subject to very large sampling errors. Nevertheless one would have little confidence in stretching the error band to bridge the big, one-sided gaps between the NAS estimates of private consumption and the household expenditure of the NSS for these two sub-groups. As imprecise point estimates with large sampling errors are a poor guide for the discovery of true values, one might, for the time being (until the results of the 43rd round of NSS become available in 1989) adopt the NAS estimates (assumption-driven, hence not objective) of private consumption expenditure on consumer durables and services. This should also oblige us to accept the finding that 9 to 10 percent difference between the NAS and NSS estimates of aggregate expenditure on all non-food (minus tobacco and liquor) items is more than fully accounted for by the differences between the estimates of the two agencies for just two sub-groups, i.e. consumer durables and transport and communication services, in 1972-73 as well as 1977-78.

4.2 The substance and pith of this validation exercise is presented in a capsule form in Table 6. Every step and number appearing in this table has already been explained in Section 3 and 4. The independent data set (NAS), it would seem fair to conclude, is far short of the touchstone quality expected of an external validator data set. A number of its components are based on such weak evidence and unverified assumptions which seriously diminish its value in a cross validation exercise. On the other hand, the NSS estimates of expenditure on minor vices, such as tobacco and intoxicants, and consumer durables and modern consumer services are of doubtful reliability. Nevertheless, despite these difficulties, which need to be overcome in both data sets, an overwhelming proportion of household consumer expenditure data of the NSS and the independent private consumption estimates of the NAS do get cross validation.

TABLE 6. CROSS VALIDATION OF CONSUMER EXPENDITURE : NSS vs. NAS

(Rs. crores)

	1972–73				1977–78			
	NSS	CSO		CSO/NSS%	NSS	CSO		CSO/NSS%
I. Food Grains								
I.1 unadjusted	13419	10363		77	19301	17561		91
I.2 NSS at CSO prices	12312	—		—	18172	—		—
I.3 adjusted for time period	—	10835		—	—	16940		—
foodgrains (adjusted)	12312	10835		88	18172	16940		93
II. Other Food								
II.1 other food adjusted for classification difference	10002	11852		118	17198	20594		120
II.2 gur and other sugars	573	1516		—	652	1460		—
II.1 minus II.2	9429	10336	907	110	16546	19137	2591	116
II.3 fruits and vegetables	1835	3097	1262	169	3228	5517	2289	171
III. All Non-Food (ANF)								
III.1 ANF unadjusted	9790	13390		137	20030	25680		128
(i) hotels and restaurants	—	−444		—	—	−757		—
(ii) imputed rents	+807	—		—	+1040	—		—
(iii) education	+745	—		—	+1023	—		—
III.2 ANF adjusted for coverage	11342	12946		114	22093	14923		113
III.3 tobacco and intoxicants	807	1481		—	1288	2061		—
III.2 minus III.3	10535	11465	930	109	20805	22862	2057	110
III.4 consumer durables and services	1284	2769	1485	216	2650	6075	3425	229

5. SOME GENERAL LESSONS AND SUGGESTIONS FOR RESEARCH

5.1 This validation exercise suggests some obvious lessons for data users in the academic community and the policy making bodies. We also have some suggestions for further research work.

5.1(a) The results presented in this paper should make it clear that it is hazardous to carry out pro-rata adjustment in the observed size distribution of consumer expenditure in a particular NSS round by multiplying it with a scalar derived from the ratio between the NAS estimate of aggregate private consumption (for some financial year) and the total household expenditure available from the NSS round. This kind of mindless tinkering with the NSS distribution, as practiced by the Planning Commission in the Seventh Plan documents, does not seem permissible either in theory or in the light of known facts.

5.1(b) It would also seem equally wrong to attempt studies of comparative trends in respective (NSS vs. NAS) aggregates, or components of these broad aggregates, by using raw data which have not been adjusted for differences in coverage, time periods, classification schemes, implicit prices, etc.[13] Quite aside from this question of genuine adjustments for negotiating comparability, the data thrown up by different cross-sectional cuts taken at varying intervals of time by the NSS and those obtained from continuous time series by the CSO are rooted in quite different objectives and operational strategies. The intention of this remark is not to suggest that the area of common intersection between the two data sets is small. On the contrary the area of this common intersection is indeed very large. Nevertheless often it is not realised that the likelihood of these two domains being coterminous (and the two methods throwing up identical estimates of total private consumption, or its main components) at many points in calender time is indeed small. We do not live in a steady-state world : the cross-sectional results are likely to have more rough edges for the comfort of those who unwittingly get used to reading smooth trends in time series data of national accounts.[14] The latter data set, one must remember, is conditioned by ad-hoc judgements, check totals, bench-marks, ratios and informal adjustments provided by many diverse official agencies outside the control of the CSO.

---

[13] Among some others, Vaidyanathan (1986) has also indulged in faulty comparisons of this kind, particularly in his Table 1, p. 135.

[14] See, for instance, Mukherjee, M. (1986).

5.1(c) The NSS estimates of the distribution of consumer expenditure over the households are controlled only by the scientific considerations explicitly incorporated into the design of the sample survey in a particular round and not by any other extraneous consideration. It is perfectly legitimate to compare the NSS estimates of per capita consumption over different rounds of the survey. The non-sampling bias in the NSS estimates, if any, has been rather small and may have remained more or less constant over time. Our conclusions on this issue, and other related matters concerning the measurement of the proportion of population below the poverty line, etc., are summed up in Appendix C, paragraph C–30, (a)(f).

5.1(d) The few major differences between the two data sets, which still remain, do not seem amenable to explanation in terms of sampling errors. There are inherent difficulties in the collection of certain types of data in the NSS ; whereas there are enormous gaps in production data and also many other difficulties in the derivation of private consumption estimates in the NAS framework. Frequent resort to unverified judgements and assumption driven estimation in national income accounting may seem unavoidable at present, nevertheless, it is unwise to assert, as some experts seem to have done[15], that judgements can do just as well or better than directly observed data.

5.1(e) It must be stressed that the conclusions of our validation exercise are based on two specific rounds of the NSS. There is need to further validate these results by examining some other rounds of the NSS. The 38th round (1983) is the obvious candidate as the detailed data for the earlier rounds in the sixties are not easily retrievable.

---

[1] See, for instance, Mukherjee, M. (1986).

Appendix A.I

UNADJUSTED CONSUMPTION EXPENDITURE IN INDIA AT CURRENT
PRICES 1972-73 AND 1977-78

(Rs. crores)

sl. no.	item	1972-73			1977-78		
		CSO	NSS	CSO/NSS %	CSO	NSS	CSO/NSS %
1.	food*	21770	23420	93	37400	36500	102
	1.1 foodgrains	10362	13418	77	17560	19302	91
	1.2 other food	11408	10002	114	19840	17198	115
B.	non-food	13390	9790	137	25680	20030	128
2.	pan, tobacco and liquor	1613	994	162	2397	1595	150
3.	clothing and footwear	2755	2482	111	6241	5475	114
	3.1 clothing	2563	2319	111	5888	5068	116
	3.2 footwear	192	163	118	353	407	87
4.	gross rent, fuel and power	2487	2271	109	4149	4462	94
	4.1 gross rent and water	1272	465	273	2079	1039	199
	4.2 commercial fuel	500	436	114	877	1052	83
	4.3 non-commercial fuel	715	1370	52	1238	2371	52
5.	furniture, furnishings—household appliances (operations and repairs)	1003	652	154	1984	1480	134
6.	medical care and health services	799	854	94	1458	1915	76
7.	transport and communication	1766	632	279	4091	1170	350
8.	recreation, entertainment education and cultural services	1402	603	232	2084	1300	160
	8.1 education*	1092	347	315	1515	492	308
9.	miscellaneous good*	1565	1302	120	3231	2633	123
	and services	(1121)	(1302)	(86)	(2474)	(2633)	(94)
	total private consumption expenditure (A + B)	35160	33210	106	63080	56530	112

*Not comparable : The expenditure in restaurants and hotels in the case of NSS is included in item 1 (Food) and under miscellaneous goods and services in NAS ; the NAS estimates for education include institutional expenditure on education which is not covered by the NSS.

Appendix A.II

(Rs. crores)

sl. no.	items	1972-73			1977-78		
		CSO	NSS	CSO/NSS %	CSO	NSS	CSO/NSS %
1.	rice	1.38	1.42	97.2	1.74	1.82	95.6
2.	wheat	0.94	1.05	90.0	1.25	1.37	91.2
3.	jowar	0.97	1.09	89.4	1.08	1.14	94.7
4.	bajra	0.99	1.05	94.3	1.27	1.24	102.4
5.	maize	0.82	0.93	88.2	1.13	1.18	95.8
6.	barley	0.99	0.98	101.0	0.98	1.04	94.2
7.	ragi	0.83	0.94	88.3	1.04	1.15	90.4
8.	small millets	0.71	0.86	82.6	0.81	1.08	75.0
9.	gram (whole)	1.42	1.27 }1.52	94.3	2.93	2.03 }2.30	127.4
10.	gram (dal)		1.77			2.57	
11.	arhar (dal)	1.50	2.26	66.4	3.43	4.04	84.9
12.	moong (dal)	2.15	2.68	80.2	2.87	3.39	84.7
13.	urd (dal)	2.08	2.63	79.1	3.01	3.41	88.3
14.	masur (dal)	1.37	2.02	67.8	2.92	3.66	79.8
15.	other pulses	1.21	1.65	73.3	2.41	2.71	88.9

Appendix A.III

UNADJUSTED* CONSUMPTION EXPENDITURE ON OTHER FOOD BY SUB-GROUPS

(Rs. crores)

sl. no.	sub-group	1972-73			1977-78		
		CSO	NSS	CSO/NSS %	CSO	NSS	CSO/NSS
1.	milk and milk Products	2705	2606	106	5227	4749	110
	1.1 milk (liquid)	1564	1760	89	3038	3173	96
2.	edible Oils	1465	1280	114	3077	2243	137
	2.1 vanaspati	307	188	162	561	322	174
3.	sugar, gur, etc.	2259	1234	183	2477	1528	162
4.	vegetables and fruits	3097	1835	169	5517	3228	171
	4.1 potato	231	347	67	778	655	119
	4.2 other tubers/cereals substitutes	191	143	133	225	155	144
5.	meat, fish and eggs	915	891	103	1690	1677	101
6.	salt and spices	467	877	53	979	1600	58
7.	miscellaneous food	884	1273	69	1630	2082	78
	total of other food	11852	10092	118	20597	17103	120

*Except for the miscellaneous food sub-group, where a partial adjustment in the NAS data for expenditure on this item in hotels and restaurants has already been affected.

Appendix B

STANDARD ERRORS OF NSS ESTIMATES OF CONSUMER EXPENDITURE

(A Supplementary Note)

B.1   Prior to computerization of the tabulation of NSS data, the estimates of standard errors (s.o's) used to be worked out from estimates relating to different sub-samples. Because of the inter-penetrating net work of samples provided in the sampling design, it was possible to obtain unbiased estimates of standard errors in this manner. Nevertheless these estimates were not efficient as they were based on very few degrees of freedom (d.f.). Quite often these estimates used to be based on a single pair of half-samples with just 1 d.f. Nevertheless, using stratum X sub-samplewise estimates or State X sub-samplewise estimates, d.f's for such estimates of s.e's could be increased to a reasonable number.

B.2   With the use of electronic computers to process NSS data, it has become possible to estimate standard errors directly from household data. However, the NSS have eschewed regular reporting of standard errors in recent years : The samples are now very large and the standard errors of estimates of consumer expenditure are expected to be very small. Also the amount of computational labour and expense involved in producing the estimates of standard errors (in multi-subject surveys, with the number of sample households being around 150,000 or more) is a real deterrent.

B.3   Although the number of sample f.s.u.s in the 27th (1972-73), 32nd (1977-78) as well as 28th (1973-74) rounds of the NSS was around 13,000, the number of households selected from each f.s.u (village/block) for the 28th round survey was restricted to two on an average. In other words, in comparison with the 32nd round, the total number of sample households in the 28th round was only around one-seventh— about 23,000 as against 158,000 households[1]. Sarma and Rao (1980) used the data of this smallest round to estimate the s.e.'s of estimates of expenditure on food, non-food, a large number of sub-groups and individual items for 17 major states of the Union. This is the only recent consumer survey round for which s.e's of estimates of consumer expenditure are available. Probably the purpose of Sarma and Rao exercise was to evaluate the precision of state level estimates corresponding to certain allocations of f.s.u.'s at the state level. They did not, therefore, bother with the computation of r.s.e's at the all-India level. However, we

---

[1]For comparative details of sample size and methods of sample selection in the 28th and 32nd round, see Table B-3.

are comparing the NSS estimates of household expenditure against the national accounting estimates of private consumption in India. We need the r.s.e's of all-India estimates of value of some aggregates, such as food, non-food and total expenditure.

B.4 Had Sarma and Rao given the estimated aggregates ($\hat{Y}_s$) of food, non-food and total expenditure along with the r.s.e (as they did in their appendix tables for many item-groups and individual items), we could have easily obtained the statewise variances $V(\hat{Y}_s)$ by calculating (r.s.e) × (estimate)^2, then summed over states (S) both the estimates and their variances to yield the all-India estimate of expenditure ($\hat{Y}$) and its variance $V(\hat{Y})$, respectively :

$$\hat{Y} = \sum_s \hat{Y}_s \text{ and } V(\hat{Y}) = \sum_s V(\hat{Y}_s).$$

B.5 In the absence of the estimated aggregates, we have to calculate $\hat{Y}_s$ and $V(\hat{Y}_s)$ from the value of r.s.e's (Table 2 of Sarma and Rao) and of A and B (Table 3 of Sarma and Rao), the first and second stage components of variance, respectively. To obtain approximate value of s.e.'s, we proceed as under :

$$V(Y_s) = \frac{A_s}{n_s} + \frac{B_s}{n_s m} = \frac{A_s}{n_s} + \frac{B_s}{2n_s} ;$$

where n_s is the average value of number of f.s.u's per stratum, which was obtained by dividing total number of f.s.u's in a state (sectorwise) by the number of strata formed in the 28th round ; and m (the number of s.s.u's per f.s.u) was equal to 2 in the 28th round.

Now $$(\text{r.s.e})_s^2 = \frac{V(\hat{Y}_s)}{\hat{Y}_s^2}$$

Therefore, $$\hat{Y}_s = \sqrt{V(Y_s)} \div (\text{r.s.e})_s$$

B.6 For each state the sum of food and non-food expenditure should agree with the independently derived total expenditure. However, in our computations this did not happen in a few states. Nonetheless for most of the states there is good agreement. The values of r.s.e's for all the 17 states pooled together are as follows :

sector	r.s.e.'s (28th round)		
	food	non-food	total expenditure
rural	0.68%	1.42%	0.74%
urban	1.20%	1.08%	1.20%

Naturally these all-India values of r.s.e's are lower than the lowest values obtained for the largest state, Uttar Pradesh (U.P.). All-India sample is manifold in size in comparison with U.P.

B.7   Our interest in this note is not to compute r.s.e's estimates of consumer expenditure directly from the household data of the 27th or 32nd round. Rather we intend to demonstrate how one can use the already available 28th round estimates to get approximate esitmates of r.s.e's for the 32nd round, as the set of strata and the sampling scheme were very similar for these two rounds. Writing the variance of all-India estimates of expenditure $(\hat{Y})$ as :

$$V = \sum_{s=1}^{17} \frac{A_s}{n_s} + \frac{1}{m} \sum_{s=1}^{17} \frac{B_s}{n_s} \cdot \qquad \dots \ (1)$$

where n_s is the average number of sample f.s.u's per stratum in the sth state, $\frac{A_s}{n_s}$ and $\frac{B_s}{m.n_s}$ are the first stage and second stage components of variance of the aggregate estimate of the sth state and m is the average number of sample s.s.u's per f.s.u. We can rewrite (1) as :

$$V = \left(\Sigma \frac{B_s}{n_s} \right) \left[\frac{\Sigma(A_s/n_s)}{\Sigma(B_s/n_s)} + \frac{1}{m} \right].$$

In the case of 28th round, variance of Y_{28} is given by

$$V_{28} = \left(\Sigma \frac{B_s}{n_s} \right) \left[\frac{\Sigma(A_s/n_s)}{\Sigma(B_s/n_s)} + \frac{1}{2} \right]. \qquad \dots \ (2)$$

Expressing this as relative variance,

$$\frac{V_{28}}{\hat{Y}_{28}^2} = \frac{\Sigma(B_s/n_s)}{\hat{Y}_{28}^2} \left[\frac{\Sigma(A_s/n_s)}{\Sigma(B_s/n_s)} + \frac{1}{2} \right] \qquad \dots \ (3)$$

Similarly the relative variance of the aggregate estimate of 32nd round, $\hat{Y}_{32}$, is given by

$$\frac{V_{32}}{\hat{Y}_{32}^2} = \frac{\Sigma(B_s'/n_s)}{\hat{Y}_{32}^2} \left[\frac{\Sigma(A_s'/n_s)}{\Sigma(B_s'/n_s)} + \frac{1}{12} \right] \qquad \dots \ (4)$$

where the symbols with prime are for the 32nd round,

$$\frac{(\text{Rel. var.})_{32}}{(\text{Rel. var.})_{28}} = \frac{\dfrac{\Sigma(B'_s/n_s)}{\hat{Y}'^2_{32}}\left[\dfrac{\Sigma(A'_s/n_s)}{\Sigma(B'_s/n_s)} + \dfrac{1}{12}\right]}{\dfrac{\Sigma(B_s/n_s)}{\hat{Y}^2_{28}}\left[\dfrac{\Sigma(A_s/n_s)}{\Sigma(B_s/n_s)} + \dfrac{1}{2}\right]}$$

Now, as the sampling design, up to the selection of first stage units, in the 28th and 32nd round was exactly the same and the periods covered were close by, it can safely be assumed that approximately

$$\frac{\Sigma(B_s/n_s)}{\hat{Y}^2_{28}} = \frac{\Sigma(B'_s/n_s)}{\hat{Y}'^2_{32}}$$

and

$$\frac{\Sigma(A_s/n_s)}{\Sigma(B_s/n_s)} = \frac{\Sigma(A'_s/n_s)}{\Sigma(B'_s/n_s)}.$$

Therefore $\dfrac{(\text{Rel var.})_{32}}{(\text{Rel. var.})_{28}}$ is approximately equal to K; where

$$K = \frac{\dfrac{\Sigma(A_s/n_s)}{\Sigma(B_s/n_s)} + \dfrac{1}{12}}{\dfrac{\Sigma(A_s/n_s)}{\Sigma(B_s/n_s)} + \dfrac{1}{2}}.$$

Hence $(\text{r.s.e})_{32} = \sqrt{K}\ (\text{r.s.e})_{28}$.

B.8 The projected values of r.s.e's for 32nd round are given in Table B-1.

B.9 In Table B-2 we have reproduced the highest and the lowest values of r.s.e's at the state level for many subgroups and items from Sarma and Rao (1980) from the 28th round data, separately from the rural and urban sectors. The name of the state is given in parentheses along with r.s.e value. With a few easily understandable exceptions, the lowest r.s.e's, as expected, are found in relatively large states (U.P. in most cases) where the sample size is large; in comparison, the smaller states turn up relatively larger r.s.e's. I do not have the time or interest in computing all India r.s.e's for the aggregates of expenditure at the sub-group level. Nevertheless, one can make the following observations to enhance the appreciation (and confidence) of certain non-statistical economists and economic statisticians in regard to the precision of all India level NSS estimates of expenditure on the sub-groups-items listed in Table B-2.

(a) The r.s.e. of estimate of expenditure at the all-India level for any sub-group/item can at most be as large (in general it will be much smaller)

as the highest estimate for the same sub-group/item computed for any state[2].

(b) The r.s.e. at all-India level can be smaller than the smallest estimate obtained at the state level for the same sub-group. Nevertheless for certain commodities/services the r.s.e. at all-India level might come out to be higher than the smallest estimate at the state level.

B.10 NSS might compute r.s.e's for a recent round. Nevertheless, the 28th round results can safely be assumed for the time being to provide the outer limits to the magnitudes of relative standard errors of the estimates of consumer expenditure for all aggregates and almost all sub-groups of items in the 32nd round, or any other recent round where the set of strata and the sampling scheme might be similar to the 28th round.

TABLE B-1. R.S.E'S (PROJECTED) OF THE 32ND ROUND ALL INDIA ESTIMATES OF AGGREGATE VALUES OF CONSUMER EXPENDITURE

item	$(r.s.e)_{28}$	K	$\sqrt{K}$	$(r.s.e)_{32}$
(1)	(2)	(3)	(4)	(5)
		Rural		
food	0.0068 (0.68%)	0.3486	0.5904	0.0040 (0.40%)
non-food	0.0142 (1.42%)	0.4057	0.6369	0.0090 (0.90%)
total	0.0074 (0.74%)	0.3998	0.6323	0.0047 (0.47%)
		Urban		
food	0.0129 (1.29%)	0.3305	0.5749	0.0074 (0.74%)
non-food	0.0198 (1.98%)	0.4293	0.6552	0.0130 (1.30%)
total	0.0129 (1.29%)	0.4588	0.6773	0.0087 (0.87%)

As the sample size in the 32nd round was many times larger than that of 28th round (about 154,000 households as against 23,000) the r.s.e.'s of the aggregate expenditure on food, non-food and total (food + non-food) in the 32nd round came out to be lower than those of 28th round.

*Thanks are due to Dr. J. K. Ghosh (Editor) for the following proof :

$$V = \Sigma R_s^2 Y_s^2 / (\Sigma Y_s)^2, \text{ where } R_s = \text{r.s.e. for } s\text{-th state}$$

$$= \frac{\Sigma R_s^2 Y_s^2}{\Sigma Y_s^2} \cdot \frac{\Sigma Y_s^2}{(\Sigma Y_s)^2}$$

The first factor, being a weighted average, is less than or equal to the maximum R_s^2 and the second factor is $\leqslant 1$, since all $Y_s > 0$. Therefore

$$V < \max R_s^2.$$

TABLE D-2. PERCENTAGE STANDARD ERROR OF ESTIMATE OF CONSUMER
EXPENDITURE (VALUE) ON ALL COMMODITIES,
SELECTED/SUB-GROUPS AND ITEMS

(highest and lowest values of r.s.e. (%) among selected states based on 28th round data)

total/subgroup/item	rural		urban	
	lowest	highest	lowest	highest
(0)	(1)	(2)	(3)	(4)
total expenditure	2.04 (UP)	4.81 (HP)	3.33 (UP)	9.72 (HP)
all food	1.87 (UP)	3.54 (HP)	2.92 (UP)	10.31 (HP)
all non-food	3.82 (UP)	9.00 (HP)	5.24 (UP)	14.20 (OR)
		Food		
milk and milk products	3.63 (Pb)	15.28(WB)	4.86 (UP)	17.21 (KL)
edible oils	2.37 (UP)	5.04 (KL)	3.67 (UP)	9.76 (KL)
(vanaspati)	5.10 (Pb)	11.11 (HP)	6.92 (Pb))	17.21 (HP)
meat, fish and eggs	4.13 (AP)	10.43 (Pb)	5.31 (TN)	17.95 (Pb)
vegetables	2.11 (UP)	5.85 (KL)	3.48 (UP)	9.83 (KL)
(potato)	2.89 (UP)	6.37 (HP)	4.31 (Mah)	6.45 (J & K)
fruits and nuts	5.32 (KL)	13.09 (WB)	6.06 (Mah)	11.42 (MP)
sugar	2.75 (Mah)	4.81 (TN)	3.58 (Mah)	8.32 (KL)
gur	3.69 (UP)	10.31 (HP)	—	—
spices	2.13 (UP)	3.20 (KL)	3.27 (UP)	6.56 (Pb)
beverages and refreshments	4.18 (Pb)	10.65 (WB)	5.00 (Mah)	10.95 (Pb)
		Nonfood		
clothing	5.59 (MP)	10.58 (Pb)	8.39 (Mah)	17.55 (KL)
footwear	9.38 (MP)	30.51 (TN)	16.61 (Pb)	36.22 (LL)
fuel and light	1.62 (UP)	2.81 (KL)	3.05 (UP)	7.43 (KL)
(electricity)	9.84 (Pb)	23.31 (UP)	8.60 (Pb)	18.95 (HP)
(firewood)	2.44 (TN)	3.96 (KN)	—	—
medicines	8.34 (Pb)	19.82 (UP)	12.11 (UP)	33.90 (Pb)
durable goods	20.79 (UP)	49.50 (AP)	31.58 (WB)	64.49 (TN)
misc. goods and services	4.65 (Pb)	6.92 (UP)	5.43 (Mah.)	10.61 (KL)

TABLE B-3. NUMBER AND METHOD OF SELECTION OF f.s.u.s AND
s.s.u.s FOR SCHEDULE 1-0 IN RECENT NSS CONSUMER
EXPENDITURE SURVEYS

	rural	urban	total (rural + urban)
(0)	(1)	(2)	(3)

28th Round (1973–74)

1. No. of FSUs	8680	4859	13539
2. Method of Selection	Randomly with probability proportional to population and with replacement	Randomly with probability proportional to size and with replacement.	
3. No. of SSUs	15467	7881	23348
4. Method of Selection	Linear systematically from the arranged frame prepared as follows: All the households were arranged into six classes on the basis of their means of livelihood (landless labour —1, cultivators —2 and others —3) and whether visited health centre or not.	Linear systematically from the arranged frame prepared as follows: All the households in a sample arranged into six classes of occupation division	

32nd Round (1977–78)

1. No. of FSUs	8216	4871	13087
2. Method of selection	Randomly with probability proportional to population and with replacement	Randomly with probability proportional to size* and with replacement.	
3. No. of SSUs	99766	58162	157982
4. Method of selection	Circular systematically from the arranged frame obtained by arranging the households in the sample village by the means of livelihood classes self-employed — non-agricultural —1, rural labour —2, and others —3.	Circular systematically from the arranged frame obtained by arranging the households into four classes on the basis of their nature of employment (self employed, not self-employed) and their per capita expenditure level.	

*Size being block population in thousands rounded upto the next higher integer.

REFERENCES

CHATTERJEE, G. S. and BHATTACHARYA, N. (1975) : *Some Observations on NSS Household Budget Data*, in Dandekar, V. M. and Venkataramaiah, P. (editors), *Data Base of Indian Economy; Vol. II*, Statistical Publishing Society, Calcutta, 97-109.

CENTRAL STATISTICAL ORGANIZATION (1980) : *National Accounts : Sources and Methods ;* Government of India, New Delhi.

DIRECTORATE of ECONOMICS and STATISTICS (1974) : *Bulletin on Food Statistics*, Ministry of Agriculture, Government of India, New Delhi.

JAIN, L. R. and TENDULKAR, S. D. (1987) : On a relationship between the real and the nominal relative disparities in the consumption of cereals in rural India, *Technical Report No. 8715*, Indian Statistical Institute, Delhi Centre (Mimeo).

KANSAL, S. M. (1965) : Preliminary estimates of total consumption expenditure in India, Indian Statistical Institute, Planning Unit, *Discussion Paper No. 5*, New Delhi. (Mimeo).

MAHALANOBIS, P. C. (1962) : A preliminary note on the consumption of cereals in India, *Bulletin of the International Statistical Institute*, **XXXIX**, No. 4, 53-76.

MINHAS, B. S., KANSAL, S. M. KUMAR, J. and JOSHI, P. D. (1986) : On the reliability of the available estimates of private consumption expenditure in India. Read at the *IARNIW Seminar on Tangible Wealth and Consumption*, Bhopal.

MUKHERJEE, M. (1969) : *National Income of India : Trends and Structure*, Statistical Publishing Society, Calcutta.

———— (1986) : Statistical information and final consumption in India and the National Sample Survey. *Economic and Political Weekly*, **XXI**, No. 5, 200-209.

MUKHERJEE M. and CHATTERJEE, G. S. (1972) : On the validity of NSS estimates of consumption expenditure, *Arthavijnana*, **XIV**, 113-121.

MURTHY, M. N. (1977) : Use of empirical studies in evaluating sample design for estimating frequency distribution. *Bulletin of the International Statistical Institute*, **XLVII**, No. 3, 191-211.

SARMA, V. R. R. and RAO, G. D. (1980) : Standard errors of estimates from the NSS 28th round consumer expenditure survey. *Sarvekshana*, **III**, No. 4, 87-106.

SUBYANARAYANA, M. H. and IYENGAR, N. S. (1986) : On the reliability of NSS data. *Economic and Political Weekly*, **XXI**, No. 6, 261-264.

VAIDYANATHAN, A. (1986) : On the validity of NSS consumption data. *Economic and Political Weekly*, **XXI**, No. 3, 129-137.