ARE EXISTING FOOD DEMAND DATA SOURCES ADEQUATE WHAT ARE THE ALTERNATIVES?

Robert E. Branson
Abbas Manban

To raise the question of data adequacy for food demand analyses is also to raise the question of adequacy for what. The only logical answer that question is to consider some additional questions such as the following. What present data sources do we have? What are the inherent purposes of the system in question? What data does the system generate? How reliable are they for the purpose intended? How reliable are they for use by economists in demand analyses, or even for market analyses for that matter?

Unfortunately, insofar as demand analyses are concerned, economists are in the position of data receivers not data generators. Therefore, it should come as no surprise that the data often fall considerably short of demand analysis requirements. Economists have been unusually complacent regarding the problem. Through this symposium an attempt is made to re-focus the attention of the profession on the two fundamentals components of demand analyses: data problems and analytical methodology. By so doing, perhaps interest will arise in their improvement. First lets identify and examine some of the present primary data systems which are to varying degrees available to us.

THE PRESENT DATA SOURCES

Present data sources may be classified into five basic systems: 1) consumer panels, 2) consumer household surveys, 3) retail store audits, 4) warehouse withdrawals and 5) a combination of separate price surveys and national product

Presented at AAEA meetings, Urbana July 27-30, 1980

disappearance data. Attention will be directed to some examples of each. Clearly it is not feasible to consider every data system in operation, so, attention will be directed to eight reasonably well known examples, well known at least among professional market research circles.

Household panels - Single Time or Continuous

Among well known commercial data generators of household panels are those operated by MRCA (Market Research Corporation of America) and NPD (National Product Data). Additionally two university or experiment station based panels are operational. One is at Griffin, Georgia, and a new one has been established in Puerto Rico.

The hallmark of household panels is data generation from a food purchaser diary maintained by the household in question. These are kept usually on a weekly basis. The item purchased, brand, size, price and kind of outlet from which purchased are recorded. Socio-economic data on the household are obtained and periodically updated. Psychometric data or information are not attained. Data are tabulated on a weekly, monthly, quarterly basis, though other bases can obviously be obtained upon special request and at additional cost. The MRCA abd NPD national panels provide regional data and selected major markets data as well as U. S. totals. Product coverage is far more complete in the Griffin panel and the Puerto Rico panels because MRCA and NPD are not inclined to develop data which is not commercially demanded.

Retail Store Audits

A number of retail store audit data services exist. A long standing and well known one in marketing circles is provided by the Nielson Company. Audit Surveys is another such service that provides national data.

The Nielson data are developed from product sales obtained from a nation-wide sample of retail stores. A four-week moving period is covered within each two-month time frame, giving six reports per year. Sales are calculated from beginning inventories, to which individual store deliveries are added and ending inventories are substracted. Data include quantity, prices of products by brand name and by package size. National regional and major market data can be provided.

Warehouse Withdrawals

The only national data service using warehouse shipments as a base is that by SAMI, a contraction for sales area market information. Coverage involves 36 SMSA markets located from coast to coast. The base reporting period is four weeks. Data include quantity and price by brand and package size of product. Though quantities are warehouse withdrawal figures, prices are retail based because food chain stores are involced on a retail value basis.

Consumer Household Surveys

Numerous household surveys are conducted each year. The ones of most interest to demand analysis economists are the BLS Consumer Expenditures Survey and the USDA Household Food Consumption studies. The data base for both is a combination of diary records kept by the households and recall reports. Historically these surveys have been made about every ten years. Included in the diary are food purchases and expenditures. Another phase obtains data on food use rather than food purchase. Prices are derived from the quantity expenditure totals reported. Socio-economic demographics information is obtained for each participating household.

EVALUATION OF PRESENT DATA SYSTEMS

The question of the adequacy of data from these various sources first must be viewed in terms of satisfying their primary purpose and secondarily their usefulness to demand analysis research. The most straightforward approach is to evaluate each data system in turn, with respect to adequacy of the survey sample involved, the product coverage included, the time period frequency of reporting, the presence of biases within the data availability of household characteristic information, and finally the practical matter of the accessibility of the data series for research economists demand analyses use.

Unfortunately, based on the six above criteria, all of the data systems available to date very clearly fall short of complete demand analyses requirements. The shortfall varies from one data source to another. The existence of this condition is a direct function of demand economists being, by and large, data receivers rather than essential collaborators in data generation. In fact there is reason to question the unfortunate isolation of which has occurred between demand analysts on the one hand and on the other hand the market or economic survey researchers who design and execute the data generating systems. Attention will now be directed toward the data shortfall problems.

Household Panel Data

A major question in using household panel data is the likely presence of substantial inherent biases in the quantities reported. Such biases are often an automatic fallout of the survey design and methodology. Demand analysis would not be seriously impaired if data biases were consistently upward or downward across all commodities by the same degree of error. Cross product elasticities particularly are affected if the rate or direction of bias among

products varies in quantity or direction. Time has not permitted a detailed analysis of this problem, so a few selected products are cited simply to illustrate the point, Table 1. Comparisons shown are not strictly valid because of each of comparability as to what is measured. The implications nonetheless are rather clear.

It would be well for all econometricians in demand analysis to remind themselves that just as the results of their analyses are no better than the analytical model employed so is it also no better than the data input. The degree of data error suggested by the above comparisons is magnified or reduced, depending on the product case, by the amount of the food used in combination products not considered in the panel data. Included are such items as ingredients in soups, TV dinners and various other further processed foods. And the latter are an increasing share of consumer purchases within recent years. Less vegetables and meats will be bought as such in the future so the problems will be compounded.

Further methodology research is needed to guide food purchase surveys procedures. Forgotten perhaps is the MRCA experience that a household placed in a food diary framework over purchases status foods and/or simply increases purchases in general in order to "look its best". That is but a natural psychological response. Therefore the diary of a new household is literally thrown away for the first three months because evidence is that it takes that long for the purchase pattern to normalize itself. Yet these are the very data, in many cases, which demand analysts consume and on which analysis are built when Consumer Expenditures Survey or the Food Purchase Study data are employed.

Even in the MRCA continuing panel, efforts to check the validity of the data have had poor results. Efforts to do so find that some elements of product flow data are lacking such as variations in stocks in wholesalers' hands. Even so it is most disturbing that one year's data may appear to be 15 or 20

Table 1. Comparison of Food Purchases Survey and Food Disappearance Data, 1965

Product	Food Consumption Study	Food Disappearance	Difference percent
	-- pounds per capita --		
Beef, lbs.	83.2	73.6	+13
Pork, lbs.	51.0	66.0	-23
Carrots, fresh lbs.	8.6	6.5	+32
Eggs	43.0	39.7	+ 8
Tomatoes, canned	7.0	4.6	+51
Corn, canned	7.5	5.4	+40
Oranges, fresh	20.8	15.9	+31
Peanuts	6.8	5.6	+21
Beans, dried	4.8	6.5	-26

Source: Johnson, Allen O., Food Consumption, Prices and Expenditures, Agricultural Economic Report Number 138, 1977 Supplement, USDA.

Food Consumption of Households by Money Value of Food and Quality of Diet, United States, North and South, U.S Department of Agriculture, Agricultural Research Service.

percent high and the next year short by a similar amount. MRCA, in fact, lays no claim to absolute accuracy. Rather their data users are advised to think in terms of relying primarily only on month change rates instead of absolute levels. This should signal an important message.

Re-examination should also be made of the length of the recall period used for food purchase or consumption data. Prior research in 1957 concluded that food quantities reported varied by as much as 30 percent depending upon the days to be recalled and indeed even upon what day of the week the question was asked. The human mind is the most phenomenal presence in the world, but a packet calculator can outstrip it in this facet every time. Given the fact that depending upon the analytical model used the demand coefficients can vary easily by 30 percent and that data errors may be of equal or larger size, it is little wonder that food policy administrators look upon the results with some trepidation as a guide to policy formation.

Unfortunately, the data from the MRCA and NPD panels, which have at least some of the data bias eliminated, such as the inflated three month initial reporting period, have not been available to university researchers.[2/] To purchase the data is prohibitively costly.

Even if food demand analysis owned the MRCA and NPD data banks, the problems would not be resolved. Usually the categories of foods are truncated to those for which a commercial market demand exists. Consequently it is an incomplete data set.

Another bias source arises within consumer household surveys. A commonly recognized error source by market researchers is that continuing consumer panels of any type - - for attitude response, for product evaluations, or

---

[2/] There is a possibility of this position being modified for the future.

purchase rates - - are not a true probability sample of households simply by virtue of the fact that households are inherently of two groups - - those who are cooperators and those who are non-cooperators. Diary keepers are likely to be motivated by different drives than are non-diary keepers. They can be more organized in their planning, more money expenditure conscious and possess other attributes which affect consumer behavior. Even our most sophisticated stratified probability sampling can not avoid this handicap.

Retail Store Audits

The major handicap of retail store audit data lies in the lack of household information. Therefore, cross sectional analyses are severely limited. Only approximations of the store service area household income and the ethnic mix of the population it serves are achievable. Any study of income variation within a census track area of a city provides doubt as to the applicability even of these approaches.

Retail audits, too, like other commercial data services are not necessarily all inclusive. Emphasis is on branded merchandise. Such food categories as meats, fluid milk and fresh fruits and vegetables and a number of baked goods are typically missing.

On balance store audit data are probably more accurate and the sample from food store audits is probably superior to that of household panels or surveys. This simply follows from each store representing, on the average, about 1000 households. The sampling problem of audits is whether or not they include all classes of food outlets, for example, the small convenience drive-in type of store and the delicatessen type unit may be omitted.

One of the more difficult aspects of store audits is the accuracy of records on drop shipments by suppliers who stock their own merchandise shelves, the presence of special emergency orders or intra-store exchanges of stock within a food chain that escape the normal record checking processes. Anyone who has conducted retail

audits in connection with marketing tests can attest to the foregoing difficulties. For a nationwide audit these errors can be compensating to some degree.

Warehouse Withdrawals

One of the added problems of data sources is that found in the SAMI warehouse withdrawal data - - the extent of warehouse coverage. Most data systems that are retail store based encounter an inability to obtain a true probability sample. Not all food chains have been willing to participate in the SAMI system. One is forced thereby to the unwanted assumption that those included are reasonably good indicators of the total market.

SAMI fortunately does make adjustments in its data to exclude shipments to store outside of the SMSA area of the market concerned. This is advantageous because some food chain warehouses service half or more of the state in which they are located. Therefore with the foregoing adjustments sales data can be related to the demographics of the market in question.

Separate Price and Consumption Surveys

Usually demand analyses for individual commodities, or products are forced to utilize data from separate sources. And these data were not meant to be homogeneous in the first place. Such was the problem faced by King and George as well as Bandow. The prices data if retail level are usually from the BLS Consumer Price Index, or if farm level are from the Prices Received by Farmers Index of the USDA. Inasmuch as only about 150 food items are priced for the CPI, compared to several thousand in a supermarket, many of those of key significance to demand analysis are missing.

The contrary side of the coin is that consumption data are limited to those for which supply and utilization data can be obtained. The demand analysts is caught, for instance, with figures that relate to beef slaughter and its conversion

to retail equivalents without the breakdown of beef even into its major product categories. In essence the analyst is faced with retail prices that relate to a limited number of forms of beef and disappearance data that are inclusive of all forms. Such a problem pervades most of the food products to varying degrees.

Though the foregoing problems exist, recognition must be given to the strength of the data systems discussed. Household panel data such as MRCA and NPD are very useful to measure levels and trends in brand shares of product lines and that is the major focus of the system. The Consumer Punchase Study, judging from the experience of the marketing research profession, probably overstates the consumption of the highly desired foods and understates the others, but it is the best so far for its intended purposes. The data in Table 1 seem to confirm that tentative conclusion about the nature of the biases. Only with reluctance are such old data cited. A search of the AAEA Journal, the Journal of Marketing Research and the Journal of Consumer Research show the subject problem has simply been ignored during the past 5 to 10 years.

Retail store audit data are reasonably good indicators of brand share positions among competing products, which, like MRCA data, is their primary aim. The same applies to SAMI warehouse withdrawal data.

Data collected for the CPI are reasonably adequate for that purpose and so are disappearance data to show trends in foods usage for many products. The key problem noted initially remains. It is simply that the purpose and use of the data are not coinciding with demand analysis purposes. Given these inadequacies the next question is whether better solutions are possible.

NEW DATA SYSTEM DEVELOPMENT

Fortunately some of the foregoing problems will be resolved by two new developments. One is the initiation in 1979 of a continuing quarterly consumer expenditure survey by the Bureau of Labor Statistics. Thereby the former ten year studies will become an ongoing quarterly survey with a rotating panel. Thereby a reasonably homogeneous prices and quantities data series will become available over time plus the inclusion of the related household socio-economic information. Faced with the more rapid adoption of food innovations because of energy costs, more women working and many other factors, a real challenge must be confronted in keeping abreast of changing household conditions and their impacting on food use behavior. The growing prevalence of microwave and heat convection ovens in food preparation, and, the increased purchases of dehydrated and pre-prepared foods are but some of the developments that must be dealt with in a constructive manner.

Unfortunately the problems of sample maintenance and of household purchase behavior conditioning effects of being a participant in a panel will still be present.

The second emerging opportunity is to see what success can be obtained in accessing, and building on, retail food store scanner data regarding food sales. More stores are adopting the system and the possibilities of using it to replace store audits is very real. Data accuracy will be greatly enhanced. Three negative aspects nontheless are faced. Only some stores with scanner equipment are presently scanning fresh meats or produce. In time this should change. The tremendous detail of the data, some 10,000 or more items, including brand and package size variations, must be somehow meaningfully dealt with. And finally the question of consumer demographies is missing, though experimental work in this area is now

underway. The tracking of purchases by a panel of consumers shopping at scanner stores is another approach or concept being tried. Left though, would still be the problem of food purchases at non-scanner establishments.

A number of coneptual possibilities for more dynamics in models of food demand emerge from the two new data systems. But that is the subject of the papers to follow. The key implication from the data systems' side is the growing importance for dialogue between data users and data generators to resolve the kinds of difficulties like those here discussed. Hopefully at some point in the future we will have a cohesive system that properly matches reliable data on quantities, prices, household characteristics (demographic and psychographic) and the environmental constraint within the household so that we can improve our understanding and predictive capabilities regarding food demand.

One final word about psychographics to which reference has been made several times without any clarification. Marketing researchers for some time have sought the link between consumer motivations and purchase behavior. With the new data systems at hand, and some concerted effort, this important dimension can be added. Whether a household is influenced by weight control, diet goats, time restraints or other motivators, these undoubtedly affect food purchase behavior, and our analytical skills should encompass these in the future. Therefore, fortunately, the future outlook is better than the performance of the past.