Scanner Data in Supermarkets: Untapped Data Source for Agricultural Economists

R.M. Nayga, Jr.*

Scanner data from supermarkets constitute a nontraditional data source for economic applications. Few agricultural economists have used scanner data in market or economic research. This paper addresses the benefits and problems associated with the use of scanner data in research. The extant literature in the agricultural economics field is reviewed and some implications for further research are discussed.

1. Introduction

Scanning in supermarkets has experienced considerable growth since its inception in the United States in July 1972 by the Kroger Company. In 1988, close to 16,000 grocery stores in the United States had already adopted scanning and the estimated dollar all commodity volume (ACV) of scanning stores as a proportion of total grocery business was about 60 per cent (*Progressive Grocer* October 1989). Scanner penetration in the 1990s is expected to approach saturation levels in major volume grocery stores (Wolfe 1990). Indeed, the increasing number of scanning systems in the grocery industry is indicative of the acceptance of this technology by the industry.

Traditional analysis of consumer demand has generally depended upon aggregate annual, quarterly, or monthly time series data of consumer prices and purchases. These data, however, do not represent current market conditions and are typically too general for product-specific decision-making. To quote Tomek (1985, pp. 913-914), "existing secondary data seem especially inadequate for studying product demand in retail markets, and fundamental work needs to be done to obtain relevant data". These traditional time series data, for instance, lack disaggregate product and price detail. On the other hand, consumer panels and consumer surveys provide more detailed data for specific products as well as sociodemographic information but they are expensive methods of data collection. A key limitation of consumer panels or surveys is their lack of price information. Prices must be imputed from reported quantity and expenditure figures. The use of such

---

*Lecturer, Department of Agricultural Economics and Business, Massey University, Palmerston North, New Zealand. The author wishes to acknowledge helpful comments made by G.R. Griffith and an anonymous reviewer. Any remaining errors or omissions are the sole responsibility of the author.

Review coordinated by Frank Jarrett.

imputations, particularly estimation of cross-sectional demand functions (Cox and Wohlgenant 1986), have been questioned by some analysts. Another limitation of the use of consumer surveys (not necessarily panels) is the lack of time continuity.

Scanner data, on the other hand, are primary data that have properties similar to cross-sectional and time-series data. The observations exist over time, usually days, as well as across various cross sectional units, typically food stores (Capps 1989). Scanner data, therefore, constitute a readily available current and timely source of product-specific information. The richness of scanner data lies in the fact that quantity, price, and hence expenditure information for a multitude of products is available on a daily basis. Hence, scanner information constitutes a non-traditional data source for economic applications.

Although scanner data have been available for several years to marketers, such data represent a new form of information for academics and food industry people. Marketers and researchers are just beginning to learn how to utilise this information source to make pricing or advertising decisions for various products. In fact, very few analyses of consumer demand have been conducted by agricultural economists using scanner data. Only since 1979 have scanner data, through refinements by manufacturers of electronic scanning check-out systems, been generated with enough reliability and consistency for application in economic research Jourdan 1981).

The tremendous potentials of using scanner data in economic and market research are addressed in this paper. In particular, the nature (benefits, problems, pitfalls) of scanner data are discussed and the extant literature in the agricultural economics field are reviewed. This paper concludes by discussing the implications for further research.

2. Present and Potential Applications in Economic Research

The introduction of scanning check-out systems into United States supermarkets in the mid-1970s opened tremendous possibilities for generating new data and for using such data in economic research and managerial decision making. Lesser and Smith (1986, p. 86) point out that with scanner data, "it is possible to do retail-level analysis routinely which previously required special tabulations". Examples of retail-level analyses requiring special tabulations include in-store pricing experiments (Doyle and Gidengil 1977), the effects of promotional programs on individual items (Hoofnagle 1965; Curhan 1974), the measurement of price elasticities (Funk *et al.* 1977; Marion and Walker 1978), the results of space allocation and display (Cox 1964; Curhan 1973; Chevalier 1985), and the effects of interactions among short-run strategy variables such as advertising, space allocation, and pricing (Curhan 1974; Wilkerson *et al.* 1982).

Supermarkets, as well as non-food retailers, are now using scanner information as a managerial tool. Top research companies

are now offering software that organises scanner data. Kiley (1990) indicated that "the realms of monthly supermarket scanner data are essentially useless without computer programs to make sense out of it".

Consequently, a number of companies and researchers have used scanner data for marketing research purposes. For instance, retailers have used scanner information to assess impacts of promotional activity, to determine optimal space allocation, and to develop sales management models. Scanner data have also been used in market research to investigate brand differentiation (Blattberg and Wisniewski 1986; Shugan 1987; Guadagni and Little 1983) and to investigate promotional effects on sales of performance (Wittink *et al.* 1988; Moriarty 1985; Eastwood *et al.* 1990). For instance, Eastwood *et al.* (1990) used scanner data to evaluate the effects of supermarket promotions and advertising on sales. Attention was focused on variable weight meat items and in the estimation of the partial impacts of the promotions on item movement.

A few demand analyses have also been conducted using scanner data (Jourdan 1981; McLaughlin and Lesser 1986; Capps 1989; Capps and Nayga 1990, 1991a, 1991b; Nayga and Capps 1991a, 1991b). Sands and Guylay (1984) acknowledged that scanner's application and benefits should increase enormously as researchers become more familiar and skilled in the management and use of scanner-based data.

A list applied studies done with the use of scanner data is presented in Table 1. Some of these studies are briefly described below.

In 1981, Jourdan estimated own-price and cross-price elasticities of demand for specific retail cuts of beef (roasts, steaks, ground beef and nonground beef) using bi-weekly data over a 25-week period from four retail food stores in Houston, Texas. McLaughlin and Lesser (1986) reported an experiment of systematically varying prices and tracking subsequent movements of potatoes, through the use of scanner data. They calculated store-specific demand elasticities in order to assess the impacts of promotional activity, to determine optimal space allocation and to develop sales management models. The common thread in these two consumer demand applications is the interaction with a single firm (although multiple stores) in a local area.

Retail demand relationships for steak, ground beef, roast beef, chicken, pork chops, ham, and pork loin were examined by Capps (1989). This research demonstrated the feasibility of using scanner data in developing short-run predictive models to anticipate sales of meat products. As well, the Center for Agricultural and Rural Development at Iowa State University, under contract with the National Live Stock and Meat Board of the United States, conducted an analysis of scanner data to measure beef consumption responses to television promotion and advertising (Schroeter 1988). Fresh beef purchases of

Table 1: Applied Studies Done Using Scanner Data		
Author	**Year**	**Objective**
Jourdan	1981	elasticity estimation for retail beef cuts
Guadagni and Little	1983	investigation of brand differentiation
Moriarty	1985	investigation of promotional effects on sales
Blattberg and Wisniewski	1986	investigation of brand differentiation
McLaughlin and Lesser	1986	examination of the effect of price variability on potato sales
Shugan	1987	investigation of brand positioning
Schroeter	1988	measurement of beef consumption responses to TV promotions
Wittink, Addona, Hawkes and Porter	1988	investigation of promotional effects on sales
Capps	1989	estimation of retail demand relationships for meat products
Capps and Nayga	1990	evaluation of effect of length of time on measured demand elasticities
Eastwood, Gray and Brooker	1990	evaluation of effects of supermarket advertising on product sales
Capps and Nayga	1991a	investigation of the demand for lean, nonlean and convenience beef products
Capps and Nayga	1991b	estimation of retail demand functions for fresh beef products
Nayga and Capps	1991a	analysis of demand for disaggregated meat products
Nayga and Capps	1991b	test of weak separability on various groups of disaggregated meat products
Capps and Lambregts	1991	estimation of demand functions for finfish and shellfish products

approximately 1800 households were monitored in Grand Junction, Colorado over the period October 1985 to July 1987. Combined with the detailed demographic information available for each of the households, scanner data provided a unique capability to assess the impact of the experimental television advertising.

The use of scanner data also permits the focus of analysis of a demand study on shorter time intervals than traditional demand analyses dependent upon aggregate annual, quarterly, or monthly time series data of purchases and prices, and also allows the analysis of more disaggregate food commodities. The limits on demand analysis can be expanded through the use of scanner data. For instance, Capps and Nayga (1990) were able to evaluate the effect of length of time on measured demand elasticities using scanner data from a retail food firm in Texas. Using disaggregate beef products and different time periods (weekly, biweekly, and monthly), they were able to examine the nature of dynamic adjustment in consumer demands for disaggregate food commodities and the sensitivity of dynamic adjustments in demands to shorter time intervals.

Capps and Nayga (1991a) also used scanner data to investigate the demand for lean, nonlean and convenience beef products for a local market in Houston, Texas. This research demonstrated the feasibility of scanner data in developing econometric models to analyse sales of beef products at the retail level. In particular, their study focussed on weekly point-of-sale purchases of 147 individual beef products, aggregated into nine major products.

The same scanner data set was then used to investigate the demand for fresh beef products (brisket, ground, loin, rib and all other beef) (Capps and Nayga 1991b). Once again, the research documented the utility of scanner data in market research. Additionally, their study showed how the use of scanner data permitted the focus of analysis on shorter time intervals (e.g. weekly) and more disaggregate beef commodities. Using the same type of analysis, Capps and Lambregts (1991) also estimated demand parameters for disaggregate finfish and shellfish products using scanner data from a retail firm in Texas.

Many studies of meat products have been based on the use of demand systems. Many of the published studies correspond to systems of less than ten aggregate commodity groups (e.g. George and King 1971; Hassan and Johnson 1976; Huang and Haidacher 1983). The problem, however, of using aggregate commodities in demand analysis is that considerable amount of information is lost concerning demands for the disaggregate commodities. Our knowledge about market demand for disaggregated commodity groups, in general, is limited. On the basis of separability tests, Eales and Unnevehr (1988) suggest that it is important to develop models for disaggregated meat products to obtain a fuller understanding of demand, rather than the traditional aggregated models of meat products.

With the use of scanner data, Nayga and Capps (1991a) were able

to fill this void by analysing the demand for disaggregate meat products in a demand system framework. In particular, they focused their analysis on 21 disaggregate fresh meat products and used a demand system approach based on the Almost Ideal Demand System. Moreover, their study also incorporated advertising effects into the complete demand system, based on the theoretical frameworks of Basmann (1956) and Baye, Jansen and Lee (1991).

Several studies which involve testing for separability in demand models have surfaced in the literature in recent years. Many of these studies have focused their analyses, once again, on broad and highly aggregated commodities (e.g. Swofford and Whitney 1987). Separability restrictions have usually been rejected in empirical work due perhaps to the use of broad commodities. In fact, Pudney (1981, p.561) states that "the empirical fruit of the theory has been disappointing, but possibly only because it has generally been applied at the wrong level of aggregation". Pudney also indicated that some information on the appropriate grouping patterns of the commodities, for instance, could be extracted from using a lower level of aggregation. Nayga and Capps (1991b) deviated from previous analyses by focusing on tests of weak separability on various groups of disaggregated meat products using scanner data. The results of their study have important implications in the creation of demand systems because of the two-stage budgeting concept.

The studies discussed above were conducted using scanner data from the United States. Although scanners are used in many countries, there has generally been no other published economic research studies done using scanner data from other countries. The results from the studies reviewed above are "location" specific and care should, therefore, be used in generalising these results to regional, national, and international levels (see the section on "Problems and Pitfalls" below).

3. Problems and Pitfalls

Overall, the research studies reviewed above encourage prospects of using scanner data in market and economic research. Despite the apparent success of using scanner data to analyse retail demand relationships, concern lies with generalising the results to regional or national levels. Scanner data from supermarkets in a particular location represent a "controlled" experimental situation. The community specific results may not contribute to broad regional or national inferences. Due to this potential limitation, the results of analyses involving scanner data should be used with care and not on a stand alone basis but as supporting evidence in conjunction with a research approach designed to conduct analyses with scanner data on a regional or national level.

Additional limitations of scanner data include (Capps and Nayga 1991a, p. 8): (1) the sheer volume of information; (2) the lack of demographic and income information; and (3) the provision of information only for food eaten at home.

Each week as few as 10 to 20 supermarkets will generate the equivalent amount of data as would a panel of 10,000 households. Price, quantity, and hence, expenditure information on a multitude of products are available on a daily basis. Consequently, considerable resources are necessary to reduce the mass of data to useful summary figures for demand analyses. The sheer volume of scanner data may be characterised as "trying to take a drink from a fire hydrant" (Capps and Nayga 1991b). Because of the potential for data overload and sometimes the problem of data integrity, empirical practitioners have been less than enthusiastic about the use of scanner data in market research.

Eastwood (1990, p.45) mentioned two practical problems in creating scanner data sets for demand and marketing research. The first relates to organising scanner data for variable weight items into consumer demand categories. The second pertains to the set of problems associated with the creation of an advertising data set that can be merged with scanner data to assess marketing strategies.

Scanner data, at least from retail food firms, also lack the dimension of consumer sociodemographic data. This socio-demographic information is essential to the derivation of income elasticities. Further, it is necessary to augment scanner data files to monitor advertising and promotional activities as well as to monitor customer counts. Competitors' actions are also important to consider in the analyses, but are extremely difficult to anticipate, measure, and evaluate. Additionally, difficulties exist in the representation of nonprice effects (merchandising schemes, coupons, services, cleanliness, product selection, and reputation for fresh meat or produce) (Capps and Nayga 1991a).

Scanner data provide information only for food eaten at home. One of the most noticeable changes in eating habits of consumers in recent years is the increased incidence of meals eaten outside the home. As the trend toward increased consumption of food away from home continues to grow, it is possible that the food service industry will consider the adoption of scanner technology in restaurants and fast food establishments to reap some of the benefits that supermarkets are getting from using scanner information.

In regard to data integrity, some food industry observers criticise scanner data as inaccurate. Some problems associated with this claim are bad symbols and poorly trained checkers. Lesser and Smith (1986) point out that scanner data misrepresent item movement if the scanning file is not rigorously maintained or if the items cannot be or are not scanned and the Universal Product Codes are not entered manually.

4. Concluding Comments and Implications for Further Research

The introduction of scanners into the supermarket checkout process has received a lot of attention in the popular press,

food marketing publications and research journals in the United States. However, few agricultural economists (not only in the United States but also in Australia and New Zealand) have realised the benefits of scanner data in market research. Scanner data have tremendous potential for use in the analysis of consumer demand for specific products. Translating these data into information for management, advertising and pricing decisions, however, remains a major concern. Since the development of effective marketing programs is a primary concern of retail food chain executives, analyses with scanner data can be used in making important pricing and advertising decisions.

The limits on traditional demand analyses can be expanded through the use of scanner data. The few existing demand analysis studies done by agricultural economists can attest to the potential uses of scanner data in research. Eastwood (1990) acknowledged that new analytical approaches to demand research are possible with scanner data and relationships among substitutes and complements can be examined to obtain better estimates of the trade offs consumers make when selecting food.

Given the enormous cost considerations of either money or physical resources, and given the potential for data overload, it might be worthwhile for analysts (like agricultural economists and marketers) to lobby heavily for the effective acquisition and organisation of scanner data. Although analysts do not have the comparative advantage in data collection, they do have the comparative advantage in data analysis. Hence, it may be appropriate for public agencies (e.g. in Australia or New Zealand) to negotiate with private firms (e.g. Information Resources, Inc.) the acquisition and organisation of scanner data.

The costs involved in the acquisition of scanner data are not trivial, but neither are the costs involved in the acquisition of data from consumer surveys or panels. Furthermore, due to the enormous information involved with scanner data, an individual researcher may not be able to efficiently collect or organise the volume of information. Individual researchers might have to band together and combine their efforts in collaboration with national retail food chains or commodity groups, especially when conducting research in a national or regional level, to become cost effective. Otherwise, individual researchers should just focus on a local retail firm with multiple stores.

Scanner data hold great promise for developing insights into both applied and theoretical research. Although the realisation of benefits from the use of scanner data is in the embryonic stage of development, analysts should concentrate on scanner data assembly, management and analysis in the next five to ten years. Capps (1989, p.759) has put it perfectly when he said, "... with proper management, scanner data may well be the ultimate data source for demand analysis at the retail level". Indeed, scanner data may become the most detailed and definitive source of retail food industry statistics available to researchers and marketing executives.

References

Basmann, R. (1956), "A Theory of Demand with Variable Consumer Preferences", *Econometrica* 24, 17-58.

Baye, M., D. Jansen and J. Lee (1991), Advertising Effects in Complete Demand Systems, Working Paper, Department of Economics, Texas A&M University.

Blattberg, R.C. and K.J. Wisniewski (1986), Analysing Intra-Category, Inter-Brand Price Competition: A Theory of Price Tier Competition, Working Paper, University of Chicago, January.

Capps Jr., O. (1989), "Using Scanner Data to Estimate Retail Demand Functions for Meat Products", *American Journal of Agricultural Economics* 71, 750-60.

Capps Jr., O. and R.M. Nayga, Jr. (1990), "Effect of Length of Time on Measured Demand Elasticities: The Problem Revisited", *Canadian Journal of Agricultural Economics* November, 499-512.

Capps Jr., O. and R.M. Nayga, Jr. (1991a), *Leanness and Convenience Dimensions of Beef Products: An Exploratory Analysis with Scanner Data*, Bulletin Number B-1693 of the Texas Agricultural Experiment Station, Texas A&M University, April.

Capps Jr., O and R.M. Nayga, Jr. (1991b), "Demand for Fresh Beef Products in Supermarkets: A Trial with Scanner Data", *Agribusiness: An International Journal* May, 241-51.

Capps Jr, O. and J. Lambregts (1991), "Assessing Effects of Prices and Advertising on Purchases of Finfish and Shellfish in a Local Market in Texas", *Southern Journal of Agricultural Economics* July, 181-94.

Chevalier, M. (1985), "Increase in Sales due to In-Store Display", *Journal of Marketing Research* 12, 426-31.

Cox, K.K. (1964), "The Responsiveness of Food Sales to Shelf Space Changes in Supermarkets", *Journal of Marketing Research* 2, 63-7.

Cox, T.L. and M. Wohlgenant (1986), "Prices and Quality Effects in Cross-Sectional Demand Analysis", *American Journal of Agricultural Economics* 68, 908-19.

Curhan, R.C. (1973), "Shelf Space Allocation and Profit Maximisation in Mass Retailing", *Journal of Marketing* 37, 54-60.

Curhan, R.C. (1974), "The Effects of Merchandising and Temporary Promotional Activities on the Sales of Fresh Fruits and Vegetables in Supermarkets", *Journal of Marketing Research* 11, 286-94.

Doyle, P. and B.Z. Gidengil (1977), "A Review of In-Store Experiments", *Journal of Retailing* 53, 47-62.

Eales, J.S. and L.J. Unnevehr (1988), "Demand for Beef and Chicken Products: Separability and Structural Change", *American Journal of Agricultural Economics* 70, 521-32.

Eastwood, D.B. (1990), "Organising Scan Data for Market Research", *Journal of Food Distribution Research* June, 45-52.

Eastwood, D.B., M. Gray and J. Brooker (1990), Using Scan Data to Evaluate Advertising: A Suggested Methodology and Preliminary Results, Paper presented during the Southern Agricultural Economics Association meetings, Little Rock, Arkansas, February.

Funk, T.F., K.D. Meilke and H.B. Huff (1977), "Effects of Retail Pricing and Advertising on Fresh Beef Sales", *American Journal of Agricultural Economics* 59, 533-7.

George, P.S. and G.A. King (1971), *Consumer Demand for Food Commodities in the United States with Projections for 1980*, Giannini Foundation Monograph Number 26, California Agricultural Experiment Station, March.

Guadagni, P.M. and J.D.C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data", *Marketing Science* 2, 203-38.

Hassan, Z.A. and S.R. Johnson (1976), *Consumer Demand for Major Foods in Canada*, Economics Branch Bulletin Number 76/2, Agriculture Canada, April.

Hoofnagle, W.S. (1965), "Experimental Designs in Measuring the Effectiveness of Promotion", *Journal of Marketing Research* 2, 154-62.

Huang, C.L. and R. Haidacher (1983), "Estimation of a Composite Food Demand System for the United States", *Journal of Business and Economic Statistics* 1, 285-91.

Jourdan, D.K. (1981), Elasticity Estimates for Individual Retail Beef Cuts Using Electronic Scanner Data, Unpublished M.S. Thesis, Texas A&M University.

Kiley, D. (1990), "Finally, Some Much-Needed Relief from Scanner Data Overload", *Adweek's Marketing Week* April 9, 34.

Lesser, W. and J. Smith (1986), "The Accuracy of Supermarket Scanning Data: An Initial Investigation", *Journal of Food Distribution Research* 17, 69-74.

Marion, B.W. and F.E. Walker (1978), "Short-run Predictive Models for Retail Meat Sales", *American Journal of Agricultural Economics* 60, 667-73.

McLaughlin, E. and W. Lesser (1986), *Experimental Price Variability and Consumer Response: Tracking Potato Sales with Scanners*, Cornell University Agricultural Economics Staff Paper No. 86-28, September.

Moriarty, M.M. (1985), "Retail Promotional Effects on Intra and Interbrand Sales Performance", *Journal of Retailing* 6, 27-47.

Nayga Jr, R.M., and O. Capps, Jr. (1991a), "A Disaggregate Analysis of Meat Products: An Application Using the Almost Ideal Demand System and Scanner Data", Technical Article Number 30181, Texas Agricultural Experiment Station, June.

Nayga Jr., R.M. and O. Capps, Jr. (1991b), "Tests of Weak Separability: The Case of Disaggregated Meat Products", Technical Article Number 30182, Texas Agricultural Experiment Station, August.

Progressive Grocer (1989), Scanning Update October, 55.

Pudney, S. (1981), "An Empirical Method of Approximating the Separable Structure of Consumer Preferences", *Review of Economic Studies* 48, 561-77.

Sands, S. and T. Guylay (1984), "Supermarket Scanners: A Marketing Tool for the Eighties", *Business* April-June, 12-15.

Schroeter, J. (1988), *Analysis of Behavior Scan Data and Beef Consumption Responses to Promotion and Advertising*, The Center for Agricultural and Rural Development, Iowa State University, July.

Shugan, S.M. (1987), "Estimating Brand Positioning Maps Using Supermarket Scanning Data", *Journal of Marketing Research* 24, 1-18.

Swofford, J.L. and G.A. Whitney (1987), "Nonparametric Tests of Utility Maximisation and Weak Separability for Consumption, Leisure, and Money", *Review of Economics and Statistics* 69, 458-64.

Tomek, W.G. (1985), "Limits on Price Analysis", *American Journal of Agricultural Economics* 67, 905-15.

Wilkerson, J.G., J.B. Mason and C.H. Paksoy (1982), "Assessing the Impact of Short Term Supermarket Strategy Variables", *Journal of Marketing Research* 19, 72-86.

Wittink, D.R., M.J. Addona, W.J. Hawkes and J.C. Porter (1988), The Estimation, Validation, and Use of Promotional Effects Based on Scanner Data, Working Paper, Cornell University, January.