Department of Agricultural, Environmental, and Development Economics
The Ohio State University
Working Paper: AEDE-WP-0043-06

Impacts of Fat and Cholesterol Information On Consumer Demand: Application of New Indexes

Wen S. Chern (The Ohio State University)
Jun Zuo (First Union Bank, North Carolina)

E-mail: chern.1@osu.edu

November 2006

Abstract: Consumers' beliefs in the benefits of reducing fat intakes, especially saturated fat, and of increasing calcium intake from such foods as dairy products, depend upon the acquired information related to diet and health. This study develops new health information measures from different sources. The monthly information indexes, constructed for 1980-93, show that the amount of consumer information related to fat and cholesterol in circulation reached the highest levels during 1989-90. The results from an empirical application show that these new indexes of consumer health information about fats and cholesterol could explain the changing patterns of consumer choice for whole milk vs. lower fat milk in the United States.

Keywords: Information, fat, demand, weight function, milk, cointegration

JEL Classification: D12, I10, C81

AEDE Agricultural, Environmental, and Development Economics

Ohio State University Extension
Ohio Agricultural Research and Development Center
College of Food, Agricultural, and Environmental Sciences

Introduction

This study develops new indexes for measuring consumer information on fat, cholesterol, and calcium, and applies them to investigate the impacts of changing health concerns on consumer demand. As medical research increasingly stresses the relationships between diet and health, particularly those findings related to blood cholesterol and heart disease, health organizations and government agencies have been trying to educate the public about the importance of healthful diets. One of the widely publicized recommendations is to reduce one's fat intake to no more than 30% of total calories. There is no doubt that the public's increasing concerns about health have influenced food consumption patterns. The surging demands for canola and olive oils can attest to this trend of changing cooking oil consumption pattern. However, it is difficult to quantify such impacts in a commodity demand model.

For investigating the impacts of health concerns and the consumer valuation of reducing dietary health risks, researchers have relied on household surveys to collect data on the consumer's knowledge, concerns and attitudes on the nutritional and health related attributes of food. These data have been used to investigate the impacts of perceived health risk on food consumption (e.g., Lin and Milon, 1993; Gould and Lin, 1994) or to estimate the consumer's willingness to pay for reducing health risk in the food supply (e.g., van Ravenswaay and Hoehn, 1991). These surveys are very useful in handling the unobserved variables of consumption perception and consumer valuation of non-market transaction goods. However, it is often difficult to validate the correlation between the knowledge/concerns stated in the survey and actual consumption behavior. It is, therefore, important to evaluate the impacts of health concerns on actual behavior based on observed consumption data.

In analyzing the impacts of health concern on observed consumption over time, one of the most difficult problems is to construct historical time-series data for the unobservable variables characterizing the consumer's knowledge, concerns, or health risk belief. One commonly used ad-hoc approach is to use a time-trend variable to measure the continuing accumulation of the knowledge about health and diet (Ippolito and Ippolito, 1984) or to use a dummy variable to capture the impact of a significant event or program affecting consumer information (Ippolito and Mathios, 1990). However, this approach has its shortcomings because the time trend can capture the impacts of all other variables which happen to exhibit increasing or decreasing trends over the study period. In order to deal with this measurement problem, Brown and Schrader (1990) made a pioneering contribution by constructing their cholesterol information index (CHOL), using quarterly time-series data for 1966-1987. They used this index to quantify consumers' health concerns about cholesterol and their impacts on egg consumption. Their index was subsequently applied by Capps and Schmitz (1991), Chang and Kinnucan (1991), Yen and Chern (1992) and Kim and Chern (1995). Recently, Chern et al. (1995) used Brown and Schrader's data in a Bayesian-type consumer information learning model to derive the mean and variance of health risk belief and applied these measures in their study of the consumption of fats and oils.

The objective of this paper is to extend the current state-of -the-art measurement of consumer information about fat, cholesterol and other nutritional attributes such as calcium. Specifically, we will develop new indexes of consumer health information for fat and cholesterol as well as calcium for the period of 1980-1993. The validity of these measures will be analyzed. These new indexes are then applied to investigating the impacts of consumers' health concerns on fresh milk choice in the United States. The results show that these new indexes perform very well in this demand model for fresh milk.

Consumer Information

In neoclassical economics, the consumer's utility is specified as a function of the quantities of goods and services purchased. The demand function is derived from maximizing such a utility function subject to a budget constraint. One important assumption in this consumer choice model is perfect information on goods and services. When utility is specified as a function of quantities, it is easy to overlook the role of consumer information. Introducing a new approach to consumer behavior, Lancaster (1971, 1991) formulated a so-called characteristics model in which utility is specified as a function of product attributes (or characteristics). He argued that consumers purchase goods for their specific characteristics. One can easily relate his model to the real world situation. For example, the increasing consumption for canola and olive oils cannot be easily explained by a sudden change (increase) in the taste or preference of consumers for these oils. However, this demand surge can be easily attributed to the increasing recognition that these oils contain important attributes for a healthy diet. Specifically, they contain a high percentage of monounsaturated fat which has been shown to lower the bad blood cholesterol or low-density lipoprotein (LDL) and at the same time to increase the good blood cholesterol or high-density lipoprotein (HDL). In this case, this attribute of fatty acid cannot be realized simply by consuming or tasting these oils. It requires knowledge which can only be gained through information. It is, therefore, easier to incorporate the role of information in Lancaster's characteristics model than in the neoclassical consumer choice model.

It is likely that the impact of information on consumers' behavior depends upon its influence on the consumer's belief and attitude. As pointed out by Kinnucan and Chang (1993), previous studies have shown that consumers respond disproportionately to negative information such as linking saturated fat and cholesterol intakes to heart diseases. Furthermore, the consumer's attitudes toward a product are formed by his/her belief, which, in turn, is affected by information. Therefore, information is the basic source for the formation of attitudes and beliefs and it may be used as an explanatory variable in addition to income and prices in a consumer demand function.

One crucial methodological problem is how to measure consumer information or more specifically, the changes in the amount and/or spread of information. In this study, we focus on the consumer information related to fat and cholesterol. The specific question raised here is how much information the consumer had at his/her disposal regarding the health risk associated with the dietary intakes of saturated fat and cholesterol during a historical period. It is recognized that the objects or sources of information may change over time. In the case of dietary fat and heart disease, the earlier information (in late 1960s and early 1970s) referred to the relationships between dietary fat and blood cholesterol, and between blood cholesterol and heart diseases. The later information (in late 1970s) referred to the difference between saturated and unsaturated fat and the relationship between HDL and heart diseases. In the early 1980s, more medical findings were obtained regarding the benefits of polyunsaturated fats. Recently in the late 1980s and early 1990s, relevant information focused on the relative benefits of monounsaturated and polyunsaturated fats. The latest debates dealt with the Omega-3 fatty acids found in fish oil, the *trans* fats found in the hydrogenated vegetable oils in margarine, and the benefits of low fat diets, particularly, the Mediterranean diet. As one can see, the nature of consumer information on fat and cholesterol has changed over time. For measuring consumer information over time, we would measure not only the amount of information but also the quality of this information. Despite this changing nature of consumer information regarding fat and cholesterol, the increases in the information base within a specific time period must be based on new information. One may argue that consumer education programs such as the 1986 National Cholesterol Education Program in the U.S. can increase consumer knowledge (or acquired information) from existing information on fat and cholesterol. However, the effectiveness of this program can not be sustained for long unless there are new studies and new information being generated.

Information Sources

Where does the consumer obtain information about fat and cholesterol? We turn to the U. S. Department of Agriculture's 1987-88 Nationwide Food Consumption Survey (NFCS) for insights into this question. Table 1 shows the survey results on the sources of health information. Note that these results are about health information not the specific information about fat and cholesterol. However, the correlation between general health information and specific ones should be very high.

The USDA survey shows that 35% of American households obtained health information from doctors, nurses or other health professionals, 47% from newspapers, magazines and books, and 45% from food package and labels.[1] interestingly, only 14%

of American households used health information from government or health organization publications and only 13% used food company publications.

The survey results indicate that one of the most important health information sources is the newspaper. Therefore, we will attempt to construct a fat and cholesterol information index based on articles pertaining to these attributes from a major U.S. newspaper, The Washington Post. Another important information source is the MEDLINE database identified and used by Brown and Schrader (1990). The MEDLINE database contains abstracts of all articles published in medical journals in the entire world. Use of the number of articles published in medical journals as a measure of consumer health information can now be justified based on the findings presented in Table 1. Since more than 35% of households obtained health information from doctors, nurses, and other health professionals and since these health professionals likely have gained their information from published medical journal articles, it would be reasonable to use the number of medical journal articles as a proxy for measuring the changes in the amount of consumer information over time. Both databases of MEDLINE and The Washington Post can be accessed from the NEXIS-LEXIS services. Appendix A provides a more detailed discussion on the MEDLINE data base and how we used NEXIS-LEXIS to obtain the numbers of relevant articles from MEDLINE and The Washington Post. The reason that we selected the Washington Post is because it is the only major newspaper that we can retrieve back to 1980.

In addition, economists argue that advertising may provide useful information about product attributes to consumers. For the empirical application, we also include an information index based on the advertising expenditure and include it as an additional information variable.

New Measures of Health Information

Brown and Schrader (1990) constructed a cholesterol information index (CHOL) for their study of shell egg consumption in the U.S. The hypothesis underlying their methodology is that consumers' attitudes toward cholesterol changed slowly as scientific information accumulated, so that an index based on articles in medical journals could serve as a proxy for information reaching consumers from many sources.

Brown and Schrader (1990) first scanned all records in English articles dealing with humans and with clinical implications on the MEDLINE database from 1966 to 1987. Approximately 8,000 records were related to cholesterol and those which did not appear relevant to the links between dietary cholesterol, serum cholesterol, and heart disease or arteriosclerosis were discarded. They discarded all Scandinavian, British, and Canadian articles based on their belief that these articles were less likely to be read by U.S. physicians. The final number of articles included in the authors' study was 896 supporting and 39 questioning the linkages.

The numbers of articles supporting and questioning the link were compiled by quarter. The CHOL index was defined as a cumulative number of articles supportive of the link minus the number of articles questioning or disputing the linkages between cholesterol and heart disease. The aggregation procedure used by Brown and Schrader can be described by the following formula:

$$CHOL_t = \sum_{i=1}^{t} (NS_i - NA_i) \qquad (1)$$

where $CHOL_t$ is Cholesterol Information Index at period t (quarterly) and NS_i and NA_i are numbers of supporting and disputing articles at period i, respectively. Their computation began in the first quarter of 1966 as i=1, the earliest time period covered in

Medline. The index was assumed to have a value of zero prior to 1966. Since this index is based on a simple accumulation of the number of articles, it has a steady increasing trend. In fact, the R^2 of a regression between the index and time trend is 0.985. Consequently, the Brown and Schrader's index often performed like a time trend variable in a demand model as shown in Chern, et al. (1995). As many economic variables are highly correlated with time, any variable highly correlated with time such as the CHOL may not be able to represent exclusively the changing pattern of consumer health information on fat and cholesterol in an econometric model. This is our main motivation for attempting to construct the alternative measures in this paper.

The new Fat and Cholesterol Information Index from MEDLINE (FCIM) represents an extension of the methodology used by Brown and Schrader . The major new developments include use of different keywords in searching, inclusion of all English articles, and a different weighting method. One major departure from the CHOL is that an article is assumed to have a finite duration and lag distribution as a source of health information. This kind of assumption has been widely adopted in previous studies of the impacts of advertising.

In order to select those articles discussing the issues of relationships among fat, blood or serum cholesterol, and heart disease or arteriosclerosis, the key words used are Fat (s), Cholesterol, and Heart disease or Arteriosclerosis. Fat is a critical new keyword, which was not used by Brown and Schrader (1990).

All English articles are scanned for selecting the relevant articles from English medical journals. Unlike Brown and Schrader(1990), we include journals from England, Canada, and other European countries because, we believe, they are important information sources used by American physicians, researchers and public media.[2]

The monthly information index is defined as:

$$FCIM_t = \sum_{i=0}^{n} w_i \, NM_{t-i} \quad \text{with} \quad \sum_{i=0}^{n} w_i = 1.0, \qquad (2)$$

where $FCIM_t$ is the fat and cholesterol information index at period t (monthly), NM_t is number of relevant articles (both supporting and disputing the link) published at period t, w_i is the weight for articles published at period i, and n is number of lagged periods. This method considers not only carryover effect but also decay effect of information. It is assumed that after an article is published, it will continue to be a source of consumer information for a finite period of time. Furthermore its impact on consumer belief and attitude formation and retention will diminish over time. Research has shown that consumers are forgetful about the information they acquire unless with repetitive information and reenforcement such as commercial advertising. Also, the maximum impact of a newspaper article may occur when it is first published. However, for an article published in a medical journal, it may take sometime before it asserts its maximum impact on consumers. It simply takes time for the information to transfer from physicians to their patients. The carryover and decay effects are captured by specifying a weight function and the total lag period. In addition, supporting or disputing articles are not differentiated as done by Brown and Schrader because the percentages of articles questioning the link were very low (less than 3%) and almost no such articles have been published since 1985 (Brown and Schrader, 1990 and Kim and Chern, 1995). Therefore, all articles are considered as equally relevant. Table 2 lists the major differences between two indexes.

The traditional methods for constructing weights are based on Almon distributed lag or polynomial lag functions. One popular polynomial lag function is quadratic or second degree polynomial used in several demand analyses for advertising effects such as Ward and Dixon (1989) and Sun and Blaylock (1993). However, this function generates symmetric weights which are rather restrictive. Therefore, the following cubic or third degree polynomial weight function is used:

$$w_i = {}_0 + {}_1 i + {}_2 i^2 + {}_3 i^3 \qquad (3)$$

where *i* is the number of *i th* lagged period. We can determine the values of these coefficients based on the following criteria: (1) the maximum weight lies somewhere between the current period (*i=0*) and the last lagged period (*i = n*), (2) the minimum weight occurs at *i = n + 1* and is set to be zero, and (3) the sum of weights over the current and lagged periods is equal to one.

Let *n* be the number of total lag periods and *m* be the lag period with the maximum weight. Based on the above mentioned criteria, we can impose the following restrictions:

$dw_i/di = 0$ at $i = m$ for defining the maximum,

$dw_i/di = 0$ at $i = n + 1$ for defining the minimum,

$w_{n+1} = 0$, and

$w_i = 1$.

From these four restrictions, we can find the solutions to j's in terms of *m* and *n*. By substitutions, the cubic weight function can be rewritten as:

$$w_i = 2a/((n+1)b) + (12m/b)\, i - (6(n+1+m)/((n+1)b))\, i^2 + (4/((n+1)b))\, i^3. \qquad (4)$$

where

$$a = (n+1)^2(n+1-3m) \text{ and} \qquad (5)$$

$$b = (n+2)\,[(n+1)^2 - m\,(2n+3)].$$

(6)

For n = 12, Figure 2 shows the weight distributions for m = 0, 1,..., 4. In general, *n* and *m* can take any finite numbers.[3] There are no easy ways to determine n and m. We experimented with many combinations of m and n: m = 0, 1,...4, and n = 0, 6, 12, 24, 36 and found that in general the trends of the FCIW are not very sensitive to the values of m and n.

The fat and cholesterol information index from The Washington Post (FCIW) can be constructed by the same procedure. Figure 3 shows the FCIM and FCIW for m = 2, and n = 12. The correlation coefficient of these two indexes is 0.824. Since these two indexes have a relatively high correlation, they show very similar trends of changing consumer information on fat and cholesterol. These trends are very different from the ever-increasing trend depicted by either the Brown and Schrader's CHOL (Figure 1) or a time trend. Specifically, both FCIM and FCIW show that the weighted numbers of articles steadily increased from 1980 to about 1989-90 and then both declined sharply during the last half of 1990 and more recent years. In 1983, there were substantial increases in the published articles reflected in both FCIM and FCIW. Another interesting observation is that the number of published articles in FCIM recovered from the bottom of 1990 and continuously increased throughout most of 1991 to first half of 1993. However, the FCIW shows that it recovered from the same bottom but the recovery lasted only through the early 1992. The Washington Post appeared to have lost interest in publishing articles related to fat and cholesterol in 1992 and early 1993 despite the continued increases in the published articles in medical journals. It is recognized that these new indexes showing declines in consumer fat and cholesterol information in recent years, may appear counterintuitive to the general belief that consumers should always become more knowledgeable about fat and cholesterol. The validity of these indexes lies in the assumption that published articles should have a finite duration and decaying effects as a source of consumer information. Since this assumption is not based on any evidence from consumers about information decay, the validity of these indexes can only be attested by their explanatory power in a demand model for such products as fluid milk, red meats, animal fats and vegetable oils. In order to validate these indexes, the FCIM is used in the following empirical demand analysis for fresh milk.

The calcium information index (CI) is constructed using the numbers of articles containing keywords of Calcium, Nutrition, and (Dairy or Milk) from The Washington Post via NEXIS/LEXIS databases and using the same cubic weight function with n = 12 and m = 0. Figure 4 shows the comparison of this calcium index (CI) with another FCIM (m=1 and n=24). The CI shows that the calcium information steadily increased in early 1980s, peaked in 1985, and has steadily declined since 1988. The trend of CI appears to be quite different from that of FCIM during the study period. Similarly, the advertising information index (ADI) is constructed by weighting (with n =12 and m = 0) advertising expenditure data for fresh milk which were obtained from Broadcast Advertisers Reports (BAR), Inc., and the Leading National Advertisers (LNA), Inc.

An Application to Milk Consumption

With increasing public concerns about fat and cholesterol content in food, consumers are more likely to purchase food with less fat and cholesterol or even fat free or cholesterol free foods. One such appropriate food is fluid milk because of its various types specifically distinguished on the basis of fat content. When a household purchases fresh whole milk (with 3.25% milk fat), as opposed to low fat milk, the household may be assumed to be either ignorant about the health benefits of low fat milk, careless about higher fat content or simply like the taste of whole milk. On the other hand, when a household purchases other fresh milk (i.e. lower fat milk with 1% or 2% milk fat or skim milk), it might be consciously concerned about the fat content in milk. The health information indexes constructed earlier are used to examine the consumer choices of fluid milk in the United States.

Our model is based on consumer participation defined in terms of the proportion or percentage of households buying a specific good. Suppose, in time t, there are H_t households, in which h_t households purchase only fresh whole milk and l_t households purchase only other fresh milk. The representative (aggregate) household attitude toward fat at time t can be captured by the following two percentage variables:

$$r_{1t} = h_t / H_t \quad \text{and} \tag{7}$$

$$r_{2t} = l_t / H_t \,. \tag{8}$$

where r_{1t} is the proportion of households that purchase only fresh whole milk at time t and r_{2t} is the proportion of households that purchase only other low fat milk at time t. From the point of aggregation, r_{1t} and r_{2t} can be regarded as indicators of a representative household's attitude toward fat content in milk in time t, if H_t is defined as population of the specific representative household.[4]

Following Lancaster (1971, 1991), goods can be classified on the basis of their characteristics such as fat content. Based on utility maximization, the demand for goods can be indirectly derived as a function of income, relative prices and good characteristics. As such, the consumer choice of a commodity can be specified as a function of income, relative prices, and product characteristics. Since product characteristics such as fat content and its relationship to health are realized through consumer information, we can specify the participation rates in (7) and (8) as functions of income, relative prices, and consumer health information.

For empirical investigation, the consumer participation model for fresh whole milk and lower fat milk can be specified in the following double-log function:

$$\log r_{it} = a_i + b_{i1} \log p1_t + b_{i2} \log p2_t + b_{i3} \log m_t + c_{i1} \log FCIM_t + c_{i2} \log CI_t$$
$$+ c_{i3} \log ADI_t + \mathrm{d}_{i1}\, \mathrm{S} + \mathrm{d}_{i2}\, \mathrm{H} + e_{it} \qquad (i = 1,2) \qquad (9)$$

where r_{1t} and r_{2t} have been defined previously; $p1_t$ and $p2_t$ are prices (Consumer Price Indexes or CPIs) of fresh whole milk and other fresh milk, respectively; m_t is average income of households; $FCIM_t$, CI_t, and ADI_t are the fat and cholesterol information index from MEDLINE (n=24, m=1), calcium information index from The Washington Post (n=12, m=0), and advertising information index (n=12, m=0), respectively; S is the seasonal dummy having a value of one for the summer months of June, July and August, H is the holiday dummy for the months of November and December, and e_{it} is the disturbance term. The coefficients b_{i1}, b_{i2}, and b_{i3} are the parameters capturing the economic effects, while c_{i1}, c_{i2}, and c_{i3} are the parameters reflecting the information effects. The seasonal effects are captured by d_{i1} and d_{i2}.

The monthly proportions of households purchasing fresh whole milk and lower fat milk and average income of households are calculated based on individual household data in the public use tapes of Consumer Expenditure Surveys (CES) conducted by the Bureau of Labor Statistics (BLS). Fourteen annual surveys from 1980 to 1993 are considered in this study. Income, CPIs, and the original advertising expenditures are deflated by CPI for all commodities. Figure 5 shows the movement of the percentages of households purchasing fresh whole milk vs. lower fat milk. As one can see, the proportions have either an increasing or declining trend during the period of January 1980 to December 1993 and thus exhibit to some extent nonstationarity. Also, other variables such as health information indexes may present nonstationarity.

Integration and Cointegration

If variables in a model are nonstationary, the basic assumptions in statistical regression are violated and the consequences of the statistical estimation and hypothesis testing are profound as evidenced by the substantial literature on "spurious regression" (Hendry 1986). However, if a linear combination of those nonstationary variables is stationary, the variables are cointegrated (Engle and Granger 1987, 1991). In this case, the spurious regression does not exist and the OLS estimates are "superconsistent" (Engle and Granger 1991). A superconsistent estimate converges to its true value at a rate T^{1} rather than the usual $T^{1/2}$ as a consequence of the infinite variance of all other linear combinations. But, the inference using the corresponding *t*-ratios is not acceptable because the they no longer have a limiting normal distribution.

Many approaches have been developed to test for nonstationarity of individual variables and cointegration among multiple nonstationary variables. Two most widely used methods, the Dickey-Fuller (DF) and Augmented Dickey-Fuller (ADF) tests (Dickey and Fuller 1979, Engle and Granger 1991), are applied in this study to identify the degree of nonstationarity (or orders of integration) for all the variables appeared in equation (10) and presence of cointegration of the model. Suppose the time series x_t needs to be tested for number of orders of integration. In order to apply the DF test or ADF test, the following equations should be estimated first using the OLS,

$$\Delta x_t = a_1\, x_{t-1} + \varepsilon_t \qquad \text{(DF)} \qquad (10)$$

or

$$\Delta x_t = a_1\, x_{t-1} + \Sigma\, \beta_j\, \Delta x_{t-j} + \varepsilon_t, \qquad \text{(ADF)} \qquad (11)$$

where ε_t is an independent and stationary error process. The DF or ADF test involves examining *t*-statistic or *t*-ratio for parameter a_1, under the null hypothesis that $a_1 = 0$. The null hypothesis implies that x_t is nonstationary, i.e., x_t follows a random walk or has a unit root. The null hypothesis is rejected if the *t*-statistic is larger than a corresponding critical value, τ_1, generated by Dickey and Fuller(1979) or MacKinnon(1990).

If the null hypothesis is rejected, the time series x_t is said to be stationary or have zero order of integration, or I(0), otherwise it is nonstationary and needed to be further tested for higher order of integration. In the nonstationary case, the first difference of x_t can be further tested using the DF or ADF test. If the null hypothesis is rejected, the first difference is said to be stationary and the level of xt is referred to have the first order of integration, or I (1). If the null hypothesis is not rejected, further tests are needed to identify the higher order of integration by additional differences.

In order to test for cointegration of the model, the DF or ADF test can be applied to the regression residuals of the model estimated by OLS. If the residuals are stationary, the model is said to be cointegrated. Table 3 reports the t-ratios of DF and ADF tests for both levels and differences of variables and for the estimated OLS residuals in equation (9). As one can see, all the variables are I (1) processes and the estimated residuals are I (0). Therefore, the model is cointegrated.

Because a cointegration exists among the dependent and independent variables in the equation (9), the model can be estimated using OLS without spurious regression and the estimated parameters will be superconsistent. However, the t-ratios can not be used for inference because they do not have the limiting normal distributions due to nuisance arising from endogeneity and serial correlation. It is necessary to correct the possible existence of endogeneity and serial correlation by employing alternative specifications of equations (9). Two Phillips' methods (Phillips 1991, Phillips and Loretan 1991) are used to correct the endogeneity and serial correlation. The first method (Phillips 1991) is applied to equation (9) by incorporating the first-order autoregressive scheme, AR (1), to correct the possible first order serial correlation. The model with AR (1) can be expressed as:

$$\begin{aligned}\log r_{it} = {} & a_i + b_{i1} \log p1_t + b_{i2} \log p2_t + b_{i3} \log m_t + \\ & c_{i1} \log FCIM_t + c_{i2} \log CI_t + c_{i3} \log ADI_t + \\ & \rho_i (\log r_{it-1} - a_i - b_{i1} \log p1_{t-1} - b_{i2} \log p2_{t-1} - b_{i3} \log m_{t-1} - \\ & c_{i1} \log FCIM_{t-1} - c_{i2} \log CI_{t-1} - c_{i3} \log ADI_{t-1}) + \varepsilon_{it} \quad (i = 1,2) \qquad (12)\end{aligned}$$

where ρ is a scalar parameter capturing possible AR (1) effect and ε_t is an i.i.d. random disturbance term with E $(\varepsilon_t) = 0$ and var $(\varepsilon_t) = \sigma^2$.

The second correction, based on Phillips-Loretan method (Phillips and Loretan 1991), can be represented by the following equation, which amends both endogeneity and serial correlation:

$$
\begin{aligned}
\log r_{it} = {} & a_i + b_{i1} \log p1_t + b_{i2} \log p2t + b_{i3} \log m_t + \\
& c_{i1} \log FCIM_t + c_{i2} \log CI_t + c_{i3} \log ADI_t + \\
& d_{i1} \Delta\log p1_t + d_{i2} \Delta\log p2_t + d_{i3} \Delta\log m_t + \\
& d_{i4} \Delta\log FCIM_t + d_{i5} \Delta\log CI_t + d_{i6} \Delta\log ADI_t + \\
& \rho_i (\log r_{it-1} - a_i - b_{i1} \log p1_{t-1} - b_{i2} \log p2_{t-1} - b_{i3} \log m_{t-1} - \\
& c_{i1} \log FCIM_{t-1} - c_{i2} \log CI_{t-1} - c_{i3} \log ADI_{t-1}) + \varepsilon_{ti} \qquad (i = 1,2) \qquad (13)
\end{aligned}
$$

where d_j's are the parameters associated with the first-differences of all independent variables. Both equations (12) and (13) are estimated by nonlinear regression approach in order to reach the goals of correction.

Estimation and Analysis

Three model specifications are thus available in estimation: the basic double-log function characterized by equation (9) - Specification 1, cointegration with the AR (1) scheme characterized by equation (12) - Specification 2, and cointegration with both AR(1) and difference variables characterized by equation (13) - Specification 3. The OLS is applied to the first specification and the nonlinear regressions are applied to the other two specifications. The estimated parameters from Specification 1 will be superconsistent but the corresponding t-ratios can not be appropriately used for inference. On the other hand, the estimated parameters from both Specifications 2 and 3 have more reliable t-ratios.

The estimated results from the three specifications for whole milk are reported in columns 2 to 4 in Table 4. As one can see, nearly all the corresponding parameters in three specifications have the same expected signs: negative b_1, positive b_2, negative b_3, negative c_1, positive c_2, and positive c_3. The correction for serial correlation is successful in both Specifications 2 and 3 because the ρ's are significant. The large differences in c_j's between Specifications 1 and 2 imply that the serial correlation may be caused by the information indexes, while the large differences of b_j's between Specifications 2 and 3 hint that the endogeneity may come from the price and income variables. Specification 3 is chosen as the participation model for whole milk in the subsequent discussion. Also, higher adjusted R^2 provides another evidence that Specification 3 may be a better choice.

The estimated results of the three specifications for lower fat milk are shown in columns 5 to 7 in Table 4. The corresponding estimated parameters also have the same expected signs: positive b_1, negative b_2, positive b_3, positive c_1, negative c_2, and positive c_3. There are no significant differences in the estimated parameter values among the three different specifications, except b_3's. The significant ρ's justify the correction for serial correlation, while the insignificant d's, associated with difference variables, imply that no correction is needed for endogeneity. The similarity of adjusted R^2's for Specifications 2 and 3 also suggests that no difference exists between these two specifications. However, in order to be commensurate with the specification choice for whole milk, Specification 3 will also be used as the chosen participation model for lower fat milk. Overall, Specification 3 is a preferred choice in terms of its reliable *t*-ratios of estimated parameters and its higher adjusted R^2's.

As one can see from Table 4, the estimated parameters b_{i1} and b_{i2}, associated with prices, have expected signs but very low *t*-ratios. Therefore, the prices of whole milk and other fresh milk do not have significant effects on their consumer participation. In other words, the percentages of households choosing either whole milk or lower fat milk are price irrelevant. The estimated parameter b_{i3}, associated with household income, has expected signs, negative for whole milk and positive for lower fat milk. It is significant for lower fat milk but not significant for whole milk. The positive b_{23} for lower fat milk means that an increase in household income will result in an increase in potential choice for lower fat milk. Specifically, a 10% increase in household income will generate a 5.6% increase in percentage of households purchasing lower fat milk.

The significant estimated parameter c_{i1}, associated with fat and cholesterol information index, has a negative sign for whole milk and a positive sign for lower fat milk. Therefore, an increase in availability of fat and cholesterol information will result in a decline in the potential choice for whole milk and an increase in the potential choice for lower fat milk. Specifically, a 10% growth of fat and cholesterol information would generate an 8% decrease in percentage of households purchasing whole milk but only a 4% increase in percentage of households choosing lower fat milk. These results may explain why the per capita consumption of lower fat milk has been increasing but that of total fresh milk has been decreasing during the last decade.

The estimated parameter c_{i2}, associated with calcium information index, has a positive significant value for whole milk, implying that an increase in availability of calcium information would result in an increase in potential purchases for whole milk. A 10% increase in the exposure of calcium information would cause a 1.4% increase in the percentage of households buying whole milk. On the other hand, the estimated parameter c_{l2} for lower fat milk is significantly negative, indicating that an increase in calcium information may induce a decline in potential choice for other fresh milk. Part of the reason for this result may rest on the fact that some consumers may associate lower fat milk with low calcium. Overall, these health information variables perform very well in the model.

The estimated parameters of c_3, associated with advertising information index, are not significant, although they have the expected signs. Therefore, the advertising expenditures do not have significant effects on consumer participation toward either whole milk or lower fat milk during this sample period.

Conclusions

This paper presents alternative measures of changing consumer information on fat and cholesterol to the one constructed by Brown and Schrader. The new indexes are based on a different set of key words and a new weighting procedure. For computing these monthly indexes, a cubic (third degree polynomial) weight function is specified. The results show that the information index based on the number of medical journal articles from MEDLINE exhibits a similar trend as the index constructed from The Washington Post. However, they are different from the index constructed by Brown and Schrader. According to these new indexes, the amount of consumer information related to fat and cholesterol steadily increased since 1980 and peaked in 1989-90 but has been declining in recent years.

The cubic polynomial weight function developed in this study is very flexible. The same procedure is also applied to constructing the calcium information and advertising expenditure indexes. The monthly indexes developed in this study can be aggregated by the same methodology to create quarterly or annual series.

The new indexes are used to investigate the impacts of health information on the consumption choices of whole milk vs. lower fat milk in the Unites States. We carefully examine the time-series of all variables used in the consumption participation model and find that the model is cointegrated. Three models with and without cointegration are specified, estimated, and compared. The econometric results suggest that it is important to correct for serial correlation and endogeneity in this cointegrated consumer participation model of fresh milk.

The empirical results show that consumer participation rates for either whole milk or lower fat milk are not significantly affected by their prices, while household income has a positive impact on the participation rate for lower fat milk and a negative effect on choosing whole milk. More importantly, the results show that there are significant effects of health information on participation rates for both whole milk and lower fat milk. The new indexes developed in this study have strong explanatory power and produce reasonable results in this model. Specifically, the fat and cholesterol information generates, as expected, a negative impact on consumer participation for whole milk but a positive effect on consumer participation for lower fat milk. Increase in calcium information would result in more household purchases for whole milk but induce a decline in the consumption choice for lower fat milk. The effects of advertising information measured by expenditures are not significant for both whole milk and lower fat milk, although the effects are positive.

Footnotes

1. Note that a household may obtain information from several sources and therefore the sum of the percentage figures in Table 1 is not equal to 100.

2. For example, according to Journal Citation Reports (JCR, 1993), The Lancet, a leading British medical journal, ranked second, British Medical Journal ranked 7th, British Medical Bulletin ranked 19th, and Canadian Medical Association Journal ranked 28th by impact factors among 115 ranked general and internal medical journals. Among 158 ranked research and experimental medical journals, European Journal of Clinical Investigation is in 8th place and Scandinavian Journal of Clinical Lab Investigation is in 20th. Also, as mentioned before, only 44% of records came from journals published in North America with 97% of these published in the United States. It is important to include sources outside the U.S.

3. It is noted that m should be equal or less than four for n=12. This can be seen from equation (6) that (n+1-3m) should be positive. This requires $m \leq (n+1)/3$. Failure of this condition will result in a negative weight (w_i) which is not acceptable.

4. There were, of course, households who purchased both whole milk and lower fat milk. They are not included in this study because their attitude toward fat content can not be clearly distinguished.

References

Brown, Deborah J., and Lee F. Schrader, "Information on Cholesterol and Falling Shell Egg Consumption." American Journal of Agricultural Economics, 73(1990):548-555.

Capps, O., Jr., and J.D. Schmitz, "A Recognition of Health and Nutrition Factors in Food Demand Analysis." Western Journal of Agricultural Economics, 16(1991):21-35.

Chang, H.S. and H. Kinnucan, "Advertising, Information, and Product Quality: The Case of Butter." American Journal of Agricultural Economics, 73 (1991): 1195-1203.

Chern, Wen S., E.T. Loehman, and S.T. Yen, "Information, Health Risk Beliefs, and the Demand for Fats and Oils." Review of Statistics and Economics, (August 1995): 555-564.

Dickey, D.A., and W.A. Fuller, "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." Journal of the American Statistical Association, 74(1979):427-431.

Engle, R. F. and C. W. J. Granger, Long-Run Economic Relationships: Readings in Cointegration, Oxford: Oxford University Press, 1991.

Engle, R.F. and C.W.J. Granger, "Co-integration and Error Correction: Representation, Estimation and Testing." Econometrica, 55(1987): 251-276.

Gould, Brian W. and H.C. Lin, “Nutrition Information and Household Dietary Fat Intake.” Journal of Agricultural and Resource Economics, 19 (1994): 349-365.

Hendry, David F., " Econometric Modelling with Cointegrated Variables: an Overview." Oxford Bulletin of Economics and Statistics, 48(1986): 201-212.

Ippolito, Pauline M. and Alan D. Mathios, "Information, Advertising and Health Choices: A Study of the Cereal Market." Rand Journal of Economics, 21(Autumn 1990): 459-480.

Ippolito, P.M. and R. A. Ippolito, "Measuring the Value of Life Saving from Consumer Reactions to New Information." Journal of Public Economics, 25 (1984):53-81.

Kinnucan, Henry W. and Hui-Shung Chang, "Advertising Evaluation: Seven Concepts and A Model." Commodity Promotion Policy in a Global Economy, ed. Walter J. Armbruster and John E. Lenz, Farm Foundation, Oak Brook, Illinois, 1993.

Kim, Dong-Kyoon and Wen S. Chern, "Health Risk Concern of Households vs. Food Processors: Estimation of Hedonic Prices in Fats and Oils." Valuing Food Safety and Nutrition, ed. Julie A. Casewell, Boulder, Co.:Westview Press, 1995.

Lancaster, Kelvin, Consumer Demand: A New Approach, New York: Columbia University Press, 1971.

Lancaster, Kelvin, Modern Consumer Theory, Brook, Vermont: Edward Elgar Publishing Co., 1991.

Lin, C. T. Jordan and J.W. Milon, "Attribute and Safety Perceptions in a Double-hurdle Model of Shellfish Consumption." American Journal of Agricultural Economics, 75 (August 1993): 724-729.

MacKinnon, James, " Critical Values for Cointegration Tests." Long-Run Economic Relationships: Readings in Cointegration, ed. R.F. Engle and C.W.J. Granger, Oxford University Press, 1991, pp.267- 276.

Phillips, P. C. B. "Optimal Inference in Cointegrated System." Econometrica, 59(1991): 283-306.

Phillips, P.C.B. and M. Loretan, "Estimating Long-run Economic Equilibria." Review of Economic Studies, 58(1991): 407-436.

Pratt, G.F., " A brief hitchhiker's guide to MEDLINE." Database , 17 (February 1994): 41- 49.

Sun, Theresa Y. and James R. Blaylock, An Evaluation of Fluid Milk and Cheese Advertising, Technical Bulletin, No. 1815, ERS, USDA, 1993.

van Ravenswaay, E. O., and J.P. Hoehn, "The Impact of Health Risk and Food Demand: A Case Study of Alar and Apples." Economics of Food Safety, ed. J.A. Caswell, NY: Elsevier Science Publish Co., 1991.

Ward, Ronald W. and Bruce L. Dixon, "Effectiveness of Fluid Milk Advertising since the Dairy and Tobacco Adjustment Act of 1983 (National Dairy Promotion and Research Board)." American Journal of Agricultural Economics, 71(1989):730-740.

Yen, S.T., and Wen S. Chern. "Flexible Demand Systems with Serially Correlated Errors: Fat and Oil Consumption in the United States." American Journal of Agricultural Economics, 74 (1992): 689-697.

Table 1. Sources of Health Information By American Households

Sources	Percentage of Households Using These Sources (%)[a]
Doctor, Nurse or Other Health Professional	35.36
Nutritionist, Dietitian, Home Economist or Extension Agent	17.69
Relatives or Friends	24.03
Radio or Television	33.02
Newspapers, Magazines or Books	47.30
Government or Health Organization Publications	14.06
Food Company Publications	23.75
Food Package or Labels	44.96

[a] Based on a sample of 4,273 households.

Source: U.S. Department of Agriculture, 1987-88 Nationwide Food Consumption Survey, Household Portion, Public Use Tape.

Table 2. Major Differences Between *CHOL* and *FCIM*

ITEM	*CHOL*	*FCIM*
Name	Cholesterol Information Index	Fat & Cholesterol Information Index
Keywords	Cholesterol & Heart disease or Arteriosclerosis	Fat(s), Cholesterol & Heart disease or Arteriosclerosis
Scanning base	Articles from U.S. journals only	Articles from all English journals
Aggregation method	Simple cumulative sum	Weighted sum over a specified period
Period	Quarterly	Monthly

Table 3. The t-ratios of DF and ADF tests[a]

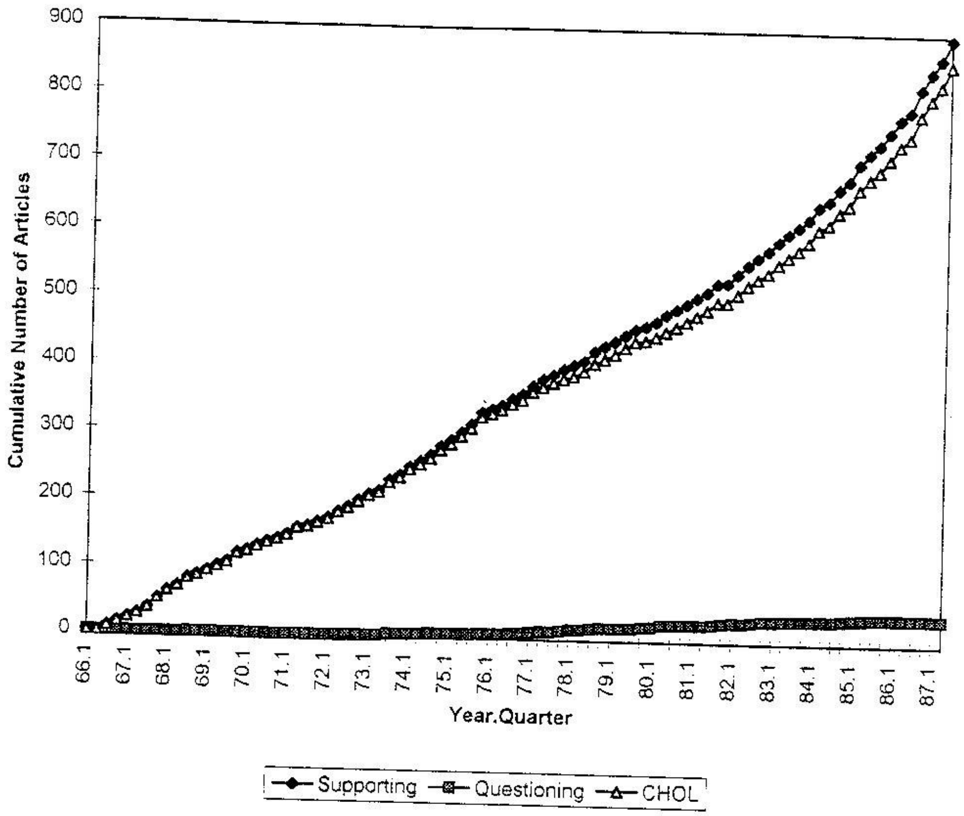
Variable	DF test		ADF test (one lag)	
	Level	Difference	Level	Difference
log r_{1t}	-0.711	-23.86	-0.975	-16.00
log r_{2t}	0.204	-19.02	0.486	-12.87
log $p1_t$	-2.574	-6.744	-1.216	-6.600
log $p2_t$	-2.467	-8.072	-1.439	-7.038
log FI_t	0.413	-13.47	0.455	-7.711
log CI_t	-2.166	-13.43	-2.131	-8.542
log ADI_t	0.807	-5.137	0.384	-5.483
e_{1t}	-6.266		-3.699	
e_{2t}	-8.245		-6.084	

[a]The numbers in this table are estimated t-ratios for level and first-difference; the critical values at the 5% level are -3.0(DF) and -3.6(ADF).

Table 4. Estimated Results of Three Specifications for Whole Milk vs. Other Fresh Milk[a]

Parameter	Whole Milk (I=1)			Other Fresh Milk (I=2)		
	Sp. (1)	Sp. (2)	Sp. (3)	Sp. (1)	Sp. (2)	Sp. (3)
a_i	-8.279	-8.306	-5.443	8.453	8.262	8.422
	(2.526)	(3.307)	(3.955)	(1.367)	(1.604)	(1.644)
b_{i1}	-3.291	-4.755	-0.367	0.727	0.542	0.523
	(2.424)	(2.569)	(4.919)	(1.312)	(1.484)	(2.055)
b_{i2}	6.337	7.661	2.691	-2.256	-2.006	-1.949
	(2.643)	(2.762)	(5.373)	(1.430)	(1.611)	(2.243)
b_{i3}	-0.279	-0.112	-0.997	0.177	0.149	0.559
	(0.267)	(0.187)	(0.692)	(0.144)	(0.131)	(0.292)
c_{i1}	-0.771	-0.586	-0.771	0.478	0.442	0.431
	(0.102)	(0.157)	(0.205)	(0.055)	(0.071)	(0.086)
c_{i2}	0.141	0.072	0.139	-0.043	-0.031	-0.046
	(0.020)	(0.034)	(0.039)	(0.011)	(0.014)	(0.016)
c_{i3}	0.012	-0.064	0.021	0.006	0.021	0.002
	(0.045)	(0.081)	(0.086)	(0.024)	(0.034)	(0.036)
ρ_i		0.606	0.596		0.351	0.354
		(0.057)	(0.056)		(0.068)	(0.068)
d_{i1}			-5.423			-0.321
			(4.719)			(2.192)
d_{i2}			3.274			1.509
			(5.008)			(2.298)
d_{i3}			0.75			-0.346
			(0.569)			(0.219)
d_{i4}			0.859			-0.285
			(0.278)			(0.165)
d_{i5}			-0.167			0.049
			(0.057)			(0.033)
d_{i6}			-0.139			-0.013
			(0.202)			(0.128)
Adj. R^2	0.752	0.844	0.861	0.799	0.823	0.827
DW	0.769	2.345	2.340	1.173	1.926	1.919

[a]The numbers in parentheses for columns with Sp. (1) are estimated standard errors using OLS; the numbers in parentheses for columns with Sps. (2) and (3) are approximated standard errors from nonlinear regressions; DW is Durbin-Watson statistics.



Source: Brown and Schrader (1990).

Figure 1. Cholesterol Information Index (CHOL)

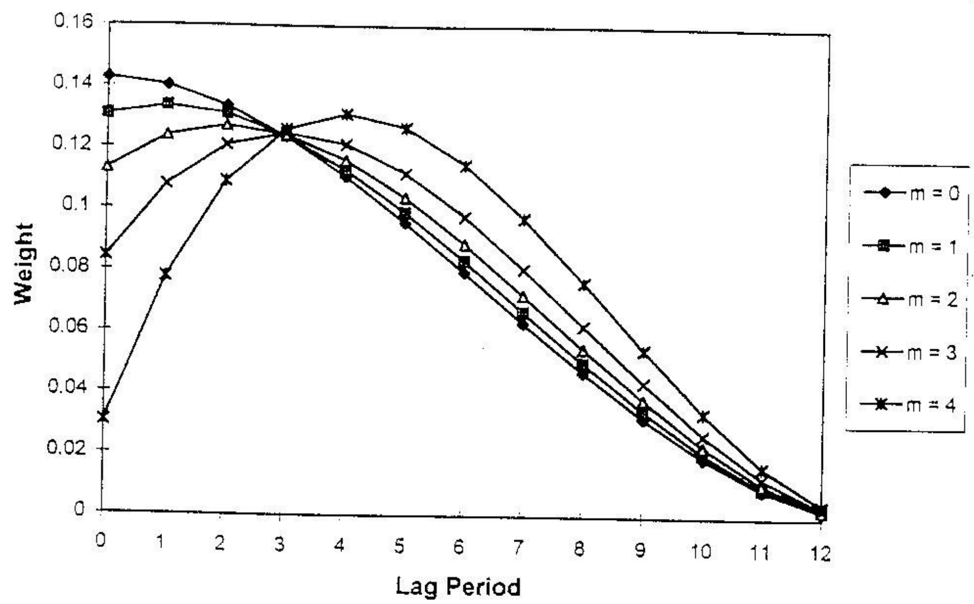


Figure 2. Weights from Cubic Function with n = 12 and Different Peak Time (m)

30

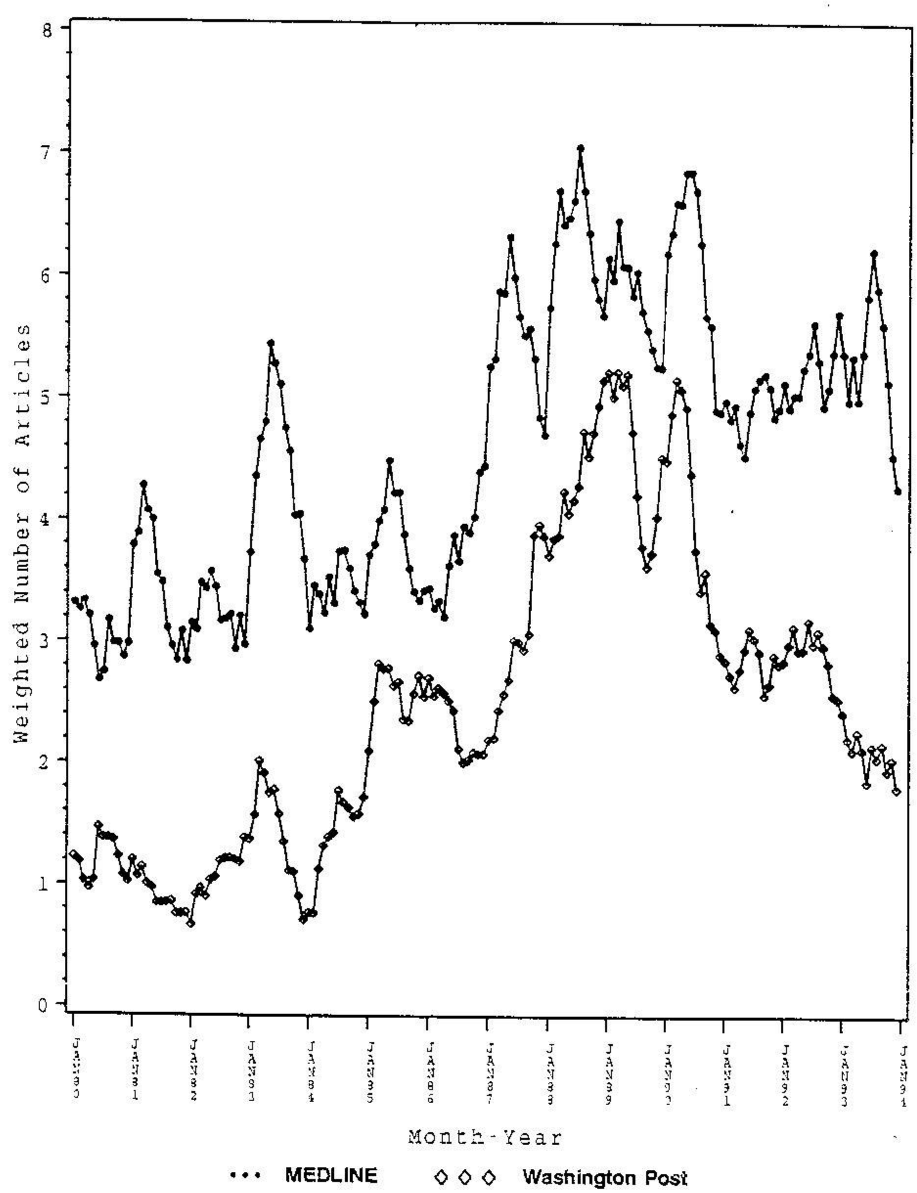


Figure 3. Fat and Cholesterol Information Indexes of FCIM and FCIW for n = 12 and m = 2

31

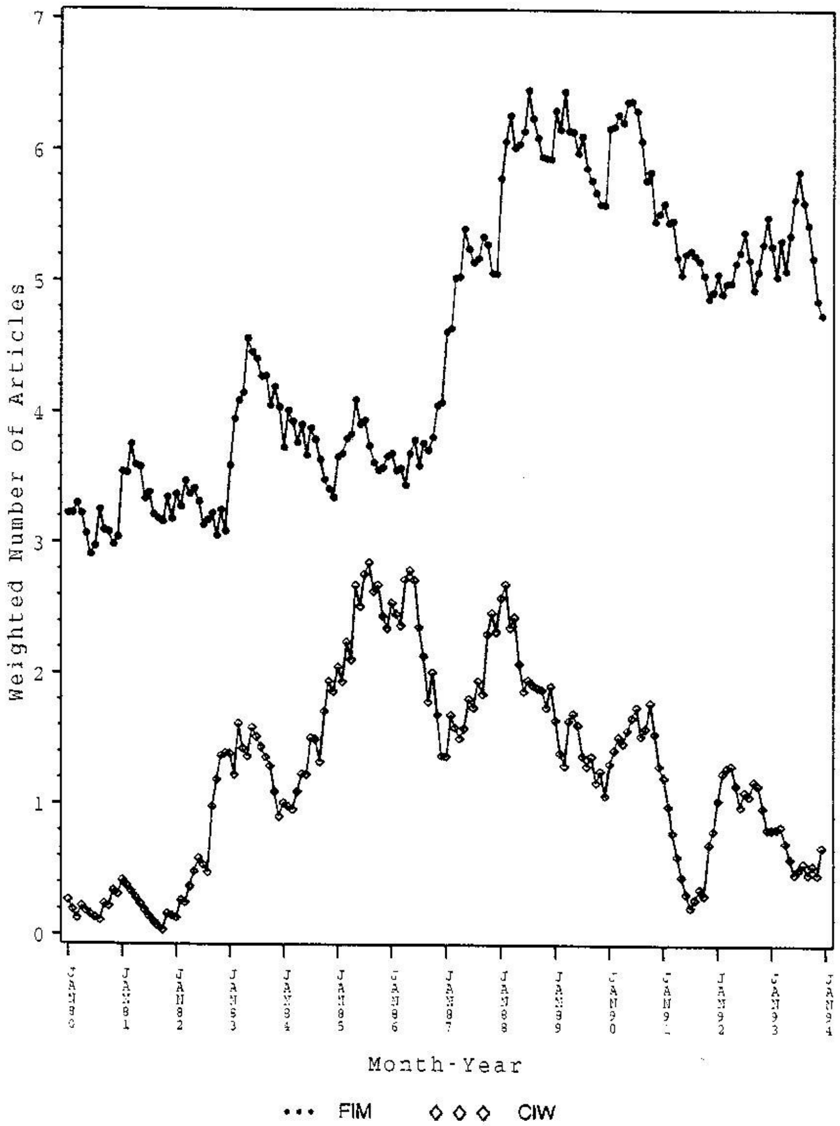
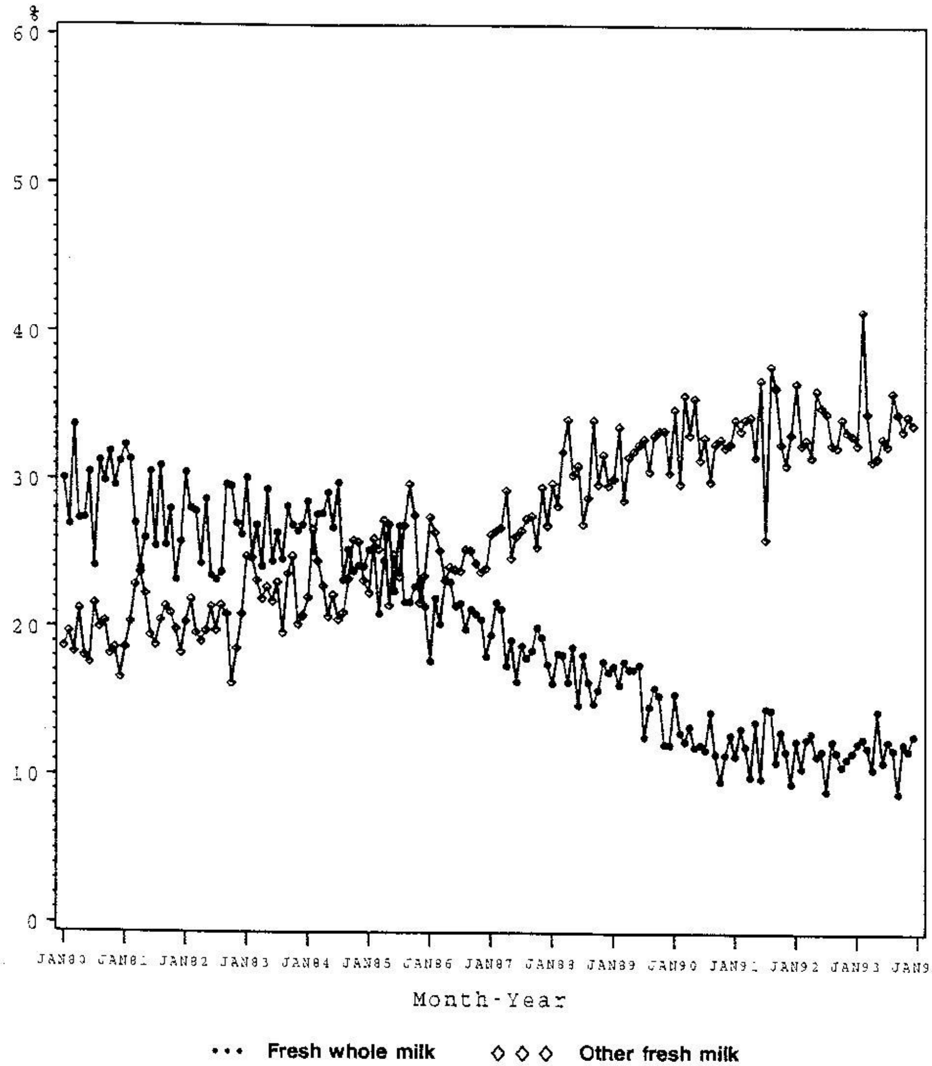


Figure 4. Comparison of Fat and Cholesterol Information Index FCIM (n = 24 and m = 1) and Calcium Information Index CI (n = 12 and m = 0)

32



Note: The weighted monthly data were calculated based on individual household data from 1980-`993 Consumer Expenditure Survey (Diary).

Figure 5. Percentages of Households Purchasing Fresh Whole Milk vs. Other Fresh Milk

33

Appendix A

MEDLINE and LEXIS/NEXIS Databases

MEDLINE, as one of the most widely-used databases in the world, is operated by Dialog Information Services. During 1992, there were over 4.5 million logins at the National Library of Medicine (NLM) alone. Over 7.4 million records from almost 3,700 journals from 1966 to 1994 are contained in the database, in which over 90% of the records were indexed as journal articles, approximately 10% of source documents were review articles, and five percent of the source items were indexed as letters (Pratt, 1994).

The records in the MEDLINE database are indexed in any one of 67 different languages. The top five accounted for 93% of the records in the database. English was first with 80%. Russian, German, Japanese, and French each added another 3-4%. Of the foreign language items, almost half had an English abstract in MEDLINE, and another 20% contained an English abstract in the original journal (Pratt, 1994).

In the MEDLINE database, 44% of the records came from journals published in North America with 97% of these published in the United States, while 47% of the records came from European journals, of which England supplied approximately one-third, with Germany and the Netherlands together publishing another 25%. Journals published in the Far East contributed about 5% of the database and 80% of these came from Japan (Pratt, 1994).

MEDLINE can be accessed in any major institutional libraries and some metropolitan libraries by either CD-ROMs or Internet. Recently MEDLINE was incorporated into the NEXIS-LEXIS databases which can be accessed freely by educational Internet users. For this study, we accessed MEDLINE through NEXIS-LEXIS.

The LEXIS-NEXIS services are premier online legal, news, and business information services in the world operated by a Dayton (Ohio) based operating division of Reed Elsevier, Inc. There are more than 5,620 databases including MEDLINE in the LEXIS and NEXIS services and more than 417 million online documents while more than 1.82 million documents are being added each week.

The LEXIS service contains major archives of federal and state statutes, regulations, case law, and public records, and has 45 specialized libraries covering all major fields of practice, including tax, securities, banking, environmental, energy, and international information.

The NEXIS service is a leading full-text news and business information service. The NEWS library, one of the many libraries in NEXIS, contains more than 2,400 full-text sources which include major news publications such as The Washington Post, New York Time, Los Angeles Times, Business Week, Fortune, and The Economist. In addition, the NEXIS service contains more than 2,000 sources of abstracts including The Wall Street Journal. Since The Washington Post has the longest data series going back to 1977 in NEXIS, it is used in this study.