On Improving Econometric Analyses of
Generic Advertising Impacts (WP99-07) March 1999

by

William G. Tomek and Harry. M. Kaiser

Dr. William G. Tomek, Professor
Department of Agricultural, Resource, and Managerial Economics
255 Warren Hall
Cornell University
Ithaca, NY 14853
Phone: (607) 255-2189
Fax: (607) 255-9984
E-mail: wgt1@cornell.edu

Dr. Harry M. Kaiser, Professor
Department of Agricultural, Resource, and Managerial Economics
349 Warren Hall
Cornell University
Ithaca, NY 14853
Phone: (607) 255-1598
Fax: (607) 254-4335
E-mail: hmk2@cornell.edu

Abstract

It is possible to obtain robust estimates of structural parameters using observational data, but it is difficult to do so. Necessary, but not sufficient, conditions are to adopt a modeling philosophy and to undertake a comprehensive evaluation of the results. Using a general-to-specific modeling philosophy, we obtained stable estimates of the long-run advertising elasticity for fluid milk. This result contrasts with an earlier, published model which did not provide stable estimates as new data points became available. It is difficult, however, to apply the general-to-specific modeling approach because it requires the researcher to specify an initial general model. But analysts are unlikely to agree on this initial model, and if this is true, then the "generality" of the model is in question. Moreover, it is a fact that the quality of the available data is sometimes insufficient to obtain the desired stable estimates.

On Improving Econometric Analyses of Generic Advertising Impacts

William G. Tomek and Harry. M. Kaiser

Can robust estimates of structural parameters be obtained from observational data? This is an important question in empirical econometrics, and is especially apt for research on the effects of, and returns to, advertising generic commodities. The evaluation of advertising programs requires estimates of retail demand functions, with particular attention to the advertising effect, net of the other factors influencing demand. If the retail demand function shifts, then it is also of interest to estimate the effect on the derived (farm-level) demand for, and ultimately on the supply of, the commodity. In sum, a complete structural model is desirable for the full evaluation of generic commodity advertising programs.

The advertising effect may have a distributed lag pattern, and consequently time-series observations are essential to measure the lagged effects. In any case, most data sets that are readily available for analysis are observational time series. Thus, as asked above, can robust estimates of the advertising effect be obtained from observational data?

A specific “yes or no” answer cannot be provided, because of the many problems related to obtaining estimates of structural parameters. The magnitude and nature of these problems vary with the economic sector under analysis. Kinnucan et al. found, for example, that the estimated effects of generic advertising of meats are fragile, although price, expenditure, and cholesterol-information coefficients were relatively robust (pp. 21f).

The objectives of this paper are to review the difficulties of obtaining stable estimates of structural parameters and to discuss an approach to obtaining more robust estimates. To facilitate discussion, we assume that the main focus of research is to obtain robust estimates of the “advertising effect” in retail demand functions. In the process, we provide illustrations using data for the fluid milk market in the United States.

We cannot provide golden rules that will always result in stable estimates, but we argue that there are some necessary conditions for high quality empirical research. These conditions may not be sufficient to guarantee robust estimates of the desired parameters; sometimes results will be fragile, notwithstanding the best efforts of the researcher. Nonetheless, we can strive to contribute to the **cumulative** knowledge about the consequences of advertising generic goods. At a minimum, the researcher should understand the nature of the fragility of the results, and in the conclusions we comment on ways to make results more cumulative.

The first major section outlines the assumptions underlying attempts to estimate structural parameters from observational data. Then, the second section provides suggestions for improving the quality of the empirical results. In a third section, we apply these principles. Finally, some conclusions are drawn.

1.0 Assumptions Underlying Estimating Structure

The problems of estimating one or more structural equations are outlined from a statistical point of view. We ask, what is being assumed, either implicitly or explicitly, in attempts to estimate structure from observational data? If the research problem requires estimation, say, of a demand equation, then quantity per capita is specified to be a function of own-price, the prices of other goods, income or total expenditures, advertising, and perhaps other variables. Data are obtained for a sample time period t = 1, 2,..., T.

To treat the resulting estimates as a structural demand equation, five key assumptions are implicitly made. First, since some variables are excluded from the equation, certain parameters are restricted to being zero. These restrictions are assumed to be correct; presumably no important variables have been omitted. In a demand system, additional restrictions are imposed, such as symmetry and homogeneity. If the restrictions are erroneous, inconsistent estimates of the parameters result, and the degree of identification of the equation is affected.

Second, the relation is usually assumed to be invariant over the sample period. Thus, the parameters of the model are assumed not to change with the passage of time. It is possible, however, to specify models that allow for changes in parameters and to test for changes. But, if the demand structure has changed and it is not appropriately modeled, then a specification error has been committed.

Third, the parameters are also commonly assumed to be structurally invariant. This means that the parameters are constant over the range of the data. A violation of this assumption, in a linear model, means that a parameter's magnitude is a function of the magnitude of a regressor (e.g., a "kink" in the relationship). An appropriate model can accommodate the lack of invariance, but this is not a common specification.

Fourth, structural analysis requires a correct classification of variables as exogenous and endogenous. For our purposes, the endogenous variables are those that we wish to explain, i.e., to model. The exogenous variables are those that are not explicitly modeled in the analysis. The issue of endogeneity can be generalized to include consideration of the importance of variables observed with error. It may be as important, or more important, to take account of errors in variables as to take account of possible simultaneity. Like the other problems, an erroneous classification of variables can seriously bias estimates of the parameters.

A fifth general assumption is that the theory underlying the model is correct. Typically, analysts are using some theory, say $Y_t = f(X_t)$ and do not necessarily consider a competing theory that $Y_t = g(Z_t)$. If the wrong theory is used, estimates of appropriate parameters will not be obtained.

Clearly, models are potentially complex. For example, the demand for milk depends on relative prices and income, but also may be influenced by changes in the age distribution of the population, changes in ethnicity of the population, and changes in perceptions of the healthfulness of milk. Advertising may influence both perceptions about health--milk is good for you--and about taste--milk tastes good. Related modeling issues are the functional form and possible distributed lag effects.

Most of us would not be surprised to find that with the passage of time, changes in income, health awareness, age distribution (say, to an older population), and advertising are correlated. It may be difficult to disentangle the separate effects of these variables. This, in turn, leads to the question of the adequacy of the data available. Do the variables really vary over the sample period? If they do not, precise estimates of their effect cannot be obtained. How collinear are the data? The explanatory variables must have some independent variability in order to obtain precise estimates of their effects.

How well do the available observations represent the underlying economic concept? If the research focus is on the advertising effect, is a time series of advertising expenditures a high quality measure of the underlying concept? Advertising dollars can be spent in different media and on different themes. These themes and media may have differing impacts on consumers. Theme A may have been a high quality educational tool that had a large effect on consumer perceptions and purchases; theme B conducted subsequently may have been much less effective. But, in the aggregate time series, both themes are represented by dollars spent. Hence, the observed regressors are not always a good measure of the concept which we want to measure, and we must be conscious of this potential errors-in-variables problem.

The problem of estimating structural parameters can also be discussed in terms of scientific logic. George Davis summarizes five sets of assumptions, which provide added insights. Briefly, the assumptions are, first, those made in the theoretical framework for the analysis. For instance, in demand analysis, it is commonly assumed that consumers maximize a static utility function subject to a budget constraint. Second, assumptions are made to bridge from abstract theory to empirical implementation. For example, to make a demand system tractable, various aggregation and separability restrictions are required.

A third set of assumptions relates to empirical implementation, including such issues as functional form and how variables are measured. Fourth, the estimator will provide consistent estimates and valid hypothesis tests only if the actual data generating process is the same as the one assumed by the estimator. Fifth, as Davis points out, the range of phenomena under consideration is always restricted in some sense. The theoretical framework, the specific modeling choices made, and so on are not exhaustive.

Thus, it is not logically possible to test all of the assumptions underlying the statistical model fitted. Hypothesis tests are necessarily conditional on some minimal set of assumptions that must be accepted as correct. As Davis states (p.1190), “The claim that there is a valid test for structural change violates the laws of logic.” Put another way, a test for structural change is a joint test of the other conditioning specifications of the model; rejection of the null hypothesis (of no structural change) may merely have identified some other problem in the model.

The foregoing discussion may suggest to some that it is hopeless to estimate a demand structure using observational data. We take a pragmatic view that it is possible, but not easy, to obtain estimates of parameters that are conditional on a specification appropriate to a specific research problem. These estimates should be interpreted and evaluated in terms of the specific objective of the research.

Suggestions for Improving Results

Modeling Process

An important starting point for model specification is a precise problem definition. What is the focus of the research? This determines the specific structural parameters, if any, that must be estimated. In the context of commodity advertising, the research problems are (1) evaluating past advertising programs and (2) making recommendations about future changes in programs and program expenditures. These foci, in turn, suggest the parameters of interest in the research.

One obvious focus is the parameter(s) measuring the advertising effect, and other parameters in the demand equation also may be relevant. Another important parameter is the own-price elasticity of supply of the commodity, which together with the advertising elasticity determine the effect of advertising on the commodity's price. In sum, we proceed assuming certain "focus parameters" exist, which are constants over a specified time period, $t = 1, 2, ..., T$. We cannot verify with certainty, however, that the model specified, using the available data, will result in estimates of the desired parameters. Rather, this is a hoped-for goal.

Research takes place in the context of received theory, the available knowledge about the economic sector under analysis, and past empirical research. Thus, a necessary condition for high quality empirical econometrics is an in-depth understanding of this information. Tomek has argued that achieving this understanding may require the duplication and updating (replication) of key pieces of past research. This can assure that all of the specific components of the previous research are understood as well as provide evidence on the "robustness" of past work. It also helps make research more cumulative by defining the differences between the current work and past analyses.

The next step, we argue, is to adopt a modeling philosophy. Here, our discussion is guided by the general-to-specific modeling methodology suggested by Hendry (Chapter 9), but the general point is to have a logical approach to modeling. For example, Leamer provides another point-of-view, but he too attempts to distinguish good from bad specification searches. A modeling philosophy helps discipline the research methods.

In general-to-specific modeling, the researcher starts by thinking of the complete set of random variables potentially relevant to the economy under investigation. Hendry (p.345) states that this vector of variables "comprises details of every economic action of every agent at time t in all regions of geographical space relevant to the analysis." This set of variables is, of course, not observable, and even if it were, it would be unmanageably large. His statement reminds us, however, that some relevant data may not be available and that the data actually used involve aggregations over time and space. Thus, in using observational data, the researcher needs to understand the data. How are they constructed? What is measured? What is missing? Also, the model specification should be logically consistent with the data; i.e., it must be possible for the chosen specification to have generated the observed data.

The realistic starting point in modeling is the subset of information which is believed relevant to the parameters of interest for this particular research problem. As noted above, theory, information about the economic sector, and prior research are important in this initial selection. The variables considered initially should be sufficiently broad that the researcher's peers think they are adequate. The complete set of variables is defined as X, which is a T x H matrix. Typically, researchers use some subset of the H variables. Thus, $X = [X_1 \; X_2]$, and commonly the initial model specification contains (say) X_1 which is a T x K matrix, where $K < H$.

In using X_1 rather than X, the assumption is that the information contained in X_2 is not essential for estimating the focus parameters in this particular research problem. The point of this distinction is that in practice, X_2 will not have been fully specified by the analyst; it is, at least partly, defined by default. As Hendry (p. 350) points out, however, it is during the foregoing steps that "an investigator's value added enters." It is the researcher's knowledge that contributes to these initial steps of modeling, and it is precisely because these steps are important that the modeling needs to be based on a precise problem statement and on a thorough knowledge of the economic sector under analysis. In Hendry's words (p.350), "Theoretical reasoning is frequently of immense help...but how one discovers useful knowledge remains an art rather than a science." This is why, as noted above, a thorough study of the work of others can contribute not only to synthesis, but to one's preparation for innovation and improved modeling (Ladd).

If some variables in X_2 should have been included in X_1, then as we know, the estimate of the focus parameter is likely biased. Put another way, if relevant variables are omitted, the researcher is not estimating the parameter of interest. For example, if changes in age distribution are affecting demand and if these changes are correlated with changes in advertising intensity, then omitting the age-distribution variable implies a model in which the advertising parameter is capturing two effects. The misspecified model does not contain the parameter of interest, the net effect of advertising. This perspective emphasizes that the model must be specified so that the focus parameter can be estimated.

In the general-to-specific modeling philosophy, economic theory is viewed as providing the long-run equilibrium relationships (Darnell and Evans, p. 78). Hence, given the tentative general model and the associated time-series observations, the modeling process must address a series of specific issues. One is whether or not the regressors are (at least) weakly exogenous. A correct classification of variables as endogenous or exogenous is necessary to obtain consistent estimates of the focus parameter. Weak exogeneity of regressors is implicitly assumed in many demand analyses, but as noted, errors-in-variables and/or simultaneity may be issues that need addressing.

Another issue is whether variables in the model are integrated. The literature suggests that many economic variables may be integrated, i.e., not stationary. It is not clear whether this is a major problem for analyses of commodity demand, but unquestionably economic data have trends. Thus, it is possible that using the levels of variables, which are integrated, would result in finding nonsense relationships, and researchers should be concerned about discriminating between true and spurious relationships in the data set. Analysts probably should explore the need to use differenced observations, perhaps as part of equilibrium error-correction specifications.

The general-to-specific modeling literature usually takes the point of view that autoregressive distributed lag (ADL) specifications are appropriate. One notation is to write $A(L)y_t = B(L)x_t + e_t$ where A(L) and B(L) are polynomial lag operators, say of order s, y_t is endogenous, and x_t is weakly exogenous. For a system of equations, the notation is generalized to think in terms of matrices and vectors. The error correction model can be viewed as ADL(1,1).

In general-to-specific modeling, one recommendation is to assume lag lengths longer than logic suggests and to test down to a more parsimonious specification. Hendry calls this process "lag truncation." Similarly, starting with a general model, restrictions on the form of the lag and on the end-points can be tested. The specification of lag relationships is likely to be very important in estimating advertising effects, and consequently whether or not the researcher is using the general-to-specific methodology, the specification of the lag structure requires explicit attention.

Functional form is still another issue in model specification, but it is closely connected with other specification issues. For example, a seeming "outlier" in a data set may reflect an omitted variable, an erroneous functional form, or an actual random error. The functional form should be consistent with the data. If the dependent variable cannot be negative, then the functional form should not permit negative forecasts of the variable. (This point is especially important in research problems where the dependent variable has clear limits, like zero and one.) We will have a little more to say about addressing functional form in the context of model evaluation.

In sum, a modeling process, like the one described above, is expected to help obtain a stable estimate of the focus parameter. The researcher starts with a general specification. The process should lead to a simpler specification, but a specification which has "parameter constancy." To be useful, this constancy should extent beyond the sample period, so that the fitted model is useful for simulations and forecasting. For analyzing advertising effectiveness, the researcher wants a robust estimate of the effect of advertising within the sample period, but also for forecasting the consequences of possible changes in advertising levels in the future. This point of view leads naturally to the topic of model evaluation.

Explicit Model Design and Evaluation

More and more, the literature on empirical econometric practice criticizes procedures that use diagnostic tests as selection criteria, which is sometimes called the simple-to-general approach of modeling. For example, a model is proposed, and a test for autocorrelated errors is conducted. If the null of zero autocorrelation is rejected, the symptom is "fixed," perhaps by using a GLS estimator or possibly by adding a variable to the model. In either case, the original model has been modified in light of the test. This is a type of data mining, i.e., pretest estimation. Consequently, the probability of type I error of subsequent tests is increasing, and it becomes meaningless to evaluate the “final” model by the same criteria that were used to select it in the first place.

In the general-to-specific methodology, Hendry (p. 361) argues that the researcher should start with “the most general, estimable, statistical model that can reasonably be postulated initially, given the present sample of data, previous empirical and theoretical research, and any institutional and measurement information available.” The general model is formulated to contain “the parsimonious, interpretable, and invariant econometric model at which it is hoped that the modelling exercise will end” (Hendry, p. 361). Thus, the general unrestricted model should be consistent with all of the pre-existing evidence. Explicit model design is intended to obtain this simpler specification from the more general specification.

At this stage, the model may contain variables with considerable collinearity, but this is not necessarily a problem unless it misleads the subsequent modeling efforts. Given the general model, the researcher should have a logical, consistent plan for simplification. The process should not be an ad hoc examination of t-ratios and signs of coefficients. Rather, the expectation is that initial, general specification will permit valid tests of logical restrictions on the model. Namely, the general specification has “assured” that the model consistent with the focus of the research is embedded within the larger model, that the variables in the model are stationary, that the functional form is consistent with the data, that a sufficient number of lagged variables are included; etc. Thus, the residuals of the equation(s) will meet the classical assumptions for hypothesis tests, and in this context, it is possible to test restrictions on the model.

Developing the series of logical tests to simplify the model is perhaps the most difficult or “artful” part of the exercise. A first step is to recognize the special cases imbedded in the general model. Then, it is possible to ask, which of the special cases are plausible for this research problem? Do the restrictions make economic sense? An ADL(1,1) model, for example, contains 10 special cases (Hendry, Table 7.1).

One example of a simplification is where the general model contains “m” lags for the advertising effect, where “m” is chosen to almost surely exceed the actual lag length, “n.” A series of tests can be conducted to see if a lag length shorter than “m” is adequate. Likewise, it would be possible to impose restrictions on the form of the lagged effect, like a polynomial structure, and test whether this is an adequate simplification. To reemphasize, the initial, general specification should permit the use of classical procedures like t, F, Wald, Likelihood Ratio, and Lagrange Multiplier tests (Charemza and Deadman, Chapter 4).

Having obtained a simpler model, it can be evaluated via a set of criteria. These criteria include (1) that the model is consistent with the data (and that the data are accurate). Hendry calls this “data admissible formulations.” (2) The model also should be consistent with theory and be identified. The foregoing criteria are often taken as “givens” by researchers, but as stressed throughout this paper, should be criteria used in modeling. A comprehensive battery of tests can be used to check many aspects of model adequacy (e.g., see McGuirk, Driscoll, and Alwang; also McGuirk, et al.). Thus, (3) these tests should check that the conditioning variables for the parameters of interest are weakly exogenous. (4) The focus parameters should be constants over the sample period (and beyond for forecasting purposes) and invariant to changes in the regressors. (5) Hendry further stresses that the "final" model should encompass rival models. A discussion of encompassing would require too much space, but encompassing basically addresses the question, "Can the reduced model explain the results of the general model from which the reduction was made? (Hendry, p. 365)" The simpler model should not have lost relevant information relative to the more general model.

For the analysis of advertising effects, parameter constancy is a key issue. It is, of course, possible that for a specific research problem, a structural break has occurred, but this possibility should be determined, in our view, by logic and not by empirical data mining. In other words, if a structural change has occurred, it should be explicitly modeled. Otherwise, the research should proceed under the assumption that it is possible to find a constant parameter for the sample period. If the estimates of the focus parameter(s) are not robust, this should be treated as a problem of specification error rather than as a structural change. Thus, we take the viewpoint that in model evaluation, stability of the estimates of the key parameters is an essential criterion.

An Illustration

In this section, we illustrate some of the principles discussed in the prior section. The illustration is based on a quarterly data set used to estimate the effect of advertising on fluid milk demand. The full data set starts in 1975.1; variable definitions are provided in the appendix.

We first duplicate the results of an earlier study, where the initial observation on the dependent variable starts in 1976.1 and ends in 1990.4 (Kaiser et al.). (Four data points are lost in lagging.) Then, the model is reestimated with revised data. Next, the results are updated by adding more recent observations, and the estimated advertising effect is not robust.

Hence, we explore whether a researcher could have built a model using data ending in 1990.4 that remains robust through 1997.4. A general-to-specific approach and associated evaluation methods are used. Given the time constraints in writing this paper and since not all of the potentially relevant data were immediately available, the analysis should be viewed as illustrating benefits and problems of a modeling philosophy, but not as the definitive way to implement the philosophy. The results suggest that it is sometimes possible to improve the "robustness" of results.

Duplication and Replication

The initial model made the per capita consumption of fluid milk a function of real price, real income, trend, a seasonal effect, and an advertising effect (Kaiser et al.). The advertising effect was modeled as a second degree polynomial, with a four quarter lag, and with end-point constraints. The variables are transformed to logarithms, and the equation in the original publication was fitted by an instrumental variables (IV) estimator. Selected results from the original fit are provided in column (1) of Table 1; we were able to exactly duplicate this result from historical data files. Then, we provide the equivalent OLS estimates, and subsequent comparisons use the OLS estimator. The long-run advertising elasticity (the sum of the current and lagged effects) is positive with a large t-ratio for both the IV and OLS estimates. Other results appear logical, and R^2 is large. The Durbin-Watson statistic suggests, however, that the model may not be adequate.

The identical model was refitted to revised data for the time period 1976.1 to 1990.4 (Table 1, column 3). These results are similar to those from the original data. But, while the absolute difference in the point estimates of the advertising elasticity for the two samples is small, the estimate using the revised data is 18 percent smaller than the original value. Subsequent updating of results provided relatively stable estimates of the long-run advertising elasticity through the next several years. A representative result is shown in column (4) of Table 1. During this period, using the original model, one could conclude that the elasticity for advertising fluid milk was about .028.

After 1992.4, however, the results start to deteriorate. The overall fit of the equation decreases; the Durbin-Watson statistic becomes smaller; most important, the magnitude of the advertising elasticity decreases, as does its t-ratio. By 1996, the estimated advertising elasticity is zero (Table 1, column 5). One possibility is that the advertising of fluid milk is indeed becoming less effective, but another possibility is that the original model is misspecified. Thus, we explore whether an alternative model has more stable estimates of the long-run advertising effect.

Model Specification

As already mentioned, the demand for fluid milk in the United States is probably influenced by a variety of factors beyond relative prices and income. The age distribution of the population is changing, as is the ethnic composition of the population. It is also possible that opinions about the healthfulness of milk are changing (either positively or negatively). With respect to price effects, one hypothesis is that price of breakfast cereals is especially important in milk demand; cereals and milk presumably are complementary goods; and increases in the price of cereals are perhaps an important factor in reducing the demand for milk. Most demand models, in contrast, have modeled the prices of substitutes (beverages). In terms of advertising, various programs have been used, and a national-wide program was started in 1985. The model specification should perhaps take account of differences in advertising programs over the sample period. Moreover, in some applications, one might distinguish between the effects of advertising brands and advertising the generic good, milk.

The modeling exercise uses the data set for the period 1975.1 through 1990.4. Because of the lagging, the initial observation for the dependent variable is 1976.4 (which is slightly different than the earlier results, Table 1). The "general" specification, which we use as a starting point, is a compromise. In a more in-depth research process, it would have been preferable to start with a truly general model, but we do test for possible omission of variables. The compromise specification is as follows. First, all variables, excepting the zero-one dummies, are transformed to logarithms, and the variables differenced. Thus, the variables represent percentage changes, and the transformed variables are assumed to be stationary. (Stationarity tests were not conducted, and an in-depth analysis would have done so.)

The differencing introduces an element of dynamics into the specification. In addition, the dynamics in demand are modeled in two ways. One and two quarter lagged values of the dependent variable (per capita consumption of milk, in logarithms and differenced) are included as explanatory variables. (Clearly, other specifications of lags would have been possible, including the ADL(1,1) discussed in the prior section.) In addition, advertising is represented by current and lagged values through six quarters; an unrestricted lag structure is used. Advertising is measured by expenditures for generic fluid milk, deflated by a media price index.

The model specification includes the real price of milk (nominal price deflated by an index of beverage prices) and real disposable per capita income (deflated by the general CPI). A possible "National Dairy Board effect" is specified by defining a dummy variable equal zero for the quarters through 1984.4 and equal to one for the quarters 1985.1 and thereafter. The hypothesis is that the introduction of national program would increase demand. The model also contains three dummy variables for seasons. Given that the variables are in first differences of logarithms, the intercept terms are capturing possible trends: the percentage change in per capita consumption if the other regressors were not changing. Variable definitions and the estimated equations are provided in appendix tables.

Turning to the initial results, the generic advertising effect for the unrestricted model fitted to the 1976 - 1990 sample has roughly a humped shape (Table 2, column 1). The coefficients for the intermediate lags tend to be larger, but are variable. Some t-ratios are large; others are small (appendix Table 2). The sum of the advertising effect is 0.0328, which is very similar to the estimate obtained in the original model (with a second degree polynomial, four quarter lag length, and end-point constraints).

Before exploring the robustness of this result, we note the following: the own-price variable has a negative coefficient but with a t-ratio of -0.4; income is positive with a t-ratio of about 1.5; and the national advertising effect is positive with a t-ratio of 2.66 (appendix Table 2). At this stage of general-to-specific modeling, however, the researcher is less interested in specific magnitudes and signs of coefficients and t-ratios and is more interested in the "quality" of the model and its possible simplifications.

In Hendry's discussion, the general model is simplified and then tested for adequacy. In our case, where the general model is not very general, we test for its adequacy and then consider one simplification. Thus, we subjected the initial model to a series of specification tests. The tests are an LM test for autocorrelated errors, a White test and an ARCH test for heteroscedasticity, several variants of RESET, and a test for whether own-price is endogenous. The autocorrelation tests include both one and two period lags; the ARCH tests included up to three lags of the squared residuals; the White test uses the current and squared values of the regressors. One RESET used the squared and cubed values of the computed dependent variable as regressors in the auxiliary regression; another RESET addressed whether omitted variables were statistically important.

The variables used in this test included three which measure age distribution (proportions of population age 5 and below, age 6 to 15, and age 16 to 19), the real price of breakfast cereal, and seven which measure brand advertising (the current and six lags of real expenditures).

Results are summarized in Table 3; test procedures are summarized in Godfrey (chapter 4); see also MacKinnon. The tentative general model “passed” all of the tests. The evidence suggests that price can be treated as weakly exogenous, that the errors are homoscedastic and not autocorrelated, and that there is no evidence of omitted variables. The latter is particularly surprising, because as noted below, logical reasons still exist to question this model specification.

Nonetheless, from a statistical viewpoint, the “general” model appears adequate. Thus, the next logical step is to see if the model can be simplified. Our principal simplification was to test for restrictions on the lag structure for generic advertising. We tested one set of restrictions; namely the lag is restricted to a second degree polynomial with end-point constraints. The six quarter lag length was retained. The “humped shaped” lag structure appeared to be roughly justified by the unrestricted results, and as it turns out, the restrictions cannot be rejected. The sum of the lags for the model fitted through 1990.4 is 0.029, with a t-ratio of 1.698.

The own-price coefficient, however, has a t-ratio of only -0.36, and the real income coefficient has a modest t-ratio of 1.448. The statistically important regressors are lagged consumption, the intercept and seasonal dummies, and the dummy representing the National Dairy Board effect. The positive effect of the National Dairy Board is logical and helpful, but otherwise the results are not very satisfying from the viewpoint of an economic explanation of changes in the consumption of fluid milk. Our analysis did not consider the effects of the advertising of competing beverages, such as colas, nor possible health concerns.

In any case, we next examined the stability of the coefficient estimates over a longer sample, and the results are stable through the sample that ends in 1997.4. The total effect (elasticity) of advertising is .029 with a t-ratio of 1.721, which is almost identical to the result for the sample ending in 1990.4. Indeed, the other results remain remarkably similar. The price coefficient is -.018 for the long sample versus -.013 for the short sample (both with small t-ratios); the income coefficients are .168 and .199 respectively. In sum, if the major objective is to obtain stable estimates of the advertising effect, this was accomplished, at least for the sample data 1975 to 1997. The advertising elasticity (.029) turned out to be almost identical to the estimate obtained for the old model, which fell apart when used with recent data.

Postmortem

The results reported in this paper support the hypothesis that advertising the generic commodity milk can increase its demand, other factors held constant. But, the per capita consumption of milk has been decreasing in recent years, and the “final” model, reported here, does not provide a satisfactory explanation about why this is happening. The intercept and dummy coefficients suggest negative trends in the changes in consumption from the second to the third (-3.7 percent) and from the third to the fourth (-2.5 percent) quarters, but a positive trend in the first quarter (3.0 percent). In sum, the model accounts for trends in a statistical sense, and selected variables, such as age distribution and price of breakfast cereals, were examined for their possible effects on demand. But, the “final” model does not provide a fundamental explanation, in terms of economic variables, for the net decrease in consumption. Until this is accomplished, most of us are not going to be fully satisfied with results such as ours, no matter how many tests of adequacy have been “passed.”

Summary and Critique

Can robust estimates of structural parameters be obtained from observational data? This paper suggests that the answer can be yes, but that it is difficult. One approach is a general-to-specific modeling philosophy, but this philosophy is not easy to implement. The demand for a commodity like fluid milk is potentially influenced by a large array of variables. Moreover, some of these variables are not easily measurable, such as possible health concerns. Thus, two or more economists, working separately, probably would not specify the same initial general model. If this is so, then it is unclear how “general” the initial specification is. We tried to make our choices clear, and in retrospect, with the time to do more research, we ourselves would have done some things differently.

Also, statistical tests of model adequacy are conditional in nature. The tests used in this paper suggested that our “general” model was statistically adequate, but other tests might have found problems. For example, we could have explored more thoroughly the combining of tests as in McGuirk, et al. On a positive note, stable estimates of the advertising elasticity were obtained over a longer sample period than had been the case with an earlier model specification. Moreover, the estimate for the recent data and model was similar to earlier results. Thus, while our results are not definitive, they contribute to the accumulation of knowledge about the effects of advertising. Our work also implies that a systematic approach to modeling may be helpful in achieving robust estimates.

The following question is implied by our discussion: if two analysts had faced the same problem with the same data set, would they have arrived at the same estimate of an advertising elasticity for fluid milk? We have suggested that one criterion for a general model is that peers view it as general, but we also think that it is difficult to obtain such general agreement. Different analysts faced with the same set of modeling issues take different approaches to them. This raises a fundamental problem for the general-to-specific modeling philosophy.

A possible approach to this lack of agreement is to have two or more teams work on precisely the same research problem. This protocol would add costs, but if the research problem involves an analysis which will influence major decisions, involving millions of dollars, then it is a justifiable strategy. In some branches of science, more than one laboratory is working on the same research problem. Perhaps more of this should occur in applied economics.

In the sciences, prizes go to clear winners. In empirical economics, it may be difficult to determine the winner, but differences in results among competing teams could stimulate thought and be the foundation for the next round of research. The cumulative effect should be beneficial.

If the foregoing proposal is not operable, then diverse estimates might be appraised through a type of "meta" analysis. At a minimum, a thorough review of the research could assist in understanding the reasons for the diverse (or similar) results that exist in the literature. In sum, we need to search for creative ways to make research more cumulative and thereby improve knowledge about factors influencing the demand for generic goods.

References

Charemza, W.W., and Deadman, D.F. (1992). New Directions in Econometric Practice. Edward Elgar.

Darnell, A.C., and Evans, J.L. (1990). The Limits of Econometrics. Edward Elgar.

Davis, G.C. (1997). The Logic of Testing Structural Change in Meat Demand: A Methodological Analysis and Appraisal. Amer. J. Agr. Econ. 79:1186-1192 (November).

Godfrey, L.G. (1988). Misspecification Tests in Econometrics. Cambridge University Press.

Hendry, D.F. (1995). Dynamic Econometrics. Oxford University Press.

Kaiser, H.M., Forker, O.D., Lenz, J., and Sun,Chin-Hwa. (1994). Evaluating Generic Dairy Advertising Impacts on Retail, Wholesale, and Farm Milk Markets. J. of Agr. Econ. Res. 44(4): 3-17.

Kinnucan, H.W., et al. (1997). Effects of Health Information and Generic Advertising on U.S. Meat Demand. Amer. J. Agr. Econ.79:13-23 (February).

Ladd, G.W. (1979). Artistic Research Tools for Scientific Minds. Amer. J. Agr. Econ. 61: 1-11 (February).

Leamer, E.E. (1978). Specification Searches: Ad Hoc Inference with Nonexperimental Data. John Wiley.

MacKinnon, J.G. (1992). Model Specification Tests and Artificial Regressions. J. Econ. Lit. 30: 102-146 (March).

McGuirk, A.M., Driscoll, P., and Alwang, J. (1993). Misspecification Testing: A Comprehensive Approach. Amer. J. Agr. Econ. 75:1044-1055 (November).

_______, ______, ______, and Huang, H. (1995). System Misspecification Testing and Structural Change in the Demand for Meat. J. Agr. & Res. Econ. 20: 1-21.

Tomek, W.G. (1993). Confirmation and Replication in Empirical Econometrics: A Step Toward Improved Scholarship. Amer. J. Agr. Econ. 75: 6-14 (October-anniversary issue).

Table 1. Coefficients of Fluid Milk Demand, Original Model, Selected Time Periods

Variables[a]	Original 1976.1 - 19 90.4 IV	Original 1976.1 - 19 90.4 OLS	Revised 1976.1 - 1990.4	1976.1 - 1992.4	1976.1 - 1996.4
Real price	-.036	-.048	-.039	-.037	-.047
	(-2.20)	(-2.98)[b]	(-2.44)	(-2.22)	(-1.76)
Real income	.252	.213	.257	.252	.236
	(6.56)	(6.77)	(7.59)	(7.15)	(3.97)
Trend	-.067	-.078	-.079	-.078	-.083
	(-13.13)	(-12.19)	(-12.75)	(-11.50)	(-7.32)
Advertising	.026	.034	.028	.029	-.003
	(8.13)	(8.06)	(6.11)	(5.95)	(-0.39)
R^2	--	.938	.938	.921	.781
D-W	1.46	1.483	1.435	1.433	0.424

[a] Per capita consumption of fluid milk dependent. All variables in natural logarithms. Intercept and seasonal variables omitted from table.

[b] t-ratios in parentheses.

Table 2. Advertising Coefficients, Fluid Milk Demand, Revised Models

	1976.4 - 1990.4[a/]		1976.4 - 1997.4	
Lag	Unrestricted	Restricted	Unrestricted	Restricted
t	.00216	.00239	.00207	.00243
t - 1	.00935	.00409	.00762	.00416
t - 2	.00314	.00511	.00100	.00520
t - 3	.00197	.00545	.00378	.00555
t - 4	.00607	.00511	.00723	.00520
t - 5	.00981	.00409	.00930	.00416
t - 6	.00031	.00239	-.00306	.00243
Sum	.03281	.02863	.02795	.02914

[a/] Identical Models fitted to two time periods. The restrictions are a second-degree polynomial lag form with end-point constraints. The restrictions cannot be rejected.

Table 3. Selected Evaluation Tests

Test	F
LM test, autocorrelation, two lags	0.558
ARCH test, heteroscedasticity, three lags	0.315
White heteroscedasticity test[a/]	0.319
RESET, cubic[b/]	0.327
RESET, omitted variables[c/]	0.804
Hausman test for endogeneity of price[d/]	0.053

[a/] Based on linear and squared regressors.
[b/] Using squared and cubed values of estimated dependent variable.
[C/] Test included 11 variables, see text.
[d/] See Godfrey, p. 149.

Appendix Table 1. Variable Definitions for the econometric model

Q1POP = per capita fluid milk demand measured in bil. Lbs. Of milkfat equivalent divided by U.S. population in mil.,
RFP = consumer retail price index for fresh milk and cream (1982-84 = 100) divided by the consumer retail price index for nonalcoholic beverages (1982-84 = 100),
INCOME = disposable personal income per capita, measured in thousand $,
TREND = time trend variable, equal to 1 for 1975, 1.....,
DUMQ1 = intercept dummy variable for the first quarter of year,
DUMQ2 = intercept dummy variable for second quarter of year,
DUMQ3 = intercept dummy variable for third quarter of year,
DUMNDB = intercept dummy variable for creation of National Dairy Board, equal to 0 for 1975.1-1984.3, and 1 otherwise,
DGFAD = brand fluid milk advertising expenditures deflated by the media price index, measured in thousand $,
A5 = percent of U.S. population 5 years or old or younger,
A615 = percent of U.S. population between 6 and 15 years of age,
A1619 = percent of U.S. population between 16 and 19 years of age,
CERCPI = consumer price index for cereal

Appendix Table 2.

LS// Dependent Variable is DQ1POP
Date: 03/01/99 Time 14:07
Sample (adjusted): 1976:4 1990:4
Included Observations: 57 after adjusting endpoints

Variable	Coefficient	Std. Error	t- Statistic	Prob.
C	-0.029073	0.009804	-2.965475	0.0050
DQ1POP(-1)	-0.523262	0.155144	-3.372751	0.0016
DQ1POP(-2)	-0.347926	0.151356	-2.298727	0.0267
DRFP	-0.015136	0.036746	-0.411922	0.6825
DINCOME	0.210724	0.140567	1.499105	0.1415
DUMQ1	0.065154	0.008818	7.388961	0.0000
DUMQ2	0.037212	0.018068	2.059611	0.0458
DUMQ3	-0.012348	0.014116	-0.874737	0.3868
DUMNDB	0.007507	0.002818	2.664324	0.0110
DGFAD	0.002157	0.003463	0.622936	0.5368
DGFAD(-1)	0.009351	0.003869	2.416732	0.0202
DGFAD(-2)	0.003137	0.004234	0.740982	0.4629
DGFAD(-3)	0.001968	0.003839	0.512656	0.6109
DGFAD(-4)	0.006074	0.003862	1.572753	0.1235
DGFAD(-5)	0.009807	0.004068	2.410756	0.0205
DGFAD(-6)	0.000313	0.003727	0.084082	0.9334

R-squared	0.960140	Mean dependent var	0.000369
Adjusted R-squared	0.945556	S.D. dependent var	0.039537
S.E. of regression	0.009225	Akaike info criterion	-9.139685
Sum squared resid	0.003489	Schwarz criterion	-8.566197
Log likelihood	195.6015	F-statistic	65.83928
Durbin-Watson stat	1.940867	Prob (F-statistic)	0.000000

Appendix Table 3.

LS// Dependent Variable is DQ1POP
Date: 03/01/99 Time 14:08
Sample (adjusted): 1976:4 1990:4
Included Observations: 57 after adjusting endpoints

Variable	Coefficient	Std. Error	t- Statistic	Prob.
C	-0.034786	0.008429	-4.126862	0.0001
DQ1POP(-1)	-0.547370	0.140849	-3.886217	0.0003
DQ1POP(-2)	-0.438681	0.140927	-3.112817	0.0032
DRFP	-0.013045	0.036187	-0.360485	0.7201
DINCOME	0.199267	0.137658	1.447547	0.1544
DUMQ1	0.065553	0.007616	8.606744	0.0000
DUMQ2	0.046170	0.015655	2.949277	0.0050
DUMQ3	-0.000399	0.012744	-0.031321	0.9751
DUMNDB	0.008292	0.002828	2.931768	0.0052
PDL01	0.002727	0.001606	1.697518	0.0962

R-squared	0.952125	Mean dependent var	0.000369
Adjusted R-squared	0.942958	S.D. dependent var	0.039537
S.E. of regression	0.009443	Akaike info criterion	-9.167006
Sum squared resid	0.004191	Schwarz criterion	-8.808576
Log likelihood	190.3802	F-statistic	103.8587
Durbin-Watson stat	1.971690	Prob (F-statistic)	0.000000

Lag Distribution of DGFAD	i	Coefficient	Std. Error	T-Statistic
>	0	0.00239	0.00000	0.00000
	1	0.00409	0.00000	0.00000
	2	0.00511	0.00000	0.00000
	3	0. 00545	0.00000	0.00000
	4	0.00511	0.00000	0.00000
	5	0.00409	0.00000	0.00000
	6	0.00239	0.00000	0.00000

	Sum of Lags	0.02863	0.00000	0.00000

Appendix Table 4.

LS// Dependent Variable is DQ1POP Date: 03/01/99 Time 14:07 Sample (adjusted): 1976:4 1997:4 Included Observations: 85 after adjusting endpoints				
Variable	Coefficient	Std. Error	t- Statistic	Prob.
C	-0.018521	0.007297	-2.538242	0.0134
DQ1POP(-1)	-0.407875	0.113961	-3.579071	0.0006
DQ1POP(-2)	-0.354584	0.108649	-3.263569	0.0017
DRFP	-0.024002	0.034449	-0.696734	0.4883
DINCOME	0.173323	0.132333	1.309744	0.1946
DUMQ1	0.050480	0.007431	6.793214	0.0000
DUMQ2	0.019820	0.013107	1.512179	0.1351
DUMQ3	-0.020417	0.009810	-2.081226	0.0411
DUMNDB	0.002322	0.002618	0.887166	0.3781
DGFAD	0.002074	0.003451	0.600820	0.5499
DGFAD(-1)	0.007617	0.003879	1.963712	0.0536
DGFAD(-2)	0.001005	0.003990	0.251881	0.8019
DGFAD(-3)	0.003777	0.003821	0.988428	0.3264
DGFAD(-4)	0.007234	0.003833	1.887246	0.0633
DGFAD(-5)	0.009303	0.003927	2.368893	0.0206
DGFAD(-6)	-0.003056	0.003643	-0.838893	0.4044

R-squared	0.933515	Mean dependent var	-0.001155
Adjusted R-squared	0.919062	S.D. dependent var	0.037604
S.E. of regression	0.010698	Akaike info criterion	-8.907408
Sum squared resid	0.007897	Schwarz criterion	-8.447615
Log likelihood	273.9551	F-statistic	64.58864
Durbin-Watson stat	1.989866	Prob (F-statistic)	0.000000

Appendix Table 5.

LS// Dependent Variable is DQ1POP
Date: 03/01/99 Time 14:09
Sample (adjusted): 1976:4 1997:4
Included Observations: 85 after adjusting endpoints

Variable	Coefficient	Std. Error	t- Statistic	Prob.
C	-0.025250	0.006686	-3.776340	0.0003
DQ1POP(-1)	-0.471881	0.108033	-4.367924	0.0000
DQ1POP(-2)	-0.405993	0.106391	-3.816049	0.0003
DRFP	-0.018306	0.034795	-0.526120	0.6004
DINCOME	0.168201	0.134635	1.249305	0.2154
DUMQ1	0.055113	0.006717	8.205265	0.0000
DUMQ2	0.032293	0.011887	2.716810	0.0082
DUMQ3	-0.011695	0.009345	-1.251451	0.2147
DUMNDB	0.002568	0.002705	0.949308	0.3455
PDL01	0.002727	0.001606	1.697518	0.0962

R-squared	0.922334	Mean dependent var	-0.001155
Adjusted R-squared	0.913014	S.D. dependent var	0.037604
S.E. of regression	0.011091	Akaike info criterion	-8.893145
Sum squared resid	0.009225	Schwarz criterion	-8.605775
Log likelihood	267.3489	F-statistic	98.96414
Durbin-Watson stat	2.008382	Prob (F-statistic)	0.000000

Lag Distribution of DGFAD	i	Coefficient	Std. Error	T-Statistic
	0	0.00243	0.00141	1.72109
	1	0.00416	0.00242	1.72109
	2	0.00520	0.00302	1.72109
	3	0. 00555	0.00323	1.72109

4	0.00520	0.00302	1.72109		
5	0.00416	0.00242	1.72109		
6	0.00243	0.00141	1.72109		
		Sum of Lags	0.02914	0.01693	1.72109