MULTIPLICATIVE MODELS WITH ZERO VALUES IN THE EXPLANATORY VARIABLES

Philippe A. NAERT* and Marcel WEVERBERGH**

1. *Introduction*

Consider the model,

$$q_i = \alpha \, x_{i1}^{\beta_1} \, x_{i2}^{\beta_2} \ldots x_{ip}^{\beta_p} \ldots x_{iP}^{\beta_P} \, e^{\varepsilon_i} \qquad (1)$$

where q_i $(i = 1,2,\ldots n)$ represents the i^{th} observation on the dependent variable, x_{ip} the i^{th} observation on the p^{th} independent or explanatory variable $(p = 1,2,\ldots,P)$, assumed nonstochastic, ε_i is the i^{th} value of the stochastic disturbance term, and α, $\beta_1,\ldots,\beta_P$ are the parameters. The observations q_i and x_{ip} are assumed to be nonnegative for all i and p . This model is generally referred to as the multiplicative model, and is extensively used in applied regression analysis. There are two main reasons for its popularity: First, explanatory variables are often assumed to interact and the multiplicative specification is one way of modeling such interaction without loss of degrees of freedom. Secondly, the model has the advantage of becoming linear upon applying a logarithmic transformation, that is,

$$\ln q_i = \alpha' + \beta_1 \ln x_{i1} + \beta_2 \ln x_{i2} + \ldots + \beta_P \ln x_{iP} + \varepsilon_i \qquad (2)$$

* Professor and Director of the Centre for Managerial Economics and Econometrics, UFSIA, University of Antwerp, Belgium, Visiting Professor, College of Business Administration, University of South Carolina

** Research assistant, Centre for Managerial Economics and Econometrics, UFSIA, University of Antwerp, Belgium.

where $\alpha' = \ln \alpha$. Thus ease of estimation is provided. We may add that in an economics or a marketing context ease of interpretation can be seen as a third advantage in the sense that the parameters β_i are elasticities. A major drawback inherent to the multiplicative specification is its inability to deal with zero values of the x's or q's, since their logarithm is minus infinity. Yet zero values for either some observations of the x's or q's, or both are quite common in empirical studies. A number of procedures for resolving this difficulty have been proposed in the literature:

i) the most commonly used procedure is to replace a vector of observation $\underline{x}_p$ or $\underline{q}$ containing at least one zero element by, $\underline{x}_p + \underline{\iota}$, $\underline{q} + \underline{\iota}$ respectively, where $\underline{\iota}' = (1\ 1 \ldots\ldots 1)$. Or to quote from SNEDECOR and COCHRAN:

   "If some 0 values of x occur log(x+1) is often used" [8], p. 329. [1]

ii) In other cases all x_{ip} are left unchanged, except for zero values being replaced by ones.

Some authors explicitly acknowledge this data adjustment process, while others do not even bother to do so. The implication being that, one, it is common practice, and two, it is generally assumed that adding one to all or some elements of an observation vector does not substantially alter the model to be estimated.

iii) Discard those observations that contain zero values. YOUNG and YOUNG [11] examined the problem of zero values in the dependent variables and demonstrated that it is more appropriate to discard those observations rather than arbitrarily replacing the zeros by ones.
   BRANDT [1] observes that in his own work he does not limit the conclusion to the independent variable only. [2]

Some comments on these procedures are in order. The third one boils down to explicitly admitting the breakdown of the model for zero values. Thus, model (1) is in fact redefined for $x_{ip} > 0$, and $q_i > 0$. As a result the researcher will, for example, be unable to make statements about q_i when $x_{pi} = 0$. If he wants to do so, he will need a separate model to deal with just that case. A second consideration which is of much practical relevance is that after discarding all observations containing zeros, we may have very few observations left. For example, with the empirical data to be used in section 3, only seven out of twenty four observations would remain. In that study it will be justified to introduce one of the explanatory variables with a one period lag as well, in which case exactly one observation would remain! It thus becomes all the more necessary to have a model which can handle zero values properly. One possibility would of course be to take recourse to one of the other two adjustment procedures (i or ii). These will lead to almost identical estimates if non-zero values are much larger than one. Both procedures, however, suffer from a major drawback, namely, an arbitrary constant (in this case one) is added. This is hard to justify for the following reasons: Not only may the magnitude of the constant be important, but also its magnitude relative to that of the non-zero observations. As a result the scaling of the observations may be critical. Adding one to each observation is obviously different if the variable is measured in dollars, or thousands of dollars, or millions of dollars. A possible solution would then be to treat the constant itself as a parameter. We will examine specification and estimation issues of such procedure in section 2. Some empirical results will be presented and commented upon in section 3. We should add here that our analysis is limited to the explanatory variable only.

2. *Specification and estimation*

Replacing the elements of $\underline{x}_p$ by $x_{pi} + k_p$ leads to the following reformulation of model (1),

$$q_i = \alpha(k_1+x_{i1})^{\beta_1} (k_2+x_{i2})^{\beta_2} \ldots (k_p + x_{ip})^{\beta_p} e^{\varepsilon_i} \qquad (3)$$

where the k_p, rather than being arbitrarily predetermined con-

stants, become themselves model parameters. Some of these may be set equal to zero. Rather than saying that this will be the case for each p whose observation vector $\underline{x}_p$ does not contain any zero elements, it might be more appropriate to set those k_p equal to zero, that theoretically should be zero. To clarify this point let us consider the following example.
Suppose retail sales of a product (q_i) is related to advertising expenditures (x_{i1}) and distribution (x_{i2}), measured by the number of retail outlets carrying the product. The index i represents time. If interaction between marketing instruments is considered important, a specification as it would typically appear in the marketing literature is, [3]

$$q_i = \alpha \, x_{i1}^{\beta_1} \, x_{i2}^{\beta_2} \, e^{\varepsilon_i} \tag{4}$$

On examining this specification, we see that it is not entirely appropriate. With zero advertising expenditures we do not necessarily expect sales to be zero. So replacing x_{i1} by $x_{i1} + k_1$ seems indicated. As far as the distribution variable is concerned, it should be clear that with x_{i2} equal to zero, sales should be zero. Thus, an improved specification would be,

$$q_i = \alpha \, (x_{i1} + k_1)^{\beta_1} \, x_{i2}^{\beta_2} \, e^{\varepsilon_i} \tag{5}$$

Specification (5) is arrived at on the basis of prior knowledge, and is independent of whether or not $\underline{x}_1$ contains any zero elements. The reader may observe that one might get into difficulties when x_{i2} equals zero. From a practical point of view this will not be a problem, since one can hardly be expected to be interested in studying the sales of a product that is not for sale.

Returning to the general formulation (3), we note that the model will be invariant under a linear transformation (translation and rescaling) of the x_{pi}. That is, the model will yield the same estimates of the parameters β_p, and the same value of goodness of fit, whatever the units of measurements of the data. Two consequences of the introduction of the parameters k_p are of particular interest. First of all, the β_p can no longer be interpreted as elasticities. The elasticity of q_i with respect to x_{pi} indeed depends on x_{pi} itself , that is,

$$\eta_{q_i, x_{ip}} = \frac{(\delta q_i / q_i)}{\delta x_{ip} / x_{ip}} = \beta_p \frac{x_{ip}}{k_p + x_{ip}}$$

The second difficulty is the added complexity in estimation and testing, following from the fact that the introduction of the parameters k_p makes the model nonlinear in the parameters. Theoretically, however, least squares estimation will still yield maximum likelihood estimates, at least under the usual assumption of normally and independently distributed disturbance terms. [4]

A large number of nonlinear estimation or function minimization procedures are in use. The special structure of (3) points to simple direct search procedures such as golden section or FIBONACCI-search. The direct search need only be applied to the parameters k_p, since for given values of the latter, the remaining parameters can be estimated by ordinary least squares. Direct search procedures may, however, become quite cumbersome if the number of nonlinear parameters (k_p) is three or more. In that event nonlinear estimation or optimization methods may seem more appropriate. For the specific problem at hand, they have the disadvantage of not taking into account the fact that, for given values of the k_p, the problem becomes linear upon applying a simple logarithmic transformation. In our computational experience several nonlinear estimation or optimization methods had problems with convergence, and in one case, to be reported upon in section 3.1, none of those applied did converge. We would, therefore, propose to make use of the special structure of (7) in developing an appropriate estimation method. One possibility is outlined below. Let us rewrite (3) as,

$$y_i = \beta_0 + \beta_1 Z_{i1} + \ldots + \beta_p Z_{ip} + \ldots + \beta_P Z_{iP} + \varepsilon_i \qquad (7)$$

where $y_i = \ln q_i$, $\beta_0 = \ln \alpha$, and $Z_{ip} = \ln(k_p + x_{ip})$. Written in matrix notation (7) becomes,

$$\underline{y} = \underline{Z}\,\underline{\beta} + \underline{\varepsilon}$$

where $\underline{y}' = (y_1\ y_2\ \dots\ y_n)$, $\underline{\beta}' = (\beta_0\ \beta_1\ \dots\ \beta_p)$,
$\underline{\varepsilon}' = (\varepsilon_1\ \varepsilon_2\ \dots\ \varepsilon_n)$
and $\underline{Z} = (\underline{Z}_0\ \underline{Z}_1\ \dots\ Z_p)$
with $\underline{Z}_p' = (\overline{Z}_{ip} \dots Z_{np})$.

Conditional on $\underline{k}' = (k_1 \dots k_p)$ the estimation of $\underline{\beta}$ is the ordinary least squares estimator,

$$\hat{\underline{\beta}}_{\underline{k}} = (\underline{Z}'\underline{Z})^{-1}\ \underline{Z}'\ \underline{y} \tag{9}$$

The unconditional least squares objective is to minimize RSS with respect to $\underline{\beta}$ and $\underline{k}$,

$$\text{RSS} = (\underline{y} - \underline{Z}\ \underline{\beta})'\ (\underline{y} - \underline{Z}\ \underline{\beta}) \tag{10}$$

Substituting (9) for $\underline{\beta}$ in (10) we obtain,

$$\text{RSS} = \underline{y}'(\underline{I} - \underline{Z}(\underline{Z}'\underline{Z})^{-1}\ \underline{Z}')\underline{y} \tag{11}$$

which has to be minimized with respect to $\underline{k}$. Or since $\underline{y}'\underline{y}$ does not depend on $\underline{k}$, minimizing (10) with respect to $\underline{\beta}$ and $\underline{k}$ is equivalent to

$$\max_{\underline{k}}\ \underline{y}'\ \underline{Z}\ (\underline{Z}'\underline{Z})^{-1}\ \underline{Z}'\underline{y} \tag{12}$$

Optimizing (12) can be expected to be more efficient than direct optimization of (10). Computational experience to be reported upon in section 3 has indeed confirmed this expectation. The first order conditions of (12) are, as is shown in the appendix,

$$\hat{\underline{\varepsilon}}'\left(\frac{\delta\ \underline{Z}}{\delta\ k_p}\right)\hat{\underline{\beta}} = 0\ \forall\ p \tag{13}$$

The optimization algorithm can then be described as follows:

Step 1. Choose initial values for the elements of $\underline{k}$
Step 2. Calculate $\hat{\beta}$ from (9) and compute the corresponding residuals $\hat{\varepsilon}$
Step 3. Substitute $\hat{\beta}$ and $\hat{\varepsilon}$ in (13) and solve for $\underline{k}$ thus obtaining new estimates
Step 4. Return to step 2 unless the new estimates of $\underline{k}$ do not differ in absolute value from the previous ones by more than a predetermined small constant δ .

3. *Empirical results*

The data base used in the empirical study below consists of 36 monthly observations on sales (q_t) of a food product. The explanatory variables are advertising expenditures (a_t) in thousands of dollars, promotion expenditures (pr_t) in thousands of dollars, price of the product per pound (p_t) in cents, and a competitive price index (p_{ct}). The data base was split up in two subsets, an estimation sample consisting of the first 24 observations, a prediction sample made up of the remaining 12 . The reader is given an idea of the structure of the data base in table 1 (the full data base cannot be printed because of the proprietary character of the data). The table indicates that there are no advertising expenditures in just two months, whereas promotion is zero in 26 of the 36 observations.

Table 1.a : summary information on advertising and promotion data in the estimation sample

	Mean ($,000)	Number of zero observations	Number of observations with value < 20 (excluding 0)
advertising	87.25	2	2
promotion	60.75	17	0

Table 1.b : summary information on advertising and promotion data in the prediction sample

	Mean (\$,000)	Number of zero observations	Number of observations with value < 20 (excluding 0)
advertising	78.08	0	1
promotion	81.08	9	0

The table further indicates that advertising expenditures are sometimes quite low even when they are not zero. We should also mention that price and competitive price showed relatively little variation during the period under study and that in addition they were highly correlated (correlation coefficient of 0.86). We will therefore use $p_t^* = p_t/p_{ct}$ in those specifications containing price.
It is of course not our objective here to arrive at the best possible specification. We just want to use the data base to illustrate some of the points made and questions raised in section 1 and 2 . Section 3.1 will deal with estimation, whereas prediction will be discussed in section 3.2 .

3.1. Estimation

3.1.1. One k_p parameter

First we examine the case with one explanatory variable and one k_p parameter. We assume here that the only explanatory variable is advertising expenditures. Model (3) then becomes

$$q_t = \alpha\ (a_t + k_1)^{\beta_1}\ e^{\varepsilon_t} \tag{14}$$

Parameters α and β_1 , were estimated using ordinary least squares for a number of different values of k_1 . The corresponding adjusted coefficients of determination R^2 are shown in table 2 .

Table 2 : R^2 as a function of k_1 (equation (14))

k_1	R^2
0.001	0.1809
0.005	0.2423
0.01	0.2610
0.05	0.3171
0.10	0.3483
0.50	0.4421
1	0.4923
5	0.6233
10	0.6778
100	0.7703
150	0.7697
200	0.7675

The results indicate that, contrary to what seems to be common belief, R^2 is highly sensitive to what constant is added to a_t. The implication being that adding an arbitrary constant, such as one, to a_t is a dangerous procedure indeed. With $k_1 = 1$, $R^2 = 0.4929$, whereas applying the algorithm described in section 2 , the maximum value of R^2 is equal to 0.7705 with a corresponding estimate $\hat{k}_1 = 114.748$. It should further be noted that adding $k_1 = 0.001$ in our case corresponds to adding one to all observations in case advertising is measured in dollars rather than in thousands of dollars. Thus applying the common procedure of replacing a_t by $a_t + 1$ and with a_t measured in dollars would have resulted in an R^2 of 0.1809 as compared to 0.7705 obtained by applying the procedure proposed in this paper. The estimated parameters $\hat{\beta}_1$ will of course also be quite different. This is to be expected since $\hat{k}_1 = 114.748$ differs a lot from $\hat{k}_1 = 1$. But also in terms of interpretation, the results are quite different. A fair comparison seems to be the average advertising elasticities derived from these estimates.[5] Table 3 shows the results.

Table 3 : $\hat{\beta}_1$ and average advertising elasticity $\bar{\eta}_{q,a}$ (equation (14))

	$\hat{\beta}_1$	$\bar{\eta}_{q,a}$
$k_1 = 1$	0.124	.1226
$k_1 =$ 114.748	0.603	.2604

It should be clear that R^2 will not always vary so dramatically as in the example above. To show this, let us look at the following specifications,

$$q_t = \alpha(a_t+1)^{\beta_1}(pr_t+k_2)^{\beta_2}(pr_{t-1}+k_2)^{\beta_3} e^{\varepsilon_t} \qquad (15)$$

$$q_t = \alpha(a_t+1)^{\beta_1}(pr_t+k_2)^{\beta_2}(pr_{t-1}+k_2)^{\beta_3} e^{\beta_4 D_t} e^{\varepsilon_t} \qquad (16)$$

$$q_t = \alpha(a_t+1)^{\beta_1}(pr_t+k_2)^{\beta_2}(pr_{t-1}+k_2)^{\beta_3}(p_t^*)^{\beta_5} e^{\varepsilon_t} \qquad (17)$$

In the above equation pr_{t-1} represents lagged promotion. In promotion months consumers tend to buy more than they normally do. Sales in a post promotion month are therefore expected to be below their normal level. The expected sign of β_3 is negative. In equation (16) a dummy variable D_t is introduced. It takes on a value of zero except when both t and t-1 are promotion months. It is expected that β_4 will be negative. Finally, β_5 being the price parameter, it is also expected to be negative. Table 4 shows the estimated parameters $\hat{k}_2, \hat{\beta}_1 \ldots \hat{\beta}_5, R^2$ and the value of R^2 obtained with $k_2 = 1$. The latter is represented by R_i^2 . The table further indicates the number of iterations, n_i, needed to find the least squares estimates.
To examine stability of the estimates, each model was estimated first with 23 observations (n = 23) (not 24 because of the lagged promotion term), then 35 observations (n = 35). Goodness of fit improves but not by as much as in the first example.[6]

Table 4 : estimates of models (15), (16) and (17) with $n = 23$ and $n = 35$

Model	n	n_i	$\hat{k}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	R^2	R_i^2
(15)	23	13	561.28	.04	1.03	-.49	-	-	.861	.825
(15)	35	27	1135.85	.03	2.02	-.99	-	-	.912	.876
(16)	23	23	268.09	.04	.58	-.26	-.07	-	.857	.840
(16)	35	25	703.18	.03	1.34	-.64	-.05	-	.910	.887
(17)	23	20	425.50	.04	.81	-.39	-	-1.17	.866	.831
(17)	35	34	736.75	.03	1.36	-.69	-	- .99	.915	.881

Looking at the estimated coefficients it would appear that the advertising, price, and dummy variable coefficients show reasonable stability, when the last twelve observations are added, whereas the promotion related parameter estimates do not. This is partly due to the fact that the effect of higher values of $\hat{k}_2$ is compensated by having higher values (in absolute values) of $\hat{\beta}_2$ and $\hat{\beta}_3$. This is demonstrated by looking at the implied average promotion ($\bar{\eta}_{q,pr}$) and lagged promotion ($\bar{\eta}_{q,pr-}$) elasticities. As shown in table 5 these elasticities show great stability indeed. In fact taking a third decimal into account the six estimates of $\bar{\eta}_{q,pr}$ range from 0.103 to 0.106, those of $\bar{\eta}_{q,pr-}$ between 0.046 and 0.052 .

Table 5 : average promotion and lagged promotion elasticities in models (15), (16), (17)

Model	n	$\bar{\eta}_{q,pr}$	$\bar{\eta}_{q,pr-}$
(15)	23	.10	-.05
(15)	35	.10	-.05
(16)	23	.11	-.05
(16)	35	.11	-.05
(17)	23	.10	-.05
(17)	35	.10	-.05

3.1.2. Two or three k_p parameters

In models (15), (16) and (17) k_1 was arbitrarily set equal to one. In model (18) it becomes itself a parameter, thus making the number of k_p's equal to two:

$$q_t = \alpha(a_t+k_1)^{\beta_1}(pr_t+k_2)^{\beta_2}(pr_{t-1}+k_2)^{\beta_3}\ e^{\beta_4 D_t}\ e^{\varepsilon_t} \qquad (18)$$

Estimates for n = 23 are shown in table 6 , where n_i again represents the number of iterations the estimation procedure needed to converge, and R_i^2 is the R^2 corresponding to $k_p = 1$, all p . With k_1 set equal to one, and only k_2 a parameter R^2 changed 0.840 for $k_2 = 1$ to a maximum value of 0.857 .

Table 6 : estimates of models (18) and (19) with n = 23

Model	n_i	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	R^2	R_i^2
(18)	51	78.74	126.47	126.47	.25	.26	-.14	-.07	.902	.840
(19)	26	97.45	14.22	697.83	.29	.09	-.51	-.09	.904	.840

Letting k_1 be a parameter as well changes the maximum R^2 value to 0.902 . In models (15) to (18) the same constant k_2 is added to both promotion and lagged promotion. Theoretically these should be the same, since they involve the same basic variable. If we do not formally impose this as a constraint, model (19) is obtained:

$$q_t = \alpha(a_t+k_1)^{\beta_1}(pr_t+k_2)^{\beta_2}(pr_{t-1}+k_3)^{\beta_3}\ e^{\beta_4 D_t}\ e^{\varepsilon_t} \qquad (19)$$

We would then expect $\hat{k}_2$ and $\hat{k}_3$ not to be too different. This is, however, not at all the case as is seen from table 6 . Table 7 indicates that the implied average promotion and lagged promotion elasticities do not differ much between models (18) and (19). For values of promotion far from the average, however, differences may be substantial as shown in table 8 for the case of promotion.

Table 7 : average advertising, promotion and lagged promotion elasticities in models (18) and (19)

Model	$\bar{\eta}_{q,a}$	$\bar{\eta}_{q,pr}$	$\bar{\eta}_{q,pr-}$
(18)	.12	.08	-.05
(19)	.14	.07	-.04

Since theoretically k_2 should be equal to k_3 it is to be expected that the model (19) will not perform as well as (18) when applied to the prediction sample. This point will be further explored in section 3.2 .

Table 8 : $\eta_{q,pr}$ as a function of pr in models (18) and (19)

Promotion	$\eta_{q,pr}$	
	(18)	(19)
0	.00	.00
1	.00	.01
10	.02	.04
50	.07	.07
100	.11	.08
1000	.23	.09

We further observe that the goodness of fit of model (19) is about the same as that of (18). To get an idea of what the improvement would have been with advertising scaled in dollars, tens, or hundreds of dollars, model (19) was estimated for these three cases. The results presented in table 9 show once again that the improvement of fit over adding an arbitrary constant of one may be substantial.

Table 9 : R^2 for various values of k_1 , k_2 and k_3 in model (19)

Values of k	R^2
$k_1 = k_2 = k_3 = 0.001$	0.792
$k_1 = k_2 = k_3 = 0.01$	0.800
$k_1 = k_2 = k_3 = 0.1$	0.813
$k_1 = k_2 = k_3 = 1.0$	0.840
$k_1 = 97.45$; $k_2 = 14.22$ $k_3 = 697.83$	0.904

Finally it may be interesting to say a few words about our computional experience. Model (19) was estimated by an iterative method based on a first order TAYLOR expansion of the nonlinear model, [7] by the FLETCHER-POWELL method [2], and by FIACCO and MC CORMICK's Sequential Unconstrained Minimization Technique (SUMT) [3]. None of these algorithms converged in estimating model (19). Table 10 shows the estimates obtained by the FLETCHER-POWELL method after 200 iterations, with initial values of $k_p = 1$ for all p .
The table shows two sets of estimates. The first was obtained on an IBM 1130, the second on a large IBM 370. The R^2 was in both cases about 0.902 showing that the R^2 function is very flat flat for relatively large ranges of $\hat{k}_1$, $\hat{k}_2$ and $\hat{k}_3$. [8] The example clearly demonstrates the power of the special purpose algorithm based on the results derived in the appendix.

Table 10 : estimates obtained by applying the FLETCHER-POWELL method after 200 iterations

system	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_3$
IBM 1130	107.70	26.71	577.28
IBM 370	85.95	44.04	608.90

3.2. Prediction

In section 3.1 various aspects of estimation of model (3) were examined on the basis of empirical evidence. It may also be of interest to examine whether considering one or more of the k_p's as parameters leads to better predictions. In this section a number of models estimated on the basis of the first 24 (or 23) observations are applied to the prediction sample consisting of the last 12 observations. The quality of the predictions can be assessed by way of THEIL's inequality coefficient [9] pp. 32-48 ,

$$TH_1 = \frac{\sqrt{\sum_{i=25}^{36} (q_i - \hat{q}_i)^2}}{\sqrt{\sum_{i=25}^{36} q_i^2} + \sqrt{\sum_{i=25}^{36} \hat{q}_i^2}} \tag{20}$$

With $k_1 = 1$ in equation (14), $TH_1 = 0.1447$, whereas with k_1 equal to its optimal value of 114.748 , TH_1 is equal to 0.1252, or about 13.5 per cent less. [9]
This is of course no guarantee that $k_1 = 114.748$ results in the best predictions. To examine this somewhat more closely, TH_1 was calculated for values of k_1 ranging from 2 to 120 in steps of 2 . Some selected values are shown in table 11 . From these result appears that TH_1 is minimal for $k_1 = 116$, quite close to the least squares estimate indeed. Table 11 also contains the values of TH_1 corresponding to values of k_1 equal to 0.001 , 0.01 , 0.1 and 1 , this to make evaluation of the scaling effect possible.

Table 11 : THEIL's inequality coefficient (TH_1) for selected values of k_1 (model (14))

TH_1	.1578	.1557	.1526	.1447	.1423	.1351	.1279	.1254	.1253
k_1	0.001	0.01	0.1	1	2	10	50	100	110

TH_1	.1252	.1252	.1250	.1251
k_1	112	114	116	118

Disadvantage of the definition of the inequality coefficient TH_1 is that the denominator contains $\Sigma\hat{q}_i^2$, which of course varies with the model specification and parameter constraints. It may therefore be preferable to assess predictive ability by the following variant of the inequality coefficient, which has the advantage of having a denominator with constant value, that is, independent of specification and parameter constraints.

$$TH_2 = \frac{\sqrt{\sum_{i=25}^{36} (q_i - \hat{q}_i)^2}}{\sqrt{\sum_{i=25}^{36} q_i^2}} \qquad (21)$$

With $k_1 = 1$, $TH_2 = 0.2781$; with $k_1 = 114.748$, $TH_2 = 0.2394$, or 14 per cent less. TH_2 was also computed for values of k_1 ranging from 2 to 200 in steps of 2 . Here TH_2 is minimal for $k_1 = 200$ and equal to 0.2367 , only marginally less than the value of TH_2 corresponding to $k_1 = 114.748$. That this is close to the minimum value is indicated by the fact that around $k_1 = 200$, ΔTH_2 is about 0.00002 for $\Delta k_1 = 2$.

Finally we come to prediction with models involving more than one k_p parameter. To that extent, models (18) and (19) are compared with model (16). Model (18) differs from (16) to the extent that in the former k_1 is a parameter, whereas in the latter it is arbitrarily set equal to one. We therefore expect better predictions with model (18) than with (16). In model (19) the condition $k_2 = k_3$ is not imposed in estimating the parameters. Since, however, theoretically k_2 and k_3 must be equal, worse predictions are expected with model (19). These various expectations are confirmed by the results in table 12 .

Table 12 : THEIL's inequality coefficient (TH_2) for $k_i = 1$ and for $k_i = k_i^*$ (least squares estimates) in models (16), (18) and (19).

Model	TH_2	
	$k_i = 1$	$k_i = k_i^*$
(16)	0.0899	0.0591
(18)	0.0899	0.0544
(19)	0.0899	0.0935

4. *Conclusions*

In this paper we have reviewed three commonly used procedures for dealing with the problem of zero values in multiplicative models. This led us to propose a fourth procedure, applicable in cases where zero values appear only in the explanatory variables. The merits of the procedure were illustrated with an empirical example. At a minimum the results indicate that adding an arbitrary constant of one may be inappropriate, since goodness of fit, estimates and predictions appear to be quite sensitive to which constant is added. At the same time this paper provides the basis for developing efficient algorithms to implement the procedure. Further empirical studies are of course needed to fully appreciate the potential of the proposed method.

Appendix

Optimal values of the k_p are obtained by differentiating equation (12) with respect to k_p , for $p = 1,\ldots,P$, and setting equal to zero, that is,

$$\frac{\delta}{\delta k_p} (\underline{y}' \, \underline{Z} \, (\underline{Z}' \, \underline{Z})^{-1} \, \underline{Z}' \, \underline{y}) = 0 \qquad (A.1)$$

which since the k_p's appear as parts of the elements of the $\underline{Z}$ matrix, can be rewritten as[10]

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \frac{\delta}{\delta z_{ij}} (\underline{y}' \, \underline{Z} \, (\underline{Z}' \, \underline{Z})^{-1} \, \underline{Z}' \, \underline{y}) \frac{\delta z_{ij}}{\delta k_p} = 0 \qquad (A.2)$$

Equation (A.2) can also be written as,

$$\mathrm{tr} \left\{ \left[\frac{\delta(\underline{y}' \, \underline{Z} \, (\underline{Z}'\underline{Z})^{-1} \, \underline{Z}' \, \underline{y})}{\delta \, \underline{Z}} \right] \left[\frac{\delta \, \underline{Z}}{\delta k_p} \right] \right\} = 0 \qquad (A.3)$$

where $\frac{\delta}{\delta z_{ij}} (\underline{y}' \, \underline{Z} \, (\underline{Z}'\underline{Z})^{-1} \, \underline{Z}' \, y)$ is the $(i,j)^{th}$ element of $\frac{\delta(\underline{y}' \, \underline{Z} \, (\underline{Z}'\underline{Z})^{-1} \, \underline{Z} \, \underline{y})}{\delta \, \underline{Z}}$, and $\frac{\delta z_{ij}}{\delta k_p}$ is the $(i,j)^{th}$ element of $\frac{\delta \, \underline{Z}}{\delta k_p}$.

Applying the chain rule of differentiation we can write,[11]

$$\frac{\delta}{\delta z_{ij}} (\underline{y}' \, \underline{Z} \, (\underline{Z}'\underline{Z})^{-1} \, \underline{Z}' \, \underline{y}) =$$

$$\underline{y}' \left\{ \left[\frac{\delta \, \underline{Z}}{\delta z_{ij}} \right] (\underline{Z}'\underline{Z})^{-1} \, \underline{Z}' + \underline{Z} \left[\frac{\delta (\underline{Z}'\underline{Z})^{-1}}{\delta z_{ij}} \right] \underline{Z}' + \underline{Z} \, (\underline{Z}'\underline{Z})^{-1} \left[\frac{\delta \underline{Z}}{\delta z_{ij}} \right]' \right\} \underline{y}$$

$$(A.4)$$

In the above expression $\frac{\delta \, \underline{Z}}{\delta z_{ij}}$ is a n x m matrix with all elements except the $(i,j)^{th}$ equal to zero, the latter being equal to one. We can rewrite this as

$$\frac{\delta \underline{Z}}{\delta z_{ij}} = \underline{e}_i \, \underline{u}_j' \tag{A.5}$$

where, $\underline{e}_i$ is a n x 1 unit vector, that is, its i^{th} element is one, all others are zero, and $\underline{u}_j$, a m x 1 unit vector, with its j^{th} element equal to one. Using (A.5) and realizing that the first and third terms in (A.4) are the same, the latter can be rewritten as,

$$2\, \underline{y}'\underline{e}_i\underline{u}_j' \, (\underline{Z}'\underline{Z})^{-1} \, \underline{Z}' \, \underline{y} + \underline{y}' \, \underline{Z} \left[\frac{\delta(\underline{Z}'\underline{Z})^{-1}}{\delta z_{ij}}\right] \underline{Z}' \, \underline{y} \tag{A.6}$$

In the above the derivative of $(\underline{Z}'\underline{Z})^{-1}$ with respect to z_{ij} can be written as [12]

$$\frac{\delta(\underline{Z}'\underline{Z})^{-1}}{\delta z_{ij}} = -\, (\underline{Z}'\underline{Z})^{-1} \frac{\delta(\underline{Z}'\underline{Z})}{\delta z_{ij}} (\underline{Z}'\underline{Z})^{-1} \tag{A.7}$$

and $$\frac{\delta(\underline{Z}'\underline{Z})}{\delta z_{ij}} = \left[\underline{z}^i \, \underline{u}_j' + \underline{u}_j \, \underline{z}^{i'}\right] \tag{A.8}$$

where $\underline{z}^i = [z_{i1} \; z_{i2} \; \cdots \; z_{im}]'$

Using (A.7) and (A.8) and after combining terms, (A.6) becomes

$$2\, \underline{y}'\underline{e}_i\underline{u}_j' \, (\underline{Z}'\underline{Z})^{-1} \, \underline{Z}' \, \underline{y} - 2\, \underline{y}' \, \underline{Z} \, (\underline{Z}'\underline{Z})^{-1} \, \underline{z}^i \, \underline{u}_j' (\underline{Z}'\underline{Z})^{-1} \, \underline{Z}' \, \underline{y} \tag{A.9}$$

Since, $\underline{z}^i = \underline{Z}'\underline{e}_i$, and making use of the fact that both terms in (A.9) are scalars, (A.9) reduces to,

$$2\, \underline{e}_i' \, \underline{y} \, \underline{y}' \, \underline{Z} \, (\underline{Z}'\underline{Z})^{-1} \, \underline{u}_j - 2\, \underline{e}_i' \, \underline{Z} \, (\underline{Z}'\underline{Z})^{-1} \, \underline{Z}' \, \underline{y} \, \underline{y}' \, \underline{Z} (\underline{Z}'\underline{Z})^{-1} \underline{u}_j$$

$$= 2\ \underline{e}_i'\ \{\ \left[\underline{y} - \underline{Z}\ (\underline{Z}'\underline{Z})^{-1}\ \underline{Z}'\ \underline{y}\right]\underline{y}'\ \underline{Z}\ (\underline{Z}'\underline{Z})^{-1}\}\ \underline{u}_j$$

$$= 2\ \underline{e}_i'\ \hat{\underline{\varepsilon}}\ \hat{\underline{\beta}}'\ \underline{u}_j \tag{A.10}$$

Since (A.10) is the $(i,j)^{th}$ element of the matrix $\dfrac{\delta(\underline{y}'\ \underline{Z}\ (\underline{Z}'\underline{Z})^{-1}\ \underline{Z}'\ \underline{y})}{\delta\ \underline{Z}}$ the latter must be equal to $2\ \hat{\underline{\varepsilon}}\ \hat{\underline{\beta}}'$.

Using this result in (A.3) we finally obtain,

$$\frac{\delta}{\delta k_p}\ (\underline{y}'\ \underline{Z}\ (\underline{Z}'\underline{Z})^{-1}\ \underline{Z}'\underline{y}) = 2\ \mathrm{tr}\ (\hat{\underline{\beta}}\ \hat{\underline{\varepsilon}}'\frac{\delta\underline{Z}}{\delta k_p}) = 0$$

or the first order optimality conditions are,

$$\hat{\underline{\varepsilon}}'\ \frac{\delta\underline{Z}}{\delta k_p}\ \hat{\underline{\beta}} = 0, \quad p = 1,\ldots,P$$

Footnotes

1 That this is often applied in empirical work is exemplified by a study by HOUSTON and WEISS [5] p. 153.

2 This is not exactly what he writes. In fact he states: "However, I do not limit this [YOUNG and YOUNG's] conclusion to the independent variable" [1] p. 101. It is quite clear from the context, however, that he discards observations for zero values for the dependent and independent variables.

3 See, for example, NAERT and LEEFLANG [7] Sections 5.313 and 6.3 .

4 See, for example, GOLDFELD and QUANDT [4] pp. 57-58.

5 By average advertising elasticity we mean the elasticity corresponding to the average level of advertising expenditures as computed from the data.

6 Had advertising been measured in dollars, the change would have been more substantial. This is to be expected given the results obtained in table 3 , and will be illustrated again in discussing the case with three k_p parameters.

7 See MARQUARDT [6] .

8 This may also explain why adding or deleting a few observations may significantly alter the estimates, as was evident from the results in table 4 .

9 It is interesting to observe that this result materializes although there are no zero value observations for advertising in the prediction period.

10 Since in some cases the same k_p may be associated with different x's , a distinction is made here between the index representing the variable ($j = 1,\ldots,m$), and the index for the parameters k_p ($p = 1,\ldots,P$). As such the notation slightly deviates from that in the main body of the paper, making the results somewhat more general.

11 For some basic results of vector and matrix differentiation see, for example, THEIL [10] pp. 30-33 .

12 See THEIL [10] p. 33 .

References

[1] A.E. BRANDT, letter to the Editor, The American Statistician, Vol. 30, N° 2, May 1976, p. 101.

[2] R. FLETCHER and M.J.D. POWELL, "A Rapidly Convergent Descent Method for Minimization", Computation Journal, Vol 6, 1963, pp. 163-168.

[3] A.V. FIACCO and G.P. MC CORMICK, Nonlinear Programming: Sequential Unconstrained Minimization Technique, New York, John Wiley and Sons, 1968.

[4] S.M. GOLDFELD and R.E. QUANDT, Nonlinear Methods in Econometrics, Amsterdam, North-Holland Publishing Company, 1972.

[5] F.S. HOUSTON and D.L. WEISS, "An Analysis of Competitive Market Behavior", Journal of Marketing Research, Vol. II, pp. 151-155, 1974.

[6] E.W. MARQUARDT, "An algorithm for Least Squares Estimation of Nonlinear Parameters", Journal of the Society of Industrial Mathematics, Vol. 4 , pp. 431-441, 1963.

[7] P.A. NAERT and P.S.H. LEEFLANG, Building Implementable Marketing Models, Boston, Martinus Nijhoff Social Sciences Division, 1978.

[8] G.W. SNEDECOR and W.G. COCHRAN, Statistical Methods, (6th ed.), Ames, Iowa, Iowa State University Press, 1967.

[9] H. THEIL, Economic Forecast and Policy, Amsterdam, North-Holland Publishing Company, 1965.

[10] H. THEIL, Principles of Econometrics, New York, John Wiley and Sons, Inc. 1971 .

[11] K.H. YOUNG and L.Y. YOUNG, "Estimations of Regressions Involving Logarithmic Transformation of Zero Values in the Dependent Variable", The American Statistician, Vol. 29, N° 3, August 1975, pp. 118-120.