Marketing Analytics, Inc.
500 Davis Street, Suite 1010
Evanston, Illinois 60201
Tel: (847) 425-9999
Fax: (847) 425-1332
www.marketinganalytics.com

March 15, 2004

Subject: **Price Gap Ratios**

This provides a summary of Marketing Analytics' position on the use of price gap ratios in sales response models.

Background: Some modeling vendors use price gap ratios as independent variables in sales response models. That is, instead of simply putting competitive price in the model, for each major competitor, they put the ratio of competitive price divided by own price. This white paper looks at this practice to determine the potential benefits and/or risks.

Summary of Findings and Conclusions: Under normal conditions, this practice is (exactly) mathematically equivalent to the more straightforward practice of putting own price and competitive prices in the model separately. In terms of final results, it neither helps nor hurts: it's just a more complicated way of doing the same thing. However, in the degenerate case where the competitive price is constant over the modeling period, own price and gap ratio are 100% correlated and the gap ratio model can lead to unpredictable and potentially wrong results. As this practice has no benefits and does nothing other than add complexity and risk, we do not recommend it.

Details:

The most common and straightforward way of modeling own and competitive price is:

$$\text{Log}(U_1) = \beta_0 + \beta_1 \text{Log}(P_1) + \beta_2 \text{Log}(P_2)$$

…where U_1 = Own units, P_1 = Own price, P_2 = Competitive price

The proposed alternative is to replace competitive price with the ratio of competitive price to own price:

$$\text{Log}(U_1) = \beta_3 + \beta_4 \text{Log}(P_1) + \beta_5 \text{Log}(P_2 / P_1)$$

This can be rewritten as:

$$\text{Log}(U_1) = \beta_3 + \beta_4 \text{Log}(P_1) + \beta_5 \text{Log}(P_2) - \beta_5 \text{Log}(P_1)$$
$$\text{Log}(U_1) = \beta_3 + (\beta_4 - \beta_5) \text{Log}(P_1) + \beta_5 \text{Log}(P_2)$$

After remapping coefficients, this is exactly equivalent to the more straightforward specification, where:

$\beta_0 = \beta_3$ $\beta_1 = \beta_4 - \beta_5$ $\beta_2 = \beta_5$

To test this, we generated synthetic data under the following assumptions:

- 1 target SKU with 1 competitor, 20 stores, 104 weeks
- Own elasticity = -1.5, cross-elasticity = 0.5
- Prices constant at regular price, except for discounts that occur randomly with 25% probability. Discount depth is random, uniformly distributed between 10% and 40%. This price pattern occurs for both own and competitive price.
- Simulation was repeated for varying correlations between own and competitive price, from 0% correlation to 100% correlation.
- Random noise added to unit sales of ± 5%

As long as there was some variation in competitive price, both models yielded exactly the same (correct) coefficients after remapping.

Standard model spec:

Model with Price and Competitor Price 386
Prob{Comp Cut ¦ Own Cut} = 0.25 17:12 Friday, March 12, 2004

The REG Procedure
Model: MODEL1
Dependent Variable: LogUnits

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	100.92248	50.46124	80036.8	<.0001
Error	2077	1.30950	0.00063048		
Corrected Total	2079	102.23198			

Root MSE	0.02511	R-Square	0.9872
Dependent Mean	4.67498	Adj R-Sq	0.9872
Coeff Var	0.53710		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	4.60444	0.00068626	6709.49	<.0001
LogP	1	-1.49880	0.00396	-378.56	<.0001
LogCompP	1	0.49759	0.00393	126.71	<.0001

Gap ratio:

Model with Price and Gap Ratio 387
Prob{Comp Cut ¦ Own Cut} = 0.25 17:12 Friday, March 12, 2004

The REG Procedure
Model: MODEL1
Dependent Variable: LogUnits

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	100.92248	50.46124	80036.8	<.0001
Error	2077	1.30950	0.00063048		
Corrected Total	2079	102.23198			

Root MSE	0.02511	R-Square	0.9872
Dependent Mean	4.67498	Adj R-Sq	0.9872
Coeff Var	0.53710		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	4.60444	0.00068626	6709.49	<.0001
LogRatio	1	0.49759	0.00393	126.71	<.0001
LogP	1	-1.00121	0.00560	-178.89	<.0001

Standard model spec: $\beta_0 = 4.60444$ $\beta_1 = -1.49880$ $\beta_2 = 0.49759$

Gap ratio: $\beta_3 = 4.60444$ $\beta_4 = -1.00121$ $\beta_5 = 0.49759$

$\beta_0 = \beta_3$ ✓

$\beta_4 - \beta_5 = -1.00121 – 0.49759 = -1.49880 = \beta_1$ ✓

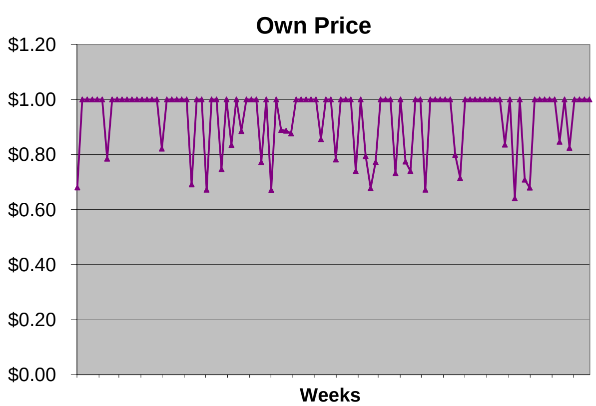
$\beta_2 = \beta_5$ ✓

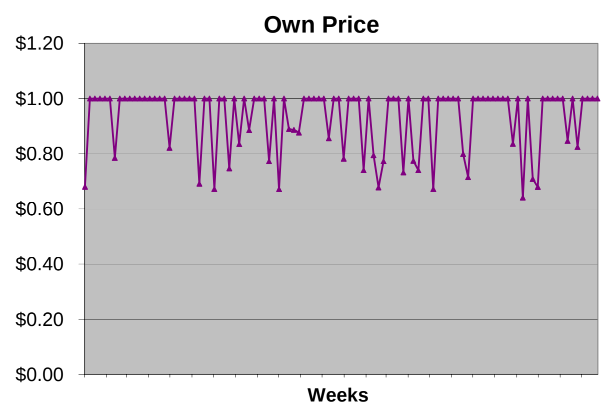
These relationships held exactly across all assumptions about price variation, correlations, discount depth, and noise in data with one exception: if there was absolutely zero variation in competitive price, the gap ratio model could yield potentially unpredictable results.

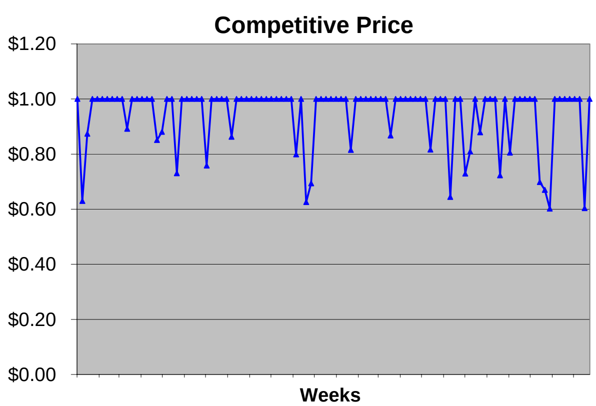
Normal Situation (Some Variation in Competitive Price)

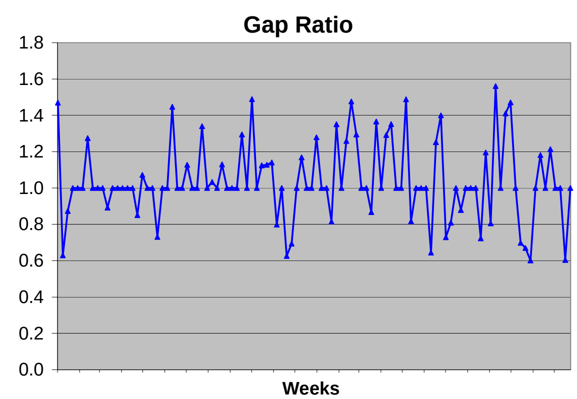
Standard Model Variables

Gap Ratio Model Variables









Under normal circumstances (above), the standard model has what are arguably more intuitive looking variables, but regardless, both models produce identical and correct coefficients.

Degenerate Situation (Zero Variation in Competitive Price)

Standard Model Variables

Gap Ratio Model Variables

In the degenerate case where competitive price does not change (above), the standard model is presented with one acceptable variable (own price) and one price with no variation (competitive price). All standard statistics routines handle this similarly: competitive price coefficient is set to zero and own price gets a reasonable estimate. While we know zero is not right, it is right for the data we have: that competitor had no impact on sales of the target product.

How the Gap Ratio model handles this case is somewhat unpredictable. If competitive price is constant, own price and gap ratio are mirrors of each other (see above). The statistics package is indifferent as to which gets a coefficient and which gets set to zero. In SAS, if you list own price first, it gets the coefficient and gap ratio is zero. But if you list the gap variable first, it gets the (wrong) coefficient, as below:

```
                    Model with Price and Gap Ratio                                  438
                    Prob{Comp Cut | Own Cut} = 0.25        17:12 Friday, March 12, 2004

                              The REG Procedure
                                Model: MODEL1
                        Dependent Variable: LogUnits

                            Analysis of Variance

                                       Sum of          Mean
Source                     DF         Squares        Square    F Value    Pr > F

Model                       1        91.67001      91.67001     108462    <.0001
Error                    2078         1.75628    0.00084518
Corrected Total          2079        93.42629

            Root MSE              0.02907    R-Square     0.9812
            Dependent Mean        4.71584    Adj R-Sq     0.9812
            Coeff Var             0.61647

NOTE: Model is not full rank. Least-squares solutions for the parameters are not unique. Some
      statistics will be misleading. A reported DF of 0 or B means that the estimate is biased.
NOTE: The following parameters have been set to 0, since the variables are a linear combination
      of other variables as shown.

                             LogP =  -LogRatio

                            Parameter Estimates

                          Parameter       Standard
    Variable      DF       Estimate          Error    t Value    Pr > |t|

    Intercept      1        4.60578     0.00071974    6399.26      <.0001
    LogRatio       B        1.49683        0.00455     329.34      <.0001
    LogP           0              0              .          .           .
```

In this case:

Gap ratio: β_3 (Intercept) = 4.60578 β_4 (LogP) = 0
β_5 (LogRatio) = 1.49683

Remapping…

Intercept: $\beta_0 = \beta_3$ ✓

Own Price Elasticity: $\beta_4 - \beta_5$ = 0 – 1.49683 = -1.49683 = β_1 ✓

Cross Price Elasticity: $\beta_2 = \beta_5$ = 1.49683 **X**

So in the degenerate case, cross price elasticity can get set to the negative of own price elasticity, which is wrong. With multiple competitors in the model, the degenerate case gets even less predictable.

Bottom line: It is not a good practice to have model results dependent on vagaries of the statistical package and the order you list variables in. As this modeling technique provides zero benefits but adds complexity and risk, we do not recommend it.

Ross-boy Link
Marketing Analytics Inc.